JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

2/2004

Internet traffic delivery over fixed and wireless networks

Special issue edited by Józef Woźniak

Editorial Board

Editorial Advisory Board

ISSN 1509-4553

Warsaw 2004

Circulation: 300 copies

Sowa - Druk na życzenie, www.sowadruk.pl, tel. 022 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Solutions for efficient delivery of Internet traffic over fixed and wireless networks attract attention of researchers, network designers and operators. Modern computer networks have to provide differentiated services, dependent on traffic profiles. During the last decade a lot of Internet research has been devoted to provisioning different quality of service (QoS) to applications. Two major QoS-oriented Internet architectures were defined: IntServ—offering per flow service and DiffServ—more scalable and offering rich service granularity.

Current research activities concentrate on a few important topics, including traffic routing algorithms, design of core networks and IP QoS mechanisms. There is a growing interest in devising bandwith sharing algorithms in the new Internet architecture, so as to ensure high bandwith utilization while maintaining some notion of fairness.

Recently, a tremendous increase in Internet traffic has been observed. Both corporate and individual users demand more bandwith and more functions with QoS guarantees. This is accompanied by growing demand for cellular, WLAN and access systems. WLANs supporting broadband multimedia communications, are being developed and deployed around the world. WLAN standards include ETSI HIPERLAN/2 and the IEEE 802.11 family.

This special issue of *Journal of Telecommunications and Information Technology* covers all these subjects. It contains a selection of papers presented at the 2nd Polish-German Teletraffic Symposium.

Józef Woźniak
Guest Editor

Paper

Fairness considerations with algorithms for elastic traffic routing

Tibor Cinkler, Péter Laborczi, and Michał Pióro

Abstract—The bit rate of modern applications typically varies in time. We consider the traffic elastic if the rate of the sources can be controlled as a function of free resources along the route of that traffic. The objective is to route the demands optimally in sense of increasing the total network throughput while setting the rates of sources in a fair way. We propose a new fairness definition the relative fairness that handles lower and upper bounds on the traffic rate of each source and we compare it with two other known fairness definitions, namely, the max-min and the proportional rate fairness. We propose and compare different routing algorithms, all with three types of fairness definitions. The algorithms are all a tradeoff between network throughput, fairness and computational time.

Keywords—elastic traffic, routing, fairness, maximum throughput, algorithms, ILP, heuristics.

1. Introduction

In modern infocommunications networks the rate of sources typically varies in time. On the one hand this is due to silence period detection of voice codecs, compression of voice and video to variable bit rate depending on the amount of information to be carried. On the other hand, the bit rate of the data that is not sensitive to delay and delay variation can be tuned according to the network conditions to maximise the throughput without affecting the delay sensitive traffic.

In the new Internet architecture there is a growing interest in devising bandwidth sharing algorithms, which can cope with a high bandwidth utilisation and at the same time maintain some notion of *fairness*, such as the max-min (MMF) [1, 2] or proportional rate fairness (PRF) [3].

Examples of elastic traffic are TCP sessions in IP networks and available bit rate (ABR) service class in asynchronous transfer mode (ATM) networks. Label switch paths (LSPs) of multiprotocol label switching (MPLS) networks are also easy to reconfigure. In all cases the rate of sources is influenced by the load of the network.

Three variants of elastic traffic optimisation can be distinguished: (1) *fixed paths*, (2) *pre-defined paths* or (3) *free paths* can be assumed. In the *fixed paths* case there is a single path defined between each origin-destination (O-D) pair and the allocation task is to determine the bandwidth assigned to each demand. In the *pre-defined paths* case we assume that between each O-D pair there is a set of admissible paths, that can be potentially used to realize the flow of the appropriate demand. In this case the allocation task does not only imply the determination of the bandwidth of the flow, but also the identification of the specific path that is used to realize the demands [4]. In the *free paths* case there is no limitation on the paths, i.e., the task is to determine the bandwidth of the traffic AND the routes used by these demands simultaneously. This novel approach, the joint path and bandwidth allocation is the main topic of this article.

Recent research results indicate that it is meaningful to associate a minimum and maximum bandwidth even with elastic traffic [5], therefore it is important to develop models and algorithms for this type of services. As an example the ABR service can be mentioned that has the minimum cell rate (MCR) lower bound and the peak cell rate (PCR) upper bound. For the bounded elastic services we propose a special weighted case of MMF notion: relative fairness (RF) that maximises the minimum rates relative to the difference between upper and lower bounds for each demand.

Considering literature, different aspects of the max-min fairness policy have been discussed in a number of papers, mostly in ATM ABR context, since the ATM Forum adopted the max-min fairness criterion to allocate network bandwidth for ABR connections, see, e.g., [6, 7]. However, these papers do not consider the issue of path optimisation in the bounded elastic environment. MMF routing is the topic of the paper [8], where the widest-shortest, shortest-widest and the shortest-dist algorithms are studied. These algorithms do not optimise the path allocation. A number of fairness notions are discussed and associated optimisation tasks are presented in [5] for the case of unbounded flows and assuming fixed routes.

Proportional rate fairness is proposed by Kelly [3] and also summarised by Massoulie and Roberts in [5]. The objective of PRF is to maximise the sum of logarithms of traffic bandwidths. While [3] does consider the path optimisation problem, it does not focus on developing an efficient algorithm for path optimisation when the flows are bounded.

Recent research activities focused on allocating the bandwidth of fixed paths. In [4] the approach has been extended such that not only the bandwidth, but also the paths are chosen from a set of pre-defined paths. The formulation of the pre-defined path optimisation problem is advantageous, since it has significantly less variables than the free path optimisation. However, its limitation is that the whole

method relays on the set of pre-defined alternative paths. If the set of paths are given in advance, setting up elastic source rates in fair way leads to suboptimal solution. Better results can be achieved if we determine the rate of elastic sources AND the routes used by these demands simultaneously. There arises a question how much resources should be reserved for each demand, and what path should be chosen for carrying that traffic in manner to utilise resources efficiently while obeying fairness constraints as well. In this paper we investigate these questions and propose exact algorithms for solving it, assuming three types of fairness definition: RF, MMF and PRF.

- *Relative fairness.* In this case the aim is to increase the rates relative to the difference between upper and lower bounds for each demand. RF is a useful subcase of the bounded MMF definition, which can be solved in shorter running time.
- *Max-min fairness.* In this case we want to maximise the smallest demand bandwidth.
- *Proportional rate fairness.* In this notion the aim is to set the rates as a result of a convex optimisation, prioritising shorter paths to longer ones.

For each of these fairness definitions, a parameter (α, β and γ, respectively) is associated with each source expressing the bandwidth of the source. In case of RF parameter α_d indicates the rate of source d relative to the difference between upper and lower bounds. In case of MMF parameter β_d indicates the bandwidth of demand d. Parameters α_d and β_d can be unique for all sources (denoted by α and β, called *uniform parameter* case), which is optimal in sense of fairness, however, typically the parameter of some demands can be increased (called *different parameters* case), which increases the rate of some sources while it does not limit the rate of other sources. The value $\min_d(\alpha_d)$ is simply denoted by $\min(\alpha)$ and $\sum_d(\alpha_d)/D$ is denoted by $\mathrm{av}(\alpha)$ where D is the number of demands in the network. Analogous notation is used for β. In case of PRF parameter γ is unique for the whole network: it indicates the sum of logarithms of traffic bandwidths.
All these fairness definitions can be investigated in the bounded case (bounds on the minimal and maximal bandwidths for each O-D pairs). In the unbounded case MMF and PRF can be optimised, while RF has no sense without bounds. All fairness definitions can be formalised with unweighted and weighted fairness measures. We formulate the unweighted case, i.e., assume that all sources have the same priority, and then extend the model for the weighted case, i.e., when the sources have different priorities.
Accordingly, the following cases will be considered in the following sections:

- relative fairness with bounds with uniform parameter (RF/B/U): in Section 2;
- relative fairness with bounds with different parameters (RF/B/D): in Section 2.2;
- max-min fairness without bounds with uniform parameter (MMF/NB/U): in Section 3;
- max-min fairness without bounds with different parameters (MMF/NB/F): in Section 3.2;
- max-min fairness with bounds with uniform parameter (MMF/B/U): in Section 3.4;
- max-min fairness with bounds with different parameters (MMF/B/F): in Section 3.4;
- proportional rate fairness without bounds (PRF/NB): in Section 4;
- proportional rate fairness with bounds (PRF/B): in Section 4.

First, we focus on the basic case of relative fairness with bounds and uniform parameter (RF/B/U) and we further enhance the method to increase network throughput by utilising the spare resources (RF/B/D). The exact formulation of the problem is presented and methods are proposed which solve them to required accuracy.

2. Relative fairness: formulation and algorithms

In this section relative fairness is considered that maximises the minimum rates relative to the difference between upper and lower bounds for each demand. The formulation relays on the integer linear programming (ILP) formulation of the unsplittable minimal cost multicommodity flow problem.
The network topology of N nodes and L links with link capacities C_l ($l = 1,2,...,L$) are given. The lower and the upper bounds for demands $d = 1,2,...,D$ are respectively m_d and M_d. Output is the capacity requirement (bandwidth) b_d of demand d: $m_d \le b_d \le M_d$, where b_d can be expressed as $b_d = m_d + \alpha(M_d - m_d)$ and where α (the parameter of RF) is a continuous variable which ensures fairness. It can take values $0 \le \alpha \le 1$. In this formulation we assume that it has the same value for all demands $d = 1,2,...,D$. A 0-1 flow indicator variable on link l of demand d is x_l^d.

Objective:

$$\max \quad \alpha\,. \tag{1}$$

Subject to constraints:

$$\sum_d x_l^d \cdot (m_d + \alpha(M_d - m_d)) \le C_l \quad l = 1,2,..,L\,, \tag{2}$$

where

$$0 \le \alpha \le 1\,, \tag{3}$$

$$\sum_{j=1}^{N} x_{ij}^d - \sum_{k=1}^{N} x_{ki}^d = \begin{cases} 1 & \text{if } i \text{ is the source of } d \\ -1 & \text{if } i \text{ is the sink of } d \\ 0 & \text{otherwise} \end{cases}, \quad i = 1,2,...,N, d = 1,2,...,D \tag{4}$$

$$x_l^d \in \{0,1\}, l = 1,2,...,L, d = 1,2,...,D\,. \tag{5}$$

Equations (2) are capacity constraints and Eqs. (4) are the well known flow-conservation constraints. Unfortunately, this is a *nonlinear* formulation, since constraint (2) is not linear. In the following subsections it will be linearised by a simple method.

2.1. Algorithms for a single α for the whole network (RF/B/U)

For configuring networks which handle elastic traffic heuristic methods are preferred since nonlinearity is hard to handle. However, in this case the following simple deterministic algorithm guarantees the quality of the results.

2.1.1. Binary search algorithm (BSA)

This algorithm is based on the idea of binary search for finding the optimal value of α between 0 and 1.

Step 1: Check the feasibility by setting $\alpha = 0$. If satisfied, check the upper bounds by setting $\alpha = 1$. If satisfied, the solution is obtained, if not, set iteration counter $k = 1$, $\alpha = 0$, $\Delta = 1/2$ and proceed to Step 2.

Step 2: Set $\alpha = \alpha + \Delta$ and run the unsplittable multicommodity flow (UMCF) subroutine (see Section. 2.1.2).

Step 3: Increment k. If UMCF was feasible set $\Delta = 1/2^k$ else set $\Delta = -1/2^k$.

Step 4: Go to Step 2 until required fairness is achieved.

This deterministic method guarantees the quality of the results, i.e., if the number of iterations is k, then the largest "unfairness" in sense of parameter α is upper bounded by $1/2^k$. In the 7^{th} iteration this unfairness will be less than 1% (0.0078125), while in the 10^{th} iteration less than 10^{-3}.

2.1.2. The unsplittable multicommodity flow subroutine

This subroutine finds the optimal routing for fixed α. This is the unsplittable multicommodity flow problem referred to as UMCF. It can be solved by an ILP solver, e.g., CPLEX. Set:

$$b_d = (m_d + \alpha(M_d - m_d)) \quad d = 1,2,...,D. \tag{6}$$

Objective:

$$\min \sum_d b_d \sum_l x_l^d \,. \tag{7}$$

Subject to constraints (4), (5) and:

$$\sum_d b_d x_l^d \le C_l \quad l = 1,2,...,L \,. \tag{8}$$

2.1.3. Adaptive search algorithm (ASA)

Instead of the BSA a faster method can be used for setting value of α. This is an extension of BSA referred to as ASA. The idea is to increase α without changing the paths. After a feasible UMCF subroutine we find a new value of $\alpha^{(k+1)}$ to be used in the forthcoming $(k+1)^{th}$ iteration, based on the paths of the current k^{th} iteration. The new alpha is calculated by the following equation derived from constraint (2):

$$\alpha^{(k+1)} = \min_l \left\{ \frac{C_l - \sum_d m_d x_l^{d,(k)}}{\sum_d x_l^{d,(k)} (M_d - m_d)} \right\} \; l = 1,2,...,L \,. \tag{9}$$

This increase of parameter α is carried out after each feasible UMCF subroutine. Adaptive search speeds up the algorithm or increases the precision of α.

2.2. Allowing slightly different values of α within a network (RF/B/D)

Since all traffics are changed equally according to the definition of parameter α, the first saturated link will limit the value of α. Therefore, an iterative approach is needed, which increases the network throughput, however, it deteriorates the fairness slightly, by offering more resources to demands not using saturated links. The idea is to set a new, higher value of $\alpha^{(k)}$ $(k = 1,2,...)$ for some demands by using free resources of yet unsaturated links in each iteration k. Note, that there are two alternatives:

Case 1: The paths of demands are determined in the first iteration. They are not changed any more, only the bandwidths.

Case 2: Both, the paths and bandwidths are improved in each iteration.

2.2.1. Case 1: Increase bandwidth

In this case the paths assigned to demands are determined within the first phase and are not changed any more. The allocations are changed only according to the following algorithm (Y_l^k represents the free capacity on link l after the k^{th} iteration):

Step 1: Set $k = 0$, $\alpha^{(0)} = \alpha$, $b_d = m_d + \alpha^{(k)}(M_d - m_d)$, $Y_l^{(0)} = C_l - \sum_d b_d x_l^d$.

Step 2: Set k++.

Step 3: Remove all saturated links and paths using these links.

Step 4: If there is no more demand left or $\alpha^{(k-1)} = \alpha^{(k-2)}$ then *Stop*, otherwise continue.

Step 5:

$$\alpha^{(k)} = \min_l \left\{ \frac{Y_l^{(k-1)} - \sum_d m_d x_l^d}{\sum_d x_l^d (M_d - m_d)} \right\} . \tag{10}$$

Step 6:

$$b_d = m_d + \alpha^{(k)}(M_d - m_d)\,. \tag{11}$$

Step 7:

$$Y_l^{(k)} = Y_l^{(k-1)} - \sum_d b_d x_l^d\,.$$

Step 8: Go to Step 2.

The new value for α is calculated by Eq. (10) that has analogous meaning to Eq. (9). Note, that this iterative procedure has to be repeated up to L times, where L is the number of links in total for the considered network, since each iteration will saturate at least one link.

2.2.2. Case 2: Increase bandwidth by rerouting

In this case both the routing of demands and allocations are changed. In each iteration (after BSA or ASA) saturated links are removed and all paths using these links are de-allocated. The link capacities should be decreased by the allocated capacity of removed demands (b_d). Now the whole algorithm should be run on the reduced graph until there are no more demands. This method has the longest running time, however, it gives the best resource utilisation. It is to be noticed that even in this case the global optimum is not guaranteed. This is because the optimal solution of BSA or ASA is not unique, and the choice of the optimal solution of BSA or ASA may influence the further development of the algorithm and its final results [4].

2.3. The weighted RF path and bandwidth allocation

If we want to prioritise some demands d then a weight factor w_d should be used. By setting $w_1 = 2w_2$ the rate allocated to demand 2 will be increased by double of the increment of demand 1.
In this case everything defined previously is valid, except that in the UMCF subroutine we should add the weight factor w_d for each demand d to the Eq. (6), as follows:

$$b_d = m_d + w_d\alpha^{(k)}(M_d - m_d)\,. \tag{12}$$

In Step 7 in Section 2.2.1 (Eq. (11)) the same should be done, and (10) (and analogously (9)) should be extended to:

$$\alpha^{(k)} = \min_l \left\{ \frac{Y_l^{(k-1)} - \sum_d m_d x_l^d}{\sum_d w_d x_l^d (M_d - m_d)} \right\}. \tag{13}$$

If we want to increase the network throughput, we can prioritise those demands which use shorter paths by setting w_d to be equal to the reciprocal value of the length of the demand, where the length is expressed in number of hops along the shortest possible path between the end-nodes of that demand. This leads to similar fairness definition than PRF. Further on we will deal with the weighted case only, assuming $w_d = 1$, $\forall d = 1,2,...,D$ for the unweighted case.

3. Max-min fairness: formulation and algorithms

First, we consider the case without bounds on the demand bandwidths, i.e., we will assume that an infinite amount of traffic is to be carried between the node-pairs. The task is to find optimal paths that allow the highest throughput, while giving the same chance to all demands, i.e., guaranteeing fairness.
Here, instead of parameter α, parameter β will be used with slightly different meaning as follows. β stands for capacity allocated to demands. In this section it will be equal for all demands $d = 1,2,...,D$.

Objective:

$$\max \quad \beta\,. \tag{14}$$

Subject to constraints (4), (5) and:

$$\beta \sum_d x_l^d \le C_l \quad l = 1,2,...,L\,. \tag{15}$$

3.1. Algorithms for a single β for the whole network

Here the UMCF algorithm described in Section 2.1.2 has to be changed only, as follows.

3.1.1. The UMCF2 subroutine

This subroutine finds the optimal routing for fixed β. If it had not been fixed, this would have been the exact formulation where β and the paths are optimised simultaneously, however, then the problem would have been nonlinear. The difference to UMCF is that $b_d = \beta$, $d = 1,2,...,D$ should be used instead of (6).
Note, that if β is a constant it can be avoided in the objective function.

Set:

$$b_d = \beta \quad d = 1,2,...,D\,. \tag{16}$$

Objective:

$$\min \left\{ \beta \sum_d \sum_l x_l^d \right\}. \tag{17}$$

Subject to constraints (4), (5) and (15).
As mentioned, this subroutine finds the optimal routing for fixed β. The value of β can be set iteratively either by the modified BS algorithm (Section 2.1.1) or by the modified AS algorithm.

3.1.2. Binary search algorithm for MMF

The BSA (Section 2.1.1) should be modified to be used for path optimisation with MMF fairness scenario as follows. The initial value of β should be set as follows:

$$\beta = \min_l \frac{C_l}{D}\,. \tag{18}$$

Then the value of β is increased iteratively by, e.g., 50–100% ($\beta^{(k)} = 1.5\beta^{(k-1)}$) while the problem can be solved. The value of β in the last k^{th} iteration will be the upper bound, while the lower bound will be its value in the $(k-1)^{th}$ iteration. Now, binary search between these two values can be used for finding β to required accuracy.

3.1.3. Adaptive search algorithm for MMF

The AS algorithm (Section 2.1.3) should be modified to be used for path optimisation with MMF fairness scenario as follows.
In this extension of BSA the idea is to increase β without changing the paths. After a feasible UMCF2 subroutine we find a new value of $\beta^{(k+1)}$ to be used in the forthcoming $(k+1)^{th}$ iteration based on the paths of the current k^{th} iteration. The new β is calculated by the following equation, derived from constraint (15):

$$\beta^{(k)} = \min_l \left\{ \frac{C_l}{\sum_d x_l^d} \right\}. \tag{19}$$

This increase of parameter β is carried out after each feasible UMCF2 subroutine, which speeds up the algorithm or increases the precision of β.

3.2. Allowing slightly different values of β within a network (MMF/NB/D)

Since the rate of all traffic is changed equally according to the definition of parameter β, the first saturated link will limit value of β. Therefore, an iterative approach is needed, which increases the network throughput, however, it deteriorates the fairness slightly, by offering more resources to demands not using saturated links. The idea is to set a new value of $\beta^{(k)}$ for yet unsaturated links in each iteration k. Now we will have different values of β for different demands or different sets of demands. Although this allows different rates to different demands, it does not really deteriorate the fairness, since demands having lower rates would not have been able to use higher rates due to bottlenecks, which can not be avoided (even re-routing does not help).

Note, that there are two alternatives analogously to 2.2.:

Case 1: The paths of demands are determined in the first iteration. They are not changed any more, only the bandwidths.

Case 2: Both, the paths and bandwidths are improved in each iteration.

3.2.1. Case 1: Increase bandwidth

In this case the paths assigned to demands are determined within the first phase and are not changed any more. The allocations are changed only, according to the following algorithm:

Step 1: Set $k = 0$, $\beta^{(0)} = \beta$, $Y_l^{(0)} = C_l - \beta^{(0)} \sum_d x_l^d$.

Step 2: Set *k*++.

Step 3: Remove all saturated links and paths using these links.

Step 4: If there is no more demand left or $\beta^{(k-1)} = \beta^{(k-2)}$ then *Stop*, otherwise continue.

Step 5:

$$\beta^{(k)} = \min_l \left\{ \frac{Y_l^{(k-1)}}{\sum_d x_l^d} \right\}. \tag{20}$$

Step 6:

$$Y_l^{(k)} = Y_l^{(k-1)} - \beta^{(k)} \sum_d x_l^d.$$

Step 7: Go to Step 2.

The new β is calculated by Eq. (20) that has analogous meaning to Eq. (19).
Note, that this iterative procedure has to be repeated up to L times in total for the considered network, since each iteration will saturate at least one link.

3.2.2. Case 2: Increase bandwidth by rerouting

In this case both the routing of demands and allocations are changed. In each iteration saturated links should be removed with all paths using these links. The link capacities should be decreased by the capacity allocated to demands removed (by β). Now the whole algorithm should be run for the reduced graph.
This method has the longest running time, however, it gives the best resource utilisation. It is to be noticed that the global optimum is not guaranteed for the reasons mentioned in Section 2.2.2.

3.3. The weighted MMF path and bandwidth allocation

In this case everything defined previously in Section 3 is valid, except that everywhere (e.g., in Eq. (15)) $w_d x_l^d$ should be written instead of x_l^d and $\sum_d w_d$ should be written instead of D in Eq. (18). Further on we will deal with the weighted case only, assuming $w_d = 1$, $\forall d = 1, 2, ..., D$ for the unweighted case.

3.4. Max-min fairness with bounds (MMF/B)

In this subsection we assume that each demand has a lower bound (m_d) and an upper bound (M_d). The lower bound is taken into account by simply modifying the capacity constraints of RF/B in the following way. β should be written instead of α, and 1 should be written instead of $(M_d - m_d)$, i.e., $(M_d - m_d)$ should be simply left out from the formulation.
The upper bound is handled by introducing an auxiliary leaf node v_d for each demand d and a new link of capacity M_d from the source node of demand d to v_d and finally

changing the source of d to v_d. Another way of handling upper bounds is to introduce extra constraints into the ILP formulations.

4. Proportional rate fairness: formulation and algorithms

The above fairness definitions ensure optimal fairness measured either relative to the upper and lower bounds (RF) or in absolute units (MMF). However, in these cases the connections spanning more distant points (more hops) will adapt their rate in the very same way, as those close to each other.
To increase the throughput the fairness criteria should be redefined in manner to prioritise connections having less hops (i.e., using less resources) to those which are more distant. F. Kelly *et al.* have proposed the concept of proportional rate fairness [3] where the objective to be optimised is the sum of logarithms of the capacities used by certain demands (e.g., b_d), while the constraints are the same as in our previous formulations.

Objective:

$$\max \quad \sum_d \lg b_d \,. \tag{21}$$

Subject to constraints (4), (5) and:

$$\sum_d x_l^d b_d \le C_l \quad l = 1,2,...,L \,. \tag{22}$$

Unfortunately, this is a convex problem that is nonlinear. To handle this problem a piece-wise linear approximation of the logarithmic function is applied by introducing an auxiliary variable f_d for each demand d as proposed in [4]. The modified objective will be

$$\max \sum_d f_d \tag{23}$$

and additionally the following constraints are given for each demand d:

$$f_d \le r_k b_d + s_k, \quad k = 1,...,K \,. \tag{24}$$

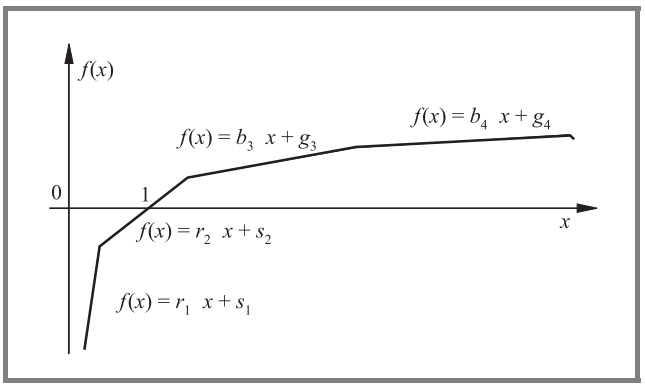


Fig. 1. The piece-wise linear approximation of the logarithmic function.

Figure 1 shows the approximation for $K = 4$ linear pieces, however, in practice more pieces can be used.
In our study the following inequalities were used:

$$f_d \le 4.023595 b_d - 2.704945, \tag{25}$$

$$f_d \le 1.386294 b_d - 1.386294, \tag{26}$$

$$f_d \le 0.693147 b_d - 0.693147, \tag{27}$$

$$f_d \le 0.305430 b_d + 0.082287, \tag{28}$$

$$f_d \le 0.109861 b_d + 1.060132, \tag{29}$$

$$f_d \le 0.034399 b_d + 2.192062. \tag{30}$$

However, after eliminating the logarithmic function from the objective, another problem occurs, namely that constraint (22) is not linear. To avoid this we introduce a new variable y_l^d, which represents the flow value of demand d on link l. By the following formulation the problem is linear, however it enables split flows.

Objective: (23).

Constraints: (24) and:

$$\sum_{j=1}^{N} y_{ij}^d - \sum_{k=1}^{N} y_{ki}^d = \begin{cases} b_d & \text{if } i \text{ is the source of } d \\ -b_d & \text{if } i \text{ is the sink of } d \\ 0 & \text{otherwise} \end{cases}, \tag{31}$$
$$i = 1,2,...,N, \;\; d = 1,2,...,D$$

$$\sum_d y_l^d \le C_l \quad l = 1,2,...,L \,. \tag{32}$$

To avoid split flows additional constraints are needed and the following final formulation is proposed: M is a large number.

Objective: (23).

Constraints: (4), (5), (31), (24), (32) and:

$$y_l^d \le M x_l^d \quad l = 1,2,...L, \;\; d = 1,2,...,D \,. \tag{33}$$

The bounded case (PRF/B) can be handled by simply introducing constraints into the above ILP formulation.
Although PRF deteriorates fairness in sense of earlier fairness definitions, it increases the throughput.

5. Comments and improvements

Although the problems have been defined here for unsplittable flows only, all the methods can be used for splittable flows as well. This even reduces the complexity, since linear programming can be used instead of integer linear programming or mixed integer programming.
When both elastic and rigid traffics coexist in a network, the model has not to be changed, only m_d and M_d values are to be set to be equal ($m_d = M_d$) for all rigid demands. However, if the problem is being solved by a mixed integer linear programming (MILP) solver it might be useful to introduce new variables instead. This will reduce the number of constraints and it will speed up solving the problem and

allow problems of larger scale to be solved. In the numerical results we will deal with elastic traffic only, since as shown this does not reduce the generality.
We believe that the proposed approach ensures the highest fairness, i.e., RF is more fair than the plain MMF. Furthermore, the formulation of the joint path and bandwidth optimisation guarantees higher (or at least equal) throughput than the one with pre-defined paths.

5.1. Iterative elastic simulated allocation (IESA)

The methods spend most of their time in the UMCF or UMCF2 subroutine. The ILP formulation of them contains LD variables and $ND+L$ constraints. Several methods have been proposed in the literature that solve the unsplittable multicommodity flow problem much faster, e.g., SA++ or CA++ in [10]. We have applied SA++ in this study that is based on simulated allocation. The main idea behind simulated allocation [11] is a randomised alternative path allocation and de-allocation of the traffic demands.
Using of SA++ is proposed in larger networks (e.g., with more than 15 nodes), which does not guarantees the optimal solution, but is much faster than the method based on ILP.

Using ILP in the UMCF subroutine is called elastic ILP (EILP), while replacing the UMCF subroutine with SA++ is called iterative elastic simulated allocation.

5.2. Elastic simulated allocation (ESA)

Simulated allocation can be used in a more sophisticated manner as well. The main point of this improvement is that after several iterations of allocations and de-allocations a special procedure called *bandwidth tuner* is called. The bandwidth tuner procedure tunes (changes) the bandwidth of each demand according to the appropriate fairness definition. For example, in case of RF it decreases the value of α if any demand can not be allocated, or increases the value of α if all demands can be allocated and more free space is available in the network. This method is called elastic simulated allocation.

5.3. Iterative heuristic for PRF (IPRF)

The ILP formulation of the PRF definition is very complex: it contains $2(L+1)D$ variables and $2ND+KD+L$ constraints. In order to speed up the calculation the following iterative heuristic method is proposed:

> *Step 0*: Find a feasible system of paths by applying MMF/NB/U or MMF/B/U. Set $k=0$ and $\gamma^{(0)}=-\infty$.
>
> *Step 1*: Increase k and find the bandwidth b_d for each demand d, according to PRF definition by solving the above problem (that is linear in this case) by an ILP solver. Let $\gamma^{(k)}$ the objective value of the problem. If $\gamma^{(k)}$ has been increased ($\gamma^{(k)} > \gamma^{(k-1)}$) then continue, otherwise *Stop*.
>
> *Step 2*: Run UMCF with bandwidths (b_ds) found in Step 1. Go to Step 1.

5.4. Shortest paths algorithm (SPA)

A simple method called shortest paths algorithm has been also implemented. It finds a shortest path for each demand and sets the bandwidth of the demand according to the appropriate fairness definition. This method is similar to those previous methods that assume fixed paths, i.e., it is not able to change the path only the bandwidth of the demands.

6. Numerical results

The tests have been carried out on six networks with different number of nodes and links (Table 1). The bounds of traffic demands have been chosen randomly so that the task was not trivial, i.e., using m_d parameters they fit into capacities, while with M_d not.

Table 1
Details of the six test networks

Details	N5	N5A	N12	N15	N25	N35
Nodes	5	5	12	15	25	35
Links	5	6	18	15	31	51
Demands	10	10	66	105	300	595

The methods have been compared according to 4 groups of criteria: computational time, network throughput, fairness parameters and hop number. The network throughput (TP) is expressed as the total of carried traffic for all demands. Fairness parameters are min(α), av(α), min(β), av(β) and γ as defined in Section 1. Average and maximal hop number, av(H) and max(H), indicate the average and maximal hops used by the system of paths.
The results are summarised in Table 2 for methods EILP, IESA and SPA on N12 which represents a relevant part of the Polish backbone. Considering running time, both EILP and IESA is about 12 times faster in case of RF/B than in case of MMF/B. The reason for this is that the addition of D new links and nodes increases the running time significantly. IESA (the heuristic method) is about an order faster than EILP. IESA yields a little worse result than EILP, but still much better then SPA in sense of throughput and fairness parameters. However, average and maximal hop numbers are higher since randomised heuristic allows longer paths. From these results it can be stated that joint path and bandwidth allocation yields better results in sense of throughput and fairness.
It is interesting to compare the fairness parameters (min(α), av(α), min(β), av(β), γ) according to the fairness that had been considered in the optimisation phase. For example, in case of RF min(α) and av(α) are relatively high compared to MMF and PRF, however, it yields lower values for min(β), av(β) and γ.
Considering PRF this yields the highest throughput, γ and also av(β), and not significantly worse min(β). Consequently, this seems to be very promising in the unbounded case. However, in the bounded case it gives very poor

Table 2

Numerical results of methods EILP, IESA and SPA for the N12 network

EILP		Time	TP	min(α)	av(α)	min(β)	av(β)	γ	av(H)	max(H)
RF/B	BS	34.1	104.9	0.083	0.083	1.249	1.590	27.153	2.20	5
	AS	16.2	105.0	0.083	0.083	1.250	1.591	27.204	2.20	5
	Case 1	16.2	151.7	0.083	0.268	1.250	2.298	45.877	2.20	5
	Case 2	39.4	152.7	0.083	0.276	1.250	2.314	46.979	2.23	5
MMF/B	BS	293.9	105.7	0.055	0.095	1.328	1.601	28.832	2.24	5
	AS	305.1	106.0	0.056	0.096	1.333	1.606	29.060	2.24	5
	Case 1	308.0	145.3	0.056	0.256	1.333	2.201	43.831	2.24	5
	Case 2	473.4	144.9	0.056	0.259	1.333	2.195	44.523	2.35	6
PRF/B		56.4	165.0	0.000	0.367	1.000	2.500	51.698	2.39	5
MMF/NB	BS	15.8	92.4	–0.100	0.070	1.400	1.400	22.204	2.20	5
	AS	15.6	92.4	–0.100	0.070	1.400	1.400	22.207	2.20	5
	Case 1	15.7	236.9	–0.100	0.660	1.400	3.590	56.391	2.20	5
	Case 2	42.6	230.6	–0.100	0.606	1.400	3.494	56.610	2.24	5
PRF/NB		47.9	243.0	0.015	0.342	1.000	3.682	59.826	2.26	5
IESA		Time	TP	min(α)	av(α)	min(β)	av(β)	γ	av(H)	max(H)
RF/B	BS	1.4	94.6	0.042	0.042	1.126	1.433	20.308	2.42	5
	AS	1.5	100.8	0.067	0.067	1.200	1.527	24.510	2.50	8
	Case 1	1.7	141.7	0.067	0.235	1.200	2.148	41.193	2.50	7
	Case 2	4.0	144.9	0.067	0.232	1.200	2.196	41.332	2.68	7
MMF/B	BS	43.3	101.0	0.043	0.074	1.258	1.531	25.669	2.59	5
	AS	31.6	89.5	0.014	0.024	1.083	1.356	17.054	2.41	8
	Case 1	31.8	135.8	0.012	0.212	1.071	2.058	36.053	2.62	9
	Case 2	155.4	137.9	0.030	0.218	1.182	2.089	39.704	2.83	7
PRF/B		0.7	146.0	0.000	0.260	1.000	2.212	38.059	2.26	5
MMF/NB	BS	0.7	84.8	–0.119	0.037	1.286	1.286	16.578	2.35	5
	AS	1.8	92.4	–0.100	0.070	1.400	1.400	22.207	2.61	8
	Case 1	1.0	225.4	–0.111	0.611	1.333	3.415	46.918	2.55	7
	Case 2	5.3	213.4	–0.100	0.566	1.400	3.233	49.946	2.56	6
PRF/NB		2.4	239.0	–0.167	0.659	1.000	3.621	51.902	2.52	8
SPA		Time	TP	min(α)	av(α)	min(β)	av(β)	γ	av(H)	max(H)
RF/B	U	0.02	28.1	0.0165	0.0165	0.0495	0.4263	–78.5047	2.14	4
	D	0.08	205.2	0.0165	0.1739	0.1155	3.1098	23.3567	2.14	4
MMF/NB	U	0.02	25.4	0.003	0.0288	0.3846	0.3846	–63.0638	2.14	4
	D	0.05	280.8	0.0036	0.4563	0.3846	4.255	38.3519	2.14	4
MMF/B	U	0.5	25.4	0.003	0.0288	0.3846	0.3846	–63.0638	2.14	4
	D	0.852	206	0.0036	0.192	0.3846	3.1212	32.6039	2.14	4
PRF/NB		0.09	297.2	0.0037	0.4967	0.1	4.503	37.2381	2.14	4
PRF/B		0.1	215.7	0.0037	0.2052	0.1	3.2682	31.3147	2.14	4

values for min(α), i.e., if RF notion is assumed to be fair, than many connections come to grief if optimised with PRF. In the bounded case RF/B (Case 2) is better in running time, throughput, hop numbers, and we believe that it is more fair than the simple, unweighted MMF. Summarised, PRF is proposed in the unbounded case, while RF in the bounded case.

In Fig. 2 min(α) of the four methods are compared. Min(α) depends on the traffic pattern, i.e., the results can only be compared within one network. EILP yields the best solution while IESA and ESA are also very close to the optimum. ESA is closer to the optimum especially in larger networks. This is a very promising heuristic method for other fairness definitions as well. EILP did not found solution in N25 and N35 in acceptable time. SPA could not solve the problem in N12 and N25, since it works with fixed shortest paths and in these cases the paths violates the capacity constraints even with the lower bounds.

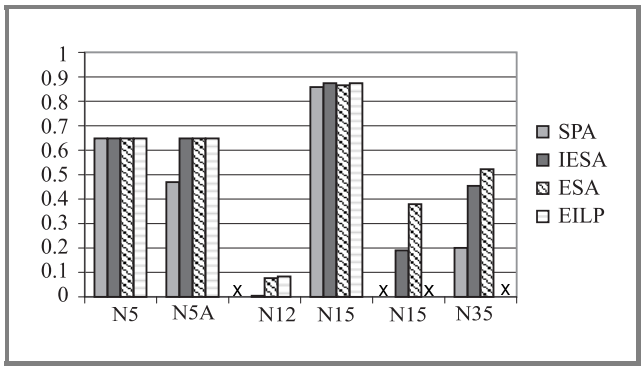


Fig. 2. Min(α) for the six test networks using algorithms SPA, IESA, ESA and ILP.

