TSG-RAN Working Group 3 meeting#7 **_TSGR3#6(99)c59_**
Sophia Antipolis, September 20-24, 1999

Agenda Item: 6.5, 19.1

Source: Alcatel

Title: Need for priority handling over Iub/Iur interfaces.

Document for: Decision

---

1. Introduction

This contribution is a complementary contribution to tdoc R3-99949 that was submitted at the RAN3#6 meeting at Sophia Antipolis. It gives more explanations on simulation results. It shows that not handling priorities at ATM or AAL2 layer leads to a huge waste of bandwidth on Iub and Iur interfaces, which is due to the mixing of real-time and non real-time user data in a single queue.

The objective delay figures proposed by Siemens in tdoc R3-99955 [6] are shown in the following table:

	8kbit/s	32kbit/s	64kbit/s	144kbit/s	384kbit/s	2048kbit/s
TN1	1msec	120msec	185msec	205msec	450msec	

These figures correspond to a single type of traffic (no priority handling). If we mix real time and non real time traffic in a single queue, then both types of traffic will be delayed by about the same amount of time. Then, if we mix a 8 kbit/s speech service with a 144 kbit/s data service, we can expect a transfer delay for the speech of about the same order of magnitude as for nrt data. In order to reach 5 ms delay for speech without handling priorities, we need less load on the ATM link, i.e. more bandwidth.
For example, simulations have shown that a traffic of only one 12.2 kbit/s speech call and one 144kbps data call at the same time requires 2.8 Mbit/s bandwidth, i.e. more than a 2 Mbit/s link when there is no priority handling (assuming transfer delay for speech data is 5 ms over Iub).
Results found by Alcatel simulations are a little bit different from Siemens results because, for example, the distribution is not the same, but it shows the same kind of issue.

Alcatel thinks that not handling priorities at AAL2 or ATM layer is not acceptable for operators, in particular over the Iub interfaces. Alcatel simulations also show that handling priorities at AAL2 layer is a little more beneficial than handling them at ATM layer.

2. Discussion

2.1. Generalities

Asynchronous Transfer Mode (ATM) and ATM Adaptation Layer type 2 (AAL2) are used as a transport layer for data streams on Iur and Iub interfaces on the user plane.

AAL2 is designed for "the bandwidth-efficient transmission of **low-rate**, **short**, variable length packets **in delay sensitive applications"** (see [1]).

Nevertheless, on the Iub and Iur interfaces, AAL2 is used for all circuit-switched and packet-switched user data. Thus AAL2 will transmit **possibly high rate** (up to 2 Mbit/s), **possibly long** (e.g. when the transmission time interval is 80 ms), variable length packets of **applications requiring different types of quality of services**.

TS22.105 ([2]) defines 4 traffic classes, which have different delay characteristics (conversational, interactive, streaming and background). In TR 23.907, traffic priority handling is mentioned for interactive and background classes.

The question **is how to handle these different traffic classes at AAL2 and ATM layers** on the Iub and Iur interfaces.

Two approaches to multiplex AAL2 channels on the Iub/Iur interfaces can be foreseen.

- First approach : No distinction of Priority is made at AAL2 and ATM layers
- Second approach : Priority handling at AAL2 or ATM layer

Simulations have been launched to compare these two approaches with two Priority classes.

2.2. Traffic profile

In the simulations, real time and non real time AAL2 connections are multiplexed on the Iub interface.

2.2.1. Real time sources (speech type)

Real time sources are of speech type.
The AAL2 SDU length is **33 octets**, taking into account the length of a radio frame of 20 ms (Transmission Time Interval) with a variable bit rate due to the AMR, plus octets for quality estimation.
The arrival process of AAL2 SDU is pseudo-periodic, with period of 20 ms (see Figure 1).

