Budapest University of Technology and Economics

Resource Management Problems in Packet Switched Networks

Szabolcs Malomsoky
High Speed Networks Laboratory
Department of Telecommunications and Telematics

Summary of the Ph. D. Dissertation

Supervisor:
Dr. Edit Halász
High Speed Networks Laboratory
Department of Telecommunications and Telematics

Budapest, Hungary
2003

Contents

1 Introduction **2**

2 Research Objectives **3**

3 Methodology **4**

4 New Results **4**

5 Application of New Results **22**

6 Acknowledgments **22**

7 Summary of Theses **23**

8 References **25**

9 Publications **27**

1 Introduction

As the development of packet switched network infrastructures advances, managing the resources of these networks becomes inevitable. Resource management in packet switched networks differs substantially from that of traditional telephone networks, because a wide range of applications and services with diverse traffic characteristics and quality of service (QoS) requirements have to be transported. The task of resource management typically includes the solution of a complex optimization problem: the fraction of packets that get lost or suffer from unacceptable delays must be kept below a given threshold, while the amount of used resources should be minimized.

In the 80-90s, the ATM (Asynchronous Transfer Mode) technology was developed to cope with the new requirements on resource management. ATM has been equipped with the basic tools of resource management, such as resource allocation and service differentiation, for example [1]. On the other hand, the dramatic growth of the Internet in the late 90s made it clear, that developing a general framework of QoS support and resource management for the Internet can result in a more widely used option of multi-service networking.

Although the original goal of ATM, namely to provide a global, multi-service network, did not prove to be viable, ATM is used in the transport infrastructures of backbone networks. One example is the land-line transport network of UMTS [2, 3] (Universal Mobile Telecommunication System). UMTS may be regarded as the future successor of today's GSM (Global System for Mobile): the objective with UMTS is to enhance the capabilities of current cellular systems to be able to deliver high quality multimedia content, to become connected to the Internet, etc. [4, 5]. WCDMA (Wideband Code Division Multiple Access) [6] is the radio interface of UMTS. WCDMA essentially operates with so called radio access bearers (RAB), which are basically packet switched radio connections with dedicated resources. Due to requirements of user mobility and radio interface timing, the major QoS requirements on these connections are fast connection set-up and low packet delay, respectively. To support the bandwidth efficient transmission of low bit-rate, delay-sensitive applications (typically conversational voice over an ATM network), ITU-T has standardized a new ATM Adaptation Layer, AAL2 [7, 8, 9]. Also fast AAL2 connection set-up has been an important goal of AAL2 protocol design, therefore the ATM/AAL2 infrastructure is a suitable transmission technology for the UMTS Terrestrial Radio Access Network (UTRAN) [10, 11]. The IP based transport infrastructure of UTRAN is currently being specified in 3GPP [12].

Since UTRAN is a connection oriented network (transporting the RAB connections), the strict delay requirement must be ensured by connection admission control (CAC). CAC decides whether a newly arrived connection can be accepted such that the delay requirements of all connections in the system are met.

According to the needs of the WCDMA radio interface, the connections are subject to soft handover. It means that a mobile can communicate with several base stations at the same time. This technique reduces the overall system interference, but puts additional demand on the land-line network. Several questions arise related to this extra traffic load: How large is it? How does it depend on user mobility? How is it possible to minimize it? Such problems could be analyzed by a proper traffic and mobility model for WCDMA systems.

In the backbone network (or core network) connecting access networks with other access networks or external networks, best-effort traffic (typically Internet traffic) is not transported in RAB connections. Therefore, this type of traffic is not subject to CAC. For aggregated best-effort traffic, resource allocation in the backbone network may be solved by a measurement-based, dynamic bandwidth control algorithm, which is able to adapt the allocated bandwidth to actual traffic demand such that the packet loss ratio is kept under a predefined threshold.

2 Research Objectives

The objective of my dissertation is to give answers and solutions to three resource management problems arising in: (1) UTRAN transport networks, (2) WCDMA systems, and in (3) general ATM backbone networks.

- In the second Chapter, a connection admission control algorithm is proposed, which is applicable in the UTRAN Iub interface connecting base stations and radio network controllers.
- In the third Chapter, a mobility and traffic model for WCDMA systems is presented, which can be used to analyze the additional traffic load on the access network, which is caused by soft handovers.
- In the fourth Chapter, a real-time bandwidth control algorithm is presented, which can be used in ATM backbone networks to dynamically adjust the bandwidth of constant bit rate VPs (Virtual Paths) carrying best-effort traffic, such that the loss ratio of ATM cells is kept under a predefined level.

3 Methodology

The results of Thesis 1 were obtained through analytical methods. For the validation of the queueing model computer simulation was used. Admissible regions provided by the algorithm were also compared to simulation results.

Thesis 2 provides a numerical method for calculating several connection level performance measures in UTRAN. The results were obtained through modeling and analytical considerations.

The results of Thesis 3 are founded on analytical basis. However, the aim was to propose practically tractable methods, therefore different simplifying assumptions and approximations were needed. The validation of the bandwidth control algorithm was performed using simulated as well as measured traffic traces.

4 New Results

Thesis 1 : Connection admission control in UTRAN [J1, J2, J3, J6, P2, P4]

While performance and resource management of the WCDMA radio interface is thoroughly discussed in the literature (see e.g., [6]), few work is dedicated to the performance evaluation, traffic control and provisioning of the transmission infrastructure of UTRAN. Up to the knowledge of the author, a CAC algorithm for the Iub interface, which works in the multi-service scenario and fulfils all practical requirements (e.g., on limited computational complexity and high precision) has not yet been presented. In this thesis, I present a connection admission control algorithm, for which background information and modeling details are given below.

The algorithm makes its decisions based on traffic descriptors and QoS parameters. In UTRAN, the traffic of a connection can be approximated by a periodic ON-OFF model (see Figure 1), and described by three parameters: the packet size b, the packet inter-arrival time TTI (transmission time interval) and the activity factor α, which is a number between 0 and 1, and can be interpreted as the average ratio of the ON state compared to the sum of the average ON and OFF period lengths. The sources can be classified based on the traffic descriptors and the QoS parameters. The number of traffic classes is denoted by K. The number of class i connections in the system is N_i. The traffic descriptors of class i are the following: b_i, TTI_i and α_i. The constant rate of the server (the capacity) is C. The packet scheduling principle is FIFO (first-in-first-out).

The CAC algorithm should ensure for each class that the probability of that a packet is delayed more than its target delay is kept below a small target value:

$$\Pr\{D_i > \tilde{D}_i\} \leq \tilde{\varepsilon}_i \quad i = 1, \ldots, K,$$

where D_i is a random variable describing the sum of the times spent by the packet in the queue and in the server. It will be referred shortly as "delay". $\tilde{D}_i$ is the delay requirement, $\tilde{\varepsilon}_i$ is the allowed maximum probability of delay requirement violation.

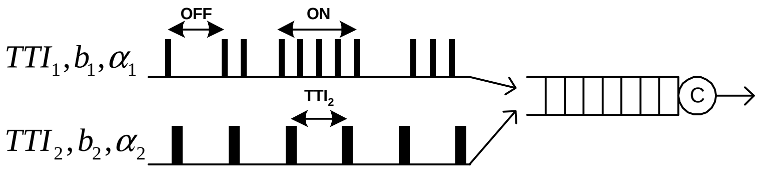

Figure 1: A queueing system with two periodic ON-OFF sources

According to the UTRAN system model, a queueing model is needed that is accurate if the delay requirements are strict (5-15 ms) and the buffer size is small (e.g., smaller than 20 ms). The queueing model of this section is developed for these conditions.