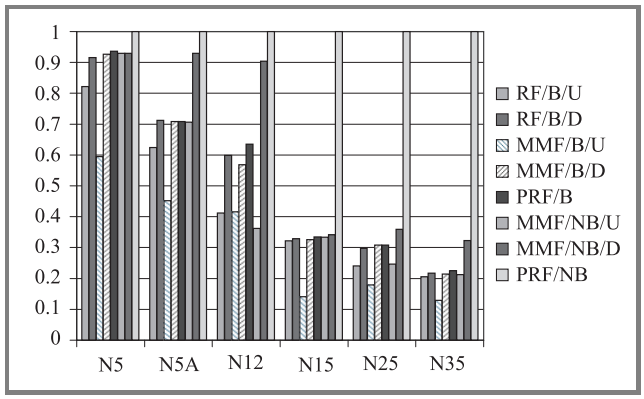


Fig. 3. Throughput of six test networks assuming eight fairness definitions.

In Fig. 3 the throughput is normalised to PRF/NB for each network. Trivially the unbounded (NB) cases always yield higher throughput than the bounded (B) case, and the case allowing different (D) parameter yields higher throughput than the case with uniform (U) parameters. RF and MMF have similar throughput, while PRF has higher throughput, especially in the unbounded case. The efficiency of PRF is very convincing in larger networks, since in this case longer paths obtain significantly less bandwidth that makes space for many short paths.

Table 3
Computational time of ILP and IESA for six test networks and five fairness definitions

EILP	N5	N5A	N12	N15	N25	N35
RF/B	0.18	0.65	39.4	8873	-	-
MMF/B	0.96	1.64	473	-	-	-
PRF/B	0.19	0.25	56.4	80.5	-	-
MMF/NB	0.09	0.18	42.6	30.6	-	-
PRF/NB	0.14	0.22	47.9	38.3	-	-
IESA	N5	N5A	N12	N15	N25	N35
RF/B	0.03	0.03	4.0	2.5	36.7	85.6
MMF/B	0.15	0.15	155	123	3768	30240
PRF/B	0.01	0.01	0.7	0.3	2.2	8.7
MMF/NB	0.02	0.02	5.3	2.8	60.6	91.7
PRF/NB	0.01	0.01	2.4	3.8	27.1	100

The computational time of EILP and IESA for five fairness definitions is compared in Table 3. In case of EILP it was acceptable only in networks having up to 15 nodes. IESA is faster, however in case of MMF/B further speed up is required.

7. Conclusion

A wide range of algorithms has been proposed, which are all a tradeoff (compromise) between network throughput, fairness and computational time.

In all cases the obtained results were better (in sense of fairness and throughput) than for the case of fixed and predefined alternative paths, however, the running time was longer. Joint optimisation of paths and bandwidths appeared to be always better. We have shown that unused capacities can be further utilised to increase the throughput without deteriorating the fairness in its strict sense. We propose to apply relative fairness notion in the bounded case and proportional rate fairness in the unbounded case. Methods based on ILP are proposed for smaller (less than 20 nodes) networks and iterative heuristics for larger networks.

These methods can be used in any centralised resource management system in the new Internet architecture for configuration of ATM, IP and MPLS networks which will carry elastic traffic.

Acknowledgements

The authors thank Gábor Fodor (Ericsson Research, Sweden) for initiating discussions and bringing our attention to this area.

T. Cinkler is supported by the OTKA grant D42211 and by the János Bólyai foundation.

P. Laborczi is supported through a European Community Marie Curie Fellowship.

References

[1] D. Bertesekas and R. Gallagher, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1987.

[2] A. Charny, D. Clark, and R. Jain, "Congestion control with explicite rate indication", in *IEEE ICC'95*, Seattle, USA, 1995.

[3] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communications networks: shadow prices, proportional fairness and stability", *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[4] G. Fodor, G. Malicskó, M. Pióro, and T. Szymański, "Path optimization for elastic traffic under fairness constraints", ITC17, Salvador da Bahia, Brazil, 2001.

[5] L. Massouline and J. W. Roberts, "Bandwidth sharing: objectives and algorithms", in *IEEE INFOCOM'99*, New York, USA, 1999.

[6] Y. T. Hou, H. H.-Y. Tzeng, and S. S. Panwar, "A simple ABR switch algorithm for the weighted max-min fairness policy", in *IEEE ATM Worksh.'97*, Lisboa, Portugal, 1997.

[7] H. Qingyanga and D. W. Petr, "Global max-min fairness guarantee for ABR flow control", in *IEEE INFOCOM'98*, San Francisco, USA, 1998.

[8] Q. Ma, P. Steenkiste, and H. Zhang, "Routing high-bandwidth traffic in max-min fair share networks", in *SIGCOMM'97*, Cannes, France, 1997.

[9] S. P. Abraham and A. Kumar, "A new approach for asynchronous distributed rate control for elastic sessions in integrated packet networks", *IEEE/ACM Trans. Netw.*, vol. 9, no. 1, 2001.

[10] P. Laborczi and P. Fige, "Static LSP routing algorithms for MPLS networks", in *Networks 2000*, Toronto, Canada, 2000.

[11] P. Gajownicek, M. Pióro, and A. Arvidsson, "VP reconfiguration through simulated allocation", *NTS*, vol. 13, pp. 251–260, 1996.

Tibor Cinkler received his M.Sc. ('94) and Ph.D. ('99) degrees from the Budapest University of Technology and Economics, Hungary, where he is currently Associate Professor at the Department of Telecommunications and Media Informatics. His research interests focus on routing, design, configuration, dimensioning and resilience of IP, MPLS, ATM, NG-SDH and particularly of WR-DWDM based multilayer networks. He has been involved in a few related European and Hungarian projects (ACTS, COST, ETIK, IKTA) and he is member of ONDM, DRCN, Opticomm, EUNICE, ICOCN, etc. Scientific Committees. He is author of over 60 referred scientific publications and of 3 patents.
e-mail: cinkler@tmit.bme.hu
Department Telecommunications and Media Informatics
Budapest University of Technology and Economics
Magyar tudósok krt. 2
H-1117 Budapest, Hungary

Péter Laborczi received his B.Sc. and M.Sc. degrees in 1999 and his Ph.D. degree in 2002 in Computer Science from the Budapest University of Technology and Economics in Hungary, Department of Telecommunications and Telematics. Since 2002 he has been a Marie Curie Postdoc research fellow at Arsenal Research in Vienna, Austria. His research fields include configuration, optimization, routing, protection and restoration of infocommunication (mainly MPLS and WDM) networks and similar areas of transport technologies.
e-mail: peter.laborczi@arsenal.ac.at
Arsenal Research
Business Area of Transport Technologies
Faradaygasse 3, Objekt 219C
1030 Vienna, Austria

Michał Pióro is a Professor at the Institute of Telecommunications, Warsaw University of Technology and at the Department of Communication Systems, Lund University. He received a Ph.D. degree in telecommunications in 1979 and a D.Sc. degree in 1990, both from the Warsaw University of Technology. In 2002 he received a Polish State Professorship. During 1984–87 and 1997 he had been with the Lund Institute of Technology, leading research projects for Ericsson Telecom AB in network design. In 1986, 2001 and 2003 he served as a senior expert for ITU. During 1990-1991 he had been working for Alcatel SESA, Spain, as a consultant in dynamic routing strategies. In Fall'2002 semester he was a visiting scholar and Professor at University of Missouri-Kansas City, USA. Prof. Pióro wrote three monographs and over 100 papers presented in international telecommunications journals and conference proceedings. He is a Technical Editor of *IEEE Communications Magazine* and a member of program committees of many important telecommunication conferences. His research interests concentrate on modelling, design and performance evaluation of telecommunications networks.
e-mail: mpp@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

An exact algorithm for design of content delivery networks in MPLS environment

Krzysztof Walkowiak

Abstract—Content delivery network (CDN) is an efficient and inexpensive method to improve Internet service quality. In this paper we formulate an optimisation problem of replica location in a CDN using MPLS techniques. A novelty, comparing to previous work on this subject, is modelling the network flow as connection-oriented and introduction of capacity constraint on network links to the problem. Since the considered optimisation problem is NP-complete, we propose and discuss exact algorithm based on the branch-and-cut and branch-and-bound methods. We present results of numerical experiments showing comparison of branch-and-cut and branch-and-bound methods.

Keywords—content delivery network, optimization, branch-and-cut algorithm.

1. Introduction

In recent years we observe a tremendous increase in data traffic, caused mainly by the growth of the Internet as well as introduction of many new services. Concurrently, corporate and individual users demand more bandwidth and more functions with quality of service (QoS) guarantees. The existing Internet sometimes cannot cope with all challenges that are to be addressed by computer networks in near future. Therefore, new solutions are being developed to overcome most of problems now being encountered by major players of the telecommunication world. Operators focus on new ideas and concepts to enable radical transformation of networks and service infrastructures. In order to achieve a success, the service provider should: develop an efficient transport network; offer and constantly change a huge number of value-added, improved services; construct business plan to make profits delivering those services.

Content delivery network (CDN) is an interesting and robust method to improve the Internet quality. CDN uses many servers offering the same content replicated in various locations. User-perceived latency and other quality of service parameters can be easily and inexpensively improved by various techniques of Web content caching. Every replicated system must deal with two fundamental issues—distributing requests to object replicas and deciding on placement of replicas. In this work we focus on the second problem. The issue of distributing requests to object replicas is strongly discussed in the literature.

In this work we address problems of CDN design in multiprotocol label switching (MPLS) environment. The MPLS approach proposed by the Internet engineering task force (IETF) is a networking technology that enables traffic engineering and QoS performance for carrier networks. MPLS is a connection-oriented technique, which is becoming a popular solution for backbone networks and must be taken into account in the design of Web replication system. Since the considered optimization problem is NP-complete, we propose an exact algorithm using the branch-and-cut approach. It must be noted that results of this work can be applied also to networks using other connection-oriented technologies (e.g., asynchronous transfer mode—ATM) or connectionless protocols (e.g., IP).

The paper is organized as follows. Section 2 presents a brief description of CDNs and Web server caching issues. In Section 3 we report on the previous work in the field of replica placement problems. In Section 4 we formulate an optimization problem of replica location in a CDN using MPLS. Section 5 contains an exact algorithm solving the replica location problem. In Section 6 we present and discuss results of numerical experiments. Last section concludes this work.

2. Content delivery networks and Web caching

Content delivery networks are defined as mechanisms to deliver a range of content to end users on behalf of origin Web servers. The original information is offloaded from source sites to other content servers located in different locations in the network. For each request, the CDN tries to find the closest server offering the requested Web page [16]. CDNs deliver the content from the origin server to replicas located much closer to end-users. The set of content stored in CDNs servers is selected carefully. Therefore, the CDNs' servers can approach the hit ratio of 100%. It means that almost all requests to replicated servers are satisfied. CDNs techniques are based on caching and replication of Web content. The general architecture of CDN system can be found in [23].

Caching is a technique typically applied to bring parts of an overall data set closer to its processing site [3]. A Web cache is an application residing between Web servers providing various content and clients that want to fetch the information [27]. Caching employs the knowledge acquired by several analyses on servers' access logs and by looking

into Web users behavior. Caching can reduce latency experienced by end users when trying to fetch some documents through their Web browser.

Replication can be considered as a kind of caching. Nevertheless, there is some dissimilarity. Replication presumes storing of an object at a place that cannot see the object, while caching is storing of an object at a place that sees the source object. It means that a cache notices both hit and miss requests. Since requests to replicated server arrives only if that server is believed to have a replica of the requested object, the replica notices only hits. In the presented sense, replica is sometimes called push cache [25]. Replication is perceived also as a caching system with only one source Web server generating content, while standard caching must serve a great number of Web servers [17].

An important issue to resolve is the choice between static and dynamic replica placement. In the static replica placement the system administrator, according to observed access and traffic statistics, decides where replicas should be located. Dynamic replica placement assumes that the system monitors access to various servers and adapts set of replicas to changing requirements [25].

One of the most important issues of Web caching is the mechanism used for requests redirection. Transparent replication assumes redirecting a client's request for a document to one of the physical replicas. The most popular practical and theoretical approaches of requests redirection: client multiplexing, IP multiplexing, DNS indirection, HTTP redirection and anycast, peer-to-peer routing have been discussed in [23, 25, 28].

Web caching and replication in CDNs are becoming popular for many reasons. The most important are [6, 28]: reducing the cost of using the Internet, reducing the latency of WWW, bandwidth will always have some cost, non-uniform bandwidth and latencies, network distances grow, bandwidth requirements continue to increase, hot spots in the Web will continue, costs of communication exceed costs of computations, traffic engineering requirements, the need for survivability.

For more information on WWW please refer to [34, 35].

3. Related work

An important issue in the design of robust and survivable CDN is the replica placement. In this section we examine the previous work on replica placement problems. For the context of this paper we are interested in static replica placement. The main problem of static replica placement is to develop effective algorithms for replica location. Some previous authors have developed such algorithms. According to [24], the first work in this area is [19]. Li *et al.* formulate in [19] a problem of proxies' location in a tree topology with the objective function of selection of proxies cost. A dynamic programming algorithm is proposed. The objective function can be calculated as the overall network latency if the link distance is associated with the cost function.

Authors of [17] take into account the cache location problem for transparent caches. The objective function is the cost of serving demands using a cache in a given location. Since the general problem is NP-complete, Krishnan *et al.* analyze only regular topologies: homogenous line, general line and ring.

Qiu *et al.* formulate in [24] problem of the placement of web server replicas as an uncapacitated k-median problem related to the facility location problem. They restrict the maximum number of replicas, but they don't restrict the number of requests served by each replica. The goal of the optimization process is to minimize the total cost of all requests defined as a sum of a distance between origin node and destination node over all requests. A greedy algorithm and a super-optimal algorithm based on the Lagrangian relaxation are proposed.

Guha *et al.* consider in [8] a generalization of the standard facility problem and introduce the requirement for fault-tolerant mechanisms. Every demand point is served by a number of facilities instead of just one. The closest facility is the working one, while other facilities serve as backup facilities. The objective function is a weighted combination of facilities locations' costs. An algorithm using the filtering technique and fractional demands is provided.

Authors of [10] present a simple and natural greedy algorithm for the metric uncapacitated facility location problem and k-median problem.

Arya *et al.* analyze in [2] a local search heuristics for facility location and k-median problems. The main operation of the proposed algorithm is swap, which includes closing one facility and opening another; clients of the closed facility are assigned to other facilities. In [4] an improved combinatorial approximation algorithms for the uncapacitated facility location and k-median problems are proposed and discussed.

The replica placement problem can be modeled as a center placement problem. The k-HST (k-hierarchically well separated tree) approach can solve this problem [11, 23].

Jamin *et al.* propose a topology-informed placement strategy, called "transit node". This heuristic applies the outdegree—information on the number of other nodes connected to a given node. It is assumed that a node with the highest outdegrees can reach more nodes with lower latency. Therefore, the servers are placed in nodes sorted in descending order of outdegrees [12, 23].

Wierzbicki formulates in [36] the Internet cache location problem in a CDN as a mixed integer programming (MILP). New models of cache location are proposed in order to overcome the limitations of the basic model. The complexity of the MILP formulation is evaluated.

The primary concern in most of works discussed above is analyzing the replica location problem as one of well-known optimization problems: k-facility location problem, k-median problem and center placement problem. The first problem consists of assignment of clients to k facilities that can be located in network nodes. The objective is to minimize the total cost including the connection cost of each

client and the facility cost. The k-median problem generally differs from the facility problem in one thing: there is no cost for opening facilities. The main element of both discussed problems is location of k facilities, i.e., selection of k nodes of the network for hosting a facility. Since one can select the closest replica in terms of connection cost, assignment of individual clients to a particular replica is much simpler. Capacity constraints on network links are not considered. However, in a capacitated version of facility location problem there is a capacity constraint on load served by each facility. The center placement problem consists of the placement of a given number of centers in order to minimize the maximum distance between a node and the nearest center.

4. Optimization problem of CDN design in MPLS environment

We propose a different approach then in previous works. Our model is much closer to problems encountered in real computer networks. The main difference is that we take into account capacity constraints on each link of the network. In many cases networks are congested. Therefore, the capacity resources must be used in effective manner. Furthermore, we consider an MPLS network that is a connection-oriented network, i.e., the flow is modeled as a non-bifurcated multicommodity flow. Most of the work in the field of replica placement considers pure IP networks using multicommodity bifurcated flow.

In this section we formulate the optimization problem of the content delivery network design using the MPLS technique. The problem is very close to the replica location (RL) problem discussed in [30–31]. We model the MPLS network flow as non-bifurcated multicommodity flow. However, results of this work can be also applied to connection-less networks. For more information on modeling of flow in MPLS network and non-bifurcated multicommodity flows, see [7, 14, 15, 18, 26, 29].

We begin presentation of the problem by introducing the notation. We will keep the same notation for the rest of the paper.

Indices:

- i used as subscript, denotes the number of considered client of CDN,
- j used as subscript, denotes the number of considered arc or node,
- r used as subscript, denotes the number of considered selection of clients or routes,
- k used as superscript, denotes the number of a route.

Sets:

- V set of $|V|$ vertices representing the network vertices (nodes),
- A set of $|A|$ arcs representing directed links,
- R set of $|R|$ CDN's content servers (replicas); each server must be located in a network vertex,
- P set of $|P|$ CDN's clients; each client is defined by the source vertex s_i, destination vertex t_i and bandwidth requirement Q_i; for each client a set of route proposals is given,
- Π_i set of routes proposals for a client i; $\Pi_i = \{\pi_i^k : k = = 1, \ldots, |\Pi_i|\}$; each route ends in the source node of client i,
- Z_r set of location variables z_i equal to one; the set Z_r is called a selection; each selection Z_r determines the unique assignment of replicas to network nodes,
- X_r set of route selection variables x_i^k equal to one; the set X_r is called a selection; each selection X_r determines the unique set of routes between clients and replicas.

Decision variables:

- z_i binary variable, which is equal to one if a replica is located in the node i and is equal to zero otherwise,
- x_i^k binary variable, which is equal to one if the client i uses the route π_i^k and is otherwise equal to zero.

Other variables:

- f_{jr} flow in link j calculated according to routes defined in selection X_r.

Constants:

- c_j capacity of arc j,
- $C(j)$ capacity of all arcs leaving the node j,
- Q_i bandwidth requirement for a client i,
- $Q(j)$ bandwidth requirement of all clients located at node j,
- a_{ij}^k binary variable, which is equal to one if the jth arc belongs the route π_i^k and is otherwise equal to zero,
- u_{ij} binary variable that equals one if the source node of the arc i is node j,
- u_{ij}^k binary variable that equals one if the source node of the route π_i^k is node j.

We assume that traffic between a replica and a set of clients connected to one node can be aggregated to one or more LSPs. Since clients receive more data than is sent to replicas, we assume that traffic between clients and replicas is generally asymmetric and we ignore the flow from a client to a replica.

The optimization problem of replica location in a CDN is formulated as follows:

$$\min_{X_r, Z_r} D(X_r, Z_r) = \sum_{j \in A} f_{jr} \tag{1}$$

subject to

$$f_{jr} = \sum_{i \in P} \sum_{\pi_i^k \in \Pi_i} a_{ij}^k x_i^k Q_i \quad \forall j \in A, \tag{2}$$

$$\sum_{j \in V} z_j = |R|, \tag{3}$$

$$\sum_{\pi_i^k \in \Pi_i} x_i^k = 1 \quad \forall i \in P, \tag{4}$$

$$f_{jr} \leq c_j \quad \forall j \in A, \tag{5}$$

$$\sum_{j\in V}\sum_{\pi_i^k\in\Pi_i} x_i^k y_j u_{ij}^k = 1 \quad \forall i \in P, \tag{6}$$

$$z_j \in \{0,1\} \quad \forall j \in V, \tag{7}$$

$$x_i^k \in \{0,1\} \quad \forall i \in P;\ \pi_i^k \in \Pi_i. \tag{8}$$

The objective function (1) is the overall flow in the CDN generated by clients. Note that if we introduce a link metric the objective function could represent cost, network latency or other function. Equation (2) is a definition of a link flow. Constraint (3) guarantees that the number of established replicas (content servers) equals the defined number of replicas. We assume that during the CDN design we know how many replicas may be located. The number of replicas can be calculated according to the budget of CDN. The overall budget is divided by the cost of one content server. Thus, we obtain the number of replicas that can be afforded for the particular budget. Constraint (4) ensures that each client uses only one route. Constraint (5) is a capacity constraint. Constraint (6) guarantees that each selected route starts in a node that has a replica. Constraints (7) and (8) ensure that decision variables are binary ones. The condition (8) ensures that the considered flow is non-bifurcated as in MPLS networks. If we relax the constraint (8) to the formula given below, the flow becomes a bifurcated multicommodity flow:

$$0 \le x_i^k \le 1 \quad \forall i \in P;\ \pi_i^k \in \Pi_i.$$

Thus, we obtain the replica location problem for protocols using the bifurcated flow, for instance IP protocol.

Note, that in the problem Eqs. (1)–(8) we don't limit the amount of service that can be provided at any replica. According to [24], it is a reasonable assumption, since increasing the number of replica sites is much more difficult than increasing the capacity of a replica. The number of replicas is frequently given a priori due to cost and administrative reasons, while the capacity constraint can be overcome by adding more machines. Since in many cases the replication traffic and cost of replicas managing can be ignored, we ignore the cost of replica location.

The problem Eqs. (1)–(8) is NP-complete because it has more constraints then the non-bifurcated flow problem which is NP-complete according to [13].

Joint optimization of replica location, clients' assignment and routes' selection must be carried out to find a globally optimal solution of the objective function for a projected traffic demand. Since the optimization is conducted jointly over location and route selection variables, the complexity of the problem grows tremendously. An interesting approach is to partition the problem into two simpler problems: first optimize replica location and next find clients' assignment for already established replicas.

The first subproblem, called only replica location (ORL) consists of selection of $|R|$ nodes to host a replica. This problem is very close to problem RL (1)–(8). However, we don't take into account assignment of clients to replicas. Therefore, we can ignore constraints (2), (4)–(6) and (8). As an objective function we use the function $D(Z_r)$ defined as a solution of clients' assignment to replicas given by the selection Z_r.

The second subproblem is to assign each client i to one replica according to selected criterion. In the optimization problem of clients' assignment to replicas (CATR) we assume that replicas are already located in network nodes and the main goal is to assign clients to replicas minimizing the overall flow. The CATR optimization problem is formulated as follows:

$$\min_{X_r} D(X_r) = \sum_{j\in A} f_{jr} \tag{9}$$

subject to Eqs. (2), (4), (5) and (8).

The CATR problem is similar to the classical non-bifurcated multicommodity flow problem (NBMC) extensively discussed in the literature [7, 14, 33]. The main difference is that in the CATR problem besides route selection for each client we must decide on which replica the client should be assigned to. It is an additional constraint. Since the NBMC problem is NP-complete [13], the CATR problem is also NP-complete.

To solve the ORL problem we must consider many CATR subproblems. For each location of replicas, in order to find the objective function, we must estimate the network flow by assigning clients to already located replicas. For this purpose exact or heuristic algorithms can be used. If a heuristic algorithm treats at least one of ORL or CATR subproblems, the obtained solution of the RL problem cannot be called an optimal one. However, this approach can reduce size of the problem and consequently shorten execution time of the algorithm.

5. Exact algorithm

Optimization of the Web replica placement is a difficult task. In many real life cases, replicas or proxies are placed in fairly obvious nodes, e.g., the Internet service provider gateway [19]. However, in order to improve network parameters some algorithms must be applied to provide optimal or sub-optimal solutions.

As mentioned above, the RL problem is NP-complete. Therefore, heuristic algorithms not always ensure that the solution is optimal. To obtain an optimal solution an exact algorithm must be applied. To construct such an algorithm we propose to use the branch-and-cut (B&C) approach, which is a modification of the branch-and-bound method (B&B). The branch-and-bound approach has become a general solution method for various integer and mixed integer problems. The B&B algorithm is an intelligently structured search over the space of all feasible solutions. The solution space is repeatedly partitioned into smaller subsets, and a lower bound of the objective function is calculated within each subset. Subsets with bound that exceeds the best solution are excluded from further partitioning. For more information on branch-and-bound algorithms refer to [20].

Branch-and-cut is a relatively new but well accepted method proposed by Padberg and Rinaldi [22] for the traveling salesman problem. B&C algorithm is a combination of cutting plane algorithm and branch-and-bound algorithm. Cutting plane procedures are introduced into the bounding phase of B&B, enabling the branching phase to utilize the information on the known cuts, what improves the relaxation of the problem and enables calculation of more effective bounds. The B&C algorithm solves strengthened continuous relaxations of the problem, resulting in fewer analyzed nodes than for the B&B algorithm. The reader interested by branch-and-cut approach is referred to [1, 9, 21, 22].

It must be underlined that in order to find the exact solution of RL we must solve both subproblems concurrently. In this section we focus on the ORL problem and propose a branch-and-cut algorithm to solve this problem. The algorithm guarantees that we analyze the whole solution space of all possible combinations of replica location. However, for each analyzed selection Z_r we must solve the CATR subproblem. If we solve CATR by an exact algorithm, the obtained solution is globally optimal. Otherwise, if we tackle CATR with an heuristic algorithm, the solution can be claimed to be optimal.

5.1. Calculation scheme

In our branch-and-cut algorithm we start with selection Z_1 and generate a sequence of selections Z_r. In order to obtain the initial selection Z_1 we can solve the RL problem using one of heuristic algorithms proposed in [30, 31]. Each new selection Z_s is obtained from a certain selection Z_r of the sequence by complementing a normal variable z_i by a reverse variable z_k in the following way $Z_s := (Z_r - \{z_i\}) \cup \{z_k\}$. It means that we shift the replica from node i to a node k. The generating process can be represented as a branch and bound decision tree. Each node of the decision tree represents a selection. We say that the selection Z_s is a successor of the selection Z_r if there is a path from Z_r to Z_s.

For each set Z_r we constantly fix a set of nodes U_r. The state of nodes included in U_r cannot be changed. It means that nodes included in the set U_r cannot be used in the selection process. If the selection Z_s is obtained from the selection Z_r as $Z_s := (Z_r - \{z_i\}) \cup \{z_k\}$ we update the U_s as follows: $U_s := U_r \cup \{i\}$. There are two key elements of the branch-and-cut algorithm: lower bound of criterion function and branching rules. The lower bound is calculated to check if a "better" solution may be found. If the test result is negative we abandon the considered selection Z_r and backtrack to the selection Z_p from which the selection Z_r was generated. If Z_r was obtained from the selection Z_p in the following way $Z_r := (Z_p - \{z_i\}) \cup \{z_k\}$ we update the U_p as follows: $U_p := U_p \cup \{i\}$. It is a consequence of the fact that variables z_i are binary ones; and if we analyze all selections for which $z_i = 0$ we may constantly fix node i with $z_i = 1$. It must be noted that in branch-and-cut algorithm the lower bound calculation is enriched with the valid inequalities, which can "cut" the solution space.

The basic task of the branching rules is to find the variables for complementing to generate a new selection with the lowest value of criterion function possible. Since in the algorithm we change only location of replica, we use the function D given by (1) as the objective function. However, in order to calculate value of this function we must solve the CATR problem. During the branching operation of the tree we add a node i without a replica (the current variable $z_i = 0$) to the set U_r. When we backtrack, a node i hosting a replica is included in the set of fixed nodes (the current variable $z_i = 1$).

5.2. Branching rules

We define two sets as follows:

$$E_r = \left(\bigcup_{j \in (N - U_r)} \{ j : z_j = 0 \} \right),$$

$$M_r = \left(\bigcup_{j \in (N - U_r)} \{ j : z_j = 1 \} \right).$$

The set E_r comprises all nodes that are not constantly fixed for Z_r and can be selected for complementing. The set M_r includes all nodes that are not constantly fixed for Z_r and can be selected for removing a replica. Since, due to condition (3) the number of replicas must be equal to $|R|$, in the branching rule for a successor of Z_r we must remove a replica from a node hosting a replica, i.e., a node included in the set M_r and locate this replica in a node incorporated in the set E_r.

In order to explain the branching rule we introduce a new function $d_i(Z_r)$ defined as a distance from the node i to the closest replica included in the selection Z_r. To find the $d_i(Z_r)$ we consider only routes from the set Π_i. Using the $d_i(Z_r)$ we define the following function:

$$G(Z_r) = \sum_{i \in P} Q_i \, d_i(Z_r). \tag{10}$$

The function $G(Z_r)$ is only an estimation of the $D(Y_r)$. However, the main benefit of the function $G(Y_r)$ compared to $D(Y_r)$ is that it can be easily calculated. Next, we introduce the function $swap(r,i,k)$ used for selection of the variables for complementing. Without loss of generality, we assume that the selection Z_s is obtained from the selection Z_r in the following way $Z_s := (Z_r - \{z_i\}) \cup \{z_k\}$. According to the above discussion $i \in M_r$ and $k \in E_r$. The function $swap(r,j,k)$ is defined as follows:

$$swap(r,i,k) = G(Z_s) - G(Z_r). \tag{11}$$

According to definition (11), $swap(r,i,k)$ is a "gain" we obtain by moving the replica from the node i to the node k. As mentioned above, the function $G(Z_r)$ used in the definition (11) is an estimate of the objective function $D(Z_r)$. Since in the branching rule we want to generate a new selection and minimize the objective function $D(Z_r)$, we propose to use the function (11) as the decision function.

5.3. Lower bound

The simplest way to calculate a lower bound of an optimization problem is to relax some constraints in order to obtain a much simpler optimization problem in terms of computational complexity. In this case we relax the capacity constraint (5). Therefore, clients can be assigned to the closest node excluding nodes abandoned while generating the decision tree (we don't consider fixed nodes $i \in U_r$ for which $z_i = 0$).
Let N_r denote a set of fixed nodes $i \in U_r$ for which $z_i = 1$:

$$N_r = \bigcup_{j \in U_r} \left\{ j : z_j = 1 \right\} \Bigg) .$$

For the current selection $|N_r|$ replicas are constantly located. It means that the number of replicas to be located is $(|R| - |N_r|)$. These replicas can be placed only in nodes included in the set $(V - U_r)$. Let $\underline{Z_r}$ denote a set of feasible selections that can be generated from the selection Z_r. We assume that the set $\underline{Z_r}$ compromises also the selection Z_r. According to discussion presented above, the following formula defines the number of elements of the set $\underline{Z_r}$:

$$|\underline{Z_r}| = \left(\begin{array}{c} |V - U_r| \\ |R| - |N_r| \end{array} \right) .$$

Now we introduce the cutting inequality. Let $C(j)$ denote the capacity of all arcs leaving the node j:

$$C(j) = \sum_{i \in A} Q_i u_{ij} \tag{12}$$

Recall that u_{ij} is a binary variable that equals one if the source node of the arc i is node j. Due to the capacity constraint (5), a replica located in node j can serve at most $C(j)$ flow. Consequently, $C(Z_r)$ denotes the upper bound of flow that can be served by replicas located in nodes given by the selection Z_r:

$$C(Z_r) = \sum_{j \in V} y_j C(j) . \tag{13}$$

The following formula is applied as a cutting plane in the lower bound:

$$C(Z_r) \geq \sum_{j \in V} (1 - y_j) Q(j) . \tag{14}$$

The inequality (14) indicates whether or not the location of replicas given in selection Z_r can serve all demands in the network. In the right-hand side we sum bandwidth requirements of demands located in network nodes except for nodes, where replicas are placed.
Let Ψ_r denote a set of selections $Z_r \in \underline{Z_r}$ for which the inequality (14) is satisfied. Note that formula (10) defines a lower bound of the objective function for Z_r. In order to find a lower bound for the selection Z_r and all its successors we apply the following formula:

$$LB_r = \min_{Z_r \in \Psi_r} G(Z_r) . \tag{15}$$

In formula (15) we analyze all feasible (in terms of the cutting inequality and fixed variables) selections that can be generated from current selection Z_r. If inequality (14) is satisfied, we calculate the function $G(Z_r)$ for considered replica location. Otherwise, we skip the given selection. Therefore, we perform fewer calculations of $G(Z_r)$. Since to obtain the $LB(Z_r)$ we relax the capacity constraint of the problem Eqs. (1)–(8), the LB_r is a lower bound of the objective function for the selection Z_r and all feasible selections that can be generated from Z_r. The elementary operation of lower bound consists of checking the inequality (14) and if it is satisfied, we must calculate $G(Z_r)$. Otherwise, when the cut (14) fails, we don't examine the given selection any further. To find $G(Z_r)$ we must find the shortest route to the replica for every client. Checking the cut (14) is much simpler, since values of $C(j)$ and $Q(j)$ are constant.
Note that in classic B&B algorithm the following formula can be used as lower bound:

$$LB_r = \min_{Z_r \in \underline{Z_r}} G(Z_r) . \tag{16}$$

In formula (16) we don't use the cutting inequality. Therefore, for every selection $Z_r \in \underline{Z_r}$ the value of function $G(Z_r)$ must be found.

5.4. Algorithm

The problem RL (5–12) can be solved using the following algorithm. Let Z_1 denote a feasible initial solution. Set $U_1 := \varnothing$, $D^* := \infty$. The current selection is denoted by Z_r. Let LB_r be a lower bound of Z_r given by (15). We start with $r := 1$.

Step 1: Compute LB_r (15). If $LB_r \geq D^*$ go to Step 4. Otherwise if $LB_r < D^*$ go to Step 2.

Step 2: Compute $D(Z_r)$. If there is a feasible solution of $D(Z_r)$ and $D(Z_r) < D^*$ then set $D^* := D(Z_r)$. Go to Step 3. Otherwise, if there is no feasible solution of $D(Z_r)$ go to Step 4.

Step 3: If $E_r = \varnothing$ or $M_r = \varnothing$ go to Step 4. Otherwise find $i \in M_r$ and $k \in E_r$ for which the value of $swap(r,i,k)$ is lowest. Generate the selection Z_s (successor of Z_r) as follows $Z_s := (Z_r - \{z_i\}) \cup \{z_k\}$, $U_s := U_r \cup \{i\}$. Go to Step 1.

Step 4: Backtrack to the predecessor Z_p of the selection Z_r. If the Z_r has no predecessor, stop the algorithm. The selection Z^* associated with the current D^* is the optimal solution. Otherwise, if Z_r has predecessor, drop the data for Z_r and update data for Z_p as follows. If Z_r has been generated as $Z_r := (Z_p - \{z_i\}) \cup \{z_k\}$ then set $U_p := U_p \cup \{i\}$. Go to Step 1.

To obtain the value of function $D(Z_r)$ calculated in Step 2 we must solve the CATR problem for the particular location of servers given by the selection Z_r.

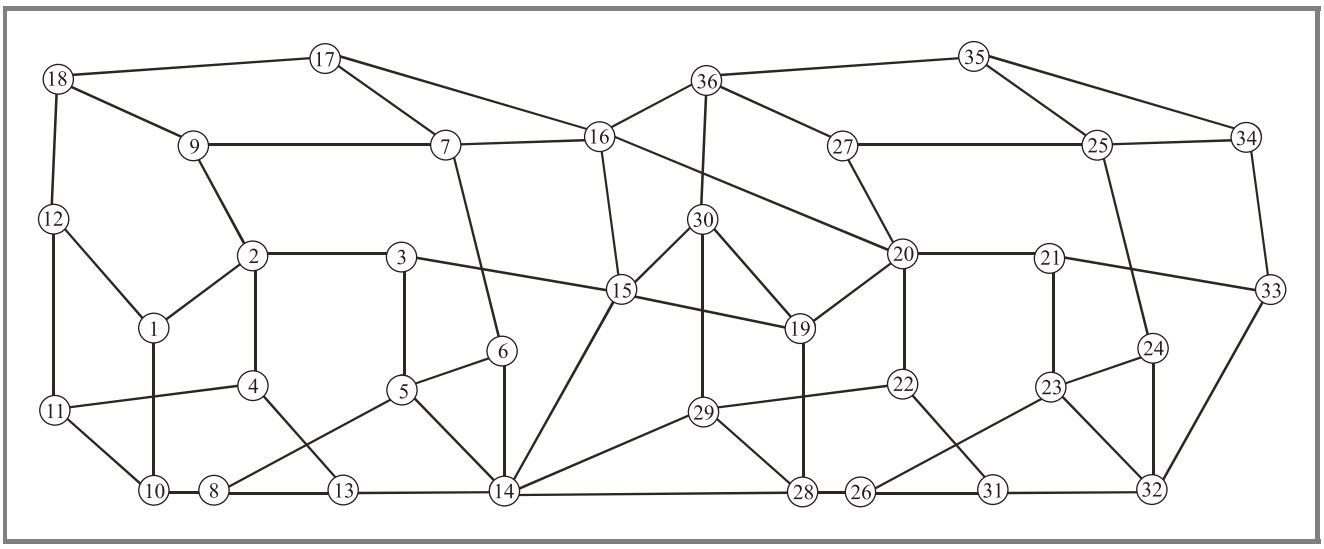


Fig. 1. Tested network.