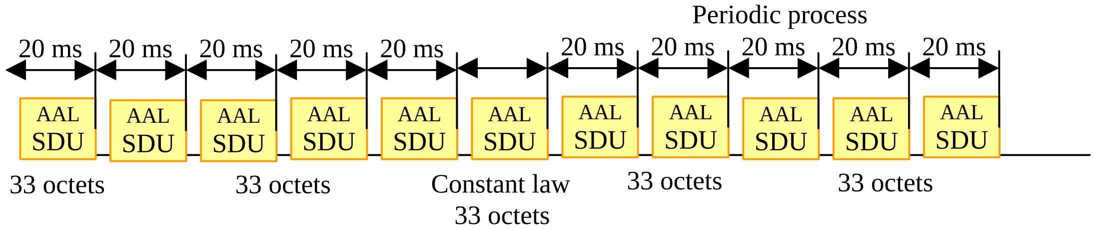

Figure 1: Speech Source

The number of speech sources varies according to the load.

2.2.2. Non real time sources (data type)

Non real time sources are data sources with cut-off Pareto arrival process, according to the model defined in [5].
The peak rate of the AAL2 connection is 144 kbit/s.
The Transmission Time Interval is 80 ms. Thus the maximum length of an AAL2 SDU is 1440 octets (144 kbit/s x 80 ms / 8 octets). This corresponds to 32 CPS Packets of 45 octets, packed into al least 33 ATM cells.
Moreover, the minimum length of an AAL2 SDU is 235 octets. The parameter of the Pareto distribution is 1.1.

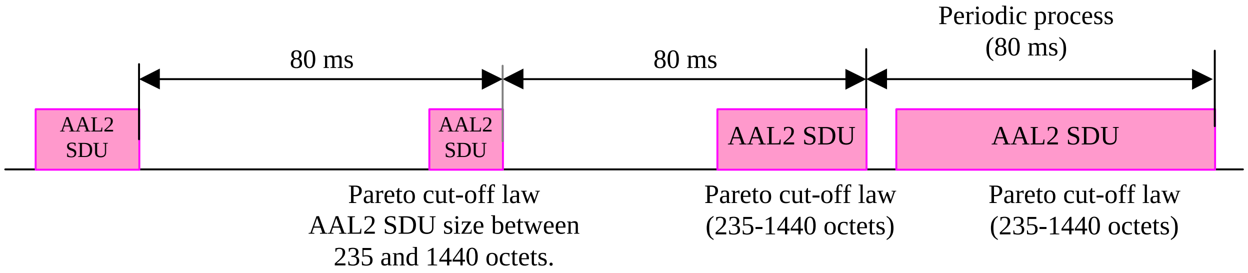

Figure 2: Data Source

The number of data sources varies according to the global load.

2.3. Systems

Three systems corresponding to the two approaches have been compared:

- First system : No distinction of Priority is made at AAL2 and ATM layers
- Second system : Priority handling at ATM layer
- Third system : Priority handling at AAL2 layer

2.3.1. First system and approach : No distinction of Priority is made at AAL2 and ATM layers

2.3.1.1. Principle

All AAL2 channels are handled in the same way regardless of the Priority requirements of upper applications.
Priority requirements at AAL2/ATM layers are not distinguished.

Thus, all AAL2 channels will be transported with the best quality of service on the Iub and Iur interfaces. They will be transported as delay sensitive connections requiring a low loss probability, even if the applications are non real time or best effort.

2.3.1.2. Simple Example

Figure 3 illustrates this approach. It depicts information to be sent on the Iub interface and on the radio interface when one speech source and one data source (described in section 2.2) are multiplexed.

Let us consider the data source.
Every 80 ms, an AAL SDU corresponding to 8 radio frames is sent on the Iub interface.
The maximum AAL SDU size is 32 CPS Packets. Thus up to 33 ATM cells should be sent every 80 ms to transmit this AAL SDU on the Iub interface.
If the **allowed transmission time on the Iub interface is short** (for example 5 ms), the required instantaneous bit rate on the Iub interface will be high (in the same example, at least 424*33 / 5 = 2.8 Mbit/s).
When the Node B receives this AAL SDU, it will transmit the unpacked data on the radio interface but the transmission time on the radio interface is 80 ms. Thus, the transmission of the same data on the radio interface requires an instantaneous bit rate of only 144 kbit/s.

