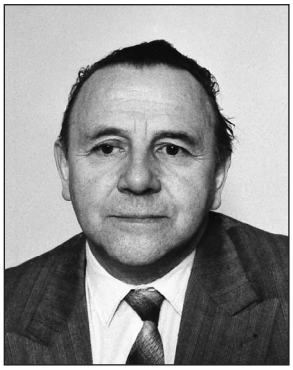
KOMUNIKÁCIE
COMMUNICATIONS



Dear reader,

We are offering you a monothematic issue of the "Communications", the journal dedicated to prospective telecommunication networks.

A modern society requires the transfer of a huge amount of data over any distance and within a very short time. Only wideband networks with a Terabit transfer rate can meet such a demanding requirement. WDM and OTM transport optical networks with a dynamic bandwidth control (by means of a control layer) and wideband access networks due to their optimal structure can successfully achieve the above objective. New 3G and 4G generations are to accomplish the task within mobile networks.

The Internet gains an increasing importance from the point of view of new generation networks. This is a reason why in order to provide various telecommunications services also in IP environment, many protocols, whose objective is to carry out high quality wideband services, are implemented in the Internet. Therefore, many research teams focus their attention on topics dealing with audio and video signal transfer.

The articles in the journal analyze the above topics in a form of theoretical studies, simulation and optimization models as well as prospective applications.

Karol Blunár

https://doi.org/10.26552/com.C.2004.4.5-12

Georgios S. Paschos – Efstathios Vagenas – Stavros A. Kotsopoulos *

MOBILITY PERFORMANCE OF A NEW TRAFFIC DECENTRALIZATION QoS ADMISSION STRATEGY

This paper proposes a new Admission Control strategy with Quality of Service (QoS) preservation for GSM networks. System initiated handover calls are used in the central cell in order to decentralize the traffic and redistribute it throughout the whole cluster. Although the channels and the cell size are fixed, the system behaves as if it had the 3G feature called soft capacity. The mobility characteristics of the users are investigated. Low speed and high speed moving users are used as input in the system and the effect of mobility on the proposed algorithm is extracted. Simulation results show that the proposed Admission Strategy is flexible, efficient and very useful in situations when the traffic is centralized.

1. Introduction

Resource management is a useful and efficient tool for network optimization. Documentation of several modern network standards contains features like call admission control (CAC), priority and Quality of Service (QoS). The effort is focused on using efficiently the available resources and simultaneously providing a constant and acceptable level of quality to the user. This is usually achieved through network algorithms called CAC algorithms.

In 3G networks, the resources are easily manipulated throughout the cluster using soft capacity. With the soft capacity feature, the central cell resources are dependent upon the neighbor traffic. Moreover, fractional loading, the utilization of resources becomes even more efficient.

In contrast, Global System for Mobile (GSM) networks have fixed channel resources and cell sizes. The implementation of soft capacity is much more challenging. Soft capacity-like characteristics in GSM networks have been studied in [1] and fractional load techniques have been studied in [2].

Another important issue is the mobility characteristics of the users. In a mobile network, the users are free to roam around the service area, and therefore cause handover traffic. While the handover mechanism is transparent to the user, it is a deciding issue in terms of admission control. It is widely accepted that handover calls are more important than new calls [3] and therefore handovers are provided a sense of priority in admission schemes. The ratio of handover attempts per new calls is called handover ratio. This ratio is related to the size of the cells, the mobility characteristics of the users and the admission strategy used.

Several studies have been made on effective CAC algorithms that can be used in GSM networks in order to provide increased QoS. The authors in [4] study guard channel admission control schemes, including new call bounding, cut off priority and new call thinning scheme. In [5] resource allocation and de-allocation is investigated for reducing the blocking probability. A thorough study of dynamic call admission policies is included in [6] and [7], where the algorithms that are presented, are adaptive to the changes of the traffic conditions and the availability of the resources in the cell. In [8] a review of the basic handover schemes that are used in mobile communications is provided. Mobility-based algorithms have been proposed in [9] and [10], where the admission algorithms are handling a call request (new or handover) according to the predictive mobility of the user.

This paper proposes a new way of introducing soft capacity into GSM networks, via system-initiated handovers. The algorithm enables soft capacity by using the unoccupied channels of the neighboring cells and improves the QoS of asymmetrically loaded systems. The performance of the proposed algorithm is tested over several user mobility cases.

2. Mobility issues

In order to calculate the handover ratio, the mobility analysis has to be performed. Mobility models have been analysed in [11], [12], [13], [14] and [15]. In [12], [13], [14] and [15], modern mobility models are proposed. However, the disadvantage of these models is the lack of analytical approach. The gamma distribution is found with the procedure of fitting the samples taken from simulations. In this analysis, the interest is focused in urban environments, since centralized traffic conditions usually appear in these cases. As in [11], it is considered that the user is moving in straight lines. The initial location is uniformly random and set (r_0, φ_0). Then, the user, randomly chooses a destination (declared by θ_i) and travels with a random velocity (declared by V_i). In case of handover call, the initial position is set on the boundaries of the cell. The four elements have the following probability density functions (pdfs):

* **Georgios S. Paschos, Efstathios Vagenas, Stavros A. Kotsopoulos**
Wireless Telecommunications Lab. – University of Patras, Patras, 26500, Greece, E-mail: kotso@ee.upatras.gr

$$f_{\varphi_0}(\varphi) = \frac{1}{2\pi}, 0 \le \varphi \le 2\pi \quad (1)$$

$$f_{r_0}(r) = \frac{2r}{R_{eq}}, 0 \le r \le R$$

$$f_{\theta_i}(r_\theta) = \frac{1}{\pi}, 0 \le \theta \le \pi \quad (2)$$

where:

$$Req = \sqrt{\frac{3\sqrt{3}}{2\pi}} R$$

$$f_{V_i}(V) = \frac{1}{V_{max}}, 0 \le V \le V_{max} \quad (3)$$

Using the above statistical results we can estimate the probability density functions pdfs of the resource holding time for a new call f_{Tn} and for a handover f_{Th}, respectively. *Tn* is the sojourn time of a new call in a cell, and *Th* the sojourn time of a handover call respectively. In [11], these are calculated and found that follow the following distribution:

$$f_{Tn} = \begin{cases} \frac{8R_{eq}}{3V_{max}\pi t^2}\left[1 - \sqrt{\left\{1 - \left(\frac{tV_{max}}{2R_{eq}}\right)^2\right\}^3}\right], 0 \le t \le \frac{2R_{eq}}{V_{max}} \\ \frac{8R_{eq}}{3V_{max}\pi t^2}, t \ge \frac{2R_{eq}}{V_{max}} \end{cases} \quad (4)$$

$$f_{Th} = \begin{cases} \frac{4R_{eq}}{\pi V_{max} t^2}\left[1 - \sqrt{1 - \left(\frac{tV_{max}}{2R_{eq}}\right)^2}\right], 0 \le t \le \frac{2R_{eq}}{V_{max}} \\ \frac{4R_{eq}}{\pi V_{max} t^2}, t \ge \frac{2R_{eq}}{V_{max}} \end{cases} \quad (5)$$

The channel holding time T is exponentially distributed [11], with mean value $1/\mu$. The pdf and the cumulative density function (cdf) are:

$$f_T(t) = \mu e^{-\mu t}, t \ge 0 \quad (6)$$

$$F_T(t) = 1 - e^{-\mu t}, t \ge 0 \quad (7)$$

We define the probability that a single call requires k handovers ($H=k$) until it is successfully completed as:

The average handover rate is calculated:

$$\lambda_H = \sum_{i=1}^{\infty} \lambda_N i P\{H = i\} = \lambda_N \overline{H} \quad (10)$$

It is evident that the handover ratio depends only on the user velocity and the cell range up to this point. The results from this analysis will be used in the teletraffic analysis.

3. QoS Estimation

The traffic load is assumed largely centralized and handled by the central cell of a seven-cell cluster. Therefore the system, in order to preserve the QoS in the central cell, initiates system handover call requests to the neighboring cells. The CAC scheme searches for available system resources every time a new call or a handover request is issued. If the ongoing traffic exceeds load threshold decided by the provider then the system scans all the established connection in order to find one that can be handed off to a neighboring cell. This decision is called QoS estimation and is based on a QoS Index denoted by QI (QoS Index).

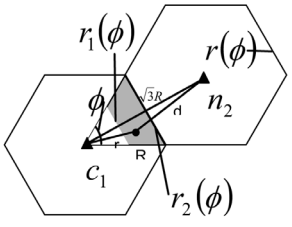


Fig. 1 Representation of parameters for calculating QI.

$$P\{H = k\} = P\{(T < T^n + kT_h) \cap (T > T_n + (k-1)T_h)\} = P\{T < T_n + kT_h\} \cdot P\{T > T_n + (k-1)T_h\} =$$
$$= \left(1 - \int_0^\infty [1 - F_T(t)] f_{T_n + kT_h}(t) dt\right) \cdot \int_0^\infty [1 - F_T(t)] f_{T_n + (k-1)T_h}(t) dt = \quad (8)$$
$$= \left(1 - \int_0^\infty e^{-\mu t}(f_{T_n}(t) + kF_{T_h}(t)) dt\right) \cdot \int_0^\infty e^{-\mu t}(f_{T_n}(t) + (k-1)f_{T_h}(t)) dt$$

Figure 1 shows the case where a mobile station is in a specific action area between the central cell and its neighbor and it receives signals from both base stations. Thus according to QoS estimation the two corresponding signal-to-interference ratios SIR_{C1} and SIR_{n2} are compared using QI as follows:

$$SIR_{C_1} = \frac{P(r)}{N + I_{adj}(d)} = \frac{A/r^2}{N + I/d^2} \tag{11}$$

$$SIR_{n_2} = \frac{P(d)}{N + I_{adj}(r)} = \frac{A/d^2}{N + 1/r^2}$$

$$QI(r) = \frac{SIR_{C_1}}{SIR_{n_2}}$$

Where r and d represent the distance of the mobile subscriber from the Base Transceiving Stations (BTS) of the central cell and its neighboring cell respectively and A is a factor, constant over r^2. Additionally, I is the adjacent channel interference and N the background noise. From the above relationship we can derive the QI in terms of the location of the mobile station in the cell:

$$QI = \frac{d^4}{r^4}\,\frac{Cr^2 + 1}{Cd^2 + 1}\text{, where } C = \frac{N}{I} \tag{12}$$

In the last equation the value of C is assumed known at all times, hence the QoS index is determined by the size and the shape of the action area. Therefore, for a specific threshold value $ThrQI$, specified by the network provider, we have the following relationship:

$$QI < Thr\ QI \rightarrow \text{the call satisfies } QI \text{ criteria} \tag{13}$$
$$QI > Thr\ QI \rightarrow \text{the call does not satisfy } QI \text{ criteria}$$

The probability of a user roaming in the action area between the central and a neighboring cell is denoted by P_{QI}. Assuming a load distribution in the cluster of $\lambda(r, \varphi)$ the probability is:

$$P_{QI} = 1 - \frac{\int_0^{r_1(\phi)}\int_0^{\pi/3} \lambda(r, \phi)drd\phi}{\int_0^{r_2(\phi)}\int_0^{\pi/3} \lambda(r, \phi)drd\phi} \tag{14}$$

In (4) $r_1(\phi)$ is the vector that represents the inside boundary of the action area and $r_2(\phi)$ is the vector representation of the boundary of the central cell and its neighboring cell, as shown in figure 1. This probability indicates that a call satisfies the QoS index QI and is available for system issued handover to a neighboring cell. The QoS estimation is a key function to the proposed call admission strategy and is performed by the QoS estimator, which is a major component of the system model.

4. Proposed CAC Scheme

One simple admission strategy is the Handover Reservation algorithm which differentiates between new call and handover call requests and assigns higher priority to the latter by reserving resources for handover calls only, [16, 17]. The algorithm then checks for spare channels. If there are unoccupied channels then the handover request is admitted, otherwise the request is dropped. Similar procedure is followed for new calls, which are admitted if there are unassigned channels available, other than the channels reserved for handover calls or blocked if all channels are occupied. Based on this simple and straight forward CAC scheme, we developed a new 3G-like CAC algorithm for 2G networks that guarantees QoS in the cluster. The proposed call admission policy, by means of thresholds controls the traffic decentralization and enables the soft capacity features, despite the fact that the cell sizes and resources in the model are fixed.

In particular, this study assumes a seven-cell cluster system with fixed cell radius. The central cell, which is the cell under study, is assumed to cover an area with excessive traffic while the neighboring cells suffer from much less loading, therefore being used as reservoirs for the excess load of the central cell. This model although theoretical, represents a common situation in real networks. The proposed admission control strategy, based on the above cell-planning model, decentralizes the traffic by handing over the excess load of the central cell to its neighbours. It has to be mentioned that the system initiated handover call requests (sih) are only issued from the central cell towards its neighbors. On the other hand, user initiated handover call requests (uih) due to the mobility of the user, who passes from one cell to another, can equally be issued by all cells in the cluster. The sihs are adequately controlled by the admission scheme in a way that QoS requirements are preserved in the cluster. Figure 2 showcases in a schematic representation, how the admission control strategy achieves traffic decentralization by initiating QoS-driven system handovers.

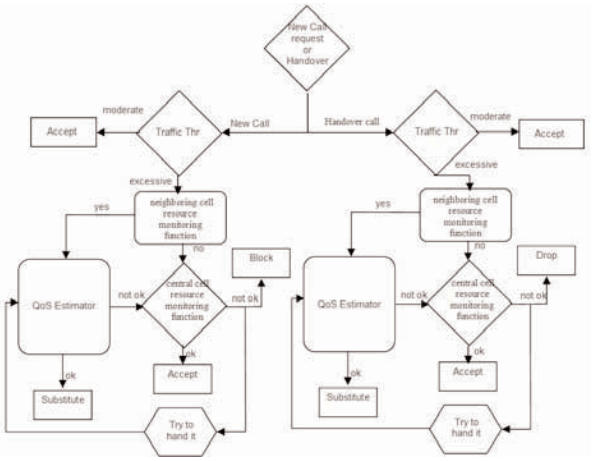


Fig. 2 The schematic diagram of the proposed admission strategy in the central cell.

Main Components of the System

The proposed call admission strategy in figure 2 consists of three major components:

- The neighboring cell resource monitoring function.
- The QoS estimator.
- The central cell resources monitoring function.

The neighboring cell resource monitoring function checks the resources of the neighboring cell in order to prevent overloading of the neighboring cells. If the neighbor resources are adequate, the QoS estimator is reached, where the QoS index is checked as the last requirement for a sih to be performed. Additionally, the central cell resources monitoring function checks the available channels in the central cell every time that a sih cannot be initiated and the issued call must be serviced within the cell. If this function finds available channels in the central cell then the call is accepted, otherwise it will be handed over to a neighboring cell as a last-ditch effort.

The CAC algorithm

The CAC algorithm is intended to be used in the central cell of a seven cell cluster. For the neighboring cells a resource reservation scheme for handover calls can be assumed, since the system ensures QoS levels in the neighboring cell by means of the resource monitoring function. Thus, only uih call requests are expected in the central cell. The analysis of the algorithm that follows is based on figure 2.

When a new call request is issued, the system checks the number of available resources in the cell. The request will be immediately admitted if the number of used channels (Ch_{used}) is below a threshold value Thr_N.

$$Ch_{used} \leq Thr_N \xrightarrow{\text{mod erate}} admit\ new\ call \tag{15}$$

When the available channels have reached the limit value Thr_N the algorithm is initiated and the neighboring cell resource monitoring function is called to perform the procedure described above.

$$Ch_{neigbor} \leq thr_{SH} \rightarrow sih\ anabled \tag{16}$$

$$Ch_{neigbor} > thr_{SH} \rightarrow sih\ disabled$$

Where thr_{SH} is a threshold for system initiated handovers and in a tolerable central-concerned system we can consider, $thr_{SH} = N$. When the neighbor cell is in a state less than thr_{SH}, the QoS estimator based on the QoS index QI either initiates a sih and the new call request is accepted in the central cell or the central cell resources monitoring function is used to allocate the call in the central cell. In the latter case, the new call request is admitted if the spare channels in the new call resource pool are below the threshold, Res_H. This value denotes the limit of the new call resource pool.

$$\{Thr_N \leq Ch_{used} \leq Res_H\} \cap \{QoS\ ok\} \xrightarrow{excessive} substitute \tag{17}$$

$$\{Thr_N \leq Ch_{used} \leq Res_H\} \cap \{QoS\ not\ ok\} \xrightarrow{excessive} admit$$

Furthermore, for a uih request a similar procedure is followed. As long as, the number of occupied channels in the central cell are below the threshold value Thr_H then the load in the cell is considered moderate and the request is admitted immediately:

$$Ch_{used} \leq Thr_H \xrightarrow{\text{mod erate}} uih\ is\ admitted \tag{18}$$

In a similar way as for new calls, when the traffic is considered excessive, the neighboring cell resource monitoring function is called and (16) enables a sih. Depending on the QoS estimation, the uih either substitutes an ongoing call which in turn is handed over to the neighbor host, or is accepted by the central cell provided the following condition about the available channels is true:

$$Ch_{used} < \xrightarrow{excessive} uih\ is\ admitted \tag{19}$$

Otherwise, the uih request will be blindly handed over to a neighboring cell as a last effort before it is dropped.

5. Teletraffic Analysis

In this section the traffic analysis is developed. The assumptions and necessary definitions are firstly stated and then applied in the mathematical analysis that follows.

Assumptions

i. Cluster traffic normalized to the *load* variable is assumed as described above.
ii. Poisson traffic with λ arrival rates and μ service rates.
iii. A theoretical seven cells per cluster analysis have been assumed with fixed cell radius R.
iv. The neighbor cells surrounding the central cell are assumed symmetrical. This assumption cannot harm the robustness of this analysis since it can be easily modified for a case study that requires asymmetrical cells or traffic.
v. Dirichlet boundaries are assumed and a random neighbor cell is surrounded by four neighbor cells and two central cells.
vi. The background noise is considered greater than the perceived adjacent channel interference when the mobile roams inside the cell which is perfectly reasonable for a typical network with large reuse factor.
vii. We assume a threshold value regarding the QoS estimator which is explained in figure 1. This graphically presented threshold is yielded from (12).

Definitions

An important feature for this analysis is the probability that a system initialized handoff request is granted by the QoS estimator, P_{acc}.

$$P_{acc} = P\{(QoS\ index\ ok) \cap (Ch_{neighbor})\} = P_{QI} \cdot \sum_{n=0}^{thrSH} P_n^n \tag{20}$$

P_{QI} is the probability for a call to fulfil the QoS index requirements. Furthermore, we define A_c as the fact that a new call is either handed onto another cell or a substitute handoff is used as compensation. The result is that the new call will not burden the available cell resources. B_c is the respective fact for a handover call case.

The probabilities of Ac and Bc are defined as α and b, and can be used as multipliers of λ in the Markov chain.

$$a = P\{A_c\} = P\{(n \geq thr_N) \cap (QoS\ Ok)\} = \sum_{n=thrN}^{N} P_n^c \cdot P_{acc} \tag{21}$$

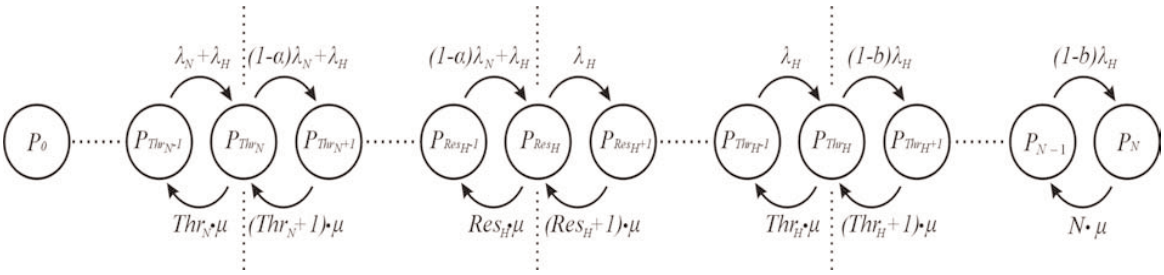


Fig. 3. The Markov chain modified for the needs of the proposed algorithm.

$$b = P\{B_c\} = P\{(n \geq thr_N) \cap (QoS\ Ok)\} = \sum_{n=thrN}^{N} P_n^c \cdot P_{acc} \quad (22)$$

The service rate μ is considered the same to all cells. The user initiated arrival rates, λc and λn are calculated from the cluster load function using a double integral. The final arrival rates, combining the transactions caused by the algorithm are:

$$\lambda_H^c = \frac{\gamma}{\gamma+1}\lambda^c - b\frac{\gamma}{\gamma+1}\lambda^c \sum_{n=Thr_H}^{N} P_n^c \quad (23)$$

$$\lambda_N^c = \frac{\gamma}{\gamma+1}\lambda^c - a\frac{\gamma}{\gamma+1}\lambda^c \sum_{n=Thr_N}^{Res_H} P_n^c \quad (24)$$

$$\lambda_H^n = \frac{\gamma}{\gamma+1}\lambda^n + a\lambda_N^c \sum_{n=Thr_N}^{Res_H} P_n^c + b\lambda_H^c \sum_{n=Thr_H}^{N} P_n^c \quad (25)$$

$$\lambda_N^n = \frac{1}{\gamma+1}\lambda^n \quad (26)$$

Where the indices c and n refer to central and neighbor cell γ is the handover ratio which is the same for both neighbor and central cells.

Markov analysis

The classical Birth-Death process is used for the central cell and a representative of the remaining 6 of this one tier analysis. The modifications due to the algorithm are showcased in figure 3.

The probability that the state n is occurring in the central cell will be as shown at the bottom of the page (28). It should be noted that apart from α and b the rest comprise a straightforward analysis. However these two percentage variables depend on the load condition of the neighbor cells.

The Markov analysis for a random neighbor cell will differ only by the fact that the iteration algorithm is not used:

$$P_n^c = \begin{cases} \dfrac{(\lambda_N^c+\lambda_H^c)^n}{n!\,\mu^n} P_0^c & 0 \leq n \leq Thr_N \\ \dfrac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{n-Thr_N}}{n!\,\mu^n} P_0^c & Thr_N < n \leq Res_H \\ \dfrac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{Res_H-Thr_N}(\lambda_H^c)^{n-Res_H}}{n!\,\mu^n} P_0^c & Res_H < n \leq Thr_H \\ \dfrac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{Res_H-Thr_N}(\lambda_H^c)^{n-Res_H}((1-b)\lambda_H^c)^{n-Thr_H}}{n!\,\mu^n} P_0^c & Thr_H < n \leq N \end{cases}$$

$$P_n^c = 1/\left(1 + \sum_{n=1}^{Thr_N}\frac{(\lambda_N^c+\lambda_H^c)^n}{n!\,\mu^n} + \sum_{n=Thr_N+1}^{Res_H}\frac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{n-Thr_N}}{n!\,\mu^n} + \right.$$
$$+ \sum_{n=Res_H+1}^{Thr_H}\frac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{Res_H-Thr_N}(\lambda_H^c)^{n-Res_H}}{n!\,\mu^n} + \quad (27)$$
$$\left. + \sum_{n=Thr_H+1}^{N}\frac{(\lambda_N^c+\lambda_H^c)^{Thr_N}((1-a)\lambda_N^c+\lambda_H^c)^{Res_H-Thr_N}(\lambda_H^c)^{n-Res_H}((1-b)\lambda_H^c)^{n-Thr_H}}{n!\,\mu^n}\right)$$

$$P_n^n = \begin{cases} \dfrac{(\lambda_N^n + \lambda_H^n)^n}{n!\ \mu^n} P_0^n & 0 \le n \le Res_H \\ \dfrac{(\lambda_N^n + \lambda_H^n)^{Res_H} (\lambda_H^n)^{n-Res_H}}{n!\ \mu^n} P_0^n & Res_H < n \le N \end{cases}$$

Where P_0^n is:

$$P_0^n = 1/\left(1 + \sum_{n=1}^{Res_H} \frac{(\lambda_N^n + \lambda_H^n)^n}{n!\ \mu^n} + \sum_{n=Res_H+1}^{N} \frac{(\lambda_N^n + \lambda_H^n)^{Res_H} (\lambda_H^n)^{n-Res_H}}{n!\ \mu^n}\right) \quad (28)$$

The equations 21, 22, 27 and 28 comprise a $2N+2$ non-linear system of equations and equal unknowns. This system could be apparently large and even larger in case of asymmetrical cells. However numerical analysis methods can be used to efficiently solve the problem.

Iteration method

In order to solve the above system of equations and extract the valuable simulation results, the iteration method of numerical analysis is utilized [18].

According to this, we make a first estimate of the unknowns α and b, ($\alpha 0$, $b0$). We use this estimate to solve the system and calculate a new set of α and b values. This procedure is called the first iteration. With each new iteration, a new set of α and b values is calculated. A point is then reached when these values are stabilized and no more iterations are needed. The method has converged to the solution of the system. We present the convergence of these two variables for several load values in the following figures 4, 5.

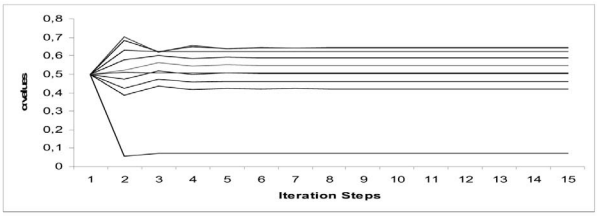


Fig. 4. The convergence of α values for several load cases.

From figures 4, 5, it follows that the method converges rapidly to the solution of the system, within five, at the maximum, iterations. Moreover, this happens regardless of the first estimate of values (α_0, b_0), or the load situation.

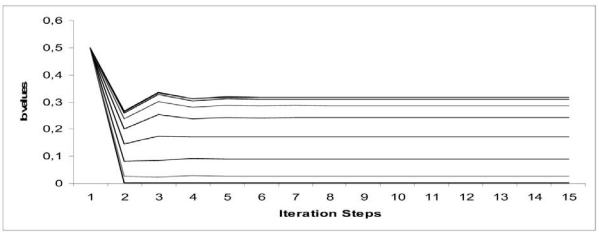


Fig. 5. The convergence of b values for several load cases.

Some measure of instability was encountered in cases of extreme loading where the system oscillates between blocking and non-blocking situation within one step. This happens because of the impact of α and b values incur when the λ/μ is very large. However, this fact cannot affect the rest of our simulation.

6. Results

The discussion of the simulation results is based on blocking and dropping probability figures stemmed from the aforementioned analysis. These two indices were chosen as measures because of their ability to grade an admission control algorithm regarding its overall performance.

The proposed system will be tested against the simplified and prevailing Erlang-B and Handover Reservation systems. Such a comparison is fair in this case, since the proposed system is an improved variation of the well-adopted Handover Reservation scheme and the only handicap of the original system is the QoS estimator which is included in the cost of the proposed one. Moreover, the Erlang-B system is always presented as a down boundary for blocking probability and upper boundary of dropping for every scheme in comparison. The traffic is assumed to have circular symmetry ($\lambda(r)$) and the load will be distributed across the radius.

$$\lambda(r,\phi) = \begin{cases} \lambda_{\max},\ 0 \le r \le R \\ \dfrac{\lambda_{\max}}{3},\ R < r \le r(\phi) \end{cases} \quad (29)$$

$$\lambda_{\max} = \frac{1}{\int_0^{\pi/3} \int_0^{r(\phi)} \lambda(r,\ \phi) dr d\phi}$$

Table 1 contains the rest of simulation inputs. The results are shown in the following figures.

The results for the central cell demonstrate the expected improvement of the system. In both cases (blocking and dropping) the proposed algorithm yields an important gain over the similar

Simulation values Table 1

	Values
Available channels	36
Average Call Duration	90 sec
α_0, b_0	0.3, 0.3
Iteration steps	15
Cell Range	500 m
High speed mean value	30 m/sec
Low speed mean value	5 m/sec

Handover Reservation algorithm. This improvement approximates one order of magnitude for heavy load situations. Moreover, the performance of the proposed algorithm challenges even the Erlang-B performance in case of blocking with heavy traffic. Figures 8, 9 show the defect of the proposed algorithm. Both probabilities are increased in case of the neighbor cell. However, for a case study of centralized traffic such as this, this deterioration is not of importance. The probabilities are practically held below 10^{-2} in any case while the traffic in the central cell is relieved only by the proposed algorithm.

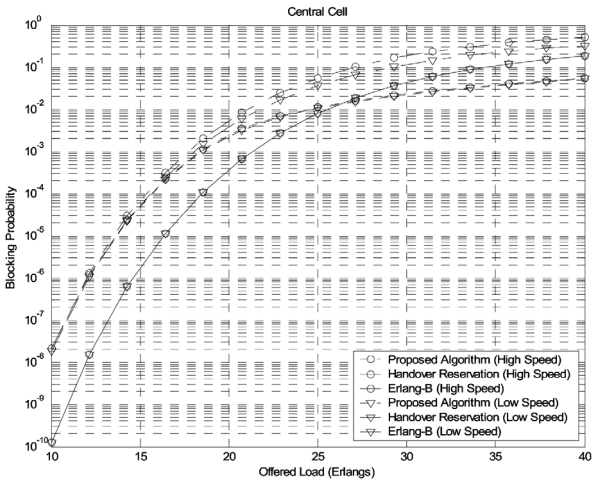


Fig. 6. Blocking Probability for the central cell ($thr_N = 0.7 \cdot N$, $Res_H = 0.9 \cdot N$ $thr_H = 0.9 \cdot N$)

7. Conclusion

The implementation of the proposed algorithm concerning the QoS driven admission strategy, can take place on the already installed systems with small interventions. The QoS indicator software structure is necessary along with some custom system messages for support. This can be achieved with the already in-use QoS strategies and measurements regarding the network inherent features. Therefore, the cost of the implementation can be regarded as minor.

Regarding the increased handover traffic, the algorithm possesses a self-defense mechanism. Due to the imposed thresholds,

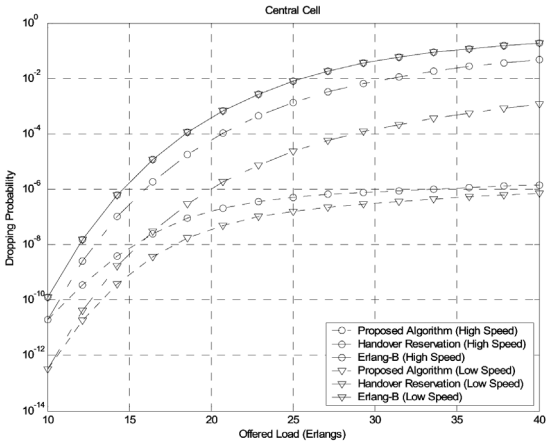


Fig. 7. Dropping Probability for the central cell ($thr_N = 0.7 \cdot N$, $Res_H = 0.9 \cdot N$, $thr_H = 0.9 \cdot N$)

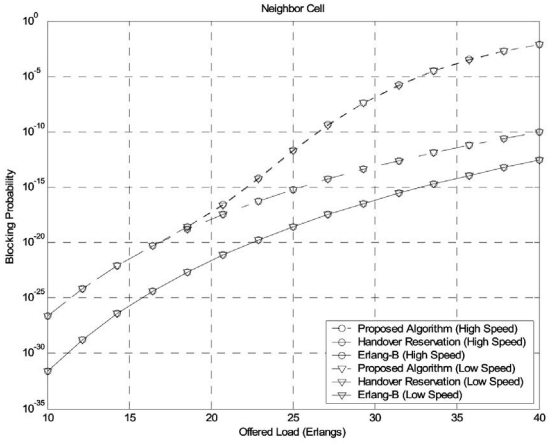


Fig. 8. Blocking Probability for the neighbor cell ($thr_N = 0.7 \cdot N$, $Res_H = 0.9 \cdot N$, $thr_H = 0.9 \cdot N$)

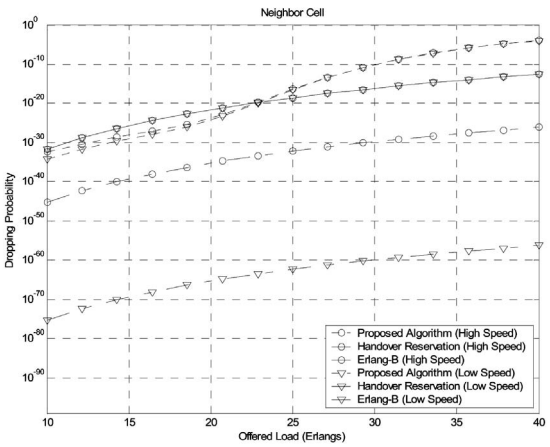


Fig. 9. Dropping Probability for the neighbor cell ($thr_N = 0.7 \cdot N$, $Res_H = 0.9 \cdot N$, $thr_H = 0.9 \cdot N$)

the sihs take place only when the system suffers from load peaks and thus, the handover traffic is increased only at times of obvious despair trying to keep the diminishing QoS steady. In addition to that, the handover reservation bandwidth acts as another countering factor and provides further defense against the handover traffic.

The drawn conclusion is that the proposed algorithm creates the soft capacity feature in a 2G system with fixed cells and resources. Hence it offers a satisfying improvement of QoS for asymmetrically loaded systems by adding an insignificantly small fraction of complexity in the system.

References

[1] AHMED, M. H., MAHMOUD, S. A.: *Soft Capacity Analysis of TDMA Systems with Slow-Frequency Hopping and Multiple-Beam Smart Antennas, IEEE Trans. Veh. Technol.*, vol. VT-51, No.4, Jul. 2002.
[2] BULDORINI, A., GIUNTINI, E., MAGNANI, N. P.: *Performance of a GSM-like system adopting the fractional loading technique,* IEEE Wireless Communications and Networking Conference, 1999.
[3] HARDY, W. C.: *Measurement and Evaluation of Telecommunications Quality of Service*, WILEY 2001.
[4] FANG, Y., ZHANG, Y.: *Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks,* IEEE Trans. Veh. Technol., vol. VT-51, Mar 2002.
[5] CHEN, W. Y., WU, J. L., LU, L. L.: *Performance Comparison of Dynamic Resource allocation with/without Channel de-allocation in GSM/GPRS Networks*, IEEE Commun. Lett., vol. 7, No. 1, Jan. 2003.
[6] WU, S., WONG, K. Y. M., LI, B. O.: *A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks,* IEEE Trans. Network, vol.10, No2, Apr. 2002.
[7] ZHANG, Y., LIU, D.: *An adaptive algorithm for call admission control in wireless networks,* IEEE Global Telecomm. Confer., GLOBECOM '01, vol.6, 25-29, Nov. 2001.
[8] LOHI, M., BALDO, O., AGHVAMI, A. H.: *Handover issues and call admission control in cellular systems,* IEEE Veh. Technol. Confer., VTC 1999 - Fall VTS 50th , vol.3, 19-22, Sept. 1999.
[9] HOU, J., FANG, Y.: *Mobility-based call admission control schemes for wireless mobile networks,* Wireless. Commum. Mob. Comput. 2001, 1:269-282.
[10] ALJADHAI, A. R., ZNATI, T. F.: *Predictive Mobility Support for QoS Provisioning in Mobile Wireless Environments,* IEEE Journal Select. Areas Commun., vol. 19, No. 10, Oct. 2001.
[11] HONG, D., RAPPAPORT, S. S.: *Traffic model and performance analysis of cellular radio telephone systems with prioritized and non-prioritized handoff procedures,* IEEE Trans. Veh. Technol., vol. VT-35, Aug. 1986.
[12] CHOI, S. G., CHO, K. R.: *Traffic Control Schemes and Performance Analysis of Multimedia Service in Cellular Systems,* IEEE Trans. Veh. Technol., vol. 52, No 6, Nov 2003.
[13] BRATANOV, P. I., BONEK, E.: *Mobility Model of Vehicle-Borne Terminals in Urban cellular systems,* IEEE Trans. Veh. Technol., vol. VT-52, No 4, Jul 2003.
[14] GUERIN, R. A.: *Channel occupancy time distribution in a cellular radio system,* IEEE Trans. Veh. Technol., vol. VT-35, Aug 1987.
[15] ZONOOZI, M. M., DASSANAYAKE, D.: *User mobility modeling and characterization of mobility patterns,* IEEE J. on Selected Areas in Comm., 15(7):1239-52, Sept. 1997. 25.
[16] KULAVARATHARASAH, M. D., AGHVAMI, A.H.: *Teletraffic performance evaluation of microcellular personal communication networks (PCN's) with prioritized handoff procedures,* IEEE Trans. Veh. Technol., vol. VT-48, No.1, pp. 137–152, Jan. 1999.
[17] FANG, Y., ZHANG, Y.: *Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks,* IEEE Trans. Veh. Technol., vol. VT-51, Mar 2002.
[18] BURDEN, R. L., FAIRES, J. D.: *Numerical Analysis*, Brooks Cole, 2000.
[19] WANG, K., LEE, L. S.: *Design and analysis of QoS supported frequent handover schemes in microcellular ATM networks,* IEEE Trans. Veh. Technol., vol. VT-50, No.4, Jul. 2001.

https://doi.org/10.26552/com.C.2004.4.13-16

Michael D. Logothetis – Ioannis D. Moscholios – George K. Kokkinakis *

TRAFFIC SAVINGS THROUGH A CAC PROCEDURE INCORPORATING A FAIR BANDWIDTH ALLOCATION POLICY FOR ELASTIC SERVICES

We apply the max-min fairness policy (MMF) for elastic services with min and max bandwidth requirements (approximated MMF) not to a single link but to a whole connection-oriented network. To achieve it we propose a Call Admission Control (CAC) procedure consisting of several scenarios dependent on the number of paths between the origin-destination (O-D) network nodes. We evaluate the global network performance at the call level by comparing the maximum Call Blocking Probability (CBP) of the proposed scheme with that of a no fair bandwidth allocation policy scheme. Simulation results show significant traffic savings in the case of the approximated MMF policy.

Keywords: *bandwidth allocation, max-min fairness, elastic traffic, call blocking probability.*

1. Introduction

Elastic services keep the call-level study in connection-oriented networks an open issue. We examine the applicability of a fair bandwidth allocation (FBA) policy to a whole connection-oriented network. From the bandwidth management point of view, a connection-oriented network is a mesh network. We consider connection-oriented (VC routing) networks, in which each path connecting an O-D pair has a certain bandwidth capacity (e.g. Virtual Path in ATM networks). We assume that the O-D path bandwidth is determined at the network design phase (NDP) and used only by calls of this O-D pair. Several FBA policies based on the MMF policy [1] have been proposed as a flow/congestion control mechanism of elastic traffic and tested either in a single link or in an isolated path. In a network most studies apply the MMF, either in the domain of flow/congestion control, dealing with low-level traffic [2] or in the routing domain [3] dealing with call-level traffic. We adopt the MMF policy with min and max bandwidth requirements (BR) (we call it approximated MMF, a-MMF) as the most appropriate for elastic traffic and use a linear programming (LP) model to present it [4]. We aim at applying the a-MMF to the call-level of a network accommodating elastic service-classes while we concentrate on the CAC process, where bandwidth is assigned to each call. We propose the application of the a-MMF at call setup among calls of the same destination taking into account the available bandwidth, *avl* of the O-D path while avoiding the bandwidth modification of in-service calls which requires a feedback control mechanism for all service-classes. To achieve our purpose we devote a fraction of the max permitted time for the completion of call-setup to the a-MMF process. The ultimate goal is to improve the global call-level network performance.

In section 2 we propose a LP model in order to provide a rigorous description of the a-MMF. In section 3 we present the CAC procedure that makes possible the application of the a-MMF to the entire network. Section 4 contains an application example for evaluation. We conclude in section 5.

2. LP model for the approximated MMF policy

The MMF policy with min and max BR of the services cannot be presented by a single LP model. Nevertheless, we propose an approximate single LP model which has the advantage of a clear and rigorous description of the a-MMF; its solution is not always an MMF vector. If one omits the min and max BR, then the resultant bandwidth vector is an MMF vector.

Let S and P be the set of network service-classes and O-D paths, respectively. Calls are conveyed by a fixed routing scheme. By C_p we denote the *avl* of O-D path $p\in P$, by S_p the set of service-classes which use O-D path $p\in P$, by n_s the number of calls of service-class $s\in S_p$ and by r_s the bandwidth which will be allocated to new calls of service-class $s\in S$. The following LP model provides a rigorous description of the a-MMF.

The constraint sets of the LP model are the following [4]:

$$\text{Network set: } \sum_{s\in S_p} n_s r_s \le C_p, \text{ for all } p\in P \quad (1)$$

$$\text{Traffic parameters set: } b_{min,s} \le r_s \le b_{max,s}, \text{ for all } s\in S \quad (2)$$

$$\text{Policy set: } r_s \ge \min\left\{\frac{C_p}{\sum_{s\in S_p} n_s}, b_{max,s}\right\}, \text{ for all } s\in S_p \quad (3)$$

$$\text{*The objective function* is: } \sum_{s\in S} w_s r_s \Rightarrow \max \quad (4)$$

* **Michael D. Logothetis, Ioannis D. Moscholios and George K. Kokkinakis**
Wire Communications Laboratory, Dept. of Electrical & Computer Engineering, University of Patras, Greece. E-mail: m-logo@wcl.ee.upatras.gr

where $b_{min,s}$, $b_{max,s}$ are the min, max BR per call of $s \in S$, respectively, and w_s is a weight (priority) associated with each $s \in S$ (if no priority is defined, then $w_s = 1.0$ for each $s \in S$).

To solve the above LP model, we use either the primal cutting-plane algorithm (an integer programming model based on the Simplex [5]) when the bandwidth is quantized or the Simplex when the bandwidth is not quantized.

Numerical Example: Consider a single link of capacity 16 and four service-classes A – D, with weights (priorities) $w_A = 2.5$, $w_B = 4$, $w_C = 0.5$ and $w_D = 1$. The number of calls arriving to the link and the corresponding bandwidth bounds of each service-class are: $n_A = 1$ with bounds $0 \le r_A \le 4$, $n_B = 1$ with $0 \le r_B \le 2$, $n_C = 1$ with $0 \le r_C \le 10$, and $n_D = 1$ with $0 \le r_D \le 4$. Then, the LP Model describing the a-MMF is:

Constraints:

(Network Set)	$r_A + r_B + r_C + r_D \le 16$
(Traffic parameters set)	$0 \le r_A \le 4$, $0 \le r_B \le 2$, $0 \le r_C \le 10$, $0 \le r_D \le 4$
(Policy Set)	$r_A \ge \min[16/(1+1+1+1), 4]$, $r_B \ge \min[16/(1+1+1+1), 2]$, $r_C \ge \min[16/(1+1+1+1),10]$, $r_D \ge \min[16/(1+1+1+1), 4]$
Objective function:	$2.5r_A + 4r_B + 0.5r_C + 1r_D \rightarrow \max$

Solution (by the Simplex): $r_A = 4$, $r_B = 2$, $r_C = 6$, $r_D = 4$. The max-min fair share algorithm presented in [6] (page 216) produces the same results.

3. CAC incorporating the approximated MMF policy

To apply the a-MMF policy to a network we consider the following: i) The route and the bandwidth of the paths between each O-D pair are predetermined. ii) Calls arrive in a node according to a Poisson process and ask for call setup; no global controller exists to manage the bandwidth allocation process. iii) Each node services the calls with a CAC procedure on a FCFS basis for an exponentially distributed holding time. iv) The bandwidth of the newly arriving calls must be determined and, in view of fairness, the bandwidth of in-service calls may be modified. This modification is difficult to be implemented because feedback control for all service-classes is required. We avoid this modification by proposing the following procedure: A fraction of the max permitted time of the completion of call setup is devoted to the a-MMF process (Fig. 1). CAC delays the call setup for at most d sec while waiting for new arrivals within d. During d, call releases are performed. At the end of d, new calls are grouped according to their destination; then the a-MMF policy is applied among those towards the same destination taking into account only the *avl* of the O-D path. Having assigned the bandwidth to each call, the call-setup process starts for all waiting calls on a FCFS basis. More precisely, the proposed procedure is as follows:

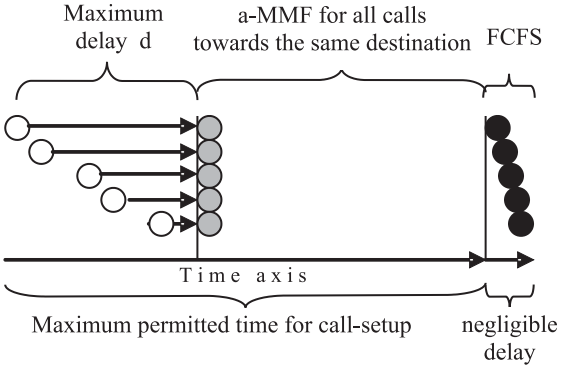


Fig. 1. The proposed new CAC scheme.

Approximated MMF Policy and Call Setup

For all calls towards the same destination:

(1) ***if*** $avl < \Sigma\, b_{min}$ ***then:*** *(select scenario a, or b, or c)*
   a. All calls are blocked and lost.
   b. Call setup is performed in the shortest path (SP) only, with b_{min}, on a FCFS basis.
   c. Call setup is performed with b_{min}, on a FCFS basis (not only in the SP). ***Endif.***

(2) ***if*** $avl \ge \Sigma\, b_{max}$ ***then:***
   Call setup is performed with the b_{max}. ***Endif.***

(3) ***if*** $\Sigma\, b_{min} \le avl < \Sigma\, b_{max}$ ***then:***
   if *multiple paths exist between* O-D ***then:***
   Call setup is performed (on a FCFS basis):
   - In the SP with b_{max}, indicating a higher priority in the use of SP
   - In the alternative paths (AP) except the last, with b_{min}, indicating a lower priority in using them.
   - In the last path according to the a-MMF policy.

   else if *a single path exists between* O-D ***then:***
   Call setup is performed according to the a-MMF policy, on a FCFS basis. ***Endif.***

4. Application example – evaluation

We consider the network of Fig. 2, which accommodates 3 service-classes. The 1st is the telephone service of 64 Kbps, while the 2nd and the 3rd are elastic services capable to change their bandwidth rates per call, from 128 to 384 Kbps, and from 768 to 1536 Kbps, respectively, in steps of 64 Kbps. We dimension (design) the network assuming an equal traffic-load among the switching pairs, which is 240, 40 and 10 erl for the 1st, 2nd and 3rd service-class, respectively, distinguishing the traffic flow directions. The elastic service-classes behave as CBR services with the max BR (at the NDP only – not at the network operation phase (NOP)). We assume a 3% grade-of-service (end-to-end CBP) for each service-class achieved by applying the trunk reservation (TR) policy to the 1st and the 2nd service-class (with TR parameters of 1472 and 1152 Kbps, respectively) to benefit the 3rd service-class. The assigned bandwidth of 53.76 Mbps to each switching pair is calculated by the Erlang Multi Rate Loss Model (with TR) [7], and split to

38.40 and 15.36 Mbps for the shortest and the alternative path, respectively. The link capacities (see Fig. 2) are produced by summing up the assigned bandwidth of the paths passing through them. At the NOP, we use the TR parameters of 704 and 640 Kbps for the 1st and the 2nd service-class, respectively.

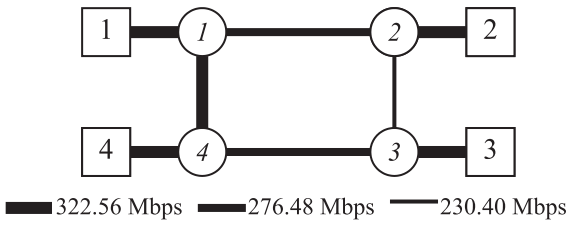


Fig. 2. Testbed network.

We have built a network simulator for simulating the call-level operations of connection-oriented networks [8]. Regarding the simulation of elastic service-classes, we model them as a concatenation of constant bit rate services, by exploiting the notion of "total offered traffic-load" (the product of: arrival rate x holding time x required bandwidth per call) [9]. For example, suppose an elastic service-class with b_{max} = 2 Mbps (per call) and b_{min} = 1 Mbps (per call). Let assume that such a call arrives at the network and initially asks to be serviced by its b_{max} = 2 Mbps for 150 sec (holding time). At that time, the network cannot provide the bandwidth of bmax, but of 1.5 Mbps. Then the holding time of this call will be extended to 200 sec, according to the total offered traffic load principle (since 2 * 150 = 1.5 * 200).

CBP's presented herein are mean values of 6 simulation runs. The simulation is done according to the proposed CAC procedure. We consider 3 values of *d*, 1.8, 2.4 and 3 sec, for the a-MMF application (we borrow these values from the possible dial-tone delay in telephony [9]). We choose the scenario **c** which is the best as far as the resultant CBP is concerned since it gives more opportunities to a call to be conveyed than **b** or **a** (similarly **b** is better than **a**) [10]. Table I comparatively presents the mean CBP of the 3rd service-class for the three scenarios and justifies the choice of **c** (lower CBP). The max CBP of the network that occurs for the 3rd service-class is shown in Fig. 3 for d = 3 sec, with or without the a-MMF (no FBA policy at all). The latter means that an arriving call immediately asks for call setup with a bandwidth between b_{min} and b_{max} (inclusive); blocking occurs when the *avl* is less than the b_{min}. The design-offered-traffic-load (corresponding to 0% of traffic fluctuation) fluctuates randomly among the switching pairs according to the uniform distribution by a max of x% (where x = 20, 40, ..., 100 – horizontal axis of Fig. 3) [7]. This type of fluctuation represents the "complete random" fluctuation between the O-D pairs, where not only the traffic fluctuation is random (in some O-D pairs the traffic increases, while in some other O-D pairs the traffic decreases), but also the choice of O-D paths where the traffic fluctuation occurs. Fig. 3 shows significant improvement of the max CBP when the a-MMF is applied. In Table II we present the CBP of the 3rd service-class for the three values of *d*. The last-column values of Table II coincide with CBP shown in Fig. 3. The confidence interval indicated in Table II is of 95%. We expected a decrease in CBP when d increases because more calls can take part in FBA each time the a-MMF is applied, utilizing better the *avl*, thus reducing the CBP. However, the results show that this happens only for very large traffic fluctuation. The 2nd service-class behaves similarly to the 3rd service-class, while the 1st service-class (CBR) behaves contrary to the elastic service-classes (Fig. 4). That is, when we get the best CBP for the elastic service-classes, we get the worst CBP for the CBR service-class.

Comparison of the three different scenarios based on the CBP. Table 1

Traffic Fluctuation (%)	Mean CBP		
	Scenario a	Scenario b	Scenario C
0	0.052	0.051	0.049
20	0.091	0.089	0.086
40	0.164	0.160	0.155
60	0.227	0.225	0.216
80	0.281	0.275	0.263
100	0.322	0.318	0.302

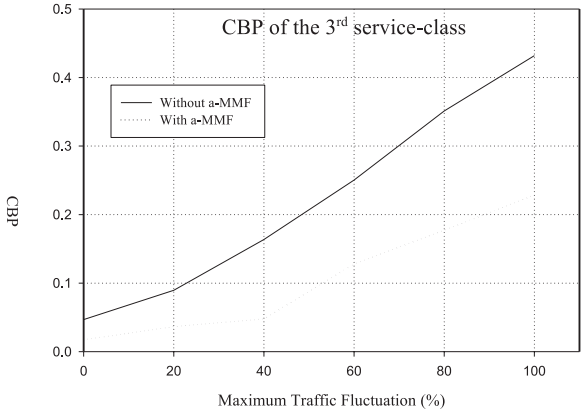


Fig. 3. Maximum CBP with and without a-MMF policy.

Maximum CBP with a-MMF policy versus d. Table 2

Traffic Fluctuation	Delay d=1.8 sec CBP (%)	Delay d=2.4 sec CBP (%)	Delay d=3 sec CBP (%)
0%	1.7 ± 0.30	1.7 ± 0.42	1.7 ± 0.29
20%	3.7 ± 0.48	3.7 ± 0.33	3.7 ± 0.48
40%	4.9 ± 0.43	4.8 ± 0.37	4.7 ± 0.45
60%	13.0 ± 1.32	12.9 ± 0.93	12.8 ± 1.06
80%	18.1 ± 1.38	18.1 ± 1.64	17.7 ± 1.53
100%	23.2 ± 1.60	22.6 ± 0.48	22.6 ± 1.57

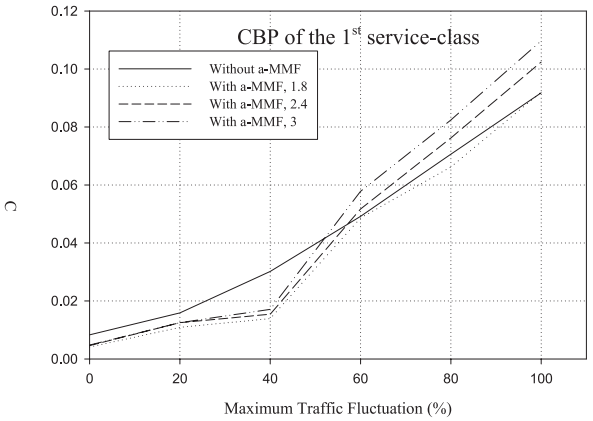


Fig. 4. Influence of the elastic traffic on the CBR traffic.

5. Conclusion

We propose a CAC procedure whereby we apply an FBA policy to the entire network. The critical part of the proposed CAC procedure is the definition of the delay. However, the high computational power of today's computer systems advocates a successful implementation of the proposed procedure. We delay calls at call setup for *d* sec, so as to apply the a-MMF policy. The available bandwidth of an O-D path is shared among calls of the same destination, whereas the bandwidth of in-service calls is not modified. Simulation results show significant lower CBP in the presence of the a-MMF policy. That is, the new CAC scheme not only achieves FBA but also improves substantially the call-level network performance. The influence of the delay *d* on CBP is small.

References

[1] HOU, Y. T., LI, B., PANWAR, S. S., TZENG, H.: *On network bandwidth allocation policies and feedback control algorithms for packet networks*, Computer Networks 34 (2000), pp. 481–501.

[2] FAHMY, S., JAIN, R., KALYANARAMAN, S., GOYAL, R., VANDALORE, B.: *On Determining the Fair Bandwidth Share for ABR Connections in ATM Networks*, in Proc. ICC'98, June 1998, pp. 1485 – 1491.

[3] MA, Q., STEENKISTE, P., ZHANG, H.: *Routing High-bandwidth Traffic in Max-Min Fair Share Networks*, in ACM Proc. SIGCOM'96, Stanford, CA, August 1996, pp. 206–217.

[4] MOSCHOLIOS, I., LOGOTHETIS, M., KOKKINAKIS, G.: *A Parametric Linear Programming Model Describing Bandwidth Sharing Policies for ABR Traffic*, in Proc. 8th International Conference on Advances in Communication & Control, COMCON 8, Rethymna, Crete/Greece, June, 2001.

[5] WAGNER, H. M.: *Principles of Operations Research.* Prentice-Hall: Englewood Cliffs, New Jersey 1969.

[6] KESHAV, S.: *An Engineering Approach to Computer Networking.* Addison-Wesley: Massachusetts 1998.

[7] LOGOTHETIS, M., SHIODA, S.: *Medium-term centralized virtual path bandwidth control based on traffic measurements*, IEEE Trans. Commun., vol. 43, pp. 2630–2640, October 1995.

[8] LOGOTHETIS, M., LIOTOPOULOS, F.: *A Batch-type, Time-true ATM Network Simulator – Design for Parallel Processing*, International Journal of Communications Systems, Vol. 15, 8, pp. 713–739, October 2002.

[9] AKIMARU, H., KAWASHIMA, K.: *Teletraffic,* Springer-Verlag: 2nd edition, Berlin, 1999.

[10] LOGOTHETIS, M., MOSCHOLIOS, I., KOTZINOS, N.: *Network Simulation with Fair Bandwidth Allocation Policy and Centralized Path Bandwidth Control for Elastic Traffic*, in Proc. SoftCOM 2002, pp. 85–89.

https://doi.org/10.26552/com.C.2004.4.17-21

Roman Jarina – Michal Kuba *

SPEECH RECOGNITION USING HIDDEN MARKOV MODEL WITH LOW REDUNDANCY IN THE OBSERVATION SPACE

Current speech recognition systems usually model a speech signal as a finite-state stochastic process, in which acoustic observations are obtained through short-term spectral analysis. The model has to deal with several thousands of speech parameters during one second of utterance. A great redundancy in the parameters makes processing computationally very expensive. We propose a combination of 2-D cepstral analysis and continuous Hidden Markov Model with a small, optimally designed, number of states and acoustic observations. 2-D cepstrum efficiently preserves spectral variations of speech and yields uncorrelated parameters in both time and frequency. The system is evaluated on isolated word recognition task in Slovak language. Promising preliminary results are presented.

1. Introduction

A majority of the state-of-art ASR systems models a speech signal as a finite-state stochastic process to handle the great variability found in human speech. Acoustic observations of speech are obtained through short-term spectral analysis. One speech feature vector, which forms one observation, usually consists of static spectral features (e.g. 13 cepstral coefficients) and their time derivatives that determine temporal variations of speech spectra. Such parameterisation is not optimal from either statistical or perceptual point of view. A great redundancy in the speech features makes further processing computationally much more costly.

Two-dimensional cepstral analysis preserves spectral variations more efficiently while also yields uncorrelated features in both time and frequency. In this paper we are giving ourselves the question whether speech signal can be "observed in time" by a much smaller number of the feature vectors (or observations) than it is common in the present ASR systems. We propose a combination of 2-D cepstral analysis and left-right continuous Hidden Markov Model with a small (optimally designed) number of states and acoustic observations. The system is evaluated on an isolated word recognition task in Slovak language. Promising preliminary results are presented.