6. Results

In this section we present results of numerical experiments. The B&C algorithm proposed in previous section was coded in C++. As mentioned above, there is a joint dependency between the replica location and the assignment of routes. Therefore, if a heuristic algorithm solves the CATR subproblem, the solution of the RL problem obtained cannot be called an optimal one. The CATR problem is very complex, even for small networks. Therefore, we decided to use a heuristic algorithm based on the flow deviation method [7] to find feasible solutions of CATR in Step 2 of B&C algorithm. Obviously, the solution of RL problem obtained cannot be called optimal. However, FD algorithm is a very effective method for solving multicommodity flow problems [4, 14, 33]. Consequently, the B&C is used as an intelligent method of searching the solution space of replica location problem.

Results presented in this section are obtained from simulations on a sample networks having 36 nodes and 128 arcs (Fig. 1). Arcs of tested network have various capacities in the range from 2 000 BU (bandwidth units) to 6 000 BU. In the experiment, it is assumed that in every network node there are 5 demands to a CDN server (replica). It means that there are overall 180 clients in the network. For a particular experiment bandwidth requirements are the same for all clients.

We have considered 6 scenarios. In Cases A, B, C and D there are 3 replicas to be located. The starting solutions indicating nodes hosting replicas are $\{2,16,32\}$; $\{1,16,32\}$; $\{1,14,32\}$; $\{2,30,32\}$ respectively for Cases A, B, C and D. In experiment E there are 2 replicas to be located and the initial solution is $\{14,25\}$. Finally, for Case F, 4 replicas are to be placed and the starting solution is $\{2,14,16,30\}$. Initial solutions are found heuristically. We studied the performance of the algorithm for increasing traffic load, examining the evolution of the network status towards a saturation condition. In particular, for every scenario we examine 26 demand patterns having the value of one client's demand between 200 BU and 450 BU.

The first objective of experiments was to investigate how increasing replicas' number changes the network overall flow. In Fig. 2 we report performance of B&C algorithm for Scenarios A, E and F for which the number of replicas to be placed is 2, 3 and 4, respectively. The *x*-axis is the total demand in the network (sum of all clients' demands), and the *y*-axis is the network flow (objective function of the RL problem). It is obvious that increasing the number of replicas decreases the network flow. More replicas means that a replica is closer to clients and the route to the replica is shorter. In Scenario E (2 replicas) the algorithm finds a feasible solution only for first 6 demand patterns. For 3 replicas, 23 of 26 considered demand patterns yield feasible result. Finally, for the last scenario having 4 CDN servers all demand patterns are satisfied.

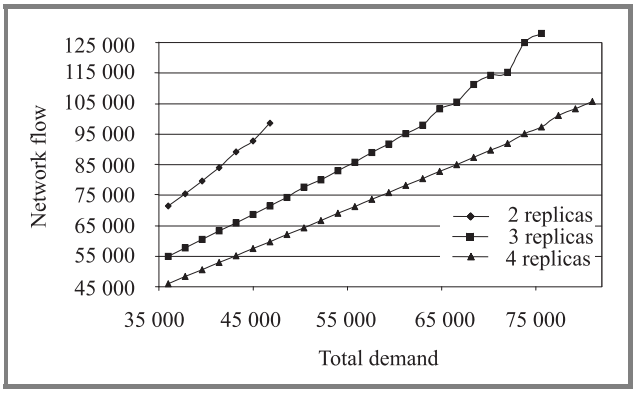


Fig. 2. Network flow as a function of total demand in the network for various number of replicas.

Since the branch-and-cut approach is a relatively new method compared to branch-and-bound algorithm, we made several tests to evaluate performance of B&C against B&B.

To obtain a B&B algorithm we slightly modified the algorithm developed in previous section and applied the lower bound given by (16). All other operations of the B&B algorithm are the same as for B&C.

First, we show how the B&C algorithm reduces the number of nodes in the solution tree of the algorithm. Figure 3 plots the number of nodes in the solution tree for Scenarios A and E as a function of total demand in the network. We show results for both algorithms: B&B and B&C. The *x*-axis is the total demand in the network. The *y*-axis uses logarithmical scale and denotes the number of nodes in the decision tree of the algorithm. We observe similar performance between the bars, reflecting A and E scenarios. For low load (according to the number of replicas), both algorithms need the same number of nodes in the decision tree. For more saturated network, B&C produces significantly less nodes than the B&B. Similar trend can be observed for other scenarios. Summarizing over all experiments, B&B algorithm produces 26 186 nodes, while for B&C the corresponding value is 19 958 nodes. The biggest difference is observed for highly saturated demand pattern in Scenario C, for which B&C needs only 104 nodes compared to 4 184 nodes of B&B. This proves that the branch-and-cut algorithm is more effective then the B&B one. For the problem considered, B&C outperforms B&B, especially for large traffic load that leads to network saturation.

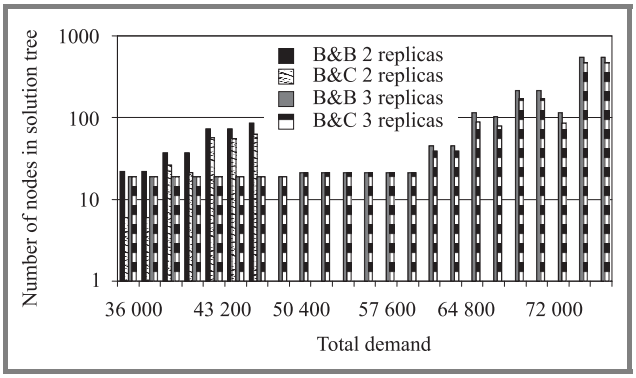


Fig. 3. Number of nodes in the solution tree for B&B and B&C algorithms.

To confirm the advantage of B&C over B&C we present further results. Recall that the main benefit of B&C algorithm is the use of cutting inequality (14) that enables reduction of calculations of function $G(Z_r)$ given by (10). Figure 4 shows the number of elementary operations performed in B&C algorithm. There are two types of elementary operations. The first type is applied when the cut (14) is satisfied and calculation of the function $G(Z_r)$ given by (10) is needed. The second operation consists only of checking the cut inequality and it is used when the cut inequality doesn't hold. The *x*-axis is the total demand in the network, and the *y*-axis denotes the number of operations. Figure 4 shows present results for Scenarios A and D. Generally the trend is the same for both cases. For low loaded networks the number of cuts exceeds the number of function *G* calculations. However, for this experiments the number of decision tree nodes is relatively small. For higher demand patterns curves become stable, number of cuts is about 3 times lower than number of the function *G* calculations. In these cases the number of decision tree nodes grows, and the lower bound is calculated for different selections and combinations of fixed nodes. Similar trend was observed for other scenarios.

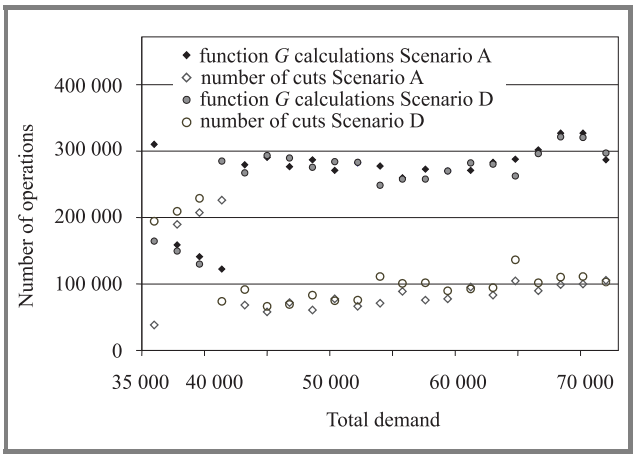


Fig. 4. Number of elementary operations for B&C algorithm.

Recall that for B&B algorithm we don't use the cutting plane. This creates additional overhead. For each analysed selection in the lower bound we must calculate the formula (10). Therefore, for the B&B algorithm the number of function *G* calculations is equal to or bigger than the sum of all elementary operations (of both types) in B&C algorithm.

Next, we present the execution time of B&C and B&C algorithms. The program implementing both algorithms was run on an IBM-compatible PC with 2 GHz Intel processor and 512 MB of RAM. It is worth remarking that decision time does not include I/O time for input of various files. It includes only the time of design output. Figure 5 depicts the decision time of both algorithms for Scenarios A and C.

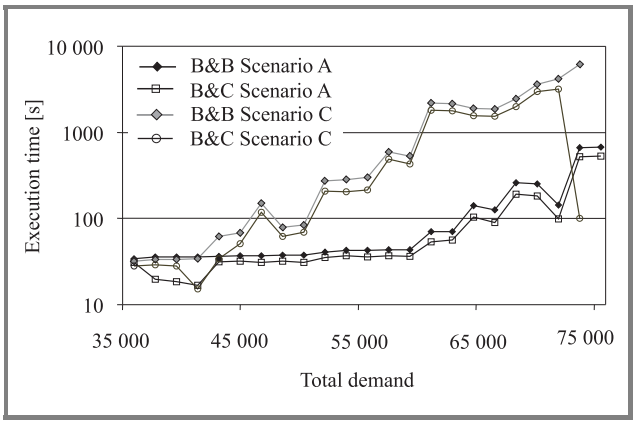


Fig. 5. Execution time of B&B and B&C algorithms for Scenarios A and C.

The *x*-axis is the total demand in the network. The *y*-axis uses logarithmical scale and denotes the decision time in seconds. We observe that B&C outperforms B&B for all demand patterns considered. The gap is similar for different network loads. It is worth remarking that when the network load grows, the execution time also increases.
Comparing Figs. 3 and 4 to Fig. 5 we can reach interesting conclusions. Analysis of B&C and B&B decision times obtained for various demand patterns shows only slight differences. On the other hand, observation of decision nodes' number and effectiveness of the cut inequality shows many differences in performance for various total loads in the network. For low loads the cut inequality works more effectively and gives relatively more positive tests. For more saturated networks the B&C produces much less solution tree nodes then B&B. Thus, these two effects combine to yield similar performance of B&C and B&B in terms of decision time for all considered demand patterns.
It should be noted that the decision time of B&B and B&C algorithms is influenced strongly by the execution time of the heuristic algorithm applied to solve the CATR subproblem. It was observed that the execution time of heuristic algorithm depends on the solved problem; the time is not constant. Therefore, analysis of the exact algorithms' decision time only from the perspective of decision tree nodes' numbers or effectiveness of cut inequality is not always sufficient.

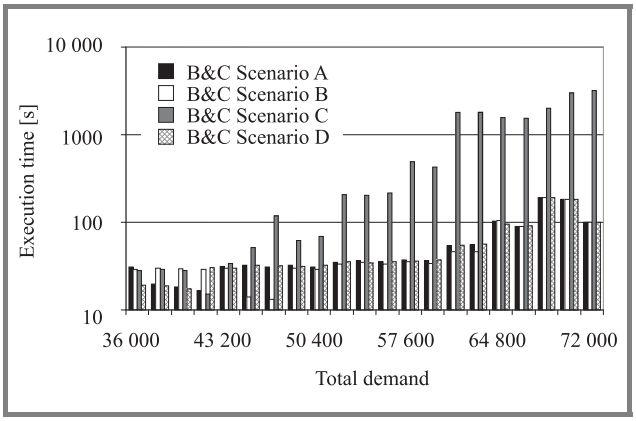


Fig. 6. Execution time of B&C algorithm for Scenarios A, B, C, and D.

Another important issue we have examined is the impact of the starting solution on the performance of the B&C algorithm. Figure 6 shows the decision time of B&C algorithm for Scenarios A, B, C and D. Recall that all these cases have 3 replicas to be located; however the initial solution of each scenario is different. The *x*-axis is the total demand in the network. The *y*-axis uses logarithmical scale and denotes the decision time in seconds. We observe very similar performance for three bar series, reflecting Scenarios A, B and D. For Case C the performance is much worse and the decision time is about 16 times longer then for other cases. This becomes evident when we analyse the quality of starting solutions applied in individual scenarios. The starting solution used in Scenario C gives an average result about 10% worse than the result obtained for B&C. The corresponding difference is 1%, 7% and 5% for Scenarios A, B and D, respectively. We can conclude that the starting solution is an important issue in the B&C algorithm, which has a strong effect on the execution time.
In summary, we must underline that experimental data showing comparison of B&B and B&C methods is reasonably well explained by the theoretical foundations of both algorithms presented in previous section.

7. Conclusion

This paper deals with the problem of replica location in a content delivery network. We have presented and discussed basic information on CDNs and MPLS. We have formulated an optimisation problem of replica location in a CDN. The network flow has been modelled as a connection-oriented flow. Furthermore, the capacity constraint has been incorporated into the model. This problem is NP-complete. The objective function is the overall flow in the network. To our knowledge, this problem has not received much attention in the literature. Using optimisation model, an exact algorithm based on the branch and cut approach has been developed. Two main operations of the algorithm: lower bound and branching rule have been discussed in detail. Results of numerical experiments have been discussed. From both experimental and analytical viewpoints, we have concluded that when applied to replicas location problem, the branch-and-cut algorithm outperforms branch-and-bound method in terms of execution time and number of analysed nodes of the decision tree. In future work we want to make more extensive tests in order to evaluate this algorithm and compare it with other algorithms.

References

[1] K. I. Aardal and C. P. M. van Hoesel, "Polyhedral techniques in combinatorial optimization II: computations and applications", *Stat. Nederl.*, vol. 53, pp. 129–178, 1999.

[2] V. Arya, N. Garg, R. Khandekar, A. Meyerson, and V. Pandit, "Local search heuristics for *k*-median and facility location problems", in *Proc. ACM Symp. Theory Comput.*, Hersonissos, Crete, Greece, 2001, pp. 21–29.

[3] M. Baentsch, L. Baum, G. Molter, S. Rothkugel, and P. Sturm, "World wide Web caching: the application-level view of the Internet", *IEEE Commun. Mag.*, pp. 170–178, June 1997.

[4] J. Burns, T. Ott, A. Krzesiński, and K. Muller, "Path selection and bandwidth allocation in MPLS networks", *Perform. Eval.*, vol. 52, pp. 133–152, 2003.

[5] M. Charikar and S. Guha, "Improved combinatorial algorithms for the facility location and *k*-median problems", in *Proc. IEEE Symp. Found. Comput. Sci.*, New York, USA, 1999, pp. 378–388.

[6] B. Davison, "The design and evaluation of Web prefetching and caching techniques", Ph.D. thesis, 2002, http://www.cse.lehigh.edu/∼brian/pubs/2002/thesis/

[7] L. Fratta, M. Gerla, and L. Kleinrock, "The flow deviation method: an approach to store-and-forward communication network design", *Networks*, vol. 3, pp. 97–133, 1973.

[8] S. Guha, A. Meyerson, and K. Munagala, "Improved approximation algorithms for fault-tolerant facility location", in *Proc. ACM-SIAM Symp. Discr. Algor.*, Washington, USA, 2001, pp. 636–641.

[9] O. Gunluk, "Branch-and-cut algorithm for capacitated network design problems", *Math. Programm.*, vol. 86, pp. 17–39, 1999.

[10] K. Jain, M. Mahdin, and A. Saberi, "A new greedy approach for facility location problems", in *Proc. ACM Symp. Theory Comput.*, Montreal, Canada, 2002.

[11] S. Jamin, Ch. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang, "On the placement of Internet instrumentation", in *Proc. IEEE INFOCOM 2000*, Tel-Aviv, Israel, 2000, pp. 295–304.

[12] S. Jamin, Ch. Jin, Y. Jin, A. Kurc, D. Raz, and Y. Shavitt, "Constrained mirror placement on the Internet", in *Proc. IEEE INFOCOM 2001*, Anchorage, Alaska, USA, 2001, pp. 31–40.

[13] R. Karp, "On the computational complexity of combinatorical problems", *Networks*, vol. 5, pp. 45–68, 1975.

[14] A. Kasprzak, *Topological Design of Wide Area Networks*. Wrocław: Wrocław University of Technology Press, 2001.

[15] L. Kennington, "A survey of linear cost multicommodity networks flows", *Oper. Res.*, vol. 26, pp. 209–236, 1978.

[16] B. Krishnamurthy, C. Wills, and Y. Zhang, "On the use and performance of content delivery networks", in *Proc. ACM SIGCOMM Internet Measur. Worksh.*, San Francisco, USA, 2001.

[17] P. Krishnan, D. Raz, and Y. Shavitt, "The cache location problem", *IEEE/ACM Trans. Netw.*, vol. 8, pp. 568–582, 2000.

[18] T. Li, "MPLS and the evolving Internet architecture", *IEEE Commun. Mag.*, pp. 38–41, Dec. 1999.

[19] B. Li, M. Golin, G. Italiano, X. Deng, and K. Sohraby, "On the optimal placement of Web proxies in the Internet", in *Proc. IEEE INFOCOM'99*, New York, USA, 1999, pp. 1282–1290.

[20] J. Mitchell and E. Lee, "Branch-and-bound methods for integer programming", in *Handbook of Applied Optimization*. Oxford: Oxford University Press, 2002.

[21] J. Mitchell, "Branch-and-cut methods for combinatorial optimization problems", in *Handbook of Applied Optimization*. Oxford: Oxford University Press, 2002.

[22] M. Padberg and G. Rinaldi, "Optimization of a 532-city traveling salesman problem by branch-and-cut", *Oper. Res. Lett.*, vol. 6, 1987.

[23] G. Peng, "CDN: content distribution networks", Tech. Rep., 2003, http://www.sunysb.edu/tr/rpe13.ps.gz

[24] L. Qiu, V. Padmanabhan, and G. Voelker, "On the placement of Web server replicas", in *Proc. IEEE INFOCOM 2001*, Anchorage, Alaska, USA, 2001, pp. 1587–1596.

[25] M. Rabinovich, "Issues in Web content replication", *Data Eng. Bull.*, vol. 21, no. 4, 1998.

[26] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture", RFC 3031, Jan. 2001.

[27] A. Vakali, "An evolutionary scheme for Web replication and caching", in *Proc. 4th Int. Web Cach. Worksh.*, San Diego, USA, 1999.

[28] K. Walkowiak, "Designing of survivable Web caching", in *Proc. 8th Polish Teletraffic Symp.*, Zakopane, Poland, 2001, pp. 171–181.

[29] K. Walkowiak, "Modelling of Web server replication system in MPLS networks", in *Proc. Inform. Syst. Model. ISM 2002*, Roznov pod Radhostem, Czech Republic, 2002, pp. 213–220.

[30] K. Walkowiak, "A new exact algorithm for Web replica location problem in MPLS networks", in *Proc. Polish-Germany Teletraffic Symp. PGTS 2002*, Gdańsk, Poland, 2002, pp. 307–314.

[31] K. Walkowiak, "Some approaches to solve a Web replica location problem in MPLS networks", in *Internet Technologies, Applications and Societal Impact*, W. Cellary and A. Iyengar, Eds. Boston [etc.]: Kluwer, 2002, pp. 61–72.

[32] K. Walkowiak, "On application of genetic algorithms to replica location problem", in *Proc. Comput. Recogn. Syst. KOSYR 2003*, Milków, Poland, 2003, pp. 445–450.

[33] K. Walkowiak, "A new approach to survivability of connection oriented networks", in *Lectures Notes in Computer Science*. Berlin, Heidelberg: Springer-Verlag, 2003, vol. 2657, pp. 501–510.

[34] J. Wang, "A survey of Web caching schemes for the Internet", *ACM Comput. Commun. Rev.*, pp. 36–46, Oct. 1999.

[35] B. Williams, "Transparent Web caching solutions", in *Proc. 3rd Int. WWW Cach. Worksh.—TF-Cache Meet.*, Manchester, England, 1998.

[36] A. Wierzbicki, "Internet cache location and design of content delivery networks", in *Lectures Notes in Computer Science*. Berlin, Heidelberg: Springer-Verlag, 2002, vol. 2376, pp. 69–82.

Krzysztof Walkowiak was born in Poland in 1973. He received the M.Sc. and Ph.D. degrees in computer science from Wrocław University of Technology in 1997 and 2000, respectively. Since 2001 he has been an Assistant Professor at the Chair of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Technology. His research interest is mainly focused on optimization of connection-oriented networks, survivability issues of ATM, MPLS, and application of soft-optimization techniques for design of computer networks, Web caching.

e-mail: Krzysztof.Walkowiak@pwr.wroc.pl
Chair of Systems and Computer Networks
Faculty of Electronics
Wrocław University of Technology
Wybrzeże Wyspiańskiego st 27
50-370 Wrocław, Poland

Paper

PFS scheme for forcing better service in best effort IP network

Monika Fudała and Wojciech Burakowski

Abstract—The paper presents recent results corresponding to a new strategy for source traffic generating, named priority forcing scheme (PFS), allowing Internet users for getting better than best effort service in IP network. The concept of PFS assumes that an application, called PFS application, sends to the network a volume of additional traffic for the purpose of making the reservations for the data traffic in the overloaded router queues along the packet path in the IP network. The emitted redundant packets, named R-packets, should be rather of small size comparing to the data packets, named D-packets. The PFS scheme assumes that the R-packets waiting in a queue can be replaced by the arriving D-packets and belonging to the same flow. In this way, the D-packets can experience a prioritised service comparing to the packets produced by a non-PFS application. Notice that the proposed solution does not require any quality of service (QoS) mechanisms implemented in the network, like scheduler, dropping, marking etc., except R- and D-packets identification and replacing. We discuss the PFS efficiency for forcing priority in the overloaded conditions. Moreover simple system analysis is also presented. Finally, the profits of using PFS scheme are illustrated by examples corresponding to FTP (TCP controlled traffic) and VoIP (UDP streaming traffic) applications.

Keywords—IP-based network, better than best effort service, priority forcing scheme.

1. Introduction

At present, the Internet users who want to get faster transfer of their data have no additional mechanisms for doing it, even if it could be associated with an additional charging. This is due to the best effort service, the only one supported by current IP-based networks. As a consequence, e.g., a file transfer protocol (FTP) user has to accept long upload/download file time when the network is overloaded. On the other hand, several attractive Internet applications are available now, like voice over IP (VoIP), netmeeting, etc., but they are rather rarely used by a user, since they require better service than this offered by best effort. More specifically, lower packet delay and lower packet losses are needed to satisfy the user. As a consequence, usefulness of these applications is limited, e.g., can be used during the time when Internet is under-loaded.

One may observe two main areas of activities for introducing QoS into Internet. The first direction is aimed at providing some QoS guarantees, similarly as it was done for ATM. In this spirit, the IP QoS network concept is investigated, which can be based on an enhancement of DiffServ [3, 4] or IntServ [5] architecture. However, this requires implementation of new QoS mechanisms at both the packet (e.g., conditioning, scheduling) as well as the network level (e.g., admission control, bandwidth broker). The example of new IP QoS architecture, based on DiffServ, is, e.g., the AQUILA concept [1, 2]. The second investigated direction is to assure for selected flows better than best effort service. The simplest approach for doing it is the implementation of priority queuing (PQ) scheduling mechanism [10] in IP routers. However, this mechanism offers much better service for high priority traffic, but may cause significant service degradation of lower priority traffic during time the router is in congestion. Another commonly used scheduling mechanism is weighted fair queuing (WFQ) [7, 10], which gives a possibility for a number of flows to get access to the link capacity proportionally to the a priori assigned weights. Other investigated way for achieving better than best effort service is to implement additional traffic control mechanisms at the application level. An example is some audio and video applications with quality adaptation mechanisms used to deal with end-to-end loss and delay variation [11]. Another proposal, named alternative best effort (ABE), involving both application and network layer, is described in [9].

The paper addresses to the strategy, named priority forcing scheme, introduced in [6]. The PFS is a proposal for achieving better than best effort service in the IP network, as it is defined, e.g., in [3]. The PFS mechanism can support an application to force prioritised packet service in IP best effort network. It assumes that the application, called PFS application, sends to the network a volume of additional traffic for the purpose of making the reservations for the data traffic in the overloaded router queues along the packet path in the network. The emitted redundant packets, named R-packets, should be rather of small size comparing to the data packets, named D-packets. According to PFS, the R-packets waiting in a queue can be replaced by the arriving D-packets belonging to the same flow. In this way, the D-packets could experience a prioritised service comparing to the packets produced by a non-PFS application. An advantage of the proposed solution is that any QoS mechanisms are implemented in the network, like scheduler, dropping, marking, etc., except R- and D-packets identification and replacing. As it was shown in [6], by using PFS a relative priority level can be reached. This paper includes recent results concerning PFS, and discusses the PFS efficiency for forcing priority in the overloaded conditions, as well as presents simple system analysis. Moreover, the profits of using PFS scheme are illustrated by considering examples corresponding to FTP (TCP controlled traffic) and VoIP (UDP streaming traffic) applications.

The rest of the paper is organised as follows. Section 2 gives short overview of PFS scheme. Section 3 presents simple system analysis. The capability of PFS for reducing packet waiting times in the case of overload conditions are discussed in Section 4. Profit from using PFS for getting better service by VoIP and FTP applications is illustrated in Section 5. Finally, Section 6 summarises the paper.

2. Overview of PFS mechanism

The PFS mechanism is designed to forcing prioritised service by a user, who wants to get better service in best effort network. It assumes that the user application besides the data packets, say D-packets, may also generate in a control way some additional packets, say R-packets, as depicted in Fig. 1. The R-packets are only generated for making the potential reservations for D-packets in the overloaded router queues. To minimise this redundant traffic in the network, which is extremely required to reduce additional load (and charging), the size of R-packets should be set as small as possible, i.e., 40 bytes for TCP and 28 bytes for UDP.

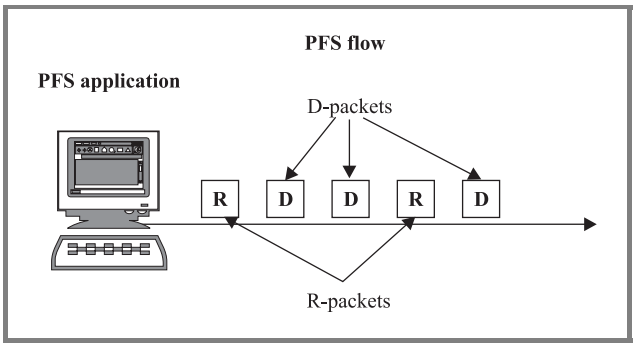


Fig. 1. Packet stream generated by PFS application (PFS flows): data, D-packets, and reservation packets, R-packets.

Since the considered network is with the only single class service, all packets in the router are served according to the FIFO discipline if no additional mechanisms exist. However in the PFS, the D- and R-packets are treated in different way (Fig. 2). For the R-packets the best effort service is assumed with a possibility of dropping them from the queue when a new D-packet arrives. For this D-packet the system is searching for the R-packet waiting in the queue (and belonging to the same PFS flow), which is the first from the top. If no R-packets exist, the D-packet is served according to the FIFO. If at least one R-packet is in the queue, the D-packet drops the R-packet and sizes its position. As a consequence, the D-packets are entitled to get better than best effort service when R-packets exist in the queue. Remark that D-packets are lost only if no R-packets exist in the queue and queue is full. One can expect that D-packets may get greater profit from PFS when more R-packets are generated to the network. Remark also that in the case of non-overloaded queue the service of D-packets is without any delay, as well as the R-packets are not dropped and are transmitted to the next router according to the routing rules. Then, the R-packet can be replaced by a D-packet only in the overloaded routers. Finally, the PFS can be effective in the situations when a bottleneck could occur at any router along the path.

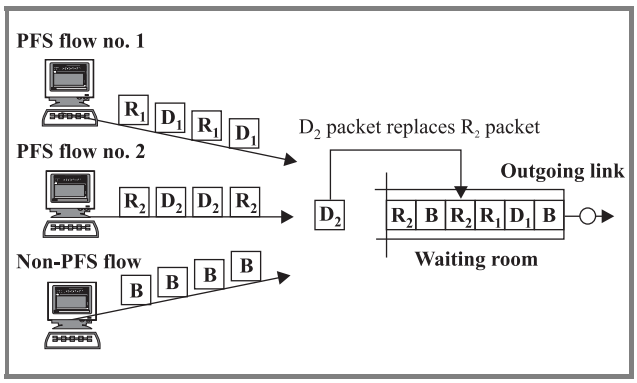


Fig. 2. Queue management for PFS mechanism: example illustrating rules for replacing R-packets in the queue by arriving D-packet.

Notice however, that implementation of PFS mechanism requires the following: (1) from application—a possibility for sending additional packets in a control way and dropping these packets (if any) at the ending-point, (2) from routers—the mechanism for distinguishing between D- and R-packets, and capabilities for replacing R- by D-packets.

3. Simple system analysis

In this section we present simple analysis of system using PFS scheme. Let us assume that the system (Fig. 3) is a single server with infinite waiting room and is fed by three types of flows, which are:

- Flow no. 1, which represents the D-packet flow emitted by a single PFS application. It is assumed as Poissonian stream with the rate λ_D and service times described by the negative exponential distribution with parameter μ_D.
- Flow no. 2, which represents the R-packet flow emitted by the PFS application generating flow no. 1. The R-packets are emitted periodically, at each T_R interval. Furthermore, let us assume that the load of this flow is negligible (R-packet size is close to 0). As it was shown in [6], sending R-packets with constant rate is the simplest and effective way for getting a profit from PFS.
- Flow no. 3, which represents the cumulative flow emitted by other sources (supported by PFS and non-supported by PFS). All B-packets are served by the system in best effort way. We assume that B-packets arrive accordingly to Poissonian low with the rate λ_B and service times described by negative exponential distribution with parameter μ_B.

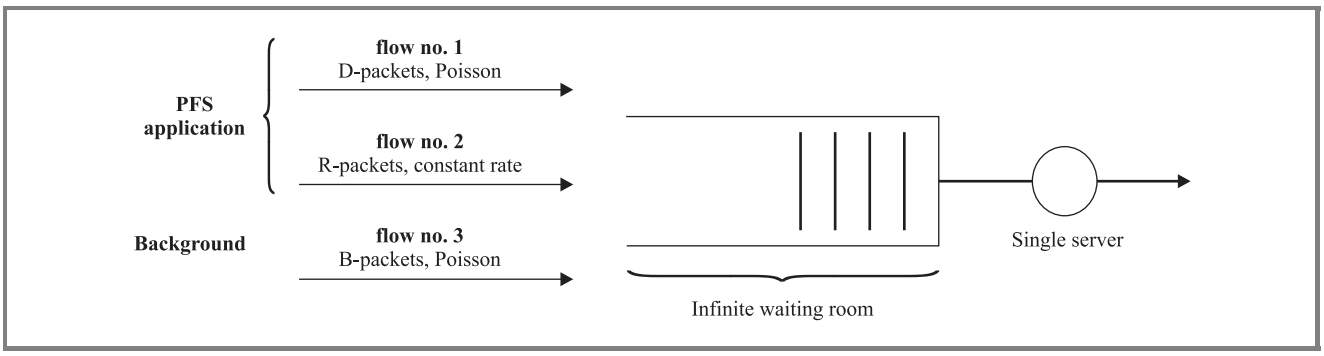


Fig. 3. Single server queue with infinite waiting room fed by PFS and non-PFS traffic.

Assuming that $\mu_D = \mu_B = \mu$, the considered system is similar to the M/M/1 queue, with the only difference that now arriving D-packet may size the R-packet in the queue (if any), and in this way get better service.

Now, we use the expression from M/M/1 system analysis, determining distribution of the packet waiting times $W_q(T)$ (e.g., [8]), which is:

$$W_q(T) = \Pr(t \le T) = 1 - \rho \cdot e^{-\mu(1-\rho)T}, \qquad (1)$$

where $\rho = (\lambda_D + \lambda_B)/\mu$ (remind that service times of R-packets are equal to 0).

Taking Eq. (1) and knowing that R-packets enter system at each T_R interval, we deduce the following approximate formula for probability that at the moment of D-packet arrival it "sees" n $(n = 0, 1, \ldots,)$ R-packets in the queue, assuming that R-packets are not replaced by D-packets:

$$\Pr{}_D^R(n) = \\ = \begin{cases} W_q(0.5\,T_R) & \text{for } n = 0 \\ W_q\left(T_R(n+0.5)\right) - W_q\left(T_R(n-0.5)\right) & \text{for } n = 1, 2, \ldots \end{cases}. \qquad (2)$$

Consider that a D-packet is entering the system at time t_0. Assuming that in this moment there are n $(n = 0, 1, 2, \ldots)$ R-packets in the queue, we deduce that the first from these R-packets arrived to the system at time $t_0 - \Delta t$, where $\Delta t = (0.5\,T_R + (n-1)T_R)$. However, during the interval Δt a number of D-packets could arrive to the system and replace R-packets. Probability that k $(k = 0, 1, 2 \ldots)$ D-packets arrived to the system during the interval Δt is done by:

$$\Pr(k, \Delta t) = \frac{(\lambda_D \Delta t)^k}{k!} e^{-\lambda_D \Delta t}. \qquad (3)$$

From Eqs. (2) and (3), we deduce approximate formula for average number of R-packets (not-replaced by D-packets) in the queue at the moment a D-packet enters the system, say N_R, which is:

$$N_R = \sum_n \sum_{i=0}^{i=n} i \Pr{}_D^R(n) \cdot \Pr\left(n - i, 0.5\,T_R + (n-1)T_R\right). \qquad (4)$$

Remark that Eq. (4) evaluates a lower bound of average number of R-packets in the queue. It can be explained in this way that in Eq. (4) we assumed that all D-packets in the queue have replaced R-packets. In fact, it is not truth since during T_R interval more than one D-packet may enter the system.

Finally, we introduce a measure allowing us to evaluate the profit coefficient (p_f) we could get from PFS. Remark that for the system without PFS, which is modelled in this case by M/M/1 system with FIFO discipline, the $p_f = 0$. The definition of the profit coefficient is as follows:

$$p_f = \lambda_B N_R T_R / \mu_B. \qquad (5)$$

Other interesting measure, illustrating the profit we could get from PFS, is the probability that a D-packet will replace R-packet, say p_s. The p_s denotes the percentage of D-packets handled in better than best effort way and it could be evaluated by:

$$p_s = 1 - W_q(0.5\,T_R). \qquad (6)$$

4. Priority forcing scheme capability in the case of overloaded queue

In this section we show effectiveness of PFS scheme for forcing priority in the queue overloaded conditions. For this purpose we consider the system from Fig. 3. We expect, that according to definition (5), the profit an application can get from using PFS is greater when number of waiting packets is growing. The ideal PFS behavior will be if we are able to provide constant waiting times for D-packets, independently of volume of submitted background traffic. Anyway, one can expect that by increasing generating rate of R-packets the effectiveness of PFS is also increased.

In Fig. 4 are presented the results showing effectiveness of PFS for forcing priority as a function of number of D- and B-packets being in the queue (L_q), at the moment a D-packet arrives, assuming that $\mu_D = \mu_B = \mu = 1$, $\lambda_D = 0.1$, $\lambda_B = 0.85$. Notice, that when $L_q = 0$, any priority forcing mechanism is needed. The Fig. 4a corresponds to the case when distance between consecutive arriving R-packets, T_R, is equal to the mean interarrival time of D-packets, $1/\lambda_D$, while Fig. 4b corresponds to the case when $T_R = 1/(2 \cdot \lambda_D)$. Four characteristics are

presented: *1*—mean number of R-packets being in the queue, *2*—reduced D-packet waiting times (number of waiting B-packets the D-packet "jumps over") thanks to PFS scheme, *3*—experienced mean D-packet waiting time using PFS, and *4*—mean D-packet waiting time for the system without PFS.

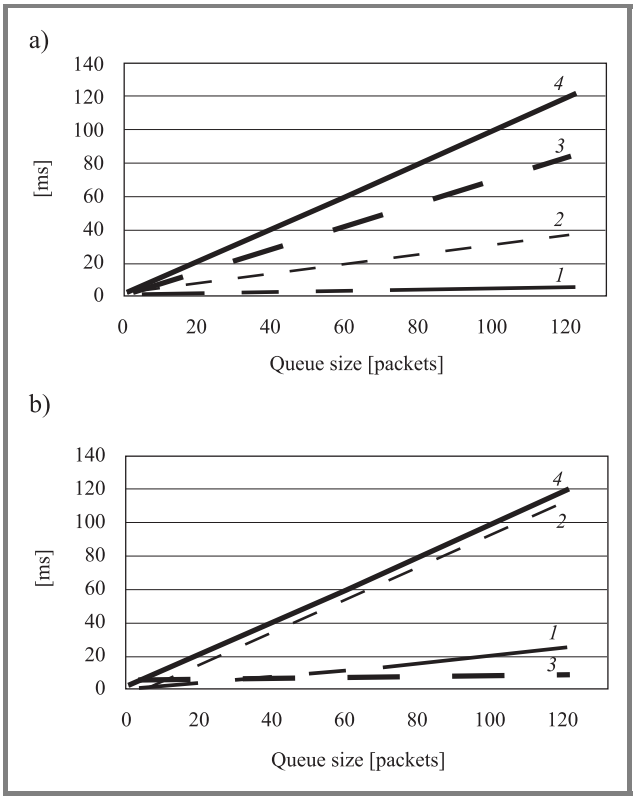


Fig. 4. Results showing effectiveness of PFS for forcing priority as a function of number of D- and B-packets being in the queue, at the moment a D-packet arrives, assuming that $\mu = \mu_D = \mu_B = 1$, $\lambda_D = 0.1$, $\lambda_B = 0.85$: (a) $T_R = 1/\lambda_D$; (b) $T_R = 1/(2 \cdot \lambda_D)$. Explanations: *1*—mean number of R-packets in the queue seen by D-packet; *2*—PFS: reduced D-packet waiting times; *3*—PFS: mean D-packet waiting times; *4*—non-PFS: mean D-packet waiting times.

