

Impact of the Timer-CU of the AAL2 Protocol on Traffic Performance within the UTRAN

Rani Makké¹, Samir Tohmé¹, Jean-Yves Cochenne² and Sophie Pautonnier³
¹*Ecole Nationale Supérieure des Télécommunications (ENST-Paris), Computer and Communication Department, 46, Rue Barrault, 75634, Paris Cedex 13, France*
²*France Telecom Research & Development, 2 Av. Pierre Marzin, 22300 Lannion, France*
³*Mitsubishi Electric ITE-TCL, 80, Av. des buttes de Coesmes, 35700 Rennes, France*

Rani.Makke@enst.fr, Samir. Tohme@enst.fr, jeanyves.cochennec@rd.francetelecom.fr,

Pautonnier@tcl.ite.mee.com

Key words: UMTS, UTRAN, QoS, ATM, AAL2.

Abstract: The UMTS (Universal Mobile Telecommunication System) is the future network for mobile communication systems. The access network of UMTS is called UTRAN (Universal Terrestrial Radio Access Network). Many works focus on the quality of service (QoS) aspects within the UTRAN and the dimensioning aspects of the links between different nodes in the UTRAN. In Release 99 of the 3GPP standard, AAL2/ATM protocol is chosen as a transport protocol within the UTRAN. The Timer-CU is an important parameter for the AAL2 protocol. High Timer-CU values lead to high transfer delays for low bitrate applications, and low values lead to an efficiency loss in the ATM connections. To optimize the choice of this value, we focus our study on the impact of the Timer-CU on the QoS parameters for real time traffic and especially voice applications.

1. INTRODUCTION

The AAL2/ATM (ATM Adaptation Layer type-2) protocol is the transport protocol within the UTRAN for the Release 99 of the 3GPP (3rd Generation Partnership Project) standard. The Timer-CU is a very important

parameter of the AAL2 protocol. The value of this timer must be carefully chosen because it has an important impact on the QoS of the traffic supported by this protocol, especially for low loaded links. This parameter also has an impact on the links utilization between Nodes within the UTRAN. In order to determine the optimal value of this parameter, we will study the QoS parameters depending on different Timer-CU values [17].

2. OVERVIEW OF THE UTRAN ARCHITECTURE

The UTRAN is the UMTS Terrestrial Radio Access Network. Its architecture [1] is very similar to the architecture of the GSM radio access network. Figure 1 represents the architecture of the UTRAN:

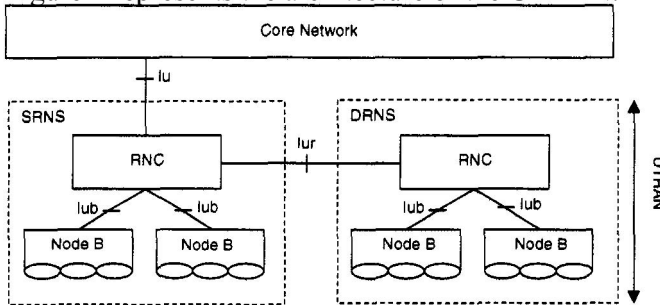


Figure 1. UTRAN architecture

The different elements of UTRAN are:

-RNC: Radio Network Controller: it controls the radio resources.

-NodeB: equivalent of the BTS (Base Transceiver Station) in GSM. The principal role of this node is to transmit data and signaling on the radio interface.

-RNS: Radio Network Subsystem: it is the access part of the UMTS network that manages the allocation and the release of the radio resources for a set of cells. There is only one RNC in each RNS.

The different UTRAN interfaces are:

-Iu: inter-connection point between the RNS and the Core Network.

-Iub: interface between one NodeB and one RNC.

-Iur: interface between two RNCs.

When the mobile terminal is in soft handover state (when the terminal has several radio links with different cells corresponding to different RNS), one RNS has an inter-connection point with the Core Network. This RNS is called SRNS (Serving RNS). The other RNS is called DRNS (Drift RNS)

and can transmit all the user data flows to the SRNC that establishes the recombination. The recombination principle consists of recombining all the data flows received from one user into one data flow to the core network.

The 3GPP has chosen in Release 99 the AAL2/ATM protocol as a transport protocol within the UTRAN (on the Iub and Iur interfaces).

3. OVERVIEW OF THE PROTOCOL STACK WITHIN THE UTRAN

The data flow goes through different protocol layers and its characteristics change before entering in the AAL2 layer. In order to study the performance of the AAL2 protocol, we should have an idea about the behavior of the traffic coming from upper layers. The protocol stack architecture of the UTRAN is described in Figure 2:

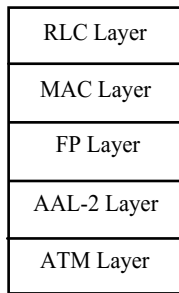


Figure 2. Protocol stack

The RLC layer [9] is transparent for voice flows. For data flows, the RLC layer segments higher layer data units into a number of RLC-PDUs (RLC - Packet Data Unit) that will be transmitted in the TTIs (Transmission Time Interval) of the MAC layer [10]. The FP (Framing Protocol) layer assembles all the RLC-PDUs transmitted in one TTI in one FP-PDU (FP- Packet Data Unit). This FP-PDU is transmitted to AAL2 layer [7,8].

