End-to-end ATM and AAL2 delay calculation in UTRAN

Gábor Horváth, Technical University of Budapest, 1521 Budapest, Hungary,
hgabor@webspn.hit.bme.hu
Miklós Telek, Technical University of Budapest, 1521 Budapest, Hungary, telek@hit.bme.hu
Csaba Vulkán, Nokia Research Center, Helsinki, Csaba.vulkan@nokia.com

Abstract *This paper proposes a stochastic model of an ATM based transport network layer of a UTRAN (UMTS Terrestrial Radio Access Network) that transmits the user data of different services such us voice, real time (RT) data and non-real time (NRT) data. Based on this stochastic model a numerical method is presented for the approximate analysis of end-to-end delay in UTRAN. A detailed discussion is provided about the "real life" traffic behavior of the considered Radio Access Network (RAN) and the applied modeling assumptions. The proposed method applies a "two-parameter" approximation, i.e., it provides not only the mean, but also the variance of the end-to-end delay. The applicability and the precision of the proposed method are demonstrated via numerical examples.*

Keywords: ATM, radio access network, traffic model, queuing network, AAL2.

1. Introduction

The UTRAN system architecture is shown in the Figure 1. A radio access network consists of one or more Radio Network Subsystems (RNS), each controlled by an RNC (Radio Network Controller). The radio network subsystems can be interconnected via the Iur interface. One RNC can control one or more base stations (Node B). They are connected to the RNC by the Iub interface. One Node B can serve one ore multiple cells. The UTRAN is connected to the core network by the Iu interface.

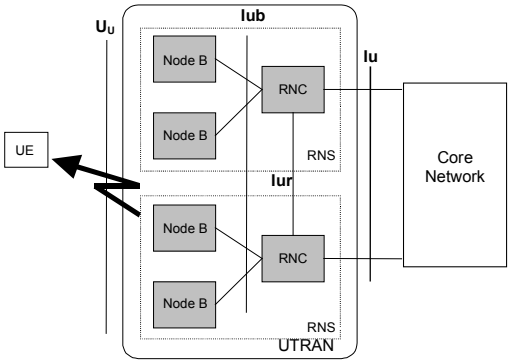

Figure 1 UTRAN system architecture

The delay budget for the user data within the UTRAN is specified in [1]. This paper focuses on ATM and AAL2 layer related elements of the user traffic end-to-end delay on the Iub interface.

The rest of the paper is organized as follows. Section 2 introduces the Iub reference model. Section 3 and 4 discuss the modeling assumptions and the analysis methods applied for the AAL2 and the ATM layers, respectively. Section 5 provides numerical examples of the proposed method and Section 6 concludes the paper.

2. ATM based network transport layer

In order to meet the QoS requirements described in [2] ATM and AAL2 has been selected as transport technology over the Iub in UMTS Release ´99. Figure 2 presents our Iub reference model for the transport layer. The AAL2 user is the Frame Protocol in UTRAN. The delay budget on the AAL2 and ATM layers is 7ms.

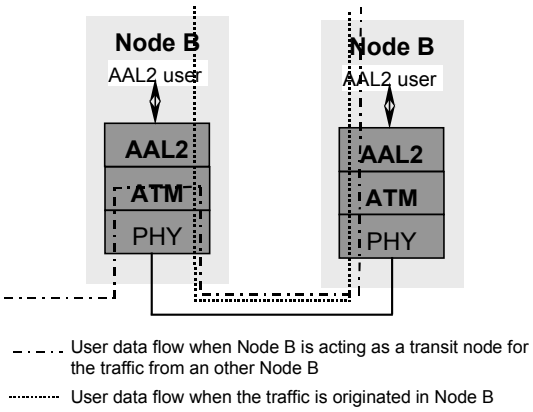

Figure 2 Reference model

The traffic directed from a Node B to the RNC is called up link traffic, and the traffic to the opposite direction as down link traffic.

Each Node B within an RNS can realize two main functions. As an ingress node, they can enter users' traffic to the network and as a core node they transmit the traffic coming from other network nodes together with the traffic entered the network at the same node.

2.1. Traffic models of different services

The following service types are considered: RT voice, RT data, and NRT data. There is a separate traffic model for each traffic classes. The main traffic parameters of the possible traffic classes are provided in the following table.