The obtained results show that by applying PFS scheme one may get essential improvement of packet delay transfer characteristics comparing to the system without PFS. The observation is that the profit gained by using PFS increases when the number of waiting packets is growing. This profit depends on the rate the R-packets are generated. Notice, that in this way we may shape the waiting times for D-packets. In the presented experiment (Fig. 4b), the waiting times for D-packets are almost constant and low, independently on the temporary queue size. In this case number of generated R-packets is double (in the average sense) comparing to emitted D-packets. This result is very promising. It appears that by appropriate setting of PFS mechanism parameters we are able to get excellent packet transfer characteristic, as, e.g., desirable by VoIP application.

5. Applying PFS to VoIP and FTP

In this section we present the simulation results showing efficiency of using PFS mechanisms to improve delay packet transfer characteristics in the case of VoIP and FTP applications. As VoIP is typical for applications emitting streaming packet flows, the FTP is for file transfer and belongs to elastic applications with TCP-controlled packet sending rate depending on network conditions.

5.1. VoIP application

Now, we show the usefulness of using PFS mechanism for getting better quality by VoIP application. The tested VoIP is sending traffic with constant bit rate equals to 64 kbit/s and fixed packet size of 100 bytes. This traffic is submitted to the network with 3 routers, as depicted in Fig. 5. The inter-router links, N1↔N2 and N2↔N3 are of 2 Mbit/s each, the capacity of access links to the routers is 10 Mbit/s. The buffer size at the output router port is fixed to 40 packets.

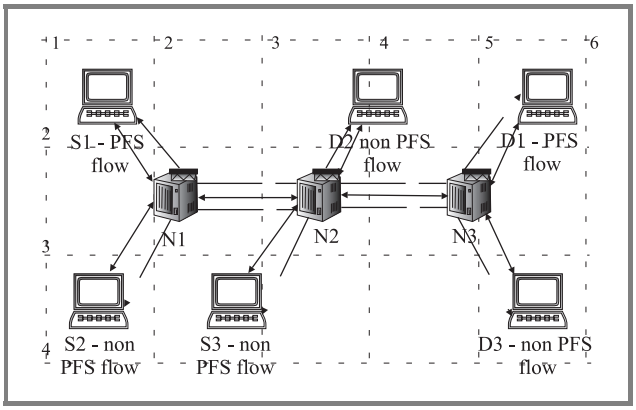


Fig. 5. Network topology for testing VoIP.

The foreground connection for VoIP is established between S1-D1 end-users and passes the routers N1, N2 and N3. The background traffic, of Poissonian type, is produced by non-PFS applications and is carried between S2-D2 and S3-D3. For this traffic the size of the packets is also constant and equals to 750 bytes. We consider three cases depending on traffic conditions in the tested network, which are:

> *Case 1.* The links N1↔N2 and N2↔N3 are both on the heavy load conditions ($\rho = 0.95$).
>
> *Case 2.* The link N1↔N2 is under heavy load conditions ($\rho = 0.95$), while the link N2↔N3 is overloaded ($\rho = 1.1$).
>
> *Case 3.* The link N1↔N2 is overloaded ($\rho = 1.1$), while the link N2↔N3 is under heavy load conditions ($\rho = 0.95$).

Let us recall that for transferring voice on acceptable level, the requirements are: for one-way delay—not more

Table 1
End-to-end B- and D-packet transfer characteristics versus R-packet flow rate (V_R)

Generation rate	Node N1/N2	PFS flow S1-D1		Non-PFS flow S2-D2		Non-PFS flow S3-D3	
V_R [kbit/s]	p_s	$D_m/D_{\max}$ [ms]	p_{loss}	D_m [ms]	p_{loss}	D_m [ms]	p_{loss}
Case 1							
0	–	54.2/206.6	$2\cdot10^{-3}$	32.4	$1.7\cdot10^{-3}$	30.5	$1.1\cdot10^{-3}$
18	0.8/0.14	48.6/204.9	$9.8\cdot10^{-4}$	32.6	$1.9\cdot10^{-3}$	30.8	$1.2\cdot10^{-3}$
36	0.88/0.62	21.1/123.2	$4\cdot10^{-5}$	34.2	$4.2\cdot10^{-3}$	31.6	$1.8\cdot10^{-3}$
54	0.92/0.77	14.7/87.6	0	34.7	$8.2\cdot10^{-3}$	32.1	$4\cdot10^{-3}$
Case 2							
0	–	116.7/219.5	$7.3\cdot10^{-2}$	32.4	$1.7\cdot10^{-3}$	92.9	$9.1\cdot10^{-2}$
18	0.8/0.2	110.1/217	$6\cdot10^{-2}$	32.6	$1.9\cdot10^{-3}$	92.5	$9.2\cdot10^{-2}$
36	0.88/0.91	51.9/136.1	$7.7\cdot10^{-3}$	34.2	$4.2\cdot10^{-3}$	89.6	$9.4\cdot10^{-2}$
54	0.92/1	18.2/123.6	$3.6\cdot10^{-5}$	34.7	$8.2\cdot10^{-3}$	79.2	10^{-1}
Case 3							
0	$-/-$	114.5/216.6	$6.7\cdot10^{-2}$	92.9	$9.3\cdot10^{-2}$	30.1	$1.2\cdot10^{-3}$
18	$0.95/10^{-3}$	107.1/216.6	$1.9\cdot10^{-2}$	91.7	$9.4\cdot10^{-2}$	30.5	$1.3\cdot10^{-3}$
36	1/0.51	25.8/139.3	$1.9\cdot10^{-4}$	80.1	$1\cdot10^{-1}$	31.8	$1.7\cdot10^{-3}$
54	1/0.71	16.1/99.5	0	70.5	$1.2\cdot10^{-1}$	32.3	$3.2\cdot10^{-2}$

p_s—the probability that D-packet replaces R-packet in the queue, p_{loss}—probability that packet is lost, D_m—mean packet transfer delay, $D_{\max}$—maximum packet transfer delay.

Table 2
End-to-end B- and D-packet transfer quality versus R-packet rate (V_R)

Generation rate	Node N1/N2	PFS flow S1-D1	Non-PFS flow S2-D2	Non-PFS flow S3-D3
V_R [kbit/s]	p_s	T [s]/G [kbit/s]	D_m [ms]/p_{loss}	D_m [ms]/p_{loss}
0	$-/-$	230.0/347.8	$56.0/8.3\cdot10^{-3}$	$54.3/6.1\cdot10^{-3}$
13	0.52/0.06	213.9/374.0	$62.5/1.6\cdot10^{-2}$	$60.3/1.1\cdot10^{-2}$
26	0.91/0.14	185.2/432.0	$80.0/3.6\cdot10^{-2}$	$82.0/2.8\cdot10^{-2}$
52	0.98/0.52	170.8/468.4	$75.3/9.2\cdot10^{-2}$	$84.2/7.1\cdot10^{-2}$

p_s—the probability that D-packet replaces R-packet in the queue, p_{loss}—probability that packet is lost, D_m—mean packet transfer delay, T—file upload time, G—TCP goodput.

than 150 ms, for packet loss ratio—less than 10^{-4}. Table 1 shows the received results, corresponding to the Cases 1, 2 and 3, illustrating the quality experienced by VoIP packets supported by PFS and without PFS, versus generation rate (constant) of R-packets (V_R). Packet size for R-packets was fixed to 28 bytes. Notice that by adding R-packets, we increase the total system load.

The presented results show that quality of VoIP application in the cases, when the IP network is under heavy load conditions ($\rho = 0.95$) is non acceptable. The packet loss rate is greater than 10^{-3} while maximum packet delay is greater than 200 ms. Anyway, by using PFS we may get acceptable quality, even if the network is overloaded ($\rho = 1.1$). Obviously, this requires greater R-packet emitting rate, which is almost 36 kbit/s (Table 1).

5.2. FTP application

In this section we show the usefulness of using PFS mechanism for getting better quality in the case of FTP application. The FTP is sending traffic using TCP protocol. This traffic is submitted, as in Section 5.1., to the tested network with 3 routers, as depicted in Fig. 6. The rates of inter-router and access links as well as the buffers of output router ports are the same as in Section 5.1.

The tested FTP connection supported by PFS mechanism is established between S1-D1 and passes the routers N1, N2 and N3. FTP client uses this connection to upload 10 Mbit file on FTP server. S2-D2 and S3-D3 constitute background traffic, each generated according to Poissonian law with the mean rate 1.5 Mbit/s and constant packet size 750 bytes, for

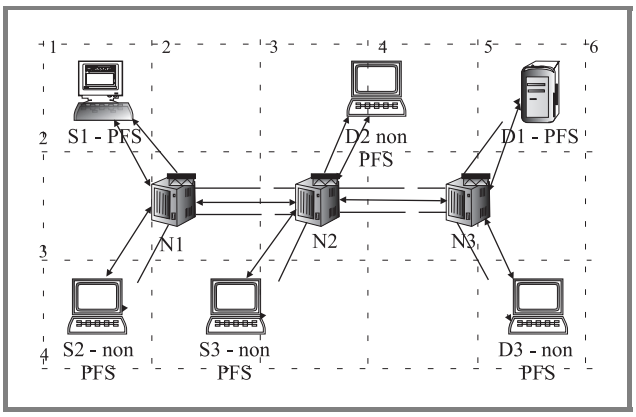


Fig. 6. Network topology for testing FTP application.

getting independent load conditions on the links N1↔N2 and N2↔N3.

Table 2 shows the received values of TCP upload file time/goodput characteristics in the case of FTP user and packet transfer characteristics (mean packet delay and packet loss rate) for Poissonian back-ground traffic, as a function of R-packet rate (V_R). R-packets are of 40 bytes each.

The presented results show, that by using PFS mechanism we can improve file upload time for FTP. Again, by increasing R-packets rate the TCP goodput is also increasing.

6. Conclusions

In the paper we presented the recent obtained results corresponding to efficiency of the PFS mechanism. Comparing to the [6], simple analysis of system with PFS was introduced and the results illustrating possibility of shaping packet delay characteristics for PFS flows were shown. Furthermore, we examined VoIP and FTP application, using PFS for improving end-to-end quality. It appeared that in both considered cases we can obtain satisfactory quality, even if the network is overloaded. Further studies are focused on detailed system analysis and implementations issues.

References

[1] A. Bąk, W. Burakowski, F. Ricciato, S. Salsano, and H. Tarasiuk, "Traffic handling in AQUILA QoS IP network", in *Proc. Quality of Future Internet Services, QoFIS 2001, Lecture Notes in Computer Science*. Springer, 2001, vol. 2156.

[2] A. Bąk *et al.*, "AQUILA network architecture: first trial experiments", *J. Telecommun. Inform. Technol.*, no. 2, 2002, p. 3–13.

[3] Y. Bernet *et al.*, "Differentiated services", Internet Draft, Feb. 1999, draft-ietf-diffserv-framework-0.2.txt

[4] S. Blake *et al.*, "An architecture for differentiated services", Internet RFC 2475, Dec. 1998.

[5] R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet architecture: an overview", RFC 1633, June 1994.

[6] W. Burakowski and M. Fudała, "Priority forcing scheme: a new strategy for getting better than best effort service in IP-based network", in *Internet Technol., Appl. Soc. Imp.* Kluwer, 2002.

[7] *Broadband Network Teletraffic. Performance Evaluation and Design of Broadband Multiservice Networks. Final Report of Action COST 242*, J. Roberts, U. Mocci, and J. Virtamo, Eds., *Lecture Notes in Computer Science*. Heidelberg: Springer, 1996, vol. 1155.

[8] D. Gross and C. Harris, *Fundamentals of Queuing Theory*, 3 rd. ed., Series in Probability and Statistics. Boston: Wiley, 1998.

[9] P. Hurley, J. Y. Le Boudec, P. Thiran, and M. Kara, "ABE: providing a low-delay service within best effort", *IEEE Network*, no. 3, pp. 60–69, 2001.

[10] S. Keshav, *An Engineering Approach to Computer Networking.* Addison-Wesley, 1997, Chap. 9.

[11] D. Wu *et al.*, "Streaming video over the Internet: approaches and directions", *IEEE Trans. Circ. Syst. Video Technol.*, March 2001.

Monika Fudała was born in Poland in 1976. She received her M.Sc. degree in telecommunications from Warsaw University of Technology in 2000. She is now a Ph.D. student. Her research interest focus mainly on traffic handling mechanisms in QoS IP networks.

e-mail: mkrol@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Wojciech Burakowski was born in Poland in 1951. He received his M.Sc., Ph.D. and D.Sc. degrees in telecommunications from Warsaw University of Technology, in 1975, 1982 and 1992, respectively. He is now a Professor in Warsaw University of Technology, Institute of Telecommunications. Since 1990, he has been involved in the European projects COST 224, COST 242, COST 257, COST 279 and AQUILA. His research interests include performance evaluation of ATM and IP networks.

e-mail: wojtek@tele.pw.edu.pl
Institute of Telecommunications
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

Methods for evaluation packet delay distribution of flows using Expedited Forwarding PHB

Sylwester Kaczmarek and Marcin Narloch

Abstract—The paper regards problem of providing statistical performance guarantees for real-time flows using Expedited Forwarding Per Hop Behavior (EF PHB) in IP Differentiated Services networks. Statistical approach to EF flows performance guarantees, based on calculation of probability that end-to-end packet delay is larger than certain value, allows larger network utilization than previously proposed deterministic approach. In the paper different methods of packet delay distribution evaluation are presented and compared. Considered cases comprise evaluation of delay distribution models for the core network and evaluation of end-to-end packet delay in the network consisted of edge node and chain of core nodes. Results obtained with aid of analytical models are compared with simulation results.

Keywords—packet delay distribution, Expedited Forwarding PHB, Differentiated Services, Service Level Specification, IP QoS.

1. Introduction

Rapidly increasing tendencies to provide services typical for traditional telecommunication networks in Internet rise new challenges for realizing services with guaranteed Quality of Service (QoS) in IP based networks. Particular field of interest is a problem of providing real-time services for streaming flows using Expedited Forwarding Per Hop Behavior (EF PHB) [6, 7, 17] in Differentiated Services (DiffServ) network [2, 24].

The paper is based on results of research effort presented in series of conference publications [18, 22, 23]. We present framework to evaluate statistical performance guarantees for flows using EF PHB and compare different methods to calculate the probability that packet delay is larger than certain value. We considered scenario with packet delay only in the network core nodes and scenario with edge node and core of the network (end-to-end delay including packet waiting in edge router). Among evaluated methods are Gaussian approximation, methods based on Large Deviation approach and approximation based on Erlang-n distribution. Despite discrepancies between presented methods, all provide possibility to evaluate statistical guarantees for packet delays of flows using EF PHB.

The paper is organized as follows. Section 2 presents previous work in the subject. In Section 3 model and methods for evaluation of packet delay distribution in the core are described. Section 4 regards influence of low priority traffic on EF packet delays in the node and presents appropriate numerical model. In Section 5 influence of packet waiting in edge node on end-to-end delay is presented together with respective analytical model. Section 6 presents configuration parameters of analysed and simulated networks together with obtained results. Section 7 concludes the paper.

2. Related work

There exist two distinct approaches to analysis of QoS for flows using EF PHB. First approach, derived from context of Integrated Services architecture [28] and represented by [1, 5], is based on deterministic bounds on performance guarantees with worst case assumptions for end-to-end delays. However, analysis in [5] led to very pessimistic bound on utilization for network with flow aggregation. The bound is order of $1/(n-1)$, where n is number of nodes the observed flow pass through. Such a small value indicates that deterministic approach cannot be applied in practice. The second approach, represented by [3], relies on statistical performance guarantees for flows using EF PHB. It allows larger level of utilization at the cost that DiffServ network assures certain packet loss ratio and guarantees that probability of packet transfer delay exceeding certain value (considered as a maximum) is smaller than certain level. That methodology is analogous to description of QoS for ATM CBR service. Also approach based on statistical performance guarantees appears in proposed standards of Service Level Specification (SLS) for DiffServ [14, 27]. An overview of the most current advances in Internet quality of service, including deterministic and statistical guarantees, can be found in [11] (see also references therein). Among other the most recent attempts to explore statistical performance guarantees is [29] where Large Deviations Theorems were applied to results obtained by the use of network calculus. However, we would like to point out that above result is limited to single node case. Thus suggested in [29] end-to-end delay calculation, which was obtained by summing bounds evaluated for single nodes in isolation, still seems to be conservative approach. In the paper we follow alternative approach to statistical guarantees based on the Better than Poisson—Negligible Jitter (NJ) conjecture, presented in [3]. That is the extension of the Negligible Cell Delay Variation notion, presented for ATM network in [4], to the case of IP environment with variable packet

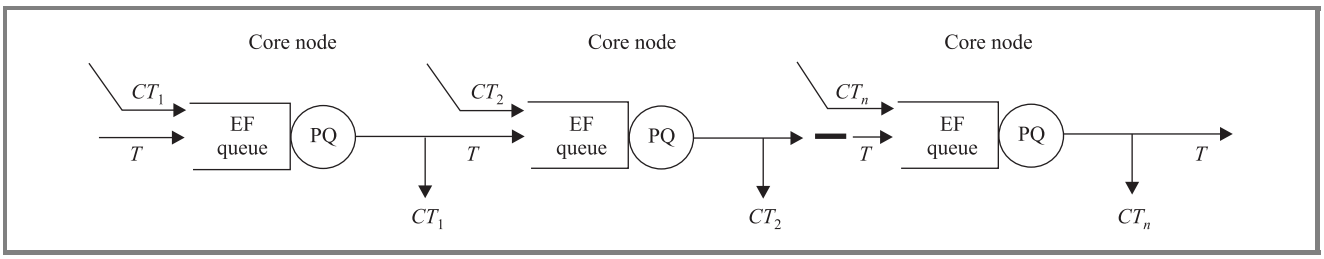


Fig. 1. Analysed network of tandem queues—core nodes case.

lengths. That approach allows radical simplification of the traffic management function, because worst case traffic inside a network can be modelled as Poisson stream of MTU size packets. Moreover, that approach is consistent with formulation of packet delay performance guarantees in the SLS specification, and allows realization of the real-time services with larger network utilization than methods based on worst case, deterministic bound on the delay. Thus we focus on statistical approaches and assume that NJ conjecture is valid.

3. Delay distribution in the core

In this section we present methods for analytical evaluation of delay experienced by packets from CBR flow in the presence of cross traffic in the core network. We considered a network (similar to presented in [3]) consisted of n FIFO queues arranged in tandem serving observed (tagged) CBR stream T passing through all queues, and interfering Poisson MTU-sized cross EF-traffic CT_i passing ith queue (Fig. 1). That type of cross traffic was chosen in order to obtain upper bound of delay distribution. We assumed that offered load ρ_T of observed traffic is relatively small in comparison with cross traffic offered load ρ_{CT}. We also assumed independence of queues in particular nodes. It should be noted that independence conjecture is reported in [3] as conservative, however it allows simplification of delay evaluation. For example, we do not have to consider the effects of distribution of queues with particular load within a chain. That assumptions are valid in the case of the core network (DiffServ network). We split packet delay into two components: deterministic service time equal to $n \cdot \tau_T$ (τ_T is observed CBR flow packet service time), and stochastic waiting time in queues modelled as a chain of n $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queues. From practical point of view, we are interested in probability that waiting delay W exceeds certain value D of delay bound. Consequently probability that end-to-end delay exceeds value $D + n \cdot \tau$ (maximum packet delay stated in SLS) can be easily obtained. Thus we can write quality of service requirement as:

$$P(W > D) \leq L, \tag{1}$$

where L is a small number, i.e., $L \in \langle 10^{-2}, 10^{-6} \rangle$ [19]. In case of analysed network, core packet delay can be written as a sum of independent random variables W_i, denoting packet waiting time in ith queue of the core network:

$$W_{core} = \sum_{i=1}^{n} W_i . \tag{2}$$

If we assume that n is large, we can apply limit theorems. First approach is based on Gaussian approximation of delay distribution. It is simple extension of the model presented in [15], in the context of CBR service in ATM to the case of variable length IP packets. In that method, Gaussian distribution of packet delays has mean:

$$\mu = \sum_{i=1}^{n} \mu_i \tag{3}$$

and variance:

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 , \tag{4}$$

where μ_i and σ_i^2 are respectively mean and variance of waiting time in ith $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queue. In next two approaches based on Theory of Large Deviations [8] we explore the fact that delay values for the probabilities of interest are largely deviated from the mean delay. Packet waiting distribution in the core network can be expressed with aid of approximation based on Chernoff theorem:

$$\log P(W_{core} \geq x) \leq -F(\theta^*), \tag{5}$$

or refinement of the Chernoff-Cramer approximation based on Bahadur-Rao theorem (local limit theorem):

$$P(W_{core} \geq x) \approx \frac{e^{-F(\theta^*)}}{\sqrt{2\pi} \cdot \theta^* \cdot \sigma(\theta^*)}, \tag{6}$$

where large deviations rate function $F(\theta^*)$ is defined as:

$$F(\theta^*) = \sup_{\theta \geq 0} F(\theta),\ F(\theta) = \theta \cdot x - \sum_{i=1}^{n} \log M_i(\theta), \tag{7}$$

and $\sigma^2(\theta)$ is second order derivate of large deviations rate function with respect to θ:

$$\sigma^2(\theta) = \sum_{i=1}^{n} \frac{M_i''(\theta) \cdot M_i(\theta) - \left(M_i'(\theta)\right)^2}{M_i^2(\theta)}. \tag{8}$$

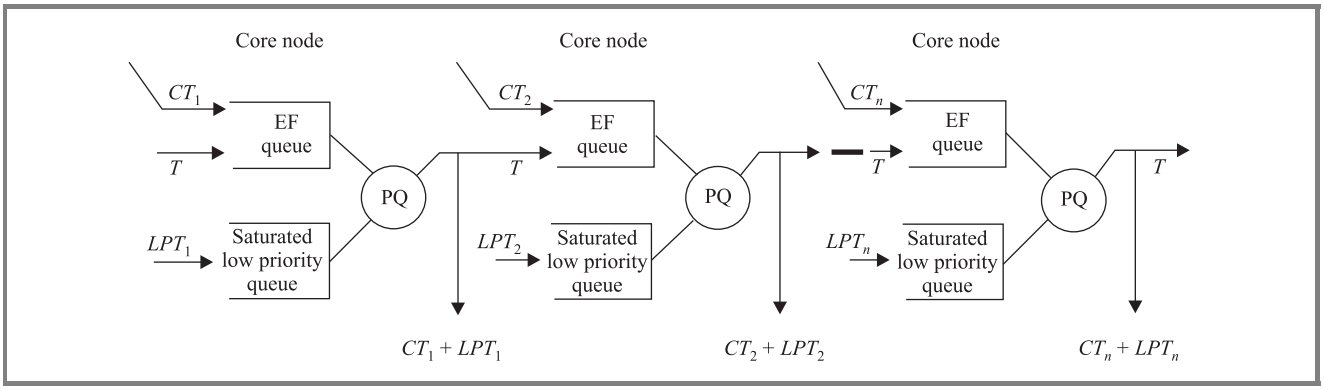


Fig. 2. Network of tandem queues with vacations—core nodes with low priority traffic.

$M_i(\theta)$ denotes moment generating function of packet waiting time in single queue for respective queuing model $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$, etc. In order to compute desired probability we have to find θ^* for which supremum of $F(\theta)$ is attained, taking into consideration moment generating function $M(\theta)$ of packet waiting time for particular queuing model. Thus θ^* is positive root of the equation with derivative of $F(\theta^*)$:

$$F'(\theta^*) = 0, \text{ where } F'(\theta^*) = x - \sum_{j=1}^{J} n_j \frac{M_j'(\theta)}{M_j(\theta)}. \quad (9)$$

It is worth noting that in general case independent random variables W_i, denoting packet waiting time in ith queue, are not necessarily identically distributed. Last of evaluated methods of packet waiting distribution in the core is based on Erlang-n distribution, and for the sake of presentation clarity will be described in the next section.

4. Influence of low priority traffic

In order to model the influence of the lower priority, non-EF traffic on EF streams performance guarantees, we extended model of core nodes and considered non-preemptive static priority queues in core nodes (Fig. 2), with multiple and exhaustive vacations, with constant vacation time equal to MTU packet transmission time [9]. In that model arrivals and services have the same characteristics as in ordinary $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queue, but in queue with vacations when high priority (EF) queue is empty, server takes vacation instead being idle waiting for EF packet to arrive for service. If queue server finds EF packets when returning from vacation, it serves them until EF queue becomes empty (exhaustive discipline), and than it takes next vacation. If there is no packet in high priority queue after returning from vacation, server takes another vacation (multiple vacation discipline). That allows modelling the real system with priority queuing and link transmitting non-EF, low priority packets, wherever there is no EF packets to transmit. This is also the worst case approach with respect to EF stream performance guarantees, because we assumed that link is saturated and there is always MTU sized low priority packet in node to send. In case of delay distribution approximation methods based on Large Deviations Theory, that extension of network node model results in application of appropriate moment generating function, regarding queuing model with vacation. Moment generating function $M(\theta)$ for $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queue with vacation can be obtained by stochastic decomposition property described in [12, 13]. Stochastic decomposition property allows to consider the waiting time in the M/GI/1 queue with vacations, as the sum of two independent components: one distributed as the waiting time in the ordinary queue in the corresponding M/GI/1 queue without vacations, and the other as the equilibrium residual time of a vacation. Thus moment generating function $M(\theta)$ for $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queue with vacation can be calculated as follows:

$$M(\theta) = \frac{U(\theta) - 1}{u\theta} M_{M/D/1}(\theta), \quad (10)$$

where $M_{M/D/1}(\theta)$ is moment generating function in ordinary $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queue and $U(\theta)$ denotes moment generating function for the vacation time. In the considered case of constant vacation time equal to MTU packet transmission time $U(\theta) = \exp(\theta \cdot u)$, where $u = MTU/C$, C is link bandwidth. In case of presented in previous section Gaussian approximation of packet delay distribution in the core, appropriate formulas for μ_i and σ_i^2 can be obtained with the aid of respective derivatives of moment generating functions $M(\theta)$ of waiting time for queues with vacations. Last of the described packet waiting distribution approximation [3] can be expressed as a sum of $n \cdot x_{\min}$ and Erlang-n distribution of mean $(MTU \cdot n)/(r \cdot C)$, where r satisfies:

$$\rho_{CT} \cdot (e^r - 1) - r = 0, \quad (11)$$

and $x_{\min}$ is defined as:

$$x_{\min} = \frac{-MTU}{r \cdot C} \cdot \log \frac{1}{K}, \quad (12)$$

where

$$K = \frac{1 - \rho_{CT}}{\rho_{CT}^2 \cdot e^r - \rho_{CT}}. \quad (13)$$

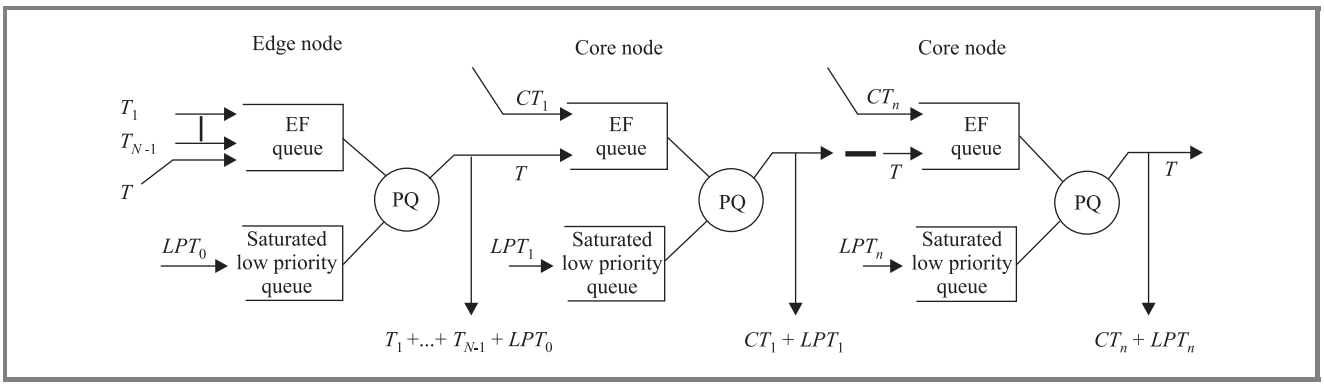


Fig. 3. Analysed network of tandem with vacations queues—edge and core nodes case.

That approach is based on presented in [25], in the context of ATM network approximation of queue size distribution:

$$P(Q > x) \approx K \cdot \exp(-r \cdot x), \tag{14}$$

with appropriate extensions to model influence of low priority traffic ($x_{\min}$ and different formula describing k). It is worth noting that Erlang-n approximation is limited to the homogenous case only and unfortunately that approach cannot be used directly to evaluate packet delay distribution for the heterogeneous case, which is typical for any practical network scenario.

5. Influence of waiting in edge node

In order to consider influence of queuing in edge node on EF packet delay, we extended evaluated network model. Model of the network core remained as previously described chain of $M/D_{MTU}/1$ queues with vacations, however we introduced edge node modelled as discrete time ND/D/1 queue with vacations. N denotes the number of CBR sources in the edge router. Each source generates packets with fixed length and constant, deterministic period P. Packet arrival instants from all N sources are independent and randomly spread with uniform distribution within the period P (assumption of random phases of the sources). We considered discrete time queuing model with time axis divided into slots. Slot duration is equal to CBR source packet transmission time τ_T in the link, between edge node and following core node. In our model one CBR stream plays the role of observed traffic T, which follows path through all nodes of considered network. The remaining $N-1$ CBR streams create background EF traffic in the ingress edge node and leave considered network path after edge node. That streams denoted in Fig. 3 as T_k, where $k \in <1, N-1>$, compete for resources in EF queue with the observed stream T only in the edge router. We also modelled influence of lower priority non-EF traffic on EF streams performance guarantees in non-preemptive priority queue of the edge node, by considering queue with multiple and exhaustive vacations, with constant vacation time equal to CBR packet transmission time. In case of analysed network model, evaluation of end-to-end packet delay distribution can be considered as a form of discrete convolution of delay in discrete ND/D/1 with vacations model, and delay distribution in chain of $M/D_{MTU}/1$ with vacations in the core:

$$P(W_{e2e} > x) = \\ = \sum_{k=1}^{N} P\left(W_{edge} = k\right) \cdot P\left(W_{e2e} > x \middle| W_{edge} = k\right), \tag{15}$$

where $P(W_{edge} = k)$ is probability that waiting delay in the edge router is equal to k slots, $P(W_{e2e} > x | W_{edge} = k)$ is conditional probability that end-to-end packet delay exceeds x conditioned on event that delay in the edge node equals k slots. Probability of packet waiting $P(W_{edge} = k)$ for discrete time ND/D/1 model is presented in [16]:

$$P\left(W_{edge} = k\right) = \frac{P}{N} P\left(Q = k\right), \tag{16}$$

where $k > 0$ and queue length distribution [16, 26] is given by:

$$P(Q > q) = \\ = \sum_{m=1}^{N-q} \frac{P - N + q}{P - m} \binom{N}{q+m} \left(\frac{m}{P}\right)^{q+m} \left(1 - \frac{m}{P}\right)^{N-q-m}, \tag{17}$$

where $q \geq 0$. Conditional probability $P(W_{e2e} > x | W_{edge} = k)$ can be determined regarding the assumption of queue independence in the network. Random variables denoting waiting time in node queues are independent and thus amount of packet delay encountered in the edge node does not influence value of delay in the chain of core nodes (queues). Because of that, we can express conditional probability as:

$$P\left(W_{e2e} > x \middle| W_{edge} = k\right) = P\left(W_{core} > x - k \cdot \tau_T\right), \tag{18}$$

and apply the approximations, describing packet delay distribution in the core of the network $P(W_{core} > x)$, presented in details in previous sections. In order to evaluate end-to-end packet delay distribution, we can also apply other method in which edge router is modelled as M/D/1 queue with vacations, where D denotes CBR source packet size (typically smaller than MTU) and core nodes are modelled as a chain of n $\mathrm{M}/\mathrm{D}_{\mathrm{MTU}}/1$ queues with vacations. In that method we can utilize approximations of delay distribution described in previous sections, but with respective modifications in formulas regarding presence of the edge node in the observed stream path. Hence, in formulas (5) and (6), W_{core} is replaced by W_{e2e} and, consequently, Large Deviations rate function $F(\theta^*)$ includes moment generating function of random variable, describing packet delay in the edge node (queue). Therefore, formula (7) should be rewritten as:

$$F(\theta^*) = \sup_{\theta \geq 0} F(\theta),\ F(\theta) =$$

$$= \theta \cdot x - \sum_{i=0}^{n} \log M_i(\theta), \tag{19}$$

where $i = 0$ regards edge node queue and $M_0(\theta)$ denotes moment generating function of waiting time in M/D/1 queue with vacations. Similarly, in case of formulas (8) and (9), range of index i should be extended to include edge node accordingly.

6. Numerical results

In order to verify accuracy of presented methods, we compared results obtained from theoretical derivations with simulation results. At first, we considered core network of 10 nodes with interconnecting links of 150 Mbit/s bandwidth and buffers for 20 packets from EF streams. We considered Poissonian cross EF-traffic with offered load ρ_{CT} of 0.1, 0.3 and 0.5 (three distinct cases), and MTU packet size equal to 1500 bytes. The observed traffic consisted of 1 CBR flow with rate 1.5 Mbit/s (offered load $\rho_T = 0.01$), and packet size equal to 100 bytes. We evaluated packet delay distribution for two scenarios. In the first scenario, a FIFO queue is dedicated to EF streams, which are the only traffic passing through the nodes. In the second scenario, we considered priority queue with vacations as described in Section 4. Figure 4 presents comparison of theoretical and simulation results for respective cases of offered load ρ_{CT} for both scenarios. We also considered influence of path length (the number of nodes the EF flow passes through). Corresponding results for ρ_{CT} of 0.3 in the network with 5 and 15 nodes are presented in the Fig. 5, respectively. In the figures one can observe that for delay distribution probabilities larger than 0.01, calculation based on Gaussian approximation provide very good results. However, Gaussian approximation provide results unacceptable from practical point of view for probabilities smaller than 0.01, i.e., calculation of delay probabilities becomes too optimistic, and comparing to simulation values packet delay distribution is significantly underestimated. For probability values smaller than 0.01 methods based on Large Deviations provide better calculation, particularly for tail probabilities. Only for delay values close to mean value, methods based on Large Deviations give moderate precision of approximation, overestimating packet delay probabilities. That comes from the fact that Large Deviations Theory is dedicated to describe rare events and tail probabilities. In all cases, method based on Bahadur-Rao (local limit theorem) approximation provides more precise results for the same queuing model than Chernoff-Cramer approximation, which should be considered as a conservative, upper bound of real delay distribution. Considering results obtained for case with different number of nodes the EF flow passes through, methods based on Large Deviations provide good approximation of packet delay distribution, even in case where the number of nodes is relatively small. That promising results was obtained regardless that limit theorems were used in formulation of proposed methods, i.e., demand for very large number of nodes. From practical point of view that feature is positive, particularly that bounds are relatively tight, regarding Bahadur-Rao approximation. However, Bahadur-Rao approximation provides delay distribution values close to simulation results for small probabilities and cannot be applied for probabilities larger than 10^{-2}. Approximation based on Erlang-n distribution is computationally very attractive and provides precise results, which are compared to results obtained by Bahadur-Rao approximation. Despite its simplicity and precision, Erlang-n approximation is limited to the homogenous case, which cannot be assured in practice for any typical network scenario.

In order to evaluate influence of delay in the edge node, we considered extended network with 1 ingress edge node and 5 core nodes. Edge node was connected to the core by 15 Mbit/s link. Links in the core had 150 Mbit/s bandwidth as in previous cases. Edge node served 6 homogenous, 1.5 Mbit/s CBR streams with packet size equal to 100 bytes, thus load of EF traffic in edge node was equal to 0.6 Erl. We considered lower priority, non-EF traffic in the edge node with packets of size 100 bytes. The observed traffic consisted of one CBR flow with rate 1.5 Mbit/s (offered load $\rho_T = 0.01$ in the core node link). In EF-queues in core nodes we considered heterogeneous scenario, regarding to offered load ρ_{CT} of Poissonian cross traffic with MTU-sized (1500 bytes) packets. The value of offered load in jth queue was equal to $0.1 \cdot j$, thus we cover the range of loads from 0.1 to 0.5 with step 0.1. Lower priority (non-EF) packets of size MTU = 1500 bytes filled remaining link capacity in every core node of the network. Simulation results for network with edge node were obtained with simulation method described in details in [18], which allows efficient evaluation of systems with ND/D/1 queues.