4. OVERVIEW OF THE AAL2 PROTOCOL

The AAL2 [13,14] is used to transport very low bitrate applications with real time constraint and variable bitrate (e.g. the compressed voice).

The AAL2 was defined to get around the problem of the ATM cell packetization delay that becomes critical for the low bitrates (at 16 Kbps, its

value is 24 ms). The solution is simple: when multiplexing several communication flows in the same ATM channel, the delay becomes reasonable for a given communication.

The AAL2 protocol consists of variable length data units called mini-cells with a maximum payload length of 45 bytes (optionally 64 bytes).

The AAL2 layer is divided into two sublayers (Figure 3): the SSCS (Service Specific Convergence Sublayer) and the CPS (Common Part Sublayer). The Segmentation and Reassembly SSCS segments higher-level data units exceeding 45 bytes (or 64 bytes) into packets with a maximum length of 45 bytes (or 64 bytes).

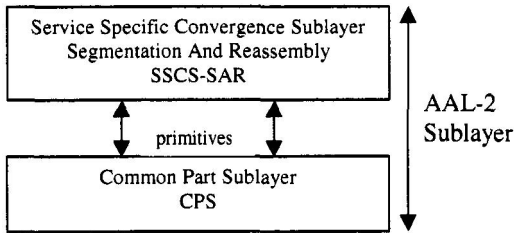


Figure 3. AAL-2 sublayers

The CPS layer overloads each mini-cell by a 3 bytes header, format is described in Figure 4.

CID	LI	UII	HEC
8 bits	6 bits	5 bits	5 bits

Figure 4. Mini-cell header format

The CID (Channel Identifier) field identifies the AAL2 connection. There are 256 possible CID values, 8 of them are reserved for signaling purpose, and the rest can be used to identify 248 different AAL2 connections. The LI (Length Indicator) field determines the mini-cell payload length. By default, the maximum length is 45 bytes, but it may be 64 bytes if there is an indication at the connection establishment procedure. The UII field is assigned to the SSCS. These 5 bits (32 codepoints) are not interpreted by the CPS sublayer and they are passed transparently from the transmitter's SSCS to the receiver's SSCS. The HEC (HeaderError Control) is used for error detection in the mini-cell header.

The ATM header allows two levels of addressing (Virtual Path Identifier VPI and Virtual Circuit Identifier VCI). Thus, it is possible to set up ATM

VPCs between AAL-2 end points and to allow them to use VCI and CID to create multiple native connections. With a 16-bits VCI field, an ATM VPC will be able to support up to 248×2^{16} AAL-2 connections.

The mini-cells are inserted in the ATM cells and one byte field (STF: STArt Field) is added at the beginning of the ATM payload. The STF contains a pointer to the first byte of the first mini-cell header in the ATM payload. Overlapping is used: one mini-cell inserted in the ATM cell can overlap onto the next cell (Figure 5). Padding may be added at the end of the ATM cell payload if there are no additional mini-cells to be inserted.

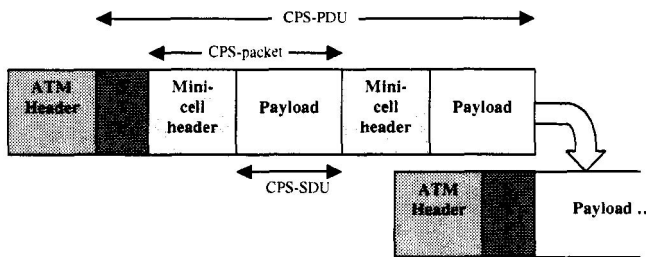


Figure 5. Overlapping

The Timer-CU is an important parameter in the AAL2 protocol. When a CPS-SDU (CPS Service Data Unit) is presented in the CPS sublayer, the CPS protocol adds the 3 bytes header to form the CPS-packet. Then Timer-CU is armed and the STF is calculated. The CPS-PDU is transmitted to the ATM layer either when it is full, i.e. the 47 byte payload field is filled, or when the Timer-CU has expired. If the Timer-CU expires before the CPS-PDU becomes full, padding is added to fill the CPS-PDU that is transmitted to the ATM layer.

The Timer-CU value is a critical parameter in the case of low bitrate traffic. A very small Timer-CU value leads to a loss of bandwidth efficiency. A very high Timer-CU value optimizes the bandwidth utilization, but leads to a higher packetization delay and consequently to a deterioration in the QoS for the real time traffic. The optimal Timer-CU value should be chosen carefully to optimize the bandwidth utilization and keep an acceptable QoS for real time traffic.

5. QOS ASPECTS

Within UTRAN, two types of flows are transported on Iub and Iur interfaces: real time applications (e.g. voice), which require low delays, and

non real time applications (e.g. web browsing, FTP, SMS), which are tolerant for transfer delay. In order to differentiate between these different application types, two schemes are envisaged:

1. Two-VC scheme: Real time applications are transported on a VPNC with stringent class of QoS, and non real time applications are transported on another VP/VC with tolerant class of QoS. In this strategy, we differentiate between real time traffic and non real time traffic at the ATM layer and all VCs are mono-service connections.