Traffic Type	Service rate (kbps)	Frame period (ms)	Bits to ATM
RT8 (sign.DCH)	8	40	384
RT12.2 (AMR) UL	12.2	20	336
RT12.2 (AMR) DL	12.2	20	320
RT7.4 (AMR) UL	7.4	20	224
RT7.4 (AMR) DL	7.4	20	208
RT64	64	40	2792
RT144	144	40	6208
RT384	384	40	16432

NRT8	8	10	144
NRT16	16	10	224
NRT32	32	10	384
NRT64	64	10	728
NRT144	144	10	1600
NRT384	384	10	4144

The detailed traffic behavior of the traffic classes are characterized by some further parameters:

1. RT voice: An on-off behavior is assumed. There is no data transmission during the talk silence intervals. The considered on/off ratio is 2, i.e., the probability of a packet arrival in a frame is 0.66.
2. RT data: Simple CBR data transmission is assumed with regular packet arrivals.
3. NRT data: WWW browsing session is assumed with packet calls and reading periods. The details of the traffic model (the distribution of number of packet calls, reading time between packet calls, number of packets in a packet call, inter arrival time between packets, mean packet size) are taken from [3].

In the analysis we consider a static traffic situation. I.e., we evaluate the end-to-end delay in case of a given number of connections of the different traffic classes between the node pairs. We do not consider the user arrival and departure process of the different traffic classes.

The analysis method is composed by two main parts. First, the delay and the traffic shaping behavior of the AAL2 layer is evaluated and than, the ATM layer behavior is analyzed. The previous is associated with the ingress node function, while the later with the core node function.

3. AAL2 layer behavior and its model

The user traffic transmitted trough the UTRAN is adapted to ATM cells by AAL2 protocol. In case of small (<< 48 byte) source packets the AAL2 protocol performs an effective multiplexing, while in case of large source packets (>> 48 byte) it completes the fragmentation of the packets. Among the considered traffic classes there are large and small packets as well. The multiplexing gain of the AAL2 protocol is tuned by the Timer_CU value. When Timer_CU =0 an ATM cell is generated form any small data fragment if the buffer is empty. When Timer_CU >0 the data segment smaller than 48 byte is delayed to wait for new packet arrivals that feeds the ATM payload full. Larger Timer_CU results in larger delay and higher multiplexing gain, i.e., fewer load to the ATM layer.

The traffic enters the network at a given node can be transmitted through one or more VCC connections. If more than one VCC connections are initiated at a node there is a predefined assignment of VCCs and mobile service connections and we model each VCC separately. The numbers of connections of different traffic types transmitted through a particular VCC are input data to the algorithm.

Summing up, the AAL2 layer analysis method evaluate the mean and the coefficient of variation (cv) parameters of the AAL2 layer delay and the traffic parameters of the ATM cell steams, that goes

through the outgoing VCCs. The AAL2 layer analysis method is based on the traffic load (number and type of connections) associated with a given VCC. The main steps of the procedure are as follow:

- generation of the modulated $\sum N_k*D_k/D/1$ queue parameters,
- calculation of the virtual waiting time,
- calculation of the mean busy and idle periods of the $\sum N_k*D_k/D/1$ queue,
- generation of the mean and squared coefficient of variation parameters of the outgoing ATM streams.

The following subsections provide the detailed descriptions of these steps.

3.1. Modulated $\sum N_k*D_k/D/1$ queue

The main difficulty in modeling the AAL2 layer traffic behavior is the complex nature of the traffic sources. On the one hand, we need to model the periodic behavior of the traffic, and on the other hand, we need to consider the on-off behavior of the RT voice connections and the alternation of packet calls and reading periods of NRT data connections. To meet these complex requirements we model the AAL2 layer with a modulated $\sum N_k*D_k/D/1$ queue [4].

The basic N*D/D/1 queue models N independent identical traffic sources with periodic arrival of fixed size packets with frame size D, i.e., the frame size equals to the transmission time of D outgoing packets. In the AAL2 layer model the outgoing packets are ATM cells and the time of a cell transmission is determined by the capacity of the outgoing link. In the practical examples D is an integer, but this assumption can be relaxed easily. The arrival instances of the N sources are uniformly distributed over the frame period. The virtual waiting time distribution (the probability that the waiting time of an incoming data unit is greater than x in the steady state), that we use to approximate the delay distribution, of the N*D/D/1 system is computed as [4]

$$Q_D^N(x) = \sum_{x<n\le N} \binom{N}{n} \left(\frac{n-x}{D}\right)^n \left(1-\frac{n-x}{D}\right)^{N-n} \frac{D-N+x}{D-n+x}$$

This basic N*D/D/1 model is extended to capture more complex traffic behavior [4]. An extension of the basic model allows analyzing the virtual waiting time of periodic traffic mixtures with different frame periods, which is the case with the considered traffic types. And another extension allows handling modulated traffic sources. A traffic source is called modulated if it does not always transmit one packet in each frame period (with probability one), but it transmits a packet with probability p (p<1). Which is the case, for example, with the RT voice sources due to their on-off behavior.