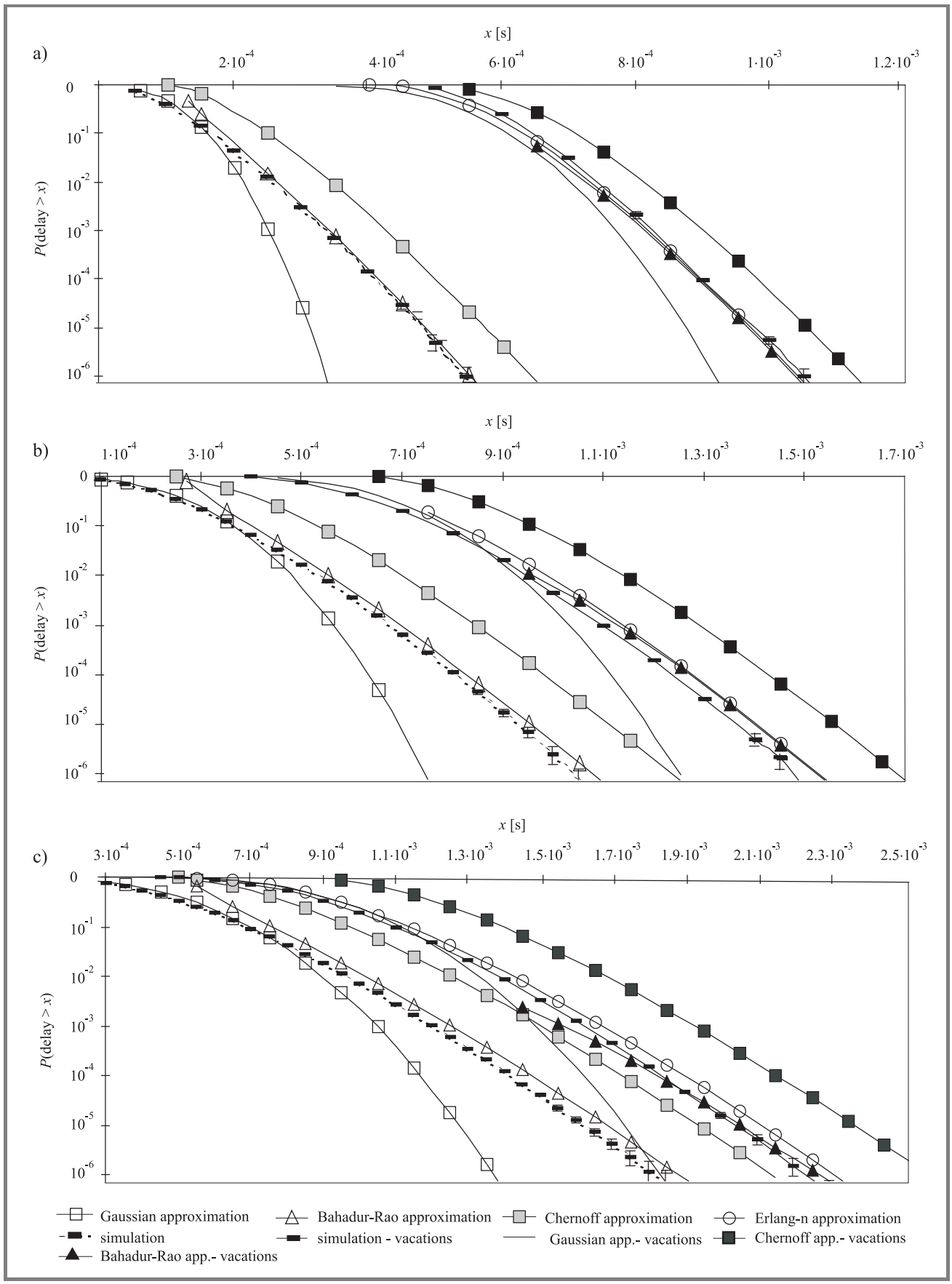


Fig. 4. Packet delay distribution of CBR flow in the core network: (a) $\rho_{CT} = 0.1$, $n = 10$; (b) $\rho_{CT} = 0.3$, $n = 10$; (c) $\rho_{CT} = 0.5$, $n = 10$.

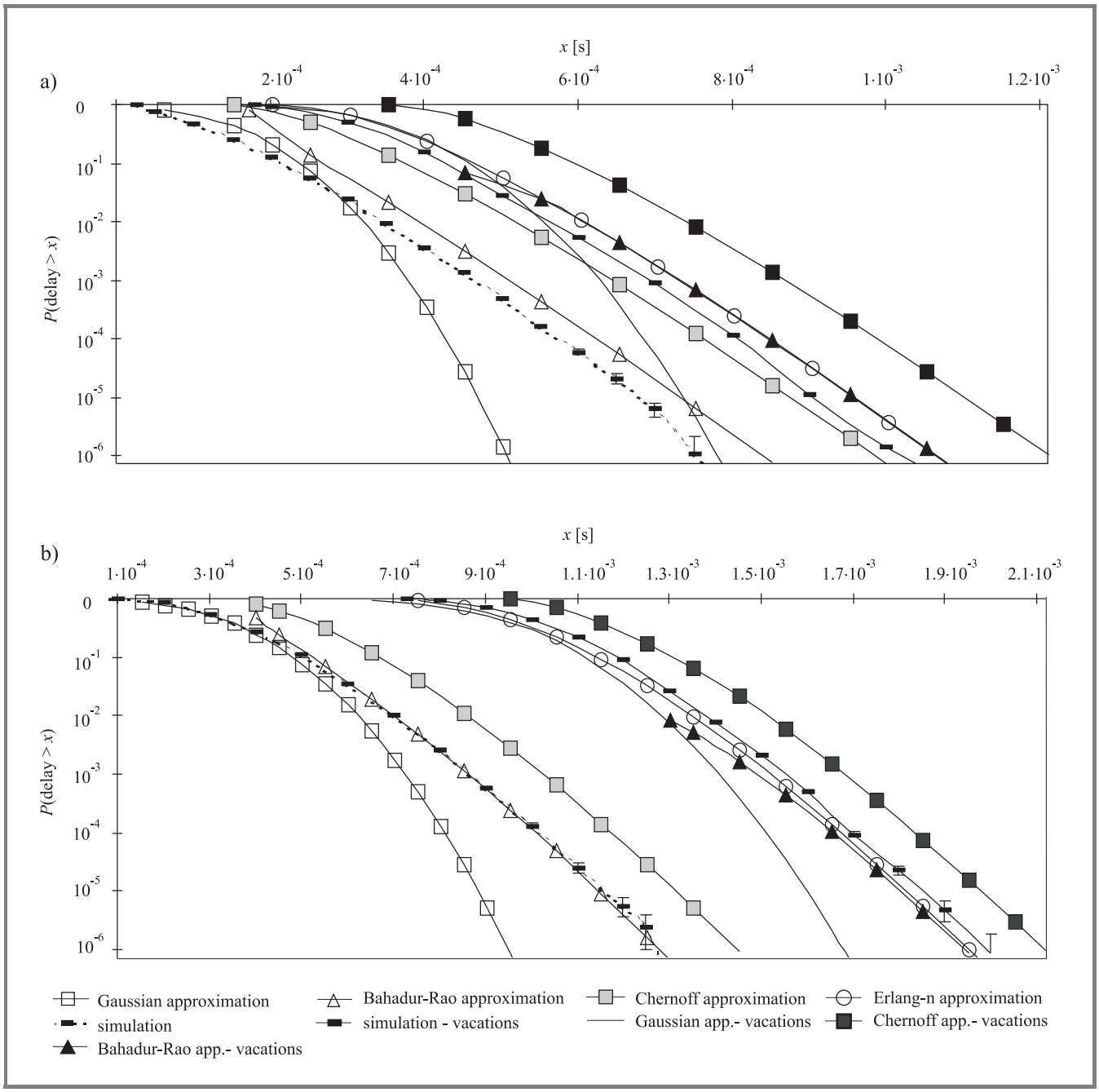


Fig. 5. Packet delay distribution of CBR flow in the core network: (a) $\rho_{CT} = 0.3$, $n = 5$; (b) $\rho_{CT} = 0.3$, $n = 15$.

In the Fig. 6 can be seen that method based on formula (14) and method in which edge router is modelled as M/D/1 queue with vacations and core nodes are modelled as a chain of n $\mathrm{M/D_{MTU}/1}$ queues with vacations provide almost similar results (under the condition of similar approximation application for calculation of delay distributions, for example, based on Bahadur-Rao theorem). However, the second approach to the calculation of end-to-end delay distribution seems to be simpler and more versatile than approximation based on formula (14), because approximation based on discrete time convolution (14) demands large number of calculations for large values of N. Alternatively, calculations with only few, significant values of $P(W_{edge} = i)$ can be applied as a form of formula (14) approximation. Moreover, in the Fig. 6 the accuracies of end-to-end packet delay distribution approximations based on different limit theorems can be compared. We would like to emphasise that precision of packet delay computation strongly depends on values provided by underlying approximation, and thus all remarks describing accuracy of approximation used to calculate packet delay distribution in the core regard calculations for end-to-end packet delay.

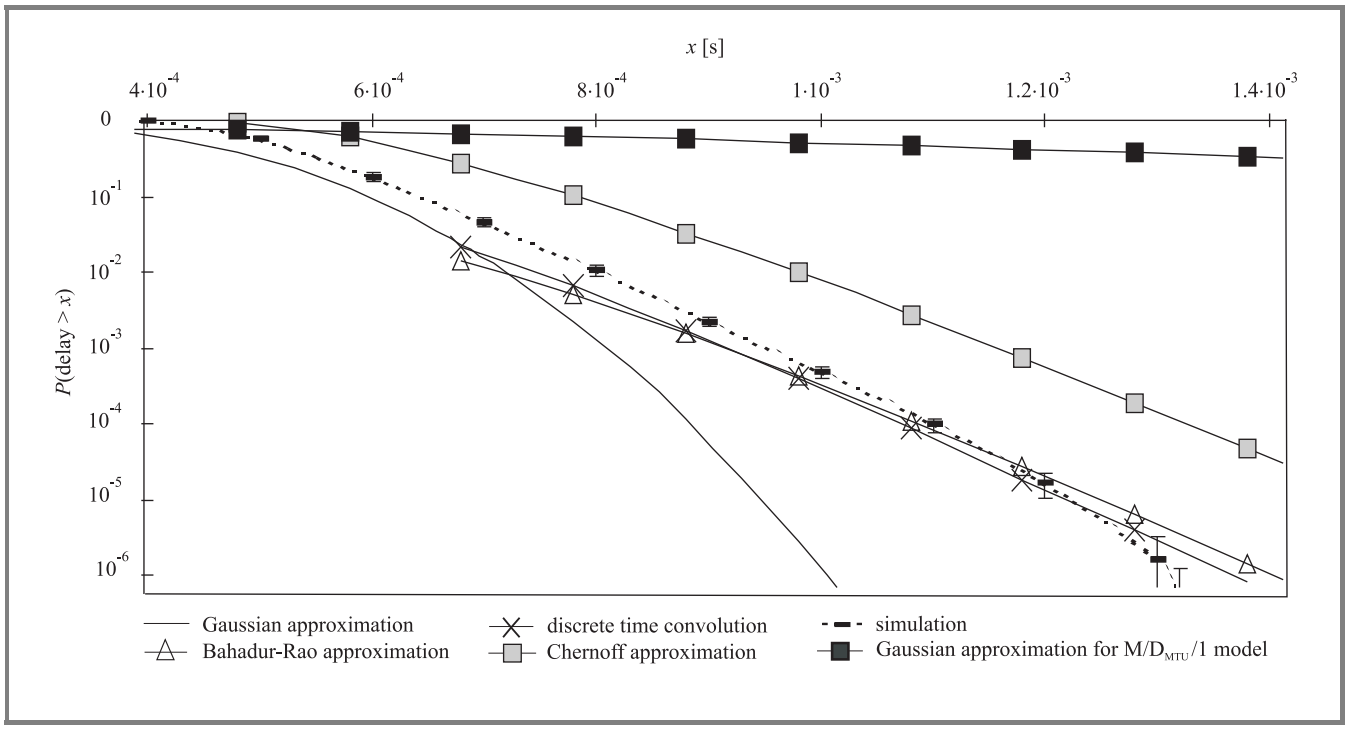


Fig. 6. End-to-end packet delay distribution of CBR flow in the network consisted of edge and core nodes.

7. Conclusions

Statistical performance guarantees allow to increase network utilization with sufficient margin of security, comparing to previously proposed deterministic approaches. Moreover, it is consistent with formulation of packet delay performance guarantees in Service Level Specification for IP QoS Differentiated Services.

Methods based on Large Deviations Theory are the most attractive from presented approaches to evaluate packet delay distribution. They provide bound on delay probabilities of packets from CBR flows, using Expedited Forwarding PHB in the region where exceeding maximum packet delay is allowed with certain, but very small probability. From practical point of view, application of that approximations allows realization of real-time services with statistical guarantees. We also extended core network to include edge node modelled by ND/D/1 queue, and applied it in evaluation of packet delay distribution. Obtained analytical and simulation results indicate that developed model allows evaluation of end-to-end packet delay distribution with accuracy which is satisfactory from practical point of view.

Knowledge of the packet delay distribution is very important in network dimensioning, network planning and traffic control algorithms. Methods presented above are used in calculation of *effective delay* [21] proposed as a new metrics for traffic control functions, for example a new EF flow admission control algorithm utilizing that notion. Consequently, precision of delay distribution calculation strongly influences accuracy of any traffic control function which relies on such approximations.

Future work in the subject should be directed toward extending model of edge router and considering more general low priority packet distribution. Also extension of presented simulation framework in order to evaluate more general EF and non-EF packet size distributions is intended.

References

[1] J. C. R. Bennett, K. Benson, A. Charny, W. F. Courtney, and J.-Y. Le Boudec, "Delay jitter bounds and packet scale rate guarantee for expedited forwarding", in *Proc. Infocom*, Anchorage, USA, 2001.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services", RFC 2475, Dec. 1998.

[3] T. Bonald, A. Proutiere, and J. W. Roberts, "Statistical performance guarantees for streaming flows using expedited forwarding", in *Proc. Infocom*, Anchorage, USA, 2001.

[4] F. Brichet, L. Massoulie, and J. W. Roberts, "Stochastic ordering and the notion of negligible CDV", in *Teletraffic Contributions for the Information Age. Proceedings of the 15th ITC*, V. Ramaswami and P. E. Wirth, Eds. Amsterdam [etc.]: Elsevier, 1997, pp. 1433–1444.

[5] A. Charny and J.-Y. Le Boudec, "Delay bounds in a network with aggregate scheduling", in *Quality of Future Internet Services. Proceedings of First COST 263 International Workshop, QofIS2000*, J. Crowcroft, J. W. Roberts, and M. Smirnov, Eds., *Lecture Notes in Computer Science*. Berlin [etc.]: Springer-Verlag, 2000, vol. 1922, pp. 1–13.

[6] A. Charny *et al.*, "Supplemental information for the new definition of the EF PHB (expedited forwarding per-hop behavior)", RFC 3247, 2002.

[7] B. Davie *et al.*, "An expedited forwarding PHB (per-hop behavior)", RFC 3246, Apr. 2002.

[8] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, London: Jones & Bartlet Publ., 1992.

[9] B. T. Doshi, "Queuing systems with vacations – a survey", *Queuing Syst. Theory Appl.*. vol. 1, pp. 29–66, 1986.

[10] "The ns Manual", K. Fall and K. Varadhan, Eds., Dec. 2003, http://www.isi.edu/nsnam/ns/ns-documentation.html

[11] V. Firoiu, J.-Y. Le Boudec, D. Towsley, and Z.-L. Zhang, "Theories and models for Internet quality of service", *Proc. IEEE*, vol. 90, no. 9, pp. 1565–1591, 2002.

[12] S. W. Fuhrmann, "A note on the M/G/1 queue with server vacation", *Oper. Res.*, vol. 32, pp. 1368–1373, 1984.

[13] S. W. Fuhrmann and R. B. Cooper, "Stochastic decompositions in a M/G/1 queue with generalized vacation", *Oper. Res.*, vol. 33, no. 5, pp. 1117–1129, 1985.

[14] D. Goderis *et al.*, "Service level specification semantics, parameters and negotiation requirements", Internet-Draft, work in progress, June 2001, draft-tequila-sls-01.txt

[15] M. Grossglauser and S. Keshav, "On CBR service" (extended version), in *Proc. Infocom*, San Francisco, USA, 1996.

[16] P. Humblet, A. Bhargava, and M. Hluchyj, "Ballot theorem applied to the transient analysis of nD/D/1 queues", *IEEE/ACM Trans. Netw.*, vol. 1, no. 1, 1993.

[17] V. Jacobson, K. Nichols, and K. Poduri, "An expedited forwarding PHB", RFC 2598, June 1999.

[18] S. Kaczmarek and M. Narloch, "End-to-end packet delay distribution of flows using EF PHB", in *Proc. 10th Polish Teletraffic Symp.*, Kraków, Poland, 2003.

[19] M. Karam and F. Tobagi, "Analysis of the delay and jitter of voice traffic over the Internet", in *Proc. Infocom*, Anchorage, USA, 2001.

[20] M. Mandjes, K. van der Wal, R. Kooij, and H. Bastiaansen, "End-to-end delay analysis for interactive services on a large-scale IP network", in *Proc. 7th IFIP Worksh. Perform. Model. Eval. ATM/IP Netw.*, Antwerp, Netherlands, 1999.

[21] M. Narloch and S. Kaczmarek, "Admission control method based on effective delay for flows using EF PHB", in *Architectures for Quality of Service in the Internet*, W. Burakowski, B. Koch, and A. Bęben, Eds., *Lecture Notes in Computer Science*. Springer-Verlag, 2003, vol. 2698.

[22] M. Narloch and S. Kaczmarek, "Methods for evaluation packet delay distribution of flows using expedited forwarding PHB", in *Proc. 2nd Polish-German Teletraffic Symp. PGTS*, Gdańsk, Poland, 2002, pp. 85–94.

[23] M. Narloch and S. Kaczmarek, "Quality of service problem in IP based network", in *Proc. First Inform. Technol. Conf.*, Gdańsk, Poland, 2003, vol. 1, pp. 301–314 (in Polish).

[24] K. Nichols *et al.*, "A two-bit differentiated services architecture for the Internet", RFC 2638, July 1999.

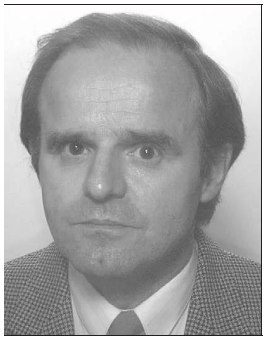
[25] *Broadband Network Teletraffic. Performance Evaluation and Design of Broadband Multiservice Networks. Final Report of COST 242*, J. W. Roberts, U. Mocci, and J. Virtamo, Eds. Heidelberg: Springer-Verlag, 1996.

[26] J. W. Roberts and J. T. Virtamo, "The superposition of periodic cell arrival streams in an ATM multiplexer", *IEEE Trans. Commun.*, vol. 39, no. 2, pp. 298–303, 1991.

[27] S. Salsano *et al.*, "Definition and usage of SLSs in the AQUILA consortium", Internet Draft, work in progress, November 2000, draft-salsano-aquila-sls-00.txt

[28] S. Shenker *et al.*, "Specification of guaranteed quality of service", RFC 2212, Sept. 1997.

[29] M. Vojnovic and J.-Y. Le Boudec, "Stochastic analysis of some expedited forwarding networks", in *Proc. Infocom*, New York, USA, 2002.



Sylwester Kaczmarek received his M.S./B.S. in electronics engineering, Ph.D. and D.Sc. in switching and teletraffic science from the Technical University of Gdańsk, Gdańsk, in 1972, 1981 and 1994, respectively. His research interests include: IP QoS and GMPLS networks, switching, routing, teletraffic and quality of service. He has published more than 125 papers.
e-mail: kasyl@eti.pg.gda.pl
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Marcin Narloch was born in Poland in 1974. He received his M.Sc. degree in telecommunications from Gdańsk University of Technology in 1998. Since 1998 he has kept assistant position at Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics. His research activities focus on traffic control in IP QoS and ATM networks.
e-mail: narloch@eti.pg.gda.pl
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Paper

Application of a hash function to discourage MAC-layer misbehaviour in wireless LANs

Jerzy Konorski and Maciej Kurant

Abstract—Contention-based MAC protocols for wireless ad hoc LANs rely on random deferment of packet transmissions to avoid collisions. By selfishly modifying the probabilities of deferments greedy stations can grab more bandwidth than regular stations that apply standard-prescribed probabilities. To discourage such misbehaviour we propose a protocol called RT-hash whereby the winner of a contention is determined using a public hash function of the channel feedback. RT-hash is effective in a full hearability topology, assuming that improper timing of control frames is detectable and that greedy stations do not resort to malicious actions. Simulation experiments show that RT-hash protects regular stations' bandwidth share against various sophisticated greedy strategies of deferment selection; as such it may contribute to MAC-layer network security.

Keywords—wireless LAN, MAC protocol, noncooperative setting.

1. Introduction

Ad hoc wireless packet networks offer the possibility of cheap on-demand interconnection of a set of stations (mobile user terminals) in an environment where any fixed communication infrastructure would turn out either physically or economically infeasible [1]. As the underlying technology matures, ad hoc networks expand from their traditional niche of disaster management and military systems to public local- and wide-area data communications and become an attractive alternative to costly wireline networks. This expansion, however, brings new design challenges that are critical to ad hoc networks' long-term survival. Namely, adherence to the deployed standard communication protocols can no longer be counted on for several reasons. Firstly, if the stations of an ad hoc network are not subjected to a common authority then there are no administrative facilities like log-in, traffic monitoring, service accessibility, conformance testing etc., which implies that punishment for a station's misbehaviour can only be enforced by other stations in a distributed fashion. Secondly, ad hoc networks guarantee a certain degree of anonymity—any station may disappear at any time (switch off, move out of range or switch identity) and reappear later pretending to be another station; therefore it need not fear any punishment that does not materialise instantly. For example, ill reputation based on cumulative statistics of past packet transmissions [2] would not make a serious disincentive to stations willing to misbehave. The emergent noncooperative design paradigm is now establishing itself as a part of wireless network security planning [3].

In this paper we focus on MAC-layer selfish misbehaviour whereby a station may depart from the standard rules of contention for the wireless channel so as to grab a larger-than-fair share of the available bandwidth. Such misbehaviour is indeed possible, contrary to the popular opinion that adherence to standard MAC protocols is only natural if the stations want to stay "synchronised" [4]. Consider the class of distributed MAC contention protocols, sometimes referred to as random token (RT) [5], that rely on random deferment of packet transmissions for collision avoidance. These include HIPERLAN/1 [6] and CSMA/CA with RTS/CTS exchange [7], later incorporated into the IEEE 802.11 MAC standard [8]. In any instance of contention the winner is the station whose deferment is extreme among the contending stations; this condition is equivalent of capturing a unique token that visits the stations in random order rather than sequentially. By manipulating the probability distribution of transmission deferment a station can easily outperform stations that apply a standard-prescribed probability distribution [9, 10].

The purpose of this paper is to propose a new protocol called RT-hash in order to prevent MAC-layer misbehaviour. Specifically, we propose to determine the winner of a contention using a public hash function of the feedback each station gets from the contention. This is hoped to confuse misbehaving stations in such a way that no modification of the probability distribution of transmission deferment should appear beneficial to them. Given that, they may resort to more sophisticated deferment selection strategies, which we attempt to anticipate and the impact of which we attempt to evaluate.

Note that the feedback from a contention is implicitly assumed to be uniform across all stations; this implies a single-hop wireless LAN (WLAN) setting and perfect channel operation. We believe that, although hidden stations and transmission errors may affect the protocol we propose, the illustrative and qualitative value of the presented results will not be diminished.

The paper is organised as follows. In Section 2 the network model and a framework for MAC-layer misbehaviour are outlined. Section 3 presents the RT-hash protocol against the background of earlier RT-like protocols and examines the requirements for the public hash function, assuming random selection of transmission deferments. In Section 4

some sophisticated selection strategies are discussed; their impact on the RT-hash protocol is evaluated in Section 5. Section 6 concludes the paper.

2. The model

We consider a number of stations interconnected by a single-channel wireless network. The proposed model reflects the general idea of an ad hoc system outlined in Section 1. We begin with the network station model and then present our model of MAC-layer misbehaviour.

2.1. The network station model

As can be expected of an ad hoc system, the station model mostly consists of non-assumptions. Namely we accept that a station need not:

- stay interconnected all the time or maintain a permanent identity; thus in general the number of stations N need not be fixed or known,
- communicate its present identity to any station other than the recipient(s) of its current transmission; thus from the viewpoint of the MAC protocol the stations are anonymous,
- interpret any transmitted data of which it is not an intended uni- or multicast recipient (except for detecting carrier on the channel); thus it can fully encrypt its communications and use any data format it has agreed on along with the current recipient(s).

To remove the dependence on a particular hearability topology, station mobility model, multihop packet forwarding protocol and traffic scenario, as well as the physical characteristics of the wireless channel, we also assume that

- all stations hear each other's transmissions directly, i.e., the network is a single-hop WLAN,
- each station always has a packet ready to send, i.e., the network operates under heavy load conditions, and
- the characteristics of the wireless channel and the attained signal-to-noise ratio ensure error-free transmission between any pair of stations.

Further, to simplify the presentation of the RT protocols, we assume that each station synchronises to a global slotted time axis. A slot allows for a transmission and reception of a MAC protocol's control frame and leaves enough time for each station to decide the type of the slot based on the feedback from sensing the channel. A station distinguishes v- or c-type slots sensed, for "void" or "carrier"; moreover, a recipient of a successful (i.e., non-colliding) transmission recognises an s-type slot, for "success", and reads its contents. This type of binary feedback facilitates collision avoidance and is employed in deferment-based MAC protocols, e.g., in the form of RTS/CTS exchange [7, 8]. We shall prefer the term *pilot/reaction mechanism* instead of RTS/CTS in reference to a generic RT protocol to stress that the format and semantics of the involved control frames may differ from those specified by the IEEE 802.11 standard.

In the pilot/reaction mechanism each station synchronises to the start of a protocol cycle, marked by a v-type slot following a packet transmission. Subsequent slots are classified by all the stations as *contention slots*, in which pilot frames can be transmitted, and *reaction slots*, reserved for reaction frames. The first slot of a protocol cycle is a contention one; any c- or s-type contention slot is followed by a reaction slot and then another contention slot; a v-type contention slot is followed by another contention slot. A station with a packet ready defers for a number of slots (the *transmission deferment*) and transmits a pilot. Further action depends on the channel feedback the station gets in the following reaction slot. The transmission deferment may vary from one protocol cycle to another as dictated by a *selection strategy*. In existing RT protocols this number is drawn from a uniform probability distribution over a range of values; such a strategy will be called *Randomiser*. It should be noted, however, that the selection strategy need not be a part of the protocol; strictly speaking, the protocol only defines the rules of contention such that all the stations can reach a consensus regarding the winner or a no-winner outcome.

2.2. Model of MAC-layer misbehaviour

A taxonomy of ad hoc station misbehaviour presented in [11] suggests distinguishing selfish and malicious misbehaviour; this roughly corresponds to rational vs. irrational motivation for departures from standard protocols. Selfish MAC-layer misbehaviour is rational in that a station may want to grab a larger-than-fair share of the available bandwidth, but will not act just to reduce other stations' throughput without a clear benefit for its own. For example, issuing unnecessary pilots or jamming other stations' transmissions "for the fun of it" is irrational in terms of own throughput and power consumption. In the context of RT protocol, selfish misbehaviour can be twofold: a station can either violate the rules of contention or adopt a selection strategy other than Randomiser. The former type of misbehaviour would have to involve improper timing of pilot/reaction frames; e.g., a station might transmit several pilots in a protocol cycle while the rules of contention allow only one. Such behaviour is detectable by means of a directional antenna and as such could in principle be immediately punished. On the other hand, tampering with a selection strategy is hard to detect and prove by other stations. Long-term statistical analysis of transmitted packets might reveal significant departures from Randomiser; unfortunately such an approach is pointless in view of the assumed station anonymity and packet encryption.

Following the approach in [9, 10], we wish to develop an RT-type MAC protocol for which no rational selection strategy will perform substantially better than Randomiser. The following framework for MAC-layer misbehaviour is assumed:

- all N stations comply with the rules of contention,
- G stations are *greedy*, i.e., willing to grab a larger-than-fair share of the available bandwidth; these are free to use any selection strategy (G need not be fixed or known to any station),
- the other N–G stations are *regular*, i.e., apply Randomiser,
- a greedy station acts in isolation, i.e., cannot coordinate its selection strategy with other greedy stations.

The last assumption may seem controversial. It is reasonable if one presumes that stations are reluctant to reveal their greedy status to others. However, collusion among some greedy stations is not unthinkable [12]. In fact, a worst-case greedy selection strategy we consider in Section 4.3 approaches a collusion scenario.

3. Random token protocols

The RT-hash protocol presented in this section breaks with the rule requiring that the winner's deferment be extreme among the contending stations. To clarify this difference we first briefly summarise some earlier RT protocols [10].

3.1. RT and RT-1s protocols

A station with a packet ready to send selects a deferment between 0 and D–1 contention slots (D is an integer parameter) and when it expires, transmits a pilot frame. The pilot contains the recipient's identity (possibly in an implicit manner, e.g., using the recipient's public or symmetric encryption key) and in addition may contain the first fragment of the packet. Since full encryption is allowed, to all non-recipients the pilot is merely a burst of non-interpretable carrier. If at some station the frame decrypts to a correct pilot, that station has just sensed an s-type slot and recognised itself as a recipient. In such a case it issues in the following reaction slot a reaction frame that is merely a burst of carrier. Having sensed the reaction slot following the pilot as c-type, the sender of the pilot continues to send the remaining part of the packet. A lack of reaction (i.e., a v-type reaction slot) marks a pilot collision and terminates the protocol cycle (Fig. 1). Note that a greedy station can do no harm by jamming a reaction frame. Nor can it benefit from issuing a reaction frame if it is not a recipient or from not issuing one if it is a recipient.

Under RT, greedy stations need only a minor alteration of the standard random selection strategy to monopolise the channel bandwidth. A modified protocol called RT-1s (for "first success") [10] is somewhat more resistant to greedy stations. Under RT-1s, stations whose pilots were not reacted to back off until the next protocol cycle, while the rest are free to transmit their pilots in subsequent contention slots (Fig. 2). Note that the back-off provision implies that the threat of misbehaviour detection using a directional antenna is serious enough to discourage greedy stations from transmitting more than one pilot per protocol cycle. Also note that the start of a protocol cycle is now marked either by the termination of a packet transmission or by D consecutive v-type slots (in this regard D is an analogue of DIFS in IEEE 802.11).

3.2. RT-hash protocol

Random token-hash aims to improve on RT-1s in the presence of greedy stations. The protocol operation is illustrated in Fig. 3. A station with a packet ready to send defers for a number of contention slots between 0 and D–1. Subsequently it transmits a pilot and awaits a reaction. Now each intended recipient transmits a reaction if an s-type slot is sensed, while refraining from reaction if a c-type slot is sensed. The presence or absence of a reaction permits the other stations to deduce that the previous slot was s- or c-type, respectively.

When D contention slots have elapsed, along with the corresponding reaction slots (if any), all stations arrive at an identical *feedback vector* $\mathbf{f} = (f_1 \dots f_D)$ with $f_i = 0, 1$ or 2 if the ith contention slot was v-, s- or c-type, respectively. Denote $S(\mathbf{f}) = \{i \,|\, f_i = 1\}$. All stations then have to agree on a unique $i^* \in S(\mathbf{f})$ designating the winning slot (and consequently the winner station), or on a no-winner outcome if $S(\mathbf{f}) = \varnothing$. In Fig. 3, $\mathbf{f} = (0\,2\,0\,1\,1\,0\,0)$ and $S(\mathbf{f}) = \{4, 5\}$, therefore $i^* = 4$ or 5. This designates station *3* or station *2* as the winner, respectively.

A unique i^* can be agreed on by defining a deterministic hash function H on the set of possible $\mathbf{f}$. The value returned by H will index into the set $S(\mathbf{f})$ written in ascending order. In the above example, $H(\mathbf{f}) = 1$ would indicate $i^* = 4$ and $H(\mathbf{f}) = 2$ would indicate $i^* = 5$. The function H should have suitable mixing properties. In particular, it should compute to uniformly distributed values for randomly chosen $\mathbf{f}$ and prevent greedy stations from easy guessing of the winning slot based on the observed prefix of $\mathbf{f}$. Of the many possible hash functions, two (denoted H_1 and H_2) have been selected for a closer examination. Let $v(\mathbf{f})$ be the numerical value whose ternary representation is $\mathbf{f}$. Then

$$H_1(\mathbf{f}) = 1 + v(\mathbf{f}) \mod |S(\mathbf{f})|$$

$$H_2(\mathbf{f}) = 1 + round\,[\pi \cdot v(\mathbf{f})] \mod |S(\mathbf{f})|, \qquad (1)$$

where $\pi = 3.14159265358979$, $|\cdot|$ denotes cardinality and *round* symbolises rounding to the nearest integer. Extensive simulation was carried out to evaluate the mixing proper-

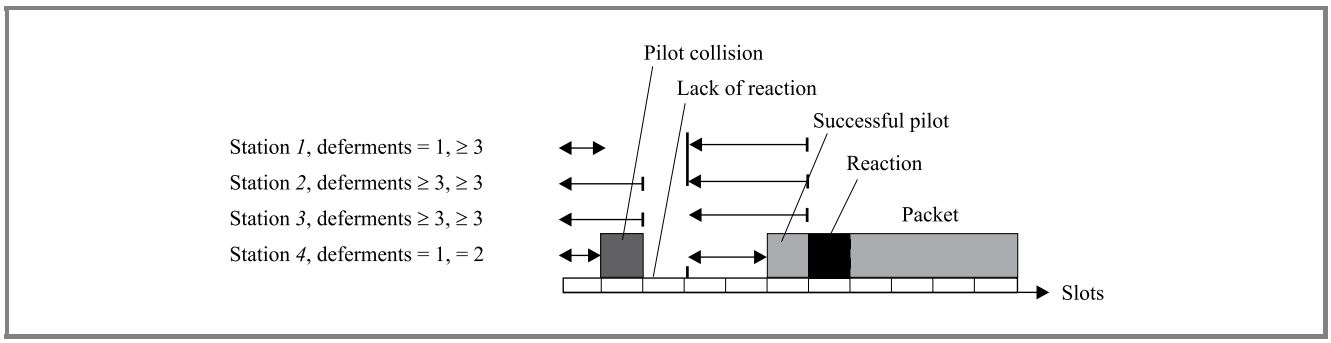


Fig. 1. RT protocol; $N = 4$, $D \geq 4$ (dashed line marks end of protocol cycle).

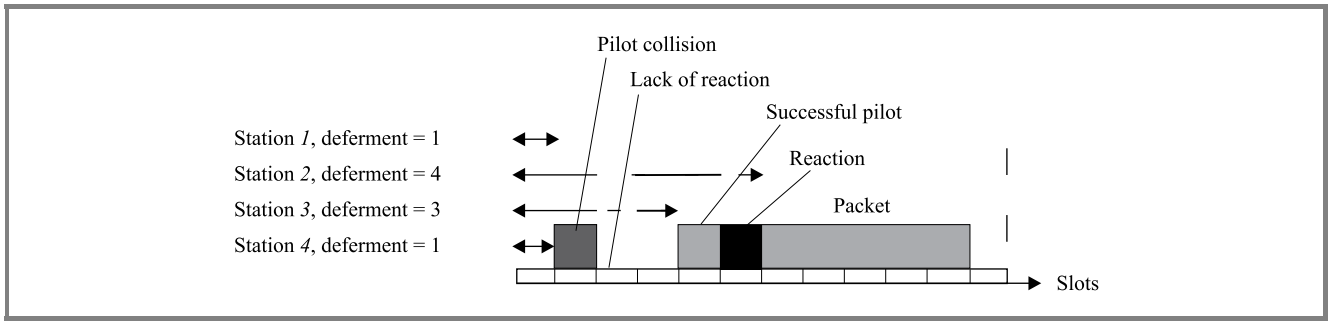


Fig. 2. RT-1s protocol cycle; $N = 4$, $D \geq 4$ (stations *1* and *4* back off, station *3* wins).

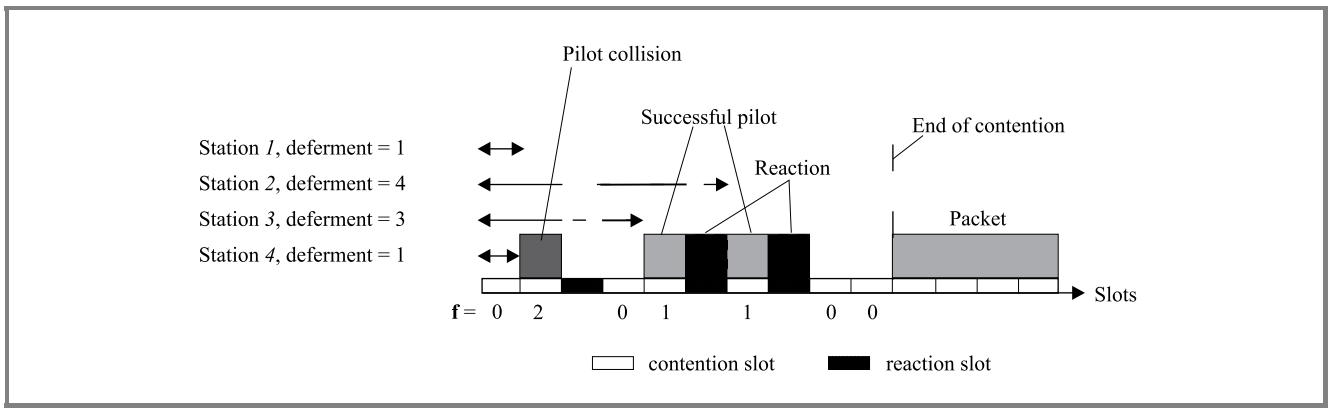


Fig. 3. RT-hash; $N = 4$, $D = 7$.

ties of H_1 and H_2, with each of N stations applying Randomiser. After 1000 instances of contention a histogram *hist* of winning slots was obtained and its normalised entropy was computed:

$$\textit{normalised entropy} = \frac{-\sum_{i=1}^{D} \textit{hist}(i) \cdot \log_2 \textit{hist}(i)}{\log_2 D}. \qquad (2)$$

For an ideal hash function, (2) should be close to unity. This was emulated by replacing a hash function by a random number generator (rand). In Fig. 4, H_1 and H_2 are compared to rand in terms of (2); H_2 was found more satisfactory and was taken for further experiments. The results presented in Fig. 4 were generated for $D = 10$; other simulation experiments show little dependence of (2) on D.