If no priority is handled at ATM or AAL2 level, all AAL2 channels shall be transported with the most stringent transfer delay (transfer delay of real time services). And the required instantaneous bit rate on the Iub interface shall be high.

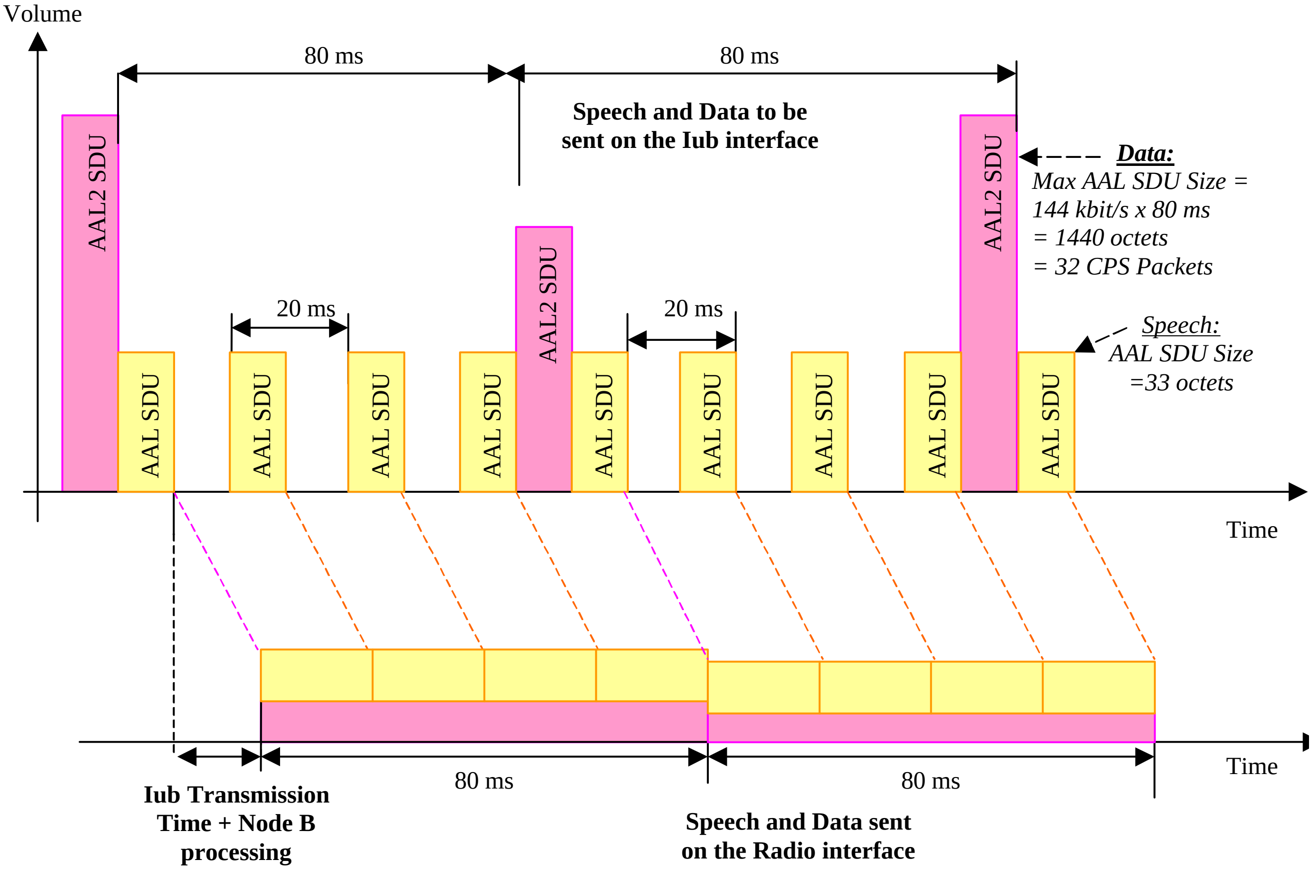

Figure 3: Multiplexing of real time and Non real time sources

2.3.1.3. Conclusion

This approach is simple and in line with [1].