2. On Markov Modelling of Speech

The most commonly used model for ASR is the first-order Markov process. The popularity of this method lies in its model simplicity, ease of training and acceptable recognition precision on certain tasks [1]. Thus time-varying characteristics of a speech signal are described through a chain of static states. Each model has a number of states that approximates the number of distinct acoustic or phonetic events in the unit being modelled. Such units are commonly words or subword units as phonemes, diphones, etc. Since details of the Markov model's operation in speech analysis must be inferred through observations of speech, the states of the model are hidden. Such model is usually referred to as Hidden Markov Model (HMM). A HMM constitutes of the state-observations and transition probabilities. The transition probabilities provide a mechanism for connection of the states, and for modelling variations in speech duration and articulation rates. The statistical distributions of speech features define acoustic observations. Mel-frequency cepstral coefficients (MFCC) are the most popular features of speech [1] [2]. The following discussion deals with the HMM with continuous distributions of speech features.

The continuous HMM is defined by a set of parameters as follows

$$\lambda = (A, B, \pi, N), \tag{1}$$

where A is the transition probabilities matrix, B is the output probabilities matrix, π is the vector of initial probabilities, and N is the total number of states (Figure 1). Let $\{y_1, y_2, ..., y_t\}$ and $\{s_1, s_2, ..., s_t\}$ be time sequences of acoustic observations and related hidden states respectively then the probability of taking a transi-

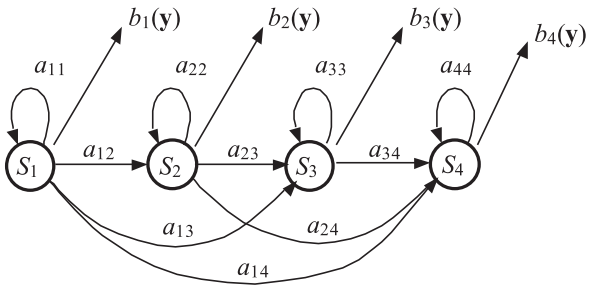


Fig. 1. Four-state left-right HMM

* **Roman Jarina, Michal Kuba**
Department of Telecommunications, Faculty of Electrical Engineering, University of Žilina, Žilina, Slovakia
E-mail: roman.jarina@fel.utc.sk, michal.kuba@fel.utc.sk

tion from the state i to the state j is $a_{ij} = P(s_t = j \mid s_t - 1 = i)$, and $A = \{a_{ij}\}$; The output probability of emitting the feature vector y_t when state i is entered, is $b_i(y_t) = P(y_t \mid s_t = i)$, and $B = \{b_i(y_t)\}$.

In the first-order MM is the assumption $P(s_t \mid s_{t-1}, s_{t-2}, ..., s_1) = P(s_t \mid s_{t-1})$. That means only neighbouring states depend on each other, and past history, except the neighbour previous state, is ignored in signal modelling. Although this simplification enables much easier computation, it represents rather inaccurate modelling since speech perception is conditional on much longer time period. Human short-memory lasts several seconds whereas ASR HMM "views" only the past speech frames. To enable to incorporate a longer past time period (2 or more previous frames), 1st and 2nd time derivates (or differences) of MFCC, referred to as *delta* (Δ) and *delta delta* (Δ^2) coefficients, are added to the speech feature set. Thus speech signal is represented by time sequence of the feature vectors. Each feature vector usually consists of 13 MFCC, which represent short-time spectrum, 13 ΔMFCC and 13 Δ^2MFCC, which represent spectral dynamics. The feature vectors are computed on a frame-by-frame basis. In such a case the one second long utterance is represented by almost 4,000 parameters (if the frame length is 10 ms).

3. Spectral Dynamics Represented by the Modulation Spectrum

Obviously, the prime carrier of the linguistic information is changes of the vocal tract shape. Such changes are reflected in changes of the spectral envelope of the speech signal. Furui has already shown that spectral transitions play very important role in speech perception [3]. If the spectral envelope is represented by a set of coefficients (e.g. MFCC or Filter-Bank energies), each coefficient varies gradually within each distinct segment of speech and thus forms the time contour (magnitude of the coefficient as a function of time). A shape of such contours, or spectral dynamics, is usually described explicitly by adding delta coefficients to the feature vector.

The procedure of delta coefficients computation can be seen as a simple FIR filtering applied on time trajectory of each of the spectral component (FB-energies in spectral domain or MFCC in cepstral domain). More general approach to filtering of these time trajectories, known as RASTA processing, was introduced and extensively studied by Hermansky et al. [4], [2]. For illustration, a spectrogram of the Slovak word "osem" is shown in Figure 1. Spectral transitions between vowels are clearly visible.

3.1 Two-Dimensional Cepstrum

The mel-frequency 2-D cepstral coefficients are computed by applying 2-D cosine transform on the block of consecutive spectral vectors (mel-FB energies) as follows [5]

$$\hat{S}_{FB}(k, m) = \log(\,|\, S_{FB}(k, m) \,|\,),\ 0 \le k \le K-1,$$
$$0 \le m \le L-1, \qquad (2)$$

where $S_{FB}(k, m)$ is the mel-spaced filter bank (FB) spectrum of the frame m, K is the number of critical-width bands and L is the number of frames used in the analysis block

$$c(u, m) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_{FB}(k, m) \cdot \cos\left[\frac{(2k+1)\pi u}{2K}\right],$$
$$0 \le u \le K-1,\ 0 \le m \le L-1, \qquad (3)$$

$$C(u, m) = \frac{1}{L} \sum_{m=0}^{L-1} c(u, m) \cdot \cos\left[\frac{(2m+1)\pi v}{2L}\right],$$
$$0 \le u \le K-1,\ 0 \le v \le L-1, \qquad (4)$$

Since not all the coefficients of the matrix $C = \{C(u,v)\}$ are needed for ASR, only selected coefficients form the *TDC (Two-Dimensional Cepstrum)* feature vector.

Spectral analysis of temporal trajectories of spectral envelopes yields the *modulation spectrum* of speech. In a TDC matrix, the dimension v (in Eq. 4) represents the modulation spectrum. Between the index v and modulation frequency in Hz is the following equation

$$\theta = \frac{F_s}{n_F} \cdot N \cdot v = \frac{v}{T}, \qquad (5)$$

where θ is the modulation frequency in Hz, F_s is sampling rate, n_F is the number of frames in the analysis block, N is the length of the frame, T is the total duration of the analysis block in seconds. The human auditory system is most sensitive to modulation frequencies around 4 Hz that reflects the syllabic rate of speech. Thus the human hearing in perception of modulated signals acts as a band-pass filter with the length of the impulse response of minimally 150-250 ms [2] which is the length of 15-20 frames. The results of speech recognition experiments have shown that the components of the modulation spectrum below 1 Hz and above 16 Hz have only a minor role in both human perception and ASR [6].

In [5], [7], we have also studied discriminative properties of TDC features on discrimination of confusable Slovak consonants. In test utterances a group of 6 consonants (3 stops and 3 fricatives) were placed between the same two vowels. We studied what components of the TDC matrix are the most important for a phoneme discrimination. We used 12 cepstral coefficients along the frequency axis (dimension u in Eq. 3–4). We confirmed that only coefficients corresponding to the modulation frequencies from the range mentioned above are important. In our case the sufficient subset of TDC coefficients was between 12×5 and 12×7 whereas the coefficients with the index $v = 0$ (i.e. $\theta = 0$ Hz) were excluded. If the coefficients with the index $v = 0$ were included into the set, the recognition rate decreased rapidly, particularly for noisy speech.

3.2 2-D Cepstrum analysis in HMM framework

For ASR task, 2-D cepstrum (TDC) was first introduced by Ariki [8] who used only one TDC matrix for each word. Recently,

Jarina [5] made several experiments in which he modelled words by a small number (1 or 3) of linearly spaced TDC matrices. He applied a multilayer perceptron to discriminate the patterns formed from these matrices. Also several works on combination of TDC and continuous HMM have been reported [9–11]. Milner [9] used the TDC for each frame while Kanedera [10] used a much longer temporal window and only a small selection of TDC, which corresponds with the range of modulation frequencies between 3 and 9.5 Hz. The authors of these experiments reported increase of recognition rate when using TDC based dynamic features rather than conventional delta features.

In this paper, we investigate the combination of 2-D cepstrum and HMM from a different point of view. HMM assumes that the acoustic observations are uncorrelated. But in reality, an intra-frame correlation (i.e. between static and dynamic features) as well as a high inter-frame correlation (i.e. between successive frames) are observed. Thus the number of observations in conventional ASR systems is over-estimated. There are about 50–100 observations for one second utterance.

TDC analysis enables modelling dynamic properties of the signal implicitly. The TDC computed via 2-D cosine transform produces almost uncorrelated set of coefficients in both frequency (index u) and modulation frequency (index v) dimensions. We hypothesise that if the TDC is applied, a much smaller number of observations is necessary (due to decorrelation properties of TDC, a high redundancy in temporal trajectories of the speech envelopes can be removed). For instance, in [5] we have shown that 22 spectral vectors could be replaced by only one TDC matrix, from which only about 60–70 coefficients are needed, unlike conventional ASR systems, in which almost 800 coefficients ($MFCC+\Delta+\Delta^2$) are required for the same duration of speech. Each observation, which is formed from TDC features, will incorporate information from several hundreds milliseconds of speech which is in accordance with the time interval of the short-time memory of human perception.

4. Experiment

We designed the continuous density HMM with a reduced number of acoustical observations represented by TDC features. The model is evaluated on a Slovak isolated digit recognition task. The speech database consists of 12 Slovak words (digits 0–9, digits 1 and 2 are spelled-out as both “jeden” and “jedna”, and “dva” and “dve” respectively) uttered by 61 speakers. The database was recorded in the Department of Telecommunications in the University of Žilina. The details about the speech database are summarised in table 1. The ASR system is designed as speaker-independent. We used a different non-overlapping sub-set of the database uttered by different group of speakers for training and testing. One HMM was created for each class.

Train and test speech database — Table 1

Recognition task	Isolated Slovak digits
Number of speakers	61 (40 for training + 21 for testing)
A/D conversion	sampling frequency = 8kHz, resolution = 8bits/sample, telephone quality
Number of records	4 records per word per speaker, 12 words 48 × 61 = 2928 records in total

4.1 ASR front-end

A speech signal is analysed in frames. The analysis procedure is depicted in Figure 2 . First, the signal is pre-emphasised by the 1st-order FIR filter with $k = 0.97$. The sliding 30 ms long window is used with the 20 ms shift. That means 1 second of the signal is split into 50 frames. FFT spectrum and 23 mel-FBE are computed for each frame. The frames are grouped into blocks of 12 frames with 6 frames overlap. TDC matrix is computed from each block (Eq. 2–4). Only one quarter of the TDC matrix forms a feature vector y (i.e. acoustic observation), which consists of 50 coefficients as follows

$$y = \{y_d\} = \{C(u,v),\ u = 1, 2, ..., 10;\ v = 1, 2, ..., 5\},$$

$$d = 1, 2, ..., 50. \qquad (6)$$

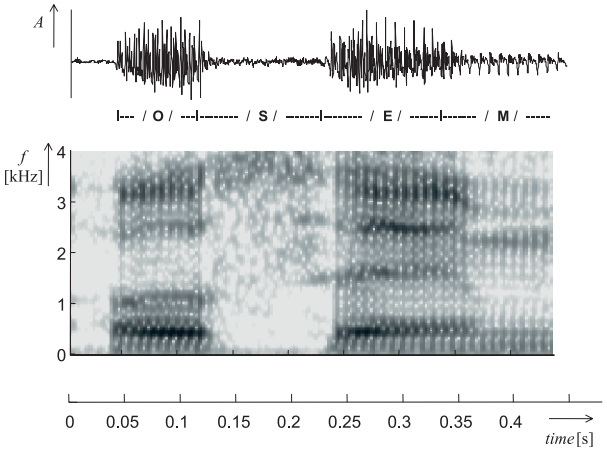


Fig. 2. Speech waveform and spectrogram of the Slovak word “osem”

Note the first row ($u = 0$) and first column ($v = 0$) of the TDC matrix are removed.

4.2 HMM Design

Let PT be the probability that a training pattern has T acoustic observations, and let T_{max} and T_{min} be maximum and minimum number of acoustic observations through all training patterns respectively. Then we propose that the number of states is given by the equation

$$N = \underset{T=T_{min}}{\overset{T_{max}}{ArgMax}} \{P_T\} \quad (7)$$

The continuous output probabilities $b_i(y)$ for each state of HMM are modelled by PDFs with a mixture of multivariate Gaussians as follows

$$p(\boldsymbol{y}\,|\,i) = \sum_{k=1}^{M} p(\boldsymbol{y}, k\,|\,i) = \sum_{k=1}^{M} p(k\,|\,i)p(\boldsymbol{y}\,|\,k, i) =$$

$$= \sum_{k=1}^{M} m_{i,k} N(y, U_{i,k}, \mu_{i,k}) \quad (8)$$

where $\boldsymbol{y}$ is the given observation, $U_{i,k}$ and $\mu_{i,k}$ are covariance matrix and mean feature vector of the k-th Gaussian component in the i-th state of HMM, and $m_{i,k}$ is the weight of k-th components. M is the number of mixture components. The term $N(y, U, \mu)$ means multivariate joint Gaussian PDF of acoustic observations $\boldsymbol{y}$, defined as

$$N(y, U, \mu) = \left[(2\pi)^{\frac{D}{2}} \sqrt{|\det U|} \right]^{-1} e^{-\frac{1}{2}(y-\mu)^T U^{-1}(y-\mu)} \quad (9)$$

where D is number of the features in one acoustic observation (D=50).

During HMM initialisation, acoustic observations of each training pattern have to be divided among N states of the model. Due to a highly reduced number of observations (see Figure 2) we proposed the following initialisation procedure: First only training utterances with the number of observations greater or equal to the number of states are selected. Their observations are allocated to the states of HMM, and centroids for each state are computed. Observations of the rest utterances are allocated to the states by a modified K–mean algorithm. Then all the observations are iteratively re-located to the states of the model by a modified K–mean algorithm. In this stage, the transition probabilities are estimated. The observations in each state are grouped to M clusters by VQ, and hyper-parameters of emission probabilities U, μ, m are estimated. A training of HMMs was performed by the well-known Baum-Welch algorithm using MLE criterion.

We tested 3 types of covariances: full, diagonal and spherical. Spherical covariance is estimated as $\boldsymbol{U} = MSE \,.\, \boldsymbol{I}$, where $\boldsymbol{I}$ is identical matrix. Mean Square Error of a cluster is given as follows

$$MSE = \underset{y}{E}\{(\boldsymbol{y} - \mu)T\,(\boldsymbol{y} - \mu)\} \quad (10)$$

4.3 Evaluation

ML classification using both the feed-forward and Viterbi algorithms were applied, and almost the same results were obtained for both methods (difference was only in speed of log-likelihood computations where Viterbi algorithm suits better). We evaluated PDF mixtures with 1, 2, 4, or 8 Gaussians. The results are summarised in Table 2.

The spherical and diagonal covariance gives similar results. The results, when full covariance is used, do not meet theoretical expectations (theoretically, the full covariance should be the most precise). But if we look closely at the third row of the table, we notice that the recognition rate falls down rapidly when a number of Gaussian components (and thus a number of parameters) is increasing. This effect has occurred because of an insufficient number of training data to tune all the parameters correctly (see Table 1, only $40 \times 4 = 160$ records per word were available for training).

Recognition rate in dependence of HMM set-up — Table 2

Number of Gauss. components	1	2	4	8
	Recognition rate			
Spherical cov.	89.8 %	92.2 %	93.2 %	92.6 %
Diagonal cov.	90.6	92.6	92.3	92.3
Full covariance	87.2	66.0	20.7	9.3

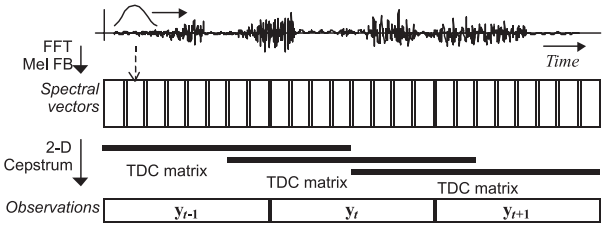


Fig. 3. Procedure of speech signal pre-processing for HMM

5. Conclusion

The ASR HMM with a reduced number of observations is proposed. In the model, one second of speech is described by only about 400 features what is ten times less than in conventional ASR systems. The performance of the system was examined on a speaker-independent isolated digit recognition task. The satisfactory results as seen in Table 2 are for the HMM with both the spherical and diagonal covariance. The best recognition rate is 93.2%. We suppose that the recognition rate will further increase if a bigger amount of training data is available. We believe that if we re-train the model on a much larger database (1000–2000 speakers is common for development of speaker-independent ASR) the model will be competitive with the ASR systems that use conventional methods. A great advantage of the proposed model is ease of computation (particularly for spherical covariance) and very fast signal processing and model training. It is suitable for application with a limited computation performance and power (e.g. mobile devices).

References

[1] O'SHAUGNESSY, D.: *Interacting with computer by voice: Automatic Speech Recognition and Synthesis,* Proceedings of the IEEE, Vol. 91, No. 9 (2003) 1272–1305.
[2] HERMANSKY, H.: *Should recognizers have ears?*, Speech Communications 25 (1998) 3–27.
[3] FURUI, S.: *On the role of spectral transition for speech perception,* J.Acoust.Soc.Am. 80(4), (1986) 1016–1025.
[4] HERMANSKY, H., MORGAN, N.: *RASTA processing of speech,* IEEE Trans. Speech Audio Process. 2(4), (1994) 578–589.
[5] JARINA, R.: *Kepstrálno-spektrálny model pre rozpoznávanie rečových signálov,* dissertation, University of Žilina (1999).
[6] KANEDERA, N, ARAI, T., HERMANSKY, H., MISHA, P.: *On the importance of various modulation frequencies for speech recognition,* Proc. Eurospeech'97, Rhodos, Greece (1997) 1079–1082.
[7] JARINA, R.: *Study of discriminative properties of two-dimensional cepstrum analysis for speech recognition,* Proc. RADIOELEKTRONIKA' 99, Brno, Czech, (1999) 168–171.
[8] ARIKI, Y., MIZUTA, S., NAGATA, M., SAKAI, T.: *Spoken-word recognition using dynamic features analysed by two-dimensional cepstrum,* Proc. IEE, Vol. 136, Pt.I, No.2, (1989) 133–140.
[9] MILNER, B.P., VASEGHI, S.V.: *Speech modelling using cepstral-time feature matrices and Hidden Markov Models,* Proc. of the IEEE conf. ICASSP '94, Vol.I, Adelaide, Australia (1994) 601–604.
[10] KANEDERA, N., HERMANSKY, H., ARAI, T.: *Desired characteristics of modulation spectrum for robust automatic speech recognition,* Proc. of the IEEE conf. ICASSP'98, Seatle, USA (1998).
[11] JANČOVIČ, P, MACHO, D., NADEU, C., ROZINAJ, G.: *Feature selection in cepstral-time matrices for clean and noisy speech recognition,* Proc. TEMPUS-TELECOMNET workshop ITTW'98, Barcelona, Spain, July (1998) 28–36.

https://doi.org/10.26552/com.C.2004.4.22-27

kamUNikácie
COMMUNICATIONS

Peter Kortiš – Vladimír Olej – Karol Blunár *

AN EUGENETIC ALGORITHM FOR DELAY-CONSTRAINED MINIMUM-COST ROUTING OF MULTIPOINT CONNECTIONS

A special kind of genetic algorithm - eugenic algorithm is presented for constructing minimum-cost tree with delay constraints. Telecommunication network is represented by an undirected graph, which uses three independent metrics: cost, delay and capacity. The efficiency of two different bit representations of individuals (adjacent matrix, list of vertices) is compared.

1. Introduction

Recent advances in telecommunication and computer technology have led to a new variety of applications based on communication of multiple parties. This brings a new dimension into the existing telecommunication networks, routing of multipoint connections. Multipoint connections are used by conference services, which need to establish point-to-multipoint and multipoint-to-multipoint connections [6]. The main goal of routing algorithm is to find connection, which fulfils conditions of specific service, as are: free bandwidth, delay, delay variation, etc. The routing algorithm has to find such a connection, which fulfils all the given conditions, and the cost of this connection is as low as possible. The cost of connection depends on the length (geographical) of the connection, number of used network resources, utilization, delay of separate links in connection, bandwidth, etc. The routing algorithm prefers less utilized parts of the network, which results in efficient usage of network capacity and consequently the blocking probability will decrease. The crux of routing of multipoint connections is described in [9], where many deterministic algorithms are presented. Nowadays, algorithms for routing of multipoint connections could be based on genetic algorithm (GA) [3, 7, 8] or eugenic algorithm (EuA) [1, 4, 5].

2. Network model and problem definition

Having a telecommunication network represented by the graph

$$G = (V, E), \tag{1}$$

where the switching nodes of telecommunication network are represented by the set of vertices V and communication links are represented by the set of edges E. Consider three non-negative real functions associated with each edge (link) e:

$$C(e) \in \Re^+ + \{0\},\ D(e) \in \Re^+ + \{0\},\ F(e) \in \Re^+ + \{0\}. \tag{2}$$

The function $C(e)$ determines the cost of the link, which depends mainly on its utilization. The link delay $D(e)$ is the delay of a data packet experiences on the corresponding link and the function $F(e)$ determines free capacity of the link.

Given a set of nodes, which have to be connected:

$$U = \{u_1, u_2, u_3, ..., u^m\};\ (U \subseteq V), \tag{3}$$

This set represents the nodes in the network, which the subscribers of one multipoint connection are connected to. Every vertex of set U must be connected with every single vertex of set U. The problem of routing multipoint connection in the telecommunication network could be formulated as the problem of finding the subgraph G', which is usually tree T. Consider the subgraph G' to be the tree $T \subseteq G$, where the set of vertices of the tree T includes all vertices from the set U. The desired tree must fulfil these two criteria:

- Free capacity of all edges must be equal or higher than the required value Φ: $\forall e \in T : F(e) \geq \Phi$, (4)
- Delay between all couples of vertices from set U must be lower or equal to the maximum acceptable delay $\Delta_{i,j}$ defined for a selected couple of vertices:
  $\forall i, j \in 1 .. m; i \neq j; P = Path(u_i, u_j) : D(P) \leq \Delta_{i,j}$, (5)
  where P is the path from vertex u_i to vertex u_j.

Cost C_T of the tree T is defined as follows:

$$C_T = 2 \sum_{e \in T} C(e) \tag{6}$$

The goal of the routing algorithm is to find the tree, which fulfils the criteria (4) and (5) and the cost of the tree is as low as possible.

If the cost of the tree is minimal, the tree is called Steiner tree and the vertices of the tree, which are not in the set U, are called Steiner vertices [2].

* **Peter Kortiš[1], Vladimír Olej[2], Karol Blunár[1]**
[1]Department of Telecommunications, Faculty of Electrical Engineering, University of Žilina, Veľký Diel, 010 26 Žilina, Slovak Republic, E-mail: Peter.Kortis@fel.utc.sk
[2]Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic E-mail: Vladimir.Olej@upce.cz

3. Eugenic algorithm

Eugenic algorithm [1, 4, 5] differs from GA [3] in many ways. The most significant difference is the method of creation of new individuals. Whereas GA creates new individuals randomly using crossover, mutation and inversion operators, the EuA creates a new individual using the analysis of selected individuals in population. Even though the analyses are deterministic, there are many random steps in the algorithm. Eugenic algorithm could be defined as follows [1, 4, 5]:

1. pop = random-population ( ) /* Randomly initialize the population pop */
   $L = 0$ /* L counts the number of individuals created by EuA */
2. **DO WHILE** ($L < L_{max}$) /* Creation of new individuals */
   $G = \{g_1, g_2, ..., g_l\}$ /* G is the set of unset genes */
   $rpop = pop$ /* rpop is the copy of population dedicated for restriction */
3. **DO WHILE** non-empty (G) /* Choose unset gene and set its allele */
4. g_g = most-significant-gene (rpop, G)
5. $g_{g,a}$ = set-allele (g_g) /* Set allele $g_{g,a}$ of the gene g_g */
6. Remove g_g from set G /* Remove unset gene g_g from the set of unset genes */
7. E = epistasis-of-population (rpop, G) /* E is epistasis of restricted population rpop */
8. p_r = probability-of-restriction (E) /* Perform restriction, if epistasis E is high */
9. **IF** (random[0,1] $< p_r$) **THEN**
10. rpop = restrict-population (rpop, $g_{g,a}$)
    END WHILE
11. U = lowest-fitness-individual (pop)
12. $pop_U = g$ /* Replace lowest-fitness individual with the new individual */
13. $L = L+1$
    END WHILE

The EuA begins by generating initial random population. The algorithm creates a new individual *g*. The algorithm continues in creation of new individuals until the maximum number of individuals L_{max} has been created. Every new individual is created step by step.

The EuA constructs a new individual step-by-step (allele-by-allele) using the analysis of selected part of population (*rpop*). In the initialization phase of the algorithm the set of unset genes is full. This set includes genes, where the alleles are still unknown and will be set in the next steps. Alleles of unset genes are not generated randomly, but according to their significance. The most significant genes are set as the first and the least significant as the last. The most significant gene and its allele is determined from the restricted population *rpop*. The allele of the most significant gene is selected randomly, but in accordance with its significance. When the allele of the gene is set, the gene is removed from the set of unset genes *G* and the restriction of population *rpop* could be performed. Whether the restriction is performed or not, depends on the amount of interdependence among a genotype's alleles called epistasis. The higher the epistasis, the higher is the probability that a population is restricted to only individuals containing selected allele of the most significant gene $g_{g,a}$. All individuals that do not have allele $g_{g,a}$ are excluded from the population *rpop*. The primary goal of restriction is to leave in population such individuals who are similar to the individuals currently being created. This technique enables to estimate which genes are more significant for creating individual and which allele is more suitable. If the gene significance analysis were performed over the entire population, there would be no regard for the relationship among separated genes. If one of the genes is set to the same allele in the entire population, then this gene is called gene without diversity. If such a gene appeared in population, its allele would never change. Because of this EuA has mechanism, which allows setting another allele for the gene without diversity. The probability of this change is equal to parameter creation rate CR. This procedure repeats until all genes of new individual *g* are set. If all alleles of newly created individual are determined, the fitness of this individual is computed and the newly created individual will replace the lowest fitness individual in the population.

To find the most significant gene and allele and compute its epistasis, it's necessary to analyze fitness of separate alleles of individuals from restricted population *rpop*. The analysis is usually based on the computation of allele fitness average of unset genes. Calculated averages are normalized to probabilities of allele presence for each gene $p(g_{g,a})$. If the probability $p(g_{g,a})$ is not one or zero and the absolute value of difference of probabilities $p(g_{g,a})$is maximal, the gene $g_g \in$ G is the most significant gene. This could be expressed as follows:

$$g_g = \underset{g_k \in G: p(g_{k,0}) \neq 0 \wedge p(g_{k,0}) \neq 1}{\operatorname{argmax}} \{ \left| p(g_{k,0}) - p(g_{k,1}) \right| \}. \quad (7)$$

The most significant allele of gene g_g is the allele with the highest probability of presence, i.e.:

$$g_{g,a} = \underset{g_{g,a}}{\operatorname{argmax}} \{p(g_{g,0}), p(g_{g,1})\}. \quad (8)$$

The amount of interdependence among a genotype's alleles (epistasis) could be calculated as follows:

$$E = 1 - \left| p(g_{j,0}) - p(g_{j,1}) \right|, \quad (9)$$

where gene $g_j \in G \setminus \{g_g\}$ is the second most significant gene after gene g_g.

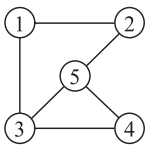
If the allele probabilities are approximately equal (i.e. approximately equal to 1/2), the epistasis goes to one and if one of the probabilities goes to zero, the epistasis will equal zero (see expression 9).

The probability of restriction p_r is commensurable to epistasis and for the majority of applications it's sufficient, if probability is equal to epistasis $p_r = E$. The result of this fact is that the restriction of the population is more likely, when the relation among alleles (epistasis) of unset genes is low. The restriction is not performed if the number of individuals after restriction would be lower than minimum acceptable number of individuals *MinN*, otherwise the restriction could be performed with the probability p_r.

4. Model design for routing of multipoint connections, simulation results and analysis.

Using the principles described in section 3, EuA for routing of multipoint connections was designed. The goal of the algorithm is to find minimum cost connection with regard to criteria (4) and (5). The proposed algorithm uses one of two chosen representations of individuals (connections): adjacent matrix and list of vertices.

The adjacent matrix is the square matrix with dimension equal to the number of nodes in the network (fig. 1). The matrix elements could only be set to one of two values (0 or 1), which determines, whether the connection between two particular nodes exists or not. If the matrix element on the *i*-th row and *j*-th column is 1, then the connection from node *i* to node *j* exists and vice versa.



$$G = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Fig. 1: The example of simple network and corresponding adjacent matrix

The list of vertices is the line vector with dimension equal to the number of nodes in the network. The vector element determines, if corresponding node must be presented in the connection or not. If the vector element is set to 1 the node is presented in connection, otherwise the corresponding node could be presented, depending on the state of the network. The tree is created using the minimum spanning tree (MST) algorithm, which respects only the delay between nodes.

The MST algorithm is described as follows [10]:

1. Given a graph G, construct a new graph $G' = (U, E')$ where delay $D(e')$; $e' = (u_i, u_j)$ is the length of the shortest path from u_i to u_j in G.
2. Construct a minimum spanning tree T' in G'.
3. Convert the edges in T' to paths in G to form a solution T.

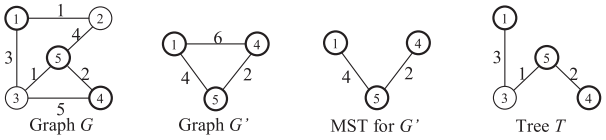


Fig. 2: Minimum spanning tree algorithma djacent matrix

The list of vertices can represent only trees, unlike the adjacent matrix, which could represent subgraphs too.

The routing algorithm consists of two phases:

- pre-processing phase
- eugenic algorithm

The main goal of pre-processing phase is to reduce the graph G to the graph G_R with edges that represent links with sufficient free capacity. The reduced graph G_R includes only edges e with free capacity equal or higher than minimum required capacity Φ i.e. $F(e) \geq \Phi$. After the reduction, the shortest paths among all vertices from set U are computed and compared with maximum acceptable delay given by parameter $\Delta_{i,j}$ (5), to ascertain if exists any suitable connection, which fulfils conditions (4) and (5). If such connection does not exist, the algorithm is interrupted; otherwise the eugenic algorithm is performed.

Given the network from fig. 3, which consists of 22 nodes interconnected with bidirectional links. The cost of every link is presented as the number beside the line. The delay of every link in the graph is 10 ms and free capacity is higher than minimum required value Φ. Nodes 6, 14, 17, 19 and 21 drawn with thick circles must be connected.

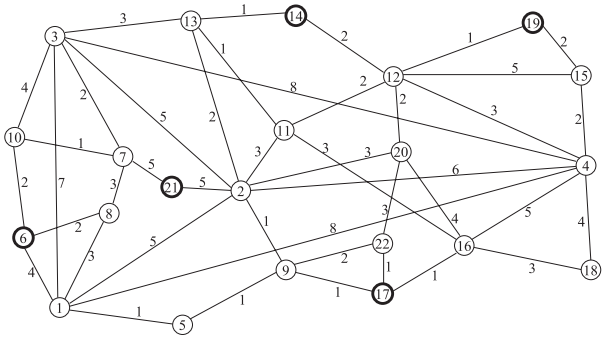


Fig. 3: The structure of the network

The proposed algorithm was tested on various network architectures, but this paper contains only results of the tests performed on the network from figure 3, because the tests on other networks give similar results. The tests were performed using both representations: adjacent matrix and list of vertices.

The cost of the resultant connection depends on these parameters as a number of generations L_{max}, size of population *MaxN*, and parameters *MinN* and *CR*. To find the effect of particular parameters to algorithm efficiency the following experiment was done. The average value of minimum, maximum and average cost of connection were calculated from 100 independent experiments for many combinations of the above mentioned parameters. Table 2 specifies the chosen values of tested parameters for both representations. The maximum acceptable delay $\Delta_{i,j}$ is set to 50 ms for all couples from the set *U*.

The sets of values for particular simulation parameters Tab. 2

L_{max}	{300; 1000}
MaxN	{10; 20; 50; 100}
MinN	{1; 0.1MaxN; 0.2MaxN; 0.3MaxN; 0.4MaxN; 0.5MaxN}
CR	{0; 0.001; 0.005; 0.01; 0.015; 0.02; 0.03; 0.05; 0.1; 0.2; 0.3}

Note: figures 4, 5, 6 and 7 depict only the average value of minimum cost because the resultant connection is represented by the lowest cost individual in the population.

The experiments show that the representation by list of vertices gives significantly better results than the representation by an adjacent matrix because the best average cost of the connection represented by the adjacent matrix is 62.86 after 1000 generations, but for a list of vertices is 48 after only 300 generations (see fig. 4 and 6). This significant difference in performance is caused by the convergence speed, which is lower for the adjacent matrix than for the list of vertices representation because the minimum cost after 300 generations is approximately 72 when the adjacent matrix representation is used. (Note: The cost in 300th generation is not included in any table of this paper because it was obtained from another simulation.)

The cost of connection is independent on a minimal acceptable number of individuals after restriction *MinN*. Stochastic char-

Min. Cost	MinN					
		1	2	3	4	5
CR	0	90	90	90	90	90
	0.001	88.86	88.78	89.32	89	89.26
	0.005	80.28	81.76	80.36	83.3	82.98
	0.01	71.96	72.48	70.4	72.84	71.84
	0.015	66.64	66.94	67.8	66.9	66.84
	0.02	64.34	65.3	63.24	66.1	65.38
	0.03	63.68	63.64	62.86	63.36	63.64
	0.05	65.42	65.76	64.84	64.82	66.06
	0.1	76.66	75.22	74.88	75.86	76.28
	0.2	88.94	89.66	89.26	88.92	89.06
	0.3	90	90	90	90	90

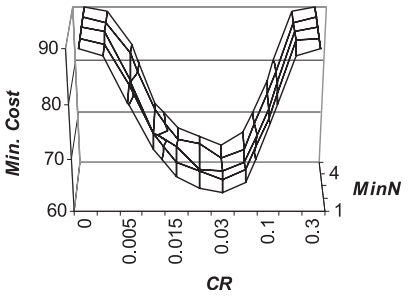


Fig. 4: Minimum cost (Adjacent matrix; MaxN = 10; L_{max} = 1000)

Min. Cost	MaxN				
		10	20	50	100
CR	0	90	90	90	90
	0.001	88.86	88.84	87.68	85.9
	0.005	80.28	76.42	73.68	84
	0.01	71.96	65.88	73.34	83.06
	0.015	66.64	64.52	75	81.92
	0.02	64.34	65.3	74.36	81.1
	0.03	63.68	66.82	76.52	79.58
	0.05	65.42	69.6	80.72	84.72
	0.1	76.66	83.6	87.52	88.98
	0.2	88.94	89.78	89.92	89.94
	0.3	90	89.96	90	90

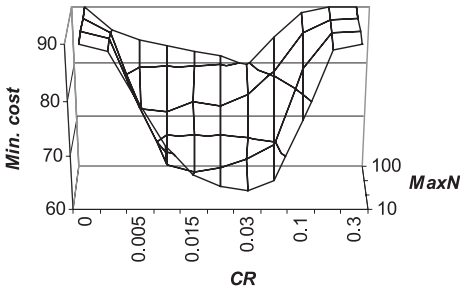


Fig. 5: Minimum cost (Adjacent matrix; MinN = 1; L_{max} = 1000)

Min. Cost	MinN						
		1	2	4	6	8	10
CR	0	50.6	51.1	50.74	50.72	50.54	50.94
	0.001	50.02	49.86	49.74	50.06	49.92	50
	0.005	49.04	48.92	48.74	48.92	48.64	48.7
	0.01	48.44	48.28	48.32	48.48	48.26	48.18
	0.015	48.16	48.12	48.2	48.16	48.18	48.28
	0.02	48.1	48.12	48.14	48.24	48.1	48.14
	0.03	48.1	48.12	48.06	48.24	48.06	48.16
	0.05	48	48.06	48	48.06	48.08	48.06
	0.1	48.06	48.14	48.02	48.04	48.12	48.04
	0.2	48.98	48.9	48.66	48.8	48.86	48.84
	0.3	50.1	49.7	49.96	49.9	49.78	49.78

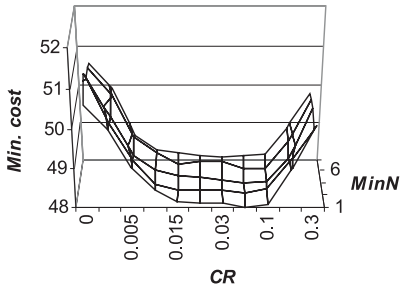


Fig. 6: Minimum cost (List of vertices; MaxN = 20; L_{max} = 300)

Min. Cost		MaxN			
		10	20	50	100
CR	0	51.64	50.6	49.62	48.54
	0.001	51.28	50.02	49.18	48.6
	0.005	49.12	49.04	48.6	48.54
	0.01	48.7	48.44	48.46	48.4
	0.015	48.66	48.16	48.74	48.36
	0.02	48.58	48.1	48.44	48.28
	0.03	48.48	48.1	48.48	48.46
	0.05	48.2	48	48.42	48.38
	0.1	48.04	48.06	48.62	48.28
	0.2	48.22	48.98	48.42	48.3
	0.3	49.46	50.1	49	48.48

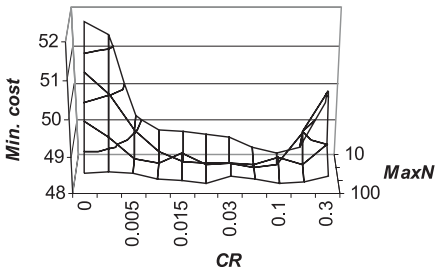


Fig. 7: Minimum cost (List of vertices; MinN = 1; L_{max} = 300)

acter of the algorithm and low number (100) of performed experiments cause small differences among minimal costs for selected values of parameter *MinN*. The total number of individuals in the population (*MaxN*) has significant effect on the cost of connection only when the adjacent matrix representation is used.

In the given adjacent matrix representation, increasing the number of individuals has negative impact on the speed of convergence to the optimal solution (see fig. 5). Despite the fact the parameter *MaxN* is significant, the parameter *CR* has bigger impact on the cost of the resultant individual. The optimal value for parameter *CR* is approximately 0.03.

If the list of vertices representation is used, the cost of the connection depends mainly on the parameter *CR*. The optimal value is approximately 0.05.

The difference in performance of both representations is illustrated in figures 8 and 9 that depict the final connections and appropriate parameters of simulations.

5. Conclusion

The performed experiments show that the representation using list of vertices gives better results than adjacent matrix representation; because the resultant connection does not contain any redundant links. The redundant links disappear when the number of generations increases, but it has negative impact on the execution time. Therefore the representation using adjacent matrix does not seem to be suitable. The mentioned kinds of representations lead to the solutions with a different number of degrees of freedom. The number of degrees of freedom for the adjacent matrix is equal to the number of links in the network $|E|$, which is higher than the number of degrees of freedom for the list of vertices representation, which is equal to $|V| - |U|$. A large number of degrees of freedom causes disproportionate increase of feasible search areas in search space and therefore the algorithm is more likely to stay in a local extreme.

MaxN	10
MinN	3
CR	0.03
L_{max}	1000

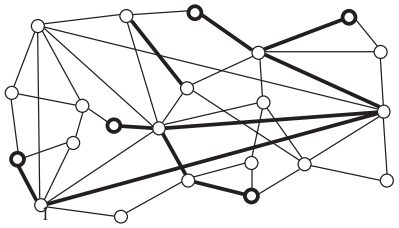


Fig. 8: Resultant connection for adjacent matrix representation (cost = 64)

MaxN	20
MinN	1
CR	0.05
L_{max}	300

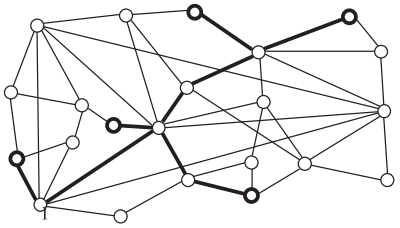


Fig. 9: Resultant connection for list of vertices representation (cost = 48)

References

[1] PRIOR, J. W.: *Eugenic Evolution for Combinatorial Optimization.* Report AI 98 - 268. Artificial Intelligence Laboratory. The University of Texas at Austin. 1998.

[2] KAPSALIS, A., RAYWARD, SMITH, V. J., SMITH, G. D.: *Solving the Graphical Steiner Tree Problem Using Genetic Algorithms.* Journal of the Operational Research Society. Vol. 44, No.4, 1993, pp. 397–406.

[3] OLEJ, V.: *Evolution Stochastic Optimization Algorithms: Genetic Algorithms and Evolution Strategies.* Electrical Engineering Journal, ISSN 1335-3632, Vol. 50, No. 9-10, 1999, pp. 266–272.

[4] OLEJ, V.: *Prediction of Gross Domestic Product Development by Frontal Neural Networks with Learning Process on the Basis Genetic and Eugenic Algorithms.* Neural Network World, International Journal on Non-Standard Computing and Artificial Intelligence, editor: M. Novák, ISSN 1210-0552, Czech Republic, Vol. 12, No. 3, 2002, pp. 279–291.

[5] OLEJ, V.: *Prediction of Gross Domestic Product Development on the Basis of Frontal Neural Networks,* Genetic and Eugenic Algorithms. 2nd Euro – International Symposium on Computational Intelligence, E – ISCI 2002, Intelligent Technologies - Theory and Applications, New Trend in Intelligent Technologies, IOS Press Ohmsha, ISSN 0922-6389, ISBN 1 58603 256 9 IOS Press, ISBN 4 274 90512 8 Ohmsha, Netherlands, 2002, pp. 309–314.
[6] ITU, T.: *Recommendation F.700: Framework Recommendation for audiovisual/ multimedia services.* July 1996.
[7] ESBENSEN, H.: *Computing Near - Optimal Solutions to the Steiner Problem in a Graph Using Genetic Algorithm* [online], 1995. http://www.daimi.au.dk/PB/468/PB-468.ps.gz>.
[8] VRBA, R.: *Možnosti tvorby viacbodových spojení v sieťach B - ISDN pomocou evolučných stochastických optimalizačných algoritmov.* Dizertačná práca, Žilinská univerzita, Elektrotechnická fakulta, Žilina, 2000.
[9] RAYWARD, SMITH, V. J., CLARE, A.: *On Finding Steiner Vertices.* Networks. Vol. 16. 1986. pp. 283–294.
[10] TANAKA, Y., HUANG, P. C.: *Multiple Destination Routing Algorithms.* IEICE Transaction on Communications. Vol. E76-B. No. 5. 1993. s. 544–552.

https://doi.org/10.26552/com.C.2004.4.28-33

kamUNikácie
COMMUNICATIONS

Róbert Hudec *

SD L-FILTER FOR FILTRATION OF IMAGES CORRUPTED BY MIXED NOISE IN FREQUENCY DOMAIN

Slower filter coefficients adaptation of adaptive LMS (Least Mean Square) filters in time domain can be improved by their constrained or unconstrained modifications, but not rapidly. Other alternative is its realisation in frequency domain, where filtered signal is de-correlated and adaptation algorithm converges faster. Based on idea of adaptation in frequency domain there was developed L-filter in frequency domain.

To the group of adaptive non-linear digital filters belong adaptive L-filters (L-F), whose outputs are linear combination of order statistics. They have a simple design and they know how to adapt own coefficients to various noise types. For powerful suppression of mixed noise there was designed a signal-dependent filter structure, which consists of two L-Fs and one SID (Spatial Impulse Detector). The former suppresses noise in homogeneous regions, the latter preserves details of a filtered image.

This paper is devoted to the design of SD LMS L-F structure in frequency domain, which uses the SMD detector (Spatial Median Detector). The aim of this design is to obtain optimal parameters of SD LMS L-F such as threshold, level and adaptation step or optimal weight coefficients.

Keywords: *adaptive LMS L-filters, Signal-Dependent structure, Spatial Impulse Detector, Fast Fourier Transform*

1. Introduction

In real filtering of noisy data by various adaptive filters, the problem of faster coefficient adaptation must by solved. Rate of convergence depends on eigen values spread of autocorrelation matrix (ratio $\lambda_{MAX}/\lambda_{MIN}$) [1, 5]. If this ratio is large, algorithm converges slowly. Faster convergence in time domain can be attained by parametric modifications of LMS algorithm (constrained or unconstrained), RLS algorithm, XLS algorithm etc. Other solution is based on de-correlation data principle in frequency domain.

For powerful suppression of mixed noise in still greyscale images there was developed a SD (Signal-Dependent) filter structure, which uses a version of L-filter (L-F) as single filter [1, 3-5]. The output of L-F is defined as a linear combination of the order statistics in the input data observation [1, 3-7]. Advantageous adaptation properties of filter design in frequency domain lead to design of SD L-F in this domain which is described in this paper.

The outline of this paper is as follows. Theory of LMS L-Filter in time and frequency domains are introduced in sections 2 and 3, and in sections 4 and 5 SD LMS L-filters in both domains are described. Achieved results are described in section 6 and they are discussed in section 7.

2. L-filter, LMS L-filter in time domain

Let a nonstationary observed signal x in each pixel be defined by the original value d and noise n

$$x(m, n) = d(m, n) + n(m, n), \tag{1}$$

where m, n designate pixel coordinates in image.

A filtration window is defined by pixel centered around the central pixel $x(m, n)$ and they are ordered in a lexicographic ordering

$$x_i = (x(m - \xi, n - \xi), ..., x(m, n), \\ ..., x(m + \xi, n + \xi))^T, \tag{2}$$

where $\Xi = 2\xi + 1$ defined a number of pixels in a row or column of square window

$$xr_i = ({}_1x_i, {}_2x_i, ..., {}_{N-1}x_i, {}_Nx_i)^T. \tag{3}$$

xr_i is the vector of ordered pixels in ascending order. The pixel ${}_kx_i$ represents *k-th* highest pixel of *i-th* input observation, where k is from the range of 1 to N ($N = \Xi^2$).

Output of L-filter (L-F) is for each as follows

$$y_i = w^T xr_i\,, \tag{4}$$

where $w = (w_1, ..., w_N)^1$ is the coefficient vector of L-F and in general, it can be obtained by Wiener-Hopf equation [1, 3-5]. In real applications, the computations of inversion of autocorrelation matrix R_i^{-1} and correlation vector p_i are more difficult, and this problem is usually solved by simple iterative algorithms. One from frequent algorithms is the LMS (Least Mean Square) algorithm, which is based on a SD (Steepest Descent) method. It minimises the MSE (Mean Square Error) parameter.

* **Róbert Hudec**
Department of Telecommunications, University of Žilina, Veľký Diel, 010 26 Žilina, Slovak Republic, tel.: +421 41 513 2210,
E-mal: robert.hudec@fel.utc.sk

Finally, the adaptation algorithm for adaptation of L-filter coefficient vector is given by

$$\hat{w}_{i+1} = \hat{w}_i + 2\mu\varepsilon_i x r_i. \tag{5}$$

3. LMS L-filter in frequency domain

Adaptation of highly corellated signals in time domain causes lower convergence of adaptation process. Its decorrelation in frequency domain makes the adaptation faster. There exist a lot of orthogonal transforms, which could be used. One of usually used transforms is fast version of DFT (Discrete Fast Fourier Transform) [2]. The scheme of adaptive L-filter in frequency domain (LF÷F) is shown in Fig. 1. It has some disadvantages as worse implementation, higher computation complexity (operation with complex numbers) [5]. Moreover, the adaptation of filter coefficient vector is the same as for L-F.

The output of LMS LF–F is given by

$$y_i = \left| IDFFT\{w^T x r_i\} \right|. \tag{6}$$

4. SD LMS L-filter in time domain

An adaptive SD LMS L-filter consists of two single adaptive LMS L-filters and one SID (Spatial Impulse Detector) [4, 5]. Its structure is shown in Fig. 2.

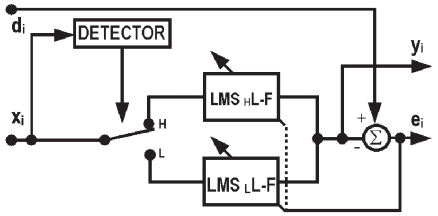


Fig. 2 The scheme of Signal-Dependent LMS L-filter in time domain

4.1 SID (Spatial Impulse Detector)

SID component serves as a switch between two LMS L–F. Moreover, if impulses are detected in the input sequence, these input data are processed by high-frequency segment. On the other hand, if in the input sequence any impulses or high-frequency components are not detected, the input data are processed by low-frequency segment. The decision rule for SD L-filter in combination with spatial impulse detectors is given by

IF

$$\sum_{k=1}^{N}{}_k D_i \geq Level \tag{7}$$

THEN $LMS_H\ L-F$

ELSE $LMS_L\ L-F$

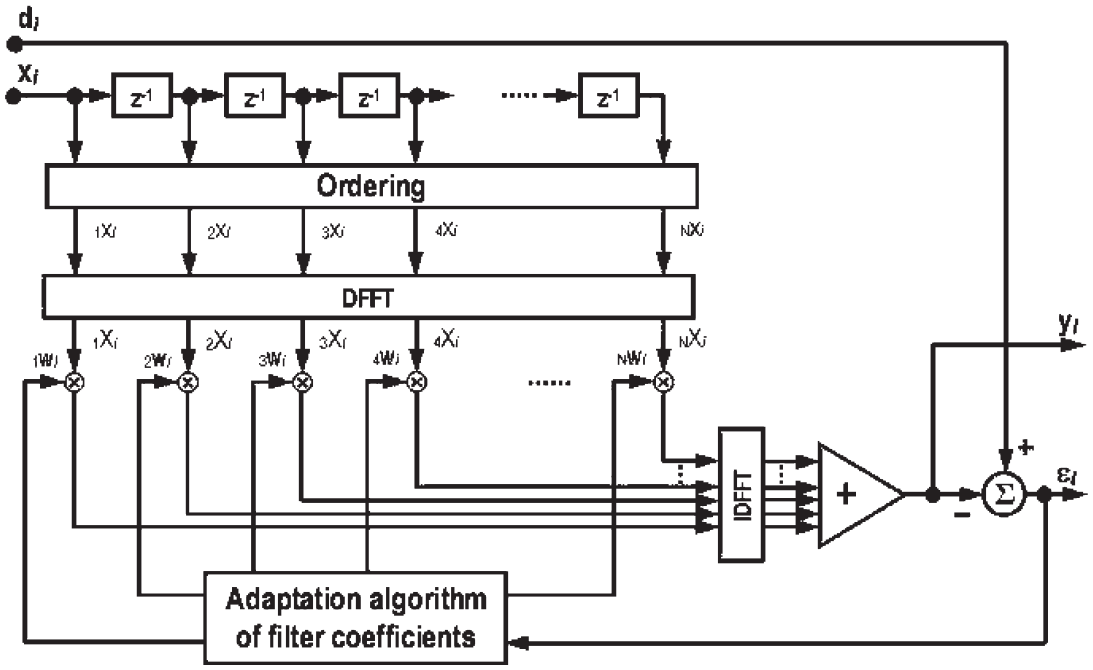


Fig. 1 The adaptive structure of LMS L-filter in frequency domain

where $_k D_i$ is the result of impulse detection for *k-th* image pixel in the *i-th* input vector. The value of level defines the number of detected impulses in the observed samples.

4.1.1 SMD (Spatial Median Detector)

The spatial median order detector (SMD) is one of the order statistic detectors family [4, 5]. It is based on the following rule

IF

$$|med\{x_i\} - {}_k x_i| \geq Tolerance \tag{8}$$

THEN $_k D_i = 1$
ELSE $_k D_i = 0$

If the difference in magnitude between k-th and median input samples is more than the value of tolerance, this input sample is marked as a detected impulse. Its great advantage is a very simple design and smaller computation complexity.

5. SD LMS L-filter in frequency domain

SD LMS LF-F contains two LMS L-F, which work in a frequency domain (LMS LF-F). Moreover, the selected SID detects impulses in time domain. The filter scheme is shown in Fig. 3. One difference between filter realisation in time domain or frequency domain is that the adaptation equations (5) use in frequency domain transformed input data and in time domain non-transformed input data. Finally, the output of LF-F is realised by formula (6).

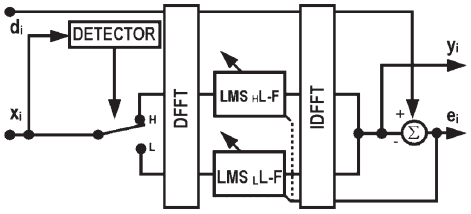


Fig. 3 The scheme of Signal-Dependent LMS L-filter in frequency domain

6. Experimental results

This section contains carried out experiments and achieved results. As the reference still image we use an image of Lena (Fig. 8a). It was corrupted by mixed noise that consists of additive Gaussian white noise with standard deviation $\sigma = 20$ and impulsive noise with probability $p = 10\%$ (Fig. 8b). The achieved filter results were evaluated mainly by the NR (Noise Reduction) parameter [3, 5]. The first experiment shows computational time dependencies from a filter window dimension and used L-filter. From the filter mask dimension point of view for filtering of homogeneous areas, the computation complexity is important. As it is seen in Fig. 4, by increasing the filter window dimension, the adaptive versions have a lower computational complexity and they are suitable for realisations.

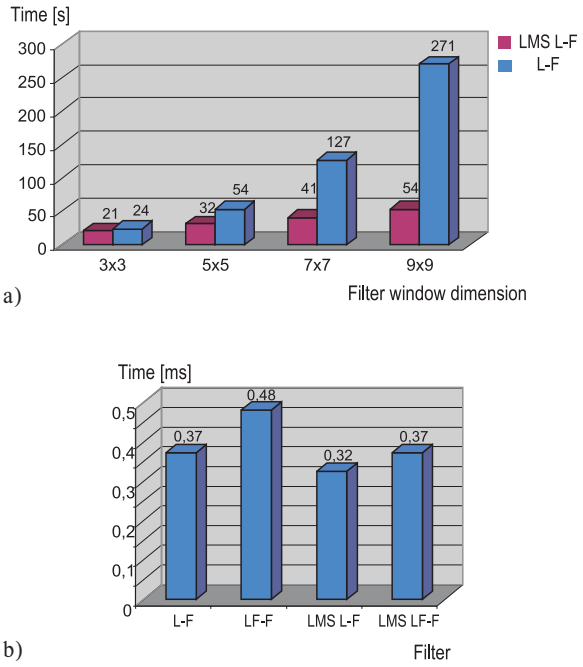


Fig. 4 Computational time,
a) different square filter window of L-F and LMS L-F,
b) one iteration of L-F, LF-F, LMS L-F and LMS LF-F

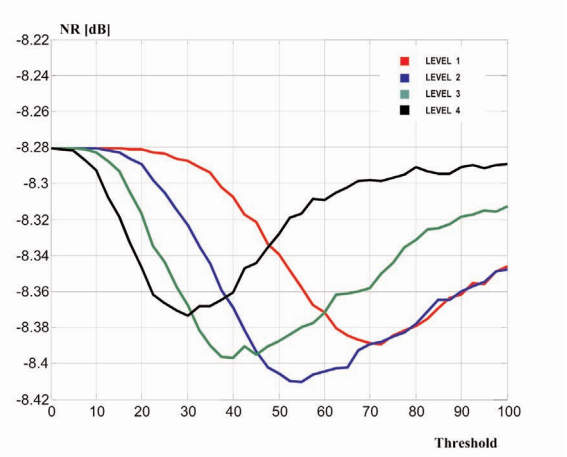


Fig. 5 NR dependencies of $_{SMD}$SD L-F for various levels

In Fig. 5 are shown NR dependencies from the threshold for various levels of $_{SMD}$SD L-F. The optimal set-up of SD L-F with SMD was achieved for the level = 2 and threshold = 55.