Note that the above results are only valid if a station does not have any additional information about the ongoing contention. This need not be the case for a greedy station, which might choose to defer a pilot transmission until a long enough prefix of **f** has been observed. The longer the observed prefix, the worse are the mixing properties of H_2 for the remaining slots. For example, for $D = 10$ the prefix (2 1 0 1 0 2) yields almost even probabilities of winning for the remaining four slots, while (1 2 0 1 0 2) yields 1%, 45%, 12% and 0%. Figure 5 depicts the normalised entropy for the remaining D–L slots, averaged over all L-long prefixes of **f**. One sees that a greedy station in-

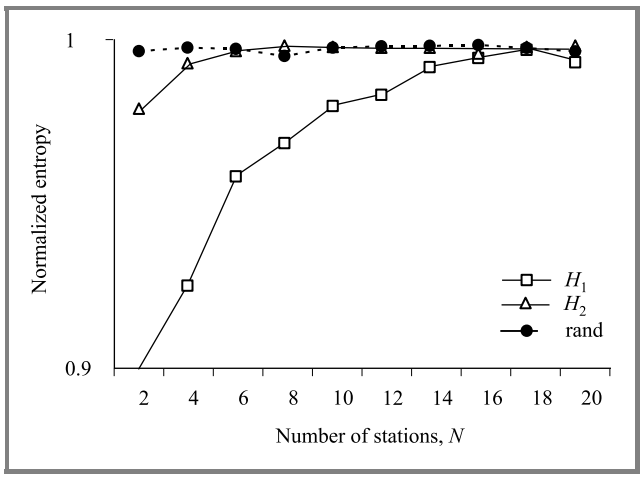


Fig. 4. Mixing properties of hash functions; $D = 10$.

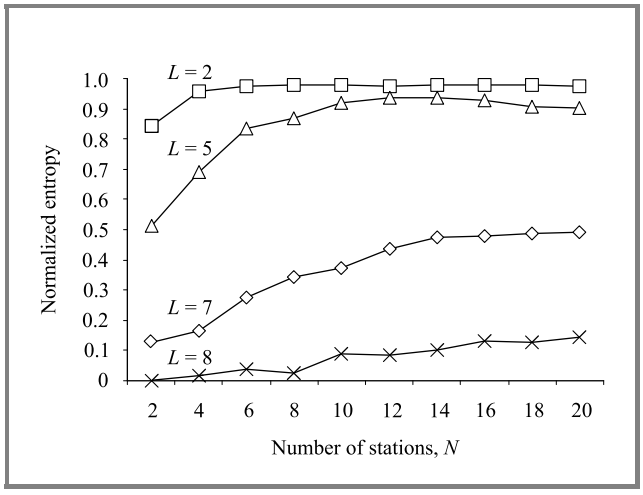


Fig. 5. Mixing properties of H_2 conditioned on L-long prefixes of **f**; $D = 10$ (note the difference in scale compared to Fig. 4).

deed can substantially improve the chances of winning based on the observed prefix. Such an attack is described in Section 4.4.

4. Selection strategies

Greedy stations self-optimise by departing from Randomiser. Because self-optimising strategies do not form a definite set, suitable heuristics should be sought; some are outlined below.

4.1. Aggressive Randomiser

While regular stations use Randomiser, greedy stations may resort to a more "aggressive" probability distribution of deferments, i.e., shifted toward short deferments (Fig. 6). We take it to be a convex quadratic function, namely $\Pr(\text{deferment} = l) = const. \cdot [1 + (l - D + 1)^2]$. The difference between Randomiser and Aggressive Randomiser is

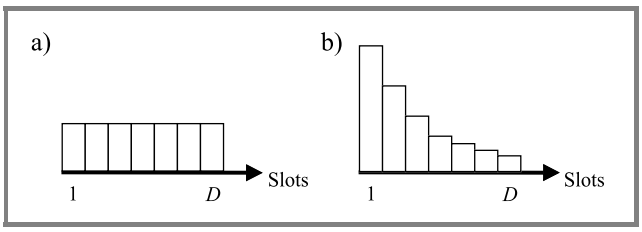


Fig. 6. Probability distributions of Randomiser (a) and Aggressive Randomiser (b).

that the former attempts to optimise the overall bandwidth utilisation, whereas the latter is self-optimising.

4.2. Optimal Randomiser

In the Optimal Randomiser selection strategy, all G greedy stations select transmission deferments at random using a common probability distribution **p** that maximises the bandwidth share they collectively obtain. In numerical experiments, a reinforced random search was applied to determine **p**, starting from a uniform distribution. In each of 10000 steps, 0.05 was tentatively subtracted from one randomly chosen element of **p** and added to another, with the new values only retained if they increased the bandwidth shares of the greedy stations.

Optimal Randomiser violates the isolation constraint since **p** is assumed common to all the greedy stations. It is considered as a worst-case scenario from the regular stations' viewpoint.

4.3. Pseudoperiodic

Suppose the G greedy stations are not subject to the isolation constraint. Firstly, they will arrange to select different deferments to avoid pilot collisions. Secondly, under RT and RT-1s, short deferments will be preferred. For $G = 4$, a strategy in Table 1 may be conceived. Each greedy station defines a deferment sequence for T consecutive protocol cycles and repeats it periodically. In the absence of regular stations this amounts to passing a token from one greedy station to another; the sequence of token holders is also periodic with period $T = 5$. The presence of regular stations may disturb this, but only through pilot collisions in slot 0.

Table 1
Deferments selected by greedy stations in periodic token passing

Greedy	Instants of contention				
	a	*b*	*c*	*d*	e
1	0	3	2	1	3
2	1	0	3	2	1
3	2	1	0	3	0
4	3	2	1	0	2

Under RT or RT-1s, Table 1 exemplifies a Nash equilibrium point [13], i.e., none of the greedy stations can improve its bandwidth share by unilaterally deviating from the specified deferment sequence. The equilibrium is not fair in that the stations are not guaranteed equal bandwidth shares (station *3* will obtain the largest on account of its double 0-slot deferment). Taking only the first four columns and $T = 4$ yields symmetry in the deferment distribution as well as in the obtained bandwidth shares.

The Pseudoperiodic selection strategy aims to reconcile the isolation constraint with the idea of token passing. Each greedy station initially applies a periodic deferment sequence $(s_1 s_2 \ldots s_T)$, $s_i \in \{0, \ldots, D-1\}$, e.g., $s_5 = 3$ means a 3-slot deferment in every 5th protocol cycle. Denote by e_i the currently observed frequency of winning the ith contention. At the end of each period, the e_i's are updated using a low-pass filter:

$$\underset{i\in\{1,\ldots,T\}}{\forall} \; e_i = \alpha \cdot e_i + (1-\alpha)\cdot last_i\,, \tag{3}$$

where $\alpha \in [0,\ 1]$ and $last_i$ equals 1 if the ith contention was won and 0 otherwise. Subsequently, with a fixed probability, the worst-performing selection is modified, i.e., s_{i^*} is replaced by a new randomly chosen value if $e_{i^*} = \min\{e_1, e_2, \ldots, e_T\}$. Thus Pseudoperiodic permanently improves the obtained bandwidth share.

Technically, the isolation constraint is violated again: the greedy stations have to apply identical T (or its multiple) since lack of synchronisation inevitably leads to more frequent collisions. Nevertheless, the choice of $T = D$ seems natural; one can also imagine a version of Pseudoperiodic with optimisation of T.

4.4. Antihash

The Antihash selection strategy (Fig. 7) was devised to exploit the reduction of the normalised entropy (2) conditioned on the observed prefix. Having observed an L-long prefix $\mathbf{f}_L$ of $\mathbf{f}$, a greedy station uses the statistics of recent

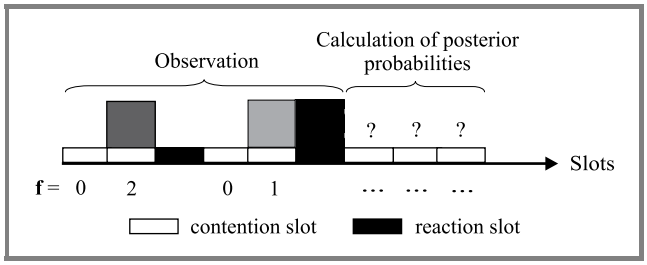


Fig. 7. Antihash; $D = 7$, $L = 4$.

protocol cycles to calculate a posterior probability distribution over possible continuations $\mathbf{f}_{D-L}$ and over the respective winning slots $H_2(\mathbf{f}_L\mathbf{f}_{D-L})$. The most promising slot between the $(L+1)$th and Dth is then determined; if it is not the $(L+1)$th one, the analysis is repeated one slot later with an $(L+1)$-long prefix observed.

5. Performance evaluation

In a series of simulation experiments, regular and greedy stations were contending for access to the medium under heavy load (all stations always had packets ready to transmit). The latter assumption implies concurrent transfer of large files, a perfect scenery for selfish behaviour. In most simulations a fixed number $N = 10$ of stations was assumed, with $D = 10$ and a variable number of greedy stations ($G \in \{0, \ldots, N\}$). Packets were of fixed size $S = 50$ slots.

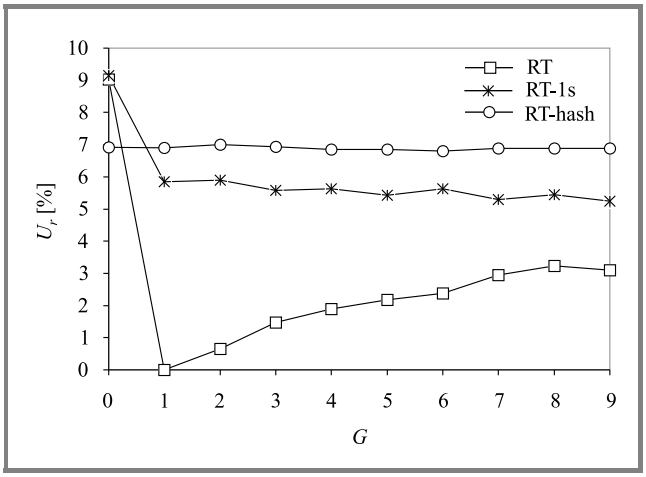


Fig. 8. Optimal Randomiser versus Randomiser.

In Fig. 8, the average regular station's bandwidth share, denoted by U_r, is plotted against G for RT, RT-1s and RT-hash. Regular and greedy stations applied Randomiser and Optimal Randomiser, respectively. $G = 0$ corresponds to an all-cooperative setting; U_r is then slightly lower than $1/N$ or 10% of the channel bandwidth on account of the scheduling penalty; this effect is stronger for RT-hash. In a noncooperative setting ($G > 0$), U_r is even lower under RT and RT-1s. While the former protocol is unacceptable, the latter seems to cope with the presence of greedy stations fairy well, especially when they use Aggressive Randomiser. Unlike RT and RT-1s, RT-hash holds its own regardless of G. Figure 8 can be explained by inspecting the optimal probability distribution $\mathbf{p}$ found by a reinforced random search. They are given in Fig. 9 for $G = 4$. Under RT, $\mathbf{p}$ is strongly biased toward short deferments; under RT-1s this bias is far less visible. Clearly, the more uniform $\mathbf{p}$ the greedy stations arrive at, the less they benefit. In this regard RT-hash is ideal since favouring or discriminating any particular deferment causes more frequent pilot collisions. Thus the selection strategies of regular and greedy stations are identical.

Figure 8 also illustrates a drawback of RT-hash, namely large overhead (as noticed at $G = 0$). This is because each contention lasts until all D contention slots (along with the corresponding reaction slots) have elapsed. For example, $S = 50$ and $D = 10$ result in a 30% overhead; clearly it decreases with S. Additional simulations of RT-hash were

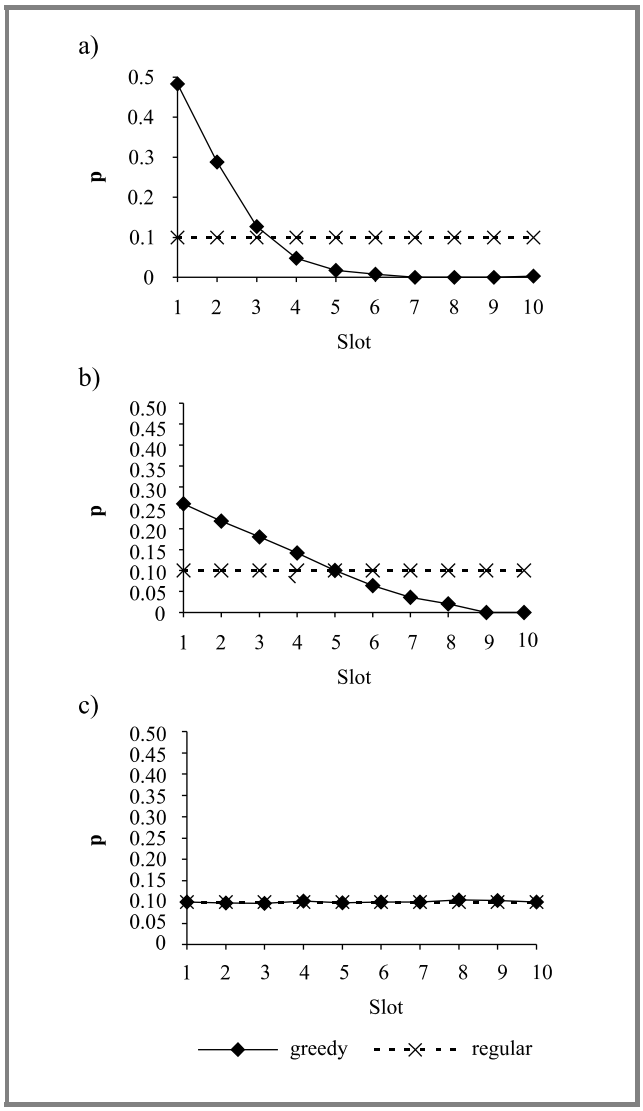


Fig. 9. Probability distribution **p** for Optimal Randomiser; $N = 10$, $G = 4$, $D = 10$: (a) RT; (b) RT-1s; (c) RT-hash.

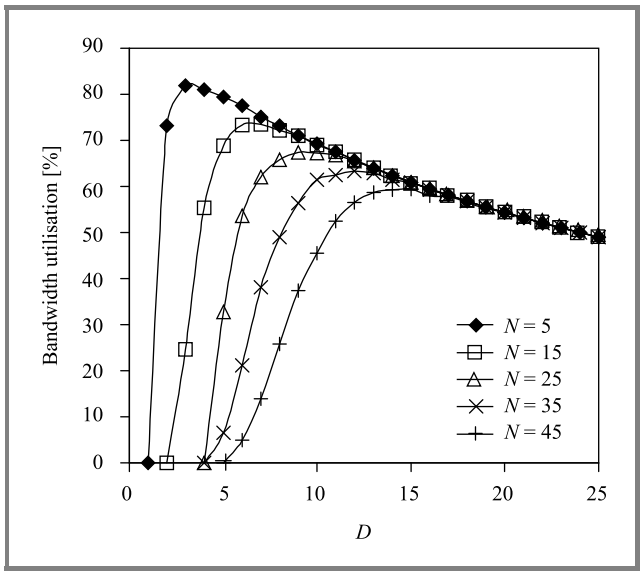


Fig. 10. Bandwidth utilisation under RT-hash.

carried out to determine the overhead under diverse parameter settings ($N = 5..45$, $D = 1..25$); the results are given in Fig. 10. For $N = 10$, the optimal D is equal to 4 and reduces the overhead to less than 20%. Unfortunately there is no uniformly optimal D across various N. A possible solution could be to make D variable and dependent on a current estimate of N, e.g., based on the observed sums $f_1 + ... + f_D$ in recent instances of contention.

In Fig. 11, regular stations used Randomiser, while greedy stations used Pseudoperiodic. For Pseudoperiodic, $T = D = 10$ and $\alpha = 0.95$ were fixed. The latter value implies that the e_i are influenced by the last few dozens of periods. The final deferment sequences at the greedy stations were found to be close to those in Table 1. The greedy stations, none of them having knowledge of the number or status of other stations, managed to establish a token passing sequence.

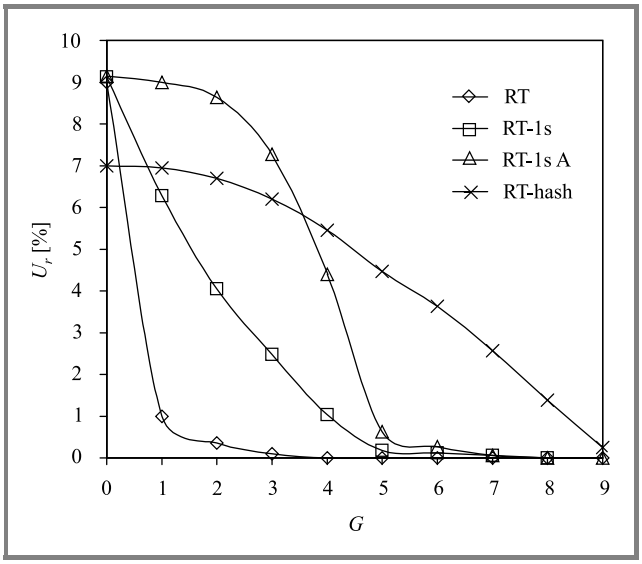


Fig. 11. Pseudoperiodic versus Randomiser.

Observe that as G increases, Pseudoperiodic turns out increasingly beneficial for the greedy stations, reflecting the fact that the token passing scheme inherently avoids pilot collisions, whereas Randomiser does not. However, the benefits depend on the protocol. RT is definitely not resistant to Pseudoperiodic; RT-1s is uniformly superior to RT, although for $G > N/2$ the results are comparably unsatisfactory. The dotted "RT-1s A" line corresponds to Aggressive Randomiser applied at regular stations under RT-1s. A better performance for $G < N/2$ is now observed; still, the range of unfavourable G remains the same. RT-hash prevents regular stations from being cut off even for larger G, at the price of an increased protocol overhead.

In each simulation run, the results presented in Fig. 11 were unfolding gradually. At the beginning, a greedy station's bandwidth share was comparable to a regular station's. Figure 12 shows a sample run under RT-1s. After about 1400 protocol cycles the greedy stations managed to establish a token passing sequence. In general, under all the considered protocols, it took greedy stations fewer

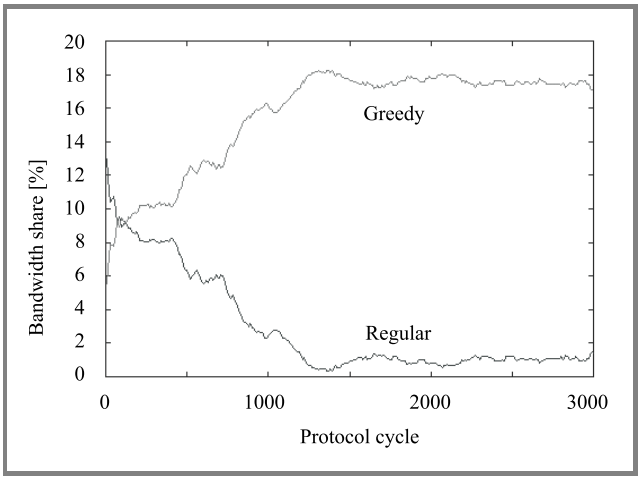


Fig. 12. Pseudoperiodic versus Randomiser under RT-1s; $G = 5$, $N = 10$.

than 2000 protocol cycles to stabilise their bandwidth shares at a high level. With 1500-byte packets and a 10 Mbit/s channel, this translates to less than 3 s. Thus a few seconds' time is enough to threaten regular stations even in changing traffic conditions. This raises the question, if Pseudoperiodic is so effective, could it be adopted as a regular station's standard strategy? The answer is no, for the following reasons:

– Pseudoperiodic is not fair in that the resulting bandwidth distribution heavily depends on the initial deferment sequences; in extreme cases, some greedy stations were observed to perform worse than regular ones,

– under RT-1s, Pseudoperiodic is not resistant to some simple selection strategies, e.g., consistent selection of a 0-slot deferment.

Finally, the performance of Antihash is given in Fig. 13. Note that this strategy might be beneficial only if there is only one greedy station; more would always collide. Therefore in our simulations $G = 1$ was fixed, with N ranging from 2 to 20. For reference, the dashed line indicates a station's bandwidth share if $G = 0$. Due to poor predictibility of the continuations $\mathbf{f}_{D-L}$, Antihash is less beneficial for larger N, though even for $N = 2$ the regular stations are not cut off. Taking a smaller D (discussed before), will make $\mathbf{f}_{D-L}$ even more unpredictable. Therefore Antihash is not a serious threat to RT-hash.

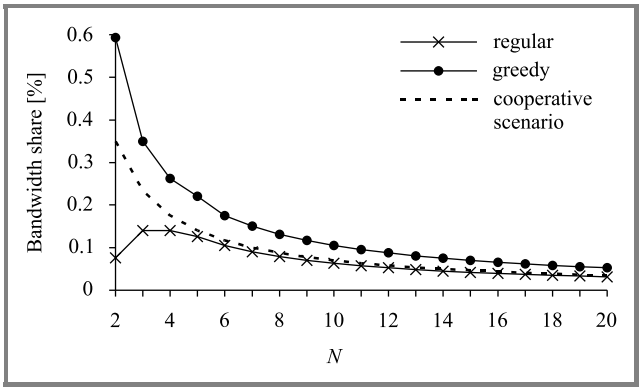


Fig. 13. Antihash versus uniform Randomiser under RT-hash; $G = 1$, $N = 2..20$, $L = 6$, $D = 10$.

6. Conclusion

A new RT-hash MAC protocol for wireless LANs has been proposed to protect regular stations from stations using greedy deferment selection strategies. Two such strategies, Optimal Randomiser and Pseudoperiodic, were considered; the former self-optimises the probability distribution of selected deferments; the latter attempts to establish a token passing-like scheme among greedy stations. RT-hash was simulated in a full-hearability configuration and compared with an earlier RT-1s protocol. Both protocols are comparably resistant to Optimal Randomiser, but only RT-hash is capable of counteracting Pseudoperiodic. A third strategy called Antihash attempted to exploit a potential weakness of RT-hash, but with minor success and only in a very restricted scenario.

References

[1] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks", *IEEE Wirel. Commun.*, vol. 9, no. 4, pp. 8–27, 2002.

[2] P. Michiardi, "CORE: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks", Res. Rep., Institut Eurecom., 2001.

[3] J. Al-Jaroodi, "Security issues in wireless mobile ad hoc networks at the network layer", Tech. Rep. TR02-10-07, University of Nebrasca-Lincoln, 2002.

[4] L. Buttyan and J. P. Hubaux, "Nuglets: a virtual currency to stimulate cooperation in self-organised mobile ad hoc networks", Tech. Rep. DSC/2001/001, Swiss Federal Institute of Technology, 2001.

[5] I. Chlamtac and A. Ganz, "Evaluation of the random token protocol for high-speed and radio networks", *IEEE J. Select. Areas Commun.*, vol. SAC-5, no. 6, pp. 969–976, 1987.

[6] "Radio Equipment and Systems (RES); HIgh PErformance Radio Local Area Network (HIPERLAN); Services and facilities", ETSI ETR 069 ed. 1 (1993-02).

[7] P. Karn, "MACA: a new channel access method for packet radio", in *Proc. 9th Comput. Netw. Conf. ARRL/CRRL Amateur Radio*, 1990, pp. 134–140.

[8] "IEEE Standard for Information Technology—LAN/MAN—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications", ISO/IEC 8802-11, 1999.

[9] J. Konorski, "Packet scheduling in wireless LANs: a framework for a noncooperative paradigm", in *Personal Wireless Communications*, J. Woźniak and J. Konorski, Eds. Boston [etc.]: Kluwer, 2000, pp. 29–42.

[10] J. Konorski, "Multiple access in ad hoc wireless LANs with noncooperative stations" in *Networking Technologies, Services and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, E. Gregori, M. Conti, A. T. Campbell, G. Omidyar, and M. Zukerman, Eds., *LNCS*. Berlin [etc.]: Springer-Verlag, 2002, vol. 2345, pp. 1141–1146.

[11] P. Obreiter, B. Koenig-Ries, and M. Klein, "Stimulating cooperative behaviour of autonomous devices—an analysis of requirements and existing approaches", Tech. Rep. 2003-1, University of Karlsruhe, 2003.

[12] P. Kyasanur and N. H. Vaidya, "Detection and handling of MAC-layer misbehavior in wireless networks", Tech Rep., University of Illinois at Urbana-Champaign, 2003.

[13] D. Fudenberg and J. Tirole, *Game Theory*. London, Cambridge (Mass): MIT Press, 1991.

Jerzy Konorski received his M.Sc. degree from the Technical University of Gdańsk (honours) and a Ph.D. degree from the Institute of Computer Science, Polish Academy of Sciences (honours). Since 1985 he has been Assistant Professor at the Department of Information Systems, Gdańsk University of Technology. He is the author of over 30 papers published in international journals or conference records, a co-author of another 10, all in the field of computer networking, and a co-editor of *Personal Wireless Communications* (Kluwer, 2000). His research interests are in performance evaluation, information systems, data transmission, operational research, distributed systems and computer networking. In 1993–96 he lectured at the EFP Franco-Polish School of New Information and Communication Technologies at Poznań and in 1994–96 was a local co-ordinator of a TEMPUS Joint European Project. In 2002 he was awarded the Erskine Fellowship at the University of Canterbury at Christchurch, New Zealand. Since 1995 he has been an IEEE member. Dr. Konorski is a Technical Program Committee member for Polish and Polish-German Teletraffic Symposia and a Scientific Council member at the Maritime Institute in Gdańsk.

e-mail: jekon@eti.pg.gda.pl
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Maciej Kurant received his M.Sc. degree from Gdańsk University of Technology (honours) in 2002. He is currently a Ph.D. student at the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. His main research interests are in ad hoc self organizing networks, survivability of optical networks and graph theory.

e-mail: maciej.kurant@epfl.ch
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland

Paper

Approximate performance analysis of slotted downlink channel in a wireless CDMA system supporting integrated voice and data services

Jacek Świderski

Abstract—This paper is concerned with the performance analysis of a slotted downlink channel in a wireless CDMA communication system with integrated packet voice and data transmission. The system model consists of mobile terminals (MT) and a single base station (BS). It is assumed that the voice (data) packet error rate (PER) does not exceed 10^{-2} (10^{-5}). With this requirement the number of simultaneous transmissions over the downlink channel is limited. Therefore, the objective of the call admission control is to restrict the maximum number of CDMA codes available to voice and data traffic. Packets of accepted voice calls are transmitted immediately while accepted data packets are initially buffered at the BS. This station distinguishes between silence and talkspurt periods of voice sources, so that data packets can use their own codes for transmission during silent time slots. Data packets are buffered in queues created separately for each destination. Discrete-time Markov processes are used to model the system operation. Statistical dependence between queues is the main difficulty which arises during the analysis. This dependence leads to serious computational complexity. The aim of this paper is to present an approximate analytical method based on the restricted occupancy urn model which enables to evaluate system performance despite the dependence. Numerical calculations compared with simulation results show excellent agreement for the average system throughput and the blocking probability of data packets for higher system loads. On the other hand, when the average data packet delay is considered, analytical results underestimate simulation and therefore only approximate system performance evaluation is possible.

Keywords—wireless CDMA system, synchronous downlink, voice and data integration, queueing analysis.

1. Introduction

The code division multiple access (CDMA) technique has been under intense research recently [1–3]. This technique can successfully be applied as an air interface in personal communication systems (PCS). In general, a PCS should provide integrated voice and data service to its users equipped with voice and data MTs. In such a system requirements for quality of service (QoS) are different for distinct traffic types. For example, a voice traffic requires real-time delivery (i.e., a guaranteed upper bound on delay with negligible fluctuations) and reasonably low packet error rate (PER). On the other hand, for a data traffic higher packet delays are allowed (i.e., packets are allowed to be buffered) and much lower PER is required.

In most papers on voice and data integration in CDMA systems call admission control for the uplink channel is analyzed [4–10]. In [4–8], silence/talkspurt period detection of the voice traffic is included in system models, while in [9, 10] speech activity is omitted. A downlink slotted CDMA channel transmitting voice and data traffic is considered in [11]. The silence and talkspurt periods of voice sources are not modeled (the voice activity factor is taken into account instead) and only one queue concentrating the whole data traffic at the BS is analyzed. This means that different destination MTs are not distinguished by the system model. Finally, in [12], approximate performance analysis of heavily loaded slotted downlink channel in a wireless CDMA system supporting integrated voice and data services is presented. The silence and talkspurt periods of voice sources are modeled and multi-packet data messages are assumed as an input, non-symmetric data traffic.

This paper is concerned with the performance analysis of a slotted downlink channel in a wireless CDMA communication system with integrated packet voice and data transmission. Its objective is to present an approximate analytical approach, which includes speech detection and makes the multi-queueing analysis of the downlink channel computationally tractable for the possibly wide range of the system load. Therefore, this paper can be considered as a contribution to the family of papers [4–12] devoted to performance analysis of voice and data integration in wireless CDMA systems.

2. System model

The system consists of MTs and a single BS. An MT can receive voice and/or data packets transmitted from the BS. It is assumed that a MT, which is able to receive voice and data traffic in the same time slot, has two receivers: one to receive voice and one to receive data packets. It is assumed that there are k MTs which are able to receive data packets in the system. Both voice and data packets are transmitted with the same rate and with the same energy per bit, thus the same pool of CDMA codes for both traffic

types is used. Voice and data packets are of the same size (n bits long) and the system is synchronized to slots of duration equal to the packet transmission time h_s. It is also assumed that bit interleaving and forward error correction codes with codeword length n are used.

The maximum of correctable bit errors for each voice (data) packet is assumed to be t_v (t_d). Typically, $t_v < t_d$ due to different quality of service requirements for voice and data traffic. It is assumed that the number of simultaneous packet transmissions is limited and accepted t_v (t_d) assures that the probability of receiving an uncorrectable voice (data) packet is less than 10^{-2} (10^{-5}).

Any downlink channel model which enables to evaluate for bit error rate (BER) as a function of k_s CDMA codes used simultaneously (i.e., BER (k_s)) can be employed. An example of such downlink channel model is presented in [13]. This model is used in this paper. Since the downlink channel is considered, all k_s simultaneously transmitted signals are both time and phase synchronous. They consist of independent and identically distributed data sequences taking values at ± 1 with equal probability. These data sequences represent voice and data packets and they are spread out with spreading codes—one code assigned to one transmitted signal. Spreading codes are sequences of independent and identically distributed random variables taking values at ± 1 with the same probability. These random variables are called chips and it is assumed that they have a rectangular pulse shape. It is also assumed that bit errors are caused only by other-user interference, allowing the effects of thermal noise on system performance to be ignored. Finally, equal received power for all signals is assumed.

Voice calls arrive at the BS according to Poisson process with a rate λ_v calls/slot. The duration of a voice call is exponentially distributed with an average $1/\mu_v$ slots and it is assumed that $1/\mu_v \gg 1$ and $1/\lambda_v \gg 1$. With the requirement that the voice PER does not exceed 10^{-2} and with proper choice of t_v, the number of simultaneous transmissions over the downlink channel is limited up to $L_v^{\max}$ ($L_v^{\max}$ is the maximum number of CDMA codes available for the voice traffic, i.e., the direct admission policy is accepted). As a result, voice call is accepted at the BS if there are less than $L_v^{\max}$ voice connections being established. Rejected voice calls are lost. If a voice call is accepted, a voice source is described by the well known ON-OFF model [5, 7, 8]. The ON state represents a voice source in a talkspurt mode (a user is speaking) while the OFF state represents a silence period during conversation. The length of both ON and OFF states is assumed to be geometrically distributed with means $1/P_{on-off}$ and $1/P_{off-on}$ slots, respectively. In the ON state, a voice source generates a stream of voice packets with constant rate equal to one packet per time slot. To serve this stream (i.e., to transmit voice packets to a given destination MT) a single CDMA code is used.

Channel capacity which is not occupied by the voice traffic can be used to serve data packets. The BS distinguishes between silence and talkspurt periods of voice sources, so that accepted data packets can use their own codes for transmission during silent time slots. Therefore, apart $L_v^{\max}$ CDMA codes used to serve voice traffic, the BS also uses k mutually orthogonal CDMA codes to serve data packets destined for k MTs.

Packets of accepted voice calls are transmitted immediately, while accepted data packets are initially buffered at the BS. Thus, voice packets can be treated as higher priority packets with respect to data packets considered as lower priority packets. Data packets are buffered in k FIFO queues created separately for each destination (i.e., for k MTs which are able to receive data packets). It is assumed that all queues have the same maximum length $L_q^{\max}$ (in packets) and that the external data packet traffic is symmetric. More precisely, a data packet destined for a given MT appears at the BS with probability λ_d in each time slot (λ_d is considered to be a data packet arrival rate in a single queue input). If the packet arrives at the BS and its queue is full, it is rejected and removed from the system. Moreover, since the probability of unsuccessful data packet transmission (because of downlink interference) is assumed to be less than 10^{-5}, retransmission of uncorrectable data packets is not considered (although it is possible in real systems).

The scheduling scheme enabling a given queue to use a CDMA code at the beginning of the next time slot t ($t = 1, 2, \ldots$) is described as follows. The scheduler calculates the number $n_v(t)$ of active voice sources in time slot t at first. Then, the algorithm determines the number on non-empty queues $k_{q\neq 0}(t)$ and the number of queues which can be served simultaneously in the same time slot t. This number is equal to $k_q(t) = \min\left[k_{q\neq 0}(t), L_v^{\max} - n_v(t)\right]$, which means that $k_q(t)$ CDMA codes can be used simultaneously to transmit data packets. Finally, the scheduler randomly determines $k_q(t)$ out of $k_{q\neq 0}(t)$ queues which can be served. The probability that a given queue can be served in time slot t is equal to $P_q(t) = \frac{k_q(t)}{k_{q\neq 0}(t)} = \min\left[1, \frac{L_v^{\max} - n_v(t)}{k_{q\neq 0}(t)}\right]$ (i.e., this probability is the same for all queues). Thus, this random mechanism assures that the scheduler is fair with respect to all non-empty queues.

With the assumed scheduler no more than $L_v^{\max}$ simultaneous voice and data transmissions in time slot t are allowed. As a result, a voice and data frame in the CDMA code domain is transmitted in each time slot. This frame has a movable-boundary for voice and data traffic. Assuming the number of accepted voice calls to be constant and equal to L_v, this voice and data boundary moves in the time slot domain within its range, following slot by slot changes of the number of active voice sources $n_v(t)$.

3. Analysis

For the assumed forward error correction codes with codeword length n and the maximum of correctable bit errors equal to t_v (t_d) for voice (data) traffic in conjunction with the interleaving technique, the probability of receiving an uncorrectable packet (i.e., the PER) when there are k_s si-

multaneous signals (i.e., k_s simultaneous CDMA codes) is given by [13, 14]:

$$P_{p,x}(k_s) = 1 - \sum_{i=0}^{t_x} \binom{n}{i} \left[P_e(k_s)\right]^i \left[1 - P_e(k_s)\right]^{n-1}, \quad x = v, d, \tag{1}$$

where index $x = v$ ($x = d$) is used for the voice (data) traffic and $P_e(k_s)$ is the known bit-error rate BER (k_s) on the downlink channel.

The PER expressed by Eq. (1) is an approximation which stems from the fact that this formula is exact only when bit errors are statistically independent. However, if bit interleaving is used (as it is assumed) one can expect that bit-to-bit error independence is hold [13]. Formula (1) is used to determine $L_v^{\max}$ (the maximum number of simultaneous voice and data packet transmissions in the downlink) for assumed transmission parameters. In order to calculate the probability $P_{p,x}(k_s)$ expressed by formula (1) it is necessary to know the probability $P_e(k_s) = \text{BER}\ (k_s)$. With the accepted assumptions this probability can be expressed as [13]:

$$P_e(k_s) = \frac{1}{\sqrt{2\pi}} \int_{\text{SNR}\ (k_s)}^{\infty} \exp\left(-\frac{u^2}{2}\right) du, \tag{2}$$

where $\text{SNR}\ (k_s) = \left(\frac{k_s - 1}{G}\right)^{-\frac{1}{2}}$ and G is the processing gain. The BER expressed by Eq. (2) is only an approximation obtained using the standard Gaussian approximation technique [13, 14]. This approximation can be good even for small number of simultaneous signals k_s, assuming relatively low processing gain G. A discussion about the accuracy of these approximation can be found in [13].