The RAB connections are modeled with independent ON-OFF sources. The ON and OFF periods are bursty, meaning that typically both are many TTI long. The long term correlation characteristics of the arrival process could not be taken into account, because traffic descriptors do not contain any information on the correlation structure of the sources, and it was also not possible to get information on this by measurements. It is not a problem in practice, because the buffer is small enough such that it fills up quickly even during a temporary overload. Therefore we can assume that the ON-OFF burst component of the queue is negligible. The system is temporarily overloaded, if so many connections are temporarily in ON state at the same time, that the server can not serve within one TTI the packets arriving in one TTI. I used the approximation that all packets arriving during a temporary overload situation violate the delay requirement.

If the server can serve during one TTI all packets that arrive within one TTI (i.e., when the overall arrival rate remains below multiplex capacity), the system behaves like the so called $\sum_i D_i/D/1$ queue: a superposition of independent periodic sources of possibly different periods and independent phases is offered to a multiplexer with a deterministic server. The time between two packet emissions of a class i stream is equal to TTI_i, and its phase with respect to a common time origin is chosen at random between 0 and TTI_i.

According to the assumptions and approximations used when setting up the queueing model, the non-overloaded and the overloaded states were examined independently, and the $\tilde{\varepsilon}_i$ target probability was divided into two parts:

$$\tilde{\varepsilon}_i = \tilde{\varepsilon}_i^{delayed} + \tilde{\varepsilon}_i^{overload},$$

where $\tilde{\varepsilon}_i^{delayed}$ is the maximum allowed probability of that the delay requirement is violated due to temporary packet congestion in the non-overloaded system, and $\tilde{\varepsilon}_i^{overload}$ is the maximum allowed probability of that the delay requirement is violated due to temporary system overload.

Thesis 1.1 : Connection admission control algorithm

By using the queueing model introduced for UTRAN, I have developed a connection admission control (CAC) algorithm, which can be efficiently used in the Iub interface of UTRAN and handles all RAB types. Based on the input parameters $\{\ b_i,\ TTI_i,\ \alpha_i,\ \tilde{D}_i$ *and* $\tilde{\varepsilon}_i,\ i = 1, \ldots, K\ \}$ *it decides whether a newly arrived connection can be accepted into the system.*

The block-diagram of the algorithm is shown in Figure 2. When a new connection arrives, the algorithm checks:

(**A**) the probability of delay requirement violation in non-overloaded system due to temporary packet congestion ($\varepsilon_i^{delayed}$), and

(**B**) the probability of delay requirement violation due to temporary system overload ($\varepsilon_i^{overload}$).

The connection gets accepted if these probabilities for all traffic types are below their respective thresholds: $\varepsilon_i^{delayed} < \tilde{\varepsilon}_i^{delayed}$, and $\varepsilon_i^{overload} < \tilde{\varepsilon}_i^{overload}$, $i = 1, \ldots, K$.

The admissible region is approximated by the intersection of a number of regions with linear borders (also referred to as hyper-planes) and one region with (generally) non-linear border. A region bounded by a hyper-plane contains the mixes, where the delay requirement of a certain class corresponding to **A** is fulfilled. The region with the non-linear border contains the mixes, which can overload the queueing system only with a small probability (requirement corresponding to **B**).

An example is depicted in Figure 3, where there are two classes. The first service has significantly stricter delay requirement, therefore the hyper-plane

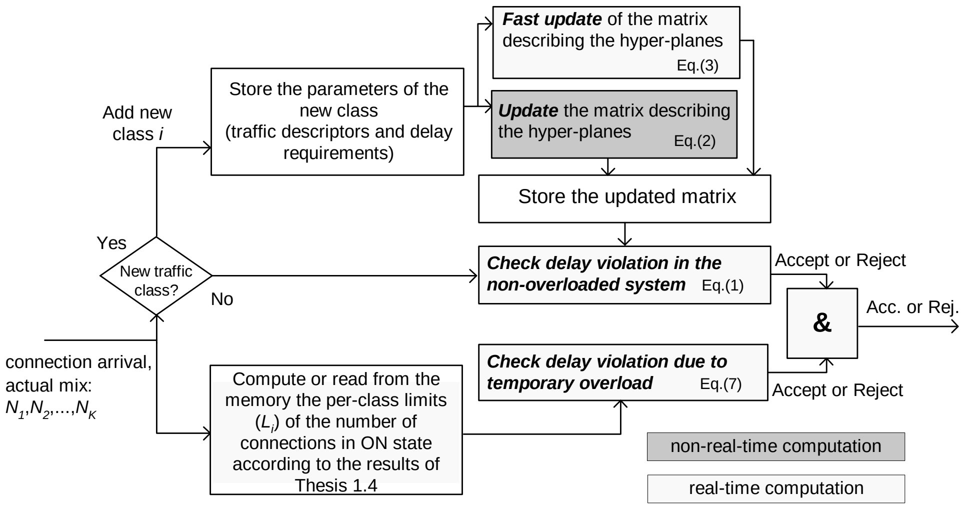

Figure 2: Block-diagram of the algorithm.

corresponding to the second service (which is not depicted) would be outside both depicted regions. In the shaded area, mixes containing connections from both classes can be accepted. In this figure, some mixes are within the delay-limited region, but outside the overload-limited region, because it is possible that at 100% utilization the delay-limit is not yet exceeded. If the activity factor of the first class is less than 1, then the maximum number of connections from the first class given that there are no connections in the system from the second class, $N_{1,max}$, increases to $N^*_{1,max}$ and the border of the overload limited region becomes non-linear.

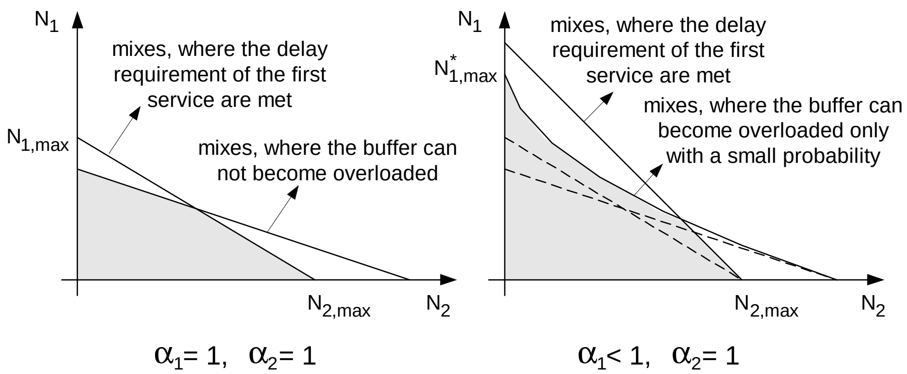

Figure 3: Construction of admissible region approximation

In the algorithm, the computation of the two types of regions is separated. Also the update (non-real-time) and decision-making (real-time) parts are separated. If a connection of a new traffic class (not yet included in the matrix describing the hyper-planes) arrives, a fast (real-time) but more approximate update of the matrix is performed.

Thesis 1.2 : Checking delay violation in the non-overloaded system

I have proposed the hyper-plane approximation for checking delay violation due to temporary packet congestion in the non-overloaded system.