The key element of the application of these extensions is to choose an adequate frame period, that is used as the *basic frame period* during the computations, and to evaluate the arrival probability of packets coming from the different sources during this basic frame period.

3.1.1. The role of the basic frame period

For the case of analyzing traffic mixtures with different frame periods, it is recommended to use the shortest frame period as the basic frame period in [4]. Using this basic frame period we need to evaluate the distribution of the number of packet arrivals from the sources with higher frame period. This approach is correctly captures the periodicity of the sources, but it has a limit of applicability. The theoretical applicability of the N*D/D/1 model is limited to the cases when the number of packets arrive during in a basic frame period is not higher than the number of packets that can be transmitted during the same period, i.e., D≥N.

Using the minimal frame period as the basic frame period might result that the mentioned stability condition does not hold for the this short basic frame periods. It means that the number of packet arrivals during short basic frame period can be larger than D, even the steady state utilization, that is equivalent with the utilization over the largest frame period, is less than 1. Indeed the shorter the basic frame period is the higher the probability of overloading this period. In practical computations the basic frame period is often chosen such that N can be greater than D, but with a very low probability. E.g., a precision requirement can be set for the N>D probability.

On the other hand, the use of a larger basic frame period does not model correctly the periodic behavior of the sources with frame period less than the basic frame period. A basic frame period larger than the shortest frame period over estimates the virtual waiting time, because it assumes a less regular arrival process (during the basic frame period) than the periodic one. Fortunately this approximation is pessimistic, i.e., the delay obtained using this approximation is higher than the 'exact' one. Since the applied analysis is based on this approximation our result upper bounds the real delay.

Based on the above assumptions the selection of basic frame period is an 'engineering' trade off between the proper modeling of periodic arrivals and the tolerable probability of overloading. The numerical experiences of the considered traffic situations suggested us the use of the basic frame period equivalent with the NRT data frame period (10 ms) in our computations.

Numerical examples demonstrate the effect of the basic frame period on the calculated virtual waiting time in Section 5.

3.1.2. Packet arrival probabilities

Having the basic frame period we need to evaluate the probability distribution of the number of packets arrive during this period. In our traffic source models the frame period of the different traffic types are integer multiple of each other, hence the subsequent discussion is restricted to this case, however, the extension to general frame period ratio is straightforward.

The most complicated case that can arise in our AAL2 model is when an on-off traffic source transmits packets of random size with a frame period different from the basic frame period. This case is analyzed using the following notations. The basic frame period is D (i.e., D times the cell

transmission time), the source's frame period is D_i. During a D_i long frame period the source transmits a packet of size X_i bytes with probability P_{on}, where X_i is a discrete random variable with distribution $S_i(n) = P(X_i = n)$ and P_{on} is the probability that the source is active. Considering the random size of the user data packet and the on-off behavior of the source the distribution of the amount of data that is transmitted in a D_i long frame period is $F_i(n) = P_{on}P(X_i = n) = P_{on}S_i(n)$ for n>0 and $F_i(0) = P_{off}$. Traffic sources with continuous transmission and/or deterministic packet size are special cases of the most complicated case, and can be analyzed based on $F_i(n)$ as well.

For example, in case of RT voice connections $F_i(n)$ takes the value of the voice packet size (e.g., 42 bytes for RT12.2 (AMR) UL) with probability P_{on} and 0 byte with probability P_{off}.

The analysis of the traffic sources proceeds with the following steps. First we round up the random packet sizes to integer multiple of ATM cells considering the size of the AAL2 header. This gives a discrete distribution of the packet size in ATM cells. Then we assume that each of these packet fragments, which correspond to an ATM cell, arrives uniformly distributed over the frame period of the source. This is the assumption that overestimates the delay a bit, because the data fragments arrives more evenly distributed over the basic frame period when $D > D_i$. Further more we assume that the fragmentation of large data packets goes together with the shaping of the segments, hence the arrival of a large user data packet (>>48 byte) does not result in a batch arrival of ATM cells due to the considered implementation of the AAL2 layer.