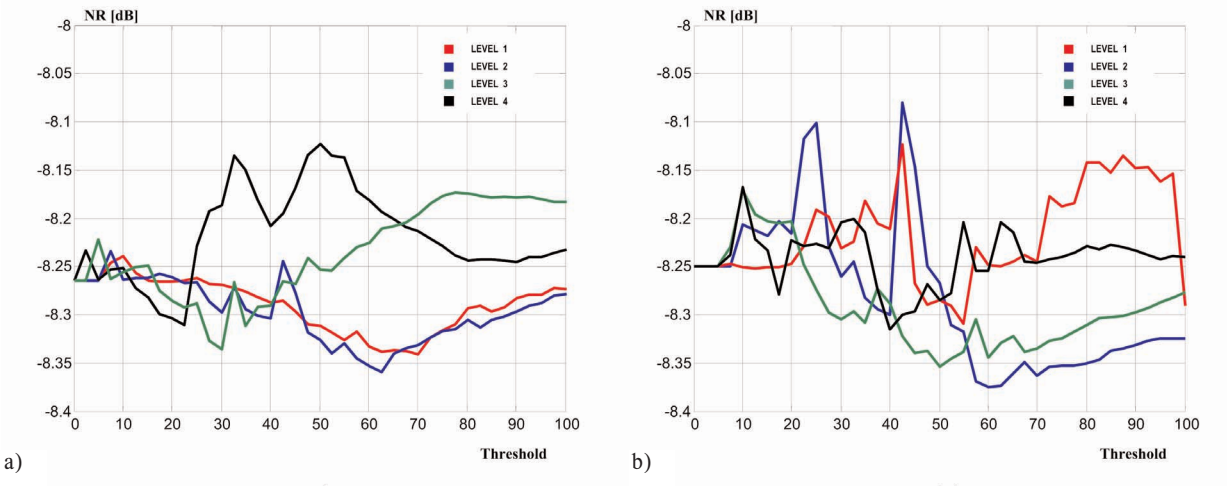


Fig. 6 NR dependencies of $_{SMD}$SD LMS L-F, $\mu = 1x10^{-7}$, a) time domain, b) frequency domain

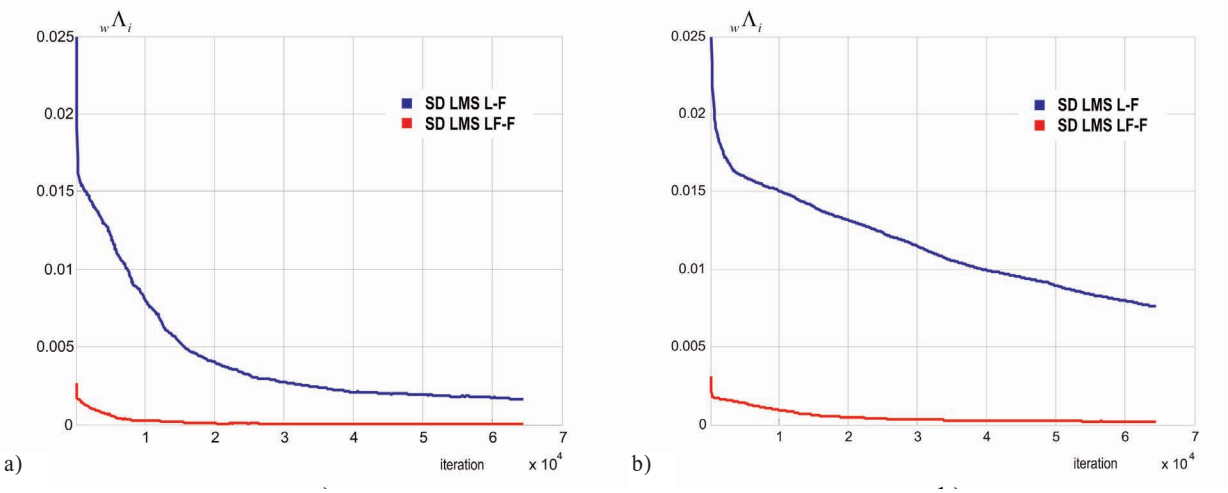


Fig. 7 Graphs of estimation error coefficient, a) $\mu = 1x10^{-7}$, b), $\mu = 1x10^{-8}$

It is known that noise is more visible in homogeneous areas than in details. From this point, better noise filtering in homogeneous areas must be realised by a larger filter window. On the other hand, a better image structure preservation must be realised by a smaller filter window. A signal-dependent filter structure enables to realise filtrations of both aims separately. In practice, they have different filter windows and they depend on image properties. In experiments, equal 3by3 square filter windows were used. In Fig.6 are shown NR dependencies from the threshold of SD LMS L-filter in both domains. It is seen that better filter results were achieved in a frequency domain where the SD filter converged faster. Both compared adaptive SD L-filters had for the level = 2 displaced a threshold of about 60.

The convergence measurement was realised by the coefficient of estimation error [1, 5]

$$_{w}\Lambda_i = \frac{1}{N}\sum_{j=1}^{N}({}_{j}\hat{w}_i - {}_{j}w_o)^2, \qquad (9)$$

where $_{j}w_o j = 1, ..., N$ consists of optimal filter coefficients for the used noise distribution and it is determined for one experiment realisation.

Convergence dependencies of SD filters in both domains for two adaptation steps are shown in Fig.7. It is evident that the version of SD L-F in frequence domain converges in first iteration under $_{w}\Lambda_i < 0.005$ for both adaptation steps. The same version in time domain converges under 0.005 level after 1500 iterations only for $\mu = 1x10^{-7}$. The limit of 0.005 interprets (at average) change of one coefficient from the optimal value about 0.3.

Finally, in Fig. 8c, 8d the filtration results of LMS LF-F and $_{SMD}$SD LMS LF-F are shown. Changes between single LF-F and $_{SMD}$SD LMS LF-F are not very visible, but after the set-up of optimal filter windows, the noise suppression effect will be evident. All experiments were realised by AMD K6-2/500MHz.

a) b)

c) d)

Fig.8 Image Lena, a) original, b) noisy, c) filtered by LMS LF-F d) filtered by $_{SMD}$SD LMS LF-F.

7. Conclusion

In this paper the signal-dependent L-filter in frequency domain was designed. The L-filter, as one of group of order statistic filters, which has suitable falicilities for various noise type suppression was used. Moreover, the signal-dependent filter structure multiplies its power. It could be assigned the fact, that for maximal noise suppression in homogeneous regions it uses a larger filter window and for edge preservation some pixels from input observation only.

Its adaptive realisation in time domain causes slower convergence of adaptation algorithm. The designed SD L-F in frequency domain employed a Spatial Median Detector in time domain, but adaptation of filter coefficient vector was realised in frequency domain. The proposed realization offered faster adaptation for the used adaptation steps.

Better filtration results could be obtained by determination of optimal detector window or by identification of optimal set-up filter windows.

Acknowledgements

This paper was supported by the state programme no. 2003 SP 51/028 09 00/028 09 10 and VEGA-1/0140/03.

References

[1] KOTROPOULOS, C., PITAS, I.: *Adaptive Nonlinear Filters for Digital Signal/Image Processing,* Control and Dynamic Systems, C. T. Leondes, Vol. 67, San Diego CA, Academic press, pp. 263–317, 1994.
[2] BRIGHAM, E. O.: *The Fast Fourier Transform and Its Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1988.
[3] KONTROPOULOS, C., PITAS, I.: *Adaptive LMS L-Filters for Noise Suppression in Image,* IEEE Transaction on Image Processing, vol.5, no. 12, pp. 1596–1609, Dec. 1996.
[4] HUDEC, R., MARCHEVSKÝ, S.: *Extension of impulse detectors to spatial dimension and their utilisation as switch in the LMS L-SD filter*, Radioengineering, vol. 10, no. 1, pp. 11–15, April 2001.
[5] HUDEC, R.: *Filtration of noisy images corrupted by mixed noise by modifications of adaptive L-filters (in Slovak),* PhD. Thesis, Košice, July 2002.
[6] MOUCHA, V., MARCHEVSKÝ, S., LUKÁČ, R., STUPÁK, CS.: *Digital filtration of image signals.* Vydavateľstvo Vojenskej leteckej akadémie M.R. Štefánika v Košiciach. Košice, str. 365 (in Slovak), December 2000.
[7] GONZALEZ, R., WOODS, R.: *Digital Image Processing,* Addison-Wesley Publishing Company, 1992.

https://doi.org/10.26552/com.C.2004.4.34-38

komunikácie
COMMUNICATIONS

Pavel Zahradník – Miroslav Vlček *

FAST ANALYTICAL DESIGN OF MAXIMALLY FLAT NOTCH FIR FILTERS

A novel fast analytical design procedure for the maximally flat notch FIR filters is introduced. The closed form solution provides recursive evaluation of the impulse response coefficients of the filter. The discrete nature of the notch frequency is emphasized. One design example is included to demonstrate the efficiency of the presented approach.

Keywords: *Notch filters, Maximally Flat, Analytical Design*

1. Introduction

The narrow band digital filters are widely used in digital signal processing. While narrow band-pass filters find their application in the detection of signals, the narrow bandstop filters are frequently used in order to remove a single frequency component from the spectrum of the signal. The narrow bandstop filters are usually called notch filters. In our paper we primarily deal with notch filters but we keep in mind the close relation between these two types of narrow band filters. The design of digital notch IIR filters is rather simple. These filters are frequently used despite of their infinite impulse and step responses, which can produce spurious signal components that are unwanted in various applications. The notch IIR filters consist of an abridged all-pass second-order section that allows independent tuning of the notch frequency $\omega_m T$ and the 3-dB attenuation bandwidth [3]. The main drawback usually emphasized in connection with FIR filters is the higher number of coefficients compared to their IIR counterparts. However, this argument is weakened continuously due to the tremendous advance in DSP and FPGA technology. The decisive advantages of FIR filters are their constant group delay and superior time response [8]. Thus the implementation of FIR filters with one hundred coefficients has a practical impact in numerous applications. A few analytical procedures for the design of linear phase notch FIR filters have recently become available [5]. The methods, which lead to feasible filters, are generally derived by iterative approximation techniques or by non-iterative, but still numerical procedures, e.g. the window technique. In our paper we are concerned with completely analytical design of maximally flat (MF) notch FIR filters. We introduce the degree formula, which relates the degree of the generating polynomial, the length of the filter, the notch frequency, the width of the notchband and the attenuation in the passbands. We derive the differential equation for the generating polynomial of the filter. Based on the expansion of the generating polynomial into the Chebyshev polynomials, the recurrent formula for the direct computation of the impulse response coefficients is derived. Consequently, the FFT algorithm usually required in the analytical design of narrow band FIR filters is avoided. The proposed design procedure is recursive one. It does not require any FFT algorithm or any iterative technique.

2. Polynomial Approximation, Zero Phase Transfer Function

Here and in the following we use the independent transformed variable w [6] related to the digital domain by

$$\omega = \frac{1}{2}(z + z^{-1})\big|_{z=e^{j\omega T}} = cos\omega T. \tag{1}$$

We denote $H(z)$ the transfer function of a notch FIR filter with the impulse response h(m) of the length N as

$$H(z) = \sum_{m=0}^{N-1} h(m) z^{-m}. \tag{2}$$

Assuming an odd length $N = 2n+1$ and even symmetry of the impulse response

$$a(0) = h(n),\ a(m) = 2h(n \pm m),\ m = 1 \ldots n \tag{3}$$

we can write the transfer function of the notch FIR filter

$$H(z) = z^{-n}\left[a(0) + \sum_{m=1}^{n} a(m)\, T_m(w)\right] \tag{4}$$

where $T_m(w)$ is Chebyshev polynomial of the first kind. The frequency response of the filter $H(e^{j\omega T})$ can be expressed by the zero phase transfer function $Q(w)$

$$H(e^{j\omega T}) = e^{-jn\omega T}\, Q(\cos\, \omega T) = z^{-n}\, Q(w)\big|_{z=e^{j\omega T}} \tag{5}$$

For $w = 0.5\,(z + z^{-1})\big|_{z=e^{j\omega T}} = \cos\, \omega T$ the zero phase transfer function $Q(w)$ represents a polynomial of the real variable w.

* **Pavel Zahradník[1], Miroslav Vlček[2]**
[1]Department of Telecommunication Engineering, Czech Technical University in Prague, Technická 2, CZ-166 27 Praha 6, Czech Republic,
E-mail: zahradni@fel.cvut.cz
[2]Department of Applied Mathematics, Czech Technical University in Prague, Konviktská 20, CZ-110 00 Praha 1, Czech Republic,
E-mail: vlcek@fd.cvut.cz

It reduces to a real valued frequency response of the zero-phase FIR filter. The zero phase transfer function $Q(w)$ of the narrow bandpass FIR filter is formed by the generating polynomial $A_{p,q}(w)$ while the zero phase transfer function $Q_A(w)$ of the notch FIR filter is

$$Q_A(w) = 1 - A_{p,q}(w). \tag{6}$$

3. Maximally Flat Notch FIR Filter

For the design of MF notch FIR filter we propose the generating polynomial $A_{p,q}(w)$ of the MF narrow bandpass FIR filter introduced in [7]

$$A_{p,q}(w) = C(1 - w)^p (1 + w)^q. \tag{7}$$

The notation $A_{p,q}(w)$ emphasizes that p counts multiplicity of zeros at $w = 1$ and q corresponds to multiplicity of zeros at $w = -1$. Forming the derivative of the polynomial

$$\frac{dA_{p,q}(w)}{\mathrm{dw}} = -Cp(1 - w)^{p-1}(1 + w)^q + \\ + Cq(1 - w)^p(1 + w)^{q-1} \tag{8}$$

and by simple manipulation of (7)

$$(1 - w)(1 + w)\frac{dA_{p,q}(w)}{dw} = \\ = -p(1 + w)A_{p,q}(w) + q(1 - w)A_{p,q}(w) \tag{9}$$

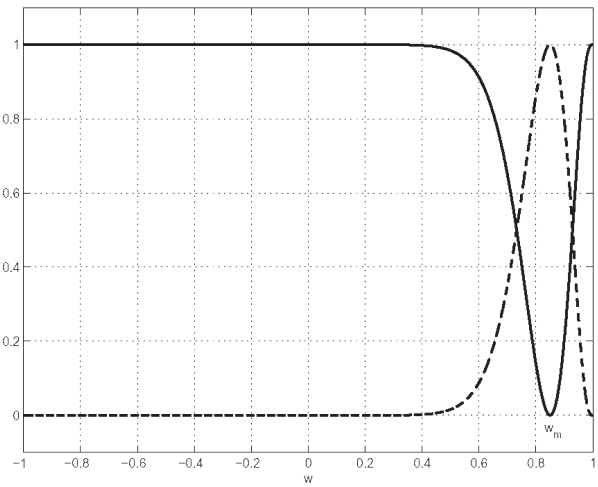


Fig. 1. Generating polynomial $A_{3,37}(w)$ (dashed) and the zero phase transfer function $Q_A(w) = 1 - A_{3,37}(w)$ of the MF notch FIR filter with extremal value for $w_m = (37 - 3)/(37 + 3) = 0.85$

we arrive at the differential equation for the generating polynomial $A_{p,q}(w)$

$$(1 - w^2)\frac{dA_{p,q}(w)}{dw} + [p - q + (p + q)w]A_{p,q}(w) = 0. \tag{10}$$

The differential equation (10) for the polynomial $A_{p,q}(w)$ forms a completely new concept in digital filter design as it provides the recursive evaluation of the impulse response coefficients of the filter described in Section 5. The normalization of the generating polynomial $A_{p,q}(w)$ constraints $A_{p,q}(w_m) = 1$ where w_m is the position of the maximum of the generating polynomial $A_{p,q}(w)$ as illustrated in Fig. 1. The normalization of the generating polynomial $A_{p,q}(w)$ results in

$$A_{p,q}(w) = \left[\frac{p + q}{2p}(1 - w)\right]^p \left[\frac{p + q}{2q}(1 + w)\right]^q \tag{11}$$

The polynomial

$$Q_A(w) = 1 - A_{p,q}(w) \\ = 1 - \left[\frac{p + q}{2p}(1 - w)\right]^p \left[\frac{p + q}{2q}(1 + w)\right]^q \tag{12}$$

represents the real-valued zero phase transfer function of the MF notch FIR filter of the real variable $w = \cos \omega T$. For illustration, the zero phase transfer function of the MF notch FIR filter $Q_A(w) = 1 - A_{3,37}(w)$ is shown in Fig. 1. The corresponding amplitude frequency response 20 log $|H(e^{j\omega T})|$ [dB]

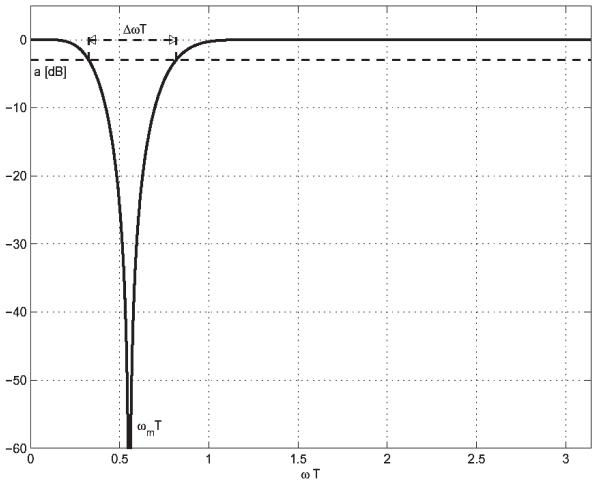


Fig. 2. Amplitude frequency response 20 log $|H(e^{j\omega T})|$ [dB] based on the generating polynomial $Q_A(w) = 1 - A_{3,37}(w)$ from Fig. 1. The parameters are $\omega_m T = 0.1766\ \pi$ and $\Delta\omega T = 0.1555\ \pi$ for the standard attenuation $a = -3.0103$ dB

is shown in Fig. 2. The transfer function of the MF notch FIR filter is

$$H(z) = \sum_{m=0}^{N-1} h(m)\, z^{-m} = z^{-n}\,(1 - A(p,q)(w)). \tag{13}$$

4. Notch Frequency and the Degree of the Maximally Flat Notch FIR Filter

The notch frequency $\omega_m T$ is derived from the minimum value of the zero phase transfer function $Q_A(w)$ (12) as

$$w_m = \cos \omega_m T = \frac{q-p}{q+p}. \tag{14}$$

The notch frequency $\omega_m T$ of the MF notch FIR filter is given from (14) by the integer values p and q exclusively. It is obvious that for the specified filter length $N = 2(p+q)+1$, exactly $p+q-1$ discrete notch frequencies $\omega_m T$ are available. From the symmetrical case $n/2 = p = q$ the degree equation

$$n \geq \frac{\log(1 - 10^{0.05a|_{dB}|})}{\log \cos \frac{\Delta\omega T}{2}} \tag{15}$$

can be derived. The relations for the integer values p, q read as follows

$$p = \left[n \sin^2\left(\frac{\omega_m T}{2}\right)\right], \; q = \left[n \cos^2\left(\frac{\omega_m T}{2}\right)\right]. \tag{16}$$

The brackets [ ] in (16) denote the rounding operation.

5. Impulse Response Coefficients of the Maximally Flat FIR Filter

We can express the generating polynomial $A_{p,q}(w)$ of the degree $n = p+q$ as the sum of Chebyshev polynomials of the first kind $T_m(w)$

$$A_{p,q}(w) = \sum_{m=0}^{n} a(m) T_m(w). \tag{17}$$

The coefficients $a(m)$ define the impulse response $h(m)$ (3) of the length $N = 2(p+q)+1$. Assuming the generating polynomial $A_{p,q}(w)$ of the MF narrow bandpass FIR filter in the sum (17) we can write

$$(1-w^2)\frac{dA_{p,q}(w)}{dw} = \sum_{m=1}^{n} a(m)(1-w^2)\frac{dT_m(w)}{dw} =$$

$$= \sum_{m=1}^{n} a(m)\frac{m}{2}[T_{m-1}(w) - T_{m+1}(w)]. \tag{18}$$

By introducing (17) and (18) into the differential equation (10) and using the recursive formula for Chebyshev polynomials

$$T_{m+1}(w) = 2\,wT_m(w) - T_{m-1}(w) \tag{19}$$

we get the identity

$$\sum_{m=1}^{n} a(m)\frac{m}{2}[T_{m-1}(w) - T_{m+1}(w)] + (p-q)a(0) +$$

$$+ \sum_{m=1}^{n} a(m)(p-q)T_m(w) + (p+q)a(0)w + \tag{20}$$

$$+ \sum_{m=1}^{n} a(m)(p+q)\frac{1}{2}[T_{m-1}(w) + T_{m+1}(w)] = 0.$$

By iterating eq. (20) we have deduced a simple recursive algorithm for the evaluation of the coefficients $a(m)$ of the generating polynomial $A_{p,q}(w)$ of the MF narrow bandpass FIR filter. The recursive algorithm is presented in Table 1. The coefficients $h(m)$ of the impulse response of the MF notch FIR filter are obtained from the coefficients $a(m)$ of the MF narrow bandpass FIR filter as follows

$$h(n) = 1 - a(0),\; h(n \pm m) = -\frac{a(m)}{2},\; m = 1 \ldots n. \tag{21}$$

6. Design of the Maximally Flat Notch FIR Filter

The goal of the MF notch FIR filter design is to find the two integer values p and q in order to satisfy the filter specification as precisely as possible. The design procedure is as follows:

1. Specify the notch frequency $\omega_m T$, maximal width of the notchband $\Delta\omega T$ and the attenuation in the passbands a [dB] as demonstrated in Fig. 2.
2. Calculate the minimum degree n (15) required to satisfy the filter specification.
3. Calculate the integer values p and q (16).
4. Check the notch frequency (14) for the obtained integer values p, q .
5. Evaluate the coefficients $a(m)$ of the generating polynomial $A_{p,q}(w)$ recursively (Table l).
6. Evaluate the coefficients of the impulse response $h(m)$ of the MF notch FIR filter (21).

Recursive algorithm for evaluation of the coefficients $a(m)$ Tab. 1

given	p, q
initialization	$n = p+q$ $a(n+1) = 0$
body (*for* $k = n+1$ *to* 3)	$a(k-2) = -\dfrac{(n+k)a(k) + 2(2p-n)a(k-1)}{n+2-k}$
(*end loop on k*)	$a(0) = -\dfrac{(n+2)a(2) + 2(2p-n)a(1)}{2n}$

It is worth of noting that a substantial part of coefficients of the impulse response $h(m)$ of the MF notch FIR filter has negligible values. From this

Impulse Response Coefficients — Table 2

m		h(m)	m		h(m)
14	74	-0.000002	30	58	0.012289
15	73	-0.000003	31	57	0.002278
16	72	0.000000	32	56	-0.019427
17	71	0.000018	33	55	-0.027483
18	70	0.000037	34	54	-0.003357
19	69	0.000010	35	53	0.042804
20	68	-0.000111	36	52	0.048063
21	67	-0.000245	37	51	-0.009353
22	66	-0.000101	38	50	-0.075616
23	65	0.000537	39	49	-0.065324
24	64	0.001173	40	48	0.029196
25	63	0.000480	41	47	0.106554
26	62	-0.002149	42	46	0.068113
27	61	-0.004302	43	45	-0.053105
28	60	-0.001388	44		0.880514
29	59	0.007135			

fact follows the possible large abbreviation of the impulse response of the MF notch FIR filter by the rectangular windowing without significant deterioration of the frequency properties of the filter as emphasized in [7].

7. Example of the Design of the Maximally Flat Notch FIR Filter

Design the MF notch FIR filter specified by $\omega_m T = 0.35\ \pi$ and $\Delta\omega T = 0.15\ \pi$ for $a = -3.0103$ dB.

Using our design procedure we get $n = [43.8256] \rightarrow 44$ (15), $p = [11.9644] \rightarrow 12$ and $q = [31.8610] \rightarrow 32$ (16). The filter length is $N = 89$ coefficients. The actual filter parameters are $\omega_m T = 0.3498\ \pi$ and $\Delta\omega T = 0.1496\ \pi$. The attenuation at the frequency

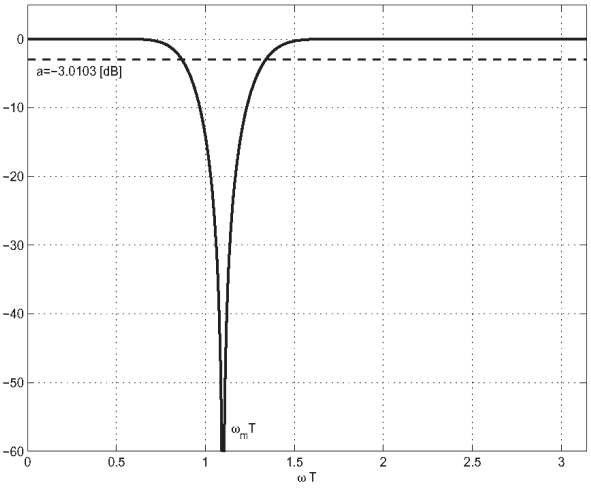


Fig. 3. Amplitude frequency response 20 log $|H(e^{j\omega T})|$ [dB] based on the zero phase transfer function $Q(w) = 1 - A_{12,32}(w)$

0.3 π amounts -168 dB. The coefficients $a(m)$ were evaluated recursively (Table 1). The coefficients of the impulse response $h(m)$ of the MF notch FIR filter were evaluated by (21). Because $|h(m)| < 10^{-6}$ for $0 < m < 14$ and $m > 74$, only the 71 central coefficients of the impulse response $h(m)$ are summarized in Table 2. The amplitude frequency response 20 log $|H(e^{j\omega T})|$ [dB] of the MF notch FIR filter is shown in Fig. 3.

8. Conclusions

Novel fast analytical design procedure for the design of maximally flat notch FIR filters was introduced. The closed form solution provides recursive evaluation of the impulse response of the filter. One example demonstrated the efficiency of the design procedure.

9. Acknowledgements

This work was supported by the grant No. CEZ: J04/ 98: 212300014, Ministry of Education, Czech Republic.

References

[1] DUTTA ROY, S .C., KUMAR, B., JAIN, S. B.: *FIR Notch Filter Design - A Review. Facta Universitatis (Niš),* Series Electronics and Energetics. Vol. 14, No. 3, 2001, pp. 295–327.

[2] PEI, S. C., TSENG, C. C.: *IIR Multiple Notch Filter Design Based on Allpass Filter.* IEEE Transactions on Circuits and Systems. Vol. 44, No. 2, 1997, pp. 133–136.

[3] REGALIA, P.A., MITRA, S. K., VAIDYANATHAN, P. P.: *The Digital All-Pass Filter: A Versatile Signal Processing Building Block.* Proceedings of IEEE, Vol. 76, No. 1, 1988, pp. 19–37.

[4] SELESNICK, I. W., BURRUS, C. S.: *Exchange Algorithms for the Design of Linear Phase FIR Filters and Differentiators Having Flat Monotonic Passbands and Equiripple Stopbands.* IEEE Trans. Circuits, Syst.-II, Vol. 43, 1996, pp. 671–675.

[5] YU, TIAN-HU, MITRA, S. K., BABIC, H.: *Design of Linear Phase FIR Notch Filters.* Sadhana, Vol. 15, Iss. 3, 1990, India, pp. 133–55.

[6] VLČEK, M., UNBEHAUEN, R.: *Analytical Solution for Design of IIR Equiripple Filters.* IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP - 37, Oct. 1989, pp. 1518–1531.
[7] VLČEK, M., JIREŠ, L.: *Fast Design Algorithms for FIR Notch Filters.* Proc. of IEEE International Symposium on Circuits and Systems ISCAS'94, London, 1994, Vol. 2, pp. 297–300.
[8] VLČEK, M., ZAHRADNÍK, P.: *Digital Multiple Notch Filters Performance.* Proceedings of the 15th European Conference on Circuit Theory and Design ECCTD'01, Helsinki, 2001, pp. 49–52.
[9] VLČEK, M., ZAHRADNÍK, P., UNBEHAUEN, R.: *Analytic Design of FIR Filters.* IEEE Transactions on Signal Processing, Vol. 48, 2000, pp. 2705–2709.

https://doi.org/10.26552/com.C.2004.4.39-47

Daša Tichá *

ALLPASS BASED ON A GENERALIZED DIVIDER PRINCIPLE

This paper deals with a new allpass structure based on a generalized divider principle. An efficient modification of the basic general divider structure using Antoniou's GICs is presented and the corresponding 1st- and 2nd-order allpass circuits are derived. A detail discussion is devoted to the optimized design of the 2nd-order allpass, taking minimization of the active elements nonidealities and dynamics optimization into account. An original design procedure is described and documented on numerical examples.

1. Introduction

All-pass filters represent a meaningful part of modern communication systems, making possible equalization of signal chain group-delay. This is important especially in the case of the data-signal or video-signal processing. As known, the basic properties of the all-pass filters are determined by all-pass transfer function (1)

$$H(s) = h\,\frac{v(-s)}{v(s)} = h\,\frac{a_0 - a_1 s + s_2 s^2 - \ldots + (-1^n) a_n s^n}{a_0 + a_1 s + a_2 s^2 + \ldots + a_n s^n} \quad (1)$$

With respect to the stability condition, the polynomial $v(s)$ has to be a Hurwitz polynomial. It is easy to derive that the transfer function (1) corresponds to the non-minimal phase circuit whose magnitude frequency response is frequency independent, determined by a multiplicative factor h. Phase response can be expressed in the form (2)

$$\varphi(\omega) = -2 \arctan \frac{\omega N(-\omega^2)}{M(-\omega^2)}, \quad (2)$$

where $\omega N(-\omega^2)$ ... represents an odd part of the polynomial $v(j\omega)$
and
$M(-\omega^2)$ ... represents an even part of $v(j\omega)$.

The group delay $\tau(\omega)$ is defined by the known formula

$$\tau(\omega) = -\frac{d\varphi}{d\omega}. \quad (3)$$

There are many ways how to realize the transfer function (1). Remember that passive realization of the group-delay equalizer is usually based on the 1st- or 2nd-order lattice and their unsymmetrical equivalent circuits – see [1], [2]. In the field of active RC circuits are well-known implementations using a single-amplifier circuit or universal biquad configurations as mentioned in [3], [4], [5] and others. The first group of circuits suffers from inappropriate sensitivity to the element value variations and amplifier non-idealities. At the same time, a parameter setting is rather complicated and does not allow simple tuning. The second group shows better sensitivity properties, but negative odd numerator coefficients are obtained as a difference of positive and negative terms. With respect to this, such a circuit is sensitive to the perfect setting and amplifier non-idealities. The mentioned circuits usually show a magnitude frequency response distortion which is caused by coefficient errors, or, in other words, by errors of transfer function pole-zero location.

To find a more suitable solution, a considerable effort has been devoted to the search for a new circuit implementation of the allpass transfer function. As shown in the following section, one of the possible ways is based on a suitable application of a generalized divider principle.

2 A generalized divider

In Ref.[6] we have presented a universal filter structure based on a generalized divider principle. The basic arrangement of the mentioned structure is shown in Fig. 1. The "black boxes" marked as GIC correspond to the generalized immitance converters characterized by the 1st-order conversion function $k_i(s) = k_i.s$, where s denotes complex frequency. Note that such a structure allows realization of any transfer function by appropriate setting of divider branches, as is evident from the corresponding transfer function (4)

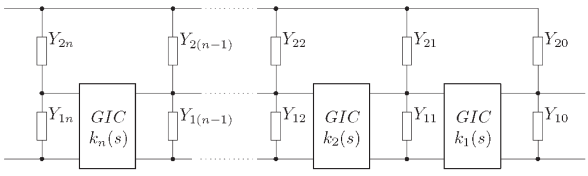


Fig. 1: Generalized divider arrangement

$$H(s) = \frac{Y_{20} + k_1(s)[Y_{21} + k_2(s)[Y_{22} + \ldots + k_n(s)[Y_{2n}]\ldots]]}{(Y_{10} + Y_{20}) + k_1(s)[(Y_{11} + Y_{21}) + k_2(s)[(Y_{12} + Y_{22}) + \ldots + k_n(s)[Y_{1n} + Y_{2n})]\ldots]]}. \quad (4)$$

* **Daša Tichá**
Department of Telecommunications, EF, University of Žilina, E-mail: ticha@fel.utc.sk

Comparing $H(s)$ to the general transfer function in the polynomial form (5)

$$H(s) = \frac{a_0 + a_1 s + a_2 s^2 + ... + a_n s^n}{d_0 + d_1 s + d_2 s^2 + ... + d_n s^n}, \quad (5)$$

it is easy to derive formulae for numerator and denominator coefficients as the functions of branch admittances and GICs conversion functions

$$a_k = Y_{2k} \prod_{i=1}^{k} k_i; \; d_k = (Y_{1k} + Y_{2k}) \prod_{i=1}^{k} k_i, \quad (6)$$

where k_i – denotes the conversion-function-multiplicative constant of the i^{th} GIC.

To apply the presented generalized divider structure in the allpass design it is necessary to accept the following restrictions:

- The equal values of numerator and denominator coefficients at the same power of s (but, in general, all the numerator coefficients can be multiplied by the transfer multiplicative constant h: $\Rightarrow a_i = h \,.\, d_i$, $i = 0, 1, 2, ..., n$).
- The odd numerator coefficients have to be negative $\Rightarrow a_j = -h \,.\, d_j, j = 1, 3, 5, ...$

The relevant problem can arise in the second condition fulfilment. The original generalized divider structure allows realization of the transfer function containing the positive numerator coefficients only, as is evident from Eq. (4). To obtain the required negative odd coefficients, it is necessary to change the sign of the corresponding branch admittances, or, to use inverted source voltage at the inputs of the upper odd branches. An efficient solution offers the use of the known Antoniou's GIC circuitry – see [4, 5]. Let us consider Antoniou's GIC in Fig. 2, whose conversion function is expressed by Eq. (7), and now modify the original circuit by adding the auxiliary port $x-x'$. An arbitrary loading impedance Z_x, connected to the auxiliary port is transformed to the "main" input port $1 - 1'$ as the negative impedance $Z_{1-1'}$, which is presented by the formula (8)

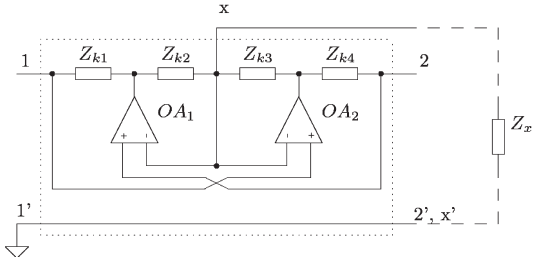


Figure 2: Antoniou's GIC circuit

$$k = \frac{Z_{k1} Z_{k3}}{Z_{k2} Z_{k4}}, \quad (7)$$

$$Z_{1-1'} = \frac{Z_{k1}}{Z_{k2}} Z_x. \quad (8)$$

Using this principle it is easy to derive the modified generalized divider structure shown in Fig. 3, convenient to the arbitrary allpass transfer realization. The odd "upper" branch admittances $Y_{2j,\ j\ =\ 1,\ 3,\ 5,\ ...}$ are connected to the auxiliary port of the corresponding odd GICs, and, with respect to the GIC transforming effect, they behave as the negative ones. The equivalent admittance of the modified branch is expressed by the equation (9)

$$Y_{2j(m)} = -\frac{Z_{2j}}{Z_{1j}} Y_{2j} = \frac{R_{2j}}{R_{1j}} Y_{2j}, \quad (9)$$

where $Y_{2j(m)}$ – denotes transformed (modified) admittance of j^{th} branch,
Z_{ij} – denotes i^{th} impedance of j^{th} GIC circuit (see Fig. 2),
Y_{2j} – denotes the original j^{th} branch admittance.

As evident the converter impedance Z_{k1}, Z_{k2} should be of the same type to save the character of the transformed admittance Z_x. It is easy to derive; the correct circuit behaviour requires arrangement of GIC circuit elements as follows:

$$Z_{k1} = R_1; Z_{k2} = R_2; Z_{k3} = R_3; Z_{k4} = \frac{1}{sC_4}. \quad (10)$$

The resulting transfer function of the improved circuit structure can be written in the form (11)

$$H(s) = \frac{Y_{20} - k_1\alpha_1 Y_{21} s + k_1 k_2 Y_{22} s^2 - k_1 k_2 k_3 \alpha_1 \alpha_3 Y_{23} s^3 + ...}{(Y_{10} + Y_{20}) + k_1(Y_{11} - \alpha_1 Y_{21}) s + k_1 k_2 (Y_{12} + Y_{22}) s^2 + k_1 k_2 k_3 (Y_{13} - \alpha_1 \alpha_3 Y_{23}) s^3 ...}, \quad (11)$$

where α_j – expresses the ratio $\alpha_j = R_{2j}/R_{1j}$, $j = 1, 3, 5, ...$ of i^{th} GIC impedances.

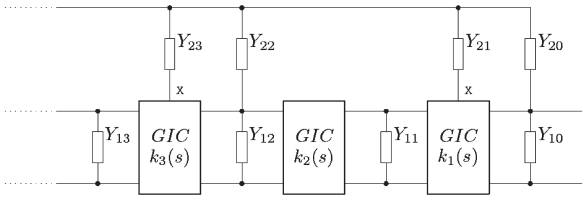


Fig. 3: Modified divider structure

As evident, the odd transfer function coefficients are now expressed in the form (12), while the even coefficients remain unchanged and correspond to the expression (6).

$$a_j = Y_{2j} \prod_{l=1,3,5..}^{j} \alpha_l \prod_{k=1}^{j} k_k; \; d_j = (Y_{1j} - Y_{2j}) \prod_{l=1,3,5..}^{j} \alpha_l \prod_{k=1}^{j} k_k,$$
$$j = 1, 3, 5, ... \quad (12)$$

The presented results indicate effective implementation of an arbitrary order allpass transfer function. With respect to the number of circuit elements (divider branches) the structure is canonical.

Note that the special case for the transfer multiplicative constant $h = 1$ leads to the minimum number elements realization. The absolute value coefficient equivalence

$$|a_i| = |d_i|$$

compared to the Eqs(6, 12) gives divider branch design conditions

$$Y_{1i} = 0,\ i = 0, 2, 4 \ldots \quad Y_{1j} = 2Y_{2j},\ j = 1, 3, 5 \ldots \tag{13}$$

It is important to say that the divider principle causes a restriction in transfer multiplicative constant value. The appropriate choice is limited to the unequality $h \leq 1$ and cannot be exceeded by no means.

In the following the particular cases of the 1st- and the 2nd-order allpass circuits will be discussed in detail.

3. The 1st and 2nd-order allpass

The simpliest version of allpass circuit is presented by 1st-order divider structure shown in Fig. 4.

The circuit transfer function was found in the form (14)

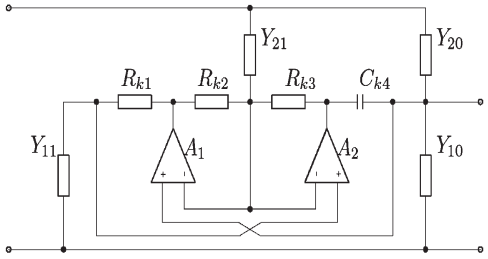


Fig. 4: 1st-order allpass circuit

$$H(s) = \frac{-Y_{21}k_1\alpha_1 s + Y_{20}}{k_1(Y_{11} - Y_{21}\alpha_1)s + Y_{20} + Y_{10}}, \tag{14}$$

where $k_1 = \dfrac{R_{k1}\, R_{k3}\, C_{k4}}{R_{k2}}$.

A comparison of the expression (14) to the general form of 1st-order allpass transfer function (15) leads to the simple design equations (16)

$$H(s) = h\,\frac{s - \sigma}{s + \sigma}, \tag{15}$$

$$Y_{10} = \sigma k_1\alpha_1\,\frac{1 - h}{h}\,Y_{21};\ Y_{11} = \alpha_1\,\frac{h + 1}{h}\,Y_{21};$$

$$Y_{20} = \sigma k_1\alpha_1 Y_{21}. \tag{16}$$

Here $\alpha_1 = \dfrac{R_2}{R_1}\cdot k_1$, and Y_{21} represent free parameters and can be chosen arbitrarily, e.g. with respect to the additional optimized design conditions.

Now, let us devote our attention to the 2nd-order allpass. The circuit diagram is shown in Fig. 5, GICs are implemented by the mentioned Antoniou's circuitry – see Fig. 2. Circuit symbolic transfer function is expressed in the "standard" form (17)

$$H(s) = \frac{a_2s^2 - a_1s + a_0}{s^2 + d_1s + d_0} = h\,\frac{s^2 - s\,\dfrac{\omega_0}{Q} + \omega_0^2}{s^2 + s\,\dfrac{\omega_0}{Q} + \omega_0^2}. \tag{17}$$

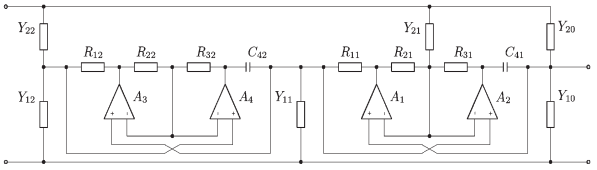


Fig. 5: 2nd-order allpass circuit using modified Antoniou's GIC

The transfer function coefficients are expressed by the formulae

$$a_1 = h\,\frac{\omega_0}{Q} = \frac{\alpha_1}{k_2}\,\frac{Y_{21}}{Y_{12} + Y_{22}},\ d_1 = \frac{\omega_0}{Q} = \frac{1}{k_2}\,\frac{Y_{11} - \alpha_1 Y_{21}}{Y_{12} + Y_{22}} \tag{18}$$

$$a_0 = h\omega_0^2 = \frac{1}{k_1k_2}\,\frac{Y_{20}}{Y_{12} + Y_{22}},\ d_0 = \omega_0^2 = \frac{1}{k_1k_2}\,\frac{Y_{10} + Y_{20}}{Y_{12} + Y_{22}}$$

and

$$h = \frac{Y_{22}}{Y_{12} + Y_{22}},\ k_i = \frac{R_{1i}R_{3i}C_{4i}}{R_{2i}},\ \alpha_i = \frac{R_{2i}}{R_{1i}},\ i = 1, 2 \tag{19}$$

Similarly to the aforementioned 1st-order case, it is possible to derive design equations of divider branches $Y_{ik\ i,k=1,2}$ from Eqs. (18). The result is

$$Y_{10} = \frac{1-h}{h}\,\omega_0^2k_1k_2Y_{22},\ Y_{11} = \frac{1 + h}{h}\,\frac{\omega_0}{Q}\,k_2Y_{22},$$

$$Y_{12} = \frac{1-h}{h}\,Y_{22},\ Y_{20} = \omega_0^2k_1k_2Y_{22},\ Y_{21} = \frac{\omega_0}{Q}\,\frac{k_2}{\alpha_1}\,Y_{22},$$

$$Y_{22} \ldots \textit{arbitrary}. \tag{20}$$

Similarly to the 1st-order case, the values of Y_{22}, k_i and α_i parameters are free and can be conformed to the optimum design conditions. In accordance with the previous considerations and derived "basic" design formulae (16, 20), the higher-order allpass can be easily created only by adding next branches to the designed circuit.

4. Optimized design of the 2nd-order circuit

The presented design equations (20) are fully valid in the case of an idealized circuit, i.e. the circuit containing idealized active elements. As known, especially the influence of amplifier finite frequency dependent gain significantly change the circuit behav-

iour, particularly increase the transfer funcion order and vary "main" poles and zeroes location. From this point-of-view, it is advisable to use the free design parameters to minimize the influence of amplifier non-idealities. A detail analysis of this topic was made in [7], in this place the main results will be summarized and optimized design algorithms presented.

Starting from the allpass general biquadratic transfer function $H(s)$

$$H(s) = \frac{N(s)}{D(s)} = h\,\frac{s^2 - s\,\dfrac{\omega_{0z}}{Q_z} + \omega_{0z}^2}{s^2 + s\,\dfrac{\omega_{0p}}{Q_p} + \omega_{0p}^2} =$$

$$= \frac{a_2 s^2 - a_1 s + a_0}{s^2 + d_1 s + d_0}\,. \qquad (21)$$

The design optimization conditions include
- minimization of ω_p and ω_z errors,
- minimization of Q_p and Q_z errors,
- dynamic optimization, i.e. equalization of maximum output voltages of all the amplifiers.

Note that the optimization conditions strongly depend on the amplifiers type used in Antoniou's converter circuitry, as documented in ref.[8]. With respect to the presented results, the transimpedance amplifier (CFOA) has been chosen as the most suitable for voltage-mode design.

The derivation of general optimization conditions requires to express parameters of the main poles and zeroes in symbolic form. But the transfer function of the real circuit is of the 6th-order, considering simple single-pole models of the amplifiers used. A symbolic evaluation of the main poles and zeroes then presents a dificult mathematical task, which can be solved only approximately, despite the modern mathematical software at disposal. The developed general algorithm is based on order reduction of the transfer function numerator and denominator polynomials neglecting the higher-order error terms. It is formed as follows:

1. *Symbolic transfer function of the real circuit is computed under consideration of finite, frequency independent amplifier gain.*
2. *Numerator and denominator of the obtained symbolic transfer function is divided by the highest power of amplifier gain:*

$$N(s) \Rightarrow \frac{N(s)}{A^m}, \qquad D(s) \Rightarrow \frac{N(s)}{A^m},$$

where A - means amplifier gain, m - denotes the highest power of A.

3. *All the terms of recalculated numerator and denominator polynomials containing power of A higher than 1 are neglected, i.e.*

$$\frac{a_i}{A^k},\ k > 1 \to 0 \text{ and } \frac{d_j}{A^k},\ k > 1 \to 0.$$

Then

$$N'(s) = s^2 + a_{11}s + a_{01} + \frac{1}{A}(a_{12}s + a_{02}),$$

$$D'(s) = s^2 + d_{11}s + s_{01} + \frac{1}{A}(d_{12}s + d_{02}).$$

4. *Gain amplifier symbol A in N′(s) and D′(s) is replaced by a frequency dependent relationship*

$$A \Rightarrow A_0\,\frac{\omega_x}{s + \omega_x} \doteq \frac{B}{s},$$

where $B = A_0\omega_x$, and ω_x means dominant pole of amplifier gain frequency response.

5. *Expressions of $N_0(s)$ and $D_0(s)$ are formally rearranged into a polynomial form and used for simplified transfer function $H_0(s)$ compilation*

$$H'(s) = h'\,\frac{N'(s)}{D'(s)} = h'\,\frac{s^3 + a'_2 s^2 + a'_1 s + a'_0}{s^3 + d'_2 s^2 + d'_1 s + d'_0} \qquad (22)$$

6. *Numerator and denominator polynomials N′(s) and D′(s) are decomposed into*

$$N'(s) = s^3 + a'_2 s^2 + a'_1 s + a'_0 =$$

$$= (s + \sigma_z)(s^2 + a_{e1}s + a_{e0}) =$$

$$= (s + \sigma_z)(s^2 + s\,\frac{\omega'_{0z}}{Q'_z} + \omega'^2_{0z})\,. \qquad (23)$$

$$D'(s) = s^3 + d'_2 s^2 + d'_1 s + d'_0 =$$

$$= (s + \sigma_p)(s^2 + d_{e1}s + d_{e0}) =$$

$$= (s + \sigma_p)(s^2 + s\,\frac{\omega'_{0p}}{Q'_p} + \omega'^2_{0p})\,.$$

Here ω'_{0p}, Q'_p represent parameters of "real" main poles, ω'_{0z}, Q'_z "real" main zeroes. σ_p and σ_z represent auxiliary "equivalent" real pole and zero without any importance for the following optimization procedure.

The third-order polynomial decomposition is possible using mathematical programs, or, in the simplified way using approximate formulae

$$\sigma_p = d'_2 - \frac{\omega_{0pid}}{Q_{pid}};\ \omega'^2_{0p} = d_{e0} = \frac{d'_0}{\sigma_p}\,, \qquad (24)$$

$$d_{e1A} = \frac{d'_1 - \omega^2_{0pid}}{\sigma_p}\,,\ \text{or}\ d_{1eB} = \frac{d'_2 d'_1 - d'_1\,d_{1id} - d'_0}{\sigma_p^2}\,, \qquad (25)$$

where $d_{1id} = \dfrac{\omega_{0pid}}{Q_{pid}}$, ω_{0pid}, Q_{pid} are parameters of the idealized transfer function (17), resp. (21), expressed in the symbolic form. The parameter Q'_p can be evaluated from Eq. (26)

$$Q'_p = \frac{\omega'_p}{d_{e1}}. \qquad (26)$$

Note that the presented simplified equations are valid under assumption

$$\sigma_p >> \omega_{0p},\ \omega_{0p} \doteq \omega_{0pid}$$

for case *A*, or $d_{e1} \doteq d_{1id}$ for case *B*. Similar equations and assumptions can be used for the evaluation of the equivalent numerator parameters σ_z, ω'_{0z} and Q'_z. As shown in Ref. [7], the sufficient accuracy is achieved when $B/\omega_0 > 10$[1], with respect to the Q - factor. The solved numerical examples confirmed small evaluation errors of $d_{e0} = \omega'_{02}$ parameters, even in the case of lower ratios B/ω_0. In the case of parameter d_{e1} or a_{e1} evaluation accuracy, case *A* results show higher values, case *B* results give lower values of computed Q - factor in comparison to the exact values. The acceptable approximation is the average value, i.e.

$$d_{e1AB} = \frac{1}{2}\,(d_{e1A} + d_{e1B}),$$

where the resulting error of the computed Q parameter is under 1% for ratio $B/\omega_0 \leq 15$.

The necessary circuit symbolic analysis and all the symbolic evaluations including the derivation of symbolic simplified transfer function parameters were made using mathematical program MAPLE V, release 5.

The optimization procedure alone is based on utilization of circuit degrees of freedom given by an additional number of optional circuit elements in comparison to the number of given design parameters. In the considered case, the optimized circuit has eight degrees of freedom, with respect to the six given parameters of the transfer function (21) and 14 optional passive elements. In the following, the procedure will be discussed in detail and the results demonstrated on typical examples.

The basic stage of the optimization procedure includes minimization of the simplified transfer function coefficient errors (27)[1]

$$\delta(a_i) = \frac{a_{ei} - a_{i(id)}}{a_{i(id)}};\ \ \delta(d_i) = \frac{d_{ei} - d_{i(id)}}{d_{i(id)}},\ i = 0,\ 1. \qquad (27)$$

The symbolic form of coefficient errors (27), evaluated using MAPLE, contains positive and negative terms; i.e. there is possible to set them to zero. As proved in Ref. [7], three of the main errors can be zeroized simultaneously in combinations (28).

Note that the conditions (29) give less applicable results. The first set of conditions leads to the additional design equations for GIC elements in the form (30)

$$R_{21} = \frac{Y_{11}R_{11}R_{31}}{Y_{11}R_{11} + 2 + Y_{21}R_{31}};\ \ R_{22} = R_{12}R_{31}\,\frac{Y_{22} + Y_{12}}{2};$$

$$R_{32} = R_{21} + R_{22}. \qquad (30)$$

The computation was made under choice of $k_1 = k_2 = \omega_0$, in conformity with the recommendation published in Ref. [8]. It is important to point out the influence of the non-corrected Q'_z - error, which causes magnitude frequency response distortion in the vicinity of frequency ω_0, as will be shown in the numerical example. To avoid this, two ways can be used to the fully correct design:

a) A predistortion of Q_z value, making final value errorless. This way is simple, because the error $\delta(a_1)$ is expressed by formula (31) evaluated under conditions (30). The improved form of design equations then includes the expressions (18) and (19) for transfer function coefficients together with optimization conditions (28). Note that the equation for coefficient a_1 in (18) is modified in the sense of the Q_z predistortion to the form (32).

$$\Delta(a_1) = a_{e1} - a_{1(id)} = \frac{\omega_0^2\,R_{31}}{B}; \qquad (31)$$

$$h\,\frac{\omega_0}{Q} = \frac{\alpha_1}{k_2}\,\frac{Y_{21}}{Y_{12} + Y_{22}} - \Delta(a_1); \qquad \alpha_1 = \frac{R_{21}}{R_{11}};$$

$$k_2 = \frac{R_{12}R_{32}C_{42}}{R_{22}}. \qquad (32)$$

The circuit element design formulae computed by MAPLE give the resulting expressions (33). It is important to say that the "free" optional parameters (R_{11}, R_{12}) influence circuit dynamic behaviour and R_{31} affects frequency response. Unfortunately, the optimum values of these elements limit to zero and, from practical design point-of-view, their values should be chosen as small as possible.

b) The second way uses modified optimization conditions (28), or (29), which keep the ω'_{0p} and ω'_{0z} parameters errorless and make the Q-errors equal, i.e.

$$\delta(d_1) = \delta(a_1) \ \Rightarrow\ \delta(Q'_p) = \delta(Q'_z).$$

To avoid an additional group-delay error at frequency ω_0 caused by Q-errors, the design can be combined with previously applied predistortion of coefficients a_{e1} and d_{e1}.

The corresponding basic set of design equations is shown in (34).

$$\delta(a_0 = 0;\ \delta(d_0) = 0;\ \delta(d_1) = 0 \ \Rightarrow\ \delta(\omega'_{0p}) = 0;\ \delta(\omega'_{0z}) = 0;\ \delta(Q'_p) = 0, \qquad (28)$$

$$\delta(a_0 = 0;\ \delta(d_0) = 0;\ \delta(a_1) = 0 \ \Rightarrow\ \delta(\omega'_{0p}) = 0;\ \delta(\omega'_{0z}) = 0;\ \delta(Q'_z) = 0. \qquad (29)$$

[1] As evident from (23), the coefficient errors indirectly express the errors of ω_0 and Q parameters as well.

$$Y_{10} = \frac{\omega_0^2\, Y_{22}(1-h)}{h};\; Y_{11} = \frac{\omega_0 Y_{22}(B + B\,h + R_{31}\,Q\,\omega_0)}{h\,Q\,B};\; Y_{12} = \frac{Y_{22}(1-h)}{h};$$

$$Y_{21} = \left(\frac{\omega_0^3\, R_{31}\, Q}{B^2\, h} + \frac{(2h+1)\omega_0^2}{h\,B} + \frac{(h+1)\omega_0}{R_{31}Q}\right)R_{11}Y_{22} + \frac{2\omega_0 Q}{B} + \frac{2h}{R_{31}};\qquad Y_{20} = \omega_0^2 Y_{22};$$

$$R_{21} = R_{11}R_{31}C_{41};\; R_{22} = \frac{R_{12}Y_{22}R_{31}}{2h};\; R_{32} = R_{21} + R_{22};$$

$$C_{41} = \frac{B\omega_0 Y_{22}}{(R_{31}\,Q\omega_0^2 + B(h+1)\omega_0)Y_{22}R_{11} + 2h\,Q\,B};$$

$$C_{42} = \frac{(R_{31}\,Q\omega_0^2 + B(h+1)\omega_0)Y_{22}R_{11} + 2h\,Q\,B}{(R_{31}\,Q\omega_0^2 + B(h+1)\omega_0)R_{11}R_{12}Y_{22} + 2\,h\,B(Q\,R_{12} + \omega_0 R_{11})}$$

$$k_1 = k_2 = \omega_0;\qquad Y_{22}\,,\; Y_{11}\,,\; R_{31}\,,\; R_{12}\text{ - optional parameters.} \tag{33}$$

$$h\frac{\omega_0}{Q} = \frac{\alpha_1}{k_2}\frac{Y_{21}}{Y_{12}+Y_{22}} - \Delta(a_1),$$

$$\frac{\omega_0}{Q} = \frac{1}{k_2}\frac{Y_{11}-\alpha_1 Y_{21}}{Y_{12}+Y_{22}} - \Delta(a_1),$$

$$h\omega_0^2 = \frac{1}{k_1k_2}\frac{Y_{20}}{Y_{12}+Y_{22}},\;\; \omega_0^2 = \frac{1}{k_1k_2}\frac{Y_{10}+Y_{20}}{Y_{12}+Y_{22}}, \tag{34}$$

$$h = \frac{Y_{22}}{Y_{12}+Y_{22}},\; k_1 = \frac{R_{11}R_{31}C_{41}}{R_{21}},\; k_2 = \frac{R_{12}R_{32}C_{42}}{R_{22}},$$

$$\alpha_1 = \frac{R_{21}}{R_{11}},\; k_1 = k_1 = \omega_0\,,\; \delta(a_0) = 0\,,\; \delta(d_0) = 0\,,$$

$$\delta(d_1) = \delta(a_1)$$

A symbolic solution of the Eqs. (34) was made using MAPLE. The free design parameters are the same as in the case a), i. e. Y_{22}, R_{11}, R_{31} and R_{12}.

The discussed design procedure will be now demonstrated on the numerical example. Let us consider the frequency normalized allpass transfer function $H(s)$ assigned by parameters $h = 1.0$, $\omega_0 = 1.0$, $Q = 4.0$,

$$H(s) = 1.0\,\frac{s^2 - 0.25s + 1}{s^2 + 0.25s + 1} \tag{35}$$

proposed to the group-delay equalization in the denormalized frequency range 10^0 MHz. The circuit realization presumes CFOA AD 844 as the active element. Normalized amplifier main parameters are $R_T = 400$, $B = 20$.

The simplest design version corresponding to the "basic" design equations (18), (19), (20) and (30) leads to the divider element values $Y_{10} = 0$, $Y_{11} = 0.50000$, $Y_{12} = 0$, $Y_{20} = 1.0$, $Y_{21} = 8.50000$, $Y_{22} = 1.0$; and GIC passive RC-network components $R_{11} = 0.25$, $R_{21} = 0.00735$, $R_{31} = 0.25$, $C_{41} = 0.11765$, $R_{12} = 0.25$, $R_{22} = 0.03125$, $R_{32} = 0.03860$, $C_{42} = 3.23809$.

An obtained circuit analysis confirmed correctness of design procedure and acceptable accuracy of the developed algorithm for transfer function order reduction. The evaluated parameters of the "full" and simplified transfer functions are summarized in the Table 1.

The non-zero parameter $\Delta(a_1)$ causes magnitude frequency response error -0.446 dB at frequency $\omega_0 = 1$. Simultaneously group-delay error $\delta(\tau) = 2.71\%$ arises at the same frequency.

An improved design procedure using Eqs. (33) gives the following results under the same initial conditions (choice $R_{11} = R_{31} = R_{12} = 0.25$, $Y_{22} = 1.0$).

A basic design results — Table 1

Function	ω_{0p}	Q_p	ω_{0z}	Q_z	Q'_{zAB}	σ_p	$\Delta(a_1)$
ideal	1.0	4.0	1.0	4.0	-	-	0
"full"	0.999982	4.000107	1.000001	4.210523	-	-	0.0124998
simplified	0.999999	4.000000	1.000024	4.210475	4.210526	69.29936	0.012500

Resulting parameters with ae1 predistortion Table 2

Function	ω_{0p}	Q_p	ω_{0z}	Q_z	Q'_{zAB}	σ_p	$\tau(1)$
ideal	1.0	4.0	1.0	4.0	-	-	16.00000
"full"	0.999982	4.000107	1.000001	3.999997	-	-	16.01271
simplified	0.999999	4.000000	1.000024	3.999951	4.000000	69.301956	16.01271

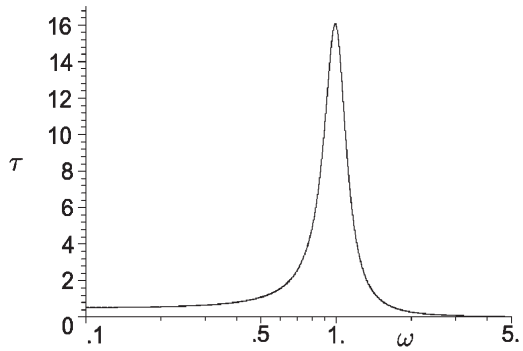


Fig. 6. Group-delay frequency response

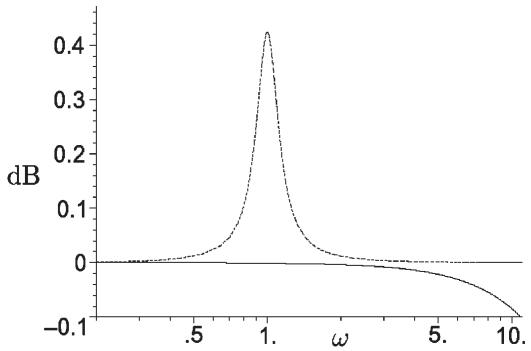


Fig. 7. Magnitude frequency response

Element values: $Y_{10} = 0$, $Y_{11} = 0.51250$, $Y_{12} = 0$, $Y_{20} = 1.0$, $Y_{21} = 8.93813$, $Y_{22} = 1.0$; $R_{11} = 0.25000$, $R_{21} = 0.00734$, $R_{31} = = 0.25000$, $C_{41} = 0.11747$, $R_{12} = 0.25000$, $R_{22} = 0.03125$, $R_{32} = = 0.03859$, $C_{42} = 3.23900$.