Nevertheless, with this approach, **no statistical multiplexing** can be done at AAL2 and at ATM level. **The required ATM bandwidth on the Iub and Iur interface is larger** than if some statistical multiplexing is done.

Transmission resource inside the UTRAN should be optimized. Decreasing the ATM required bandwidth for the same net bit rate could allow an operator to decrease its transmission and operating cost.
This major drawback leads to a second approach: Priority requirements of upper applications are handled at AAL2 or ATM layer.

2.3.2. Second Approach : Priority handling at AAL2 or ATM layer

Priority is handled at AAL2 or ATM level. Priority classes should be defined at AAL2 level in order to distinguish Priority requirements of every AAL2 connection.

Two ways to handle Priority at AAL2 or ATM level can be foreseen.
1. Priority handled at AAL level: AAL2 channels with the different Priority requirements are multiplexed in the same VCC. VCC will be transported with the best quality of service inside ATM network.
2. Priority handled at ATM level: A VC connection is dedicated to one Priority. AAL2 channels with the same Priority requirements are multiplexed in the same VCC.

Statistical multiplexing gain is possible.
The required bandwidth of the VCC carrying AAL2 channels is lower than if no statistical multiplexing is done. This is shown in the simulation results below.

2.4. Simulation results

Simulations have been launched to compare the three following systems.
- No priority handling
- Priority handled at ATM layer
- Priority handled at AAL2 layer

When Priority is not handled or handled at AAL2 layer, the service category of the VC connection is Constant Bit Rate. The Peak Cell Rate of this VCC is 2 Mbit/s. A 2 Mbit/s VCC is chosen, as a representative example of an Iub interface link.
When Priority is handled at ATM layer, two VCC are needed (one for each Priority). The sum of the Peak Cell Rate of these two VCC is 2 Mbit/s.

The percentage of real time traffic is 25%. A set of simulations has been run with the global load varying between 0.05 and 0.9. The global load is defined as the ratio between the actual resulting bit rate of all established connections and the bit rate of the VCC. The global load is a fundamental parameter, as it roughly represents the efficiency of the transmission.

The maximum delay is supposed to be the 10^{-5} quantile of the delay distribution (more precisely, the probability that the transfer delay is greater than the max. is less than 10^{-5}). The delay is the one introduced in the packing queue between AAL2 and ATM layer.

In ATM networks, a load of **0.8** or even higher is often considered as a reasonable objective for peak hours.

The results of the simulations show that even if the load is only equal to **0.6**, the goal of 2 ms is obtained with AAL2 priorities. A goal of 5 ms is obtained with ATM priorities.

	No priority Handling	Priority handled at ATM Layer	Priority handled at AAL2 Layer
Mean delay of real time octets	2.4 ms	0.8 ms	0.6 ms
Mean delay of non real time octets	4.5 ms	6 ms	5 ms
Max delay of real time octets	**29 ms**	**5 ms**	**1.3 ms**
Max delay of non real time octets	**30 ms**	**49 ms**	**37 ms**

Table 1: Mean and Max delay – Load = 0.6

Simulations show that if no priority is handled, real time AAL2 connections encounter unacceptable delays (between 7 and 48 ms if load varies between 0.05 and 0.9).
The 7 ms figure can be simply explained as follows. The maximum AAL SDU size is 1440 octets, packed into 33 ATM cells. If a speech frame arrives just after an AAL SDU of maximum size has been queued on the VCC, it will have to wait 33 cell time units before starting to be transmitted. 33 cell time at 2 Mbit/s corresponds to about 7 ms. The probability of this event is high enough to hit the 10^{-5} quantile of the delay distribution of the speech connection.
Thus, it is not possible to accept both one data connection following the Pareto distribution above (144 kbit/s Peak rate) and one speech connection on a 2 Mbit/s VCC, while keeping a transmission delay over the Iub interface less than 7 ms for the speech connection, if the priority is ignored at ATM and AAL level.