Applying the hyper-plane approximation, the necessary condition of accepting the traffic mix $(N_1, N_2, \ldots, N_K)$ is

$$\sum_{i=1}^{K} \frac{TN_{jj}}{TN_{ij}} \cdot N_i \leq TN_{jj} + 1 \qquad j = 1, 2, \ldots, K, \tag{1}$$

where TN_{ij} is defined as the maximum number of connections from class i assuming that a *single packet* from class j would fulfil the delay requirement of class j. I have approximated by TN_{ij} the maximum number of class i connections if *one additional connection* from class j is present in the system. The proposed formula for determining TN_{ij} values is the following:

$$TN_{ij} = \max\left\{ N_i \;\middle|\; \sum_{n_i=0}^{N_i} \Pi_i(n_i)\ \Pr\left\{ D_i > \tilde{D}_j - \frac{b_j}{C} \;\middle|\; N_i^{act}(t_0) = n_i \right\} \leq \tilde{\varepsilon}_j^{delayed} \right\}, \tag{2}$$

where $\Pi_i(n_i)$ is the probability that n_i connections of class i are active (are in ON state) at the same time. $N_i^{act}(t_0)$ is the number of active class i connections at (an arbitrary) time t_0. $\tilde{\varepsilon}_j^{delayed}$ is the allowed maximum probability of that packets from class j get delayed more than the target maximum delay $\tilde{D}_j$ due to temporary packet congestion in the non-overloaded system.

I have validated the hyper-plane approximation for UTRAN-specific parameters using simulation and mathematical analysis. For connections with the same TTI, I have shown that the delay-limited part of the constructed admissible region (i.e., the hyper-planes) can be approximated by Eq.(4) in Thesis 1.4, which is practically linear in N_i. I have also shown that having different TTIs for different classes does not invalidate the hyper-plane approximation. In this case, the complementary distribution function of the virtual waiting time can not be obtained analytically [13]. Therefore, I have given simple upper and lower bounds and I have shown that (1) for strict delay requirements (i.e., as $x \to 0$ in Eq.(4)) the difference between the bounds disappears and (2) for loose delay requirements ($x \to \min_i TTI_i$) it is already the region with the non-linear border that constrains the admissible region.

Thesis 1.3 : A closed form approximation for calculating the values of the TN matrix

I have given a closed form approximation for calculating the values of the TN matrix:

$$TN_{ij} \approx \frac{C\ (TTI_i + \alpha_i\ y)}{\alpha_i\ b_i} \left(1 - \frac{b_i\ \ln\left(\tilde{\varepsilon}_j^{delayed}\right)}{2\ C\ y}\right)^{-1}, \quad y = \tilde{D}_j - \frac{b_j}{C}. \tag{3}$$

I have approximated the arrival process by a Gaussian process (Brownian bridge), and showed that for this arrival process, assuming that all classes are characterized by the same packet inter-arrival time (TTI), the delay-limited part of the admissible region can be approximated as:

$$\sum_i N_i \left(\alpha_i\ \rho_i + \frac{\alpha_i\ \rho_i^2}{C}\ \frac{\gamma\ TTI}{2\ x}\right) \leq C + \frac{C\ x}{TTI}\ \frac{\sum_i N_i\ \alpha_i^2\ \rho_i^2}{\sum_i N_i\ \alpha_i\ \rho_i^2}, \tag{4}$$

where $\rho_i = b_i/TTI_i$ is the bit rate of an active class i connection, x is the delay criterion, $\gamma = -\ln(\tilde{\varepsilon}_i^{delayed})$, and $\tilde{\varepsilon}_i^{delayed} = \tilde{\varepsilon}_j^{delayed}$ for all $i, j = 1, \ldots, K$. From this approximation Eq.(3) directly follows.

Using simulations I concluded that the accuracy of approximation Eq.(3) is acceptable even if the number of multiplexed connections is low and thus the continuous approximation of the arrival process is less accurate.

Thesis 1.4 : Checking delay violation due to temporary system overload

Based on the queueing model introduced for UTRAN, I have proposed a fast calculation method for checking delay violation due to temporary system overload. The method exploits the statistical gain only within the different classes. Therefore I have also proposed an extension for this method to take into account at least partially the statistical gains among classes.

Based on the queueing model introduced for UTRAN, I concluded that to check delay violation due to temporary overload I need to find a solution of the bufferless ON-OFF model (see e.g., [14]), which can be efficiently used in the CAC. Unlike existing methods, I used a method that enables to carry out the necessary computations independently for the traffic classes. Based on numerical investigations I concluded that the method introduced below performs well for parameters that are typical in UTRAN.

Assume that the new connection arrives from class i. I have introduced L_i, which is the smallest number of class i connections being in ON state such that the requirement on the temporary overload is met. We will also refer to this value as the "per-class limit" of the number of connections in ON state. Its values are calculated as follows:

$$L_i = \min\left\{ L \,\middle|\, \sqrt[K_a]{1-\tilde{\varepsilon}_i^{overload}} \leq \frac{1}{N_i\ \alpha_i} \sum_{k=0}^{L} k\ \Pi_i(k) \right\}, \tag{5}$$

and for the other classes:

$$L_l = \min\left\{ L \,\middle|\, \sqrt[K_a]{1-\tilde{\varepsilon}_i^{overload}} \leq \sum_{k=0}^{L} \Pi_l(k) \right\}, l = 1, 2, \ldots, K_a, l \neq i, \tag{6}$$

where K_a is the number of traffic classes with activity factor smaller than one ($\alpha_i < 1$), $\tilde{\varepsilon}_i^{overload}$ is the target maximum probability that packets from class i get delayed due to temporary system overload, and $\Pi_i(k)$ is the probability that the number of active connections from class i is k. For always active traffic classes ($\alpha_i = 1$), the per-class limit equals the number of connections in the system, i.e., $L_i = N_i$.

It is noted that this method can be implemented efficiently by storing the resulting L_i values in memory (the computational complexity can be reduced to a memory-read).

When a new class i connection arrives to the system, and the actual connection mix becomes $(N_1, N_2, \ldots, N_i, \ldots, N_K)$, one needs to calculate or read from the memory the new L_i value and check the following inequality:

$$\sum_{i=1}^{K} L_i\ \rho_i \leq C, \tag{7}$$

which is the necessary condition of accepting $(N_1, N_2, ..., N_K)$.

It is obvious that using Eq.(5) and Eq.(6), only the statistical gain of multiplexing sources from the same class is exploited. Keeping the property that L_i values can be obtained independently from each other, but taking into account partially statistical gains from multiplexing different classes, I have proposed the following method.

1. Find L_i^* for all i using Eq.(5) and Eq.(6) with

$$N_i^* = N_i + \sum_{\alpha_k \le \alpha_i,\ k \ne i} \min\left(1,\ \frac{\rho_k}{\rho_i}\right)\ N_k; \quad k = 1, \ldots, K, \tag{8}$$

   and calculate the statistical multiplexing gain for class i as:

$$MG_i = (N_i^* - L_i^*)/N_i^*. \tag{9}$$

2. Repeat until MG_i values are no longer increasing:

   - Consider the classes with $\alpha_k > \alpha_i$ and $\rho_k < \rho_i$, $\forall\ i,\ k$. If $MG_k > MG_i$, then let $\alpha_i' := \alpha_k$ and calculate MG_i' executing step 1 with the temporary activity factor value α_i'. If the resulting $MG_i' > MG_i$, then let $MG_i := MG_i'$. (At the end of this step reset the original value of α_i.)
   - Consider the classes with $\alpha_k > \alpha_i$ and $\rho_k \ge \rho_i$, $\forall\ i,\ k$. If $MG_k > MG_i$, then let $MG_i := MG_k$.

3. Finally, $L_i = N_i(1 - MG_i)$ for all i.

The multiplexing gain values, MG_i, $i = 1, \ldots, K$, have been modified in steps 1 and 2 such that also statistical multiplexing effects between class-pairs are taken into account. Therefore, L_i values calculated in step 3 can be smaller than the ones obtained using Eq.(5) and Eq(6).

Thesis 2 : Mobility and traffic analysis for WCDMA networks [C1]

I have extended the model of Nakano, Saita and Sengoku [15] for WCDMA systems. My contribution is that I allow overlapping cells and consider soft handover instead of hard handover. I use the extended model for estimating the user-plane traffic on the WCDMA Iur interface, the soft handover intensities in the cells, etc.