The corresponding magnitude a group-delay frequency responses are shown in Fig. 6, 7. In Fig. 7 the dotted line denotes the frequency response of the predistorted transfer function.

Class b) design, based on Eqs. (34) gives similar results. Divider admittance values are $Y_{10} = 0$, $Y_{11} = 0.52500$, $Y_{12} = 0$, $Y_{20} = 1.0$, $Y_{21} = 4.26250$, $Y_{22} = 1.0$, and GIC passive elements $R_{11} = 0.25$, $R_{21} = 0.015396$, $R_{31} = 0.25$, $C_{41} = 0.246334$, $R_{12} = = 0.25$, $R_{22} = 0.0625$, $R_{32} = 0.077896$, $C_{42} = 3.20941$. Note that the design uses the same values of the optional elements to obtain comparable results to class a) versions. Similarly to the previous cases the designed circuit was simulated and analysis results are summarized in Table 3. For illustration, the τ -error frequency response is shown in Fig. 8 and magnitude frequency response in Fig. 9. As it can be observed, the correction of magnitude frequency response is worse in comparison to the class a) design.

Parameters of the class b) design Table 3

Function	ω_{0p}	Q_p	ω_{0z}	Q_z	Q'_{zAB}	σ_p	$\tau(1)$
ideal	1.0	4.0	1.0	4.0	-	-	16.00000
"full"	0.999614	4.000916	1.000002	3.999986	-	-	16.02677
simplified	0.999637	4.000559	1.000049	3.999899	3.999999	34.608310	16.02677

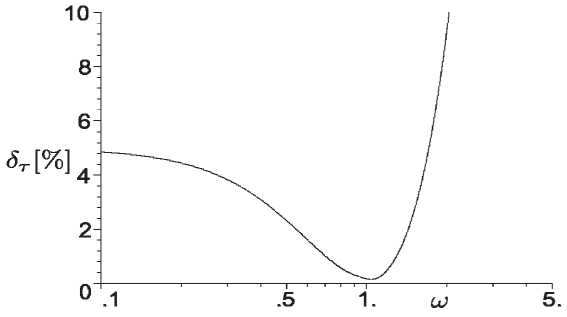


Fig. 8: Group-delay error response

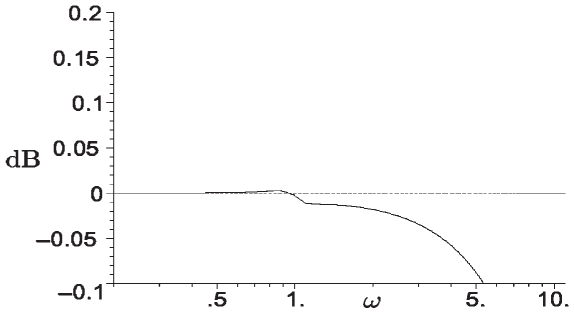


Fig. 9: Magnitude frequency response

Circuit parameters for the compromise design Table 4

Function	ω_{0p}	Q_p	ω_{0z}	Q_z	Q'_{zAB}	σ_p	$\tau(1)$
ideal	1.0	4.0	1.0	4.0	-	-	16.716
"full"	0.998381	4.461324	1.000172	4.478866	-	-	17.912
simplified	0.998212	4.457200	1.000232	4.479514	4.480210	14.426514	17.912

Comparing the evaluated circuit parameters, it is possible to render some partial conclusions:

- Both the class a) versions give similar results.
- Sensitivity of transfer function parameters to the amplifier GBW is significantly influenced by suitable chioce of R_{31} value. This fact is in agreement with general theory of current-feedback circuits. The lesser value of R_{31} makes circuit frequency range wider.
- Class b) design leads to a higher sensitivity to the amplifier GBW, which is evident from the comparison of the first parasitic poles of the resulting "full" transfer function, or indirectly, by comparing σ_p values. To improve the frequency properties and gain conformable results, it is necessary to reduce the R_{31} value approximately by half.

The higher stage of the optimized design includes additional dynamic optimization in the sense of the equalization of amplifier maximum output voltages. Dynamic analysis disclosed inappropriate overshoot of the fourth amplifier output voltage in the vicinity of frequency $\omega = 1$. To improve circuit dynamic properties, the set of design equations was extended about conditions (36) expressing the request of equal amplifier output voltages at frequency ω_0

$$Mod_{A1}(\omega_0) = Mod_{A2}(\omega_0) = Mod_{A3}(\omega_0) = Mod_{A4}(\omega_0). \quad (36)$$

Here $Mod_{Ai}(\omega_0)$, $_{i\,=\,1,\,2,\,3,\,4}$ expresses symbolically evaluated modul of partial transfer function

$$\left|H_{Ai}(j\omega_0)\right| = \left|\frac{V_{oi}(j\omega_0)}{V_{in}(j\omega_0)}\right| \quad i = 1, 2, 3, 4 \quad (37)$$

corresponding to the amplifier outputs at frequency ω_0.

A solution of the extended design equations provided the following results:

- The "full" set of equations containing Eqs. (33) and the additional dynamic conditions (36) is unsolvable, the requests to the ω_0 - and Q-errors minimization negate dynamic equalization.
- The full dynamic optimization allows only ω_{0z}-error zeroing, the remaining errors are uncorrected. These can be minimized by a suitable choice of the optional value of R_{32}, limit value is $R_{32} \to 0$.
- Compromise solution including ω_{0p} and ω_{0z} errors zeroing and partial dynamic equalization seems to be the most acceptable. Optimum results were achieved considering conditions

$$Mod_{A1}(\omega_0) = Mod_{A3}(\omega_0) = Mod_{A4}(\omega_0).$$

Maximum output voltage of the 2nd amplifier is in this case lower than in others.

- In general, dynamic optimization deteriorates frequency properties and leads to the higher errors of resulting group-delay and magnitude frequency responses. A fully acceptable solution is achievable using a more sophisticated optimization strategy, e.g. using evolutionary algorithms, or using a current-mode design.

A numerical illustration of the results obtained by using a compromise design is shown in table 4. Here the optional parameters were chosen $R_{31} = 0.125$, $R_{32} = 0.75$ and $Y_{22} = 1.0$. Note that the ratio R_{32}/R_{31} influences circuit frequency properties and the chosen value corresponds to the local optimum. The calculated dynamic overshoots attain to +13.5 dB. An additional predistortion of transfer function improving the final parameters was not made, but it would be possible. Fig. 10 illustrates the results of the equalization of the amplifier maximum output voltages. As evident, the simplified approach gives acceptable accuracy of the dynamic optimization.

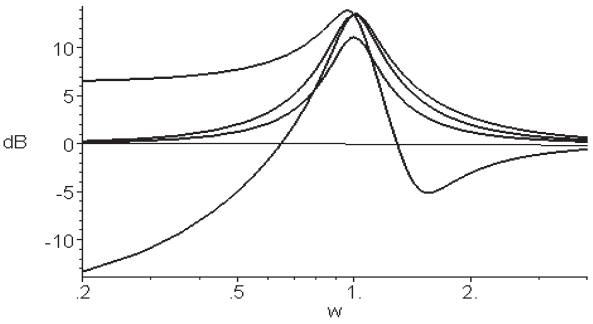


Fig. 10: Results of the dynamic optimization

5. Conclusions

The aim of this paper is to mention a new possibility of the allpass design. The described circuit is original, gained by the modification of the earlier published general divider structure. The use of CFOA warrants a wide frequency range, acceptable for design of phase equalizers in video- or fast-data-signal processing systems. The developed design procedures give improved solutions without noticeable group-delay and magnitude frequency response errors.

At this moment multicriterial optimization including circuit dynamics is not fully solved. This problem is a topic of the future research and its solution is posible using a current-mode design or by application of other variants of GIC circuits, e.g. GICs based on current conveyors.

References

[1] KVASIL, J.: *Lineární obvody*. NADAS, Praha, 1967.
[2] PLOSZEK, A., VOLNER, R.: *Elektrické analogové obvody*. VŠD Žilina, 1991, ISBN 80-7100-045-0
[3] HERPY, BERKA: *Active RC Filters.* Akademiai Kiado, Budapest, 1985
[4] SCHAUMANN, R., GHAUSI, M. S., LAKER, K. L.: *Design of Analog Filters.* Prentice Hall, 1990
[5] HASLER, M., NEIRINCK, J.: *Electric Filters.* Artech House, inc., Dedham, USA, 1986. ISBN 0-89006-186-6
[6] MARTINEK, P., TICHÁ, D., BOREŠ, P.: *A Contribution to the Design of Filters Based on a Generalized Divider.* In: Proc. of the ECCTD'97 Conference, Technical University Budapest, Budapest, 1997, Vol. 2, pp. 573–577 ISBN 963-420-523-2
[7] TICHÁ,D.: *Integrované filtre v telekomunikačnej technike* (in Slovak). PhD. Thesis, ŽU Žilina, 2000
[8] MARTINEK, P.: *An Optimized Design of Circuits Containing GICs.* Proc. of the 8th International Czech-Slovak Scientific Conference Radioelektronika 98, FEECS TU Brno,1998, Vol.1,pp. 42–45 ISBN 80-214-0983-5.

https://doi.org/10.26552/com.C.2004.4.48-52

kamUNikácie
COMMUNICATIONS

Vladimír Klapita – Zuzana Švecová *

OPTIMAL LOCATION OF SERVICE CENTRES UNDER UNCERTAIN COSTS

It is fairly questionable to estimate future costs in the problems of strategic decision. The uncertainty may cause that a resulting solution could be very inefficient considering current costs. In this report we try to find an approach, which enables to determine a unique solution of a location problem under uncertain costs so that the solution be resistant to future changes. We deal with a sensitivity analysis and with a connection of an exact mathematical programming method and the theory of fuzzy sets.

1. Introduction

In managerial experience we can find a problem of service centre optimal location. Location of those objects like manufactures, distributive and shopping centres, supply depots markedly affects the costs of material flows in creative logistic networks. The location of centers is so much complicated because there is not only one logistic chain but a whole distributive network.

Determination about a location or non-location of a service centre in some areas will affect the systems effectiveness for next several years. For finding the optimal solution it is possible to apply an exact method, but only for the known costs. When we solve location problems, for the most of them we have no real future costs, only their gross estimates. So it is necessary to deal with the approach of a location problem solving under uncertain costs.

This paper deals with a possible method of finding the optimal location of service centres under uncertain costs represented by fuzzy numbers.

2. Location problem formulation

A service centre can be set up only in some places from the finite set of possible locations, which requires standby costs. In the system are also costs of satisfying customer demands from some of located centres, which depend on quantity of requirements. The goal is to minimize complete costs of the system. So we have a difficult combinatorial problem of determination of a located service centres number.

There is securing freight traffic from one or more primary centres to customers in the distribution system. This freight traffic could be linear (without transshipments) or combined with transshipments in some centres called terminals, which are often warehouses or buffer stocks. The structure of distribution system is figured out by a set of primary centres, customers, terminals and flows of goods among them.

The location problem is a problem of optimal location of service centres on the given part of the transportation network.

The incapacitated location problem is conceived as follows:

The transportation network is given with customers in the nodes $j \in J$ and localities $i \in I$, in which it is possible to locate serve centres. Let's also assume that also one centre located in the node from the set I is able to serve all customers (see Fig. 1). The task is to minimize complete costs, which include standby costs f_i paid for each location of the service centre in i and variable costs c_{ij} of demand satisfaction b_{ij} of a customer j from the terminal i. The variable costs for satisfying demand b_j of customer $j \in J$ $c_{ij} = (e_1 d_{Si} + e_0 d_{ij} + g_i) b_j$ consist of charges e_1 for import from the primary centre S to the terminal i, costs g_i for transshipment in the transshipment i and charges e_0 for freight traffic from i to the customer j. The haul between the primary centre S and the terminal i is d_{Sj} and between the terminal i and the customer j is d_{ij}. The condition is that all the customers have to be served, or more precisely have to be assigned to some of the located terminals.

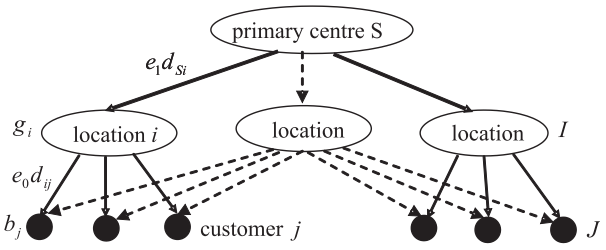


Fig. 1: Two designs of distribution system with transshipments

Having introduced 0–1 variable $y_i \in \{0,1\}$ for each $i \in I$, which models the decision if the terminal is located at i or not, and variable $z_{ij} \in \{0,1\}$ for each pair i, j, $i \in I, j \in J$, which assigns

* **Vladimír Klapita, Zuzana Švecová**
Department of Railway Transport, Faculty of Economics of Transport and Communication, University of Žilina, E-mail: Vladimir.Klapita@fpedas.utc.sk, Zuzana.Svecova@fpedas.utc.sk

the customer j to the terminal location i, we can set the following model of the complete cost minimization.

$$\text{minimize } f(y, z) \sum_{i \in I} f_i y_i + \sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij}$$

$$\text{where } c_{ij} = (e_1 d_{Si} + e_0 d_{ij} + g_i) b_i \quad (1)$$

$$\text{subject to } \sum_{i \in I} z_{ij} = 1 \quad \text{for} \quad j \in J \quad (2)$$

$$z_{ij} \leq y_i \quad \text{for} \quad i \in I, j \in J \quad (3)$$

$$z_{ij} \geq 0 \quad \text{for} \quad i \in I, j \in J \quad (4)$$

$$y_i \in \{0, 1\} \quad \text{for} \quad i \in I. \quad (5)$$

In the model above, the objective function (1) represents the complete costs of distribution system. Constraints (2) ensure that each customer demand has to be satisfied from exactly one terminal location, constraints (3) force placement of a terminal at location i whenever a customer is assigned to the terminal location i, constraints (4) ensure the location of terminal in every locality from which demands of some customers are satisfied.

3. Analysis of the existing approaches

In strategic decision problems it is difficult to estimate future values of standby or/and variable costs. In this case, the estimation of future costs is inaccurate. Considering confidential variables, which model determination about (un)location of terminals, the resulting solution can be economically inefficient in the view of the future costs. For example, the growth of f_i or e_1 creates a change of system structure of locations number and change of customers assignment (see Fig. 1). As a consequence, the estimation of expected costs by one numeric value is risky. Uncertain costs can be in that case described by an expectant interval of change of coefficient f_i or c_{ij}, (but the uncertainty is too big) or by fuzzy number, which gives us more information about charges.

There are two approaches how to overcome the uncertainty.

First of them is a classical sensitivity analysis [2], which tells us how the optimal solution changes when some of the parameters has other value than the one which was calculated.

If the uncertain parameter f_i changes in the interval $\langle f_i^1, f_i^3 \rangle$, by dividing this interval into m parts we will have $m+1$ location problems with the known costs (6). But the result of sensitivity analysis is not a unique solution.

Fig. 2: Interval of standby costs

$$\text{minimize } f(y, z) = \sum_{i \in I} f_i y_i + \sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij} \quad f_i \in \langle f_i^1, f_i^3 \rangle \quad (6)$$

Another approach uses the theory of fuzzy sets, where the uncertain value q is described by a possible interval and membership function μ_q (see Fig. 3) – it is a power of applicability of given element to q. This membership function has a triangular form.

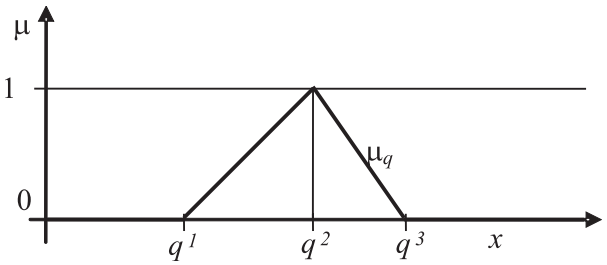


Fig. 3: Membership function μ_q

According to fuzzy arithmetic rules, fuzzy numbers can be mutually added, subtracted and multiplied and divided by a real number without loss of the triangular form. When the coefficients q_i of an objective function $F = q_1 x_1 + q_2 x_2 + ... + q_n x_n$ of a linear programming problem are triangular fuzzy numbers, then the value of the objective function for a given set of variable values $x = \langle x_1, x_2, ..., x_n \rangle$ is also a triangular fuzzy number:

$$F(x) = \langle F^1(x), F^2(x), F^3(x) \rangle = \left\langle \sum_{i=1}^{n} q_j^1 x_j, \sum_{i=1}^{n} q_j^2 x_j, \sum_{i=1}^{n} q_j^3 x_j \right\rangle \quad (7)$$

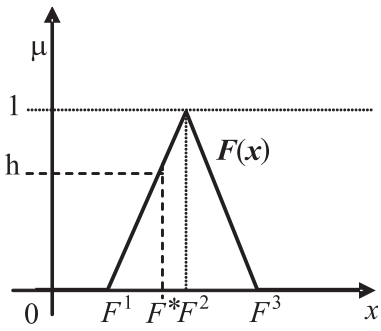


Fig. 4: Membership function of fuzzy number F

The existing approach [4], which uses the theory of fuzzy sets looks for a solution for thr given level of satisfaction h which is given by an expert (see Fig.4). So we solve the original task, but with a changed objective function describing uncertain costs:

$$\text{minimize } F^*(x) = F^1(x) + h(F^2(x) - F^1(x)) \quad (8)$$

The result of this method is a concrete determination, but credibility of the associated result depends on an expert's ability and his experience in determining a suitable level of satisfaction.

4. Concept of location problem solving

One method for finding a concrete determination about the service centres location, which is not dependent on an expert's

ability, is the fuzzy algorithm [3]. This approach is based on introducing the fuzzy set F_s, which expresses an assertion that "value of F is small" with the membership function shown in Fig. 5, where F^{min} and F^{max} denote respectively minimal values of $F^1(x)$ and $F^2(x)$ over a set of feasible solutions of the problem.

In this approach we searched a feasible solution x^*, for which the membership function of the fuzzy set "$F(x)$ *and* F_s" obtains the maximal value h (see Fig. 5).

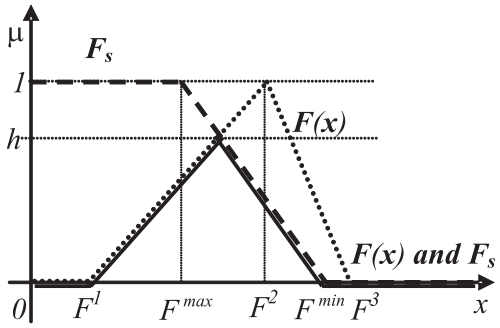


Fig. 5: Membership function of fuzzy sets F(x), F and their intersection F(x) and F_s

The maximal value h of the membership function of the fuzzy set $F(x)$ *and* F_s and for the given x has to satisfy the following equality in the cases when $F^1(x) \leq F^{max}$ holds.

$$F^1(x) + (F^2(x) - F^1(x))h = F^{max} - (F^{max} - F^{min})h \qquad (9)$$

In other cases h can be set to zero. For the former case we get

$$h(x) = \frac{F^{max} - F^1(x)}{F^2(x) - F^1(x) + F^{max} - F^{min}} \qquad (10)$$

and we seek for x^* maximizing $h(x)$, which is a non-linear programming problem. The following numerical process [3] obtains an approximate solution of the problem.

1* Set $\underline{h}$ to an initial positive value near zero.
2* Minimize the following objective function $F^1(x) + (F^2(x) - F^1(x))\underline{h}$ over the set of feasible x and denote $x^*(\underline{h})$ the associated optimal solution
3* Compute $h(x^*(\underline{h}))$ according to (10).
4* If $|h - h(x^*(\underline{h}))| < \epsilon$ then stop else set $h = h(x^*(\underline{h}))$ and go to step 3.

As it can be noticed in Fig. 5 or derived from the expressions (9), the direct fuzzy approaches make use only of the left hand side of the membership function. It means, that part of fuzzy number from F^2 to F^3 is not taken into account (see Fig. 5).

To overcome this weakness of the above-mentioned fuzzy approaches, there is another fuzzy approach [1], which makes use of the membership function on its whole range. This approach resembles way in which random coefficients are processed, when their distribution of probability is known. In this probabilistic-like approach the interval [0, 1] of possible values of the membership function is divided by real numbers of an arbitrary chosen finite set $H \subset [0, 1]$. Then for each fuzzy coefficient c from the location model the values $c^1, c^2, ..., c^r$ are determined, so that the constraint $\mu_c(c_k) \in H$ holds for $k = 1, 2, ..., r$. This is possible concerning fact that the level of satisfaction of a fuzzy number centre is 1.

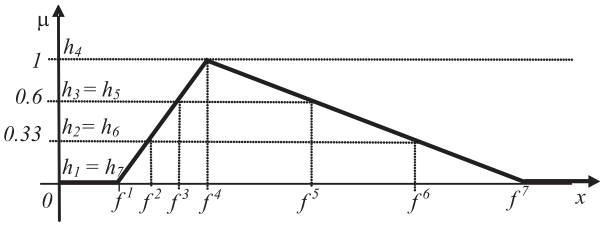


Fig.6: Level of satisfaction assignment to values of membership function

Then we minimize the weighted sum function over the feasible solutions D.

$$min\left(\frac{\sum_{k=1}^{m+1} \sum_{i \in I} h_k f_i^k y_i}{\sum_{k=1}^{m+1} h_k} + \frac{\sum_{k=1}^{m+1} \sum_{i \in I} \sum_{j \in J} h_k c_{ij}^k z_{ij}}{\sum_{k=1}^{m+1} h_k} \right)$$

subject to $(y, z) \in D$ (11)

The operating name of this method is *weights 2*.

In the case, we don't have more accurate information about uncertain costs, it means $h_k = 1$ for $\forall k = 1, 2, ..., m+1$ the method is named *minisum 2*.

If we use results of classical sensitivity analyses in the weighted sum function and find for which of those results is its value minimal, it is the method minisum 1. When we have nonzero weights, the method is named weights 1.

To compare and verify both approaches, there was implemented branch and bound method and built a software tool for sensitivity analysis and fuzzy processing of the location problem [1]. Functionality of the program was tested on 90 examples making use of the whole road system of Slovakia with 2906 dwelling places and 71 possible terminal locations. This way, in accordance with the primary source selection at 10 big towns of Slovakia 10 basic problems with predefined parameters f, e_1, e_0 were obtained By three types of modification done independently with each of the three parameters, there 90 benchmarks were obtained, which were used in the experiments.

An average locations number for the method *weights 2* is 9 ± 0.9.

Average objective function value for method *weights 2* is 32294723 ± 3190392 Sk.

Differences of average locations number between *weights 2* and other methods — Table 1

method	difference of average locations number from 9 ± 0.9
weights1	0.01 ⇒ 0.12 %
classical fuzzy method	0.49 ⇒ 5.39 %
fuzzy algorithm	1.09 ⇒ 12 %
minisum1	0.06 ⇒ 0.61 %
minisum2	0.06 ⇒ 0.61 %

Differences of average value of objective function between *weights 2* and other methods — Table 2

method	differences of average objective function value from 32294723 ± 3190392 Sk
weights1	174298 ⇒ 0.54 %
classical fuzzy method	660291 ⇒ 2.04 %
fuzzy algorithm	1619143 ⇒ 5.01 %
minisum1	172483 ⇒ 0.53 %
minisum2	99119 ⇒ 0.31 %

5. Conclusion

The fuzzy algorithm computes unique solution, which is not dependent on expert's ability (like classical fuzzy method) and is resistant to future changes.

We have compared these approaches:

- *sensitivity analysis* and its usage by methods *minisum 1*, *minisum 2*, *weights 1*, *weights 2*,
- *classical fuzzy method*,
- *fuzzy algorithm*.

If a triangular fuzzy number describes the uncertain costs of location problem, both methods *weights 2* and *fuzzy algorithm* are correct ways of finding the design of distribution system. These methods give similar results (see tables 1 and 2). There is difference 12% in number of placed terminals (average is 9 terminals). We suggest to perform both approaches and resulting design take into account only if results of these methods differ slightly. In opposite case, we suggest to perform an additional cost analysis and make the fuzzy cost more precise

Comparing the methods *minisum1* and *minisum2* with weights2 is only the reference example, because these approaches don't take wights into account.

One of the program outputs is a graphical representation of the solution, so the user can find out the stability of the optimal solution (see Fig. 7).

It is possible to change the values of a choosen parameter and also the method – analysis of sensitivity, classical fuzzy approach or fuzzy algorithm. The results are: object time, value of objective function, optimal number of terminals and their names and also associated customers.

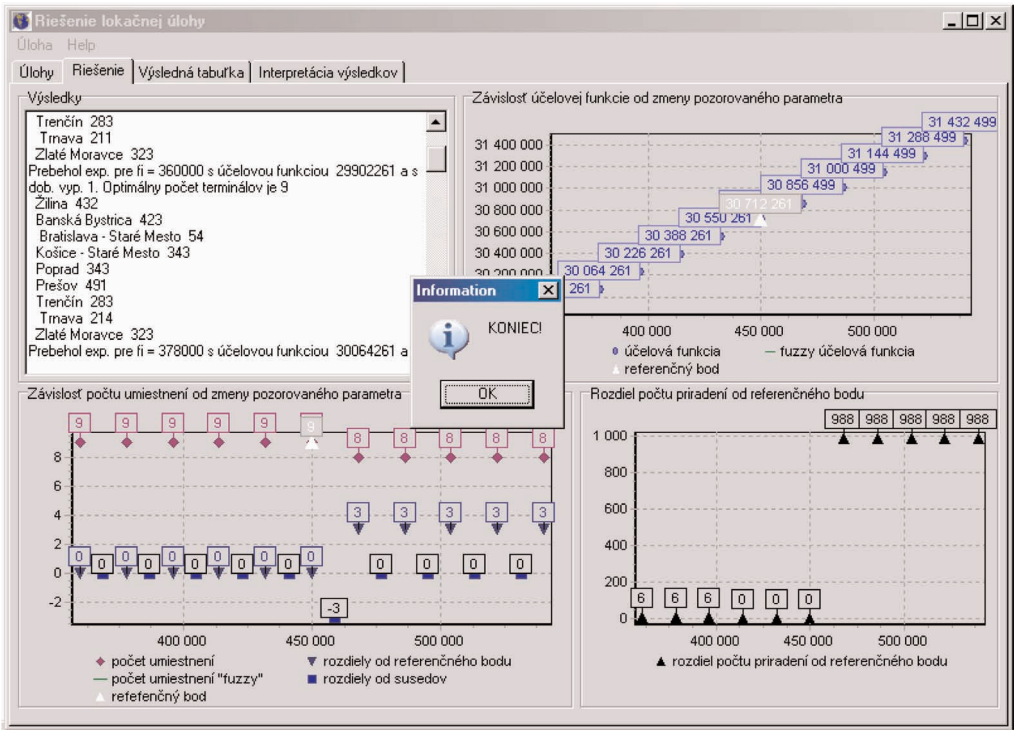


Fig. 7: Graphical output of the program

References

[1] ŠVECOVÁ, Z.: *Solution of location problems at uncertain costs* (in Slovak), Thesis, Faculty of Management Science and Informatics, University of Žilina, (2003), 90 p.
[2] JANÁČEK, J: *Optimalizace na dopravních sítích,* ŽU-Žilina, 2002, 248s.
[3] JANÁČEK, J: *Location problem under uncertain costs and charges,* In proceedings of the 18th conference "Mathematical methods in economics", 13.-15. 9. 2000, Prague, s.85–90.
[4] TEODOROVIČ, D., VUKADINOVIČ, K.: *Traffic control and transport planning: A fuzzy sets and neural networks approach,* Boston, 1998, 387 s.

https://doi.org/10.26552/com.C.2004.4.53-57

Róbert Jankovský – Radoslav Odrobiňák *

TECHNO-ECONOMIC EVALUATION OF BROADBAND ACCESS ALTERNATIVES FOR IP SERVICES

This paper deals with a techno - economic evaluation of broadband upgrade strategy for IP services. This scenario is focused on the FTTcab/VDSL architecture with ATM versus Ethernet approaches. Target market is residential customers and small business customers in urban and dense urban areas. Installed First Cost (IFC) and Net present Value (NPV) as main techno - economic results are discussed regarding common assumptions. The results show that the choice of technology (Ethernet or ATM) has almost no effect on the cost level and profitability of the cases.

1. Introduction

VDSL, as one of many broadband alternatives for upgrading the existing access network offers up to 50 Mbps to the customers. Advanced residential customers using the ADSL technology will expect faster applications and services in the near future. High capacity of VDSL require short copper loops and therefore fiber to a certain extent is needed in the access network. Two main technologies are candidates for new evolutions and IP-based services. The first is based on Gigabit Ethernet switching as technologies largely deployed in LAN is expected be cost effective. The second is based on Full Service Access Network (FSAN), whose main objective is accelerate optical penetration in the access network. It defined Broadband Passive optical Network (BPON), based on ATM as a low-cost fiber based access alternative.

In this paper TONIC and TITAN [2] tool for techno-economic evaluations of broadband access networks was used. Ten years deployment of both strategies is simulated on urban and dense urban areas with 65 536 users and six classes of customers are defined depending on bit rates request or services demand.

Comparison scenario is realised for FTTCab/VDSL architecture with the two approaches (Ethernet versus ATM).

2. Model case description

Model in urban area is built on the Metropolitan Area Network (MAN) between Central Echange location and customer premises. Scenario depicted in figure 1 corresponds to the FTTCab/VDSL architecture, using either Gigabit Ethernet or ATM technology. Optical infrastructure is deployed between Central Exchange and Cabinet and copper cables are used from Cabinet to the customer locations. The urban area is modelled as an area of 8 km^2 with an average copper loop of 300 m with the mean subscriber density 2048 subscribers/km^2 and the dense urban area is modelled as an area of 2.5 km^2 with the subscriber density 6553 subscribers/km^2.

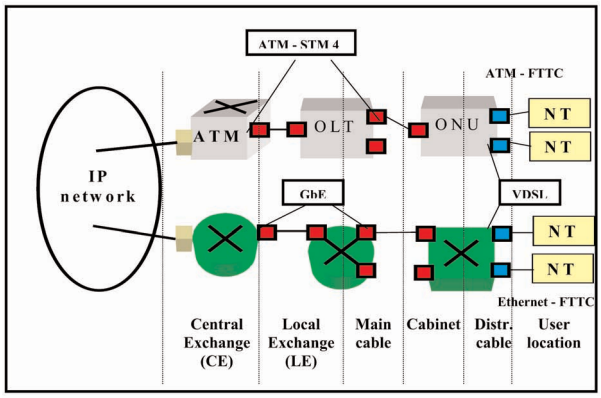


Fig. 1 FTTCab/VDSL architecture

In the Gigabit Ethernet approach, switches are deployed in Cabinet, Local Exchange (LE) and Central Exchange (CE). From CE to the Cabinet are the connections considered as Gigabit Ethernet connections and for the subscribers, as an Ethernet connec-

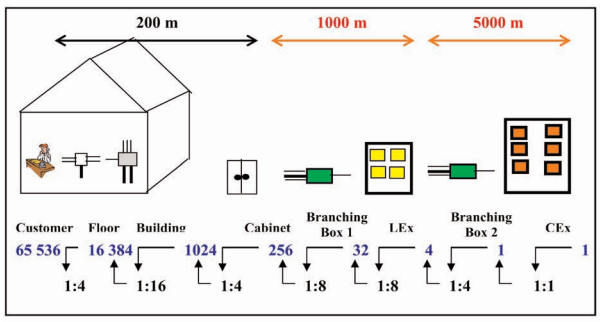


Fig. 2 Example of dense urban area

* **Róbert Jankovský, Radoslav Odrobiňák**
Department of Telecommunications, University of Žilina, Faculty of Electrical Engineering, Veľký diel, 010 26 Žilina, Slovak Republic,
E-mail: jankovsky@fel.utc.sk, odrobinak@fel.utc.sk

tions over DSL CPE (Customer Premises Equipment) with e.g. one 10BaseT downlink interface [5].

In the second case, the OLT is connected to the backbone via an ATM switches and handles ONU through a point-to-point STM4 link. VDSL modem is installed in the customer locations.

We consider two types of areas: urban and dense urban areas for our scenario. The geometric model [6] can be used to calculate the cable and duct lengths. This model is layered by flexibility points (FPs) and link levels. In our case, each Central Office encompasses four Local Exchange.

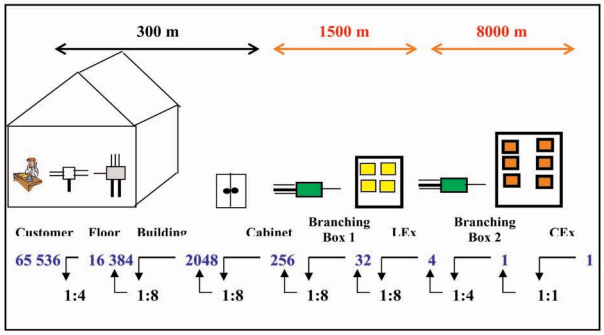


Fig. 3 Example of urban area

Now we must define the main parameters for both dense urban and urban areas. The dense urban area has shorter cable lengths than the urban area due to higher density of customers. So the number of the buildings is also different. The other parameter of infrastructure is for example duct availability, where we assumed the higher availability in the dense urban area than in the urban area.

3 Service requests

In this paper, users are divided into two main classes: residential and business customers. Type of service is dependent on a user profile. Besides, these two main classes can be divided into three subclasses. The range of services required by a business customer is even wider compared to residential customers (e.g. they require higher bit rates comparing to services dedicated to residential customers). Business subscribers often require symmetric bit rates while residential customers need high asymmetric bit traffic.

Broadband access forecast shows the number of incumbent operator subscribers in figure [7]. We assumed that total service penetration is 55% and the market share of incumbent operator is 70%. Then the total number customers of incumbent operator is 65 536 * 50% * 70% = 22 938. The monthly charge for asymmetric 1 Mbps connection is set to 55_ and for symmetric 2 Mbps is set to 100_ in normalized values.

Residential customers Table 2

Service class	Applications	Maximum bit rates dowvstream (Mbps)–asymmetric
Class 1	Fast Internet	1
Class 2	Fast Internet	2
Class 3	Fast Internet, One video channel	6

Business customers Table 3

Service class	Applications	Maximum bit rates dowvstream (Mbps)–asymmetric
Class 1	Fast Internet, Intranet	2
Class 2	Fast Internet, Intranet, VoIP	8
Class 3	Fast Internet, Intranet, VoIP, Data transmission	24

Provisioning represents 10%, Network operations 10% and Sales and marketing 15% of total revenues. For all calculations in this paper, we assumed that Discount rate is set to 10% and Tax rate equals to zero. I used the cost database which contains current or estimated list price information for modelling of network costs [4]. The modelling process was realized in macro Excel language and in Visual Basic programming language. The numbers of incumbent operator customers are presented in figure 4 and 5.

4 Our scenario results

We suppose 256 cabinet in each area with fiber-based technology up to the cabinet locations in our scenario. The connection between the cabinet and user locations is based on VDSL modem technology. Either it's Ethernet protocol based technology or ATM cell based in the case of FSAN. The difference between scenarious is in switching technology used.

Area description Table 1

Area	Subscrib. /km^2	Cable length CE-user	Surface	Number of buildings	Subscrib. per building	Duct availability Lex-Cab	% Bus.	% Res.
Urban	2 048	6 200	8	2 048	32	60%	60%	60%
Dense urban	6 553	9 800	2.5	1 024	64	90%	40%	40%

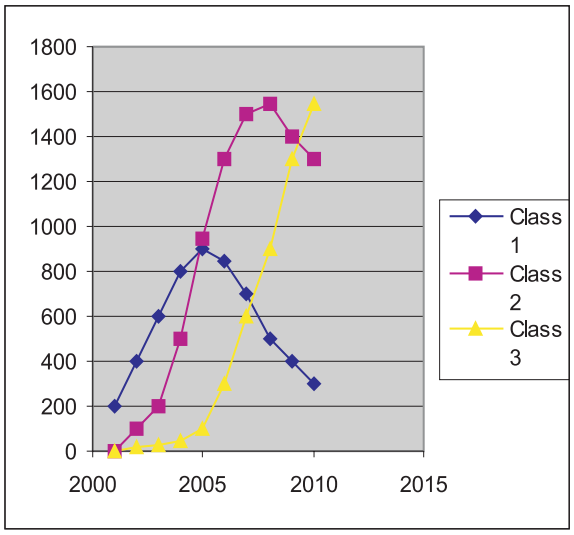


Fig. 4 Residential Service class

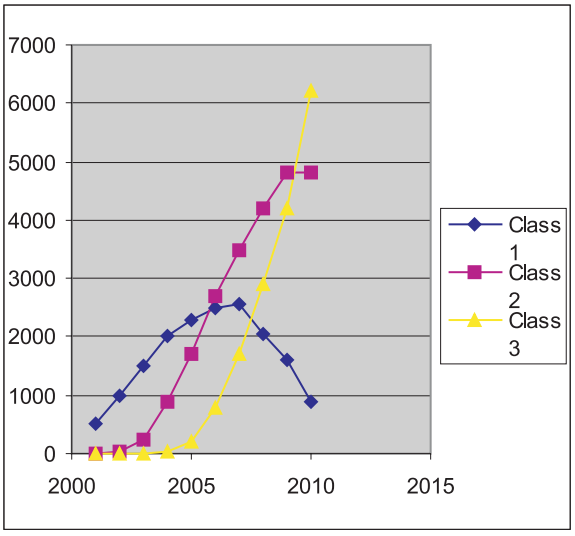


Fig. 5 Business Service class

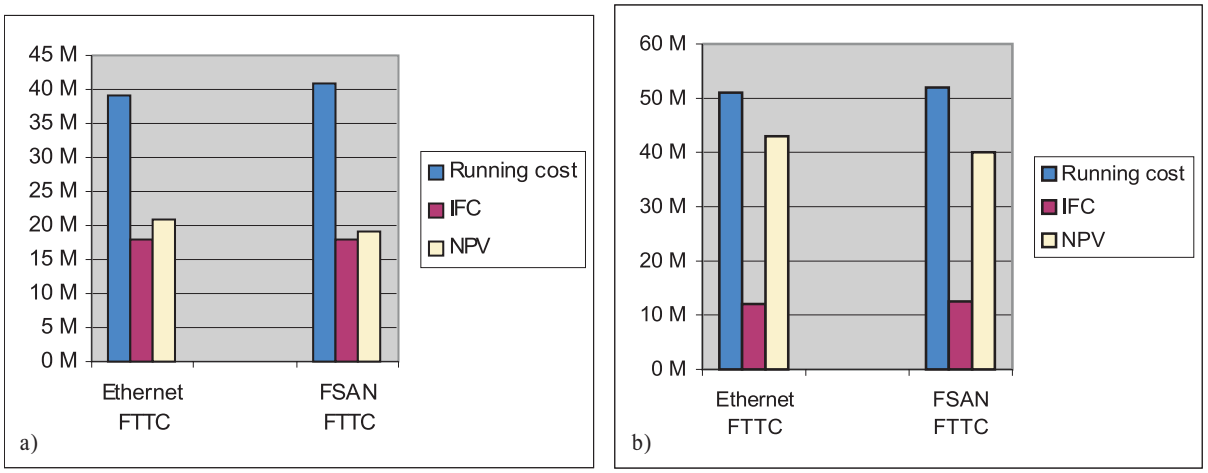


Fig. 6 Costbreakdown for a) the urban area, b) the dense urban area

In figure 6 we can see only a slight difference for the same areas. But if we compare the areas, the NPV is higher in the dense urban than in the urban area. The investments are lower in the dense urban area and revenues are higher because 60 % of customers are business customers. Operation&administration costs related to revenues are part of running costs and they are a bit higher in the dense urban area. On the other hand the NPV is twice higher in the dense urban area than in the urban area.

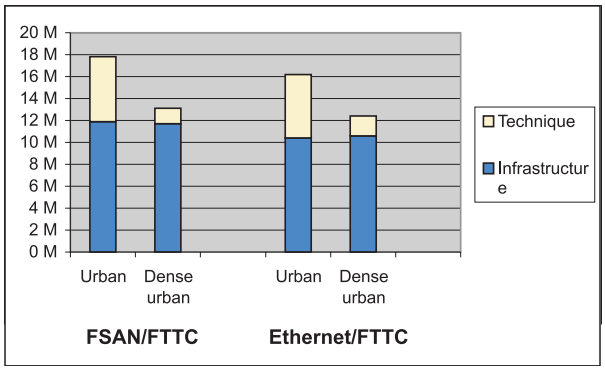


Fig. 7 IFC Breakdown

Figure 7 depicts IFC Breakdown into infrastructure and technique discounted investments. The difference in the same area is relatively small, up to 12 %. As we can expect investments in urban areas are higher than in dense urban areas due to infrastructure costs.

Figure 8 shows that the advantage of the Ethernet technology compared to FSAN/FTTC appears just in the beginning of the study period. The cash balance comparison in general is similar for both technologies. The Pay Back Period of the FSAN architecture can be reached after 6.2 years versus 6.0 years for Ethernet in urban areas and 4.6 years for FSAN versus 4.1 years for Ethernet in dense urban areas.

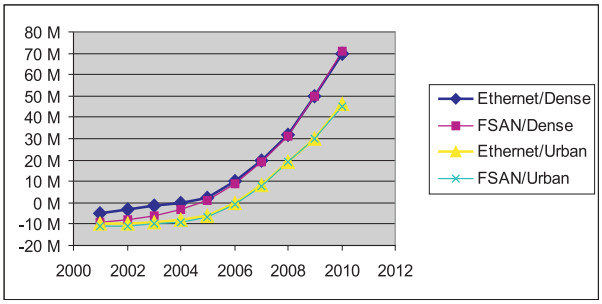


Fig. 8. Cash balance comparison

5. Component cost versus time

The cost evolution of every component in a special cost database is described by the following parameters:

- Reference year
- Price at a given reference year
- Volume class, which gives the cost of the component as a function of time with three parameters: $n_r(0)$, ΔT, and γ.
- $n_r(0)$, which is the accumulated production volume by $t = 0$
- ΔT, which is the time it takes for the accumulated production volume to increase from 10% to 90% of the total accumulated production volume
- γ, which describes the asymmetry of the applied logistic curve
- Learning curve class, which gives the cost of the component as a function of produced volume with one parameter, K: optimistic or pessimistic

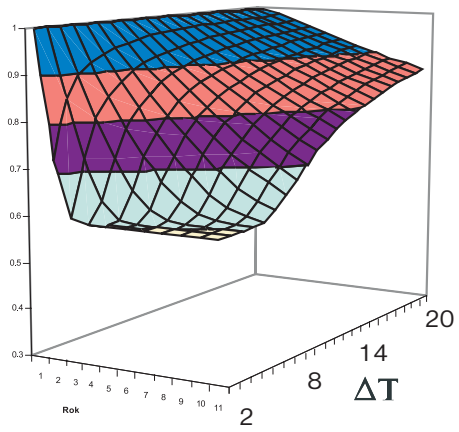


Fig. 9 The impact ΔT on the normalised price, if $n_r(0) = 0{,}001$

The final expression for price versus time in general is:

$$C(t) = C(0)*\left[n_r(0)^{-1}*\left(1 + e^{\left\{ \ln[n_r(0)^{-1}-1]-\left[\frac{2*\ln 9}{\Delta T}\right]*t \right\}} \right)^{-1} \right]^{\log_2 * \gamma}$$

Five parameters are the input to the formula:
$C(0)$ – price in the reference year 0
$n_r(0)$ – relative accumulated volume in year 0

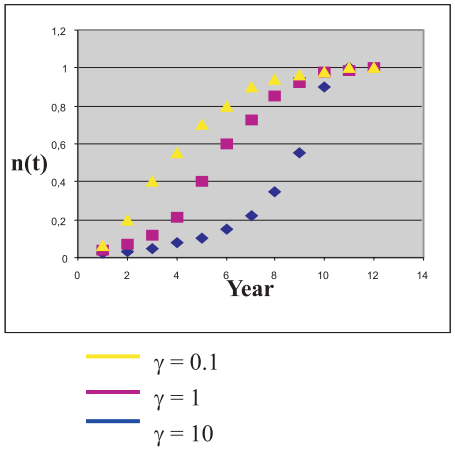


Fig. 10 The impact ΔT on the normalised price, keeping ΔT = 5 years

ΔT – time for accumulated volume to grow from 10% to 90%
K – learning curve coefficient
γ – asymmetry of the logistic curve

6 Conclusion

This paper presents the results obtained from simulation of techno – economic evolution broadband alternatives. These simulations are based on experiences obtained in TITAN and TONIC project and they are realized in the programming language C++ and in Excel tools.

For two areas (dense urban and urban), scenarios focused on the FTTCab/VDSL architecture with two approaches (Ethernet versus ATM) are realized for comparison. With the given assumptions FTTC solutions for urban and dense urban areas result in a positive NPV (net present value) with a payback period between approximately five and six years. The positive NPV is mainly driven by the existing infrastructure in terms of ducting systems, the housing infrastructure and the short total connection distance to the customers.

The most important variables are tariffs, the network operation costs, the access equipment costs and customer penetration or their density. The point-to-point Ethernet FTTC architectures seem to be a plausible solution as a first migration step after ADSL for both dense urban a urban areas. The results for FTTC show that choice of technology (Ethernet or ATM) has almost no effect on the cost level and profitability of the cases.

The figures for investment breakdown show just a small difference between different technologies. The differences have nothing to do with the protocol architecture, but rather with interface granularity and fiber consumption.

Figure 11 depicts the possible target architectures for broadband technologies in light of their integration to Next-Generation Networks (NGN).

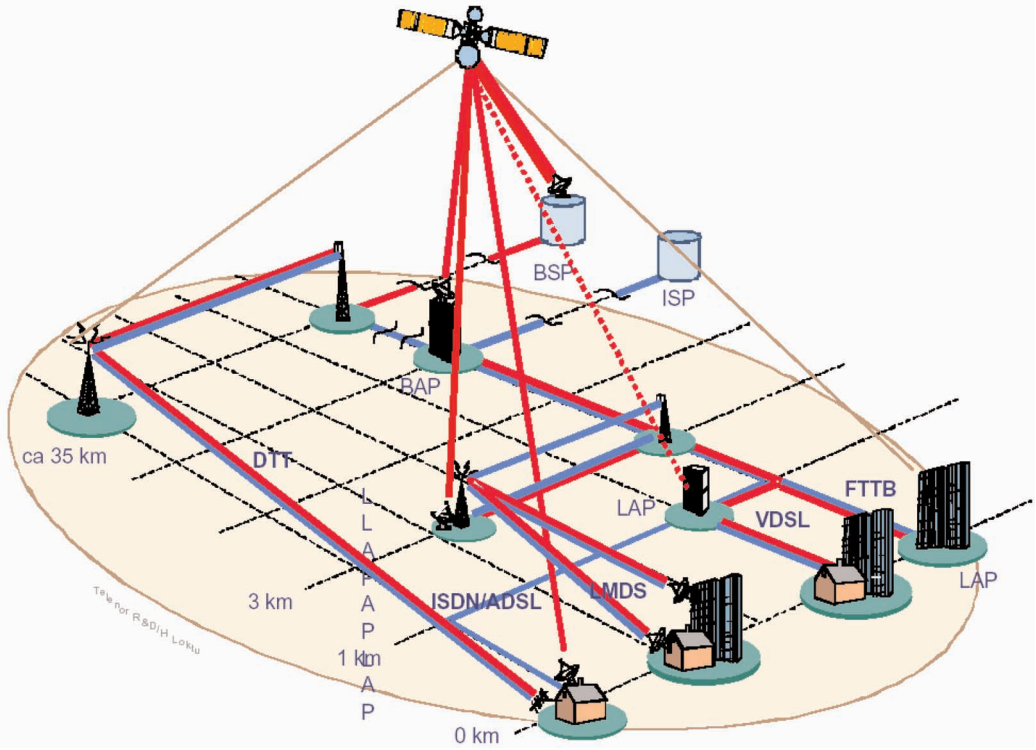


Fig. 11 Target architecture for broadband technologies

References

[1] MONATH, T., ELNERGAARD, N., CADRO, PH., DEVOLDERE, P.: *"Devirable 4: First results on broadband network solutions for new IP services offered in the fixed networks"*, IST-2000-25172 TONIC.
[2] IMS, L.A. (editor): *"Broadband access networks - Introduction strategies and techno-economic analysis",* Chapman-Hall, January, 1998.
[3] STORDHALL, K., RAND, L.: *Long Term Forecast for broadband Demand,* Telektronik, vol. 95, no. 2/3, 1999.
[4] Deliverable 8, *"Market Models for IP Services",* IST-2000-25172 TONIC, May 2002.
[5] MONATH, T. et al.: *"Economics of Ethernet Based Access networks for Broadband IP Services",* XIV Int'I. Symp. Services and Local Access, Seoul, Korea, Apr. 14–18, 2002.
[6] JANKOVSKÝ, R.: *"Návrh geometrických modelov pre prístupové siete"*, medzinárodná konferencia, ELEKTRO, 2004.
[7] Stordhall, K., Ims, L. A., Moe, M.: *"Broadband Market – The Driver for Network Evolution"*, Proc. Networks 2000, Toronto, Canada, Sept. 10–16, 2000.

https://doi.org/10.26552/com.C.2004.4.58-62

komunikácie
COMMUNICATIONS

Vladimír Wieser – Tomáš Marek *

UMTS UPLINK SIMULATOR

This paper presents the link level simulator developed at the Telecommunication Department of the University of Žilina. The simulator was developed as part of the project VEGA No. 1/0140/03 "Effective radio resources management methods in next generations of mobile communication networks" and it will be used in the more complex radio interface simulator of the next generation of mobile communication network built at our department. Students studying some problems about spread spectrum methods will also use the simulation model in some subjects, e.g. Digital communications, Mobile Radio Networks.

The simulator was developed in MATLAB Simulink and is made on a block principle, which allows adapting and configuring the model as necessary.

In this contribution we present the model solution and some simulation results in an uplink direction. The intention of the contribution is not to present new and original results (similar simulators were previously developed in many institutions [1,2]), but to give an overview what will be realized at our department in the field of mobile communication networks.

1. Introduction

The UMTS (Universal Mobile Telecommunication System) is the third generation mobile communication system, which is scheduled to start service in Europe as a part of IMT-2000 (International Mobile Telecommunication). It is designed to provide a wide range of services to mobile and stationary users in a variety of working environments and application areas.

The UMTS system is based on two different solutions: W-CDMA (Wideband Code Division Multiple Access), which uses frequency division duplex (FDD) and TD-CDMA (Time Division CDMA) designed for time division duplex (TDD). The presented simulation model is based on FDD mode [4].

The simulator is designed to evaluate the Bit Error Rate (BER) and Frame Error Rate (FER), which can be measured as a function of signal to interference ratio (SIR), spreading factor (SF), propagation channel (8 channel types), scrambling code, interleaver and Forward Error Correction (FEC) type.

Multiple access interference is generated by up to 8 interfering terminals with adjustable parameters.

2. Description of the simulator

The general structure of the radio chain is almost the same for voice and data services, so we used some simplification in the construction of time frames to be able to compare the results.

From a general point of view, *the transmission part* of the radio chain includes the following modules: source of binary data, CRC (Cyclic Redundancy Code) and tail bits insertion, channel coding, rate matching, interleaving, spreading, scrambling and modulation (Fig. 1) [1].

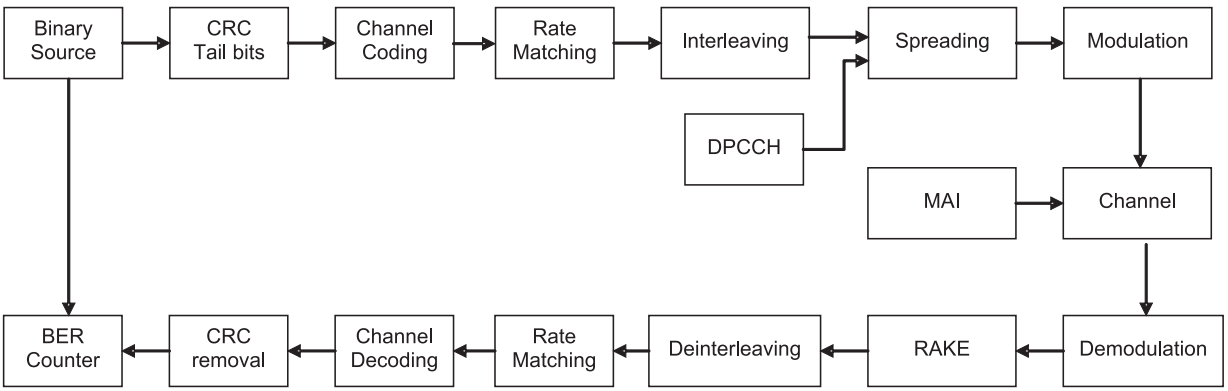


Fig. 1 Block diagram of the UMTS uplink radio chain

* **Vladimír Wieser, Tomáš Marek**
Telecommunication Department, University of Žilina, tel. 041 513 2260, E-mail: wieser@fel.utc.sk, kauly@szm.sk, student

The model blocks were created by 3GPP and ETSI specifications [5–9, 12].

Because of a very complex uplink frame structure for different services (transmission rates) we used a simplified version designed for voice services and altered only a spreading factor, which defines the transmission rate used in the channel.

After inserting 16 bits of CRC code and 8 tail bits the stream is channel coded with convolution code of coding rate 1/2, 1/3 or none (code constraint K = 9). Puncturing or repetition, to meet the closest rate for the chosen spreading factor SF does rate matching. The signal is interleaved (possibility to change the interleaver depth and the time, for which the maximum group error spans) and after multiplexing with a control channel DPCCH (SF = 256) it is spread in a spreading modulator.

The spreading is realized in the same manner for desired and interfering users (Fig. 2). After multiplying data and control channel by different channelization codes (Orthogonal Variable Spreading Factor – OVSF Codes) the resultant summing signal is multiplied by the complex scrambling code (unique signature of the user).

The *propagation channel* is made up of a multipath-fading generator described in details in [3]. The propagation environments and models defined by ETSI and ITU-R [11] are:

1. Indoor / Low Range Outdoor – Office A, Office B, Outdoor to Indoor & Pedestrian A.
2. Suburban Outdoor – Outdoor to Indoor & Pedestrian A, Outdoor to Indoor & Pedestrian B, Vehicular A.
3. Rural Outdoor – Vehicular A, Vehicular B.

These models are used in propagation simulation of useful and interference signals, respectively. The model allows switching off the short-term fading of each interfering source and setting up the propagation loss. It is also possible to model a propagation channel defined by the user who can set up a number of paths (up to six), delay and propagation loss for each path and speed of mobile terminal.

Multiple access interference (MAI) is modeled by up to eight interfering mobile terminals, which can be configured as it is shown in Fig.3.

The main window of the simulation model is in Fig. 4.

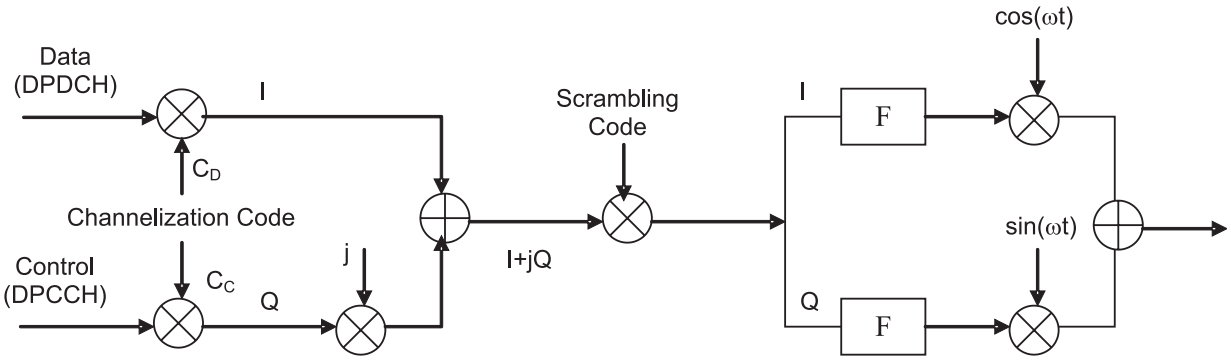


Fig.2 Uplink spreading and modulation

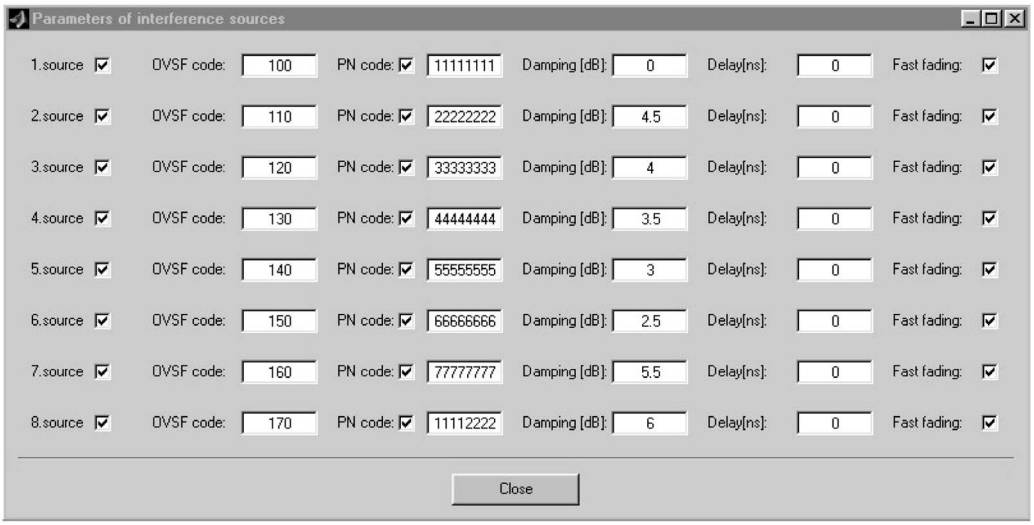


Fig. 3 Set up of interfering sources

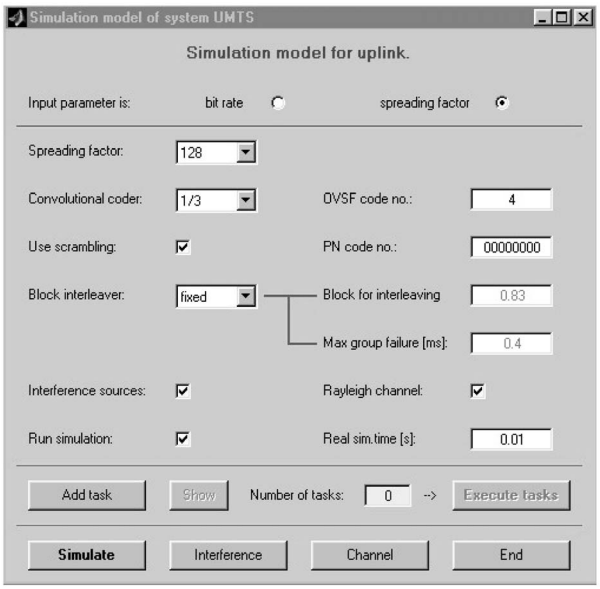


Fig.4 The main simulator's window

The *receiving part* consists of a Rake receiver with four branches tuned on the channel paths according to a chosen propagation model. De-interleaver and Viterbi decoder follows the Rake receiver, which is shown in Fig.5.