As it is discussed above the selection of the basic frame period a numerical problem, hence we need to consider both cases, when $D > D_i$ and when $D_i > D$. When $D > D_i$ the distribution of the number of cells arrive in a basic frame period from source *i* is calculated as the D/D_i times convolution of the distribution given for the D_i frame period: $\hat{F}_i(n) = \underset{D/D_i}{\otimes} F_i(n)$. If $D_i > D$ we the distribution of the number of cells arrive in a basic frame period based on the uniform arrival distribution over D_i. This results: $\hat{F}_i(n) = \sum_{k \geq n} \binom{k}{n} (D/D_i)^n (1 - D/D_i)^{k-n} F_i(k)$. Having the distribution of the number of packets arrive from source *i* in a basic frame period the distribution of the total number of packets is calculated by convolution for all sources: $\hat{F}(n) = \underset{i \in sources}{\otimes} \hat{F}_i(n)$

The implementation of the numerical convolution is usually difficult, computationally intensive task. In our analysis this step of the computation is quite quick, since we do not need to sum too many random variables and their support are very limited (e.g., the RT voice source takes 2 potential value, and the RT data takes only one value for their frame period).

Finally, the virtual waiting time distribution is calculated based on $Q_D^N(x)$ as $Q_D(x) = \sum_N Q_D^N(x)\hat{F}(N)$.

To consider the overhead/multiplexing gain of the AAL2 layer we applied a further approximation. The approximation of the effect of Timer_CU = 0 is based on the idle buffer probability that is $P_{idle} = Q_D(0)$. The approach discussed below can also be used for the approximation of the Timer_CU > 0 case based on P_{idle} and the idle time distribution of the buffer.

When Timer_CU = 0, we assume that a packet fragment smaller than an ATM cell is transmitted in an whole ATM cell if the buffer is idle when the fragment needs to be transmitted, and it is fully multiplexed with other small fragments otherwise. To calculate the number of ATM cells used to transmit user data packets we round up the packet fragment of size *a* (*a* times the ATM cell payload) to an ATM cell with probability P_{idle}. With probability $1-P_{idle}$ this packet fragment generates only *a* ATM cell.

In case of Timer_CU > 0 the P_{idle} value is replaced with the probability that the buffer is idle and the idle time is larger than Timer_CU. The idle time distribution of an N*D/D/1 queue is provided, e.g., in [4].

For the calculation of the virtual waiting time considering the multiplexing gain of AAL2 layer (defined by Timer_CU) the P_{idle} probability is needed. We applied an iterative procedure starting from the worst case to calculate the virtual waiting time. The initial worst case assumption is that all used data fragments generate a complete ATM cell. In each steps of the iterative procedure we refine the packet arrival distribution according to the calculated multiplexing gain. The finally obtained virtual waiting time distribution $\hat{Q}_D(x)$ is the base of the subsequent calculations.

3.2. Calculation of VCC traffic parameters

Two sets of parameters are calculated associated with VCCs, the AAL2 layer packet delay and the traffic parameters of the outgoing ATM stream.

3.2.1. Calculation of AAL2 layer packet delay

First of all it should be noted that in this step we calculate only the AAL2 queuing delay and the fragmentation delay that comes from the shaping of the packet fragments has to be considered additionally. Based on the virtual waiting time distribution $\hat{Q}_D(x)$ the mean and the second moment of virtual waiting time are calculated as $E(T) = \int_x \hat{Q}_D(x)dx$ and $E(T^2) = \int_x x\hat{Q}_D(x)dx$, and the squared coefficient of variation of the waiting time distribution is $cv_T = \left(E\left(T^2\right) - E^2(T)\right)/E^2(T)$.

3.2.2. Calculation of VCC stream parameters

The ATM cell stream transmitted through the outgoing VCC is considered to be an on-off traffic process. As long as the AAL2 buffer is idle there is no ATM cell transmission and when the buffer is busy ATM cells are transmitted at the rate of the VCC capacity. To characterize the traffic parameters of this outgoing traffic we calculate the mean idle and busy time of the AAL2 buffer [4]. Let A be the mean number of packets arrives to the buffer during the basic frame period. $A=\sum_n n\hat{F}(n)$. The on time of the outgoing stream, that is the mean busy time of the AAL2 buffer, is given [4] as $T_{on}=D/(D-A+1)$, while the mean off time of the outgoing stream, that is the idle time of the AAL2 buffer, is $T_{off}=D(D-A)/A(D-A+1)$, where A, D, T_{on} and T_{off} are in cell transmission time unit.