Formula (2) is used to calculate $P_{p,d}(k_s)$ expressed by Eq. (1). Moreover, the average BER for voice (data) traffic after decoding can be evaluated using the following formula [15]:

$$P_{e,x}(k_s) = \frac{1}{n} \sum_{i=t_x+1}^{n} \binom{n}{i} \left[P_e(k_s)\right]^i \left[1 - P_e(k_s)\right]^{n-i}, \quad x = v, d, \tag{3}$$

where index $x = v$ ($x = d$) is used for the voice (data) traffic.

Since it is assumed that the length of a data packet (this length is equal to one time slot) is much shorter than the average duration of a voice call (voice connection), i.e., $1/\mu_v \gg 1$, and it is also much shorter than the average interarrival time between two consecutive voice calls, i.e., $1/\lambda_v \gg 1$, the CDMA system under consideration can be modeled at either the voice call level or at the data packet level, depending on the particular system performance aspect being investigated [5, 12].

3.1. System performance at the data packet level

Due to the mentioned assumptions, it is reasonable to assume that for a constant number L_v of accepted voice calls the queueing system reaches the steady state before L_v changes. In order to evaluate efficiency of the system, some performance measures for the steady state are given below.

The average data packet throughput (in packets per time slot) can be expressed as:

$$\hat{\lambda}_{d|L_v} = \sum_{n_d=0}^{kL_q^{\max}} \Pr(n_d) \sum_{l=1}^{\min(n_d, L_v^{\max}, k)} l t(n_d, l), \tag{4}$$

where $\Pr(n_d)$ is the probability that there are n_d data packets at the BS, and $t(n_d, l)$ is the conditional probability of l simultaneous data packet transmissions given n_d packets at the BS (index $d|L_v$ means that data throughput is obtained by conditioning on the number of accepted voice calls L_v).

Next, the data packet blocking probability can be obtained as:

$$PB_{d|L_v} = 1 - \frac{\hat{\lambda}_{d|L_v}}{k\lambda_d}. \tag{5}$$

Finally, using Little's formula, the average data packet delay can be calculated as:

$$T_{d|L_v} = \frac{\overline{m}}{\hat{\lambda}_{d|L_v}}, \tag{6}$$

where $\overline{m} = \sum_{n_d=1}^{kL_q^{\max}} n_d \Pr(n_d)$ is the mean number of data packets at the BS (also including data packets in service).

In order to calculate the mentioned performance measures it is necessary to obtain the following probability distributions:

$$\{t(n_d, l) : n_d = 0, 1, \ldots, kL_q^{\max},\ l = 0, 1, \ldots, \min(L_v^{\max}, k)\},$$

$$\{\Pr(n_d) : n_d = 0, 1, \ldots, kL_q^{\max}\},$$

$$\{\Pr(n_v) : n_v = 0, 1, \ldots, L_v\}.$$

These probability distributions can be obtained using an analytical approach described below.

Two discrete-time Markov processes are used to model the system operation. Namely, with the accepted assumptions, once voice calls are admitted at the BS, their behavior depends only on their traffic sources and therefore the state evolution for the voice traffic is independent not only of the data traffic, but of the states of queues at the BS as well. Therefore, the system is considered as two statistically independent subsystems, each of which can be described by a discrete-time Markov chain. The first chain represents behavior of voice sources, while the second chain describes all (k) queues at the BS.

The state space of the chain representing the voice sources can be expressed as $\mathbf{S}_v(t) = \{n_v(t) : n_v(t) = 0, 1, \ldots, L_v\}$, where $n_v(t)$ is the number of active voice users (i.e., speaking users) in time slot t. The probability of a transition

from state $n_v(t)$ to state $n_v(t+1)$ can be expressed by the following formula [12]:

$$\Pr\big(n_v(t+1)\big|n_v(t)\big) = \sum_{x=0}^{\alpha}\binom{\alpha}{x}p^x(1-p)^{\alpha-x}\times$$

$$\times\quad\binom{L_v-\alpha}{|n_v(t+1)-n_v(t)|+x}q^{|n_v(t+1)-n_v(t)|+x}(1-q)^{\beta-x}, \quad (7)$$

where

$$\{\alpha,\beta,p,q\} =$$

$$=\begin{cases}\{n_v(t),L_v-n_v(t+1),P_{on-off},P_{off-on}\}, & n_v(t+1)\geq n_v(t)\\ \{L_v-n_v(t),n_v(t+1),P_{off-on},P_{on-off}\}, & n_v(t+1)<n_v(t)\end{cases}.$$

One can notice that from a given state $n_v(t)$ there are always L_v+1 possible transitions to all feasible states $n_v(t+1)\in\mathbf{S}_v(t)$.

Since all k queues at the BS are served by one common scheduler, these queues are always statistically dependent. Therefore, these queues cannot be analyzed separately and a system of $(1+L_q^{\max})^k$ linear equations has to be solved [12, 16]. However, in practice, the problem is intractable because of its computational complexity. In this situation an approximate analytical approach is used. The queueing system which stores and serves data packets at the BS is analyzed by expanding a linear approximate model proposed in [16] and then used in [17]. It is assumed that the system state depends only on the aggregate number n_d of data packets in the queueing system. This assumption allows to construct a Markov chain with $kL_q^{\max}+1$ linear equations. The state space of this Markov chain can be expressed as $\mathbf{S}_d(t)=\{n_d(t): n_d(t)=0,1,\ldots,kL_{q,i}^{\max}\}$, where $n_d(t)$ is the number of data packets at the BS in time slot t. The probability of a transition from state $n_d(t)=i$ to $n_d(t+1)=j$ can be obtained as:

$$\Pr\big(n_d(t+1)=j|n_d(t)=i\big)=P_{i,j}=\sum_{l=0}^{\min(k,i)}a_{j-i+l}(i)\,t(i,l)\ , \quad (8)$$

where $a_j(i)$ is the probability of j packets arriving in a time slot given i packets in the queueing system (i.e., at the BS), and $t(i,l)$ is the conditional probability of l data packet transmissions given i packets in the queueing system. To calculate the transition probabilities the restricted occupancy urn model is used [16].

Let $R(i,k,L_q^{\max})$ be the number of ways of distributing i indistinguishable balls (or data packets) among k distinguishable urns (or queues) under $L_q^{\max}$-(occupancy) restriction, given by:

$$R(i,k,L_q^{\max})=\sum_{j=0}^{k}(-1)^j\binom{k}{j}\binom{i+k-j\,(L_q^{\max}+1)-1}{k-1}. \quad (9)$$

Let also $R(i,k,L_q^{\max}\,|\,b_r=u)$ be the number of ways distributing i indistinguishable balls (or data packets) among k distinguishable urns (or queues), so that exactly u urns (or queues) have exactly r balls (or packets) under $L_q^{\max}$-restriction, derived as:

$$R(i,k,L_q^{\max}\,|\,b_r=u)=$$

$$=\binom{k}{u}\sum_{j=0}^{k-u}(-1)^j\binom{k-u}{j}R(i-(u+j)r,k-(u+j),L_q^{\max}). \quad (10)$$

Therefore:

$$\Pr(b_r=u\,|\,i)=\frac{R(i,k,L_q^{\max}\,|\,b_r=u)}{R(i,k,L_q^{\max})}. \quad (11)$$

Using the restricted occupancy urn model, the probability $a_j(i)$ can be expressed as:

$$a_j(i)=\sum_{u=j}^{k}\frac{R(i,k,L_q^{\max}\,|b_{L_q^{\max}}=k-u)}{R(i,k,L_q^{\max})}\binom{u}{j}\lambda_d^j(1-\lambda_d)^{u-j}. \quad (12)$$

Next, the conditional probability $t(i,l)$ can be calculated as:

$$t(i,l)=\sum_{d=l}^{\min(k,i)}\frac{R(i,k,L_q^{\max}\,|\,b_0=k-d)}{R(i,k,L_q^{\max})}s(d,l), \quad (13)$$

where $s(d,l)$ is the probability of l data packet transmissions in a time slot, given d non-empty queues. For the analyzed queueing system this probability can be obtained as:

$$s(d,l)=\begin{cases}1, & (d=l)\wedge(l\leq\gamma)\\ \sum\limits_{n_v=0}^{\min(L_v^{\max}-l,L_v)}\Pr(n_v), & (d=l)\wedge(l>\gamma)\\ \Pr(n_v=L_v^{\max}-l), & (d>l)\wedge(l\geq\gamma)\\ 0, & \text{otherwise}\end{cases}, \quad (14)$$

where $\gamma=L_v^{\max}-L_v$.

3.2. System performance at the voice call level

Due to the accepted assumptions, at the voice call level the downlink CDMA channel can be modeled as a $M/M/m/m$ queueing system, where $m=L_v^{\max}$. Applying this model, the probability that there are L_v voice connections served in the system is given by [18]:

$$\Pr(L_v)=\frac{\frac{(\lambda_v/\mu_v)^{L_v}}{L_v!}}{\sum\limits_{i=0}^{L_v^{\max}}\frac{(\lambda_v/\mu_v)^i}{i!}}, \quad (15)$$

and the blocking probability of voice calls $PB_v(\rho_v)$ is given by Erlang's loss formula, i.e., $PB_v(\rho_v)=\Pr(L_v=L_v^{\max})$, where $\rho_v=\lambda_v/\mu_v$. On the other hand, the data packet throughput, the data packet blocking probability, and the average data packet message delay which have been determined by conditioning on L_v at the data packet level have to be obtained now by taking the distribution of L_v into

account. Therefore, the basic performance measures at the voice call level can be expressed as follows:

$$\hat{\lambda}_d(\rho_v,\lambda_d)=\sum_{L_v=0}^{L_v^{\max}}\hat{\lambda}_{d|L_v}(\lambda_d)\Pr(L_v)\,, \tag{16}$$

$$PB_d(\rho_v,\lambda_d)=\sum_{L_v=0}^{L_v^{\max}}PB_{d|L_v}(\lambda_d)\Pr(L_v)\,,$$

and finally

$$T_d(\rho_v,\lambda_d)=\sum_{L_v=0}^{L_v^{\max}}T_{d|L_v}(\lambda_d)\Pr(L_v)\,,$$

where $\hat{\lambda}_{d|L_v}(\lambda_d)$, $PB_{d|L_v}(\lambda_d)$, and $T_{d|L_v}(\lambda_d)$ are determined by formulas (4), (5), and (6), respectively. One can notice that system performance evaluation at the voice call level seems to be a simple task and therefore only system performance at the data packet level is farther considered.

4. Numerical example

In order to illustrate the presented approximate performance analysis and to verify its validity, an integrated wireless voice and data CDMA system described by the parameters given in Table 1 is considered.

Table 1
System parameters

Parameter	Symbol	Quantity	Unit
Processing gain	G	16	
Voice and data packet length	n	255	[bit]
Maximum number of correctable bit errors per data packet	t_d	18	
Maximum number of correctable bit errors per voice packet	t_v	12	
Slot duration	h_s	20	[ms]
Average voice call duration	$1/\mu_v$	7500	[slot]
Maximum number of voice connections	$L_v^{\max}$	5	
Transition probability of a voice source from ON- to OFF-state	$P_{on\text{-}off}$	1/17	
Transition probability of a voice source from OFF- to ON-state	$P_{off\text{-}on}$	1/22	
Data packet arrival rate in a single queue input	λ_d	0.1–0.9	[pack./ slot]
Number of MTs accepting data	k	10	
Maximum length of a queue of data packets	$L_q^{\max}$	4	
Number of voice connections served by the system	L_v	1, 3, 5	

It is assumed that the processing gain is relatively low ($G=16$) and only five signals can be transmitted simultaneously (this is explained below). For these values the approximate BER analysis is accurate enough [13], but on the other hand relatively strong error-correction capabilities of coding is required (e.g., BCH code can be used with $n=255$ and with $t_v=12$ and $t_d=18$, for voice and data packets, respectively) to satisfy different quality of service for both traffic types. Namely, it is assumed that the system has to ensure the probability of receiving an uncorrectable voice (data) packet to be less than 10^{-2} (10^{-5}). One can find that $P_{p,v}(5)=6.3\cdot10^{-3}$ and $P_{p,d}(5)=8.2\cdot10^{-6}$ provided that formula (1) is used. Since $P_{p,v}(6)>10^{-2}$, therefore in the sequel it is assumed that at most five simultaneous transmissions are allowed on the downlink (i.e., $L_v^{\max}=5$). For $k_s=5$ the average BER after decoding for voice (data) traffic calculated using (3) is equal to $P_{e,v}(5)=3.3\cdot10^{-4}$ ($P_{e,d}(5)=6.2\cdot10^{-7}$). Queueing model validity is evaluated by simulation. The BS is simulated as a system of k queues with slotted service time and the single common scheduler, which enables/disables the use of CDMA codes for transmission of data packets in each time slot. The method of independent replications is used [19]. A 95% confidence interval is generated by applying standard arguments based on the t-distribution.

A systematic investigation of the accuracy of the analytical approach is presented in Figs. 1, 2, and 3. More precisely, numerical calculations are compared with the simulation results for different numbers of voice connections, i.e., $L_v=1$, 3, and 5. In all calculations the maximum queue length is chosen to be $L_q^{\max}=4$. The total system load of data packet traffic $k\lambda_d$ changes with respect to $\lambda_d\in[0.1,\ 0.9]$. High system load is assumed, since the number k of destination MTs accepting data packets is chosen to be 10. An excellent agreement between numerical and simulation results is obtained for the average data throughput (Fig. 1). A different situation occurs when the average data packet delay is analyzed. Namely, Fig. 2 shows that numerical calculations underestimate simulation results for the assumed range of λ_d. It can also be noticed that the higher the system load, the better the approximation

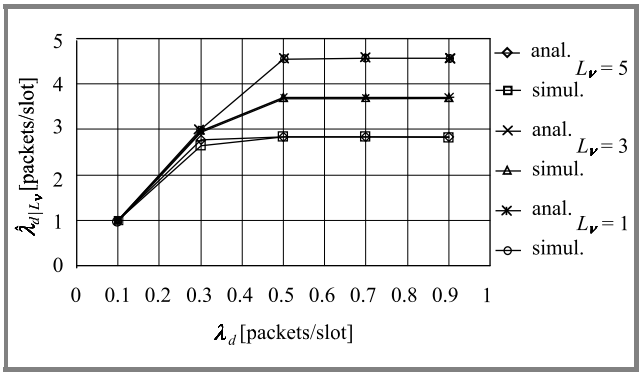


Fig. 1. Average data packet throughput $\hat{\lambda}_{d|L_v}$ versus system load λ_d ($L_q^{\max}=8$, $L_v=1,\ 3,\ 5$).

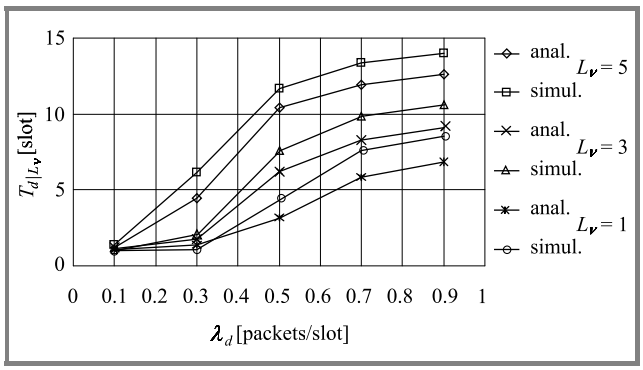


Fig. 2. Average data packet delay $T_{d|L_v}$ versus system load λ_d ($L_q^{\max} = 8$, $L_v = 1, 3, 5$, $h_s = 20$ ms).

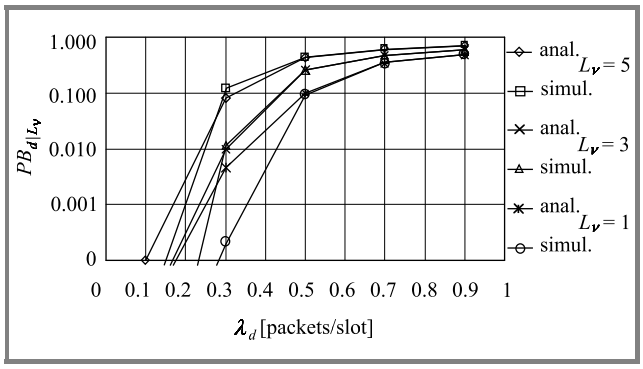


Fig. 3. Data packet blocking probability $PB_{d|L_v}$ versus system load λ_d ($L_v = 1, 3, 5$).

(when $\lambda_d = 0.9$ the approximation error is less than 20%). A similar observation can be done when the blocking probability of data packets is considered. However, one can notice that the approximation error is very low (it is less than 2%) only when the average packet input rate is higher than the output packet service rate, i.e., the total system load $k\lambda_d$ is such that $k\lambda_d > \left[L_v^{\max} - \sum_{n_v=1}^{L_v} n_v \Pr(n_v)\right]$, and the probability that queues are not empty is very low. This occurs for $\lambda_d > 0.46$, 0.37, 0.28 and for $L_v = 1$, 2, 5, respectively.

5. Conclusions

An approximate performance analysis of slotted downlink channel in a wireless CDMA system transmitting voice and data packets has been presented. Unfortunately, analyzed queues at the BS are statistically dependent, since a common scheduler allows them to use their allocated CDMA codes in subsequent time slots. The Markov analysis requires a system of linear equations to be solved. In practice however, with the accepted system model, the problem is computationally intractable, because the number of equations is unacceptably large. In order to make the analysis computationally tractable the restricted occupancy urn model has been used to estimate the probability distribution of data packets in the analyzed system. Numerical calculations, compared with simulation results, show excellent agreement for the average system throughput and the blocking probability of data packets for higher system loads. On the other hand, when the average data packet delay is considered, analytical results underestimate simulation and therefore only approximate system performance evaluation is possible.

It has also been shown that performance results strongly depend on the number of simultaneous voice connections L_v in the system, even when silence periods of voice sources are used for data transmission. These results can be averaged, using probability distribution of the number of simultaneous voice connections. Therefore, system performance as a function of both data and voice traffic load can easily be evaluated.

References

[1] M. B. Pursley, "The role of spread spectrum in packet radio networks", *Proc. IEEE*, vol. 75, no. 1, pp. 116–134, 1987.

[2] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading, MA: Addison-Wesley, 1995.

[3] R. Prasad and T. Ojanpera, "An overview of CDMA evolution toward wideband CDMA", *IEEE Commun. Surv.*, vol. 1, no. 1, pp. 2–28, Fourth Quarter 1998.

[4] N. D. Wilson, R. Ganesh, K. Joseph, and D. Raychaudhuri, "Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN", *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 870–884, 1993.

[5] T. Liu and J. A. Silvester, "Joint admission/congestion control for wireless CDMA systems supporting integrated services", *IEEE J. Select. Areas Commun.*, vol. 16, no. 6, pp. 845–857, 1998.

[6] W. Yue and Y. Matsumoto, "Output and delay process analysis for slotted CDMA wireless communication networks with integrated voice/data transmission", *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1245–1253, 2000.

[7] J. M. Capone and L. F. Merakos, "Integrating data traffic into CDMA cellular voice system", *Wirel. Netw.*, vol. 1, pp. 389–401, 1995.

[8] R. Fantacci and L. Zoppi, "A CDMA wireless packet network for voice/data transmissions", *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1162–1166, 1997.

[9] M. Soroushnejad and E. Geraniotis, "Multi-access strategies for an integrated voice/data CDMA packet radio network", *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 934–945, 1995.

[10] E. Geraniotis, M. Soroushnejad, and W.-B. Yang, "A multi-access scheme for voice/data integration in hybrid satellite/terrestrial packet radio networks"", *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1756–1767, 1995.

[11] Y. Yang and E. Geraniotis, "Admission policies for integrated voice and data traffic in CDMA packet radio networs", *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 654–664, 1994.

[12] J. Świderski, "Approximate performance analysis of heavily loaded slotted downlink channel in a wireless CDMA system supporting integrated voice/data services", accepted for publication in *IEEE Trans. Wirel. Commun.*

[13] R. K. Morrow Jr. and J. S. Lehnert, "Bit-to-bit error dependence in slotted DS/SSMA packet systems with random signature sequences", *IEEE Trans. Commun.*, vol. 37, no. 10, pp. 1052–1061, 1989.

[14] A. Polydoros, A. Anastasopoulos, T.-B. Liu, P. Panagiotou, and C.-M. Sun, "An integrated physical/link-access layer model of packet radio architecture", California PATH Res. Rep. UCB-ITS-PRR-94-20, Oct. 1994.

[15] L. B. Milstein, T. S. Rappaport, and R. Barghouti, "Performance evaluation for cellular CDMA", *IEEE J. Select. Areas Commun*., vol. 10, no. 4, pp. 680–689, 1992.

[16] A. Ganz and T. Chlamtac, "A linear solution to queueing analysis of synchronous finite buffer networks", *IEEE Trans. Commun*., vol. 38, no. 4, pp. 440–446, 1990.

[17] H. Okada, T. Yamazato, M. Katayama, and A. Ogawa, "Queueing analysis of CDMA slotted ALOHA system with finite buffer and finite population assumptions", in *Proc. IEEE Int. Conf. Commun*., Atlanta, USA, 1998, pp. 407–411.

[18] L. Kleinrock, *Queueing Systems*. Vol. 1, *Theory*. New York: Wiley, 1975.

[19] *Computer Performance Modeling Handbook*, S. S. Lavenberg, Ed. New York: Academic Press, 1983.

Jacek Świderski was born in Gdańsk, Poland, in 1951. He received the M.S. and Ph.D. degrees in electronic engineering from the Technical University of Gdańsk in 1974 and 1982, respectively. He is now with the Institute of Power Engineering, Gdańsk, as a Research Staff Member. His main current interest is in the area of computer communication systems including modeling, analysis, and performance evaluation.

e-mail: j.swiderski@ien.gda.pl
Institute of Power Engineering, Div. Gdańsk
Mikołaja Reja st 27
80-870 Gdańsk, Poland

Paper

Analysis of bandwidth reservation algorithms in HIPERLAN/2

Józef Woźniak, Tomasz Janczak, Przemysław Machan, and Wojciech Neubauer

Abstract—This paper focuses on performance of channel access methods in the HIPERLAN/2 standard. It discusses commonly used approaches to bandwidth allocation and presents a modified algorithm for effective bandwidth management based on pre-scheduled resource grants. Simulation results show that the new algorithm ensures much higher throughput compared to the standard method.

Keywords—HIPERLAN/2, MAC, bandwidth allocation, QoS.

1. Introduction

Wireless LANs have gained market acceptance over the last several years, partly because of the increased demand for wireless communications and advances in portable computers and networking technology. Although the first WLAN solutions were used as a cordless replacement for Ethernet networks, it is clear that in the near future WLANs will have to keep up with the growing demand for bandwidth-consuming multimedia traffic. In particular that means effective support for service differentiation.

Today, two institutes lead the development and standardization of wireless LANs, namely the Institute of Electrical and Electronics Engineers (IEEE) and European Telecommunications Standards Institute (ETSI). Since 1996, when IEEE first published the 802.11 standard [6] offering 2 Mbit/s data rate, a lot of work has been done to increase the transmission speed. The IEEE 802.11b supplement, published in 1999, enhances available data rates to 5 Mbit/s and 11 Mbit/s. Another IEEE group developed the 802.11a version, which exploits the orthogonal frequency-division multiplexing (OFDM) technique to achieve data rates up to 54 Mbit/s. Although IEEE 802.11 defines a number of new physical layers, the originally proposed MAC protocol remains untouched. This approach promotes compatibility, but it retains the legacy MAC layer that lacks effective quality of service (QoS) support.

While IEEE worked on 802.11, ETSI proposed the high-performance radio LAN (HIPERLAN/1) standard, offering up to 18 Mbit/s data rate [1]. Unlike the 802.11 solution, the HIPERLAN/1 MAC protocol uses frame priorities, and thus it has means to support quality of service differentiation. This allows HIPERLAN/1 to effectively transmit a variety of information types, such as data, video and voice. In spite of this advantage, HIPERLAN/1 devices were never introduced. Therefore, ETSI developed the HIPERLAN Type 2 solution [2] which had a number of attractive features as compared to 802.11a. One of them is higher throughput. As both 802.11a and HIPERLAN/2 use the same OFDM coding technique they reach the same maximum data rates of 54 Mbit/s at the physical layer. However, from user perspective the maximum throughput of HIPERLAN/2 is 42 Mbit/s, whereas throughput of 802.11a is only around 18 Mbit/s. The other significant differences are on MAC layers. IEEE 802.11 implements distributed CSMA with collision avoidance. HIPERLAN/2 (H/2) employs a central controller to coordinate transmissions. Like the first version, HIPERLAN/2 inherently supports quality of service differentiation.

Although, IEEE 802.11 seems to prevail on the market now, it is possible that the advantages of HIPERLAN/2 will attract attention of customers, especially in the face of growing throughput demand [8].

In this paper, we analyse the efficiency of bandwidth allocation methods in HIPERLAN/2. We also introduce a new allocation scheme that significantly increases the overall network throughput.

This paper is organized as follows:

- Section 2 outlines the HIPERLAN/2 architecture.
- Section 3 discusses the original bandwidth allocation scheme.
- Section 4 describes the modified algorithm.
- Section 5 presents simulation results that compare performance of both solutions.
- Section 6 concludes the paper.

2. HIPERLAN Type 2

HIPERLAN/2 is intended for short-range, high-speed radio communication systems with data rates from 6 Mbit/s to 54 Mbit/s. It connects portable devices with broadband networks based on IP, ATM, and other technologies [11]. H/2 specifications cover two lowest layers of the OSI model, namely the physical (PHY) and the data link control (DLC) layer. Figure 1 illustrates the protocol stack. Furthermore, the DLC specification is split into two parts:

– basic MAC protocol [3],

– radio link control services (responsible for connection maintenance).

A convergence sublayer between DLC and upper layers adapts the H/2 to existing network architectures. The convergence options currently considered are:

– packet based networks, either Layer 2 (Ethernet) or Layer 3 (IP),

– cell based networks (ATM) standardized by the ATM Forum,

– UMTS based services, developed by the 3GPP partnership project.

The medium access control (MAC) protocol operates in a centralized manner, where a single node controls data transmission in a subnetwork cell. In a business environment, all terminals communicate using a fixed access point (AP). In a home environment, the capability of direct link communication is provided within a single subnet [4]. HIPERLAN/2 uses this mode to create ad hoc subnetworks without relying on the cable infrastructure. In this case, a central controller (CC) is dynamically selected from the portable devices. The CC node is capable of supporting multi-media applications by providing mechanisms to handle QoS reservations and allocate bandwidth, same as AP [5]. There is also a proposed solution for inter-subnet forwarding [9, 10].

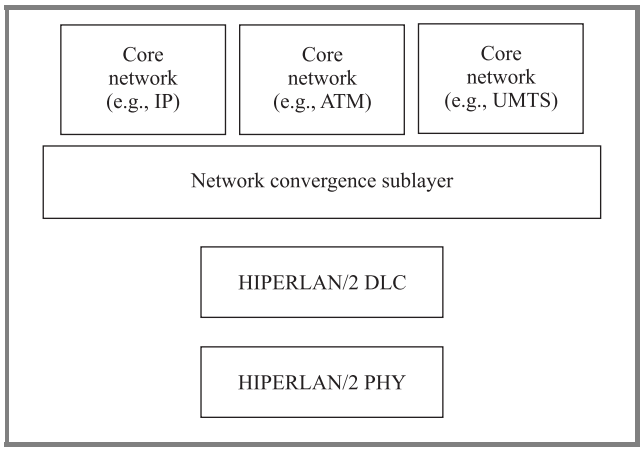


Fig. 1. HIPERLAN/2 protocol stack overview.

In both modes, the medium access control protocol employs a dynamic TDMA/TDD scheme. The MAC frame has a fixed length of 2 ms. An AP or CC node generates the frame structure, as illustrated in Fig. 2.
Each MAC frame comprises:

- **Fixed broadcast channel (BCH)**. This is a downlink channel that identifies AP/CC and a subnet. Broadcast channel denotes the beginning MAC frame.

- **Variable-length frame channel (FCH)**. It describes the structure of the MAC frame in terms of resource grant (RG) messages. RGs hold information about the assigned short channels (SCHs) and long channels (LCHs) in data transmission phases. RGs also inform that a MAC frame contains data that must be received by a particular mobile station in the subnetwork.

- **Variable-length access feedback channel (ACH)**. The ACH stage informs terminals about the results of the contention-based access attempts in RCH slots of the previous frame. ACH includes a bit field; if a bit corresponding to RCH is set the transmission was successful. Otherwise, a collision has occurred.

- **Downlink (DL), direct (DiL) and uplink (UL) data transmission phases**. These phases include SCH and LCH slots. The SCH slot is used to control traffic, while the LCH slot carries user data traffic. If a direct mode is used, the MAC frame can also contain a DiL phase, inserted between the DL and UL phases.

- **At least one random channel (RCH)**. The RCH slots make it possible for the terminal to send unsolicited control information to AP/CC, such as resource reservation (RR) messages. Nodes access RCH on a contention basis, according to the back-off algorithm. Additionally, stations use RCHs for the first contact with AP/CC to register themselves in a subnet.

The existing resource reservations and a scheduling algorithm determine the structure of a MAC frame, except for RCH. As a result, to maximize the overall bandwidth, the number of RCH slots should be kept to a minimum. However, the optimal number of RCHs depends on the current traffic load. In particular, it should be dynamically adjusted to the number of stations using these slots.
The RCH access algorithm employs a contention window scheme. Initially, a station sets the window size to n. If a station fails to access the RCH slot, it changes the window size according to the number, a, of unsuccessful attempts. Next, the station selects a random number from the range $< 0,\ CW_a >$:

$$CW_a = \begin{cases} 256 & 2^a \geq 256 \\ 2^a & n < 2^a \leq 256 \\ n & n \geq 2^a \end{cases} .$$

The selected value corresponds to the RCH slot in the current MAC frame or in one of the successive frames. After the transmission, station waits for positive feedback in FCH.

Bandwidth allocation. To control the allocation of network resources, AP or CC needs to know the state of its own buffers and buffers in the mobile terminals (MTs). MTs report their buffer states in resource request messages sent to AP or CC. Optionally, MT negotiates a fixed capacity allocation upon connection establishment. In this case, a mobile terminal does not need to explicitly request resources with RR messages.

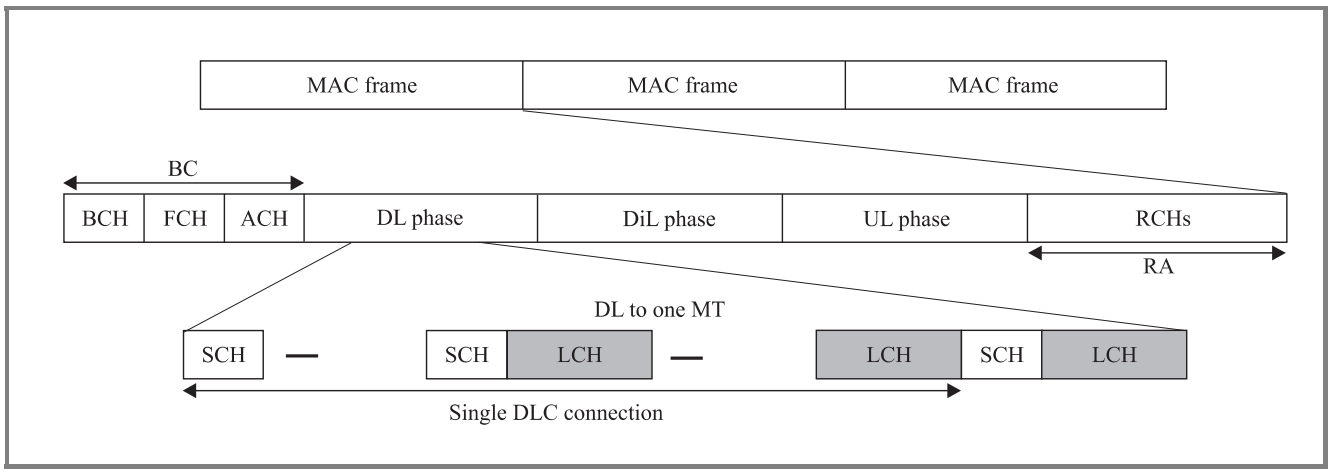


Fig. 2. Generic structure of a MAC frame in HIPERLAN/2.

In response to RR messages (or negotiated capacity), the AP/CC node allocates resources to mobile terminals. The central node should allocate the resources according to the buffer states and, if required, take quality of service parameters into account. The allocated resources are announced in resource grant messages, transmitted in the FCH phase. According to [9], three classes of MAC scheduling algorithms are considered:

- static resource allocation,
- dynamic resource allocation,
- priority-based scheduling (a combination of the above).

The static allocation guarantees transmission of a specified number of LCHs and SCHs per MAC frame. A disadvantage of this method is the low flexibility in changing capacity demand. The H/2 standard provides two algorithms of this class:

- fixed slot allocation (FSA),
- fixed capacity agreement (FCA).

In both functions the estimated capacity is negotiated while connecting. FSA is only available in home extension of the H/2 standard [4]. It assumes that a fixed part of a MAC frame is assigned to the particular connection. If the resource demand changes, the allocation can be adjusted by a connection modification procedure. The FCA function periodically assigns a number of LCHs and SCHs for a connection. The agreement can be adjusted by using RR messages.

Implementations of dynamic bandwidth allocation in the H/2 standard are based on resource request messages generated by stations. RR messages convey the current state of the DLC transmission buffer. Each station maintains a separate logical buffer for every DLC connection. According to the H/2 standard, RR messages can be carried in SCH slots granted to a station or in a randomly chosen RCH slot. However, ETSI does not recommend the use of RCH for RR messages.

There are several methods proposed for an AP/CC station to perform dynamic allocation; such as round-robin and its modifications [7]. Non-exhaustive round-robin assigns one LCH for every DLC until the MAC frame is completely filled. Exhaustive round-robin serves the first connection completely until the next connection is considered.

Priority-based scheduling is a combination of static and dynamic resource allocation. For example, the H/2 Ethernet convergence layer recommends using of 802.1p traffic priorities. In this case, a scheduling algorithm allocates a different number of LCHs for every DLC (based on connection priority).

3. Standard bandwidth allocation

This section describes the proposed resource allocation method in the HIPERLAN/2 standard. The method assumes that RR is carried in SCH if it has been assigned by the scheduling algorithm in AP. Otherwise, MT selects the RCH slot to send a resource request message to AP, as shown in Figs. 3 and 4. In response, AP allocates one additional SCH (exclusively) to the RR message.

Several simulation experiments have been carried out to verify the effectiveness of the standard method. In each experiment, a scenario with one hundred MTs and one AP was examined. Every mobile station established one best effort DLC connection with AP. The best effort traffic was modelled as a Poisson source of fixed-length packets. Moreover, all stations generate traffic streams of the same intensity. The traffic stream is injected under Ethernet service specific convergence sublayer. The lengths of the SSCS frames are 1500 bytes, 500 bytes and 60 bytes.

Figure 5 illustrates throughput characteristics of the 1500-byte-long SSCS frames. The expected behaviour is that net-

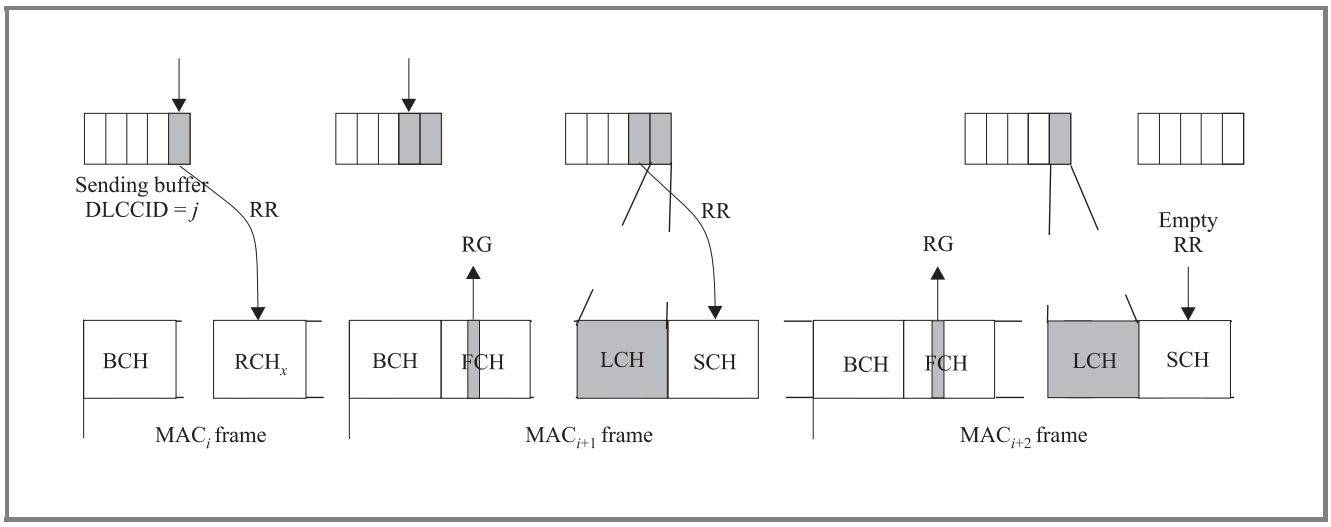


Fig. 3. Original algorithm.