Thesis 2.1 : Model extension

By extending the model presented in [15], I have set up a mobility and traffic model for analyzing WCDMA systems.

My objective was to analyze the UTRAN system on connection-level, such that the results of the analysis are applicable to practical problems (e.g., dimensioning of the transport network). For this, i needed a model, which takes into account user mobility and operates with realistic parameters. I recognized that the model of Nakano [15] fulfils these requirements. However, this model can only be applied for CDMA based systems, if it can be extended to include the modeling of soft handovers.

In the original model that works for hard handovers, cells along the roads of the road system can be analyzed independently, because a call is handled by a single cell at a time, i.e., a handover event results in a cell change. This is not the case with soft handovers, where a call may use the resources of more cells at a time, depending on how the user moves.

As it can be seen in Theses 2.2, 2.3 and 2.4, the model of Nakano can be extended for soft handovers. However, the complexity of the calculations significantly increases.

I have assumed that the road network is covered by *overlapping* cells. I have introduced the notion of the *soft handover regions* (SHR), which are the overlapping areas of cells. Soft handover can happen if a mobile moves to an SHR covered by more cells. Along route r we have soft handover regions:

$$SHR_1^r, \ldots, SHR_j^r, \ldots, SHR_{J_r}^r,$$

listed in order of appearance.

To be able to formalize the connection between SHR and cell parameters, for each SHR_j^r I have defined the *characteristic set* as:

$$\mathcal{Z}_{j,r} = \{\text{cells containing } SHR_j^r\} = \{z_1^{j,r}, \ldots, z_{K_{j,r}}^{j,r}\},$$

where SHR_j^r is overlapped exactly by cells $z_1^{j,r}, \ldots, z_{K_{j,r}}^{j,r}$. I have implemented the model for circle-shaped omni-cells, in that case the number of those cells is $K_{j,r} = 1$, 2 or 3.

Figure 4 gives a picture about the elements of the extended model and the used notation.

Thesis 2.2 : Calculating soft handover region parameters

Using the extended model, I have calculated the following soft handover region parameters:

- I have determined the *soft handover probability of new calls* of class c from SHR_j^r (the probability that a call of class c generated in SHR_j^r successfully reaches the boundary of SHR_j^r and SHR_{j+1}^r):

$$p_c^{new}(SHR_{j+1}^r \mid SHR_j^r) = \frac{h_c}{T_0^{j,r}}(1 - e^{-\frac{T_0^{j,r}}{h_c}}), \quad \text{for } 1 \leq j < J_r, \quad (10)$$

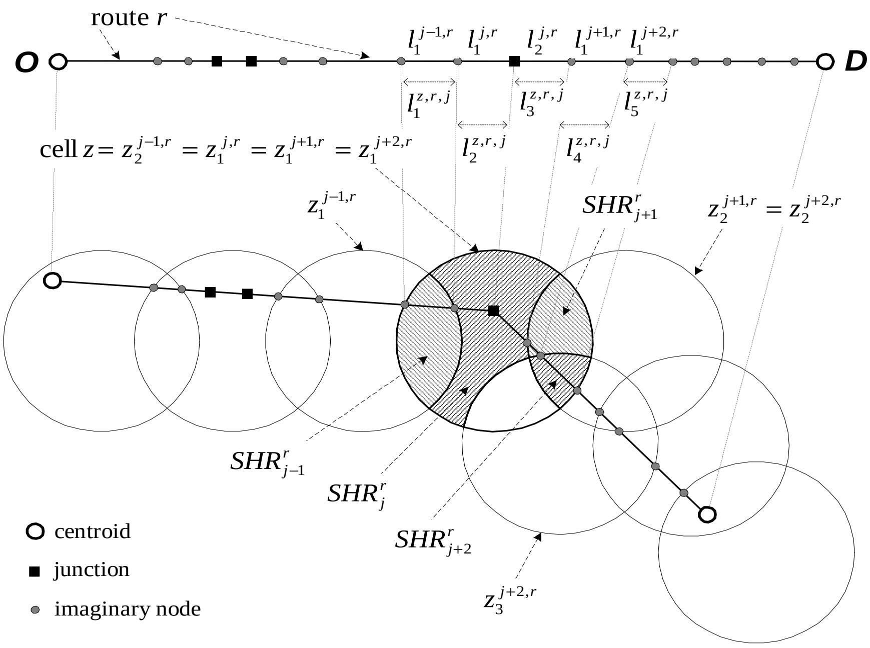

Figure 4: Hierarchy of sections on route r

where $T_0^{j,r}$ is the traveling time through SHR_j^r on route r, and h_c is the mean holding time of a call of class c.

- I have determined the *soft handover probability of soft handover calls* of class c from SHR_j^r, i.e. the probability that while moving on route r the call of class c enters SHR_j^r (so it is "handed over" from SHR_{j-1}^r to SHR_j^r) and successfully reaches the boundary of SHR_j^r and SHR_{j+1}^r (it is "handed over" to SHR_{j+1}^r):

$$p_c^{ho}(SHR_{j+1}^r \mid SHR_j^r) \;=\; e^{-\frac{T_0^{j,r}}{h_c}}, \quad \text{for } 1 < j < J_r. \tag{11}$$

I have determined the boundary conditions for the soft handover probabilities as follows:

$$p_c^{ho}(SHR_2^r \mid SHR_1^r) \;=\; e^{-\frac{d_r}{v(l_1^{1,r})h_c} - \frac{T_0^{1,r}}{h_c}}, \tag{12}$$

where the first link of route r (link - between two junctions, denoted by l in Figure 4) is "lengthened backwards" to reach the boundary of one of the cells of $\mathcal{Z}_{1,r}$, and the lengthened part is d_r long. $v(l_i^{j,r})$ is the speed on link $l_i^{j,r}$ (for the definition of a link see Figure 4).

Calls in $SHR^r_{J_r}$ are not "handed over" to further soft handover regions, therefore:

$$\begin{aligned} p_c^{new}(SHR^r_{J_r+1} \mid SHR^r_{J_r}) &= 0, && (13)\\ p_c^{ho}(SHR^r_{J_r+1} \mid SHR^r_{J_r}) &= 0. && (14) \end{aligned}$$

I have derived the *Z-set call arrival rates* to be able to estimate the inter-RNC traffic (the user-plane traffic on the Iur interface). Z is a set of cells, by definition: $Z \subseteq \mathcal{Z}_{j,r}$. For example, if $\mathcal{Z}_{j,r} = \{1,2\}$, then $Z \subseteq \mathcal{Z}_{j,r}$ means that $Z \in \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$. The Z-set is similar to the "active set" in CDMA systems. The active set contains the radio cells the mobile could be connected to, because the measured pilot signal strength is sufficient. The Z-set contains the cells the mobile is actually connected to.

- I have defined the *Z-set arrival rate of new calls* of class c in SHR^r_j on route r as the proportion of the arrival rate of new calls of class c in SHR^r_j on route r that get connected to the base station set $Z \subseteq \mathcal{Z}_{j,r}$. I calculated this as follows:

  $$^Z\lambda_c^{new}(SHR^r_j) = \lambda_c \; Q_r \; T_0^{j,r} \; {}^Z P_c(SHR^r_j), \quad \text{for each } 1 \le j \le J_r, \quad (15)$$

  where λ_c is the arrival rate of calls of class c, Q_r is the vehicular traffic volume on route r (vehicle/hour), and

  $$^Z P_c(SHR^r_j) = \prod_{z \in Z} \big(1 - B_c(z)\big) \prod_{s \in \mathcal{Z}_{j,r} \setminus Z} B_c(s), \quad (16)$$

  where $B_c(z)$ is the blocking probability of c-type calls in cell z.