Unlike other Rake simulators [1, 2] we implemented a possibility of switching off each of the Rake fingers, which gives us the option to simulate the situation with more paths than is the number of Rake fingers. This situation is the most probable one in real channel and environment with many reflectors in the surroundings of the receiver.

3. Simulation results

The UMTS system uses orthogonal OVSF codes for channelization purposes (spreading codes) and PN codes (Pseudo-noise) as scrambling ones. The intention of scrambling codes is to improve bad autocorrelation and cross-correlation functions (in the system without synchronization) of OVSF codes and to differentiate users in the uplink direction.

To show the system behavior in different situations we simulated the situation when the system uses the same OVSF codes in each neighbor cell.

The most inconvenient situation arises when all mobiles (useful and interfering) are situated on the cell boundary. We simulated three different dependences. In all simulations we used Rayleigh fading channel (Clarke's model).

Fig.6 shows the dependence of BER on the number of interfering sources for three different situations. The output transmission power of the useful mobile and the interfering one are the same ($P_{out\_MS} = P_{int}$).

The importance of scrambling codes using is obvious because of the cross-correlation among the same orthogonal OVSF codes. If the system uses the same scrambling code for each mobile, BER is worse than in the "normal" situation (each mobile uses its own

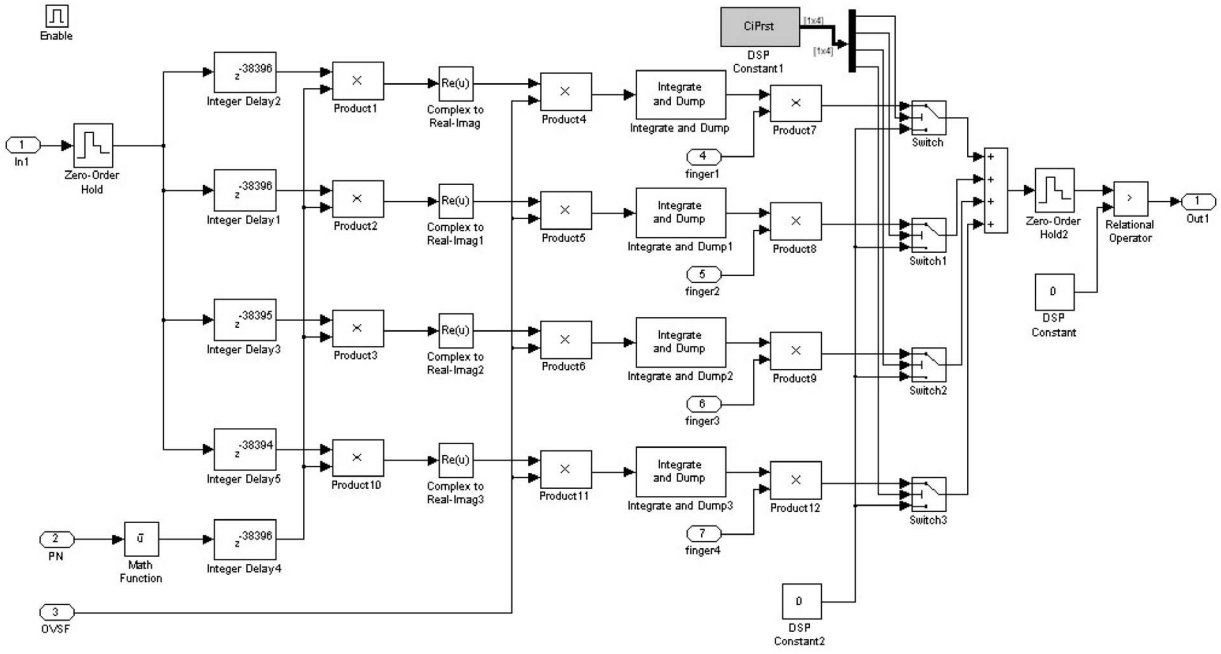


Fig. 5 Rake receiver

scrambling code), but it is much better than in a situation when the system doesn't use scrambling codes at all.

The next simulation (Fig. 7a, b) shows the dependence of BER = *f* (SIR) and FER = *f* (SIR) for the spreading factor SF = 4 (all mobiles), scrambling codes are different and mobiles use convolutional coding with the code rate R = 1/2, R = 1/3, R = 0 (without coding) together with an interleaver. To simulate the worst situation when coding + interleaving is used, the simulation was realized in Vehicular A channel (Rural Outdoor), mobile's speed 500 km/h with channel coherence time t_{coh} = 0.2 ms.

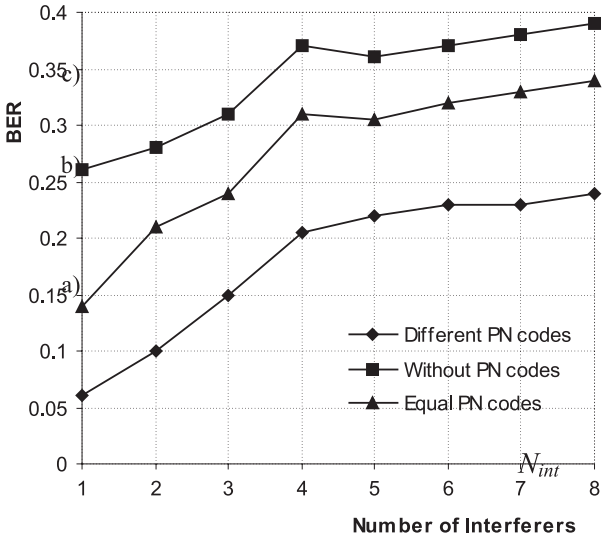


Fig. 6 The dependence of BER = f (N_{int}) for three different situations: a) the system uses different scrambling codes for each mobile, b) the same scrambling code is used in all mobiles in the system, c) spreading without scrambling codes at all

The results show very bad performance of convolutional coding with R = 1/2 in a situation with very strong interference (8 interferers with the same OVSF code). It is obvious that very small spreading factor (SF = 4) in the worst situation is incapable to create required processing gain ensuring the minimum value E_b/I_0 in the receiver.

In addition, the system without convolutional coding and interleaving has better performance than the coded system.

The situation is different in FER simulation (Fig.7b) because the system without coding will have at least one error in each frame (FER = 1).

To confirm our presumption about insufficient processing gain to suppress mutual interference, the simulation with SF = 32 was done (Fig. 8). The advantage of convolutional coding with interleaving is obvious.

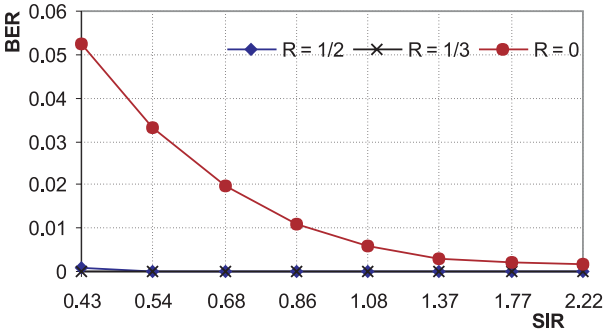


Fig.8 The dependence BER = f (SIR) with SF = 32

The last simulation shows the dependence of BER = *f* (SF) and FER = *f* (SF) for R = 1/2, 1/3, 0 with SIR = 0.45 (const.) (Fig. 9). This situation may arise when all mobiles are at the cell boundary. The poor performance of the convolutional coding with R = 1/2 up to SF = 32 is obvious. On the contrary, the performance of coding with R = 1/3 is poor only up to SF = 16.

Very promising results of coding with R = 1/3 are shown in Fig.9b where the improvement of performance with the coded system is obvious.

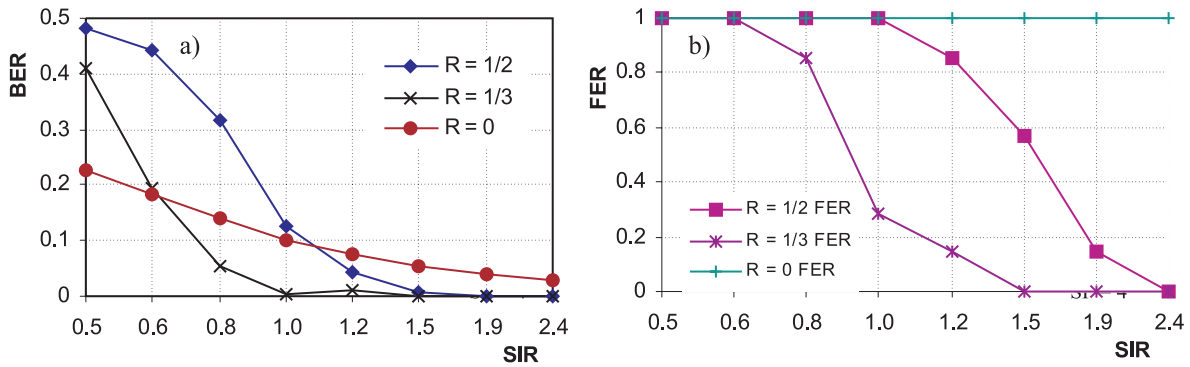


Fig.7 The dependence a) BER = f (SIR) and b) FER = f (SIR)

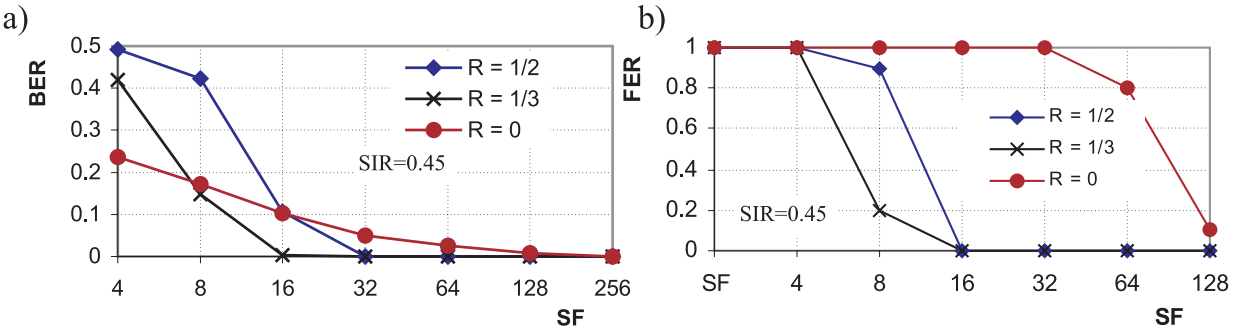


Fig. 7 The dependence a) BER = f (SIR) and b) FER = f (SIR)Fig. 9 The dependence a) BER = f (SF) and b) FER = f (SF)

4. Conclusion

In the article we presented the uplink UMTS simulator created in MATLAB Simulink with interactive GUI (Graphical User Interface). The simulator allows varying many important simulation parameters of the UMTS system (SF, coding rate, interleaver, parameters of channel and interfering sources, OVSF and scrambling codes, etc.).

The simulations were done for a critical situation with heavy mutual interference in the system. It was shown that convolutional coding with interleaving is not capable of repairing errors when the transmission rate is high (SF = 4).

References

[1] ROMANO, G., MELIS, B., MASTROFORTI, M.: *Link Level Performance of the W-CDMA Radio Interface for UMTS.* Rapporti Tecnici CSELT, vol. 28, 2000, pp.421–434. (model CDMASIM).
[2] FAKHRUL, A.: *Simulation of third generation CDMA systems.* [diploma work], Virginia Polytechnic Institute & State University, Blacksburg, Virginia, USA, 1999.
[3] MAREK, T., PŠENÁK, V.: *Simulácia digitálnych modulácií v mobilnom rádiovom kanáli.* [ŠVOČ práca], Žilinská Univerzita, Žilina, 2003.
[4] CASTRO, J. P.: *The UMTS network and radio access technology.* John Wiley & Sons, Ltd., England, 2001.
[5] 3GPP TS 25.101: *UE Radio transmission and reception* (FDD). V6.2.0 (2003-09).
[6] 3GPP TS 25.201: *Physical layer - general description.* V5.2.0 (2002-09).
[7] 3GPP TS 25.211: *Physical channels and mapping of transport channels onto physical channels* (FDD). V5.5.0 (2003-09).
[8] 3GPP TS 25.212: *Multiplexing and channel coding* (FDD). V5.6.0 (2003-09).
[9] 3GPP TS 25.213: *Spreading and modulation* (FDD). V5.4.0 (2003-09).
[10] 3GPP TR 25.944: *Channel coding and multiplexing examples.* V4.1.0 (2001-06).
[11] ETSI TR 101 112: *Selection procedures for the choice of radio transmission technologies of the UMTS.* V3.2.0 (1998-04).
[12] ETSI UMTS XX.07: UTRA FDD; *Physical layer procedures.* V1.3.1 (1999-02).

https://doi.org/10.26552/com.C.2004.4.63-70

Martin Klimo – Tatiana Kováčiková – Pavol Segeč *

SELECTED ISSUES OF IP TELEPHONY

Technologies for transporting voice over networks based on IP protocol, known under various names, e.g., Voice Over IP (VoIP), Voice on the Net (VoN), IP telephony (below IP telephony), made a great progress in the last decade. IP telephony has become one of the most important and the most rapidly developing Internet technologies that is offering a technical as well as economical alternative to the existing telecommunication networks. The paper points three issues that from the authors' point of view seem to be crucial for a wide deployment of IP telephony in European networks: Development and implementation of new IP telephony services based on the SIP (Session Initiation Protocol), Interworking between IP-based network signalling protocols and signalling protocols used in traditional SCNs (Switched Communication Networks) and the quality of voice transmission over IP networks. The paper also gives some examples of practical solutions of the problems mentioned above.

1. Introduction

Technologies for transporting voice over networks based on IP protocol, known under various names, e.g., Voice Over IP (VoIP), Voice on the Net (VoN), IP telephony (below IP telephony), made a great progress in the last decade. IP telephony has become one of the most important and the most rapidly developing Internet technologies that is offering a technical as well as economical alternative to the existing telecommunication networks. Dynamics of the IP telephony development, successful handling of a standardisation process, increasing offer of IP telephony products, successful implementation of IP technologies, and especially, offer of new attractive communication services on the market, seem to be the key steps in a process of convergence of different services and networks on the IP technology basis.

IP telephony as a service does not mean only classical telephone services like we know them from classical circuit-switched networks (PSTN – Public Switched Telephone Network, ISDN – Integrated Switched Telephone Network, GSM – Global System Mobile), but it offers also opportunities for creation of new attractive services based on integration of telecommunication and data worlds. In this "integrated, or converged world", other types of communication tools, apart from voice, may be used. This includes video, application sharing, virtual reality, etc.

Having taken into account the extreme opportunities of IP telephony for creation of new services, it is necessary to provide the right tools to speed up their development and implementation. There is a strong relation between the tools for service creation and a protocol that will be used for service delivery. In other words, the properties of a protocol have a strong impact on the opportunities for service creation.

In connection with new services provided over IP networks, the Session Initiation Protocol (SIP) is nowadays mentioned as the most perspective one. The main reason for that is its simplicity, the same principles with other Internet protocols (e.g., HTTP, SMTP) as well as the fact that it can be used as a control mechanism for multimedia multiparty connections, new types of services (e.g., Instant Messaging, etc.) as well as a signalling mechanism in Next Generation Networks (NGNs).

Despite the fact that there are some operators that are already offering IP telephony service based on the SIP (e.g., MCI in the USA, Telia in Europe), SIP technology still has to solve a number of issues before it can be widely deployed in European networks. Among those that are very topical, are: resources for the right tools for the development of new services that would advance the service of IP telephony at a qualitatively higher level (e.g., SIP – http integration), possibilities for interworking of existing network types and terminals with IP-based networks and new types of terminals for IP telephony, as well as the Quality of Service (QoS) for IP telephony.

2. Development and implementation of new IP telephony services based on the SIP protocol

The importance of Internet within the world telecom industry has grown rapidly in several last years. There is evident renaissance of communications visible mainly in the Internet world and feasible thanks to a suite of new signalling and control protocols, which are based on completely different principles as the centralized control systems used in the traditional circuit-switched systems. SIP – Session Initiation Protocol, as a signalling protocol developed to set up, modify and tear down multimedia sessions over Internet. Thanks to it close relations to the HTTP and SMTP protocols, the traditional telecom signalling and control models for telephony could be moved towards the Internet and web-based protocols. SIP does not provide only seamless integration of telephony and conferencing with many other WWW and messaging applications,

* **Martin Klimo, Tatiana Kováčiková, Pavol Segeč**
Department of InfoCom Networks, Faculty of Management Science and Informatics, University of Žilina

but benefits also from new forms of communications and new service features, like presence, instant messaging or mobility. Due to these facts SIP belong definitely among the key IP protocols.

In the Department of InfoCom Networks, the University Infoline (UI) service has been developed and implemented. The UI service integrates voice SIP and web services. This kind of services is related to the category of enhanced "click-to-call" services. The developed service is fully operational and the function of the service was tested in the framework of the international research project Eurescom P1111 NGOSSIP [1] through Internet. For the service purposes a SIP communication infrastructure has been built up at the Department of Information Networks.

SIP (Session Initiation Protocol)

SIP is a signalling protocol developed to set up, modify and tear down multimedia sessions over the Internet. SIP provides signalling and control functionality for a large range of multimedia communications. The main functions are: location of parties, invitation to service sessions, and negotiation of session parameters. To accomplish this, SIP uses a small number of text-based messages, which are exchanged between the SIP peer entities, i.e., SIP User Agents in user terminals. In time of messages exchange messages can traverse the network entities like proxy servers or redirect servers, which are used for support, like address resolution, routing calls to other entities, etc.

SIP only defines the initiation of a session; all other parts of the session are covered by other protocols. The session itself is then described in two levels. The SIP itself contains the parties' addresses and protocol processing features. The description of the media streams exchanged between the parties of a session is defined by another protocol. The Session Description Protocol (SDP) is used for this. The SDP defines text-based media description format that can be carried in the SIP message body. The message body is transparent to SIP; thus, any other session description can be carried (e.g., a weblink). From this point of view, SIP sessions are not limited to telephony calls or conferences only, but they can include any session description, for example web link, picture, e-mail address.

Programming SIP services and features

For defining services, SIP took a different approach as compared to standard telephony. Implementation of SIP services can be done in general:

- Using SIP baseline protocol mechanisms
- Defining extensions (headers, methods-SIP call control framework)
- Dedicated programming tools (SIP-CPL, SIP-CGI, SIP-servlets, Java applets, JAIN APIs, Parlay).

From another point of view, SIP services can be programmed either by the trusted (such as administrators), or by the untrusted (such as end users) users.

The example of the UI service shows the implementation of SIP service by the use of Java programming language and the proprietary SIP API by the trusted users.

Independently on a programming tool, service creation has the meaning of the implementation of a service logic that guides behaviour of each of the system elements and controls a specific message flow or reacts on a message request. In the UI service, a service logic is implemented on the SIP application server (which assemblies the functions of the SIP registrar, SIP proxy, Third Party Call Control module) side. Thus, the service logic is a Java programme, which directs the application server's actions based on inputs from the SIP messages exchanged between SIP elements, the service logic and localization database and the data got from the Web page.

Integration of web and SIP technologies

The UI Service integrates dynamic web technologies (Java Server Pages, Java Applet) with SIP. This kind of web and SIP integration brings a new communication environment, which provides new communication features allowing to program a vast number of new services, from the traditional telephone service to advanced multimedia communication services. Such services may include, for example, click-to-dial services, SIP profile management through a web interface, presence services, etc., which provide the user web and SIP data integrated into a web page.

Due to a similar client-server nature of both technologies, integration of SIP and web technologies shows the following aspects.

A client side aspect is *the first* one. This includes adding SIP multimedia and signalling capabilities into the client side and integration of these capabilities together with web. It can be done through SIP integration into the client web software (i.e. browsers). Browsers become universal desktop tools with the multimedia capability. To achieve such effect, incorporation of some kind of a proprietary SIP enabled plug-ins into a browser technology, or in a case of Java UA (standalone SIP UA or lightweight Java Applet SIP UA) adding support for the JAVA technology to a client operating system, is required.

The second aspect is the integration of web server technology with SIP server technology. This can be done, for example, through some other technologies, e.g. through database sharing.

The last aspect resides in a design and supplying of the information flows between client sides and server entities involved in a service. For example the service UI is composed of web browsers, a web server, SIP UAs, SIP registrar and proxy servers, a database and the information flows among them.

The service University Infoline

There is no doubt, that WWW service is the most popular Internet service, and popularity and utilization of this service still grows rapidly. This has been enabled by the fast development of dynamic web technologies (for example JSP), which are changing browser technology to complex desktop tools. On the other hand, SIP is bringing new multimedia features to these communications environments.

The University Infoline service has been developed by the Department of Information Networks at the University of Zilina. The service integrates both web and SIP technologies. The service

belongs to the group of click2call services and enables web browsing simultaneously with the establishment of SIP (Session Initiation Protocol) calls with persons who are responsible for a content of the web page (University / Educational Centre / Restaurant / Infoline etc.) The service is implemented as a part of distance learning environments of the departmental network, but in general, it may be implemented and used as any infoline service.

From a user point of view in the UI service, a student retrieves courses, which are listed in the web page of the University web space. Browsing a list of courses, a student wishes to get the detailed information on the course, provisions, etc. When he clicks on the hyperlink, a window with the course description appears together with the presence information of a tutor and a proposal to establish a call to a tutor of a course. If he accepts the offer, he becomes the SIP caller. The tutor becomes a callee.

Up to now, it looks more or less like a "classical click2call" scenario. In the frame of the service, we were interested in different states of a SIP call, which may occur between a student and a tutor and how to handle these states according to the tutors' personal preferences (system view). Under "tutors' personal preferences", we understand the possibility of a tutor to setup the time of day when he/she is available and when he/she wishes to accept calls initiated from the web on his/her personal SIP UA (User Agent). Thus, the following states of a tutor has to be distinguish and handle:

1. Tutor is online and available to accept calls initiated by a student from the web
2. Tutor is online, but busy with another call
3. Tutor is online, but not available to accept calls from the web
4. Tutor is offline.

In case 1., the SIP call between the tutor and the student is established.

In case 2. – 4., some options are made available for a student how to proceed. The options include providing a tutor contact addresses, i.e. e-mail address and SIP address of a called tutor, which can be used by a student to contact a tutor manually later and/or the possibilities, which allow a student to leave his/her contact addresses to the infoline system (e-mail or SIP address of a student, to which the system will sent the e-mail or establish a SIP call in the case a called tutor becomes online and available.

A complete service operates then in the following way (Fig. 1).

(1) Administrator creates a tutor account. (2) Data is recorded to a DB through JDBC. (3)(4) New tutor registers to his service web account through web and modifies his settings. (5) SIP registration of the tutor. (6) Registrar records the registration data to the DB. (7) Student is browsing through the list of courses and selects one of them. (7) Http request for a course is sent. (8) Web server loads SIP presence info of a tutor from DB. (9) Web page with course description and SIP presence sent back. In the case of

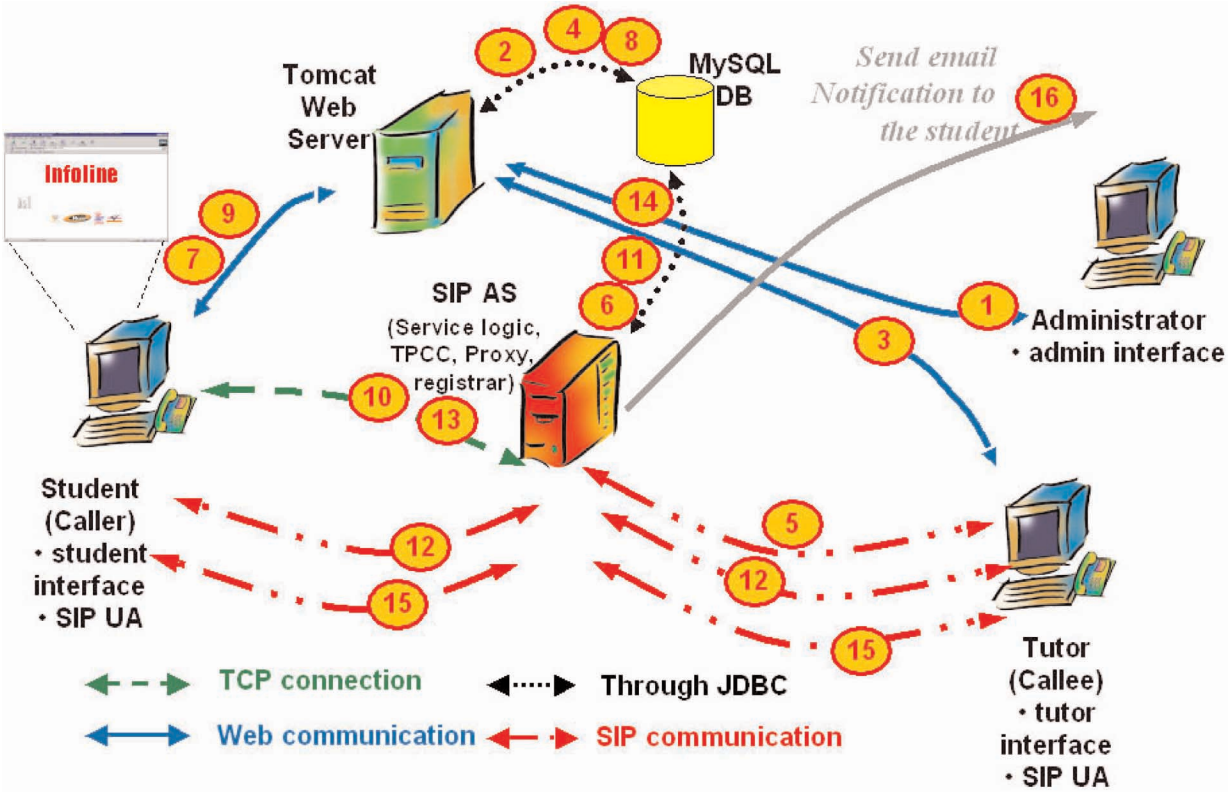


Fig. 1 Architecture of University Infoline service

a "clicking" on the call button, the Java Applet is loaded in the same way. (10) Applet loading finished; the applet establishes socket connection to the SIP AS and initiates a SIP call. (11) Service logic tests tutor status. (12) Tutor is online and available; TPCC establishes a call. (13) Tutor is not available; student is leaving his contact addresses. (14) Service logic records the student contact data to the DB. (15) Tutor is becoming online and available – system establishes the communication based on the student contact addresses (16) And / or system sends e-mail to the student

The Service has been designed as hardware and operating system platforms independent. Therefore all the used server software components (database, web and SIP servers) are available for more OS platforms. The service logic has been developed using Java and the StarSIP API supported by the StarSIP application server. As a web technology has been used JSP together with JavaApplet technology.

Efficient programming of new communication services is a key issue for Internet telephony. With SIP, services can be created that combine elements from telephony and other web applications such as email, messaging, the Internet and video streaming. Using SIP, services like "click-to-call" become possible, in which, the user profiles can be managed through a web interface. On an example of the University Infoline service, we present a category of new enhanced "click-to-call" services, which allow a user browsing through the web pages not only to establish a SIP call, but which enable, via the service logic, to handle different states of a call based on the preferences of calling and called parties. The service is simply integrable to any type of voice-enhanced e-commerce services.

3. Interworking between IP-based network signalling protocols and signalling protocols used in traditional SCNs (Switched Communication Networks)

Historical operators have widely deployed SCN (or TDM) networks over decades for providing basic call services. With the popularity of the Internet and the low price of IP-oriented equipment, a new orientation for basic call has emerged: IP telephony. Even though SIP is not limited to IP telephony, SIP has raised big interest in telcos and telecom equipment manufacturers research and development laboratories. Two worlds for basic call now coexist: TDM networks and IP-oriented networks. If a subscriber in a world wants to communicate with a subscriber in the other world, the two worlds need to interwork.

For customers two kinds of SIP phones exist: soft-phone and hard-phone. Soft-phone runs on operating system (such as Windows XP) of a computer connected to the Internet. Hard-phone implements SIP stacks on phone-like equipment that is connected to an Ethernet plug.

A customer using SIP phones cannot limit his phone calls to other SIP-based users, interest is in interworking with other telephone technology such as PSTN, GSM and other IP technology such as H.323. Anyhow final destination for a phone call cannot be pre-determined when using SIP technology (when interworking gateways are implemented in the network).

With the emergence of the Internet, a new kind of network architecture has appeared: Next Generation Networks (NGNs). A NGN is built on packet networks in order to transport both data and voice. The convergence of data and voice over the same typology of network (packet) will decrease investments by telcos. There are many definitions of NGN, however, two of the principal characteristics of the NGN are usually common:

- it is be based on a unique and shared packet-based network (IP, ATM, ...)
- transport layer and control layer are independent.

This brings also the necessity of the separation of control functions for bearer control, call control and service control functions and leads to:

- replacement of the traditional switching systems by new types of functional elements at the control layer
- appearance of new signalling protocols between different functional elements.

The full NGN architecture comprises network elements needed for the provision of traditional telephony services and advanced next-generation applications [Eurescom Project P1109: Next Generation Networks: the service offering standpoint]. They include Media Gateway, Signalling Gateway and Call Controller (Media Gateway Controller)

Media Gateway (MG) converts media and framing protocols provided in one type of network to the format required in another type of network. The Media Gateway terminates the bearer control protocols and contains bearer terminations. It also contains media manipulation equipment (e.g., transcoders, echo cancellers, or tone senders).

The MG should also provide a Signalling Gateway function when it terminates call signalling from access networks and has to relay it to the Call Server. The Media Gateway may therefore provide terminations such as Q.931 on the PSTN's side and IUA (ISDN User Adaptation layer) over SCTP/IP on the NGN side.

The Media stream policy/classification functions (e.g. control of traffic contracts: peak cell rate, average cell rate etc. in ATM, QoS class / priority management for Diffserv / RSVP in IP) are also provided by the Media Gateway.

Signalling Gateway (SG) is the network entity responsible to forward call control signalling by converting the transport mechanism of the incoming signalling to an appropriate ongoing transport mechanism (e.g., SS7 call signalling over MTP onto SS7 over IP).

In order to provide this function the Signalling Gateway may exist as a separate physical entity or reside within the same physical entity that support Media Gateway function depending on the

particular network scenario. In the case that the Call server terminates call signalling from access network, the Signalling Gateway function is useless.

Call Server, or Media Gateway Controller (MGC) provides basic call control including call routing (routing tables, address translations between different numbering plan formats, routing information retrieval from external devices), call signalling process (SIP, H.323, ISUP, MGCP, etc.) and H.248-like Media Gateway Controller functions.

It should also provide more advanced call control functions like third party call control and CLASS services.

In addition, it must provide standard and open interfaces towards Application Servers to enable call related event triggering, service and policy control (e.g., personalized QoS policies, AAA policies, etc.).

In the NGN architecture, the Call Server presents the central node that supports the intelligence of communication.

Examples of Call servers are call agents, softswitches, SIP Server and H.323 gatekeepers.

Thanks to decomposition of the network functions into layers the physical implementation and geographical localization of the functional elements mentioned above is left on the choice of constructors and network operators. This also allows to optimize the network resources and to dimension them separately.

Call control protocols

Call control protocols allow the establishment, in general based on the user demand, the communication between two terminals or between a terminal and a server. The two candidate protocols are H.323 and SIP.

Media Gateway control protocols

The necessity to interconnect the traditional telephone networks into NGN as well as the flexibility allowed by the separation of the NGN transport and control layers have lead to the distinction of the functions of the Media Gateway and Call Server. This was the reason for developing a protocol, which allows the Call server to control the MG. In fact, it is a protocol, which allows for co-ordination between the transport and control layers. So far, two protocols for controlling the MG have been proposed: MGCP (Media Gateway Protocol) and MEGACO/H.248. These are relatively low-level device-control protocols that instruct an MG to connect streams coming from outside a packet network onto a packet stream such as the RTP (Real-Time Transport Protocol). Both protocols follow the master/slave approach based on the distribution of the network gateway functions into “more intelligent” (master) “and less-intelligent” (slave) parts. This approach enables centralization of application intelligence in relatively fewer control servers (MGCs in the MGCP and MEGACO/H.248 terminology) on the one hand and highly-cost and performance optimised gateway devices (MGs) on the other hand.

Signalling protocols between Call Servers (MGCs)

There are two types of signalling protocols that might be used between MGCs:

- at the level of a core network (BICC, SIP-T, H.323)
- for the interconnection with the existing PSTN/ISDN networks, via the transport of PSTN/ISDN signalling over an IP-based network (SIGTRAN).

The SIGTRAN (Signalling Transport) protocol suite was developed within the IETF to allow an interworking between SS7 network elements and IP-based elements. Its primary purpose is to address the transport of PSTN/ISDN signalling over IP networks, taking into account the functional and performance requirements of the PSTN/ISDN signalling. To interwork with the PSTN/ISDN, IP-based networks need to transport signalling such as DSS1 (Digital Subscriber System), or SS7 (e.g., ISUP, SCCP) messages between IP nodes such a SG (Signalling Gateway), a MGC (Media Gateway Controller), and a (MG) Media Gateway.

The architecture that was defined by SIGTRAN WG consists of three components:

- A standard IP
- A common signalling transport protocol – a protocol that supports a common set of reliable transport functions for signalling transport. A SCTP (Stream Control Transmission Protocol) has been defined for this purpose by the IETF and endorsed by the ETSI.
- An adaptation sub-layer that supports specific primitives, such as management indications, required by a particular signalling application protocol. The adaptation layers are different for different splits of the protocol. A number of different adaptation layers are being developed by the SIGTRAN group:
  - IUA: The boundary is Q.921/Q.931 [2]
  - M2UA: The boundary is MTP2/MTP3 [3]
  - M2PA: The boundary is MTP2/MTP3, but for a symmetric scenario [4]
  - M3UA: The boundary is MTP3/user part [5]
  - SUA: The boundary is SCCP/SCCP user [6]

The adaptation layers mentioned above have been recently endorsed by the ETSI TC SPAN.

As mentioned above, the Session Initiation Protocol (SIP) is a “hot” candidate for being a call control protocol for new IP-oriented services. Even though SIP is not yet widely deployed in public networks, when this case occurs SIP needs to interwork with call control protocols in existing SCN networks such as ISUP, DSS1, or INAP. Within the EURESCOM project P1111 N-GOSSIP [1], the Department of InfoCom Networks from the University of Žilina was involved in, scenarios for SIP interworking with existing signalling protocols (ISUP, INAP, DSS1) were analysed.

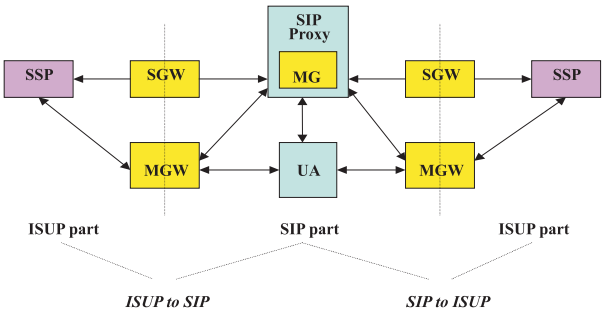
Voice calls do not always have to originate and terminate in the PSTN (via MGCs). They may either originate and/or terminate in SIP phones. The alternatives for call origination and termination suggest the following possibilities for calls that traverse through an IP network:

1. PSTN origination – IP bridging – PSTN termination
   It is the situation in which a SIP network connects two instances of the telephone network. A telephone call originates in the PSTN and a SS7 ISUP message is dispatched to the MGC,

which is the point of interconnection with the PSTN network. In this case, the MGC is the point of origination for message flows for this call over the IP network. The call is passed through proxies that route the call to the appropriate PSTN interface. The MGC that interconnects to the PSTN at the output of the IP network is the point of termination of the IP message flow. This MGC then uses ISUP to communicate with the PSTN at the terminating end.
The role of the SIP in the IP network is:
- to determine the appropriate point of termination
- to establish a session between the points of origination and termination in order to carry the call through the IP network.

2. PSTN origination – IP termination
This is a case when a call originates from the PSTN and terminates at a SIP phone.
3. IP origination – PSTN termination
A call originates from a SIP phone and terminates in the PSTN. There is no telephony interface at call origination.
4. IP origination – IP termination: this is a case of pure SIP, with no PSTN involvement.
Thus, from a functional point of view, there are three distinct elements in a SIP VoIP network offering PSTN inter-connection:
- The originator of SIP signalling
- The terminator of SIP signalling
- The network of SIP proxies that routes calls from the originator to the terminator.

Based on this, the general SIP-ISUP interworking architecture was proposed as depicted in the following Figure:



Note: there might be no SGW between the ISUP part and SIP part. In that case it is the MGC that represents the interface between the ISUP and SIP parts.

Within this task, the protocol interworking issues including SIP-ISUP and ISUP-SIP issues were studied as well [1].

4. Impact of packet loss and jitter to the quality of a voice transmission

If the transmitted analogue signal has changed its shape for any reason during its transmission over the network, the difference between a received signal and transmitted signal can be assumed as noise. To explain the main idea of this research project part, we will use only the second order characteristic Signal-to-Noise Ratio (SNR), to argue for its use in the network design and management. For simplicity we do not present results for voice, but just for test signal given by the band-limited noise with constant power spectra within this band.

In the sampling theorem the samples are taken regularly, but due to a random delay in the packet switched network the samples are taken irregularly by the reconstruction procedure. The received (reconstructed) signal will be

$$r(t) = \sum_k s(k\Delta)\Phi(t - k\Delta - \tau_k),$$

Let us suppose that this irregularity expressed by differences (delays) from regular sampling points, creates the point process $\{\tau_k, k \in Z\}$. Let $\{\tau_k, k \in Z\}$ be an ergodic, stationary, random point process with the characteristic function

$$G(\omega) = E[e^{j\omega\tau_k}],\ \omega \in \Re,\ k \in Z.$$

If $s(k\Delta)$ are pairs of independent variables and test signal has the constant power density then the SNR defined by the ratio of signal average power and noise average power in logarithmic scale

$$SNR = 10\log \frac{\sigma_S^2}{\sigma_N^2}, \text{ [dB]}$$

becomes

$$SNR = -10\log 2\left(1 - \frac{1}{2\Omega}\int_{-\Omega}^{\Omega} ReG(\omega)d\omega\right), \text{ [dB]}$$

Differences between transmitted signal and received signal in which lost samples were replaced by sample approximations, were studied as noise and SNR was used as the objective measure by Jayant and Christensen. More general study of the noise properties, when the lost sample is replaced by a linear combination of other available samples, was obtained in our research. We restrict explanation to the simplest case again. Only zero stuffing (lost sample is replaced by zero) is assumed, and only SNR is used for the noise description. Under these assumptions the noise caused by lost samples (or packets) gives Signal-to-noise ratio expressed in dB

$$SNR = -10\log(p), \text{ [dB]}$$

where p is packet (sample) loss probability.

It should be remarked that the analysis of sample loss is easier than the analysis of jitter and using additional noise properties, for example, psofometric weighting is possible.

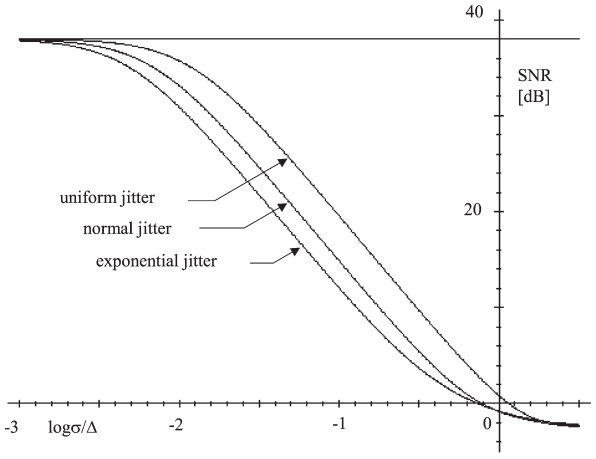
Until now, the jitter or packet loss was taken as the only source of noise. The approach, in which the impact of the voice distortion sources are given by SNR, allows to calculate the impact of more independent noise sources. Let $n_1(t)$, ..., $n_M(t)$, $t \in \Re$ be independent noises with average powers σ_{N1}^2, ..., σ_{NM}^2. Then the total signal-to-noise ratio is

$$SNR = 10 \log \frac{\sigma_S^2}{\sum_{i=1}^{M} \sigma_{Ni}^2} = -10 \log \sum_{i=1}^{M} 10^{-\frac{SNRi}{10}},$$

where $SNRi = 10\log \frac{\sigma_S^2}{\sigma_{Ni}^2}$, $i = 1, 2, ..., M$. As an example the speech signal is taken with 8-bit □-law ideal quantizing, which produces the noise of SNR=38dB. When this signal is also attacked by the independent jitter with exponential distribution, the total SNR is approximately

$$SNR \approx -10 \log 2\left(1.000079 - \frac{\Delta}{\pi\tau} arctg - \frac{\pi\tau}{\Delta} \right), \text{ [dB]}.$$

A similar function can be obtained for the jitter with uniform and normal distribution, as shown in the following figure.



It is recommended for multiplicative noise to use "signal to signal-correlated-noise ratio", called Q when expressed in decibels. Taken into account subjective factors, it is recommended to combine Q values on a 15log10 basis, i.e.

$$Q = -15\log \sum_{i=1}^{M} 10^{-\frac{Q_i}{15}}$$

In the previous part we assumed that only distortion sources are present and no tool for their compensation is available. These tools are: lost sample/packet replacing techniques and jitter buffer. Lost sample/packet replacing methods are well known from other applications (signal processing, forward error correction codes), but jitter buffer is a new element in telephone network and should be studied carefully. If delay was not a very serious problem in circuit switched telephone networks, compression/decompression, coding/decoding, packetisation and especially waiting in internal network buffers are additional sources of delay. We have no compensator of lost time until now, then it is on the network designers to calculate all delays very carefully.

The total delay occuring in the transmission chain can be divided to two parts: the fixed part of delay and the variable part of the delay – jitter J. If the sum of these delays exceeds a given limit, the delayed packet/sample is discarded by the network i.e. it is lost. To eliminate sample jitter, jitter buffers (play-out buffers) are used. In the case of nonempty jitter buffer, samples are waiting in the queue, and they are played-out in regular intervals. If the buffer is empty when a sample should be sent, a signal gap occurs until the delayed sample is available (no recovery of lost sample is assumed). This residual delay produces again noise, and the previous results may be applied. It is not our aim to study play-out buffers behaviour in its completeness in this contribution. We just take as an example the case, when the delays are independent, i.e. when gaps create a Bernoulli process. Let p_0 be the probability that a buffer is empty when the sample should be played out, and the distribution of the residual delay i.e. the interval since the sample should be played out is known. Then,

$$SNR = SNR_{res} - 10 \log p_0$$

This formula shows, how the waiting for the delayed sample improves SNR, compared to the policy when the empty buffer generates zero samples, but just in time. There exist several methods how to improve this result. Firstly, other lost sample replacing methods bring an advantage, but of course, only if their gain in SNR, compared to the zero stuffing is greater than SNRres. Secondly, probability of an empty buffer p_0 can be decreased if samples are played out slower than the sampling rate, when the buffer is coming to be empty. The static control policy of play-out sample rate can be used, i.e. $\Delta_n = \Delta + \delta_n$, $n = 1, 2, ...$ inter-sample interval is used, when the sample should be played out, and the buffer contains n samples. See research reports for details.

In this presentation of the research project we assume static control policy of playout sample rate, i.e. we assume that $\Delta_n = \Delta + \delta_n$, $n = 1, 2, ...$ intersample interval is used, when the sample should be played out, and the buffer contains n samples (infinite buffer capacity is assumed). If the buffer is empty at the playing out instant, the sample will be played out as soon as it comes, i.e. intersample interval $\Delta_1 + \tau_{res}$ occurs. The marginal characteristic function of delay is

$$G(\omega) = E\{e^{j\omega\tau}\} = \sum_{n=1}^{\infty} p_n e^{j\omega\delta_n} + p_0 E\{e^{j\omega\tau_{res}}\}, \omega \in \Re$$

and the noise power spectral density, is

$$\hat{n}(\omega) = 2\sigma_S^2 \hat{\Phi}(\omega)\left[1 - p_0 - \sum_{n=1}^{\infty} p_n \cos \omega\delta_n \right] + p_0 \hat{n}(\omega)_{res},$$

$\omega \in \Re$.

The average noise power is

$$\sigma_N^2 = \frac{2\sigma_S^2}{\Delta}\left[1 - p_0 - \sum_{n=1}^{\infty} p_n \frac{sin\pi\epsilon_n}{\pi\epsilon_n} \right] + p_0 \sigma_{N\,res}^2,$$

where $\epsilon_n = \frac{\delta_n}{\Delta}$, $n = 1, 2, ...$ is a relative sample delay. The previous formula can be used for SNR calculation.

ITU-T has proposed a complex method, known as E-model, for voice transmission quality evaluation. All basic network performance parameters of a telephone network are considered. If the voice is transmitted over the packet switched network (IP or ATM), new network performance parameters may increase degradation of the voice quality, the most important being packet loss and rest jitter. There are several approaches to evaluating the impact of packet loss on the transmission quality. Two of them were taken into account. The first one (proposed by ITU-T Rec. G.113 Annex 1) incorporates packet loss to the equipment impairment factor, and the second (described in the paper) incorporates packet loss and rest jitter to the simultaneous impairment factor Is. These models give quite a different evaluation and should be improved. The second approach, used in our research, incorporates packet loss and rest jitter to the simultaneous impairment factor Ie. This approach allowed to combine more degradation factors and respected the fact, that influences of jitter and packet loss are not additive.

5. Conclusions

In this contribution we focused on some topical problems of IP Telephony. Having used an example of the "University Infoline: service, we demonstrated the capabilities of new tools for the development of services that advance the service of IP telephony at a qualitatively higher level (SIP and HTTP integration SIP s HTTP). Another problem that is necessary to be solved is interworking of existing network types and terminals with IP-based networks and new types of terminals for IP telephony. In the frame of the Quality of Service (QoS) for IP telephony we proposed analytical models for the noise caused by packed loss and jitter. Influence of packet loss and jitter was incorporated to the E-model introduced by ETSI and ITU-T for an evaluation of voice transmission quality in telephone networks.

References

[1] N-GOSSIP: *Next-Gen open Service Solutions over IP,* http://www.eurescom.de/public/projects/P1100-series/P1111/default.asp],
[2] draft-ietf-sigtran-rfc3057bis: http://www.ietf.org
[3] RFC 3331: http://www.ietf.org
[4] draft-ietf-sigtran-m2pa: http://www.ietf.org
[5] RFC 3332: http://www.ietf.org
[6] draft-ietf-sigtran-sua: http://www.ietf.org

https://doi.org/10.26552/com.C.2004.4.71-73

Vladimír Hottmar – Michal Kuba *

THE MICROCOMPUTER CONVERTER OF THE TELEX SIGNAL

This paper deals with a structure and a description of the microcomputer converter of the telex signal (MCTS) that we developed for the Slovak Telecom. The microcomputer converter was developed within the International contract between Hungary Telecommunication Association Matav, Slovak Telecom in Bratislava, and Department of Telecommunications at the University of Zilina. We made the following tasks: design, debugging and testing of the resident program of the converter. After evaluation of the prototype by the company Matav 150 pieces of the converters, were manufactured and then successfully implemented in the TELEX exchange in Slovak Telecom in Bratislava for continuous operation.

1. An implementation of the MCTS modules into the system of the telex signalling

The microcomputer converter of telex signal emulates the telex subscriber signalling. The module consists of two identical single-chip microcomputers and supporting peripheries, which enable their implementation into the system of the telex signalling. Two subscribers of the telex operation are connected to one module, which consists of two converters. The system enables to connect up to 150 subscribers of the telex network. The modules MCTS realize the conversion from the signalling of the telex terminals that are located in Hungary, to the signalling of the telex exchange AXB20 situated in Bratislava. The conversion of the signalling is realized in the time and character domain. MCTS manages the signalling for the calling subscribers (the outgoing direction) and the called subscribers (the incoming direction) in different ways. It also manages the signalling of successful national and international calls as well as unsuccessful calls. It handles timing and ordering of the character sequences in both directions: "subscriber – telex exchange" and "telex exchange – subscriber" for all types of the calls.

The telex exchange AXB20, the transmission system of the telex signals between Budapest and ST Bratislava and the microcomputer converter of the telex signalling are shown in Fig. 1. It is seen from Figure 1 that through the microcomputer converter telex terminals in Hungary are connected to the telex exchange AXB20, by which they are controlled.

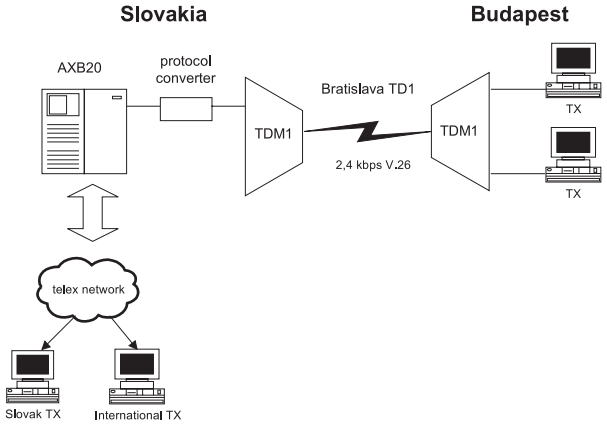
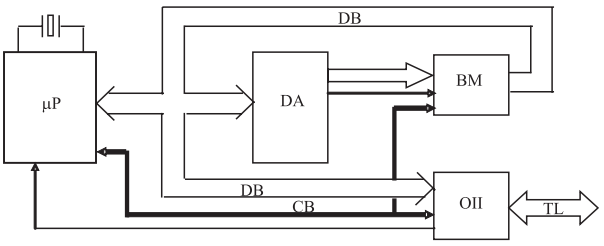


Fig. 1 Telex transmission system with the converter

2. Architecture of the MCTS module

An architecture of the microcomputer converter of the telex signals is based on the single-chip microcomputer of 8051 family and its supporting peripheries (See Fig. 2). The storage space consists of the fixed program memory EPROM with the total capacity of 16kB and the data memory of 2kB. The circuits of the system interface are optically isolated. The optical isolation interface block OII consists of isolation optoelements as well as circuits of the combination logic and the power elements. They form the electrical interface for connecting MCTS to the transmission system of the telex exchange AXB20 and the interface TDM1.



Legend: μP – the single-chip microcomputer, DA – the decoder of the address, BM – the block of the memories (RAM, EPROM), OII – the optical isolated interface, DB – the data bus, CB – the control bus, TL – the telex line

Fig. 2 The microcomputer converter of telex signal

* **Vladimír Hottmar, Michal Kuba**
Department of Telecommunications, Faculty of Electrical Engineering, University of Žilina, Žilina, Slovakia,
E-mail: vladimir.hottmar@fel.utc.sk, michal.kuba@fel.utc.sk

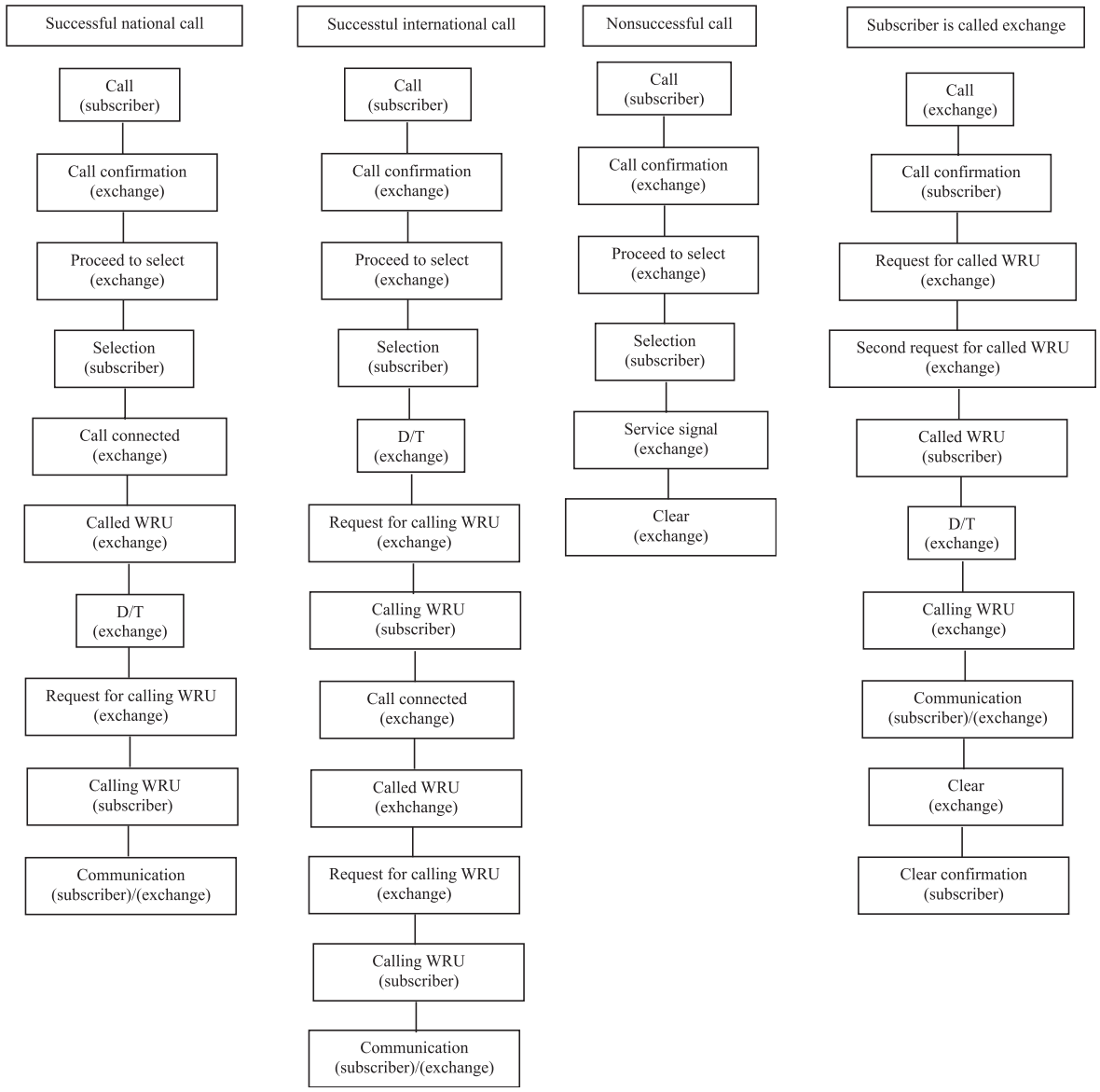


Fig. 3 The signalling procedures preformed by the converter

3. Converter activity description

The microcomputer converter controls the telex signalling and the converter`s activity is initialized by calling the subscriber, or by the exchange when the subscriber is called. The converter performs several types of signalling (See Figure 3) in dependence on the type of call as follows:

a) The successful national call, when the subscriber is calling. It represents the outgoing direction.
b) The successful international call, when the subscriber is calling. It represents the outgoing direction.
c) The unsuccessful call of the subscriber. The connection is not established and the exchange sends the terminal an appropriate service signal to log off the connection.
d) The subscriber is called by the exchange for the national and international calls.

4. Conclusion

Hungary Telecommunication Association Matav asked the Slovak Telecom to modify their telex signal (TS). For this purpose we developed the microcomputer converter of the telex signal (MCTS). The converter was developed within the International contract between Hungary Telecommunication Association Matav, Slovak Telecom in Bratislava, and the Department of Telecommunications of the University of Zilina. After evaluation of the prototype by the company Matav, 150 pieces of the converters

were manufactured and then successfully implemented in the TELEX exchange in Slovak Telecom in Bratislava for continuous operation. Currently the system has been working for one and half year without failures.