2. Single-VC scheme: Real time and non real time applications are transported on the same VP/VC. The differentiation between services is done by scheduling mechanisms at AAL2 layer. This is the case of multi-service connection.

In the first scheme, the real time VC is more critical because of the low bitrate of the flow in the case of low loaded VC. The non real time VC is less critical because the data flows are not very sensitive to delay. Furthermore, data frames are as long as they fill the ATM cells without waiting for a long packetization time. Thus, the study of the real time VC is more stringent.

In the second scheme, the case of a low loaded VC transport real time and non real time flows is critical. In fact, considering priority scheduling mechanisms between real time and non real time applications at the AAL2 level, first real time packets are inserted in the ATM cells and afterwards, data packets (if any in data buffers) will fill the rest of the cell payload field and thus minimize the delay (packetization time) for real time packets.

In this paper, our study is focused on the impact of the Timer-CU value on the QoS parameters for the voice traffic with the two different schemes described above.

6. TRAFFIC MODEL

In order to evaluate the AAL2 protocol performance, a traffic model is needed for voice flows entering in the AAL2 layer. At the source level, a voice model is defined by the 3GPP for the AMR (Adaptative MultiRate) codec [4,5,11]. This model consists of an ON/OFF model with exponential distribution of the ON and OFF periods. The time interval between two packets is a constant value. Table 1 represents the model parameters.

Table 1: voice model parameters

Parameter	Distribution	Value
ON period length	Exponential	3 sec (ETSI model) or 352 ms
OFF period length	Exponential	3 sec (ETSI model) or 352 ms
Time-interval between two packets	Constant	20 ms
AMR Packet size	Constant	Depending on the AMR coding mode
SID packet size	Constant	39 bits

There are different AMR coding mode. Table 2 represents these different modes with their corresponding packet size.

Table 2: AMR coding mode

AMR coding mode	Throughput	Packet size
AMR12.20	12.20 Kbps	244 bits
AMR10.20	10.20 Kbps	204 bits
AMR7.95	7.95 Kbps	159 bits
AMR7.40	7.40 Kbps	148 bits
AMR6.70	6.70 Kbps	134 bits
AMR5.90	5.90 Kbps	118 bits
AMR5.15	5.15 Kbps	103 bits
AMR4.75	4.75 Kbps	95 bits

In fact, the traffic coming from the voice source passes through the stack protocol described in paragraph 3. The traffic form is lightly changed when entering in the AAL2 layer. The RLC protocol is transparent for voice traffic. At the MAC layer, one voice packet is transmitted on the radio channel in each TTI (Transmission Time Interval). The TTI value for voice connections is 20 ms. Thus, the FP layer receives one voice packet each 20 ms. It adds the FP overhead and send this packet to the AAL2 layer. The traffic entering the AAL2 layer is similar to the source traffic model with a different packet size (native packet size + MAC and FP headers). Ten signaling bytes are sent every 300 ms.

For data traffic, we consider in our study Web browsing sources. A web traffic model is defined by the 3GPP [4]:

Web browsing session consists of a sequence of packet-calls. The number of packet-calls in a session is a geometrically distributed random

variable with a mean of 5. Each packet-call consists of a number of packets, which is a geometrically distributed random variable with a mean of 25. Packet-calls are separated by an interval called reading-time which is a geometrically distributed random variable with a mean of 412 seconds. Packet size is a Pareto with cut-off distributed random variable. The normal Pareto distribution (without cut-off) is defined by:

$$f_x(x) = \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}}; x \geq k$$

Packet size is defined by: PacketSize = min(P,m), where P is the normal Pareto distributed random variable ($\alpha=1.1$, $k=81.5$ bytes) and m is the maximum allowed packet size. The time interval between two consecutive packets inside a packet-call is a geometrically distributed random variable with a mean depending on the type of the UDD source. Table 3 represents this time interval for different UDD types.

Table 3: UDD types

UDD types	UDD 8Kbps	UDD 32Kbps	UDD 64Kbps	UDD 144Kbps	UDD 384Kbps	UDD 2048Kbps
Average inter-arrival time between packets	0.5	0.125	0.0625	0.0277	0.0104	0.00195

This model is applicable on the Iu interface but it is not directly applicable on the Iub and Iur interfaces. In fact, RLC flow control mechanisms shape the traffic coming from upper layers so that the data throughput on the air interface does not exceed the defined bitrate for the chosen UDD (Unconstrained Delay Data) mode. Furthermore, the TTI parameter gives a periodic pattern to the traffic transported on the radio channels and entering in the AAL2 layer. The RLC protocol splits upper layer packets into TBs (Transport Blocks) with a predefined size and adds a 2 bytes header. In each TTI, the MAC layer sends a certain number of TBs on the radio channel without exceeding the bitrate of the UDD mode (in our model). This number is determined by a dynamic resource allocation algorithm at the RRC layer (Radio Resource Controller). At the FP (Frame Protocol) layer, all TBs transmitted in the same TTI for one user are assembled in one FP-PDU that is transmitted to the AAL2 layer after adding FP overhead. The UDD traffic entering in the AAL2 layer has the pattern represented in Figure 6.