- I have defined the *Z-set arrival rate of soft handover calls* of class c entering SHR^r_j (denoted by $^Z\lambda_c^{ho}(SHR^r_j)$) as the proportion of the arrival rate of soft handover calls of class c entering SHR^r_j that get connected to the base station set $Z \subseteq \mathcal{Z}_{j,r}$. It consists of two parts, namely the newly initiated calls of class c in SHR^r_{j-1} and then "handed over" to SHR^r_j, and the calls of class c "handed over" from the previous soft handover region to SHR^r_{j-1} and then "handed over" further to SHR^r_j for $1 < j \le J_r$.

  If SHR^r_j is covered by more cells than SHR^r_{j-1}, then the Z-set can be *unchanged*, if the new cells block the connection setup request, or it can be *extended* with new cells, if at least one of the new cells do not block the request. Using the notations; for $Z \nsubseteq \mathcal{Z}_{j,r} \setminus \mathcal{Z}_{j-1,r}$, if $\mathcal{Z}_{j-1,r} \subseteq \mathcal{Z}_{j,r}$, I calculated the Z-set arrival rate of soft handover calls as follows:

$$\begin{aligned} {}^{Z}\lambda_c^{ho}(SHR_j^r) = \; & p_c^{new}(SHR_j^r \mid SHR_{j-1}^r) \; {}^{Z \cap \mathcal{Z}_{j-1,r}}\lambda_c^{new}(SHR_{j-1}^r) \quad \times \\ \times & \prod_{z \in \mathcal{Z}_{j,r} \setminus \mathcal{Z}_{j-1,r}} \left\{ \big(1 - B_c(z)\big)\mathcal{I}_{\{z \in Z\}} + B_c(z)\mathcal{I}_{\{z \notin Z\}} \right\} \quad + \\ + \; & p_c^{ho}(SHR_j^r \mid SHR_{j-1}^r) \; {}^{Z \cap \mathcal{Z}_{j-1,r}}\lambda_c^{ho}(SHR_{j-1}^r) \quad \times \\ \times & \prod_{z \in \mathcal{Z}_{j,r} \setminus \mathcal{Z}_{j-1,r}} \left\{ \big(1 - B_c(z)\big)\mathcal{I}_{\{z \in Z\}} + B_c(z)\mathcal{I}_{\{z \notin Z\}} \right\}. \end{aligned} \tag{17}$$

The Z-set can not be *entirely changed* to a new set, i.e., for $Z \subseteq \mathcal{Z}_{j,r} \setminus \mathcal{Z}_{j-1,r}$, if $\mathcal{Z}_{j-1,r} \subseteq \mathcal{Z}_{j,r}$, I got:

$$ {}^{Z}\lambda_c^{ho}(SHR_j^r) \; = \; 0. \tag{18}$$

If SHR_j^r is covered by less cells than SHR_{j-1}^r, then the Z-set can be reduced or emptied. If the Z-set is *reduced but not emptied*, i.e., for $Z \neq \varnothing$, if $\mathcal{Z}_{j-1,r} \supseteq \mathcal{Z}_{j,r}$, I got:

$$\begin{aligned} {}^{Z}\lambda_c^{ho}(SHR_j^r) = \; & p_c^{new}(SHR_j^r \mid SHR_{j-1}^r) \sum_{Z \subseteq S \subseteq \mathcal{Z}_{j-1,r}} {}^{S}\lambda_c^{new}(SHR_{j-1}^r) \quad + \\ + \; & p_c^{ho}(SHR_j^r \mid SHR_{j-1}^r) \sum_{Z \subseteq S \subseteq \mathcal{Z}_{j-1,r}} {}^{S}\lambda_c^{ho}(SHR_{j-1}^r). \end{aligned} \tag{19}$$

If the Z-set is *emptied*, i.e., for $Z = \varnothing$, if $\mathcal{Z}_{j-1,r} \supseteq \mathcal{Z}_{j,r}$, I got:

$$\begin{aligned} {}^{\varnothing}\lambda_c^{ho}(SHR_j^r) = \; & p_c^{new}(SHR_j^r \mid SHR_{j-1}^r) \quad \times \\ \times & \sum_{\varnothing \neq S \subseteq \mathcal{Z}_{j-1,r} \setminus \mathcal{Z}_{j,r}} {}^{S}\lambda_c^{new}(SHR_{j-1}^r) + p_c^{ho}(SHR_j^r \mid SHR_{j-1}^r) \quad \times \\ \times & \sum_{\varnothing \neq S \subseteq \mathcal{Z}_{j-1,r} \setminus \mathcal{Z}_{j,r}} {}^{S}\lambda_c^{ho}(SHR_{j-1}^r). \end{aligned} \tag{20}$$

For the completion of this recursive formula I have given the initial condition as:

$$ {}^{Z}\lambda_c^{ho}(SHR_1^r) \; = \; {}^{Z}\lambda_c^{0}(SHR_1^r) \frac{Q_r}{\sum_{\text{route } h \text{ starts in } SHR_1^r} Q_h}, \tag{21}$$

for $Z \subseteq \mathcal{Z}_{1,r}$, where ${}^{Z}\lambda_c^{0}(SHR_1^r) = \lambda_c^{0}(p_1^r) \; {}^{Z}P_c(SHR_1^r)$ (here p_1^r is the first centroid of route r). (I have assumed that the arrival rate of calls coming from outside the target area to SHR_1^r is divided among the routes starting in the particular soft handover region in the proportion of the vehicular traffic volumes on these routes.)

- I have calculated the *Z-set traffic load* for $Z \subseteq \mathcal{Z}_{j,r}$ for the call class c for each soft handover region SHR_j^r (defined as the proportion of the traffic load that is induced by the calls that get connected to the base station set $Z \subseteq \mathcal{Z}_{j,r}$) as:

$$
\begin{aligned}
{}^{Z}Load_c(SHR_j^r) = & \qquad (22)\\
& {}^{Z}\lambda_c^{new}(SHR_j^r)\ h_c^{new}(SHR_j^r) + {}^{Z}\lambda_c^{ho}(SHR_j^r)\ h_c^{ho}(SHR_j^r),
\end{aligned}
$$

  and thus *the l-leg offered traffic load for any area* (some soft handover regions together) can be calculated as the sum of those Z-set offered traffic loads of soft handover regions in the area for which $|Z| = l$, for $l = 0, 1, 2, 3$.

Thesis 2.3 : User-plane traffic estimation at the Iur interface

Using the extended model and the calculated soft handover region parameters, I have estimated the user-plane traffic at the Iur interface.

Let $RNC_1 \subseteq \{1, \ldots, N_{cell}^1\}$ and $RNC_2 \subseteq \{1, \ldots, N_{cell}^2\}$ represent disjunct sets of cells, which contain the cells connected to RNC_1 and RNC_2, respectively. Using Eq.(22), I have obtained a lower bound ($A = 1$ in Eq.(23)) and an upper bound ($A = 2$) for the traffic generated on the Iur interface between the two RNC-s:

$$
\begin{aligned}
Iur^A(RNC_1, RNC_2) = & \sum_{z_1 \in RNC_1, z_2 \in RNC_2} \ \sum_{\mathcal{Z}_{j,r} \ni z_1, z_2} \ \sum_{Z \subseteq \mathcal{Z}_{j,r}} \ \sum_{c=1}^{C} {}^{Z}Load_c(SHR_j^r) \ \times \\
\times & \quad \left\{ \mathcal{I}_{\{z_1, z_2 \in Z,\ |Z|=2\}} + A \cdot \mathcal{I}_{\{z_1, z_2 \in Z,\ |Z|=3\}} \right\}, \qquad (23)
\end{aligned}
$$

where all those Z-set traffic loads of the soft handover regions are summed up, which are both in RNC_1 and RNC_2 (see the first two summations) for sets $Z \subseteq \mathcal{Z}_{j,r}$ that contain a cell of each RNC ($z_1 \in RNC_1, z_2 \in RNC_2$), so the proper 2 and 3-leg traffic loads (see the third summation and the indicator functions) for each call class (see the fourth summation). Note that we need the blocking probabilities here implicitly in the Z-set traffic loads through the Z-set call arrival rates.