We calculate the first two moment of the inter-arrival time of the outgoing ATM cell stream as $E(T_A)=(T_{on}+T_{off})/T_{on}$ and $E(T_A{}^2)=(T_{off}{}^2+T_{on}+2\cdot T_{off})/T_{on}$. These are the parameters that we refer to as the 'two parameter' description of the outgoing ATM cell stream and they are used in the ATM layer analysis of the RNS.

4. ATM layer behavior and its model

The ingress node functions are modeled and analyzed in the previous section. In this section we analyze the core node functions of the UTRAN. We assume that the transmission layer of the UTRAN is a homogeneous (in the sense that the network is composed by identical node) ATM network and we perform a network wide analysis of this ATM network in this step.

The applied modeling approach is commonly referred to as queuing network (QN) model, but the application of this general modeling approach requires specific considerations about the modeled system. These considerations are essential to generate the QN model of a given telecommunication systems, because different considerations result in completely different QN models of the same telecommunication system.

Our basic assumption on the ATM switches of the RNS is that their bottleneck are the capacity of the outgoing links, i.e. the cells are queued only at the output buffers of the outgoing links. Our assumption is supported by the common opinion that the main bottleneck of radio access networks is the transmission capacity (in contrast with backbone network). Hence we model only the output buffer of the ATM switches and ignore any other queuing delay in the ATM switch (such as the input buffer and the central buffer delay). Further more, we assume that the buffer is large enough and the network is designed properly, such that there is no cell loss at the output buffer. In the QN model it means infinite buffer at the queues.

There are several approximation methods to evaluate the performance parameters of QNs. Some of these approximations allow calculating the first two moments of the performance parameters based on the 'two parameter' description of the QN model. Since our goal is to characterize the first two moment of the end to end delay we applied one of these 'two parameter' approximation methods

referred to as Queuing Network Analyzer (QNA) [5]. This widely used method completes the analysis of the QN as it is provided in [5], hence we focus only on the generation of the QN model and the post-processing of the results. To define the QN model we need to define its traffic parameters and topology.

4.1. QN traffic parameters

The 'two parameter' description of the traffic enters the QN nodes from outside are characterized based on the AAL2 layer behavior as discussed in the previous section. The service process of the queues is quite simple to determine due to the fixed size ATM cells. We have a deterministic service time in each node (i.e., the cv of the service time is 0) and the service time is calculated based on the outgoing link capacity. The service time is the time to transmit 53 bytes through a link of the given capacity. A common feature of QN models is the 'random' routing of the packets between the consecutive queues. The traffic routing between the queues is described by a so called routing matrix, whose *i,j* element gives the probability that the packet leaves queue *i* joints queue *j* next. The traffic routing of UTRANs might be very simple because its topology is often tree structured. It results that the routing matrix of the QN model contains a lot of 0 and 1. Further details on the routing matrix is provided with the QN topology.

4.2. QN topology

The QN model of an RNS is a network of queues that are not necessarily identical with the nodes of the RNS. Instead, based on our output buffer assumption, a queue is associated with all output buffer of the ATM switches.

The core network of the considered RNS is composed by directed transmission links and ATM switches. Due to the simple topology (usually tree structured) of UTRANs the network traffic between given node pairs in a given direction can be easily mapped to the transmission links. From stochastic point of view the network traffic that goes through different links can interact each other only at the network nodes. But based on the assumption that the bottleneck of network nodes is the capacity of the outgoing link (and not the switching fabric) the link independent traffic relations are assumed to be stochastically independent as well. Utilizing this stochastic independence we can decompose the network level analysis. We introduce independent QNs for the up link and down link traffic.

4.2.1. QN model of up link traffic

Based on the considered tree structure of the RNS only traffic multiplexing is at the ATM switches of the up link direction and there is only one outgoing link at each switch. The multiplexing of independent traffic steams is based on the QNA method [5]. The single outgoing link of the ATM switches means that we have as many nodes of the RNS as many queues in its QN model. It is a

special case when the nodes of the RNS are mutually univocally related with the queues of the QN model. We assume that the traffic streams are not delayed at the destination node. Hence the service time of the destination node is set to infinitesimally small. A simple example of the up link direction and its QN model is shown in Figure 3.