References:

[1] Technical specification, requirements based on the Hungarian attributes.
[2] Transmission system SZT 2400.
[3] Test minutes of ST telex model, Matav, Budapest, 28 November 2001.

https://doi.org/10.26552/com.C.2004.4.74-79

kamUNikácie
COMMUNICATIONS

Juraj Smieško *

HYPEREXPONENTIAL MODEL OF TOKEN BUCKET SYSTEM

In the present paper we apply the Theory of Markov Chains for a formal description of a hyperexponential model of Token Bucket System (TBS). This paper extends the topic covered in [3] by replacing Exponential model of human speech by a more precise hyperexponential model.

In the first section we deal with analysis of the Hyperexponential model of On-Off Source of human speech. In the second section we deal with steady-state analysis of Token Bucket system and we solve the balance equations for probabilities of states. In the third section we use real parameters of VoIP and we calculate characteristic of TBS and in the last section the values of probability of packet loss are approximated by an exponential regression function. We can directly compute values of probability of packet loss using this exponential function.

1. Hyperexponential model of On-Off Source

The On-Off Source Model of human speech is composed of two periods, where On period belongs to the active periods in human speech and the Off period belongs to the silence. There are two main states On and Off and the source switches from one state to another.

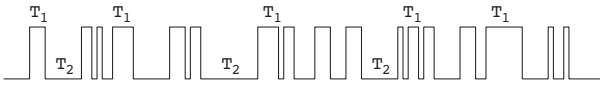


Fig. 1 On-Off model of human speech

Talk-spurt duration is modelled by the variable T_1 and pause duration is modelled by the variable T_2. The probability densities of T_1 and T_2 are modelled by two weighted geometric distribution functions.

Every increment of these variables T_1 and T_2 is equal to 5 ms and then the average talk-spurt duration is $ET_1 = 227$ms and the average pause duration is $ET_2 = 596$ms (see [1]).

$$T_1 \sim f_1(k) = c_1(1 - U_1)U_1^{k-1} + c_2(1 - U_2)U_2^{k-1}$$

$c_1 = 0.60278$ $\quad c_2 = 0.39817$
$U_1 = 0.92446$ $\quad U_2 = 0.98916$

$$T_2 \sim f_2(k) = d_1(1-W1)W_1^{k-1} + d_2(1 - W_2)W_2^{k-1}$$

$d_1 = 0.76693$ $\quad d_2 = 0.23307$
$W_1 = 0.89700$ $\quad W_2 = 0.99791$
for $k = 0, 1, 2, \ldots$.

The best approximation of these processes is the Hyperexponential model. The discrete geometrical distributions are approximated by continuous exponential ones and the rest of this model has completely the same nature as the geometrical one. In the new model we can keep the same averages.

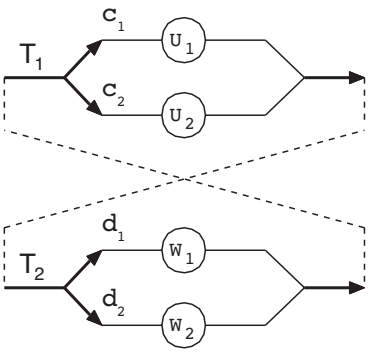


Fig. 2 Hyperexponential On-Off Source

ON: $T_1 \sim f_1(t) = c_1\alpha_1 e^{-\alpha_1 t} + c_2\alpha_2 e^{-\alpha_2 t}$
$c_1 = 0.60278$ $\quad c_2 = 0.39817$ $\quad \alpha_1 = 0.01511$ $\quad \alpha_2 = 0.00217$

OFF: $T_2 \sim f_2(t) = d_1\mu_1 e^{-\mu_1 t} + d_2\mu_2 e^{-\mu_2 t}$
$d_1 = 0.76693$ $\quad d_2 = 0.23307$ $\quad \mu_1 = 0.02060$ $\quad \mu_2 = 0.00042$

Fig. 3 and Fig. 4 show approximation of hypergeometrical distributions (points) by hyperexponential distributions for On and Off periods.

We construct a transmission diagram of Markov model of On-Off Source (loops are omitted):

The rate matrix of this Markovov chain is

$$Q_0 = \begin{pmatrix} -\alpha_1 & 0 & d_1\alpha_1 & d_2\alpha_1 \\ 0 & -\alpha_2 & d_1\alpha_2 & d_2\alpha_2 \\ c_1\mu_1 & c_2\mu_1 & \mu_1 & 0 \\ c_1\mu_2 & c_2\mu_2 & 0 & \mu_2 \end{pmatrix}$$

* **Juraj Smieško**
Department of Mathematical Methods, Faculty of Management Science and Informatics, University of Žlina

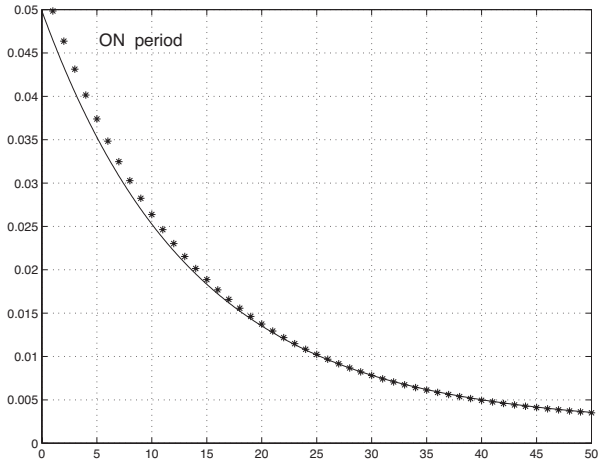


Fig. 3 Approximation for On period

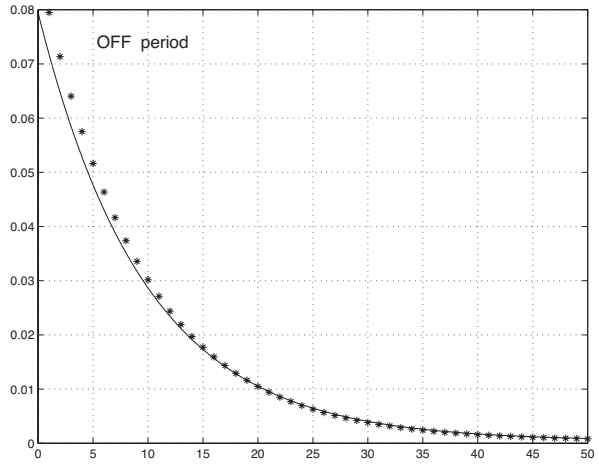


Fig. 4: Approximation for Off period

We solve the balance equations $\pi \,.\, Q_0 = 0$, where $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$:

$$\pi_2 = \frac{c_2\alpha_1}{c_1\alpha_1}\,\pi_1 \qquad \pi_3 = \frac{d_1\alpha_1}{c_1\mu_1}\,\pi_1 \qquad \pi_4 = \frac{d_2\alpha_1}{c_1\mu_2}\,\pi_1$$

We compute the probability π_1 from the normalisation condition $\sum_{i=1}^{4} \pi_i = 1$

$$\pi_1 = \frac{c_1\alpha_2\mu_1\mu_2}{(c_1\alpha_2 + c_2\alpha_1)\mu_1\mu_2 + (d_1\mu_2 + d_2\mu_1)\alpha_1\alpha_2}$$

The probabilities for real parameters of VoIP are $\pi = (0.04902, 0.22546, 0.04579, 0.67973)$.

Let P_{ON} and P_{OFF} be probabilities that the Source is in state On or Off:

$$P_{ON} = \pi_1 + \pi_2 = 0.27448 \qquad P_{OFF} = \pi_3 + \pi_4 = 0.72552$$

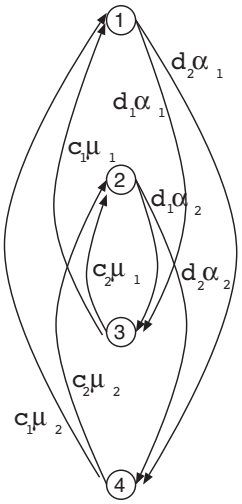


Fig. 5: Markov chain for hyperexponential model

Notice that these probabilities are very similar to those in the Exponential model (see [3], $P_{ON} = 0.27586$ and $P_{OFF} = 0.72414$). But in the Hyperexponential model we have a more detailed structure of On-Off source than in the Exponential model.

2. Steady-state Analysis of Token Bucket System

The Token Bucket System is related to VoIP problems. We have formed the steady-state analysis of VoIP under the Token Bucket Control. Our main problems will be to compute the probability characteristics of TBS and to find relation between probability of packet loss and bucket depth. At first we will analyze the working of TBS in general.

There is a flow of packets entering the TBS. The Token Bucket System is generating tokens (marks) and then it marks each packet with one of them. Only the marked packets will go through network. For practical reasons we can assume a limited bucket with depth "n". When the bucket is empty (there are no tokens), the TBS cannot mark any packet, which is lost then. Let P_{lst} be probabilities of this random phenomenon and we will call it "probability of packet loss".

We will deal with the Hyperexponential model of On-Off Source of human speech (see above). When the Source is in On state it starts generating flow of packets, which represents human speech. The packet rate in usual VoIP systems (using G.729A) is 50 p/s. We will assume that the flow of packets is modelled by Poisson process $N_p(t)$ with rate $\eta = 0.05\ p/ms$.

In general there can be any token rate. Usually it is the same as an average number of packets entering the TBS. The flow of tokens will be modelled by Poisson process $N_p(t)$ with a rate λ:

$\lambda = \eta \,.\, P_{ON} + 0 \,.\, P_{OFF} = 50\ p/s \,.\, 0.27448 = 13.724\ p/s =$
$= 0.01372\ p/ms$

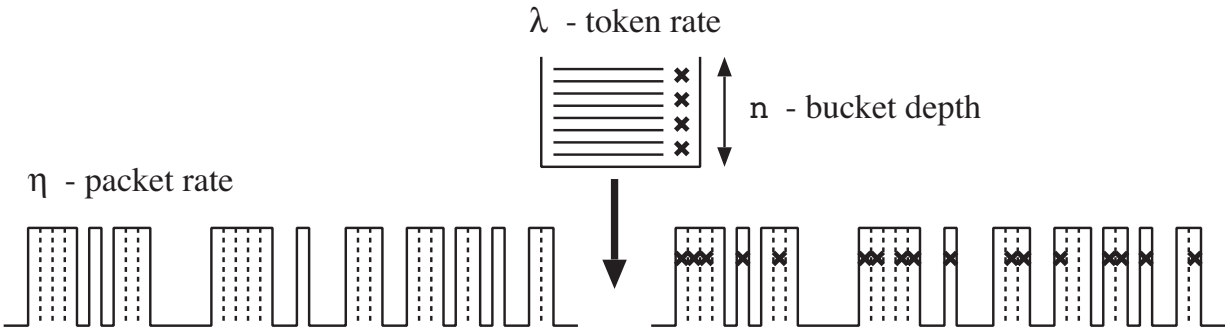


Fig. 6: Token Bucket system

In the Token Bucket System we have three elementary random phenomena entering the packet, generating token and switching between On-Off states in Source. We have assumed that each of them is Poisson process, and because of that we can model this TBS by Markov chain.

Probability of an increasing number of tokens in the bucket by one during "short" time Δt is $P(N_T(\Delta t) = 1) = \lambda\Delta t + o(\Delta t)$. If the bucket contains k tokens, probability of a decreasing number of tokens in the bucket by one during "short" time Δt is $P(N_P(\Delta t) = 1) = \eta\Delta t + o(\Delta t)$. If there is an arriving packet and the bucket is empty, we cannot mark it and this packet is lost. Probability of packet loss was named P_{lst}.

We will construct a transition diagram of the whole system. Component columns of the diagram refer to a number of tokens in the token bucket. We assigned this as "k-level" for $k = 0, \ldots n$. For easy reading loops in the transition diagram are omitted:

Let X_k and Y_k be probabilities that in the TBS are k tokens and Source is in On states and let Z_k and W_k be probabilities that in the TBS are also k tokens, but Source is in Off states. Let P_k be a probability of k-level. We see that $P_k = X_k + Y_k + Z_k + W_k$ and $P_{lst} = X_0 + Y_0$. Taking from the Theory of Markov Chains [2] we can write the following balance equations for state probabilities of steady-state Markov chain:

for $k = 0$

$$0 = -(\lambda + \alpha_1)X_0 + c_1\mu_1 Z_0 + c_1\mu_2 W_0 + \eta X_1$$
$$0 = -(\lambda + \alpha_2)Y_0 + c_2\mu_1 Z_0 + c_2\mu_2 W_0 + \eta Y_1$$
$$0 = d_1\alpha_1 X_0 + d_1\alpha_2 Y_0 - (\lambda + \mu_1)Z_0$$
$$0 = d_2\alpha_1 X_0 + d_2\alpha_2 Y_0 - (\lambda + \mu_2)W_0$$

for $k = 1, \ldots, n-1$

$$0 = \lambda X_{k-1} - (\lambda + \alpha_1 + \eta)X_k + c_1\mu_1 Z_k + c_1\mu_2 W_k + \eta X_{k+1}$$
$$0 = \lambda Y_{k-1} - (\lambda + \alpha_2 + \eta)Y_k + c_2\mu_1 Z_k + c_2\mu_2 W_k + \eta Y_{k+1}$$
$$0 = \lambda Z_{k-1} + d_1\alpha_1 X_1 + d_1\alpha_2 Y_k - (\lambda + \mu_1)Z_k$$
$$0 = \lambda W_{k-1} + d_2\alpha_1 X_1 + d_2\alpha_2 Y_k - (\lambda + \mu_2)W_k$$

for $k = n$

$$0 = \lambda X_{k-1} - (\alpha_1 + \eta)X_n + c_1\mu_1 Z_n + c_1\mu_2 W_n$$
$$0 = \lambda X_{k-1} - (\alpha_2 + \eta)Y_n + c_2\mu_1 Z_n + c_2\mu_2 W_n$$
$$0 = \lambda Z_{k-1} + d_1\alpha_1 X_n + d_1\alpha_2 y_n - \mu_1 Z_n$$
$$0 = \lambda W_{k-1} + d_2\alpha_1 X_n + d_2\alpha_2 Y_n - \mu_2 W_n$$

We want to solve the balance equations $p \,.\, Q = 0$, where $p = (X_0, Y_0, Z_0, W_0 \ldots X_n, Y_n, Z_n, W_n)$. The rate matrix Q is more complicated, so we will write it to the block matrix:

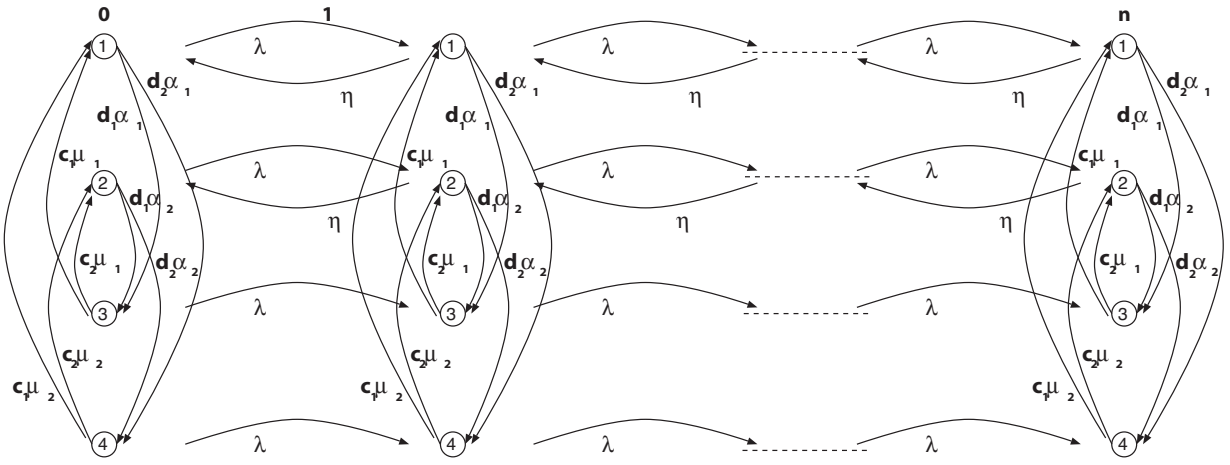


Fig. 7: Markov chain for Hyperexponencial TBS

$$Q = \begin{pmatrix} A_0 & \Lambda & 0 & \dots & 0 & 0 & 0 \\ \Gamma & A & \Lambda & \dots & 0 & 0 & 0 \\ 0 & \Gamma & A & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A & \Lambda & 0 \\ 0 & 0 & 0 & \dots & \Gamma & A & \Lambda \\ 0 & 0 & 0 & \dots & 0 & \Gamma & A_v \end{pmatrix} = \begin{pmatrix} A+\Gamma & \Lambda & 0 & \dots & 0 & 0 & 0 \\ \Gamma & A & \Lambda & \dots & 0 & 0 & 0 \\ 0 & \Gamma & A & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A & \Lambda & 0 \\ 0 & 0 & 0 & \dots & \Gamma & A & \Lambda \\ 0 & 0 & 0 & \dots & 0 & \Gamma & A+\Lambda \end{pmatrix}$$

$$A_0 = \begin{pmatrix} -(\lambda+\alpha_1) & 0 & d_1\alpha_1 & d_2\alpha_1 \\ 0 & -(\lambda+\alpha_2) & d_1\alpha_2 & d_2\alpha_2 \\ c_1\mu_1 & c_2\mu_1 & -(\lambda+\mu_1) & 0 \\ c_1\mu_2 & c_2\mu_2 & 0 & -(\lambda+\mu_2) \end{pmatrix} \qquad A_0 = \begin{pmatrix} -(\alpha_1+\eta) & 0 & d_1\alpha_1 & d_2\alpha_1 \\ 0 & -(\alpha_2+\eta) & d_1\alpha_2 & d_2\alpha_2 \\ c_1\mu_1 & c_2\mu_1 & -\mu_1 & 0 \\ c_1\mu_2 & c_2\mu_2 & 0 & -\mu_2 \end{pmatrix}$$

$$A = \begin{pmatrix} -(\lambda+\alpha_1+\eta) & 0 & d_1\alpha_1 & d_2\alpha_1 \\ 0 & -(\lambda+\alpha_2+\eta) & d_1\alpha_2 & d_2\alpha_2 \\ c_1\mu_1 & c_2\mu_1 & -(\lambda+\mu_1) & 0 \\ c_1\mu_2 & c_2\mu_2 & 0 & -(\lambda+\mu_2) \end{pmatrix} \qquad \Gamma = \begin{pmatrix} \eta & 0 & 0 & 0 \\ 0 & \eta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \Lambda = \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$$

We assigned Q_0 as rate matrix of the Hyperexponential On-Off Source. It can be easily seen that:

$$A_0 = A + \Gamma,\ A_n = A + \Lambda \text{ and } Q_0 = \Gamma + A + \Lambda$$

We can determine recurrent formulas for the probabilities W_i and Z_i:

$$\begin{aligned} &(\lambda+\mu_2)W_0 = d_2\alpha_2 Y_0 + d_2\alpha_1 X_0 \\ &(\lambda+\mu_2)W_k = d_2\alpha_2 Y_k + d_2\alpha_1 X_k + \lambda W_{k-1} \quad k = 1, ..., n-1 \\ &\mu_2 W_n = d_2\alpha_2 Y_n + d_2\alpha_1 X_n + \lambda W_{n-1} \end{aligned}$$

$$\begin{aligned} &(\lambda+\mu_1)Z_0 = d_1\alpha_2 Y_0 + d_1\alpha_1 X_0 \\ &(\lambda+\mu_1)Z_k = d_1\alpha_2 Y_k + d_2\alpha_1 X_k + \lambda Z_{k-1} \quad k = 1, ..., n-1 \\ &\mu_1 Z_n = d_1\alpha_2 Y_n + d_1\alpha_1 X_n + \lambda Z_{n-1} \end{aligned}$$

But the equations for probabilities Yi and Xi are very complicated:

$$\begin{aligned} &\eta_1 X_1 + c_1\mu_2 W_0 + c_1\mu_1 Z_0 = (\lambda+\alpha_1) X_0 \\ &\eta_1 X_{k+1} + c_1\mu_2 W_k + c_1\mu_1 Z_k + \eta Y_k = (\lambda+\alpha_1) X_k + \\ &+ \lambda W_{k-1} + \lambda Z_{k-1} + \lambda Y_{k-1} \qquad k = 1, ..., n-1 \\ &c_1\mu_2 W_n + c_1\mu_1 Z_n + \eta Y_n = \alpha_1 X_n + \lambda W_{n-1}\ \lambda Z_{n-1} + \lambda Y_{n-1} \end{aligned}$$

$$\begin{aligned} &\eta_1 X_k + \eta_1 X_k = \lambda W_{k-1} + \lambda Z_{k-1} + \lambda Y_{k-1} + \lambda X_{k-1} + \\ &k = 1, ..., n \end{aligned}$$

If we assign $p_k^{on} = X^k + Y_k$ and $p_k^{off} = Z_k + W_k$ we have these interesting relations:

$$p_k^{on} = \frac{\lambda}{\eta} \cdot p_{k-1}^{off} + \frac{\lambda}{\eta} p_{k-1}^{on} \qquad k = 1, ..., n$$

3. TBS with real parameters of VoIP

To deduce an explicit formula for probability of packet loss $P_{lst} = P_{lst}(n) = X_0 + Y_0$ is unreal, but there is no problem to compute values of P_{lst} for real parameters of our Token Bucket System with the Exponential On-Off Source:

$c_1 = 0.60278$	$c_2 = 0.39817$
$\alpha_1 = 0.01511\ \text{ms}^{-1}$	$\alpha_2 = 0.00217\ \text{ms}^{-1}$
$d_1 = 0.76693$	$d_2 = 0.23307$
$\mu_1 = 0.02060\ \text{ms}^{-1}$	$\mu_1 = 0.00042\ \text{ms}^{-1}$
$\lambda = 0.01372$ p/ms	$\eta = 0.05$ p/ms

For example the reader can see for himself the difference between $n = 4$ and $n = 5$:

n	X_k	Y_k	Z_k	W_k	P_k
0	0.0237	0.1507	0.0153	0.0113	0.2010
1	0.0098	0.0453	0.0116	0.0150	0.0817
2	0.0060	0.0164	0.0075	0.0167	0.0466
3	0.0048	0.0079	0.0050	0.0177	0.0354
4	0.0046	0.0051	0.0064	0.6191	0.6352

n	X_k	Y_k	Z_k	W_k	P_k
0	0.0222	0.1475	0.0146	0.0108	0.1951
1	0.0091	0.0444	0.0111	0.0143	0.0789
2	0.0054	0.0162	0.0070	0.0158	0.0444
3	0.0042	0.0080	0.0046	0.0167	0.0335
4	0.0040	0.0052	0.0034	0.0174	0.0300
5	0.0041	0.0041	0.0049	0.6047	0.6178

For real use it is enough to have the bucket depth $n = 1, ..., 10$. Now we will increase the bucket depth n and calculate characteristics of models:
$P_{lst}(n)$ – probability of packet loss or probability of an empty bucket in time of arriving packet
$\lambda \, . \, P_{lst}(n)$ – average number of lost packets
P_n – probability of full bucket (bucket will refuse tokens)
EK – average bucket depth
$H = EK/n$ – token bucket usage

n	$P_{lst}(n)$	$\lambda \, . \, P_{lst}(n)$	P_n	EK	χ
1	0.2079	2.852 *p/s*	0.7574	0.7574	75.5%
2	0.1892	2.596 *p/s*	0.6893	1.4696	73.5%
3	0.1803	2.474 *p/s*	0.6567	2.1543	71.8%
4	0.1744	2.393 *p/s*	0.6351	2.8218	70.6%
5	0.1697	2.328 *p/s*	0.6178	3.4779	69.6%
6	0.1655	2.271 *p/s*	0.6027	4.1256	68.8%
7	0.1617	2.219 *p/s*	0.5888	4.7664	68.1%
8	0.1582	2.170 *p/s*	0.5758	5.4015	67.5%
9	0.1548	2.124 *p/s*	0.5636	6.0314	67.0%
10	0.1517	2.081 *p/s*	0.5519	6.6566	66.6%

4. Relation between Probability of packet loss and Token Bucket depth

The most interesting characteristic is probability of packet loss $P_{lst}(n)$. In Fig.7 we compared the values of probability of packet loss $P_{lst}(n)$ for the Exponential model and Hyperexponential model (see [3]).

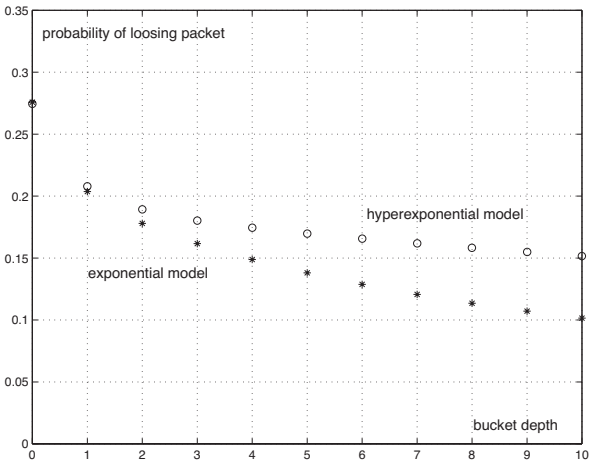


Fig. 8: Compared between exponential and hyperexponencial model

The approximation by hyperexponencial distributions is the best approximation and therefore the Hyperexponential Model of TBS is more accurate in modelling the real TBS than the Exponential model, but for computation of state probabilities we had to use numerical methods.

The approximation by exponential distributions is not so "good", but for this model we have gained recurrent formulas for state probabilities and values of characteristics that we get from the Exponential model can be used as lower estimation of characteristics of the real TBS.

Now we use approximation by the regression function $f(n)$ by the least squares method to find relation between Probability of packet loss $P_{lst}(n)$ and the Token Bucket depth n. The Regression function must satisfy these conditions to gain solid approximation:

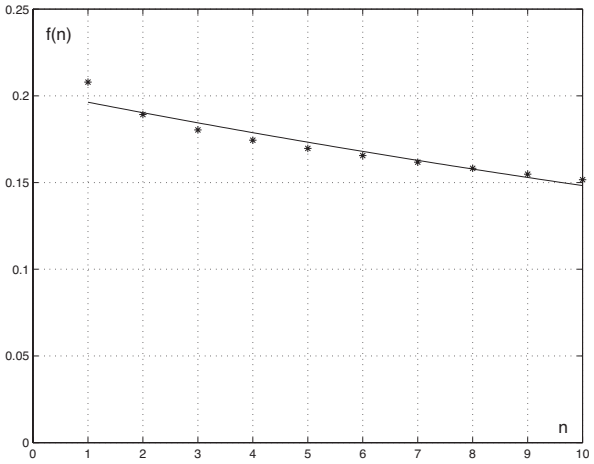
$$\lim_{n\to\infty} f(n) = 0, \;\; \nexists n_0; f(n_0) = 0 \;\text{ and }\; \forall n_1 < n_2; f(n_1) > f(n_2)$$

If we want few parameters of regression function, there is an ideal exponential function $f_t(n) = ae^{bn}$.

We will measure the quality of approximation by the square root of sum residuals:

$$\varepsilon = \sqrt{\sum_{k=1}^{n} [P_{lst}(k) - f(k)]^2}$$

For real use it is enough to have the bucket depth $n = 1, ..., 10$. Then the exponential regression function is $f(n) = 0.2025.e^{-0.0312n}$ with $\varepsilon = 0.0144$ and with the mean error $\varepsilon/n = 0.0014$.



n	$f(n)$	$P_{lst}(n) - f(n)$	n	$f(n)$	$P_{lst}(n) - f(n)$
1	0.1963	0.0116	6	0.1680	−0.0024
2	0.1903	−0.0010	7	0.1628	−0.0011
3	0.1844	−0.0041	8	0.1578	0.0004
4	0.1788	−0.0044	9	0.1529	0.0019
5	0.1733	−0.0036	10	0.1482	0.0034

Fig. 9: Approximation by exponential function for n = 1, ..., 10

The exponential function is approximation (except $n = 1$) with a maximum error $4.4 \cdot 10^{-3}$. If we are satisfied with this precision, we can exchange $P_{lst}(n)$ for $f(n)$ for $n = 2, ...,10$.

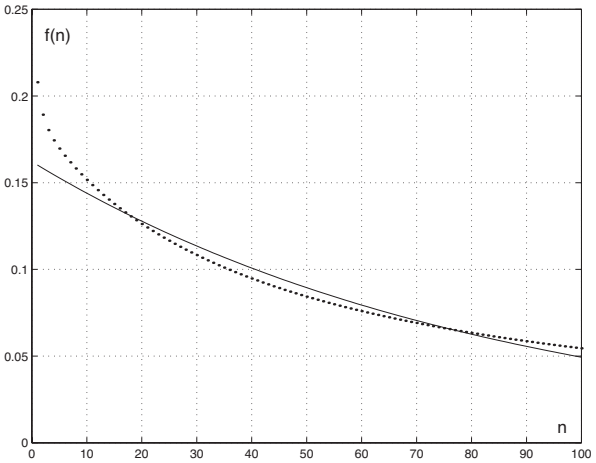


Fig. 10: Approximation by exponential function for n = 1, ..., 100

If we use the bucket depth $n = 10, ..., 100$, the exponential regression function is $f_{lst}(n) = 0.1621.c^{0.0119n}$ with $\varepsilon = 0.0807$ and with the mean error $\varepsilon/n = 0.0008$.

This approximation is very "good" for the depth $n = 11, ..., 100$ with a maximum error $6.4 \cdot 10^{-3}$.

For real use we are satisfied with the function:

$$f(n) = \begin{cases} 0.2745 & \text{for } n = 0 \\ 0.2079 & \text{for } n = 1 \\ 0.2025 \cdot e^{-0.0312n} & \text{for } n = 2, ..., 10 \\ 0.1621 \cdot e^{-0.0119n} & \text{for } n = 11, ..., 100 \end{cases}$$

5. Conclusion

We executed steady-state analysis of Hyperexponential model of Token Bucket System and created balance equations for states of Markov chain. Analytical solution led to complicated recurrent formulas for state probabilities. Therefore we used numerical methods to compute probability of a losing packet. We approxiamted these results using the exponential function. Finally we obtained the function for direct calculation of this probability.

References:

[1] ITU-T Recommendation P. 59, *Artificial Conversational Speech,* 1993.
[2] PEŠKO & SMIEŠKO, J.: *Stochastics Models of Operation Analysis* (in Slovak), Žilinská univerzita, Žilina 1999.
[3] SMIEŠKO, J.: *Exponential Model of Token Bucket System,* Communications – Scientific Letters of the University of Zilina, 4/2003, 66–70.

https://doi.org/10.26552/com.C.2004.4.80-83

komunikácie
COMMUNICATIONS

Peter Kvačkaj – Ivan Baroňák *

SUBMISSION TO CAC

Asynchronous transfer mode (ATM) is the base for future telecommunication network. The greatly variable requirements for different applications, especially in video and data traffic, make high demands on the development of traffic control in ATM networks. Speaking about Quality of Service (QoS), in ATM there is a traffic control mechanism, which ensures the standard requested by a new application or service. In the following paper an introduction to traffic control mechanism in ATM known as a connection admission control (CAC) and appropriate CAC method are presented.

Keywords: ATM, traffic contract, Quality of Service QoS, CAC method

1. Introduction

ATM is technology designed for broadband communication infrastructure. It ensures transmission of synchronous and asynchronous signals and provides a wide scale of services and applications. It links together advantages of fast packet switching and asynchronous time division. ATM provides universal communication environment for voice services, data and video transmission.

2. Traffic management

Each of the services has its own requirements for transmission. That is why the ATM must regard these requirements in the case of connection assembly and then during the transmission.

Traffic management used in the ATM technology has an important role before the connection is completed. Traffic management realizes the process of agreement between the user and communication network on QoS – *traffic contract*.

Traffic contract is a deal between the user and communication network. Traffic source characteristics are described with *traffic parameters*. After the realization of contract some of them become points of agreement, so called *QoS parameters*:

- *Peak Cell Rate* – PCR,
- *Sustainable Cell Rate* – SCR,
- *Maximum Burst Size* – MBS,
- *Minimum Cell Rate* – MCR,
- *Cell transfer delay* – CTD,
- *Maximum Cell transfer delay* – MaxCTD,
- *Cell loss ratio* – CLR.

Five categories of services in ATM are defined according to these parameters:

- *Constant Bit Rate* – CBR,
- *real time Variable Bit Rate* – rt-VBR,
- *non real time Variable Bit Rate* – nrt-VBR,
- *Available Bit Rate* – ABR,
- *Unspecified Bit Rate* – UBR.

The user guarantees not to exceed the QoS parameters presented in the contract (for example peak cell rate), but the network guarantees it only in the case of the QoS conditions adherence. The requirements of different applications are greatly variable. Consequently it is necessary to admit mechanisms which prevent the network from congested states and from signal degradation. These mechanisms are part of traffic management and they can be divided into two groups.

1. *Reactive mechanism* reacts only in cases when the capacity of a link is exceeded.
2. *Preventive mechanism* tries to prevent the network from overload situation with statistical estimation of possible on-coming network status in the case of new connection admission or with traffic monitoring and controlling.

In this paper we review possibilities of one of the preventive mechanism: the connection admission control CAC.

3. The chosen method of analysis

CAC method deals with the question whether a network node can admit a new connection or not with respect to QoS requirements of new connection and the connections in multiplex.

Some of the methods are based on mathematical models with exploitation of knowledge in the field of probability and statistics. Other methods are based on the traffic measurements or on the buffer exploitation analysis in the network nodes. CAC methods are also implemented in the environment of neural networks or using fuzzy systems.

Some of the methods need a specific traffic model others use a set of traffic parameters, for instance PCR and SCR.

* **Peter Kvačkaj, Ivan Baroňák**
Faculty of Electrotechnical and Informatics, STU, Ilkovičova 3, 812 19 Bratislava, Tel.: +421-2-68279604.
E-mail: kvackaj@ktl.elf.stuba.sk, baronak@ktl.elf.stuba.sk

3.1 Effective bandwidth method

This principle is one of the most popular. This method is difficult to compute in real conditions, so there is need to use simplified formulae for effective bandwidth estimation for on-off traffic models. For the buffer size B and a number of N connections, in this case the estimated effective bandwidth for i-th connection is:

$$C_i = R\,\frac{y - B + \sqrt{(y+B)^2 + 4yaB}}{2y} \tag{1}$$

where

$$y = (-\ln \varepsilon)\left(\frac{1}{\beta}\right)(1-a)R \tag{2}$$

where R is source peak cell rate, a is a source activity factor and $1/\beta$ is burst length.

The question of admitting a new connection is expressed in this simple formula

$$\sum_{i=1}^{N} C_i(\varepsilon_i) \le C \tag{3}$$

The simplified formulas (1, 2, 3) are not exact when the buffer size is small or moderate. Either estimation of effective bandwidth is higher than the one used in real traffic. That is why compromise was made using the continuous traffic model and calculating effective bandwidth from the Gaussian approximation. The use of continuous traffic model fails to account the statistical multiplexing gain, but the result is a lower bandwidth estimation. The required bandwidth for the Gaussian approximation equals

$$C_B = \sum_{i=1}^{N} \lambda_i + \left(\sqrt{-2\ln(\varepsilon) - \ln(2\pi)}\sum_{i=1}^{N}\delta_i\right) \tag{4}$$

where λ_i is the mean rate of cell interarrival time and σ_i is the variation of cell interarrival time for each i-th connection. The result is the required bandwidth C_B given by the Gaussian approximation. Finally, the estimated bandwidth for i-th connection is

$$\min\{C_i, C_B\} \tag{5}$$

The result of this estimation for each connection is: admission or rejection of the new connection.

3.2 Simulation of effective bandwidth method

3.2.1 Traffic model

It is necessary to create an on-off traffic model to verify functionality of the effective bandwidth method. A source is able of transmission on two possible states: the source is active and transmits with cell rate equal to its QoS parameter PCR, or is passive and does not transmit.

Traffic generated with the source is characterized with parameters of peak cell rate (PCR) and sustainable cell rate (SCR). If the traffic is monitored during the time interval T, the formulae

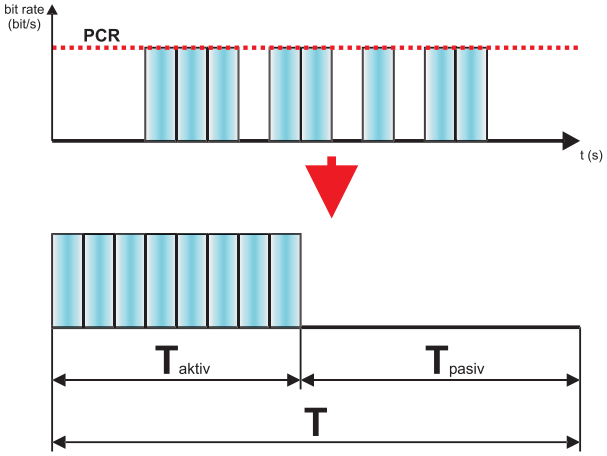


Fig. 1 Time of source active state

demonstrated below are obtained. For specification of time T the following condition must be taken into account

$$T >> \frac{1}{PCR} \quad \text{(s)} \tag{6}$$

The time of active state of the source is obtained from its traffic parameters

$$T_{aktiv} = \frac{SCR}{PCR}\cdot T \quad \text{(s)} \tag{7}$$

and the time of source passive state

$$T_{pasiv} = T - T_{aktiv} \quad \text{(s)} \tag{8}$$

The parameters mentioned above are used for designation of the following parameters – cell burst length

$$\beta^{-1} = T_{aktiv} \quad (\text{s}^{-1}) \tag{9}$$

and the source activity factor

$$a = \frac{\theta}{\theta + \beta} \tag{10}$$

where

$$\theta = 1 \backslash T_{pasiv} \quad (\text{s}^{-1}) \tag{11}$$

The parameters mentioned above are necessary for estimation of the effective bandwidth for traffic generated with the certain source.

3.2.2 Method simulation

An abstract traffic model is characterized by the vector si for each i-th connection

$$s_i = (x_1, x_2, ..., x_i, ..., x_M) \text{ for } i = 1, 2, ..., N$$
$$\text{where } x_i \in \{0, PCR_i\} \tag{12}$$

in the case of N connections, where M stands for a number of discrete cell rate measurements during the time interval T and x_l represent concrete values. Supposing two cases of traffic, in the first case the PCR value for each connection is generated with the formula

$$PCR_i = (c/N) \cdot k \tag{13}$$

in the latter case the PCR value is different for each connection

$$PCR_i = (c/2^i) \cdot k \tag{14}$$

where N stands for a number of connections and the value of constant k is a set to exceed capacity of an output link. SCR can gain values between 0 and PCR (0 and PCR value included) of this connection using the formula

$$SCR_i = \frac{\sum_{l=1}^{M} x_l}{M} \tag{15}$$

The simulation parameters are:

- buffer size $B = 1$ *cell*,
- output link capacity $C = 155$ Mbit/s
- number of connections $N = 100$,
- number of discrete patterns of connections cell,
- required cell loss probability $CLP = 0.1$ for each of the connections.

The results of simulation are presented in graphs (Figure 2 and 3). From left to right there are:

1. PCR and SCR generated values.
2. Effective bandwidth estimation for each of the connection.
3. Generated traffic with output link capacity.
4. Estimation of needed capacity for certain number of connections in multiplex.

Each of the N connections is characterised with its peak cell rate PCR and sustainable cell rate SCR (1. Connection requested PCR and SCR). All the connections in the multiplex generate VBR traffic (3. VBR data flow). The peak cell rate PCR and the sustainable cell rate SCR are input parameters for calculations of

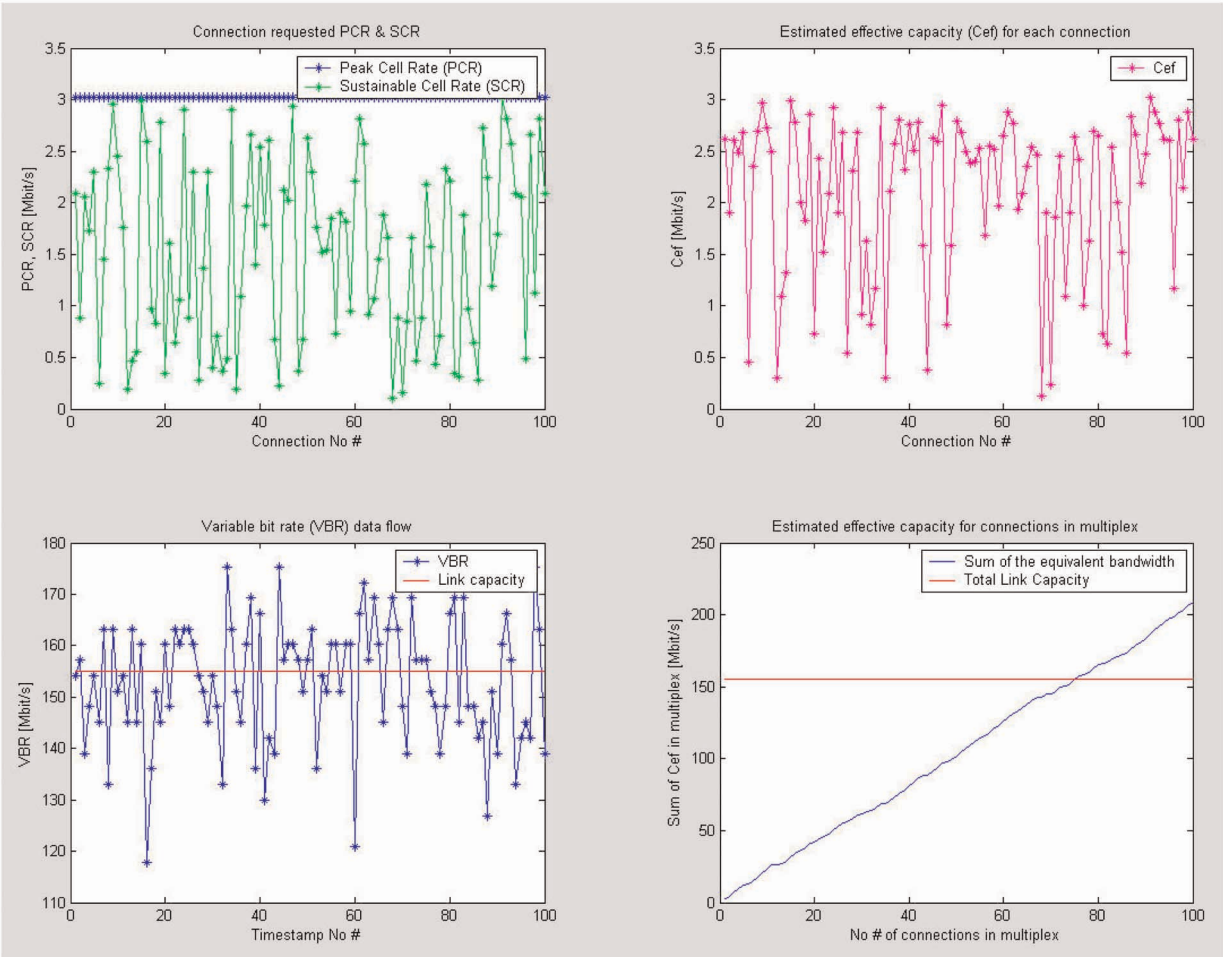


Fig. 2 Simulation results for first case of traffic state (constant PCR)

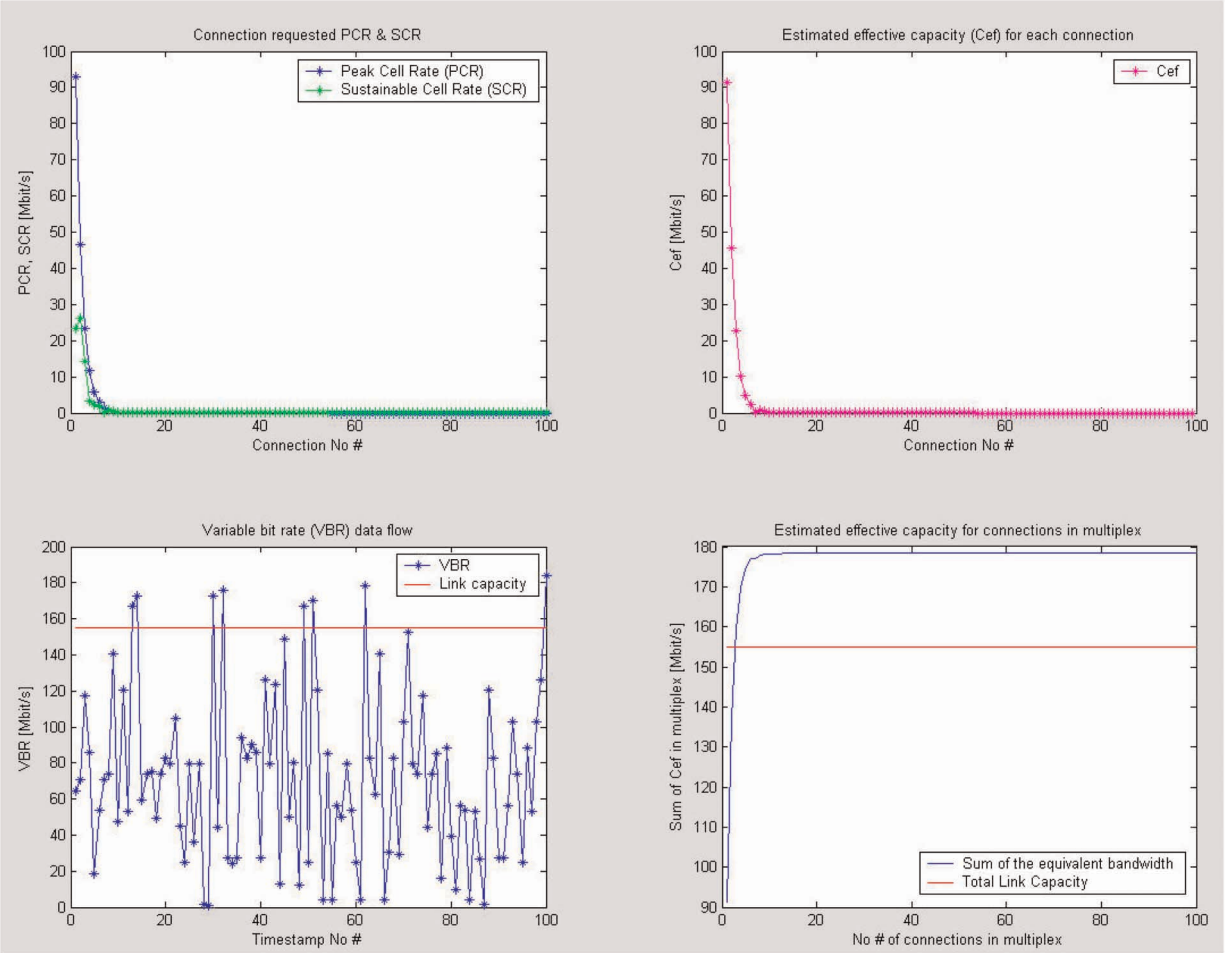


Fig. 3 Simulation results for second case of traffic state (variable PCR)

the CAC method. As a result, the CAC method returns an estimation of the needed effective bandwidth for each connection (2. Estimated efficient capacity). The summation of estimated bandwidths can not exceed the output link capacity or the connections will be rejected (4. Estimated efficient capacity for connections in multiplex).

Estimation of the CAC method in the first case enables admission to multiplex for approximately 76 connections, others will be rejected.

In the second case the first two connections take whole output link capacity. Admission for other connection will be rejected.

4. Conclusion

A properly created abstract model enables to visualize results obtained from simulation. By means of the results we can assume whether the CAC method is suitable for given environment or not. In real systems the choice of a suitable CAC method is a strategic point for effective exploitation of link capacity and QoS guarantee.

References

[1] BAROŇÁK, I., KAJAN, R.: *Management of Broadband ATM Network* (in Slovak). In.: TELEKOMUNIKACE 97, International Scientific Conference Brno, VUT, 12 June 1997, pp. 6–12.

[2] BAROŇÁK, I.: *The Pilot Project - ATM Network*. In: Bridging East&West Through Telecommunications. Acapulco, Mexico, 22–25 May 2000, pp.760–763.

[3] GELENBE, E., MENG, X., ONVURAL, R.: *Diffusion based statistical call admission control in ATM,* Journal Performance Evaluation, 1996, No. 27&28, pp. 411–436.

https://doi.org/10.26552/com.C.2004.4.84-88

kamUNikácie
COMMUNICATIONS

J. Slovak – C. Bornholdt – B. Sartorius *

A WAVELENGTH-PRESERVING ALL-OPTICAL 3R REGENERATOR

A novel all-optical 3R regenerator operated in the optically clocked scheme (OC-3R) is presented. The regenerator basically comprises two semiconductor components: an ultra-long semiconductor optical amplifier (UL-SOA) for effective use of fast nonlinearities and a self-pulsating PhaseCOMB-laser for all-optical clock recovery. The dynamic behaviour of the key devices is investigated theoretically and experimentally. An excellent regenerative performance of the assembled OC-3R is demonstrated by eye analysis and by bit error rate (BER) measurements at 40 Gbit/s.

1. Introduction

In future high-speed global optical networks the data signals will be transparently switched through the nodes without additional opto-electronic conversion. In this case all-optical 3R regeneration (re-amplification, re-timing, re-shaping) is a key function in order to make networks resistant to distortions derived from optical cross-connects and transmission links [1]. A scheme of a 3R regenerator with its three functional elements is depicted in Fig. 1. First, the power level of the degraded signal is increased in the amplification stage. The re-timing function is performed by the clock signal, which is recovered from the pseudo random bit sequence (PRBS) data signal. Finally, a decision stage with a threshold function suppresses accumulated noise and equalises amplitude fluctuations. A completely restored signal is obtained at the regenerator output.

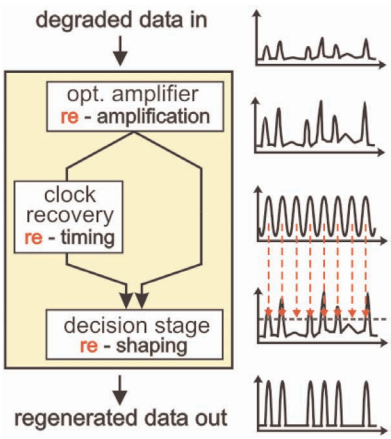


Fig. 1 Principle scheme of a 3R regenerator [1]

Several regenerator architectures have been demonstrated in different experiments [2, 3]. However, most solutions apply a large number of optical components and require high power levels inside the regenerator. In this paper we present a novel regenerator scheme that is based on only two compact semiconductor devices: a 4 mm ultra-long semiconductor optical amplifier (UL-SOA) for effective exploitation of fast non-linearities and a self-pulsating DFB laser (PhaseCOMB) for all-optical clock recovery. An optically clocked technique is applied to stabilise the gain dynamic in the UL-SOA and to perform the regeneration function on the data signal [4]. The application of compact semiconductor devices results in low power consumption and integration possibility on a planar lightwave circuit platform. We investigate theoretically and experimentally the characteristics of the key components and demonstrate the performance of the assembled optically clocked 3R regenerator at 40 Gbit/s.

2. Optically clocked all-optical 3R regenerator

The experimental set-up of the optically clocked regenerator (OC-3R) is shown in Fig. 2. An ideal 40 Gbit/s RZ PRBS signal at 1550 nm was first degraded in an optical noise unit in order to simulate the distortions derived from a transmission link. At the regenerator input the data signal was amplified by a linear semiconductor optical amplifier (LOA). A split-off part of the signal was injected into the self-pulsating PhaseCOMB laser for all-optical clock recovery. The PhaseCOMB laser synchronised to the injected data signal and emitted a 40 GHz stable sinusoidal pulse stream at 1562 nm. The second part of the data signal was delayed by half a bit relative to the clock pulses. Both signals were simultaneously

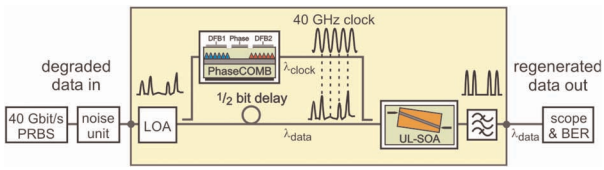


Fig. 2 Set-up of the all-optical OC-3R regenerator

* **J. Slovak, C. Bornholdt, B. Sartorius**
Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany

launched into the ultra-long SOA. The clock signal was blocked at the regenerator output while the data signal was transmitted on its original wavelength. The quality of the data signal was analysed by jitter analysis using an electrical sampling oscilloscope and by BER measurements. In the following, the key devices of the OC-3R are characterised in more detail.

3. Ultra-long semiconductor optical amplifier (UL-SOA)

An ultra-long semiconductor optical amplifier with a length of 4 mm has been developed and manufactured (Fig 3). It is based on a BH structure with a 1.5 μm InGaAsP strained bulk active layer for polarisation insensitive operation. Tilted facets are applied additionally to the antireflection coating in order to minimise unwanted reflections. The UL-SOA chip is packaged into a Peltier-cooled module.

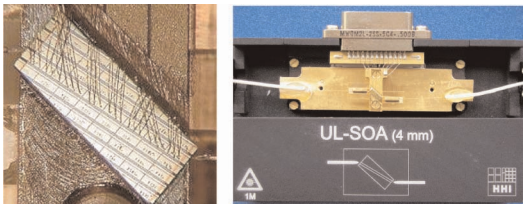


Fig. 3 The chip of the 4 mm UL-SOA (left-hand side) and the device packaged into a module (right-hand side)

3.1 Numerical results on UL-SOA gain dynamic at 40 Gbit/s

The dynamic behaviour of an SOA was simulated using the VPI Transmission Maker program. First a single ideal 40 Gbit/s RZ data signal (Fig. 4a) was injected into the SOA with a length of 4 mm. Strong signal degradation caused by patterning effects can be observed (Fig. 4b). Due to the relatively slow inter-band effects [5] the data pulses are amplified differently depending on the previous data sequence passing through the SOA. The first "1" bit after a long "zeros" sequence is amplified more then a "1" within a bit sequence of "ones". Next, the optically clocked principle was applied. In this case, an additional 40 GHz sinusoidal signal is launched into the SOA simultaneously with the data signal.

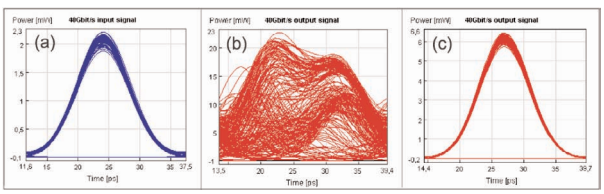


Fig. 4 Eye diagrams of the ideal 40 Gbit/s RZ signal at the SOA input (a), and of the output data signal without additional SOA stabilisation (b) and by applying the OC-technique (c)

The clock pulses are delayed by half a bit relative to the data pulses, which is the crucial condition of this principle. The corresponding data signal at the SOA output is shown in Fig. 4c. The slow gain dynamic is suppressed by the strong clock signal. Thus the high signal quality is preserved.

The behaviour of the optically clocked SOA was simulated also for degraded 40 Gbit/s PRBS data signals (Fig. 5a). The length of the SOA varied from 0.5 mm to 8 mm and the corresponding quality of the data signal at the SOA output was analysed (Fig. 5b-f). A signal improvement mainly for longer devices can be observed. The interaction between data and clock leads to the regenerative function due to fast intra-band effects in the semiconductor medium. However, comparing the simulation results for the 4 mm and the 8 mm SOA reveals only a relatively small improvement of the quality. Note, for practical use the SOA should be as short as possible. Thus a 4 mm UL-SOA has been manufactured for experimental analysis.

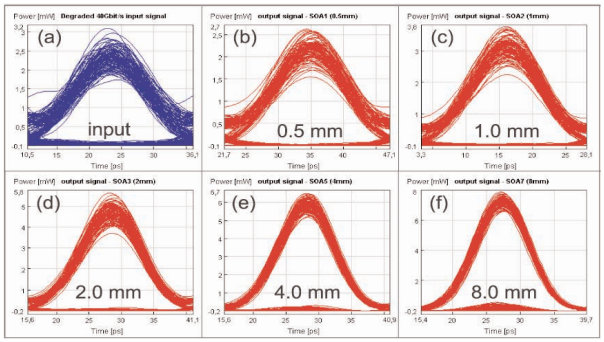


Fig. 5 Degraded data signal at the SOA input (a) and the corresponding output signals for different SOA lengths (b-f) by applying the OC-principle

3.2 Experimental investigation of the UL-SOA performance at 40 Gbit/s

The dynamic behaviour of the 4 mm UL-SOA at 40 Gbit/s was investigated experimentally. We analysed the results using an

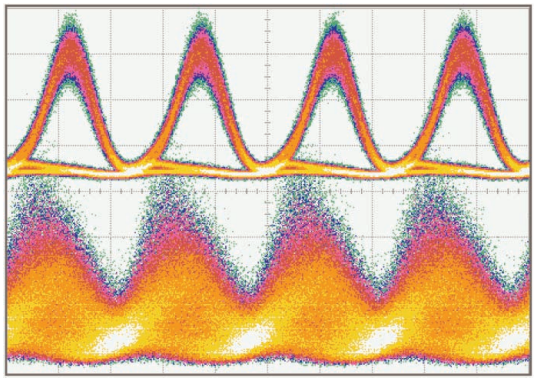


Fig. 6 Eye diagram of 40Gbit/s PRBS RZ signal (upper trace) and the data signal at the UL- SOA output (lower trace)

electrical sampling oscilloscope with precision time base. First, a high-quality 40 Gbit/s RZ data signal at 1550 nm (Fig. 6 upper trace) was launched directly into the UL-SOA. As expected from the simulation results, strong signal degradation due to patterning effects occurs (Fig. 6 lower trace).