Notice that, using the extended model, we could determine the exact value for the traffic on the Iur interface, but it is contagious to formulate it (the 3-leg traffic gives one or two-legged traffic on the Iur interface depending on the route structure in the corresponding SHR).

Thesis 2.4 : Calculating cell parameters

Using the extended model and the soft handover region parameters, I have calculated the distribution of the channel occupancy time, the offered traffic load and the soft handover intensity in a cell.

- I have derived the *distribution of the channel occupancy time* in the cell for the newly initiated calls of class c:

$$Pr\{\tau_c^{z,r,j} < t\} \;=\; 1 - \frac{T_0^{z,r,j} - t}{T_0^{z,r,j}} e^{-\frac{t}{h_c}}, \tag{24}$$

  where $\tau_c^{z,r,j}$ is the channel occupancy time in cell z on the section of route r that contains SHR_j^r, and $T_0^{z,r,j}$ is the traveling time through cell z on the section of route r that contains SHR_j^r. (The index j is needed, because it is possible that route r goes through the area of cell z more than once.)

- I have calculated the *offered traffic load* for a cell, which is composed of all the newly initiated call intensities in the cell, the newly initiated and handover call intensities in the SHR preceding the cell on some route and successfully reaching the boundary of the cell and finally the call arrival rates from outside the target area to the cell each multiplied by the corresponding mean channel occupancy times:

$$\begin{aligned} Load_c(z) = & \sum_{\mathcal{Z}_{j,r} \ni z} h_c^{new}(z^{(r,j)})\ \lambda_c^{new}(SHR_j^r) + \sum_{\mathcal{Z}_{j,r} \ni z \notin \mathcal{Z}_{j-1,r}} h_c^{ho}(z^{(r,j)}) \quad \times \\ \times & \ \big\{p_c^{new}(SHR_j^r|SHR_{j-1}^r)\big(\lambda_c^{new}(SHR_{j-1}^r) - {}^{\varnothing}\lambda_c^{new}(SHR_{j-1}^r)\big) \quad + \\ + & \ \ p_c^{ho}(SHR_j^r|SHR_{j-1}^r)\big(\lambda_c^{ho}(SHR_{j-1}^r) - {}^{\varnothing}\lambda_c^{ho}(SHR_{j-1}^r)\big)\big\} \quad + \\ + & \sum_{\mathcal{Z}_{0,r} \ni z} h_c^{ho}(z^{(r,j)})\lambda_c^0(p_1^r), \end{aligned} \tag{25}$$

  where $z^{(r,j)}$ is the cell identified by the cell identifier z, and (r,j) means that the computation of the mean holding times is done on route r over the section, which is covered by cell z and contains SHR_j^r.

  The offered traffic load and the blocking probability (obtained with a multi-dimensional Erlang-B formula) are not independent of each other, therefore it is not possible to get them explicitly. Therefore, I have solved numerically the related simultaneous system of equations.

- Finally, I have derived the *soft handover intensity* (mean number of handover requests in the cell (to the cell)) that consists of the non-blocked newly initiated calls in the SHR region preceding the cell on some route and the non-blocked soft handover calls in the same SHR that reach the boundary of the cell and the call arrival rates from outside the target area (if the cell is on the boundary of it):

$$
\begin{aligned}
{}^{in}\lambda_c^{ho}(z) = & \sum_{\mathcal{Z}_{j,r}\ni z \notin \mathcal{Z}_{j-1,r}} \big\{ p_c^{new}(SHR_j^r|SHR_{j-1}^r)\big(\lambda_c^{new}(SHR_{j-1}^r) - \\
& - {}^{\varnothing}\lambda_c^{new}(SHR_{j-1}^r)\big) + p_c^{ho}(SHR_j^r|SHR_{j-1}^r)\big(\lambda_c^{ho}(SHR_{j-1}^r) - \\
& - {}^{\varnothing}\lambda_c^{ho}(SHR_{j-1}^r)\big)\big\} + \sum_{\mathcal{Z}_{0,r}\ni z} \lambda_c^0(p_1^r).
\end{aligned} \tag{26}
$$

Thesis 3 : Real-time VP bandwidth control [J4, J5, C5, C6, P8]

I proposed an algorithm, which adjusts dynamically the bit rate of an ATM Virtual Path (VP) such that the cell loss ratio (CLR) in the output buffer associated with the VP remains below a pre-determined threshold. In this algorithm I have applied the buffer monitoring method of Vidács [16], which is able to accurately measure the cell loss ratio in the output buffer of an ATM switch for both short and long-range dependent traffic. For the on-line adaptation of the VP bandwidth I have introduced the notion of state-space representation of a single server queue, and I have applied Bayesian regression analysis to estimate the state variable of that system.

Thesis 3.1 : VP bandwidth control algorithm

I have proposed a VP bandwidth control algorithm, and showed that it converges iteratively to the optimal bandwidth value for both short-range and long-range dependent traffic.

In [16], the following simple approximation has been derived:

$$
\frac{1-\rho}{\rho} \approx d \ \log p_K, \tag{27}
$$

where ρ is the link utilization ($\rho = A/C$, A is the long-term average rate of a stationary process, and C is the constant bit rate of the VP), and d is a coefficient, which is constant given that the utilization and the buffer occupancy probability, p_K, do not change with time.

In the bandwidth estimation algorithm, the utilization and the buffer occupancy probability are periodically measured. The measurement intervals are indexed, for example $\rho(n)$ denotes the utilization in the n-th measurement interval.

As it is shown in Thesis 3.2, I have given a simple state space representation of the queueing behavior, in which $d(n)$ (that corresponds to coefficient d) is the system state variable. I have proposed the following equation for controlling the bit rate of the VP:

$$C(n+1) = \hat{A}(n+1)(1 + \hat{d}(n) \log CLR_{obj}), \tag{28}$$

where $C(n+1)$ is the VP bandwidth in the $n+1$-th control interval, $\hat{A}(n+1)$ is the estimated average traffic in the $n+1$-th control interval predicted in the n-th control interval, $\hat{d}(n)$ is the estimate of the system state variable in the n-th control interval and CLR_{obj} is the target cell loss ratio.

Assuming that the packet arrival process is stationary, such that $\hat{A}(n) = A$, and considering the coefficient d as a function of the VP bandwidth C, the sufficient condition of that the control converges to the optimal VP bandwidth is:

$$\left|\frac{\partial d(C, p_K(C))}{\partial C}\right| < \left|\frac{\partial d(C, CLR_{obj})}{\partial C}\right|, \tag{29}$$

where

$$d(C, p_K(C)) \approx \frac{C - A}{A \log p_K(C)}, \tag{30}$$

and

$$d(C, CLR_{obj}) = \frac{C - A}{A \log CLR_{obj}}. \tag{31}$$

The above condition together with the robust estimation provided in Thesis 3.2 allows us to use the method for controlling the bandwidth of VPs carrying either *short* or *long range dependent* traffic. This property ensures that the method works for traffic patterns that appear in practice (examples for this can be found in [J5]).

Thesis 3.2 : Estimation of the state variable of the queueing system

I have provided the state space representation of the considered single-server queue and I have set up a Kalman Filter to estimate the state variable $d(n)$.