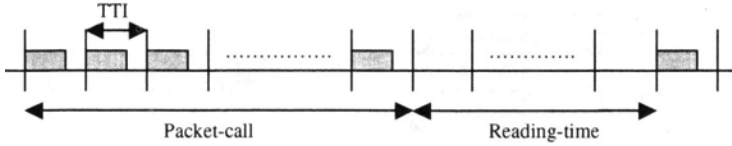


Figure 62. UDD traffic pattern at AAL-2 layer

In each TTI, one FP-PDU is sent to the AAL2 layer. The FP-PDU size depends on the UDD mode and on the number of TBs sent in the TTI. The number of TBs sent in each TTI depends on the radio link utilization: if it is low loaded, the RRC algorithm increases the number of TBs allowed to each user in the TTI because there is a free bandwidth. If the radio link is very loaded, the RRC algorithm decreases the allowed number of TBs for each user in order to share the bandwidth between all data users. As our study is focused on the impact of the Timer-CU on traffic performance in the case of low bitrate traffic, we can consider that the radio bandwidth is low loaded. Therefore, the UDD sources can transmit data at the maximum bitrate allowed to each type. Thus, the FP-PDU size is calculated corresponding to the maximum UDD bitrate. A packet-call corresponds to the download of a file which has a Pareto distribution size with a parameter $\alpha = 1.1$, mean file size = 12 000 bytes, minimal file size $k = 1\ 858$ bytes, maximal file size = 5 000 000 bytes [12].

6.1 Simulation model

In our study, we consider two simulation scenarios according to the two possible solutions proposed in paragraph 5 (Figure 7). In the first scenario (a), we separate real time traffic and non real time traffic into two different VCs. As real time traffic is more critical in term of delay, we will study the impact of the Timer-CU on traffic performance for the VC transporting voice flows. In the second scenario (b), we consider one VC in which we aggregate real time and non real time flows. In this case, a scheduling mechanism is required at AAL2 layer. The AMR12.2 mode is used as a voice traffic model. Two data traffic model are used: UDD64Kbps and UDD384Kbps.

For voice channels, the TTI value is equal to 20 ms. For data channels, we assume a TTI value of 40 ms.

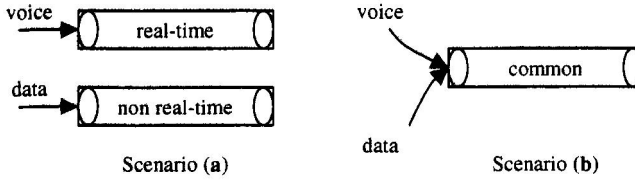


Figure 7. Two different scenarios

7. SIMULATION RESULTS

The QoS parameters evaluated are transfer delay and delay variation. The filling ratio is an important parameter that gives an idea about the efficiency of the link. It is defined by $FR = NB/47$, NB is the utile payload (in bytes) of the CPS-PDU, and 47 is the length of the CPS-PDU payload field.

Scenario (a): we consider a DBR VC with PCR = 2Mbps [15,16]. For the 352ms/650ms voice model, we represent the filling ratio in Figure 8 for different Timer-CU values. The 95-percentile delay and the standard deviation (StdDev) of delay are measured and represented in Figure 9.

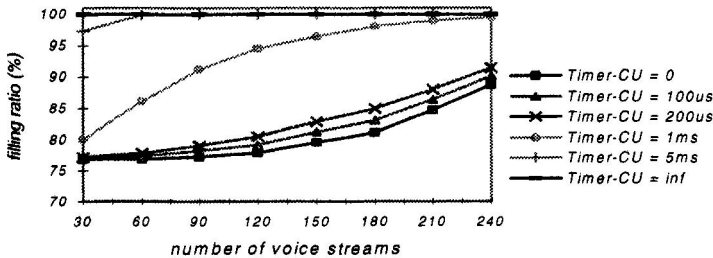


Figure 8. Filling a ratio (352ms/650 model)

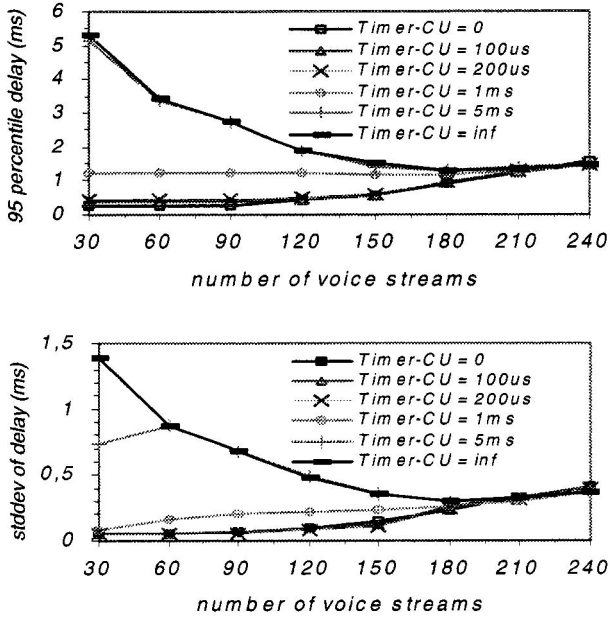


Figure 9. 95-percentile delay and StdDev of delay (352ms/650ms model)

The same parameters are measured for the model 3s/3s. The filling ratio is represented in Figure 10; the delay and StdDev of delay are represented in Figure 11.