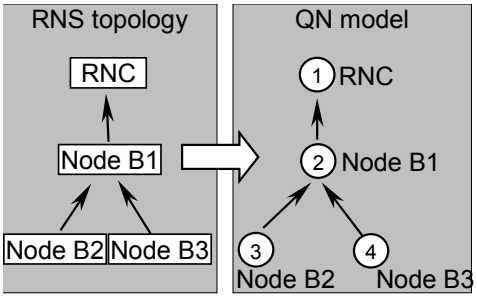

Figure 3 QN model of the up link direction

4.2.2. QN model of down link traffic

In the down link direction there might be more than one outgoing links of the ATM switches, hence we might need to associate more than one queue with an RNS node. The difficulty of representing the traffic behavior of the RNS with a QN model can be demonstrated with the simple example in Figure 4.

The 2 output directions of Node B1 are modeled with 2 buffers. But there is only one link from the RNC to Node B1, while it seems that the QN model represents 2 links. It is not the case. The topology of the QN model talks about the direction of packets through the network, which is different from the RNS topology. It can be interpreted as the output buffer and the output link of the RNC is modeled by queue 1 of the QN model. Similar to the up link case the destination nodes do not delay the packets.

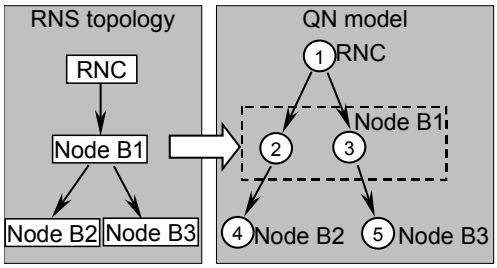

Figure 4 QN model of the down link direction

There is an additional difficulty in modeling the down link direction. If there are more than one output directions at the RNC the QN model splits into independent parts as it is demonstrated in Figure 5.

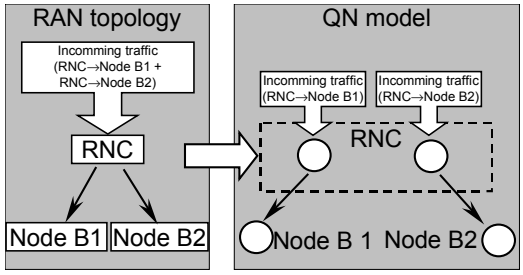

Figure 5 QN model of the RNC with multiple output directions

5. Numerical example

To demonstrate the applicability and the numerical properties of the proposed numerical method a simple example network, shown in Figure 6, is analyzed. Particularly, the up link delay of the Node B2 – RNC relation is evaluated.

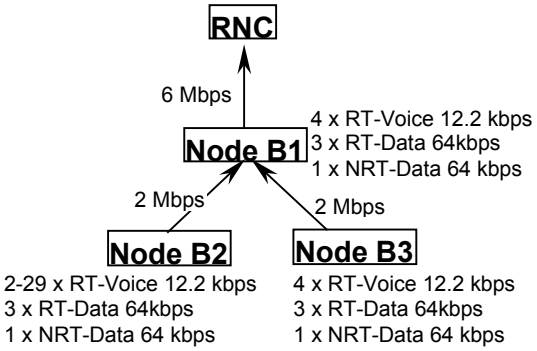

Figure 6 Example network

The traffic scenario is similar in each Node B, i.e., there are 4 voice (12.2 kbps), 3 RT-Data (64 kbps) and 1 NRT-Data connections (64 kbps) to the RNC. The number of the voice connections in Node B2 has been incremented from 2 to 29 connections. One VCC is established for the up link traffic at each Node B. The mean ATM and AAL2 layer delay and their standard deviation have been calculated for the Node B2 and RNC relation. The example network with the given traffic configuration was evaluated also by simulation.

In **Error! Reference source not found.** and **Error! Reference source not found.** the mean and the variance of the AAL2 delay is depicted as a function of the number of voice connection from Node B2 to RNC. As it is expected the approximate analytical results are above the corresponding simulation results due to the previously discussed nature of the approximation. The error of the mean delay is less than 10%. The error of the mean AAL2 delay predicts a higher error in the variance as it can be seen in Figure 8.