Based on Eq.(27) I have assumed the following relation for the measurement intervals:

$$\frac{1 - \rho(n)}{\rho(n) \log p_K(n)} = d(n) + \sigma(n), \tag{32}$$

where $\sigma(n)$ is Gaussian white noise with mean 0, representing the measurement error. To take into account traffic pattern variation and to ensure the convergence of the bandwidth to the objective, I have allowed the coefficient d to change with time according to the following equation:

$$d(n) = d(n-1) + \omega(n), \tag{33}$$

where $\omega(n)$ is also a Gaussian white noise with mean 0.

In the terminology of the control system Eq.(32) and Eq.(33) are called the "state space representation" of our single server queue, and the variable $d(n)$ is called the state variable of the system.

Bayesian regression analysis (for an introduction, see [17]) can be applied to estimate the state variable $d(n)$. The state variable can be recursively estimated at the end of every measurement epoch by the following Kalman Filter formulation:

[State Renewal]

$$\begin{aligned} \hat{d}(n) &= \hat{d}(n|n-1) + K_n \left\{ \frac{1-\rho(n)}{\rho(n)\log p_K} - \hat{d}(n|n-1) \right\}, \\ K_n &= \frac{D(n|n-1)}{D(n|n-1) + \Sigma(n)}, \\ D(n) &= (1-K_n) D(n|n-1), \\ \Sigma(n) &\stackrel{\text{def}}{=} Var\{\sigma(n)\}. \end{aligned} \tag{34}$$

[Projection]

$$\begin{aligned} \hat{d}(n|n-1) &= \hat{d}(n-1), \\ D(n|n-1) &= D(n-1) + \Omega(n), \\ \Omega(n) &\stackrel{\text{def}}{=} Var\{\omega(n)\}. \end{aligned} \tag{35}$$

Here $\hat{d}(n)$ and $\hat{d}(n|n-1)$ are respectively estimates of the state variable $d(n)$ at the n-th and $(n-1)$-th measurement epochs, and $D(n)$ and $D(n|n-1)$ are respectively the variances of $\hat{d}(n)$ and $\hat{d}(n|n-1)$. Initial value of the state variable can, for example, be given under the assumption that A_t is a stationary process, and we have [18]:

$$d = -\frac{1}{2K} \lim_{t\to\infty} \frac{Var\{A_t\}}{E\{A_t\}}. \tag{36}$$

In practice, to implement the above Kalman Filter formulation it is not a trivial question how to set and/or measure the noise variances $\Sigma(n)$ and $\Omega(n)$. Based on the characteristics of the system and the task of the algorithm I have proposed the following settings:

- I proposed to set the variance of the measurement noise as follows:

$$\Sigma(n) = \left(\frac{a(1-\rho(n))}{(a-1)\rho(n)\log CLR_{obj}} \right)^2, \tag{37}$$

  where a is a number between 0 and 1. According to my experience, $a = 0.25$ is an acceptable value in practice.

- I have determined the variance of the state noise as follows:

$$\Omega(n) = \frac{s^2\Sigma(n)}{(1/b-1)^2(1-\rho(n))^2 - s(1/b-1)(1-\rho(n))}, \tag{38}$$

  where the following rule has been set: at any desired $\rho(n)$, if (due to measurement uncertainties) we estimate such a small $\log p_K(n)$ as $\log p_K(n) = b \cdot \log CLR_{obj}$ for some $b > 1$, the relative change in the VP bandwidth should be less than s (for example, $s = 5\%$). Note that we set the state noise variance for the *desired* $\rho(n)$, because we intend to regulate the behavior of the system when it is stabilized, i.e., the variations of the state are due to measurement uncertainties. According to my experience, $b = 2$ is an appropriate choice.

I have validated the method with the proposed settings using simulation, and concluded that it works correctly for both generated and real ATM traffic.

5 Application of New Results

The research work related to Thesis 1 and Thesis 2 was sponsored and the results have been applied by Ericsson's Product Unit Wideband Radio Network (PU WRN).

The CAC algorithm has been patented [P2]. Two other patent applications are directly related to this algorithm ([P4] and [P6]). The algorithm has been implemented in Ericsson's UTRAN transport node products, which contain AAL2 multiplexers.

The traffic and mobility model of Thesis 2 has been implemented in a software tool at Ericsson. It has not been used to solve practical problems yet.

Methods presented in Thesis 3 have been implemented in a simulator. Using this simulator it was possible to try the algorithm for both simulated and measured traffic. A patent application has been filed [P8].

6 Acknowledgments

During my PhD studies, I have been a member of HSNLab, where Dr. Tamás Henk, the head of the laboratory, helped me to start my work.

The work on Thesis 1 and Thesis 2 was done at the Traffic Analysis and Network Performance Laboratory, Ericsson Research. I wish to thank Dr. Miklós Boda, head of the Research and Development Division at Ericsson Hungary, for his support and encouragement.

The results of Thesis 3 were achieved at the Multimedia Networks Laboratories, Nippon Telegraph and Telephone Corp., Tokyo, Japan. I would like to acknowledge Dr. Hiroshi Saito for his supervision during this time.

During the years I have learned a lot from the co-operation with Dr. Gábor Fodor and Dr. Attila Vidács. Dr. Attila Vidács deserves special thanks for efficiently helping me to finalize my dissertation.

7 Summary of Theses

Thesis 1 Connection admission control in UTRAN [J1, J2, J3, J6, P2, P4]

Thesis 1.1 Connection admission control algorithm

By using the queueing model introduced for UTRAN, I have developed a connection admission control (CAC) algorithm, which can be efficiently used in the Iub interface of UTRAN and handles all RAB types. Based on the input parameters $\{\, b_i,\, TTI_i,\, \alpha_i,\, \tilde{D}_i$ *and* $\tilde{\varepsilon}_i,\, i = 1, \ldots, K \,\}$ *it decides whether a newly arrived connection can be accepted into the system.*

Thesis 1.2 Checking delay violation in the non-overloaded system

I have proposed the hyper-plane approximation for checking delay violation due to temporary packet congestion in the non-overloaded system.

Thesis 1.3 A closed form approximation for calculating the values of the TN matrix

I have given a closed form approximation for calculating the values of the TN *matrix.*

Thesis 1.4 Checking delay violation due to temporary system overload

Based on the queueing model introduced for UTRAN, I have proposed a fast calculation method for checking delay violation due to temporary system overload. The method exploits the statistical gain only within the different classes. Therefore I have also proposed an extension for this method to take into account at least partially the statistical gains among classes.

Thesis 2 Mobility and traffic analysis for WCDMA networks [C1]

Thesis 2.1 Model extension

By extending the model presented in [15], I have set up a mobility and traffic model for analyzing WCDMA systems.

Thesis 2.2 Calculating soft handover region parameters

Using the extended model, I have calculated the following soft handover region parameters:

- the soft handover probabilities of calls generated *or* moving through *a given region,*

- the Z-set intensities of calls generated *or* moving through *a given region,*

- the Z-set offered traffic loads.

Thesis 2.3 User-plane traffic estimation at the Iur interface

Using the extended model and the calculated soft handover region parameters, I have estimated the user-plane traffic at the Iur interface.

Thesis 2.4 Calculating cell parameters

Using the extended model and the soft handover region parameters, I have calculated the distribution of the channel occupancy time, the offered traffic load and the soft handover intensity in a cell.

Thesis 3 Real-time VP bandwidth control [J4, J5, C5, C6, P8]

Thesis 3.1 VP bandwidth control algorithm

I have proposed a VP bandwidth control algorithm, and showed that it converges iteratively to the optimal bandwidth value for both short-range and long-range dependent traffic.

Thesis 3.2 Estimation of the state variable of the queueing system

I have provided the state space representation of the considered single-server queue and I have set up a Kalman Filter to estimate the state variable $d(n)$.

8 References

[1] ATM Forum. *ATM Forum Traffic Management Specification Version 4.0*, 1996.