Even if the maximum allowed delay for speech is 20 ms, the achievable load is 0.3 if no Priority is handled (against 0.9 if Priority is handled at AAL2 or ATM level).

If Priority is handled at ATM level, the achievable load is **0.6** if the delay encountered by non real time octets should be less than 50 ms and if the delay encountered by real time octets should be less than 5 ms.
If Priority is handled at AAL2 level, the achievable load is even better with the same performance objectives.

These results have shown that the approach without priority handling is not acceptable, and that a priority handling at AAL2 or at ATM layer is recommended.

3. Conclusion and proposal

It is proposed to handle Priority at AAL2 or at ATM layer on the Iub and Iur interfaces. This has to be specified in TS 25.426.

The node B should know the priority associated with an AAL2 channel or an ATM connection. Since there is no field concerning priority in [4], it is proposed to add a new parameter in the Radio Link Set Up and Radio Link Reconfiguration messages, in order to indicate the priority level to be given to the Iub transport bearer associated to each DCH transport channel. This has to be specified in TS25.433, and text proposal is provided in Tdoc 951/99.

3.1. Proposed figures for delay budget

We propose to modify objective values of figures proposed by Siemens as shown below, with the addition of a note.

	AMR 12.28kbit/s	32kbit/s	64kbit/s	144kbit/s	384kbit/s	2048kbit/s
TN1	51msec*	120msec	185msec	205msec	450msec	

* The objective for AMR 12.2 kbit/s real-time traffic shall be reached for any mix of real time and non real time traffic.

3.2. Change proposal to TS 25.426

Changes are proposed in section 4.2.

4.2 Transport Layer

Asynchronous Transfer Mode (ATM) [2] and ATM Adaptation Layer type 2 (AAL2) [3, 4] are used as a transport layer for DCH data streams on Iur and Iub interfaces. Service Specific Segmentation and Reassembly (SSSAR) sublayer for AAL2 is used for the segmentation and reassembly of AAL2 SDUs.

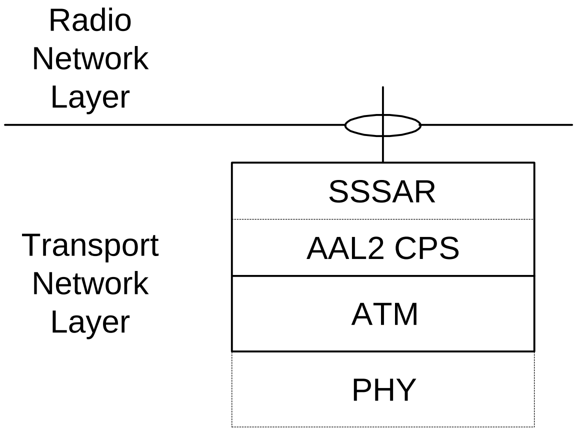

Figure 1. Transport network layer for DCH data streams over Iur and Iub interfaces.

Priority between the different transport bearers set up for each DCH shall handled at AAL2 or at ATM layer, using the transport priority field provided by the Radio Network Layer (NBAP), when the transport bearer is set up.

4. References

[1] ITU-T Recommendation I.363.2 “B-ISDN ATM Adaptation Layer Type 2 Specification”

[2] TS22.105, “Technical Specification Group Services and System Aspects. Service aspects, Services and Service Capabilities”

[3] ITU-T Recommendation I.356 “B-ISDN ATM Layer Cell Transfer Performance”

[4] ITU-T Recommendation Q.2630.1 “AAL Type 2 Signalling Protocol”

[5] TR 101-112 – Selection Procedures for the choice of radio transmission technologies of the UMTS

[6] Tdoc R3-99955, Siemens, AAL2 Packetisation and De-packetisation Delay