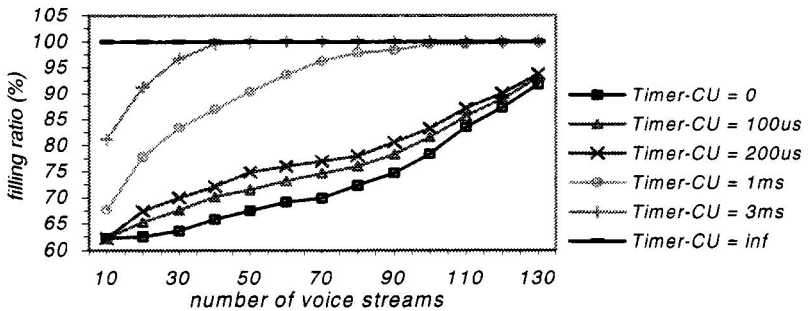


Figure 10 Filling ratio (3s/3s model)

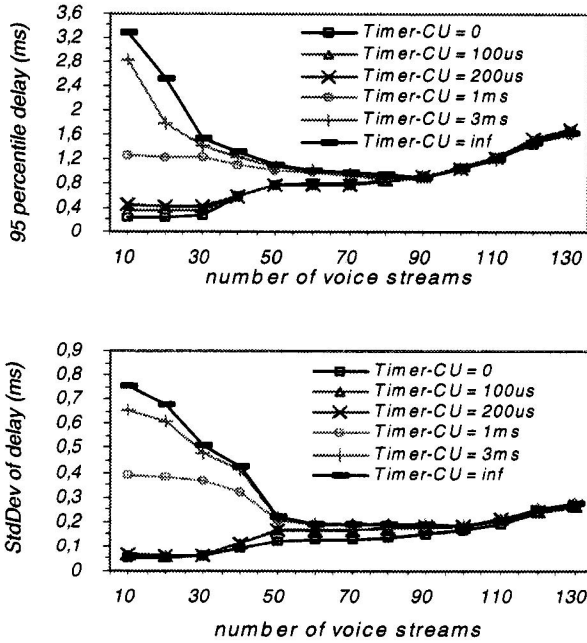


Figure 11. 95-percentile delay and StdDev of delay (3s/3s model)

When the number of voice streams becomes large, the QoS parameter values become the same for all Timer-CU values. The impact of the Timer-CU is very clear for a low loaded VC.

In order to evaluate accurately the impact of the Timer-CU in the case of low loaded VC, we fixed the load of the VC ($\rho < 0.1$) and we measured the QoS parameter values depending on the Timer-CU value for different VC PCR. The Timer-CU values are chosen as a multiple of the period T ($T=1/PCR$).

In Figure 12, we represent the 95-percentile delay and the standard deviation of delay against the Timer-CU value and for different VC PCR (the unit of the Timer-CU is $T=1/PCR$). The QoS parameters increase with the Timer-CU value; VCs with smaller PCR value generate greater delay and delay variation for high Timer-CU values. In order to optimize the QoS of voice traffic, low Timer-CU values are suitable.

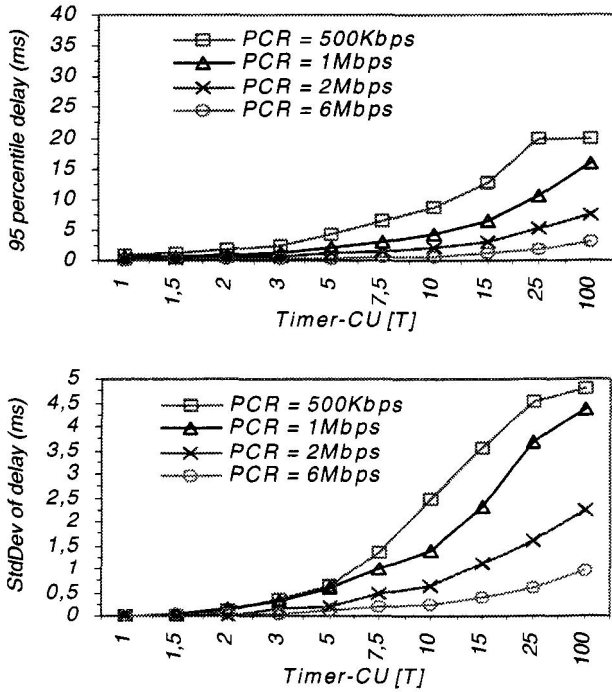


Figure 12. 95-percentile delay and StdDev of delay (voice only)

The filling ratio is represented in Figure 13. Such all VCs have the same load, the filling ratio is the same for different PCR values and it increases with the Timer-CU value.