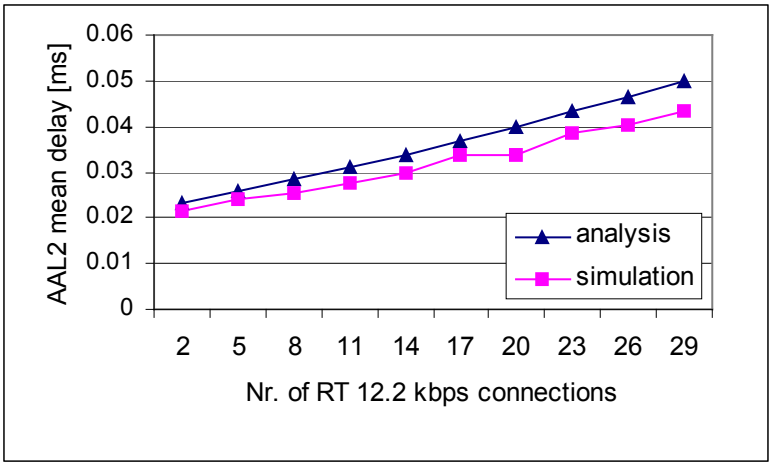

Figure 7 **AAL2 delay - mean value**

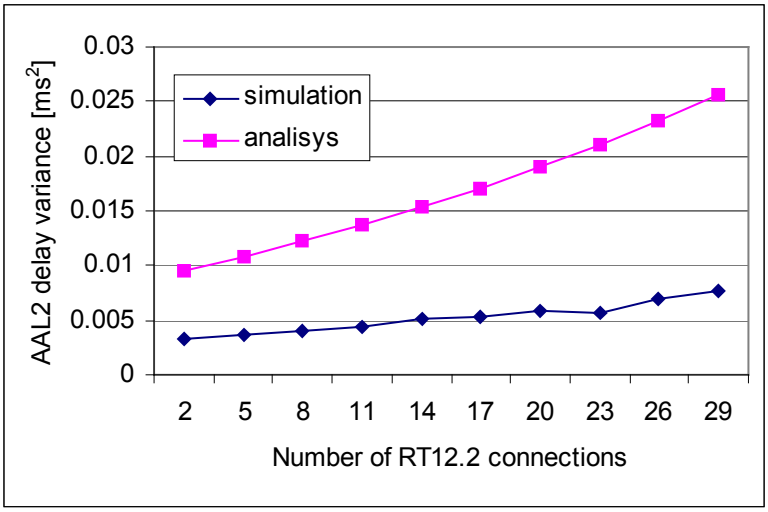

Figure 8 **AAL2 delay variance**

Section 3.1.1 discusses the role of the basic frame period. To investigate its effect on the quality of the approximation Figure 9 and 10 depicts the approximate mean and variance of the AAL2 delay as a function of the basic frame period. The applied numerical method is quite insensitive to the basic frame period in the range of 5 to 20 ms, but the approximation gets incorrect as the basic frame period decreases below 5 ms. It is because the probability that more cells arrive than can be transmitted in a basic frame period gets significant.