[2] Göran Eneroth and Martin Johnsson. ATM transport in cellular networks. In *International Switching Symposium (ISS'97)*, Toronto, Canada, September 1997.

[3] Hiroshi Nakamura, Hisakazu Tsuboya, Masatomo Nakano, and Akihisa Nakajima. Applying ATM to mobile infrastructure networks. *IEEE Communications*, January 1998.

[4] *Introduction to 3G Mobile Communications.* Artech House, Norwood, USA, 2001.

[5] *Third Generation Mobile Systems, http://www.ericsson.com/3G.*

[6] Tero Ojanpera and Ramjee Prasad. *Wideband CDMA for Third Generation Mobile Communications.* Artech House, Norwood, USA, 1998.

[7] ITU-T. *AAL Type 2 Signalling Protocol (Capability Set 1)*, 1999.

[8] ITU-T. *AAL Type 2 Signalling Protocol (Capability Set 2)*, 2000.

[9] John H. Baldwin, Behram H. Bharucha, Bharat T. Doshi, Subrahmanyam Dravida, and Sanjiv Nanda. AAL–2 — a new ATM Adaptation Layer for small packet encapsulation and multiplexing. *Bell Labs Technical Journal*, April 1997.

[10] G. Eneroth, G. Fodor, G. Leijonhufvud, A. Rácz, and I. Szabó. Applying ATM/AAL2 as a switching technology in 3rd generation mobile networks. *IEEE Communications Magazine*, 37(1), 1999.

[11] 3GPP Technical Specification TR 25.426. *UTRAN Iur and Iub Interface Data Transport & Transport Signalling for DCH Data Streams*, 1999.

[12] 3GPP. *IP Transport in UTRAN*, March 2002.

[13] I. Norros, J. W. Roberts, A. Simonian, and J. T. Virtamo. The superposition of variable bit rate sources in an ATM multiplexer. *IEEE Journal on Selected Areas in Communications*, 9(3), 1991.

[14] C. Rasmussen, J. H. Sorensen, K. S. Kvols, and S. B. Jacobsen. Source-independent acceptance procedures in atm networks. *IEEE Journal on Selected Areas in Communications*, 9(3), April 1991.

[15] K. Nakano, K. Saita, and M. Sengoku et al. Mobile communications traffic analysis on a road system model. *Performance and Management of Complex Communication Networks, International Federation for Information Processing (IFIP), Kluwer Academic Publishers*, 1998.

[16] A. Vidács. *Fractal characterization of network traffic: from parameter estimation to application, PhD dissertation.* BUTE-DTT, 2000.

[17] G. Welch and G. Bishop. An introduction to the Kalman Filter, TR 95-041, http://www.cs.unc.edu/~welch/kalman/kalmanintro.html, Department of Computer Science, University of North Carolina at Chapel Hill, NC 27599-3175.

[18] S. Shioda and H. Saito. Real-time cell loss ratio estimation and its application to ATM traffic controls. In *proc. IEEE INFOCOM*, Kobe, Japan, 1997.

9 Publications

Journal Papers

[J1] **Sz. Malomsoky**, S. Rácz, Sz. Nádas. Connection admission control in UMTS radio access networks. accepted to *Elsevier Computer Communications*, June 2002.

[J2] **Sz. Malomsoky**, Sz. Nádas, B. Sonkoly. UMTS hozzáférési hálózatok teljesítőképesség vizsgálata. *Híradástechnika*, August 2002.

[J3] **Sz. Malomsoky**, Sz. Nádas, B. Sonkoly. Performance Evaluation of UMTS Terrestrial Radio Access Networks. *Journal on Communications*, July 2002.

[J4] **Sz. Malomsoky**, A. Vidács, H. Saito. Real time VP bandwidth control for long range dependent traffic. *International Journal of Communications Systems*, (12):229-247, 1999.

[J5] A. Vidács, **Sz. Malomsoky**, H. Saito. A simple adaptive bandwidth control for real traffic. *Advances in Performance Analysis*, 2(1):21-44, 1999.

[J6] Sz. Nádas, S. Rácz, **Sz. Malomsoky** and S. Molnár. Connection Admission Control in All-IP UTRAN. *Submitted to IEEE Journal on Selected Areas in Communications, special issue on All-IP wireless networks*, date of submission: February, 2003.

Conference Papers

[C1] **Sz. Malomsoky**, A. Szlávik. Mobility and traffic analysis for WCDMA networks. *International Conference on the Performance and QoS of Next Generation Networking*, Nagoya, November 2000.

[C2] G. Fodor, G. Malicskó, **Sz. Malomsoky**. A joint radio-IP resource reservation scheme in ALL-IP 3rd generation networks. *IEEE Wireless Communications and Networking Conference*, Chicago, September 2000.

[C3] **Sz. Malomsoky**, G. Tóth, Sz. Nádas, P. Zarándy. Simulation based GPRS network dimensioning. *ITC Specialist Seminar on Mobile Networks*, Lillehammer, March 2000.

[C4] G. Fodor, G. Leijonhufvud, **Sz. Malomsoky**, A. Rácz. Comparison of call admission control algorithms in ATM/AAL2 based 3rd generation mobile access networks. *IEEE Wireless Communications and Networking Conference*, New-Orleans, September 1999.

[C5] A. Vidács, **Sz. Malomsoky**, H. Saito. Real-time cell loss ratio estimation for bursty and self-similar traffic. *International Conference of the Performance and Management of Complex Communication Networks (PMCCN97), Workshop 2*, Tsukuba, November 1997.

[C6] **Sz. Malomsoky**, A. Vidács, H. Saito. Bandwidth control and its applicability based on queue length monitoring. *International Conference of the Performance and Management of Complex Communication Networks (PMCCN97), Workshop 2*, Tsukuba, November 1997.

[C7] A. Faragó, T. Cinkler, V.T. Hai, **Sz. Malomsoky**. Joint planning of the physical and logical configuration for ATM networks. *Networks'96*, Sydney, November 1996.

Presentations

[R1] **Sz. Malomsoky**. Traffic planning in a WCDMA network. *The 3GSM World Congress*, Cannes, February 2001.

Patents

[P1] **Sz. Malomsoky**, Sz. Nádas, S. Rácz. Efficient Traffic Concentrator. Patent Application filed in October 2002.

[P2] **Sz. Malomsoky**, Sz. Nádas, S. Rácz. Connection admission control in packet-oriented, multi-service networks. Patent Application filed in March 2002.

[P3] **Sz. Malomsoky**, Sz. Nádas, S. Rácz. Protocol multiplexing. Patent Application filed in March 2002.

[P4] **Sz. Malomsoky**, I. Szabó, S. Rácz. Facilitating reliable connection admission control for telecommunications system using AAL2 signaling. Patent Application filed in March 2001.

[P5] Pál Zarándy, **Sz. Malomsoky**. Method to transfer parallel TCP connections of an UMTS subscriber. Patent Application filed in December 2000.

[P6] **Sz. Malomsoky**. Randomized packet arrival's process in UTRAN. Patent Application filed in October 2000.

[P7] F. Máthé, **Sz. Malomsoky**. Transcoding data in a packet switched communication network supporting radio interfacing by selecting a transcoding processor and/or network portion. Patent Application filed in August 1999.

[P8] **Sz. Malomsoky**, A. Vidács, H. Saito. Virtual path bandwidth control apparatus and virtual path bandwidth dimensioning method. Patent Application filed in December 1997.

[P9] W. Holender, **Sz. Malomsoky**. Adaptive virtual path dimensioning method especially for paths defined on telecommunications network using entropy rate function as blocking measure and balancing loads on links by equalizing blocking probabilities and determining allocation of physical resources. Patent Application filed in July 1995.