The VC utilization ratio is defined as: $VC_utilization = cell_rate/PCR$, where $cell_rate$ is the number of cells transmitted on the VC per second, and PCR is the peak cell rate of the VC DBR.

This parameter is represented in Figure 14. For low Timer-CU values, the VC is more loaded because of the padding added to partially filled cells. In order to optimize the efficiency of the VC bandwidth, high Timer-CU values are appropriate.

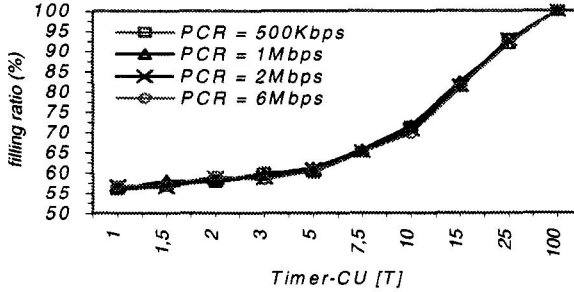


Figure 13. Filling ratio (voice only)

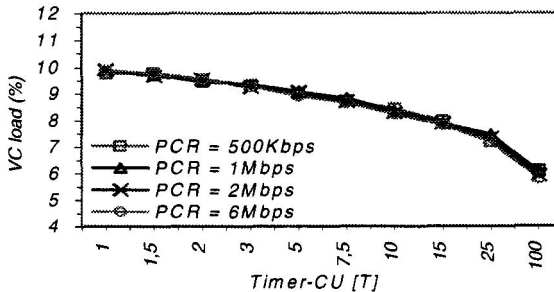


Figure 14. VC utilization (voice only)

The optimal Timer-CU value that guarantees the required QoS and gives in acceptable bandwidth utilization should be chosen carefully depending on the PCR value.

Scenario (b): in this scenario, we consider a VC DBR in which we aggregate voice and data flows. We consider two web traffic types: UDD64Kbps and UDD384Kbps. At AAL2 layer, a priority scheme is done to differentiate between voice and data flows: Data packets are served only when voice buffers are empty. The QoS parameters are measured for different traffic combinations.

Figure 15 represents the filling ratio in the case of low loaded VC ($\rho < 0.1$) with 20% AMR12.2 and 80% UDD64Kbps. Figure 16 represents the delay for voice and data packets.

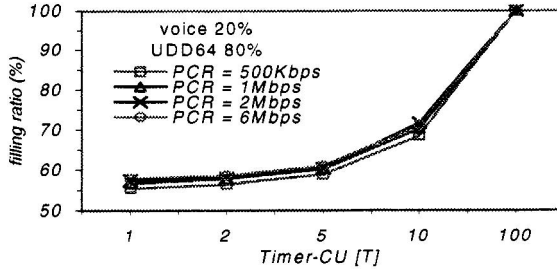


Figure 15. Filling ratio. Multiservice VC (UDD64)($\rho < 0.1$)

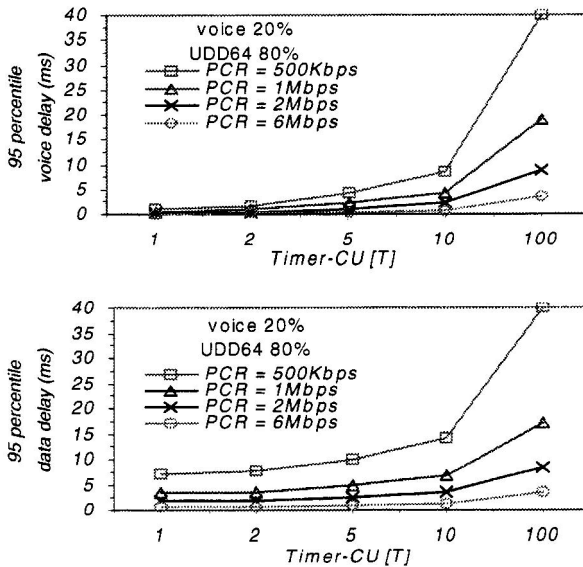


Figure 16. Voice and data delay (UDD64). Multi-service VC ($\rho < 0.1$)

The AAL2 bandwidth efficiency is the same for all PCR values and it increases with Timer-CU value.

The voice packet delay increases with Timer-CU value. Because of the small number of voice users, it is probably that all sources are in the OFF state; a packet arrived at the end of the ON state of one user and does not fill the ATM cell payload field, will wait for a long time before be sent to the ATM layer (in the case of high Timer-CU value).

For data flow, all mini-cells have a length of 45 bytes and can be transmitted at the peak cell rate of the VC. Only the last mini-cell may have

a length less than 45 bytes and thus it should wait for the arrival of another packets or the expiration of the Timer-CU.

Figures 17 and 18 represent the same parameters in the case of UDD384Kbps traffic. (20% voice and 80% data).