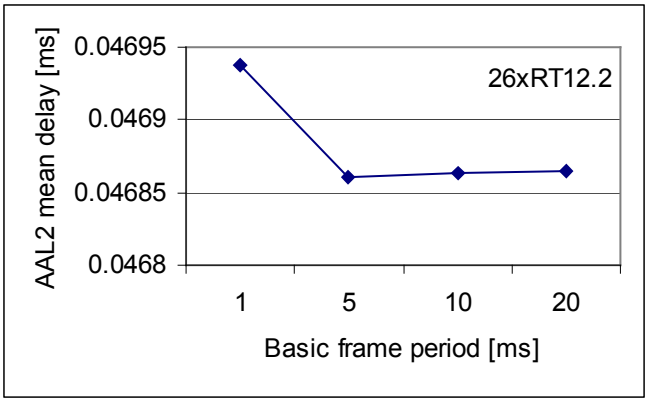

Figure 9 Calculated mean AAL2 delay as a function of the basic frame period

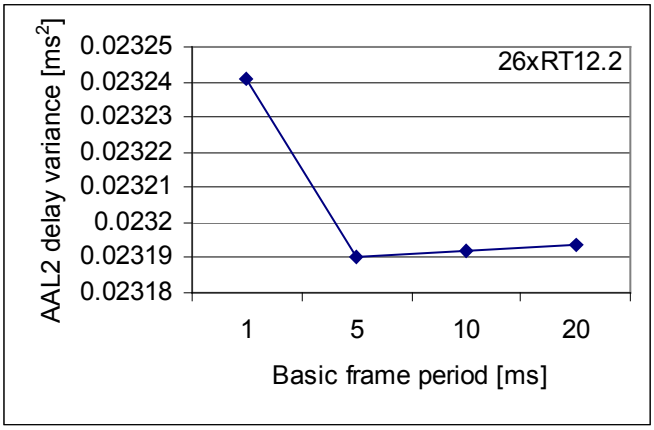

Figure 10 Calculated AAL2 delay variance as a function of the basic frame period

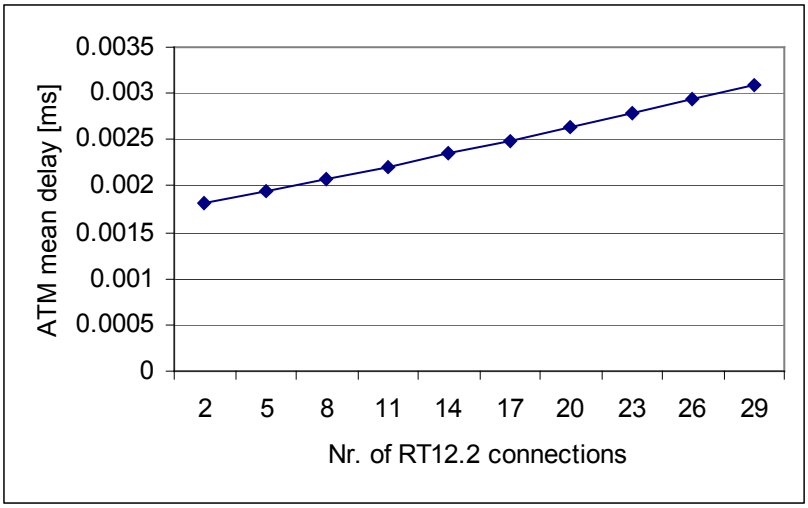

Figure 11 Calculated mean ATM delay as a function of the load

Figure 11 provides the mean ATM delay. The figure suggests that the ATM delay plays less significant role in small networks like the considered example, because only few streams are multiplexed.

At the considered range of the traffic load the ATM delay is an order of magnitude lower than the AAL2 delay. Hence, we did not depict the overall end-to-end delay of the Node B2 – RNC relation because it is practically identical with the AAL2 delay depicted in Figure 7 and 8.

Finally we need to mention the difficulties of the analysis of the considered example using simulation. The difficulty comes by the wide rage of time scales that needs to be considered in the analysis. E.g., the NRT traffic sources transmit packets in 10 ms frames, while the average on period of a voice connection is 3 s. The precise simulation of this system requires the evaluation of at least 60 s of model time. We found that the accurate overall simulation analysis of our small example network including the AAL2 and ATM layer behavior was not possible in a feasible running time.

In contrast, it is a valuable feature of the proposed numerical method that it practically insensitive to the range of time scales. The complexity of the packet size distributions and the number of traffic sources determine the computational complexity of the analysis method.

6. Conclusion

The paper presents a stochastic model of UTRAN with different service classes and traffic behaviors. Several technical details of ATM based UTRAN are considered in the proposed model with high accuracy.

Based on the introduced stochastic model a numerical method is developed for the approximate analysis of the end to end delay. The proposed numerical method has a low computational complexity and it provides reasonable accurate results with negligible response time. The proposed method is a promising alternative of discrete event simulation when negligible response time is required. A numerical example demonstrates the applicability of the proposed method.

References

[1] Technical Specification Group (TSG) RAN; Delay Budget within the Access Stratum TR 25.932 V1.0.0 (2000-05)

[2] TSG services and System Aspects; QoS Concept and Architecture 3G TG 23.107 V3.3.0 (2000-06)

[3] UMTS 30.03 ver. 3.2.0 TR 101 112 (1998-04)

[4] Broadband network teletraffic. Cost 242 final report, Eds.: J Roberts, U Mocci, J. Vitarmo, Springer, LNCS 1155, 1996.

[5] W. Whitt: The queuing network analyzer Bell system technical journal, vol. 62/9, 1983.