The optically clocked technique was applied in order to stabilise the dynamic of the UL-SOA. An additional 40 GHz sinusoidal signal at 1560 nm was injected into the device. The clock signal was generated using an external modulator electrically driven by a 40 GHz synthesizer. In this case, no signal degradation is observed at the output of the semiconductor device (Fig. 7 lower trace). The carrier density in the ultra-long SOA is periodically modulated by the clock signal and not by the pattern envelope of the PRBS signal. This validates the analysis performed by simulations.

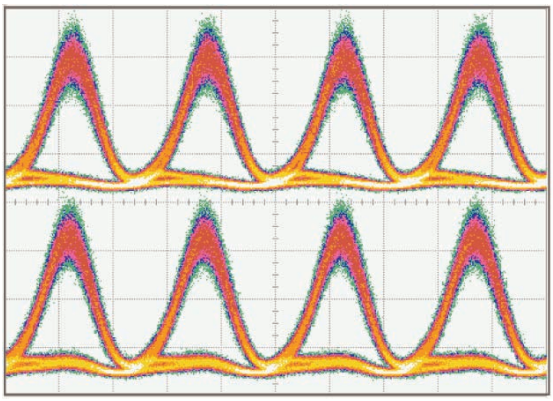


Fig. 7 40 Gbit/s PRBS RZ signal at the UL- SOA input (upper trace) and the output data signal (lower trace). The UL-SOA is stabilised by an additional 40 GHz clock signal

3.3 Jitter reduction function of the optically clocked UL-SOA

It was shown that the quality of the ideal data signal remains preserved at the output of the UL-SOA applying the optically clocked technique. Next, we continuously degraded the data signal by additional time jitter and measured the corresponding jitter of the output signal. The power levels of the 40 Gbit/s data signal and 40 GHz clock signal at the UL-SOA facet were -2 dBm and +4 dBm respectively. The results are summarised in Fig. 8. The rms-jitter of the input signal varies in a wide range from 400 fs to 1400 fs. A strong jitter reduction function of the optically clocked UL-SOA can be observed. The rms-jitter of the output signal increases only slightly and remains below 700 fs. The jitter improvement is related to the clock signal with its high time stability. The periodic gain modulation of the ultra-long SOA by the stable sinusoidal signal effectively shapes the data pulses and pushes them towards the correct time position.

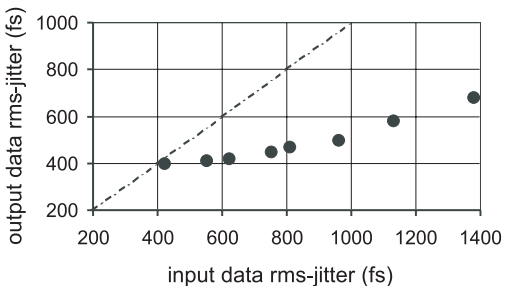


Fig. 8 Rms-jitter of the PRBS RZ signal measured by a sampling oscilloscope at the input of the optically clocked UL-SOA and the corresponding time jitter of the output data signal

4. Self-pulsating PhaseCOMB laser for all-optical clock recovery

The self-pulsating PhaseCOMB-laser (SP-laser) is a ridge waveguide three section InGaAsP DFB-laser comprising two DFB-sections and an integrated phase tuning section [6]. Each section is driven separately by a direct current. It is a compact device with a total length below 1 mm. The facets of the PhaseCOMB are antireflection coated and the device is packaged into a temperature controlled module.

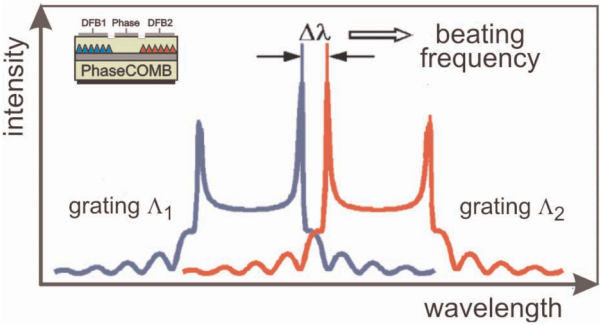


Fig. 9 Design concept of the PhaseCOMB-laser

The superposition of the lasing modes of both DFB-lasers generates a beating signal with a self-pulsation frequency given by their spectral distance (Fig. 9). The frequency can be continuously tuned in a wide range (28 GHz – 50 GHz) simply by the three direct currents [7]. If the frequency is adjusted close to the data rate, the self-pulsation synchronises to the injected PRBS signal

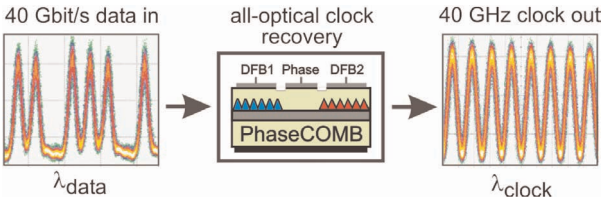


Fig. 10 Operation principle of the self-pulsating laser

and the SP-laser emits a sinusoidal pulse stream, which is stable in amplitude and time (Fig. 10). The locking to the data signal is polarisation and wavelength insensitive within the whole C-band.

4.1 Jitter reduction function of the PhaseCOMB laser

Similar to the previous characterisation of the optically clocked UL-SOA jitter analysis of the self-pulsating laser was performed. The 40 Gbit/s PRBS data signal at 1550 nm, degraded by additional time jitter, was launched into the PhaseCOMB laser. The SP-laser synchronised to the data signal and recovered a 40 GHz sinusoidal clock signal. The results are summarised in Fig. 11. Note that the time stability of the clock signal is preserved. Even for a data signal with a high jitter of 1400 fs, the rms-jitter of the clock remains well below 600 fs. Thus the excellent retiming function of the PhaseCOMB-laser is verified.

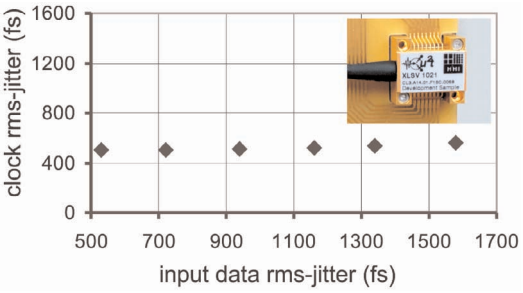


Fig. 11 Time stability of the PhaseCOMB-laser. (The packaged device as insert)

5. Optically clocked all-optical 3R regenerator

Finally, the performance of the assembled optically clocked 3R regenerator (Fig. 2) was investigated. The ideal 40 Gbit/s RZ signal (1550 nm) with a jitter of 520 fs was first additionally degraded by amplitude noise and timing jitter. The strong amplitude fluctuations and the rms-jitter of more than 1200 fs are clearly visible (Fig. 12a). The PhaseCOMB laser locked to the data signal and emitted a stable pulse stream at 1562 nm with a rms-jitter of only 526 fs (Fig. 12b). According to the optically clocked principle, the degraded data signal and the recovered clock signal were delayed by half a bit relative to each other and launched into the ultra-long SOA. The power levels of the data signal and clock at the UL-SOA input were -2 dBm and +4 dBm respectively. The quality of the data signal at the regenerator output is clearly improved (Fig. 12c). The amplitude and time stability of the regenerated output pulses is only marginally affected by the degraded input signal. The power fluctuations are effectively suppressed, the rms-jitter is reduced from 1220 fs to about 650 fs. Furthermore, the regenerated signal has clearly larger eye opening.

Next, bit error rate measurements (BER) were performed to evaluate the system performance of the OC-3R regenerator. The results are shown in Fig. 13. First an ideal back-to-back signal with

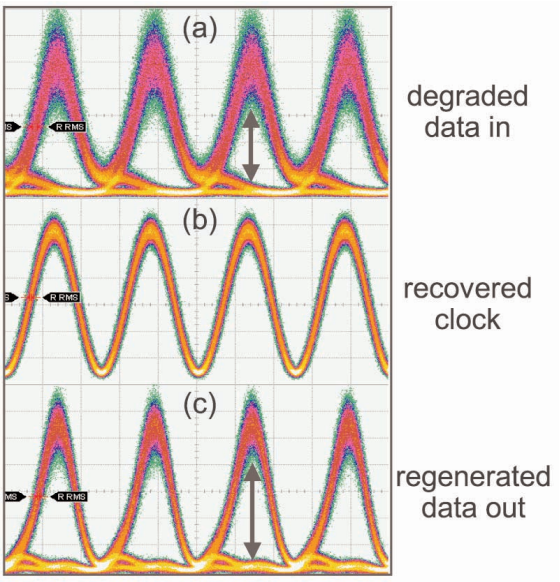


Fig. 12 Degraded 40 Gbit/s PRBS RZ signal (a) the recovered 40 GHz clock signal (b) and the data signal at the OC-3R regenerator output (c)

520 fs jitter was injected into the bit error rate detector and the BER curve depicted by filled circles was measured. A strongly degraded signal with a jitter of 1220 fs was then applied. The BER curve for the 40 Gbit/s data signal at the regenerator input is shown by the filled squares and for the output signal by the open squares. The signal quality is clearly improved by the OC-3R regenerator and a negative penalty of more then -2 dB is obtained at the bit error rate of 10-9. Note that there is no indication of an error floor down to a BER of 10-11. Further, it can be observed that the BER curve measured for the regenerated signal is close to the BER curve of the ideal input data signal. Thus, the excellent features of the optically clocked 3R regenerator shown by eye analysis were verified by the bit error rate measurements.

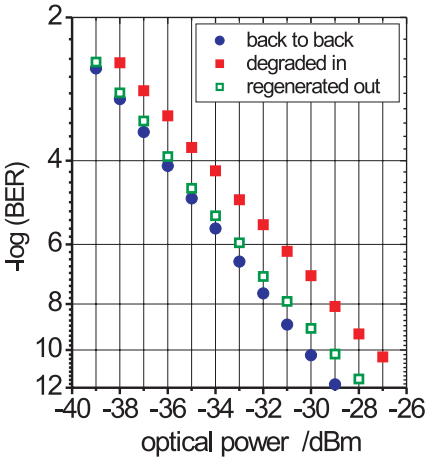


Fig. 13 Performance evaluation of the OC-3R by BER measurements

6. Conclusions

An all-optical 3R regenerator operated in the optically clocked scheme at 40 Gbit/s has been presented. It was based on two compact semiconductor components: the ultra-long SOA and the self-pulsating PhaseCOMB laser. The dynamic behaviour of the UL-SOA was investigated numerically and experimentally and a good correspondence was found. Applying the optically clocked principle, the carrier density in the SOA was effectively stabilised by the clock signal, resulting in jitter reduction function at the UL-SOA output. The self-pulsating DFB-laser applied for all-optical clock recovery synchronised even to strongly degraded data signals and emitted a stable sinusoidal pulse trace. An excellent regenerative function of the assembled optically clocked 3R regenerator was demonstrated. The rms-jitter of the degraded 40 Gbit/s data signal was reduced from 1220 fs down to 650 fs. The BER curve of the regenerated signal showed a negative penalty of -2 dBm.

The use of semiconductor devices makes the OC-3R regenerator very compact, easy to handle and attractive for application in optical signal processing systems. It will pave the way for competitive and cost effective all-optical 3R schemes.

References

[1] SARTORIUS, B.: *All-Optical 3R Signal Regeneration*, ECOC ' 00, Munich, Germany, We 9.4.1
[2] LAVIGNE. B. ET AL.: *Cascade of I00 optical 3R regenerators at 40 Gbit/s based on all-active Mach-Zehnder interferometers*, ECOC'01, Amsterdam, Nederland, We.F.2.6
[3] NISHIMURA, K.: *Novel All-optical 3R Regenerator Using Cross-Absorption Modulation in RF-Driven Electroabsorption Waveguide*, ECOC 2001, Amsterdam, Nederland, We.F.2.4
[4] SLOVAK, J. ET AL.: *Optically Clocked Ultra Long SOAs: A Novel Technique for High Speed 3R Signal Regeneration*, OFC 2004, Los Angeles, USA, WD4
[5] GIRARDIN, F. ET AL.: *Gain recovery of bulk semiconductor optical amplifiers*, IEEE Photon. Technol. Lett., vol.10, pp. 784–786, 1998.
[6] SARTORIUS, B.: et al.: *All-optical clock recovery module based on self-pulsating DFB laser,* Electronics Lett., 20th August 1998, Vol. 36, No. 17
[7] SARTORIUS, B.: *All-Optical Clock Recovery for 3R Optical Regeneration,* OFC 2001, Anaheim, California, MG 7, invited.

https://doi.org/10.26552/com.C.2004.4.89-93

Miroslav Voznak *

EXTENT OF SERVICES SUPPORTED BY Q-SIGNALING OVER IP

QSIG is a signaling protocol used for interconnection of various telecommunication systems in corporate networks. These networks are composed of homogeneous or heterogeneous elements. Basic procedures of connection and supplementary services are transferred through messages, which are defined in recommendations ETSI or ISO. QSIG can be used for transport proprietary signaling between the same PBXs and is supported in voice gateway of solution of Voice over IP too and environment of IP network is often chosen for interconnection between PBXs. The main aim of this paper is to acquaint specialists with the result of the QSIG compatibility test between Siemens and Cisco. Siemens produces one of the most used telecommunication systems in the Czech republic and Cisco systems routers are the most used in IP solutions. In order to provide all supplementary services in telecommunication, Siemens developed a new protocol, called CorNet-NQ, which is a superset of QSIG and uses messages tunnelling through QSIG. For realization of this test the QSIG version PSS1V2 in agreement with the ISO standards was used. [1]

1. Introduction

QSIG is a signaling protocol for controlling the establishment, maintenance and clearing of calls between PINXs, nodal entity known as Private Integrated services Network eXchange. It provides an extremely powerful method of connecting PINX equipment in a corporate network. QSIG is not a proprietary standard. It is an open, international standard and is supported by the world's leading PBX suppliers. Twelve of the world's leading PBX manufacturers signed a Memorandum of Understanding (MoU) concerning the development and support of QSIG. The MoU, which came into effect on 1st February 1994, commits the signatories to facilitate the performance of interoperability tests and to:

- incorporate a Primary Rate interface (as defined in ETS 300 011) into its products,
- support a Basic Call (as defined in ETS 300 172),
- implement the generic procedures (as defined in ETS 300 239),
- implement QSIG supplementary services as far as each signatory considers to be economically viable,
- and participate in interoperability testing with the other signatories.

QSIG standards are developed within ECMA in Technical Committee 32 (TC32) for Communications, networks and systems interconnection. TC32 began to work on QSIG in 1988 and most of these standards were published also as ISO/IEC International Standards, endorsed by ETSI as European Standards and implemented by all major PBX vendors. [2]

2. QSIG protocol stack

QSIG standards specify a signaling system at the "Q" reference point, which is primarily intended for use on a common channel;

Layer	Standards			Description
Layer 4-7	Application mechanism			End to End protocol network transparent
Layer 3	Standards depend on the value of supplementary services			QSIG procedures for supplementary services
	ISO/IEC 11582, ETS300 239, ECMA165			QSIG Generic Functional Procedures
	ISO/IEC 11574/11572, ETS300 171/172, ECMA142/143			QSIG Basic Call
Layer 2	ECMA141, ETS300 402			Interface-dependent protocols
Layer 1	Basic Access (BRA) ETS300 011 I.430	Primary Access (PRA) ETS300 012 I.431		
Medium	Copper Wire	Copper Wire	Optical fibre	

Figure 1: Reference model QSIG

* **Miroslav Voznak**
Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science, Department of Electronics and Telecommunications,
Tel: +420 596991699, E-mail: miroslav.voznak@vsb.cz

e.g. G.703 interface. Within the public ISDN the two end PINXs are connected through two reference points using different ISDN protocols, namely DSS1 at the "T" reference point and ISUP within the public ISDN at the "N" reference point. For private ISDNs, only one protocol is necessary as the QSIG protocols have sufficient functionality to be used both within the network at transit nodes and on the outside at access nodes.

The architecture of QSIG signaling protocol agrees with the architecture RM OSI, see Figure 1.

The standard for Basic Call is ECMA 143 (ISO/IEC 11572), QSIG generic support for supplementary services is defined in ECMA-165 as a toolkit on which signaling for support of supplementary services can easily be built (ISO/IEC 11582). The QSIG protocol stack is identical in structure to the DSS1 protocol stack. Both follow the ISO reference model. Both can have an identical layer 1 and layer 2 (LAPD). However, at layer 3 QSIG and DSS1 differ.

In the last years significant part of work inside TC32 is concerned with the interoperation of Private Integrated Services Networks (PISN) with IP Networks, in particular with the two following aspects:

- Interworking of PISNs and IP networks via Gateway, primarily between QSIG and H.323 and between QSIG and SIP,
- Connection of PISN components via IP network infrastructures (i.e. QSIG tunnelling).

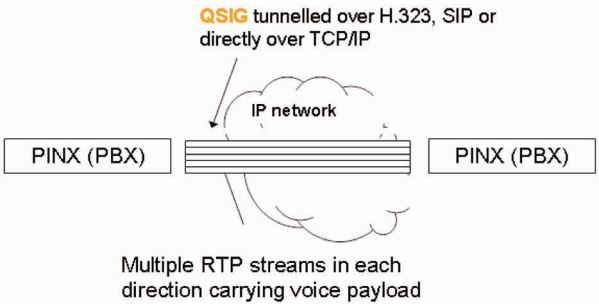


Figure 2 : Tunnelling QSIG over IP network.

Inside the TC32 a new Task Group TG-17 was founded, known as TC32 – TG17, for standardization in the area of IP-based multimedia communications. This group considers possible future work in the area of NGN (Next Generation Networks), where close cooperation with ETSI is anticipated.

New ECMA standards allow tunnelling QSIG over IP network, see Figure 2, these standards are listed below.

Interworking between QSIG and H.323:

- *ECMA-332 – basic call*
- *ECMA-307 – generic support for supplementary services*
- *ECMA-308 – call transfer supplementary services*
- *ECMA-309 – call diversion supplementary services*
- *ECMA-326 – call completion supplementary services*

Interworking between QSIG and SIP:

- *ECMA-339 – basic call*
- *work in progress on call transfer and call diversion services*

Tunnelling of QSIG over IP (for PBX interconnection):

- *ECMA-333 – tunnelling of QSIG over H.323*
- *ECMA-336 – tunnelling of QSIG directly over TCP/IP*
- *work in progress on tunnelling QSIG over SIP*

Figure 3 shows that for private ISDNs, the QSIG protocols have sufficient functionality to be used both within the network at transit nodes and on the outside at access nodes. Hence QSIG is used between all three PINXs. The "Q" reference point is the logical signaling point between two PINXs. The physical connection to the PINXs is made at the "C" reference point. The Intervening Network (IVN) can be either dedicated channels (analogue or digital) or switched connections (for Virtual Private Networks).

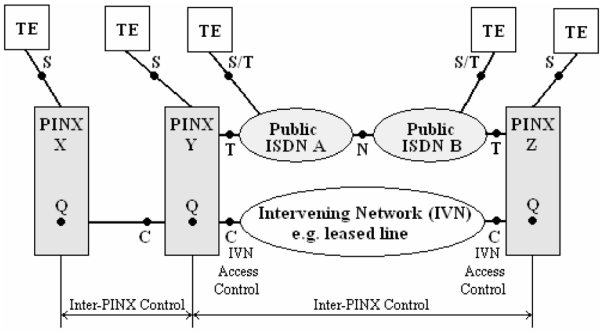


Figure 3: Reference model extended for corporate networks

3. Conditions of QSIG test interoperability

User signaling services can be divided into two groups:

- *BASIC CALL* – to transfer the control information required for the set-up, monitoring and release of connection and the identification of a subscriber's number,
- *SUPPLEMENTARY SERVICES* – to exchange signaling information for the control of the supplementary services and additional network features (ANFs)

System components used in test *Tab. 1*

PBX model	Siemens HiPath4000
PBX Release	Version 1-09
Signaling	QSIG (PSS1V2/ISO)
Interface	ISDN BRI
Voice gateway	Cisco 1751-V
Gateway release	IOSTM 12.3(6a)
VoX protocol	ITU-T H.323

In this QSIG test interoperability individual supplementary services were tested and signalling messages were traced between interconnected equipment in all signalling path. [3], [4], [5], [6].

Test configuration

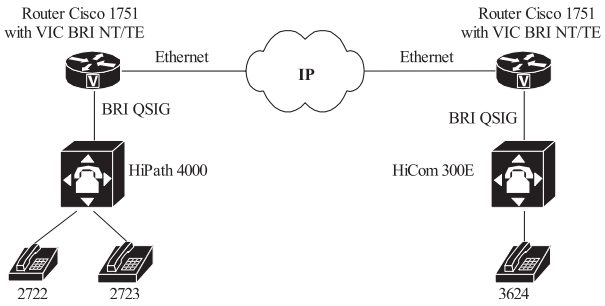


Figure 4: Used model of tested network.

3.1 Configuration of the voice gateway

Set up Notes

The Cisco 1751-V gateway with BRI-ISDN supports protocol QSIG, when the value of Switch-type parameter is set to *basic-qsig*. The network side on the BRI is set by the following commands:

- isdn layer1-emulate network,
- isdn protocol-emulate network.

Significant part of configuration in Cisco voice gateway is shown below as an example.

```
/* for interface BRI

interface BRI0/0
 no ip address
 isdn switch-type basic-qsig
 isdn overlap-receiving T302 5000
 isdn protocol-emulate network
 isdn layer1-emulate network
 isdn incoming-voice voice
 isdn static-tei 0
 isdn timeout-signaling
!
voice-port 0/0,
 compand-type a-law
 cptone CZ
 bearer-cap 3100Hz
!
```

```
/* for outgoing and incoming route

dial-peer voice 2 pots
 destination-pattern ....
 direct-inward-dial
 port 0/0
!
dial-peer voice 11 voip
 destination-pattern 36..
 session target
ipv4:195.113.113.156
 codec g711alaw
```

3.2 Configuration of the PBX

Set up Notes

The Siemens HiPath 4000 supports the protocol QSIG when the value of parameter Protvar in TDCSU command is set to *PSS1V2*, and the user side on the BRI is set by the following parameters:

- *MASTER = N*,
- *SMD = N*.

A significant part of configuration in Siemens HiPath 4000 is shown below as an example.

```
                    /* TDCSU - command for trunk configuration

+------------------------ DIGITAL TRUNK (FORMAT=L) --------------------------+
| DEV      = S0CONN         PEN      =  1-01-103-1   TGRP     = 157         |
|-------------------------------------------------------------------------|
| PROTVAR  = PSS1V2         INS      = Y             SRCHMODE = CIR         |
| COTNO    = 11             COPNO    = 10            DPLN     = 0           |
| ITR      = 0              COS      = 3             LCOSV    = 1           |
| LCOSD    = 1              CCT      = CISCO 157     DESTNO   = 157         |
| SEGMENT  = 8              DEDSCC   =               DEDSVC   = NONE        |
| FACILITY =                DITIDX   =               SRTIDX   =             |
| TRTBL    = GDTR           SIDANI   = N             ATNTYP   = TIE         |
| CBMATTR  = NONE           NWMUXTIM = 10            TCHARG   = N           |
| SUPPRESS = 0              DGTPR    =               CHIMAP   = N           |
| ISDNIP   =                ISDNNP   =                                      |
| PNPL2P   =                PNPL1P   =               PNPAC    =             |
| TRACOUNT = 31             SATCOUNT = MANY          NNO      =             |
| ALARMNO  = 0              FIDX     = 1             CARRIER  = 1           |
| ZONE     = EMPTY          COTX     = 11            FWDX     = 1           |
| DOMTYPE  =                DOMAINNO =               TPROFNO  =             |
| INIGHT   =                                         CCHDL    =             |
| UUSCCX   = 16             UUSCCY   = 8             FNIDX    = 1           |
| CLASSMRK = EC     & G711    & G729OPT              SRCGRP   =             |
| TCCID    =                                                              |
|-------------------------------------------------------------------------|
| MASTER   = N              SMD      = N             CNTRNR   = 0           |
| BCNEG    = Y                                                            |
+-------------------------------------------------------------------------+

/* next commands RICHT, LDAT and WABE are needed, closed numbering scheme was used
                                in configuration
```

3.3 Decoding of QSIG

To decode the QSIG messages, SW application the WinVisu was used. It supports CORNET-NQ decoding protocol too and all the contents of signaling are in detail displayed in accordance with the recommendation of the chosen protocol, see Figure 5.

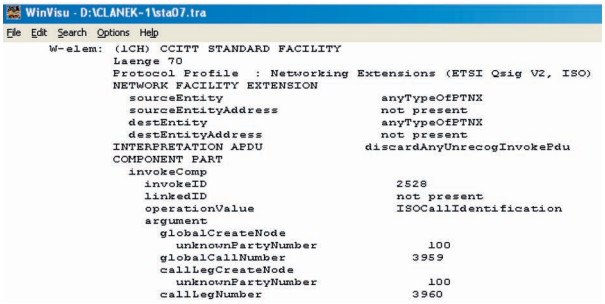


Fig. 5. Screen of WinVisu window

4 Test result

In the table below, the tested supplementary services are listed including the results related to each of these services.

Only one service (of all 22 supplementary services), Call Completion Busy Subscriber – CCBS did not work properly because the SETUP message was not transferred correctly through the IP network.

Information element in SETUP message with CCBS

W-elem: (1CH) CCITT STANDARD FACILITY
Laenge 48
Protocol Profile : Networking Extensions (ETSI Qsig V2, ISO)
NETWORK FACILITY EXTENSION
sourceEntity endPTNX
sourceEntityAddress not present
destEntity endPTNX
destEntityAddress not present
INTERPRETATION APDU rejectAnyUnrecInvokePdu
COMPONENT PART
invokeComp
invokeID 23280
linkedID not present
operationValue ISOCcbsRequest
argument
numberA
presAllowedNumber
unknownPartyNumber 2722
numberB
unknownPartyNumber 3624
service 0x04 0x03 0x80 0x90 0xa3 0x7d 0x02 0x91 0x81

Table of tested supplementary services Table 1

Description		Support
Call Completion	Busy subscriber – CCBS	NOK
	No Reply – CCNR	OK
Call Forwarding	Busy subscriber – CFBS	OK
	No Reply – CFNR	OK
	Unconditional – CFU	OK
Call Intrusion – CI		OK
Identification	CLIP	OK
	COLP	OK
	CNIP	OK
	CONP	OK
Identification Restriction	CLIR	OK
	COLR	OK
Conference – CONF		OK
Call Transfer – CT		OK
Call Waiting – CW		OK
Do Not Disturb – DND		OK
Do Not Disturb Override – DNDO		OK
HOLD		OK
Message Waiting Indication – MWI		OK
Direct Dialing In- DDI		OK
Multiple subscriber number – MSN		OK
Advice of charge – AOC D/E		OK
Call Deflection – CD		Not tested
Path Replacement – PR		Not tested
Subadressing – SUB		Not tested
User-User Signaling – UUS		Not tested

This supplementary service was not transferred because the voice gateway incorrectly processed the CCBS signaling request in SETUP message. On the other hand, the CCNR – Call Completion to Subscriber (in idle state) was transferred properly as it was indicated through MWI facility (Message Waiting Indication). Supplementary CD service was not tested as it is not supported by HiPath 4000.

References:

[1] ISO/IEC Standards, http://www.iso.ch/iso/en/ittf/PubliclyAvailableStandards
[2] ETS 300 172, *Inter-exchange signaling protocol,* ETS Institute, France, 1996
[3] PETERS J., DAVIDSON J.: *Voice over IP Fundamentals,* Cisco Press, Indianopolis USA, 2000, ISBN 1-57870-168-6
[4] VOZNAK, M.: *Comparison of H.323 and SIP Protocol Specification,* International Conference "Research in Telecommunication Technology 2003", STU Bratislava, September 2003, p. 45-47, ISBN 80-227-1934-X
[5] HiPath 4000 Service Manual, Siemens AG, Munich, 2002, P31003-H3101-K104-2-20
[6] Cisco Systems Inc., http://www.cisco.com

https://doi.org/10.26552/com.C.2004.4.94-97

Christian Julien *

ETSI ACTIVITIES ON NEXT GENERATION NETWORKS

The contribution deals with the current activities in ETSI (European Telecommunications Standards Institute) on Next Generation Networks (NGN). The first part includes the general information on ETSI and the creation of the new Technical Committee TISPAN (Telecommunications and Internet converged Services and Protocols for Advanced Networks) which is responsible for all aspects of the wired networks like signalling protocols or systems interworking and which is currently moving towards definition and specification of the NGN. The next part is aimed at the current structure of the TC TISPAN and finally, the key objectives within the TISPAN Releases are presented.

The European Telecommunications Standards Institute (ETSI) is an independent, non-profit organization, whose mission is to produce telecommunications standards for today and for the future.

Based in Sophia-Antipolis in the south of France, ETSI unites 688 members from 55 countries, and brings together manufacturers, network operators and service providers, administrations, research bodies and users – providing a forum in which all key players can contribute.

ETSI's Members determine the Institute's work programme, allocate resources and approve its deliverables. As a result, ETSI's activities are closely aligned with market needs and there is wide acceptance of its products. ETSI's standards are built on consensus.

ETSI is structured in different working groups called Technical Committee (TC). Each of them is responsible for a given technical area like: Satellite, Radio Communications, Fixed Networks Communications, Safety, and Mobile Communications.

One of them, ETSI Technical Committee TISPAN (Telecommunications and Internet converged Services and Protocols for Advanced Networks) was created in September 2003 as the result of the combination of TC SPAN (Services and Protocols for Advanced Networks) and ETSI Project TIPHON (Telecommunications and IP Harmonization Over Networks). The decision to merge the two former bodies has been taken by the ETSI Board in view of maximizing the synergies between the top down approach of TIPHON in their definition of technology independent features network architecture to support service capabilities and the bottom up approach followed by SPAN in their workplan to produce their specifications enabling a smooth evolution of legacy networks towards IP-based networks to support a wide range of new generation services. The objective of TISPAN is indeed to produce a first set of specifications that are awaited by the ETSI members in order to provide a confident basis for the development and implementation of first steps of NGN systems supporting truly innovative multimedia services.

During the first half of 2003, EP TIPHON advanced significantly the support of multiple media applications, providing extended security features, user authentication and information confidentiality. These efforts were aimed at maximizing the service capabilities and network features harmonization with other ongoing standards developments, particularly with the Universal Mobile Telecommunications System (UMTS™) under the auspices of the Third Generation Partnership Project (3GPP™). In addition plans for a proposed activity were discussed in order to further enhance the set of fundamental service capabilities (such as presence management, messaging) aiming at enabling ubiquitous multimedia communications.

Concurrently, up to summer 2003, TC SPAN continued to actively contribute to the activities of the International Telecommunication Union Telecommunications Standardization Sector (ITU-T), and to work jointly with the Third Generation Partnership Project (3GPP™) particularly for the development of interfaces with service creation environment and application platforms through Application Programming Interfaces (APIs) providing Open Services Access (OSA). These sustained activities aimed at ensuring that fixed networks requirements are well accommodated by the emerging network architectures and technologies.

Combining both helped strengthen synergies and confirmed the ETSI crucial role in rolling out Multi-service/Multimedia Network technologies in Next Generation Networks (NGN) environments at the global level.

TC TISPAN is responsible for all aspects of the wired networks like signalling protocols or systems interworking. It is currently moving towards definition and specification of the Next Generation Networks (NGN). Its architectural concept is based on a 3-level structure: Service, Control and Transport. Access to NGN shall be widely open. The first approach is the one defined for the 3GPP Release 6 Architecture called IP Multimedia Services (IMS). This common approach will ease the fixed-mobile convergence, which will allow the seamless service goal. TC TISPAN has defined a so-called NGN Project with a very ambitious aim,

* **Christian Julien**
European Telecommunications Standards Institute, Secretariat, ETSI, 650, Route des Lucioles, Sophia Antipolis, France,
E-mail: Christian.julien@etsi.org

which is to produce a first set of specifications by the end of June 2005.

TISPAN held its kick-off meeting in September, bringing together the TIPHON and SPAN cultures and expertise, and combining their respective top-down and bottom-up approaches. The first set of NGN specifications for the support of a realistic set of high priority real-time conversational multimedia services and content delivery services (e.g. Video on demand, TV program distribution) in addition to Internet-based applications is expected in the year 2005 timeframe.

Its second meeting held in December resulted in the launch of the TISPAN_NGN project, taking a pragmatic phased approach. The scope of TISPAN_NGN Release 1 was initially discussed and consensus was reached to adopt the 3GPP (SIP-based) IP Multimedia Subsystem (IMS) as a component basis for the support of NGN multimedia conversational services. It is planned to work in cooperation with the 3GPP in order to define the IMS adaptations required for the support of the xDSL IP-Connectivity access network technology. The plan is to complete TISPAN-NGN Release 1 by mid-2005 according to the Telecom industry expectations.

Its third meeting held in March 2004 resulted in the consolidation of services to be included in the NGN Release 1 and in the decision to prepare a first joint session with the 3GPP relevant groups (Service Access (SA) and Core Network (CN)). A date has been agreed on 22-23 June 2004. During this common session overview of organisations, basic rules and working methods will be presented and a way to progress together will be specified.

An aggressive TISPAN work plan is being set-up, with the hope that the ambitious goal can be met to enable initial NGN deployment steps, while aiming at smooth fixed and mobile networks convergence in terms of features supported, architectures deployed and service capabilities provided. That's a lot of work ahead of the TISPAN participants; let us be part of this great challenge of contributing to the definition of the 21st century telecom network! But what is TISPAN and how is it organised to complete successfully this task?

TISPAN is structured into 7 Working Groups (WG) dealing mainly with 5 projects. Each WG is chaired by a WG chairman and each project is under the responsibility of a Project leader. WGs are: Services, Architecture, protocols, Numbering/Addressing/Routing (NAR), Quality of Service (QoS), Testing, Security and Network Management. Five projects are: Next Generation Network (NGN), Open Service Access (OSA) Dynamic synchronous Transfer Mode (DTM), Emergency Telecommunications (EMTEL) and Fixed-Multimedia Messaging Service (F-MMS).

WG1 on Services is responsible for:
- Conducting studies leading to deliverables on new generation services and applications (from the user perspective);
- Co-ordination and preparation of service descriptions for NGN, including:
  - Inter-personal real-time services,
  - Video (on-demand, streaming, unicast and broadcast),
  - Data (Storage, processing), Messaging
  - Multimedia and group communication
  - For different market segments and communities.
- Identification of generic NGN functional requirements to support service operation.

WG2 on Architecture is responsible for:
- Network Intelligence, Universal personal and terminal Mobility, NGN architectures and Home Environment functional and architectural requirements; including the definition and the use of an overall functional architecture;
- Standards development activities related to Network architecture and its evolution, including inter-working and IP matters at the service, transport and control level;
- Studies on the functions, and the reference points required for inter-working between the emerging new architecture networks and legacy networks (IN/ISDN/PSTN);
- Definition and analysis of Functional Entities, Message Sequences and Information Elements;
- Studies on the functional requirements that Service Providers have when accessing the networks of Public Network Operators (PNO) in the fulfilment of the Special Network Access directives of the European Union (EU); including functional interfaces for Third Party Service APIs;
- Studies in the field of third generation mobility systems (e.g.: IMT-2000 and UMTS) concerning inter-working with existing and emerging IP-based Core Networks. This includes Inter-access-system inter-operability, mobility and registration.

WG3 on Protocols is responsible for:
- Defining all aspects of Protocols: protocol definition, protocol requirements, protocol mapping, protocol profiles, analysis of protocols developed by other bodies, protocol extensions and inter-working specifications.

WG4 on NAR is responsible for:
- Conducting studies leading to deliverables on: numbering, naming, addressing and routeing;
- Addressing and Naming translation;
- Representing ETSI at the ENF (European Numbering Forum);
- Collaborating with ERO (the European Radio Office) on European numbering, naming and addressing issues;
- Representing ETSI at the ETNS Steering Committee.

WG5 is responsible for:
- Conducting studies leading to deliverables on:
  a) application- and user-centric quality of service (QoS) parameters, objectives and measurement methods,
  b) network-centric performance (NP) parameters, objectives and measurement methods necessary to meet the identified QoS objectives, and
  c) determining the requirements of the protocols necessary to allocate and manage network resources to meet the identified QoS&NP objectives,

  for packet-based, circuit-switched and hybrid networks supporting multimedia services.

WG6 is responsible for:
- Management and co-ordination of the development of the testing specifications for the next generation telephony;
- Providing testing specifications for TISPAN-developed specifications and profiles;
- Maintaining existing testing specifications as required;
- Tracking ongoing worldwide bake-off, interoperability, testing and certification activities of interest to TISPAN.

WG7 is responsible for:
- Conducting studies leading to deliverables on Security;
- Management and co-ordination of the development of security specifications for the next generation telephony and multimedia communications;
- Investigation of security services and mechanisms required for providing services over the Internet;
- Development of security analyses of candidate protocols and network elements to be used within the NGN framework to implement capabilities e.g., EMTEL aspects, IPv6 migration, keying strategies and methods;
- Tracking ongoing worldwide security activities of interest to TISPAN.

WG8 is responsible for:
- Developing a consistent and harmonised approach to Telecommunications Management across all disciplines and technologies under the umbrella of the ETSI Standardisation activities;
- Telecommunications Management encompasses the management of all types of telecommunication equipment, networks, services and although biased to the Telecommunications Management Network (TMN).

So far the applied TISPAN working method is to have 4 Plenary a year where all WGs work together for progressing their relevant studies. Between each Plenary usually there is an interim meeting more dedicated to any Project where the Project activity is fasten. All WG chairmen and Project leaders met regularly into the TISPAN Management group for coordinating their respective activities either by special meeting or by regular conference call. Relationship with the other bodies is maintained by a standard liaison statement procedure.

After describing TC TISPAN, let's add some words on the TISPAN-NGN project which is the purpose of this paper.

From a technical point of view, TISPAN-NGN can be considered as a multi-service/multimedia network based upon the use of packet switching technologies. Architecturally speaking, it can be envisioned as consisting of a packet based core network with interfaces to several different access networks (e.g. xDSL / Wireless LAN etc) possibly using different underlying transport technologies.

From an application viewpoint, TISPAN-NGN release #1 supports two broad types of application:
- Session based: Examples include conversational applications such as VoIP/Multimedia calls and on-line gaming as well as non-real-time conversational applications such as messaging.
- Non-session based: Examples include streaming applications such as web-casting / broadcasting and video and audio/music on demand.

While TISPAN-NGN supports the co-existence of both session and non-session based applications on a single NGN, in the relevant standards these applications are modelled as being provided through logically separate sub-systems.

A key assumption within TISPAN-NGN is that services themselves are not standardized. Instead TISPAN-NGN defines service capabilities. In such network operators are free to combine service capabilities (potentially provided by 3rd parties) to offer specific services to subscribers. Examples of service capabilities include presence, location, non-real time data transfer (e.g. email, ftp), web-browsing, authorization mechanisms supporting single sign-on, and the capability to set up sessions or streaming media with particular characteristics in support of offering conversational and/or streaming based services.

Another important activity relevant to NGN has already been initialised within Network Management group with the aim to specify at least for the NGN Release 1, the new NGN Operating Sub System (OSS). In order to prepare this task a new Specialist Task Force (STF) has been started end of March 2004 with the active participation of Ms. T. Kovacikova from University of Zilina, Slovakia. The purpose of this group of experts is to investigate the current status of the standardization in the domain of Network Management and to identify where the lack of standards for completing the work is.

As stated today within the content of Release 1, amongst the different scenarios envisaged for introducing the NGN within the current country there are:
- PSTN/ISDN replacements,
- PSTN/ISDN interworking,
- Cable network interworking,
- 3GPP interworking,
- Internet interconnection.

Also included into Release 1 the following network accesses are envisaged:
- ADSL,
- WLAN,
- Cable network,
- Corporate networks,
- UMTS.

The content of the further Releases is not yet known in detail but integration of IPv4 into IPv6 or the full Mobility will certainly be the main issues.

The current agreed schedule for the different Releases is as follows:
- Release 1: June 2005,
- Release 2: December 2006,
- Release 3: June 2008.

Finally some figures for sizing TC TISPAN:

- Currently its work programme is made of about 142 work items, 16 standards and/or technical reports have already been published;
- There are 14 exploder lists in which more than 600 delegates coming from every part of the world are registered.
- Usual amount of input contributions per Plenary is more than 200 documents;
- More than 100 delegates attend usually any Plenary.

https://doi.org/10.26552/com.C.2004.4.98-102

Miroslav Bahleda – Karol Blunár *

THE WAVELENGTH CONVERSION IN WDM NETWORKS

In this paper we deal with the problem of wavelength conversion in WDM (wavelength division multiplex) networks. In all-optical networks the data are transmitted through optical networks transparently. It means that the data still remain in the optical domain between the source and end node. The data transmission between two nodes can be without wavelength conversion or with wavelength conversion (with full wavelength conversion or limited wavelength conversion). We want to give a comprehensive paper about recent and new forms of wavelength conversion.

1. Introduction

The WDM network is an optical network to employ the wavelength division multiplex WDM technology. This technology provides transmission of data on different optical wavelengths through the same optical fiber. We will concentrate on WDM network only, where each wavelength corresponds to a data communication channel. Moreover, we deal with all-optical networks. In these networks, the optical signal still remains in the optical domain from the source node to the destination node.

An optical fiber can carry several simultaneous wavelength channels. Each wavelength has to be different in the same fiber. The number of wavelengths that each fiber can carry is limited by the physical characteristic of the fiber and the state of optical technology, which is used to combine these wavelengths onto to the fiber and isolate them off the fiber. Today the third low loss optical window (about 1550 nm) is used for the transmission, which can support about tenths wavelengths. However, it is expected to grow rapidly in the next ten years [1].

2. Wavelength Conversion

In the case that in the networks a signal is transmitted through physical links and if the signal has to use the same wavelength, then it is called transmission without wavelength conversion. The networks, which do not enable wavelength conversion, are called the networks without conversion or the non-conversion networks. If it is possible to change a wavelength to other wavelength in the network nodes, then those networks are called networks with conversion or the conversion networks.

When we use the wavelength to routed data then it is referred to as wavelength routing. A network that employs this technique is called a wavelength-routed network. A wavelength routing network consists of two types of nodes [2]:

- optical cross-connects (OXC), which connect the fibers in the networks,
- end nodes (access or edge nodes), which provide the interface between non-optical end systems and the optical systems.

The OXC provides the switching and routing functions in order to establish the connection between edge nodes. The OXC can include wavelength converters for supporting wavelength conversion.

The wavelength conversion means the change of incoming wavelength to other wavelength. We know several types of wavelength conversion. In general, any incoming wavelength can be switched to any outgoing wavelength. The number of possible outgoing wavelength on which the incoming wavelength can be switched is k, and the number of all outgoing wavelength is W at the node. In dependence on k, we know the following types of wavelength conversion [3]:

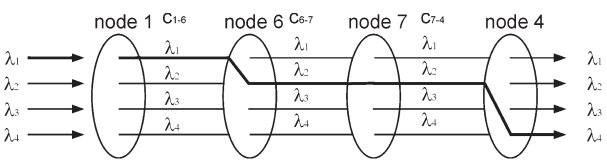


Fig. 1 Wavelength path

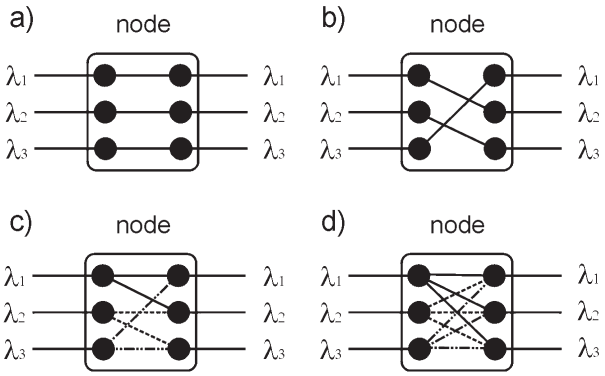


Fig. 2 Types of wavelength conversion

* **Miroslav Bahleda, Karol Blunár**
Department of Telecommunications, EF, ŽU v Žiline, Veľký Diel, 010 26 Žilina
E-mail: bahleda@fel.utc.sk, blunar@fel.utc.sk, Tel: 041 513 2227, 2204, Fax: 041 513 1520

- *no-wavelength conversion:* if the $k = 1$, where each incoming wavelength is converted only to itself (any incoming wavelength is switched to the same outgoing wavelength) (Fig. 2.a),
- *fixed wavelength conversion:* if the $k = 1$, where each incoming wavelength can be converted to other outgoing wavelength, but this outgoing wavelength is well-known (it is still the same) (Fig. 2.b),
- *limited wavelength conversion:* if the $1 < k < W$, where each input wavelength can be converted to any wavelength from a specific limited set of wavelengths (Fig. 2.c),
- *full wavelength conversion:* if the $k = W$, where any incoming wavelength can be converted to any outgoing wavelength (Fig. 2.d).

3. Network without wavelength conversion

In the networks without wavelength conversion the optical signal is transmitted from the source node to end node through all links on the same wavelength along a physical path. The session has to use the same wavelength on every fiber-hop (links). This requirement is also known as the wavelength continuity constrain.

Before the session is set up, the wavelengths, which are not occupied on all the links of the path, are found. After that we choose a wavelength out of these free wavelengths by some wavelength assignment methods. This wavelength is the same on all the links of the path. If no such a wavelength is available on all the links then the new call is blocked and, consequently, is lost. A call is accepted on a route if there exists at least one wavelength, which is simultaneously idle on all the links of that route. This means that an arrival call can be blocked if there are free wavelengths on all the links but it is not the same.

We can suppose without any mathematical analysis that the networks without wavelength conversion have the worst behavior of other networks (with full or limited wavelength conversion). The blocking probability is measured of the throughput of network and is the highest in the network with no conversion. However, the very good network routing and wavelength assignment algorithm can slightly decrease blocking probability. But these algorithms are more complicated and their complexity is significant.

In these networks the utilization of wavelengths is very small due to the wavelength continuity constrain if we want to keep the low blocking probability. If we want to increase the wavelength utilization the blocking probability will be increased very dramatically. Moreover, the blocking probability is very increased with a number of hops on a path. Therefore, we have to achieve sufficiently low utilization, for example usually about $0.1 \div 0.2$ for 10 hops and 15 wavelengths. Barry and Humblet [4] have proposed the model for evaluation wavelength utilization as a function of a number of wavelengths for the certain number of hops. There are very interesting results for the networks without wavelength conversion.

4. Network with limited wavelength conversion

The limited wavelength conversion is wavelength conversion in some node where each input wavelength can be converted to any wavelength from a specific limited set of wavelengths. We know different ways of limited wavelength conversion, which are described in next chapters.

4.1 Symmetrical limited wavelength conversion with conversion degree d

A limited range symmetrical wavelength conversion with conversion degree d is the conversion of the wavelength of the optical signal at the switch, which can convert any incoming wavelength to d adjacent wavelengths on the output side from the wavelength plane of the spectrum, as well as to the input wavelength itself (the same wavelength) [5]. The wavelength plan is an order of all outgoing wavelength. We choose the same outgoing wavelength or from d outgoing wavelength on the left or on the right side from wavelength plane. The possible outgoing wavelengths are symmetrically distributed.

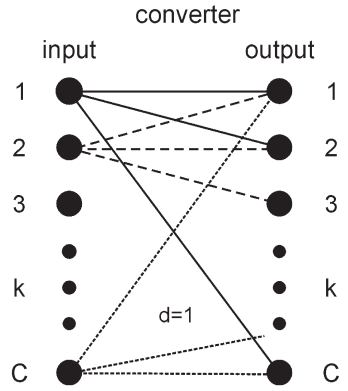


Fig. 3 The symmetrical limited wavelength conversion with conversion degree $d = 1$

This means that any wavelength can be converted to $k = 2d + 1$ outgoing wavelengths (Fig. 2). In general, if the incoming wavelength is λ_i, then this wavelength can be switched to any outgoing wavelength λ_{i-d}, ..., λ_i, ..., λ_{i+d}, where λ_i, for $i = 1, 2, 3, ...,$ n are possible outgoing wavelengths for the conversion degree d.

If $d = 0$ then it is the case without wavelength conversion. Otherwise, if d is very high it is the full wavelength conversion case. Usually $d = 1$ or $d = 2$.

This case of limited wavelength conversion is the best for the wavelength converter technology. The time needed to set up the switching elements is shorter than in other cases. It means that the technological requirements are no so complex.

4.2 Non-symmetrical limited wavelength conversion with conversion degree d

In this case, the input wavelength can be converted to the same wavelength or to one on the left side (the right side) from the wavelength plan. It means that it is possible to switch any input wavelength to $k = d + 1$ output wavelengths.

For example, if the incoming wavelength is λ_i, then the wavelength can be switched to the following set of wavelength λ_{i-d}, ..., λ_i, in the case "conversion on the left side" (Fig. 4.a), or λ_i, ..., λ_{i+d}, in the case "conversion on the right side" (Fig. 4.b).

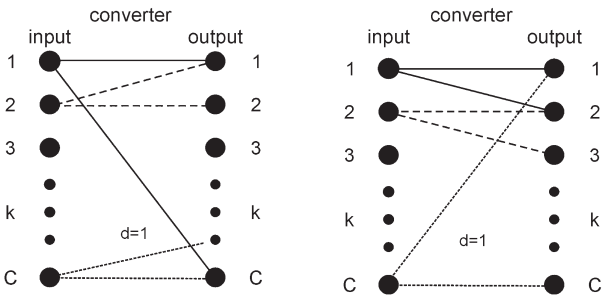


Fig. 4 The non-symmetrical limited wavelength conversion with conversion degree d = 1
a) "on the left side" b) "on the right side"

4.3 Random limited wavelength conversion

The random limited wavelength conversion is generalization of limited wavelength conversion when any input wavelength is converted to the limited set of wavelengths, which is well-known. However, the selection of wavelengths for any set of wavelength is random from the wavelength plan.

4.4 Partial limited wavelength conversion

In the standard architecture, at first, each wavelength is switched in space. Then, if possible, it is converted to the same or another wavelength. However, in networks with conversion each output port has its converter (Fig. 5).

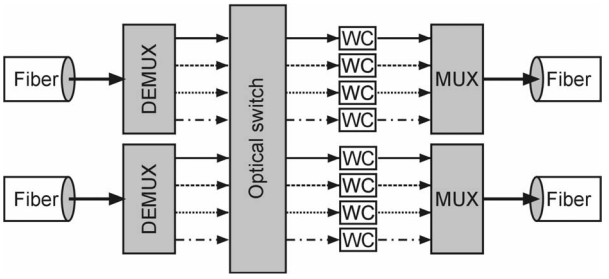


Fig. 5 Non-partial wavelength conversion

In this case, the limited number of wavelengths can use wavelength converters at the same time. There is a set of wavelength converters, which are shared by all the output ports. These converters are ordered to a converter bank. The idea of this architecture is to save the number of used wavelength converters. Therefore only the wavelengths requesting wavelength conversion are transported through the converter bank. Each wavelengths converter can provide full or limited wavelength conversion.

We know different architectures of partial limited wavelength conversion [6]:

- share-per-node wavelength-convertible switch architecture: The converter bank is shared for all the output wavelengths of any fiber. However, it means that we need the second space optical switch, which is smaller (Fig. 6).
- share-per-link wavelength-convertible switch architecture: The converter bank is shared for all output wavelengths of the fiber link. However, it means that we need more converter banks (Fig. 7).

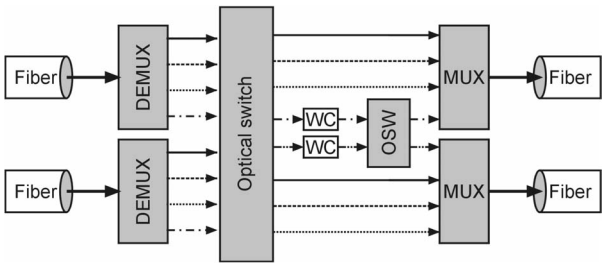


Fig. 6 Share-per-node wavelength-convertible switch architecture

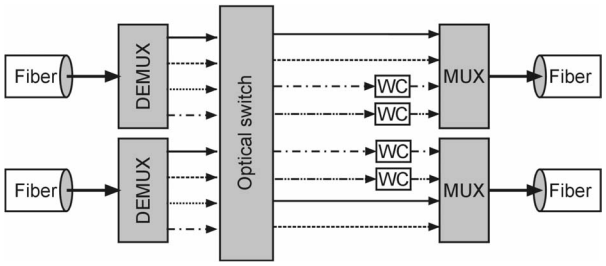


Fig. 7 Share-per-link wavelength-convertible switch architecture

4.5 Sparse limited wavelength conversion

If some nodes of the network have a wavelength converter, but not all, then it is called sparse limited wavelength conversion (Fig. 8) [7]. And the wavelengths converter can provide either full or limited wavelength conversion.

A lightpath has to use the same wavelength along several links in a segment between two nodes, which are equipped with converters. However, it is possible to use a different wavelength for a lightpath along the different segment of links.

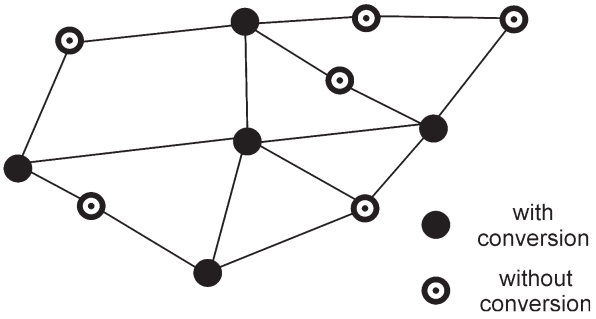


Fig. 8 Sparse wavelength conversion network

4.6 The different ways for building the networks with limited conversion

The networks with limited wavelength conversion do not enable full conversion but the conversion with some restriction is possible. This restriction can be realized at the node, where any input wavelength can be connected to a limited set of output wavelengths. This case is called the node limited wavelength conversion. Otherwise, the next restriction at the node is partial wavelength conversion. The other restriction is the limited number of nodes enabling wavelength conversion. All other nodes do not enable wavelength conversion. It is called spare wavelength conversion.

The all-optical networks with limited wavelength conversion can be divided to the following cases:

- limited wavelength conversion due to restriction at the node:
  - all nodes of WDM network use converters only with limited wavelength conversion,
  - all nodes of WDM network use converters only with partial wavelength conversion,
- sparse (limited) wavelength conversion due to restriction in the network:
  - only a few nodes employ wavelength converters with limited wavelength conversion (other nodes employ either without or full wavelength conversion),
  - only a few nodes employ wavelength converters with full wavelength conversion, other nodes do not enable wavelength conversion,
- in the case of multi-fiber WDM networks, the network do not enable wavelength, but the same wavelength can be employed on the different fiber on the same link.

Further, the network with limited wavelength conversion can originate as a combination of the limited wavelength conversion due to restriction at node with sparse wavelength conversion due to restriction in the network.

5. Network with full wavelength conversion

In networks with full conversion, a full conversion is enabled to all nodes. It means that any incoming wavelength can be changed to any outgoing wavelength in all switches.

In that case we choose a wavelength out of the free wavelengths on the first hop by some wavelength assignment algorithm. This is repeated on the subsequent hops of the path. In case there is no free wavelength on some link of the path the arrival call is blocked and lost [8].

It is clear that these networks are the most flexible of other cases and these networks have the lowest loss. However, full conversion networks are the most technologically difficult because optical switches have to enable to switch any input with every output. This switching has to be very fast and it is still a complex problem today.

It is clear that the networks with full wavelength conversion have the best performance. However, it has been found that good routing policy in network without wavelength conversion can decrease the blocking probability. And moreover, the network with limited wavelength conversion has the blocking probability very close to blocking probability of network with full wavelength conversion. The network with the sparse wavelength conversion uses far less converters than the networks with full wavelength conversion. But if we use good placement algorithms for location converters the performance of sparse wavelength conversion is similar to the one in the case of networks with conventional wavelength conversion.

6. Conclusion

We presented different possibilities of wavelength conversion in WDM network. We wanted to provide you with an overview of the current state and new promising technology in this area. Moreover, we want to indicate how blocking probability depends on the wavelength conversion and which case is the best in term of blocking probability. But which is also important is effectiveness and value.

The present state of technology does not provide the full wavelength conversion in the real networks because it is very expensive. The networks with limited wavelength are very popular and probably they will become more used in the future. There are many possibilities how to apply the limited wavelength conversion into the real networks. If the sparse wavelength conversion is used, it is not necessary to equip each node of the network with the wavelength converter. And if the partial wavelength conversion is used, it is not necessary to place as many converters into node as in case of the full conversion. Both approaches mentioned save the wavelength converter as the most expensive part of WDM switch. And if the limited wavelength conversion is used in the node, we do not need so complicated wavelength converters. It means that all the restrictions mentioned save financial resources.

In future, we will study the limited wavelength conversion in more detail and we will design the network model for computing blocking probability in WDM networks. Moreover, we want to determine the optimal placement for the wavelength converters in networks with sparse limited wavelength conversion.