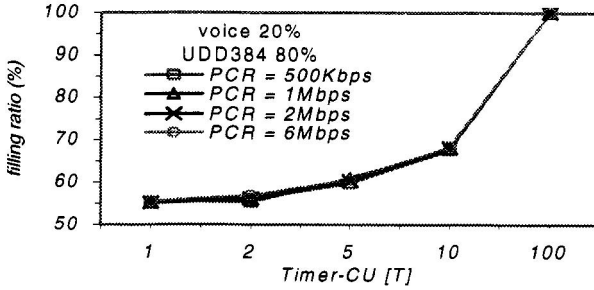


Figure 17. Filling ratio (UDD384)($\rho < 0.1$)

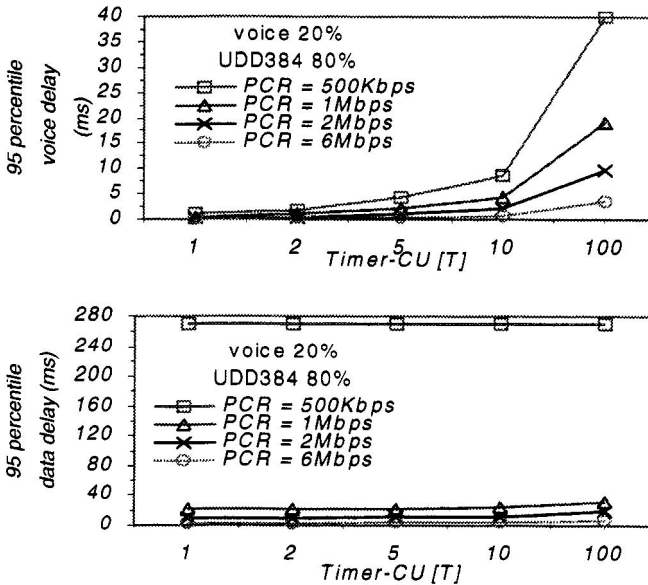


Figure 18. Voice and data delay (UDD384)($\rho < 0.1$)

With another traffic combination, 80% voice and 20% data, and another load ($\rho < 0.2$), we calculate the performance parameters in the case of UDD64

and UDD384 data traffic. Figure 19 represents the filling ratio and Figure 20 represents the delay in the case of UDD64 data traffic.

In the case of UDD384 data traffic, the same parameters are represented in Figure 21 and 22.

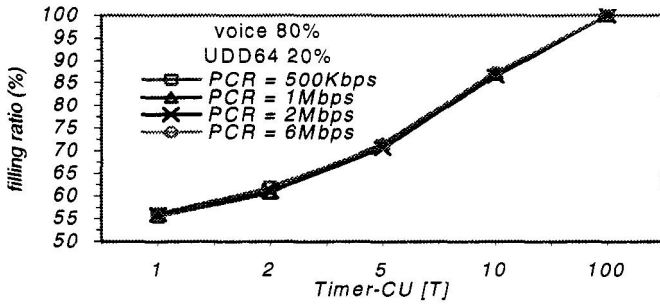


Figure 19. Filling ratio (UDD64)($\rho < 0.2$)

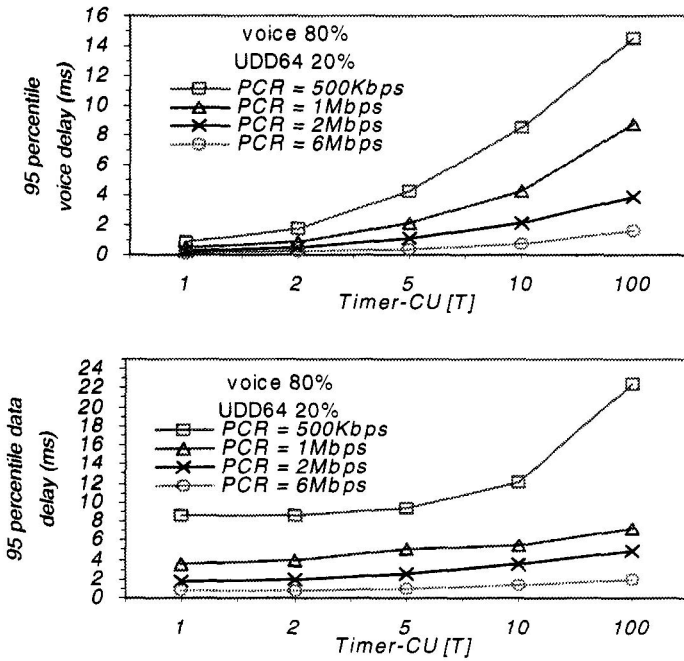


Figure 20. Voice and data delay (UDD64)($\rho < 0.2$)

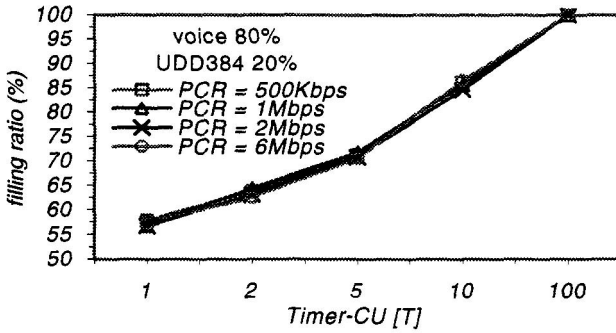


Figure 21. Filling ratio (UDD384)($\rho < 0.2$)