References

[1] GIRARD, A.: *"Guide To WDM Technology & Testing",* EXFO Electro-Optical Engineering Inc., Quebec City, Canada, 2000
[2] ROUSKAS, G. N.: *"Routing and Wavelength Assignment in Optical WDM Networks",* Wiley Encyclopedia of Telecommunications, John Wiley & Sons, 2001
[3] RAMAMURTHY, B., MUKHERJEE, B.: *"Wavelength Conversion in WDM Networking",* Journal of Selected Areas in Communication, Vol. 16, September 1996, pp. 1061–1073
[4] BARRY, R. A., HUMBLET, P. A.: *"Models of Blocking Probability in All-Optical Networks with and Without Wavelength Changers",* IEEE Journal on Selected Areas in Communication, Vol. 14, No. 5, June1996, pp. 858–867
[5] TRIPATHI, T., SIVARAJAN, K. N.: *"Computing Approximate Blocking Probabilities in Wavelength Routed All-Optical Networks with Limited-Range Wavelength Conversion",* Proceedings, IEEE INFOCOM '99, pp. 329–336
[6] Chu, X., Liu, J., Zhang, Z.: *"Analysis of Sparse-Partial Wavelength Conversion in Wavelength-Routed WDM Networks",* IEEE INFOCOM '04, Hong Kong, March 2004
[7] SUBRARNANIAM, S., AZIZOGLU, M., SOMANI, A. K.: *"All-Optical Networks with Sparse Wavelength Conversion",* IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 4, NO. 4, AUGUST 1996
[8] SRIDHARAN, A., SIVARAJAN, K. N.: *"Blocking in All-Optical Networks",* IEEE INFOCOM 2000.

https://doi.org/10.26552/com.C.2004.4.103-106

komunikácie COMMUNICATIONS

Peter Vestenický *

SOLUTION OF SOME TECHNICAL PROBLEMS IN MARKER AND MARKER LOCATOR DEVELOPMENT

This paper gives a short description of selected technical problems and their solutions in development of a marker locator. The principle of underground engineering network equipment marking by markers and principle of marker location is explained. Next, mathematical analysis of signal intensity received from an excited marker and its dependence on distance is described. Moreover, influence of over current induced by lightning into a marker is analyzed and practically measured. These facts were used to increase reliability and lifetime of markers.

1. Introduction

At present large amount of technical equipment which create engineering networks are hidden under the ground surface, for example telecommunication and power cables, gas and water pipes, sewerage pipes etc. Construction of new underground networks, their repairs, construction of other objects, earthworks etc. require precise marking out of the engineering networks in a given place. Combinations of geodetic and electromagnetic methods are used to determine the position.

One of frequently used electromagnetic methods of the position determining is the marking by markers. Essentially, the marker is a resonant circuit, which is tuned on the concrete frequency different for different marked equipment (83 kHz for gas pipes, 101.4 kHz for telecommunication cables, 121.6 kHz for sewerage pipes, 145.7 kHz for water pipes and 169.8 kHz for power cables). The marker is buried under the ground together with engineering network equipment and marks important points (connection, turning, crossing, point of reparation) or route. The marker can be whenever localized by special device – locator.

2. Principle of marker – locator system operation

The marker is resonant circuit connected according to Fig. 1. The locator periodically transmits an excitation signal which excites damped harmonic oscillations in the marker. Frequency of damped oscillations is given by a marker tuning and this signal is received and evaluated by a locator (Fig. 2).

Signals recorded by a digital oscilloscope directly in the locator are depicted in Fig. 3, where trace #1 shows the excitation current in the locator antenna and trace #2 shows received response from the marker, i.e. damped harmonic oscillations. The marker is localized by a method which is shown in Fig. 4.

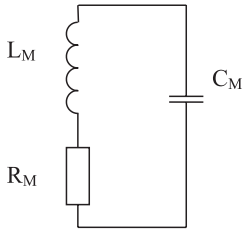


Fig. 1. Connection of marker

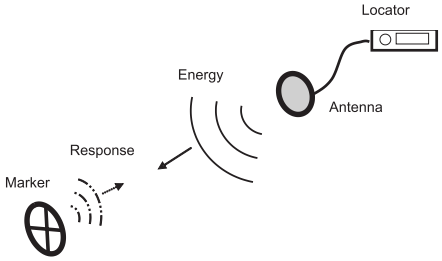


Fig. 2. Principle of marker - locator system operation

3. Dependence of received signal from marker on distance from locator

The transmitting antenna of locator is essentially a circular loop coil with radius r_{LT} (with area S_{LT}) and with N_{LT} turns. The excitation current I_L flows by this coil and excites magnetic field. Intensity of the magnetic field in the distance x from the antenna and in its axis direction is given (according to [1]) by formula

$$H_L(x) = \frac{N_{LT} I_L S_{LT}}{2\pi (r_{LT}^2 + x^2)^{3/2}} \quad (1).$$

* **Peter Vestenický**
Department of Control and Information Systems, Faculty of Electrical Engineering, University of Žilina. Veľký diel, 010 26 Žilina, Slovakia.
E-mail: peter.vestenicky@fel.utc.sk, Tel.: +421-41-5133345

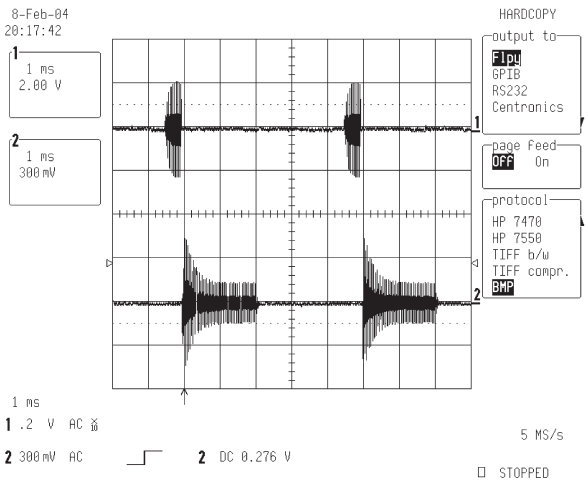


Fig. 3. Excitation and response of marker

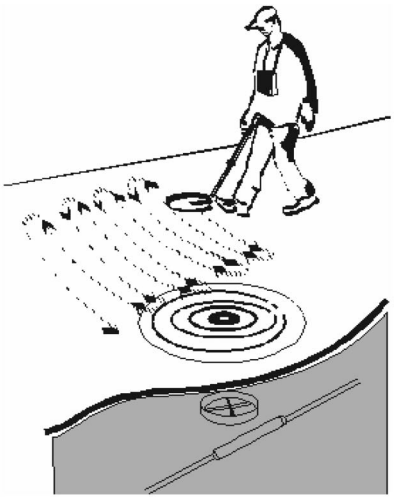


Fig. 4. Principle of marker localization

Assuming that the excitation current is harmonic and has frequency ω in the marker coil which has N_M turns, area S_M, inductivity L_M and its axis is identical with the locator antenna axis the voltage

$$U_M = N_M \frac{d\Phi(x)}{dt} = \mu\omega N_M S_M H_L(x) \tag{2}$$

is induced, where $\Phi(x)$ is magnetic flux and μ is environment permeability. This voltage causes the current

$$I_M = Q_M \frac{U_M}{\omega L_M} = Q_M \frac{\mu N_M S_M H_L(x)}{L_M} \tag{3}$$

in the marker coil, where Q_M is the quality factor of marker resonant circuit. The current I_M creates magnetic field with intensity H_M, which in receiving locator antenna place (i.e. again in distance x from marker) induces the voltage U_L in receiving locator coil. These magnitudes can be calculated from next formulas:

$$H_M(x) = \frac{N_M I_M S_M}{2\pi(r_M^2 + x^2)^{2/3}} \tag{4}$$

$$U_L = \mu\omega N_{LR} S_{LR} Q_{LR} H_M(x) \tag{5},$$

or, after combining (1), (3), (4) and (5)

$$U_L = \frac{\mu^2 \omega N_{LR} N_M^2 N_{LT} Q_{LR} Q_M S_{LR} S_M^2 S_{LT} I_L}{4\pi^2 (r_{LT}^2 + x^2)^{3/2} \cdot (r_M^2 + x^2)^{3/2} \cdot L_M} \tag{6},$$

where separate symbols denote:

N_{LR} receiving locator coil number of turns
S_{LR} area of receiving locator coil
Q_{LR} quality factor of locator input tuned circuit.

The marker voltage U_M and receiving coil voltage U_L calculated from (2) and (6) are graphically depicted in figures 5 and 6 in dependence on distance x. For calculation these parameters were used:

- marker parameters
  - number of turns $N_M = 18$
  - inductivity $L_M = 169\ \mu H$
  - coil radius $r_M = 0.1\ m$
  - coil area $S_M = 0.0628\ m^2$
  - quality factor $Q_M = 70$
- transmitting locator coil parameters (shall be chosen subject to [2])
  - number of turns $N_{LT} = 30$
  - coil radius $r_{LT} = 0.1\ m$
  - coil area $S_{LT} = 0.0628\ m^2$
  - excitation current $I_L = 0.2\ A$
- receiving locator coil parameters
  - number of turns $N_{LR} = 50$
  - coil radius $r_{LR} = 0.075\ m$
  - coil area $S_{LR} = 0.03534\ m^2$
  - quality factor $Q_{LR} = 15$
- other parameters
  - permeability $\mu = 4\pi.10^{-7}\ H/m$
  - frequency $\omega = 637\ krad/s$ ($f = 101{,}4\ kHz$).

As (2) depends on (1), it is obvious that the marker induced voltage depends reverse proportionally on the third power of distance x. Similarly, it is obvious that voltage on the receiving coil, i.e. induced back from excited marker depends reverse proportionally on the sixth power of distance x. Moreover, these simplified calculations do not consider that the amplitude of marker voltage falls exponentially with time (damped harmonic oscillations of marker).

4. Calculation of voltage induced into marker by current shock wave

Markers can mark not only non-metallic, but also metallic pipes and cables in an open field. For this reason we must consider the influence of the current waves which spread in a metal conductor on the current and voltage loading of marker components. This shock wave can be created by lightning current or short circuit in power cable. Amplitude of lightning current can achieve up to 50 kA [3]. Considering the ground and metallic conductor

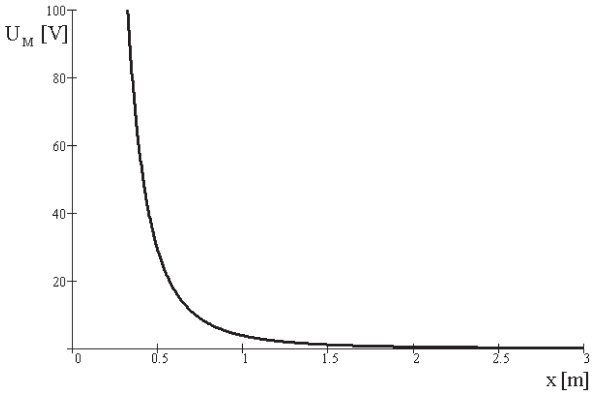


Fig. 5. Voltage on marker

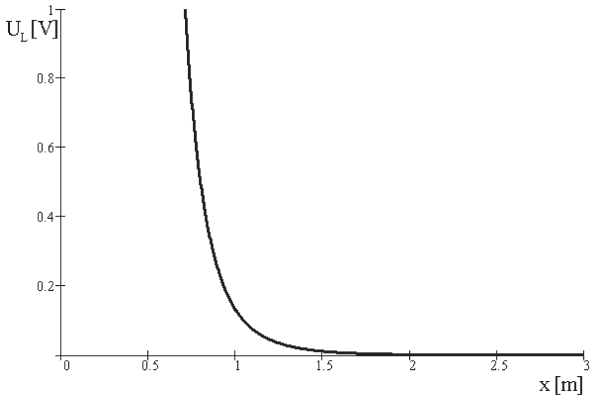


Fig. 6. Voltage on receiving coil

damping factor, we assume about current wave with amplitude 2 kA.

Time dependence of such shock wave corresponds approximately with Fig. 7 and can be mathematically modelled by subtraction of two exponentials [3]

$$i(t) = I(e^{-\frac{t}{t_2}} - e^{-\frac{t}{t_1}}) \tag{7},$$

where T_h is front time, T_m is half-rear time and I_{MAX} is maximum value of current in the time T_h. For measuring and calculation the current wave with $T_h = 8\ \mu$s and $T_m = 20\ \mu$s are being used. Corresponding parameters t_1, t_2 and I must be calculated from a nonlinear equation system

$$\frac{t_1 t_2}{t_2 - t_1} \cdot \ln\left(\frac{t_2}{t_1}\right) = T_m \tag{8a}$$

$$I\left(e^{-\frac{T_h}{t_2}} - e^{-\frac{T_h}{t_1}}\right) = I_{MAX} \tag{8b}$$

$$I\left(e^{-\frac{T_m}{t_2}} - e^{-\frac{T_m}{t_1}}\right) = \frac{I_{MAX}}{2} \tag{8c}.$$

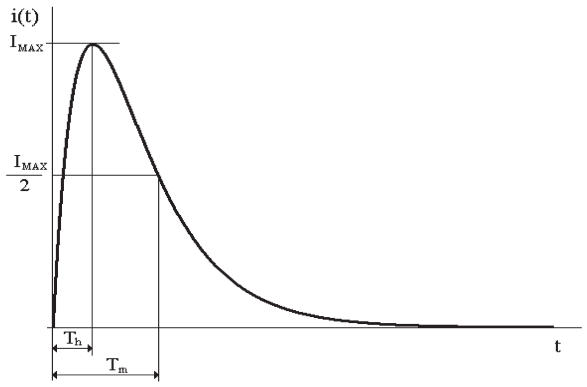


Fig. 7. Shape and parameters of current wave

The numerical solution of the system (8) for given $T_h = 8\ \mu$s, $T_m = 21.5\ \mu$s (if $T_m = 20\ \mu$s the system diverges) and $I_{MAX} = 2000$ A gives the parameters of shock wave model: $t_1 = 7.18\ \mu s$, $t_2 = 8.95\ \mu s$ and $I = 24.7\ kA$.

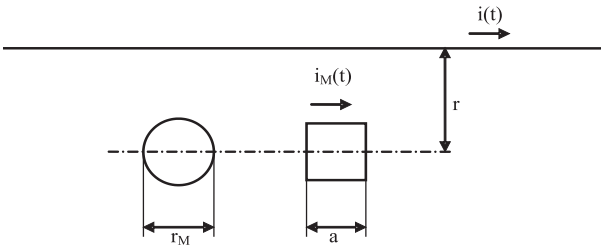


Fig. 8. Arrangement of conductor and marker

We assume that the markers of circular and square shape lay in the distance r from a straight infinitely long conductor by which the current wave i(t) with parameters described above flows (Fig. 8). The current i(t) creates in marker area magnetic flux

$$\Phi(t) = \int_{S_M} \mu N_M \frac{i(t)}{2\pi x}\, dS_M = \int_{r-\frac{a}{2}}^{r+\frac{a}{2}} \mu N_M \frac{i(t)}{2\pi x}\, a \cdot dx =$$

$$= \frac{\mu a N_M}{2\pi} \cdot \ln\frac{2r + a}{2r - a}\, i(t) = M i(t) \tag{9},$$

where the circular marker was replaced by a square marker with equivalent area for simplifying the surface integral calculation,

$$a = \sqrt{\pi} \cdot r_M \tag{10}$$

is length of equivalent square marker side and

$$M = \frac{\mu a N_M}{2\pi} \cdot \ln\frac{2r + a}{2r - a} \tag{11}$$

is mutual inductivity of a straight conductor and marker coil.

As the marker is essentially resonant circuit, the time dependence of current $i_M(t)$ induced into the marker circuit by influence of current wave $i(t)$ (formula (7)) can be calculated from a differential equation

$$L_M \frac{d^2 i_M(t)}{dt^2} + R_M \frac{di_M(t)}{dt} + \frac{1}{C_M} i_M(t) = M \frac{d^2 i(t)}{dt^2} \qquad (12)$$

with respect of initial conditions $i'_M(0) = 0$ and $i_M(0) = \frac{M}{L_M} I \cdot \left(\frac{1}{t_1} - \frac{1}{t_2}\right)$. Voltage on the marker capacitor is given by formula

$$u_c(t) = \frac{1}{C_M} \int_0^t i_M(\tau) d\tau \qquad (13).$$

The equations (12) and (13) were solved for next parameters:

- marker coil inductivity $L_M = 1\ mH$
- marker coil resistance $R_M = 10\ \Omega$
- number of turns $N_M = 54$
- condensator capacity $C_M = 1.621\ nF$
- distance of marker from conductor $r = 0.3\ m$
- length of equivalent square marker side $a = 0.177\ m$
- mutual inductivity of marker coil and conductor $M = 1.162\ \mu H$
- current wave $i(t)$ given by formula (7) with parameters $t_1 = 7.18\ \mu s$, $t_2 = 8.95\ \mu s$, $I = 24.7\ kA$

and the result of solution is graphically shown in Fig. 9.

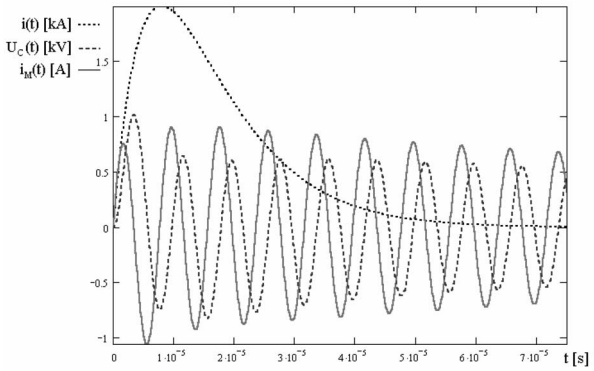


Fig. 9. Time shapes of current wave i(t), marker capacitor voltage $u_C(t)$ and marker current $i_M(t)$

The current induced in the marker by current wave was experimentally measured with utilization of a pulse generator generating the current pulse with parameters $T_h = 8\ \mu s$, $T_m = 20\ \mu s$, $I_{MAX} = 2000\ A$. The current shape was recorded by a digital oscilloscope as a voltage fall on 0.47 Ω resistor and is in Fig. 10. The

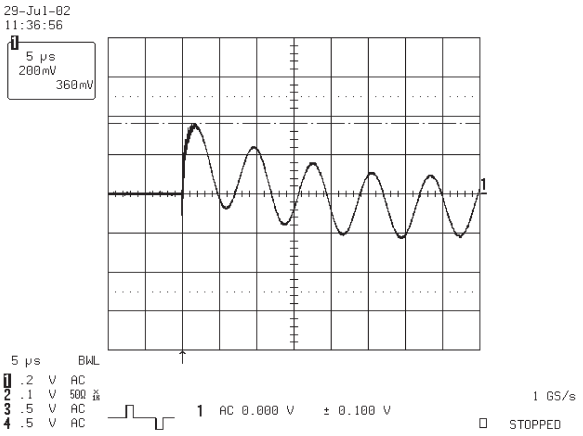


Fig. 10. Really recorded time dependence of marker current represented by voltage on 0.47 Ω resistor

maximum value of current recalculated from voltage (360 mV) and resistance (0.47 Ω) is 0.76 A.

5. Conclusion

The mathematical calculations described in this paper were used in marker locator development, which was realized by the Department of Control and Information systems for the Komplex company as the contract no. EF/28/97.

The calculation of voltage induced from an excited marker back to the receiving antenna showed that this voltage is reverse proportionally dependent on the sixth power of distance. This fact limits practically achievable distance from which the marker can be localized to the value of 2.5, maximum 3 m.

The calculation and measuring of voltage and current in the marker induced from current wave flowing by a metal conductor of marked equipment showed that these influences cannot be neglected. For increasing the marker reliability and lifetime (marker lifetime must be greater than 40 years) the marker capacitor must be high voltage and current type. Therefore the capacitors with working voltage of 1500 V are used. Similarly, insulation of the marker coil conductor must be carefully chosen.

At present the development of a new marker type based on RFID (Radio Frequency Identification) technology is in progress. The RFID technology enables to supplement new function of marker – identification of underground device by a numeric code.

References

[1] GEHRING, U., ROZ, TH.: *RFID made easy. EM Microelectronic-Marin SA.* Marin, Switzerland, 2000
[2] ETSI EN 300330-1 (V1.3.2) *Electromagnetic compatibility and Radio spectrum Matters (ERM)*; Short Range Devices (SRD); Radio equipment in the frequency range 9 kHz to 25 MHz and inductive loop systems in the frequency range 9 kHz to 30 MHz; Part 1: Technical characteristics and test methods
[3] ROUS, Z.: *Overvoltage Protection of telecommunication lines and equipment* (in Czech). NADAS, Praha, 1981

https://doi.org/10.26552/com.C.2004.4.107-109

Radoslav Odrobiňák – Milan Dado – Róbert Jankovský *

OPTICAL SOA-BASED SWITCHING

All-optical switches are fundamental building blocks for high-speed optical networks and basic blocks towards all-optical networks for future communications that utilize Optical Time Division Multiplexing (OTDM) techniques of signal with single channel data rate of hundreds Gbps. It seems that the most promising technology will be based on Semiconductor Optical Amplifiers (SOA). We can arrange these components to some interferometric architectures in which SOAs achieve efficient optical switching rate with low control energy and big optical sampling bandwidth. The interferometric optical switch geometries are presented and characterized in this paper. We discuss their limitation on temporal width of switching window and some of their applications.

1. Introduction

In fiber optic communications we can see a continuing growth in transmission capacity and speed basically due to advances in development of optical components and systems in a physical layer (new components are still needed for optical signal processing), for example laser diodes – DFB-lasers, DBR-lasers, Mode-locking lasers...; new hybrid optical amplifiers – Raman/EDFA; and so on. Further, we can see that development in optical components leads up to miniaturization and integration and, therefore, in the future we will prefer those devices witch can satisfy this condition while remaining cost competitive. For future generation of optical networks to utilize the full bandwidth of optical fiber we expect such systems (Dense Wavelength Division Multiplexed – DWDM), which will have real bit-rate of individual channels out of range of electronics components. Now practically whole signal processing with optical stream is performed in electrical domain or rarely in hybrid devices – Fig. 1.

In such devices we have to convert an input optical signal to electrical domain where we can perform required processing and then convert it back to an optical signal. The odds are that such devices have their disadvantages (this double conversion is time-consuming and expensive). Hence we want to overcome or reduce O-E-O conversion. Illustration of such processing can be seen in figure 2.

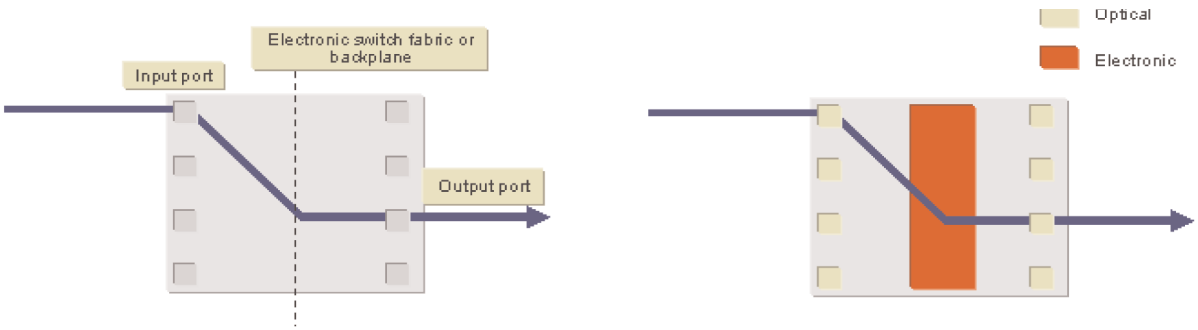


Fig. 1. Illustration of electronic and hybrid optical signal processing

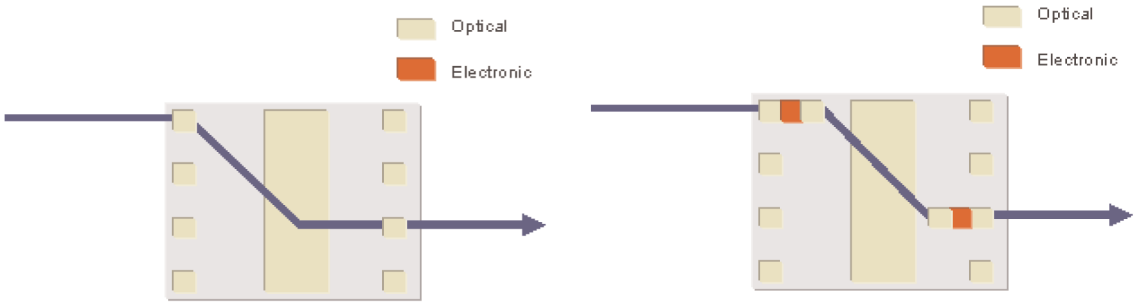


Fig. 2. Illustration of all-optical and hybrid signal processing

* **Radoslav Odrobiňák, Milan Dado, Róbert Jankovský**
Department of telecommunications, University of Zilina, Veľký diel, 010 26 Žilina, Slovakia, E-mail: odrobinak@fel.utc.sk

In the present time, the developers produce many architectures for all-optical signal processing (but not yet in optimized fabricable application) and SOAs with long recovery times (> 100 ps) seem to demonstrate efficient interferometric all-optical switches to deliver switching and demultiplexing on Tbps data-rate streams. These nonlinearities are typically based on resonance excitation in an actively biased optical amplifier or passive semiconductor nonlinear waveguides. Extensive experimental [1–6] and theoretical [6] analyses have been performed on various configurations of these devices.

2. SOA architectures

In this section, we review some of the most promising applications of ultra-fast, SOA-based optical switches under investigation at the University of Zilina. We introduce the architectures for all-optical switches and characterize their performance.

where c_{fiber} is speed of light in fiber.

Using a similar operating principle it is possible to form other interferometric configuration based on SOA. These architectures improve the integerability and performance of the device, although they require active stabilization if constructed from discrete components. Two variations of the switch in Mach-Zehnder interferometer (MZI) configuration are shown in figure 3. In the absence of the control signals, the MZI is balanced so that the data signals are rejected from the output port. When control pulses are injected into the interferometer, a differential phase shift is briefly introduced between the two arms of the interferometer causing the data pulse to be switched to the output port. The advantage of the Colliding Pulse MZI (CPMZI) is requirement of no output filter. The nominal switching window of CPMZ and SMZ are:

$$\tau_{win} = \frac{2\Delta x_{SOA}}{c_{fiber}} \qquad \tau_{win} = \Delta t_{cs} \qquad \text{respectively.}$$

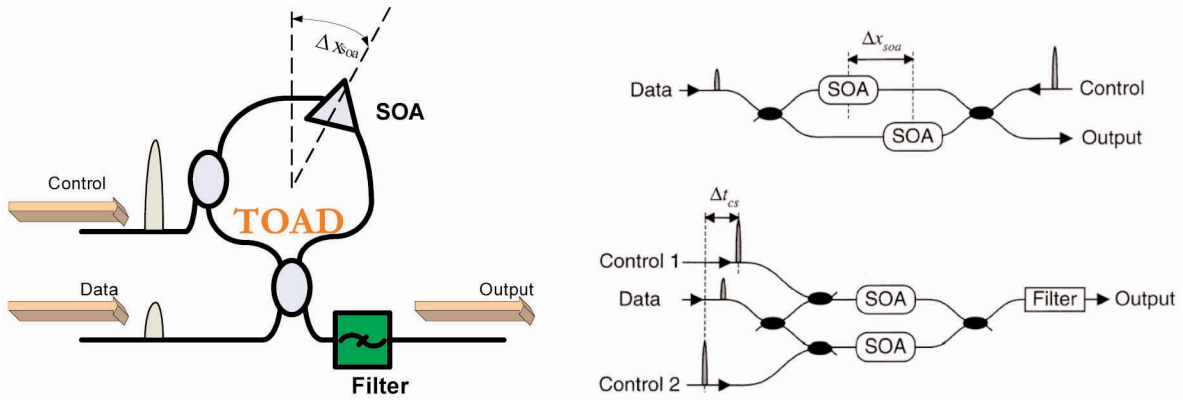


Fig. 3. The main SOA-based architectures: TOAD, CPMZ and SMZ

The main SOA-based architectures are (figure 3): TOAD – Terahertz Optical Asymetric Demultiplexer, CPMZ – Colliding Pulse Mach-Zehnder and SMZ – Symmetric Mach-Zehnder. The TOAD is based on Sagnac interferometer. It was discovered that the temporal width of the switching window could be linearly controlled by adjusting the displacement of the SOA from the midpoint of the loop. Due to dynamics of this configuration, the switching window actually closes earlier than the recovery time of the SOA as the SOA is moved closer to the midpoint. In the absence of a control pulse, data pulses enter the fiber loop, pass through the SOA at different times as they counter-propagate around the loop and recombine interferometrically at the coupler (data is reflected due to same changes in SOA). In the presence of control pulse switching can occur, because control pulse saturate SOA and change its index of refraction, therefore we can achieve different phase shift between two counter-propagating data pulses. The temporal duration of the switching window is determined by the offset of the SOA, Δx_{SOA}, from the center position of the loop.

$$\tau_{win} = \frac{2\Delta x_{SOA}}{c_{fiber}}$$

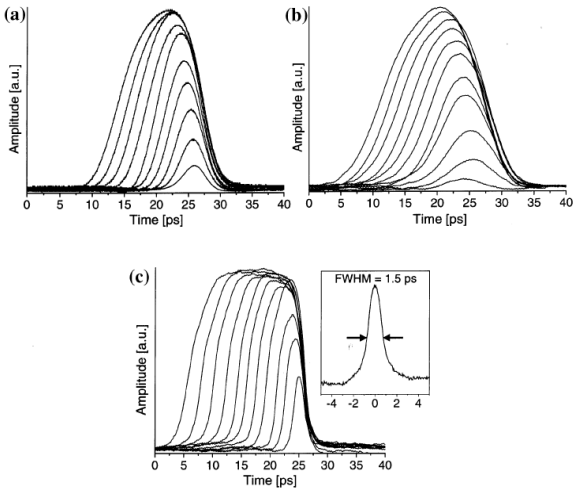


Fig. 4. Optical switching window width simulation results, a) TOAD, b) CMPZ and c) SMZ architecture

3. Simulation of the switches

While many theoretical models have been developed to understand the ultrafast temporal response of SOAs [7–9] we present a summary of the simulated performance of the three geometries introduced in the previous section. The switching windows are acquired by using a scanning pump-probe apparatus described in [6].

The switching window provides information regarding the shape, amplitude, and temporal width of the optical transfer function. This characterization is useful for determining the optical demultiplexing and sampling bandwidth of the switch. For the results described here, various pulse energies, width, and repetition rates were used to demonstrate the effect, where these parameters affect the switching performance.

The shortest switching window achieved by TOAD was about 3.8 ps (FWHM), by CPMZ about 8 ps and by SMZ even 1.5 ps. This performance indicates that the SMZ structure may be suitable for demultiplexing from 660 Gbps data stream. Reducing the optical switching window further to subpicosecond regimes may be difficult due to SOA gain compression. Further investigation and measurements of these dynamics are important for future research.

4. Summary

In this paper we presented new architectures as TOAD, CPMZ and SMZ interferometrical switches. We briefly introduced their performance and operating principle. Using mathematical simulation in VPIprofessionalTool® software we investigated their parameters such as optical temporal window width. We demonstrated that SMZ architecture is ultrafast operating switch able to demultiplex 660 Gbps data stream.

References:

[1] EISELT, M., PIEPER, W., WEBER, H. G.: *"All-optical high speed demultiplexing with a semiconductor laser amplifier in a loop mirror configuration",* Electronics Letters 29, str. 1167, 1993.

[2] SOKOLOFF, J. P., PRUCNAL, P. R., GLESK, I., KANE, M.: *"A Terahertz Optical Asymmetric Demultiplexer (TOAD)",* IEEE Photon. Technol. Lett. 5: 787–790, 1993

[3] ELLIS, A. D., PATRICK, D. M., FLANNERY, D., MANNING, R. J.: *"Ultra-high-speed OTDM networks using semiconductor amplifier-based processing nodes",* Journal of Lightwave Technol. 13, 761, 1995

[4] PATEL, N. S., HALL, K. L., RAUSCHENBACH, K. A.: *"40-Gbit/s cascadable all-optical logic with an ultrafast nonlinear interferometer",* Optics Letters 21, p. 1466, 1996.

[5] NAKAMURA, S., UENO, Y., TAJIMA, K.: *"Ultrafast (200-fs switching, 1.5-Tb/s demultiplexing) and high-repetition (10 GHz) operations of a polarization-discriminating symmetric Mach-Zehnder all-optical switch",* IEEE Photonics technology Letters 10, 1575, 1998.

[6] TOLIVER, P., RUNSER, R. J., GLESK, I., PRUCNAL, P. R.: *"Comparison of Three Nonlinear Interferometric Optical Switch Geometries. Opt. Commun.",* vol. 175: 365–373, 2000

[7] HALL, K. L., LENZ, G., DARWISH, A. M., IPPEN, E. P.: *"Subpicosecond Gain and Index Nonlinearities in InGaAsP Diode Lasers"* Optics Commun. 111, 589–612, 1994.

[8] HONG, M. Y., CHANG, Y. H., DIENES, A., HERITAGE, J. P., DELFYETT, P.J.: *"Subpicosecond pulse amplification in semiconductor laser amplifiers: theory and experiment"*, IEEE Journal of Quantum Electron., 30, p. 1122, 1994

[9] TANG, J. M., SHORE, K. A.: *Strong picosecond optical pulse propagation in semiconductor optical amplifiers at transparency,* Journal of Quantum Electronics 34, 1263, 1998.

https://doi.org/10.26552/com.C.2004.4.110-112

Igor Mihalik *

SPEECH COMPRESSION ALGORITHM BASED ON NON-EQUIDISTANT SAMPLING

The method presented here is one of the methods of time domain compression. The technique uses non-equidistant sampling. Human voice and voice conversation are characterized by transitions. During the speech frequency charactericstics change so that viariable sampling can be applied. The nonuniform sampling method which is implemented is described in this paper; its usage in speech synthesis software which is being developed at the Department of Info-Com Networks is shown.

1. Algorithm

There are several approaches to speech compression. One can divide them into 3 basic groups as shown in Fig. 1: *waveform coders* (PCM, ADPCM, ADM, DPCM), *vo-coders* (ITU standardized: G.728, G.729, G.723.1) and *hybrid coders*.

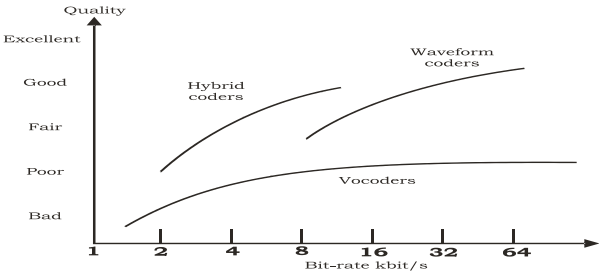


Fig. 1. Code types and quality/bit-rate dependency

The method presented in the article belongs to the group of waveform coders. Nyquist Theorem is the primary theory to keep in mind. Suppose the highest frequency component, in hertz, for a given analogue signal is f_{max}. According to the Nyquist Theorem, the sampling rate must be at least $2f_{max}$; or twice the highest analogue frequency component. The common techniques use constant sampling frequency. PCM [2], for example, uses 8 kHz sampling; this method is used in PSTN. Human speech is characterized by transitions of silence and voice. During the voice period the frequency characteristics change as show in Fig. 2. Some parts of speech contain higher frequencies; so that higher sampling frequency is required, some contain low frequencies; in that case lower sampling rate is required. This observation suggests dividing the speech signal into intervals and applying a different sampling rate to each. To effectively use transmission channels the suitable sampling frequency should be selected. The optimum is to divide the speech signal into intervals and determine the sampling frequency for each using frequency analysis to minimize the amount of transmitted data and keep the quality at a certain level.

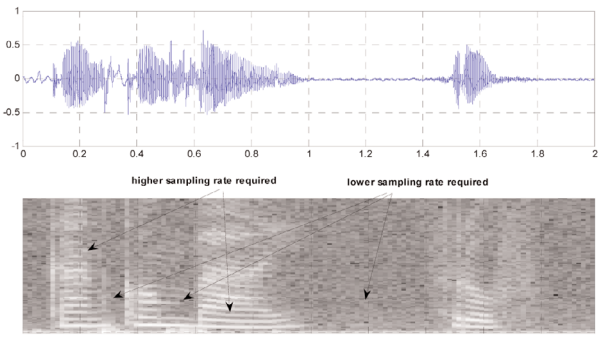


Fig. 2. Human speech: Frequency characteristics change over time. Parts of speech containing higher frequencies need to be sampled using higher sampling rates, parts containing low frequencies need to be sampled using lower sampling rates.

In the next part a method that changes sampling frequency adaptively using a mask is presented. The method works strictly in time domain and a new approach is examined.

2. Mask

The algorithm is based on a mask creation; the Mask *M* is characterized by several properties. The illustration of this mask is shown in Fig. 3. The Mask consists of a set of points, each characterized by its ID and [X, Y] coordinates:

```
typedef T_MASKPOINT struct {
        int ID;
        int X;
        int Y;
};
```

* **Igor Mihalik**
Department of Information Networks, FRI, University of Žilina, Veľký diel, 010 26 Žilina, Slovakia

$$M = \{b_1, b_2, b_3, ..., b_n\} \quad (1)$$

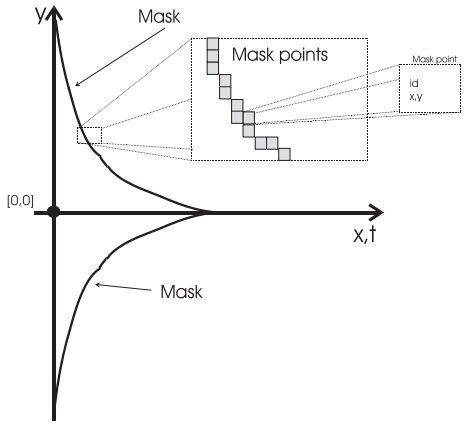


Fig. 3. The Mask consists of a set of points each characterized by its unique ID and coordinates.

In one step one point of the mask is chosen. This point relatively characterizes the signal. The effort is to select such a sequence of points so that by using them the original signal can be accurately reconstructed. The technique is demonstrated in Fig. 4. It shows input signal and its intersection with the mask at points 1, 2 and 3. Each intersection has a point with a specific ID. The sequence of IDs is then transmitted. The decompression part uses the same mask and using the IDs tries to reconstruct the original signal.

3. The Shape of Mask

It is possible to determine the shape of mask experimentally, so that an acceptable compromise between the quality and the compression ratio is achieved. Signal to Noise Ratio is used to measure the quality; for continuous signals it's defined as (2); $f(t)$ stands for an original signal, $f'(t)$ stands for a reconstructed signal. SNR represents the ratio between the energy of the signal and the energy of the noise.

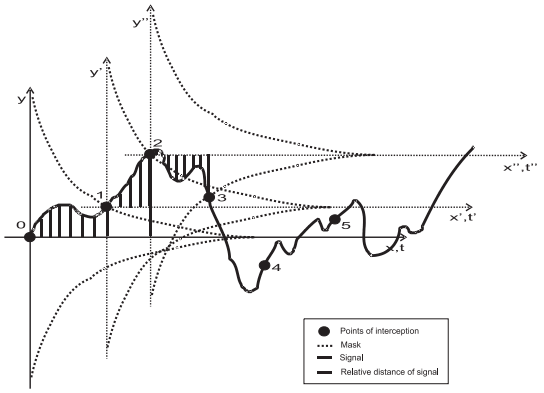
$$SNR = 10 * \log \frac{\int_0^T f^2(t)dt}{\int_0^T (f(t) - f'(t))^2 \, dt} \quad (2)$$

Compression ratio R (2) measures the ratio between the size of an original data and the size of a compressed data.

$$R = \frac{|audio\_data|}{|compressed\_data|} \quad (3)$$

4. The Results

The principle of determining the shape of the mask is based on random generation. For each generated shape (mask instance) the compression ratio and the SNR is measured. The resulting graph is shown in Fig. 5. It shows the dependency between quality and compression rate. The masks with IDs of bit-size 4, 5, 6, 7 and 8 were generated. It was not possible to generate all the possible shapes due to the time complexity, therefore an adaptive algorithm was developed [8]. The number approximates $\approx 256^{256}$ for 8-bit



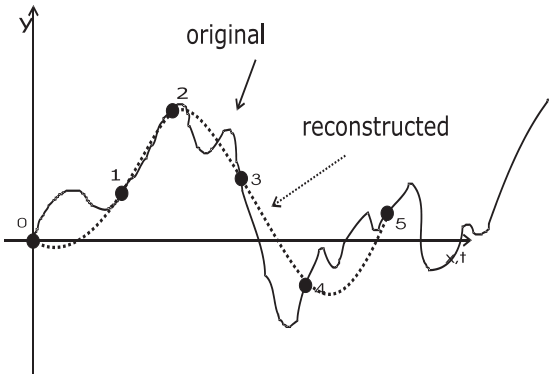


Fig. 4. A compression and a signal reconstruction

There are some elementary mask properties. In the first place, it has to be guaranteed the mask covers the signal so that the signal never moves beyond the mask boundary. The next property is the size of the mask; i.e. the number of points. The more points there are, the more IDs we need; so that we need more bits to encode the IDs for transmission. The shape of the mask is the highest factor influencing the resulting quality and the compression ratio. The method used during the signal reconstruction has a high influence, too.

IDs. For 8-bit IDs we get 256 points. Each point can be anywhere within the coordinates. The case with more than one point placed on the same coordinate eliminates the final number of possibilities. But it does not reduce the number as radically as to be able to examine all of them. For the generation exactly two points to be generated having the same X coordinate (above the X axis, below the X axis) are supposed. The mask has to be continuous and at least two points are generated with X coordinate equal to zero.

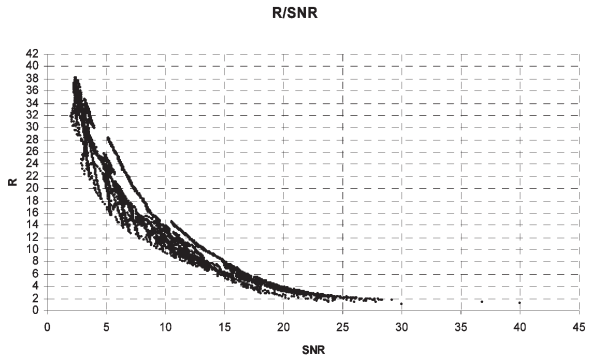


Fig. 5. Each point in graph represents one instance of mask. Each instance is characterized by achieved SNR and compression ratio (R). R/SNR dependency diagram gives an overview of possibilities of presented approach

5. Speech synthesis (TTS)

One of the methods of speech synthesis is based on usage of speech units called diphones. Using the diphones the resulting speech is created, units are concatenated together. This method is known as the concatenate method. The text-to-speech system developed at the Department of Info-Com networks uses this concatenate technique. The memory requirements are the problem bearing in mind the number of used diphones. There are about 2000 diphones used and the size of each is 10 Kbytes on average. It is necessary to minimise the memory requirements for the mobile devices such as PDAs, cellular phones and others. The compression contributes to this goal noticeably. Fig. 6 shows the structure of the diphone used in the TTS system. The diphones

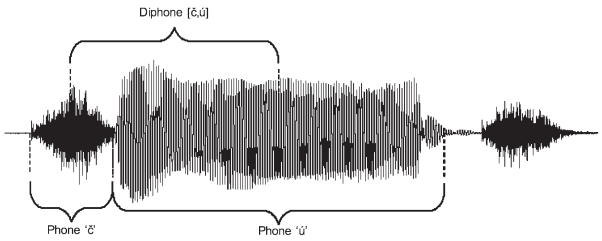


Fig. 6. A diphone structure

are stored compressed; for each experimentally determined the suitable mask shape to meet the requirements for quality (SNR > 25dB). Using compression the total amount of data was reduced to 40% of its original size.

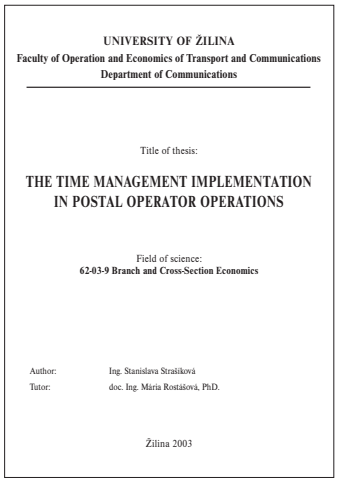
6. Conclusion

The presented technique gives acceptable results and reduces the amount of data needed for diphone storage. More experiments with the shape of the mask and the reconstruction method are required to cover a larger area of possibilities. During the experiments linear, cubic and spline approximation was used. Other methods of interpolation would also give a more complete view of the technique although noticeable improvements are not expected. The results show that presented algorithm compared to other techniques (i.e. ADPCM, that is loss-less and reduces data up to 25% of original size) does not excel and achieved compression ratio and SNR are not satisfactory. However there are not any known sources or publications that would use presented approach so this paper may contribute to future research in the area of non-equidistant sampling rate.

References

[1] MINOLI, D., MINOLI, E.: *Delivering Voice over IP networks,* ISBN 0-471-25482-7, (1998)
[2] PULSE CODE MODULATION (PCM) OF VOICE FREQUENCIES *ITU-T Recommendation G.711*
[3] 40, 32, 24, 16 kbit/s ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION (ADPCM) *ITU-T Recommendation G.726*
[4] 7 kHz AUDIO - CODING WITHIN 64 KBIT/S *ITU-T Recommendation G.722*
[5] Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3kBit/s ITU-T Recommendation G.723.1
[6] Engineering Fundamentals, Sampling Theorem and Nyquist Rate, *http://www.efunda.com/designstandards/sensors/methods/DSP_nyquist.cfm*
[7] Black, A. W., Taylor, P., Caley, R.: *"Festival Speech Synthesis System",* http://www.cstr.ed.ac.uk/projects/festival/festival-toc.html
[8] Igor Mihálik: *Návrh parametrov kompresného algoritmu,* DP reg.číslo 275/2000, (2001).

Title of thesis:	**The Time Management Implementation In Postal Operator Operations**
Author:	**Ing. Stanislava Strašíková**
Field of science:	**Branch and Cross-Section Economics**
Training institute:	**Department of Communications**
Department:	**Department of Communications**
Faculty:	**Faculty of Operation and Economics of Transport and Communications**
Name of university:	**University of Žilina**
Tutor:	**doc. Ing. Mária Rostášová, PhD.**
Viva voce:	**13th January 2004**



UNIVERSITY OF ŽILINA
Faculty of Operation and Economics of Transport and Communications
Department of Communications

Title of thesis:

THE TIME MANAGEMENT IMPLEMENTATION IN POSTAL OPERATOR OPERATIONS

Field of science:
62-03-9 Branch and Cross-Section Economics

Author: Ing. Stanislava Strašíková
Tutor: doc. Ing. Mária Rostášová, PhD.

Žilina 2003

The PhD thesis subject is the implementation of time management principles in postal operator operations. I underlined the fact that not only costs and quality but also time can become a basis for a future competitive advantage of a postal operator.

My PhD thesis aimed to design a methodical time management application procedure for a postal operator. I proposed a specific strategy how to reveal problems, implement and monitor solutions in the specified field. Although postal operator managers are obliged to make long-term efforts in order to carry out the defined strategy they will achieve efficient work practices.

This PhD thesis deals with a process of continual development in the field of postal operator time management. The development was described as a model consisting of eight the following stages: 1. time use analysis, 2. time problems identification, 3. self-assessment, 4. definition of goals and priorities, 5. programming goals into action plans, 6. daily schedules and planning, 7. improvement of time management techniques, 8. follow-up and repeated analysis.

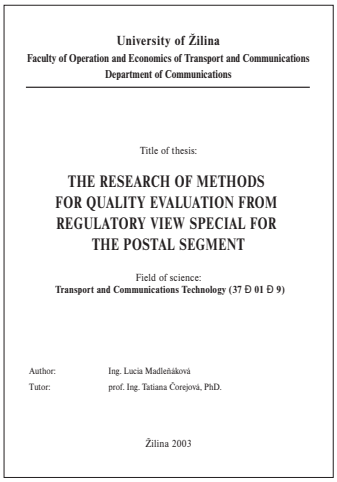
Within the process as a whole and for each stage I suggested some suitable methods and techniques (e.g. time studies, Pareto analysis, SWOT analysis, Ishikaw diagram, focus group methodology, Eisenhower principle, Balanced Scorecard, workshop of manager team) to be used as methodological tools. My proposal is aimed at searching opportunities how managers-working style can be improved to use time more efficiently.

The Slovak Post Office, state corp., has as its basic goal to provide quality services, which meet customer needs. The most important strategy of the Slovak Post Office is to strengthen more and more its position of a customer-oriented enterprise. To provide reliable, allowable and effective services, it decides about its future and assures necessary resources for employees' development and motivation.

In order to understand manager's productivity, time is a useful diagnostic tool. If organizations can realize the way they actually use their time, they will be able to make a better use of it. Focussing on middle managers in the Slovak Post Office I presented several research results concerning how effectively the managers use their time. Consequently I evaluated their economic impact for the Slovak Post Office.

The research I conducted within the Directorates of Posts Offices indicates that the middle managers should prefer to spend more time on their higher priority activities and less time on administrative tasks. The realized measurement of time shows how employees' real results are out of line with their ideal profiles. The fact highlights the need for on-going emphasis on the key core competency. This involves establishing goals, setting priorities, and managing interruptions, delegating and communicating effectively as well as overcoming procrastination, saying no and preparing for meetings.

There appears to be enormous variations in the way the managers either use their time effectively or squander it. In my opinion this topic deserves specific attention in conditions of each postal operator too. Issues of particular importance are the identification of priorities, logical sequencing of work, delegation and the need for managers to avoid wasting time.



University of Žilina
Faculty of Operation and Economics of Transport and Communications
Department of Communications

Title of thesis:

THE RESEARCH OF METHODS FOR QUALITY EVALUATION FROM REGULATORY VIEW SPECIAL FOR THE POSTAL SEGMENT

Field of science:
Transport and Communications Technology (37 Đ 01 Đ 9)

Author: Ing. Lucia Madleňáková
Tutor: prof. Ing. Tatiana Čorejová, PhD.

Žilina 2003

Title of thesis: Research of methods for quality evaluation from regulatory view special for the postal segment
Author: Ing. Lucia Madleňáková
Field of science: Transport and Communications Technology
Training institute: Department of Communications
Department: Department of Communications
Faculty: Faculty of Operation and Economics of Transport and Communications
Name of university: University of Žilina
Tutor: prof. Ing. Tatiana Čorejová, PhD.
Viva voce: 10th December 2003

Summary

Conditions of entrepreneurship for national operators, which are influenced by the successive liberalization, are changed in the framework of the creation of the uniform European market. Activities of the public postal operator of the Slovak Republic were also affected by these changes. The public postal operator has run into pressure of the growing competition. It is caused by the influx of new providers. In order to secure providing uniform services of the required standard, activities of the postal market are regulated by the regulatory office which came into being as a result of EU requirements in the framework of law harmonization of both the Slovak Republic and the European Union.

In order to evaluate the status of the postal service market, a regulator must have knowledge of the financial management of the public operator, as well as knowledge about the quality of all services provided. The main aim of the thesis was to set a methodological procedure for monitoring and evaluation of the postal service quality which is necessary for the regulator, in this procedure there were considered all requirements of the customer, regulator, an the operator with regard to their mutual relations and the market position.

At first it was necessary to define the basic concepts and definitions in the services branch, first of all the postal services, and the management of service quality. By means of the producers' analysis, accesses and methods of the quality management used not only in a native country but also abroad, it was possible to prepare the methods analysis which are determined for measuring of transport term as the most important criteria influencing the quality of a transferring procedure. On the basis of analysis some recommendations for determination of the range, choice of connections of the final measuring points with regard to the national and international regulations, recommendations for creations for criteria of an optimal network of respondents, as well as a proposal for securing the continual system measurement were set.

Measuring and monitoring the criteria concerning the quality of postal services are done for that purpose how to find out fulfilling of requirements set by the regulator. The results of measuring are compared with standards, which are represented by the limits of normality for purposes of procedure diagnostics in the post. Their deflection can cause both positive and negative customer reactions. It is possible to find out places where potential malfunctions will have an affect on the quality of the final postal service.

The basic characteristic feature of designed model is also enables to apply non-standard procedures of the quality management, which are based on diagnostics, analysis and therapeutic bringing the implementation of correctional methods. Such a model enables application of suitable methods for clearly defined problems including methods and techniques needed for measuring and evaluation of requirements concerning the uniform service quality. All is done to satisfy valid legislation.

On the basis of facts given in the thesis it can be pointed out that this presentation contains uniform knowledge from the branch of the quality management with the orientation toward the postal services. The methods and procedures described in this presentation are in accordance with the thesis recommendation, and they will open a space for further studying of regulators' and operators' relations with regard to fulfilling customers' requirements. The main goal is to achieve his/her satisfaction.

COMMUNICATIONS – Scientific Letters of the University of Žilina
Writer's Guidelines

1. Submissions for publication must be unpublished and not be a multiple submission.
2. Manuscripts written **in English language** must include abstract also written in English. The submission should not exceed 7 pages (format A4, Times Roman size 12). The **abstract** should not exceed 10 lines.
3. Submissions should be sent: **by e-mail** (as attachment in system Microsoft WORD) to one of the following addresses: *holesa@nic.utc.sk* or *vrablova@nic.utc.sk* or *polednak@fsi.utc.sk* **with a hard copy** (to be assessed by the editorial board) **or on a 3.5″ diskette** with a hard copy to the following address: Zilinska univerzita, OVaV, Moyzesova 20, SK-10 26 Zilina, Slovakia.
4. Abbreviations, which are not common, must be used in full when mentioned for the first time.
5. Figures, graphs and diagrams, if not processed by Microsoft WORD, must be sent in electronic form (as GIF, JPG, TIFF, BMP files) or drawn in contrast on white paper, one copy enclosed. Photographs for publication must be either contrastive or on a slide.
6. References are to be marked either in the text or as footnotes numbered respectively. Numbers must be in square brackets. The list of references should follow the paper (according to **ISO 690**).
7. The author's exact **mailing address of the organisation where the author works, full names, e-mail address or fax or telephone number,** must be enclosed.
8. The editorial board will assess the submission in its following session. In the case that the article is accepted for future volumes, the board submits the manuscript to the editors for review and language correction. After reviewing and incorporating the editor's remarks, the final draft (before printing) will be sent to authors for final review and adjustment.
9. The deadlines for submissions are as follows: September 30, December 31, March 31 and June 30.
10. Topics for the year 2005: 2/2005 – Transport – technologies, management and economics, 3/2005 – Security engineering.

POKYNY PRE AUTOROV PRÍSPEVKOV DO ČASOPISU KOMUNIKÁCIE – vedecké listy Žilinskej univerzity

1. Redakcia prijíma iba príspevky doteraz nepublikované alebo inde nezaslané na uverejnenie.
2. Rukopis musí byť **v jazyku anglickom**. Príspevok by nemal prekročiť 7 strán (formát A4, písmo Times Roman 12 bodové). K článku dodá autor **resumé** v rozsahu maximálne 10 riadkov (v anglickom jazyku).
3. Príspevok prosíme poslať: **e-mailom**, ako prílohu spracovanú v aplikácii Microsoft WORD, na adresu: *holesa@nic.utc.sk* alebo *polednak@fsi.utc.sk* príp. *vrablova@nic.utc.sk* (alebo doručiť na diskete 3,5″) **a jeden výtlačok** článku na adresu Žilinská univerzita, OVaV, Moyzesova 20, 010 26 Žilina.
4. Skratky, ktoré nie sú bežné, je nutné pri ich prvom použití rozpísať v plnom znení.
5. Obrázky, grafy a schémy, pokiaľ nie sú spracované v Microsoft WORD, je potrebné doručiť buď v digitálnej forme (ako GIF, JPG, TIFF, BMP súbory), prípadne nakresliť kontrastne na bielom papieri a predložiť v jednom exemplári. Pri požiadavke na uverejnenie fotografie priložiť ako podklad kontrastnú fotografiu alebo diapozitív.
6. Odvolania na literatúru sa označujú v texte alebo v poznámkach pod čiarou príslušným poradovým číslom v hranatej zátvorke. **Zoznam použitej literatúry** je uvedený za príspevkom. Citovanie literatúry musí byť **podľa STN 01 0197 (ISO 690)** „Bibliografické odkazy".
7. K rukopisu treba pripojiť **plné meno a priezvisko autora a adresu inštitúcie v ktorej pracuje, e-mail adresu** alebo číslo telefónu event. faxu.
8. Príspevok posúdi redakčná rada na svojom najbližšom zasadnutí a v prípade jeho zaradenia do niektorého z budúcich čísiel podrobí rukopis recenziám a jazykovej korektúre. Pred tlačou bude poslaný autorovi na definitívnu kontrolu.
9. Termíny na dodanie príspevkov do čísiel v roku sú: 30. september, 31. december, 31. marec a 30. jún.
10. Nosné témy v roku 2005: 2/2005 – Doprava – technológie, riadenie a ekonomika, 3/2005 – Bezpečnostné inžinierstvo.

komunikácie
COMMUNICATIONS

VEDECKÉ LISTY ŽILINSKEJ UNIVERZITY
SCIENTIFIC LETTERS OF THE UNIVERSITY OF ŽILINA
6. ROČNÍK – VOLUME 6

Adresa redakcie:
Address of the editorial office:
Žilinská univerzita
Oddelenie pre vedu a výskum
Office for Science and Research
Moyzesova 20, Slovakia
SK 010 26 Žilina
Tel.: +421/41/5620 392
Fax: +421/41/7247 702
E-mail: polednak@fsi.utc.sk, holesa@nic.utc.sk

Každý článok bol oponovaný dvoma oponentmi.
Each paper was reviewed by two reviewers.

Časopis je excerpovaný v Compendexe
Journal is excerpted in Compendex

Vydáva Žilinská univerzita
v EDIS – vydavateľstve ŽU
J. M. Hurbana 15, 010 26 Žilina
pod registračným číslom 1989/98
ISSN 1335-4205

It is published by the University of Žilina in
EDIS - Publishing Institution of Žilina University
Registered No: 1989/98
ISSN 1335-4205

Objednávky na predplatné prijíma redakcia
Vychádza štvrťročne
Ročné predplatné na rok 2005 je 500,– Sk

Order forms should be returned to the editorial office
Published quarterly
The subscription rate for year 2005 is 500 SKK

Jednotlivé čísla časopisu sú uverejnené tiež na:
http://www.utc.sk/komunikacie
Single issues of the journal can be found on:
http://www.utc.sk/komunikacie