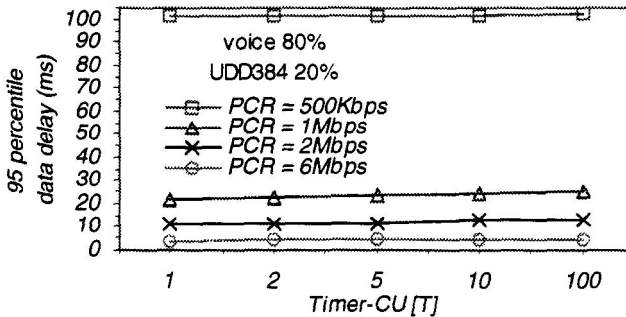
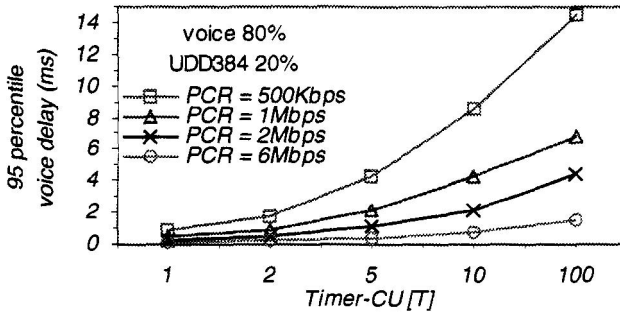


Figure 22. Voice and data delay (UDD384)($\rho < 0.2$)

8. CONCLUSIONS

The Timer-CU is an important parameter in the AAL2 protocol. It widely affects the performance of the AAL2 connections in the case of low loaded

VCs. It must be carefully chosen to have a compromise between transfer delay and bandwidth efficiency. In our study, we showed that high Timer-CU values lead to a QoS degradation, and very low values lead to a bandwidth efficiency loss. In the case of high loaded VC, the Timer-CU value doesn't have an important impact on traffic performance.

At same VC load (low load), the filling ratio is approximately the same for the mono-service VC (voice only) and the multi-service VC (voice + UDD64/UDD384). It increases with the Timer-CU value. For different PCR values, the filling ratio is the same at equal loads. The filling ratio increases with the load of the VC.

The transfer delay for voice packets is affected in the case of a multi-service VC that has the same load of a mono-service VC. In fact, data traffic has a bursty pattern, and long data packets arrive to the data buffer and serve to fill the ATM cells when there are no additional voice packets. The packetization delay of voice packets decreases when data traffic is aggregated with voice traffic in the same VC. The optimal Timer-CU value chosen in the case of mono-service VC gives better performances for voice traffic in the case of multi-service VC (for the same load). The optimal Timer-CU value is a function of the VC load and the period T corresponding to the PCR of the DBR VC. When the VC load increases, the transfer delay decreases for high Timer-CU values, but when the VC becomes very loaded, the delay increases due to the congestion in the buffers, and Timer-CU will not have an important impact on performance parameters.

The optimal Timer-CU value for a low loaded VC may be chosen in the case of mono-service VC because it is the worse case, and we will be sure that this value will give good performances in the case of multi-service VC.

REFERENCES

- [1] ETSI TS 125.401 UTRAN Overall Description.
- [2] 3G TS 23.107 QoS Concept and Architecture
- [3] 3G TS 22.105 Services and Service Capabilities.
- [4] ETSI TR101.112 V3.2.0(1998-04) Selection Procedures for the choice of radio transmission technologies of the UMTS.
- [5] ETSI TS 126.101 AMR Speech Codec Frame Structure.
- [6] 3G TS 25.301 Radio Interface Protocol Architecture.
- [7] 3G TS 25.427 UTRAN Iub/Iur Interface User Plane Protocol.
- [8] 3G TS 25.435 UTRAN Iub Interface User Plane Protocols for Common transport Channel Data Streams.
- [9] 3G TS 25.322 RLC Protocol Specification.
- [10] 3G TS 25.321 MAC Protocol Specification.
- [11] 3G TS 26.071 AMR Speech Codec; General Description.
- [12] 3GPP TSG-RAN WG3 Ad Hoc Meeting #2, Paris, France, 6th – 8th November 2000.

- Transport Network Bandwidth Utilization. Motorola.
- [13] ITU-T 1.363.2, B-ISDN ATM Adaptation Layer Specification: AAL Type2.
 - [14] ITU-T 1.366. 1, Segmentation and Reassembly Service Specific Convergence Sublayer for the AAL type 2.
 - [15] ITU-T 1.356 B-ISDN ATM Layer Cell Transfer Performance.
 - [16] ITU-T 1.371 Traffic Control and Congestion Control in B-ISDN.
 - [17] Minicel project, ENST-Paris, France Telecom-Research and Development (FTR&D), Mitsubishi Electric ITE.