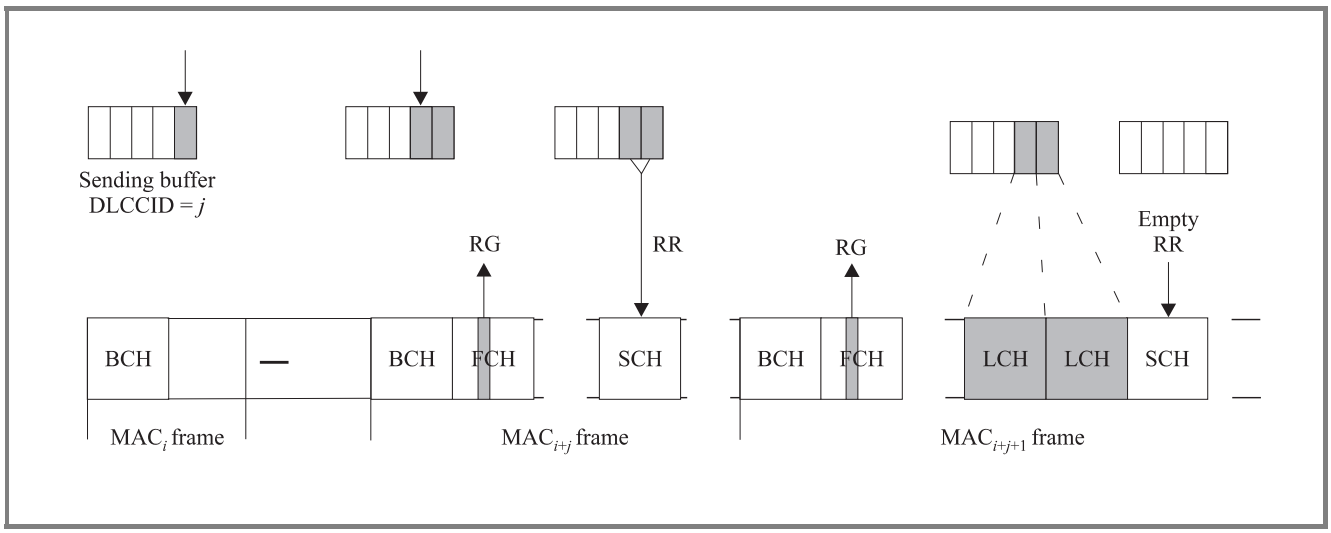


Fig. 4. Modified algorithm.

work throughput will be proportional to the total traffic load up to the network saturation point, and then constant beyond the saturation point. However, as shown in Figs. 5, 6, and 7, even 5 RCH slots are not enough to assure the desired throughput characteristics before the saturation point. Basically, if the number of RCH slots is too small, stations have to wait a long time before sending RR messages. Moreover, because of the enormous number of collisions in the RCH phase, a lot of shared bandwidth is wasted. The network resources are utilized better as the number of slots increases. However, there is an obvious trade-off between the number of slots in RCH and the bandwidth available for the user data. The optimal number of RCH slots can be estimated as a balance point between the low number of collisions and the high bandwidth utilization.

Unfortunately, the optimal number of slots strongly depends on the length of the SSCS frame (as shown in Figs. 6 and 7). Furthermore, in case of heterogeneous traffic the problem becomes even more complicated, because the optimal number of RCH slots for 1500 byte frames is not the same as for 60 byte frames. As a result, adjusting the optimal number of slots to current load conditions becomes a non-trivial issue.

Given such conclusions, we developed a solution that does not rely on dynamic adjustment of RCH slots. Instead, the modified algorithm tries to minimize the number of RCH users, thus decreasing the frequency of collisions and the amount of wasted bandwidth.

4. Modified algorithm

In the standard algorithm, the problem of bandwidth degradation escalates as the SSCS frame becomes shorter. This is because an RR message is sent on every transmission buffer change. From the simulation experiments it is clear

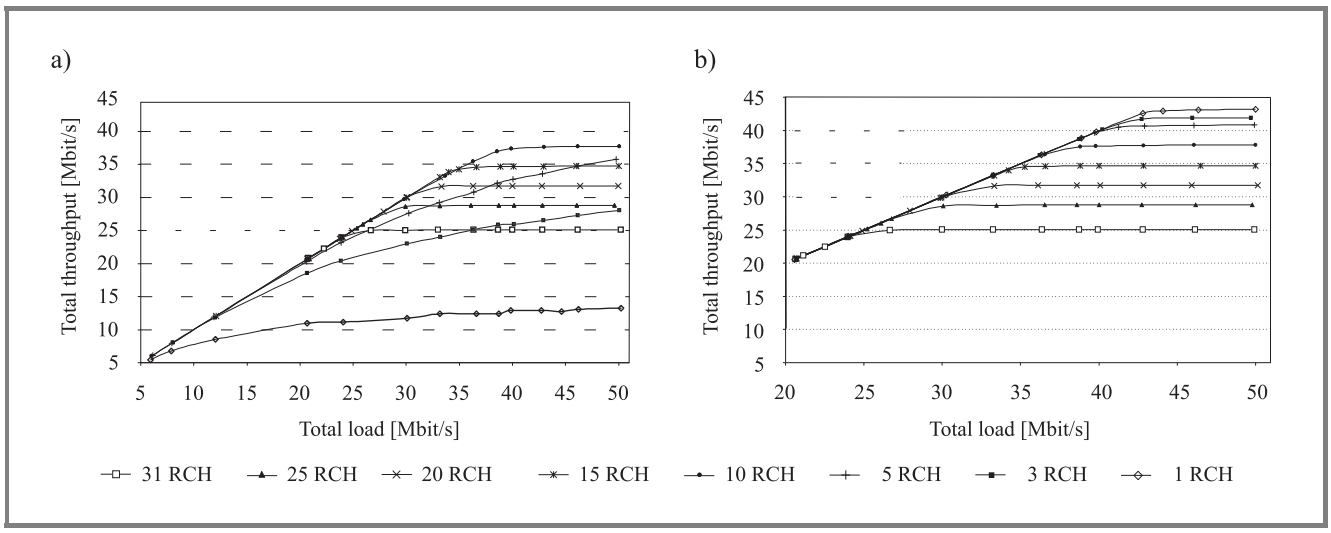


Fig. 5. Comparison of standard (a) and modified (b) bandwidth allocation methods for 1500-byte packets.

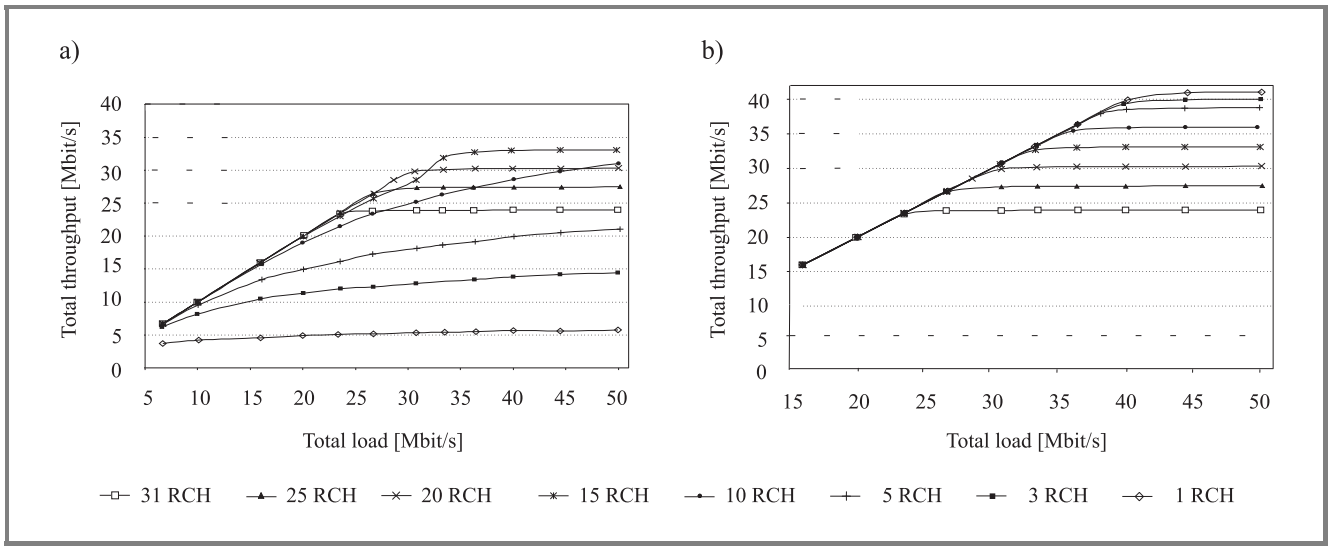


Fig. 6. Comparison of standard (a) and modified (b) bandwidth allocation methods for 500-byte packets.

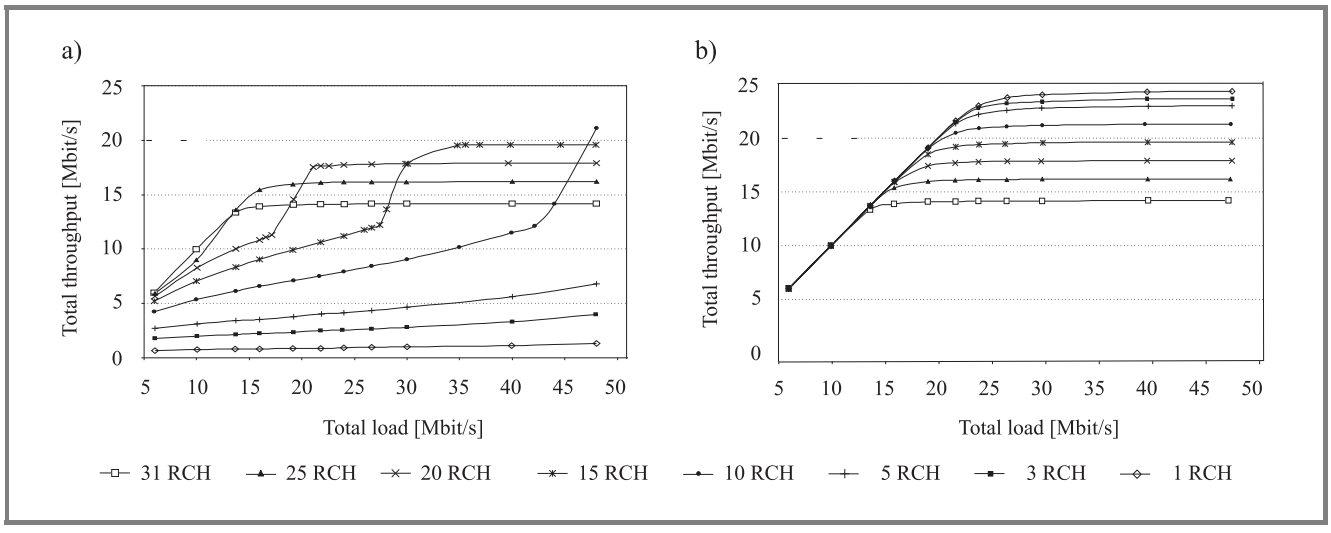


Fig. 7. Comparison of standard (a) and modified (b) bandwidth allocation methods for 60-byte packets.

that stations tend to use RCH rather then SCH channels, as data packets become shorter. The increasing number of RCH access attempts results in more collisions, which in turn causes non-optimal bandwidth utilization.

The modified method (Fig. 4) assumes that stations do not use RCH slots to send RR messages. To avoid the overflow of station transmission buffers, AP periodically polls MTs for their buffer status. To poll a station, AP arbitrarily allocates SCH for inactive DLC connections. A connection is considered inactive if there were no resources granted to it in the previous MAC frame. In a granted SCH channel, a station reports the number of LCHs needed to service the buffered frames. The modified method increases the chance that a station uses SCHs instead of RCH to communicate resource requests. However, RCH cannot be eliminated completely (for example, RCH is used for DLC connection set-up).

AP allocates SCH to poll stations using the round robin algorithm. The algorithm periodically allocates one SCH for each inactive connection. If there are no resources in the current MAC frame, the algorithm allocates SCH to poll a station in the next MAC frame.

The cooperation of standard and modified methods is also possible. If a station buffer state changes, the "backoff" timer is set and the station waits for SCH granted to send RR. If the timer reaches zero, and no SCH has been allocated, the station sends RR in RCH.

5. Simulation results

The modified method ensures better utilization of network bandwidth as shown in Figs. 5, 6, and 7. The new method demonstrates advantages, regardless of the frame length. Moreover, since RCH slots do not carry RR messages, the adjustment of number of slots is outside the scope of the scheduling algorithm.

In addition, the standard method has a higher probability of dropping incoming packets in comparison to the modified method. This is because of collisions in RCH slots. In the H/2 standard, RR messages convey information on one DLC connection transmission buffer. The number of access attempts in RCH is proportional to the number of DLC connections, not the number of mobile terminals.

6. Conclusions

This paper presented the current state of work on bandwidth allocation in the HIPERLAN/2 standard. A new solution was also proposed, based on pre-scheduled resource grants. The performance characteristics—obtained via simulation—prove that our algorithm ensures higher throughput compared to the standard method.

Moreover, the new proposal remains almost completely independent of the traffic pattern, making this strategy particularly suitable for supporting real-time traffic with strong QoS requirements. At the same time, the new priority-based transmission scheme does not introduce any organizational overhead to the MAC protocol.

References

[1] "Broadband Radio Access Networks (BRAN); HIgh PErformance Radio Local Area Network (HIPERLAN) Type 1; Functional Specification", ETSI EN 300 652 V1.2.1 (1998-07).

[2] "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; System Overview", ETSI TR 101 683 V1.1.1 (2000-02).

[3] "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions", ETSI TS 101 761-1 V1.1.1 (2000-04).

[4] "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 4: Extension for Home Environment", ETSI TS 101 761-4 V1.3.1 (2001-07).

[5] J. Habetha, A. Hettich, J. Peetz, and Y. Du, "Central controller handover procedure for ETSI-BRAN HIPERLAN/2 ad hoc networks and clustering with quality of service guarantees", in *1st IEEE Ann. Worksh. Mob. Netw. Comput.*, Boston, USA, 2000.

[6] "IEEE Standard for Information Technology—LAN/MAN—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications", ISO/IEC 8802-11, 1999.

[7] N. Jason, V. Chris, and Z. Hua, "Virtual-time round-robin: an O(1) proportional share scheduler", in *Proc. USENIX Ann. Tech. Conf.*, Boston, USA, 2001.

[8] A. Kadelka, A. Hettich, and S. Dick, "Performance evaluation of the MAC protocol of the ETSI BRAN HIPERLAN/2 standard", in *Proc. Eur. Wirel.*, Munich, Germany, 1999, pp. 157–162.

[9] J. Peetz, "Quality of service support in HIPERLAN/2 multihop ad hoc networks based on forwarding in the frequency domain", Aachen University of Technology, 2001.

[10] J. Rapp, "Increasing throughput and QoS in a HIPERLAN/2 system with co-channel interference", in *IEEE Int. Conf. Netw. ICN*, Colmar, France, 2001, Part I, pp. 727–736.

[11] B. Walke, N. Esseling, J. Habetha, A. Hettich, A. Kadelka, S. Mangold, J. Peetz, and U. Vornefeld, "IP over wireless mobile ATM—guaranteed wireless QoS by HIPERLAN/2", *Proc. IEEE*, vol. 89, pp. 21–40, 2001.

Józef Woźniak received his M.Sc., Ph.D. and D.Sc. degrees in telecommunications from the Faculty of Electronics, Gdańsk University of Technology (GUT), Poland, in 1971, 1976 and 1991, respectively. In 2001 he became a Professor. Prof. J. Woźniak has rich industrial and scientific experience. In February 1984 he participated in research work at the Vrije Universiteit, Brussel. From December 1986 to March 1987 he was a Visiting Scientists at the Dipartimento di Elettronica, Politecnico di Milano. In both cases he was working on modelling and performance analysis of packet radio networks. In 1988/89 he was a Visiting Professor at the Aalborg University Center, lecturing on computer networks and communication protocols. Prof. Józef Woźniak is author or co-author of more than 170 scientific papers and co-author of four books.

He is also co-editor of 4 conference proceedings and co-author of 4 student textbooks and great number of unpublished scientific reports. His scientific and research interests include network architectures, analysis of communication systems, network security problems, mobility management in WATM as well as LAN and MAN operational schemes together with VLANs analysis.
e-mail: jowoz@pg.gda.pl
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Tomasz Janczak was born in 1975 in Gdańsk, Poland. He received the M.Sc. degree in computer science from Gdańsk University of Technology (GUT) in 1999. Since 1999, he has been working towards the Ph.D. degree at the same university. His research work focuses on wireless local area networks (WLANs), including topics covered by IEEE 802.11, HIPERLAN and Bluetooth solutions. He has published several technical papers on supporting quality of service and service fairness in wireless networks with dynamically changing topologies. In 2001, he received the M.Sc. degree in management and economics from TUG. At present, he works as a system engineer in Intel Corporation, at the R&D networking site located in Gdańsk.
e-mail: janczak@eti.pg.gda.pl
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Przemysław Machan received M.Sc. in computer science (2001) and M.Sc. in information management (2003) degrees from Gdańsk University of Technology (GUT), Poland. Currently, he is studying towards Ph.D. at GUT. His research work includes IP and WLAN mobility, QoS in WLANs and WLAN architectures. He is also working as a software engineer in Intel Corporation, at R&D networking site located in Gdańsk.
e-mail: przemac@thenut.eti.pg.gda.pl
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Wojciech Neubauer was born in 1977 in Gdańsk, Poland. He received the M.Sc. degree in computer science from Gdańsk University of Technology (GUT) in 2001. Since 2001, he has been working towards the Ph.D. degree at the same university. At present, he works as a software engineer in Intel Corporation, at the R&D networking site located in Gdańsk. His research interest focuses on security of wireless local area networks (WLANs).
e-mail: newo@wp.pl
Gdańsk University of Technology
G. Narutowicza st 11/12
80-952 Gdańsk, Poland

Regular paper

An adaptive hidden Markov model for indoor OFDM based wireless systems

Christos V. Verikoukis

Abstract—Detailed physical layer simulation of orthogonal frequency division multiplexing (OFDM) systems requires programs that execute too slowly due to long coherence time of the indoor mobile channel. Evaluation of higher layers of such systems is simplified if suitable models for reproduction of channel errors statistics are available. An adaptive hidden Markov model (HMM) for indoor OFDM based systems that accurately reproduces error statistics of the real system with less computational effort than the exact simulation is presented in this paper. The standard HMM methodology has been modified in order to reproduce the periodicity in the error positions of the OFDM systems. The proposed model is validated by comparison of three statistical parameters: number of errors, length of the errors run and length of the error—free intervals in a frame of bits.

Keywords—hidden Markov model, OFDM, multicarrier transmission.

1. Introduction

Orthogonal frequency division multiplexing (OFDM) is a special case of multicarrier transmission, where a single stream is transmitted over a number of lower rate subcarriers. In this way, several modulated orthogonal carriers with overlapping spectra are transmitted in parallel. The parallel approach has the advantage of spreading out a frequency selective fade over many symbols. It has been proposed as the modulation type for several broadband indoor wireless systems in order to combat frequency selective fading. ETSI-BRAN has selected this type of transmission for HIPERLAN/2 [1] while OFDM has been chosen for the IEEE 802.11a [2] and the recently approved IEEE 802.11g standards in the 5.2 GHz and 2.4 GHz frequency ranges respectively. OFDM proposals have been also presented for the 4G wireless systems [3].

One of the indoor radio channel characteristics is that the impulse response is slowly changing in comparison with system bit rate. Therefore, long time consuming programs are needed to simulate all the possible channel conditions, which make the real time emulation of OFDM systems virtually impossible and hampers simulation of higher layers. One method to overcome this problem is to model the physical layer using a HMM [4].

The HMMs are useful tools for modelling stochastic random processes that are general enough to approximate various statistical data, which explain their popularity in many applications such as speech recognition, image analysis, control theory, biology, communications theory, queuing theory, etc. In several works, channels with memory as well as the error sequences in digital communication systems are modelled using HMMs [5–9].

In an indoor OFDM based wireless communication system, the bit errors appear on the subcarrier where the notch of the frequency selective fading is centered, and, due to the slow varying nature of the indoor channel, the fading affects the same subcarrier for duration of several OFDM symbols. Therefore, a periodicity equal to the number of bits in an OFDM symbol is observed in the position of the errors after demodulation.

In this paper, an adaptive HMM suitable for characterization of indoor OFDM based systems is proposed. The merit of the adapted HMM is ensured by applying the model to the uplink of a typical indoor physical layer based on OFDM. The paper is organized as follows: Section 2 presents HMM basics and Section 3 describes in detail the proposed HMM. In Section 4 a typical OFDM physical layer is analyzed, while Section 5 presents some simulation results. Finally, conclusions are presented in Section 6.

2. Hidden Markov model basics

The HMM is an extension of the Markov model concept which includes the case where the observation is a probabilistic function of the state rather than deterministic one. The resultant model is a double stochastic process which is not observable. An HMM can be characterised by the following parameters [4]:

- **N**, the number of states in the model. The individual states are defined as $S = \{S_1, S_2, \ldots, S_N\}$ and the state at instant t as q_t.
- **D**, the number of the different observation symbols per state. The observation symbols correspond to the physical output of the system being modelled. The individual symbols are defined as $V = \{v_1, v_2, \ldots, v_D\}$.
- $\mathbf{A} = \{a_{ij} 1 \leq i, j \leq N\}$, the state transition probability matrix, where

$$a_{ij} = P\left[q_{t+1} = S_j \,|\, q_t = S_i\right]. \tag{1}$$

- $\mathbf{B} = \{b_j(k)\ 1 \le j \le N,\ 1 \le k \le D\}$, the observation symbol probability matrix, where

$$b_j(k) = P\left[v_k \text{ at } t \,|\, q_t = S_j\right]. \tag{2}$$

- The initial state distribution vector $\pi = \{\pi_i\}$, $1 \le i \le N$, where π_i is the probability for the initial state to be S_i, that is

$$\pi_i = P\left[q_1 = S_i\right]. \tag{3}$$

Giving appropriate values to the parameters N, D, A, B, and π, one can use the HMM as a generator to produce the observation sequence $O = \{O_1, O_2, \ldots, O_T\}$, where O_T is one of the symbols of V, and T is the number of observations in the sequence. From the above it is concluded that an HMM requires both model parameters (N and D), the observed symbols and three probability measures (A, B and π) to be specified. For convenience, the following notation is used to indicate the complete set of model parameters:

$$\lambda = (\mathrm{A},\ \mathrm{B},\ \pi).$$

3. The description HMM adopted

In order to use a HMM in communication systems it is necessary to divide the received data into constant length packets, where each one is characterised by the number of errors it contains. The HMM reproduces the correlation between consecutive packets by forcing the current state to depend on the state of the previous packet. In our context, a reasonable choice is to identify a packet of the HMM with a coded packet plus medium access control (MAC) headers.

A HMM consists of a Markov chain that has a number of states, each state representing a range of number of errors within the packet. Choosing the correct number of states is of critical importance to obtain an accurate model. A small number of states obligates the HMM to group together very distinct observations, while with too many states the training simulation may not provide enough distinct events to estimate the parameters of each state correctly. By using the statistical information from exact physical layer off-line simulations for a specified number of states, the range of numbers of errors in every state and the probability of each state are tentatively defined. It is noted that the states should be as equiprobable as possible. Next, also by the exact simulation, the "transition matrix" filled with the probabilities of passing from one state to another is obtained. This matrix is used as the input to the HMM program that generates a sequence of errors statistically similar to the real one.

When running the HMM, the exact number of errors to place in a received packet is determined by sampling a random variable with a mean equal to the mean number of errors of the current state. In a conventional HMM, the errors are distributed inside the packet with uniform statistics [10]. This is not adequate for an indoor OFDM based physical layer because, as already mentioned, the errors appear in periodic clusters. Our method includes obtaining, from the training simulation, the probability density function (p.d.f.) of the error positions in the real packets, conditioned by the fact that the notch of fading is centered on one of the subcarriers. This is repeated for every subcarrier. Then, when running the model, the errors are distributed inside the packet, in the adequate carrier, according to the previously stored p.d.f.'s.

To check the performance of the proposed model, some validation tests are carried out, consisting of analysing the errors introduced by the HMM within the packets and comparing them with the training simulation. The validation is based on the comparison of p.d.f. of the following three statistical parameters:

- number of errors in a frame of bits,
- length of error runs in a frame of bits,
- length of error—free intervals in a frame of bits.

In case when large deviation in the validation test is obtained after choosing the number of states N, the initial number of states is increased and the procedure is repeated.

Once the given physical layer is well trained, the statistical information generated by the off-line simulation is no longer needed. Therefore, the HMM parameters are sufficient to reproduce the error distribution of the physical layer in future cases.

4. Physical layer description

A typical indoor physical layer based on OFDM is used in the simulations (Fig. 1). It provides a 20 Mbit/s (uncoded) wireless link occupying a bandwidth of 25 MHz and a carrier frequency of 5.2 GHz. OFDM with 16 subcarriers and QPSK modulation on each one has been selected. An inverse fast Fourrier transformation (IFFT) and an oversampling factor of 4 are used to generate the time domain samples transmitted during one OFDM symbol. Oversampling is necessary to avoid aliasing in the generated signal spectrum.

A cyclic prefix is added by copying some samples from the end of each OFDM symbol to the beginning in order to protect the symbol from the echoes of the channel. The use of a time domain raised cosine (roll-off = 0.5) windowing of each OFDM symbol is necessary to reduce the adjacent channel interference. The raised cosine window lasts for 15.6% of the OFDM symbol at the beginning and at the end of each symbol. Thus the total symbol is made up of the windowing time, a FFT period of 1.28 μs and a 160 ns guard time (cyclic period). After modulation the signal is up-converted to the RF channel band. Before being transmitted, the signal passes through a nonlinear amplifier. To limit the distortion due to the power amplifier without reducing too much its efficiency, a 3 dB back-off is applied.

A three-ray, time-varying radio channel with a delay spread of 100 ns in a picocellular environment has been chosen for

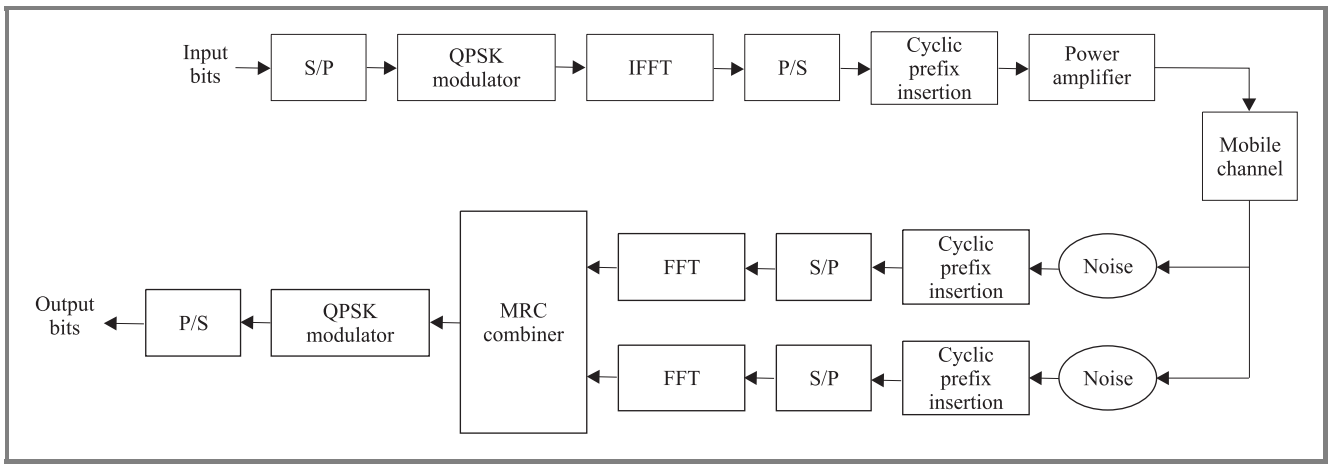


Fig. 1. Block diagram of a typical OFDM physical layer.

simulation while, complex-valued additive white Gaussian noise (AWGN) model is assumed.

At the receiving end, the signal is demodulated coherently and the bits are detected using minimum distance criterion. To estimate the transfer function of the channel a pilot sequence of 10 known OFDM symbols are transmitted with ideal back-off once every 1000 OFDM symbols. Since the indoor channel is slowly time variant, the pilot sequence is known a priory by the receiver. In this way, two antennas spatial diversity with two independent receivers whose outputs are combined using maximal ratio combining (MRC) is applied while, ideal time and frequency synchronism are assumed.

5. Simulation results

Off-line simulations of the physical layer for the up-link assuming mobile terminal speed of 1 m/s have been carried out. In order to take all possible values of the channel, the simulation time must be 100 times the coherence time. This means that at least $3 \cdot 10^6$ FFT blocks are simulated for every E_b/N_0 value.

Figure 2 shows the mean bit-error-rate (BER) versus the mean E_b/N_0 values. The simulation with linear amplifier and ideal channel estimation and the theoretical curve are compared in order to demonstrate the correct operation of the simulation tool.

Validation results of the HMM for a packet length of 512 bits are presented in this section. Every packet contains the data payload, the error correction and detection bits and all necessary medium access control headers.

Validation results for a mean E_b/N_0 value of 10 dB, corresponding to a mean BER value of $6 \cdot 10^{-2}$, are analysed first. Figure 3 allows to compare the length of the error-free intervals obtained by a HMM with 16 states and the real system. It can be seen how the proposed HMM correctly reproduces the periodicity of error positions in the real sequence. In Fig. 4 the p.d.f. of the number of errors, for different numbers of states in the HMM is shown. It

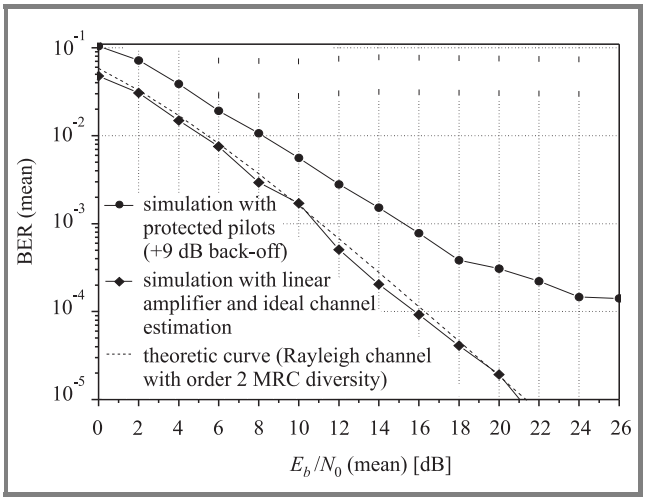


Fig. 2. Mean E_b/N_0 versus BER characteristics.

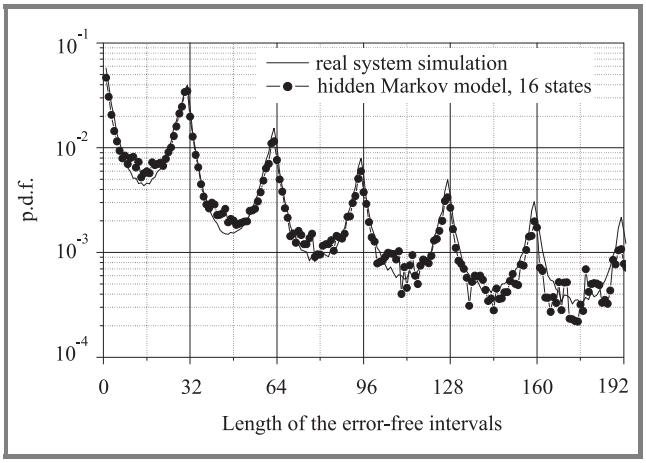


Fig. 3. Probability density function (p.d.f.) of the length of the error-free intervals for a mean E_b/N_0 of 10 dB.

is obvious that a HMM with 8 states gives better accuracy than one with 4 states only. Increasing the number of states to 16 does not give significant performance improvement.

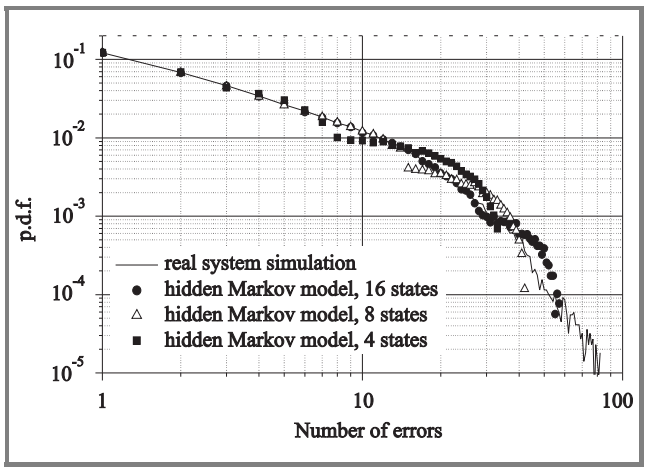


Fig. 4. Probability density function (p.d.f.) of the number of errors for a mean E_b/N_0 of 10 dB.

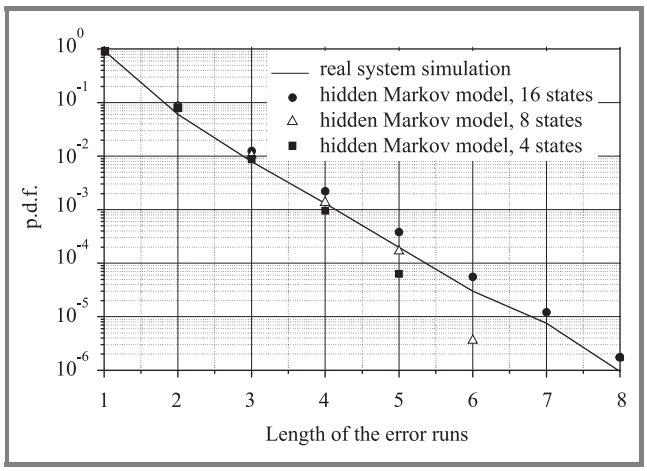


Fig. 5. Probability density function (p.d.f.) of the length of the error runs for a mean E_b/N_0 of 10 dB.

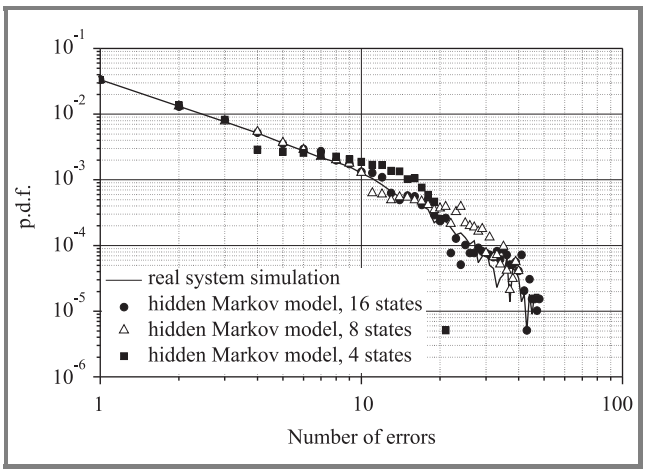


Fig. 6. Probability density function (p.d.f.) of the number of errors for a mean E_b/N_0 of 18 dB.

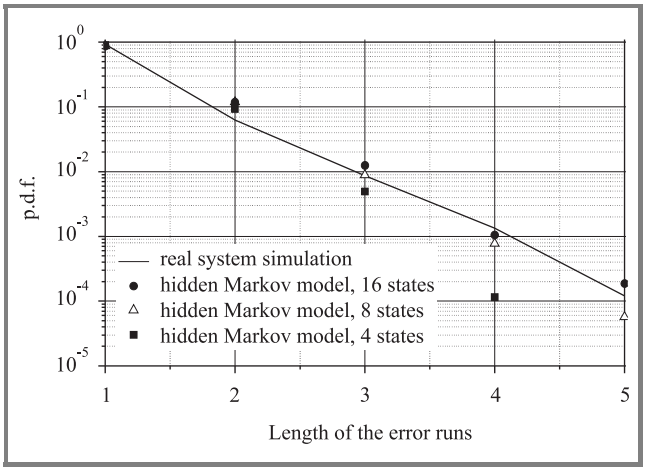


Fig. 7. Probability density function (p.d.f.) of the length of the error runs for a mean E_b/N_0 of 18 dB.

Finally, Fig. 5 compares the p.d.f. of the error runs length. As we can see, 16 states are preferable because the error runs are better reproduced than with 8 or 4 states. Consequently, the use of 16 states is suggested for this E_b/N_0 value.

It has to be noted that when increasing the number of states more parameters have to be saved, increasing simulation time. Thus, it is sometimes better to have a less accurate, but faster HMM. This fact can be seen in Figs. 6 and 7, where the p.d.f. of the error runs and the number of errors in a packet are presented for a mean E_b/N_0 value of 18 dB, corresponding to a mean BER of $3 \cdot 10^{-4}$. It is evident in these figures that using a HMM with 16 states gives no significant improvement of HMM accuracy. Thus, use of 8 states is preferable because simulation runs faster.

It is concluded from the above results that the proposed HMM reproduces the error distribution of the examined OFDM based physical layer with a good accuracy. In case of any modification to the physical layer parameters or the indoor propagation channel model the procedure described in Section 3 must be repeated, producing another set of HMM parameters. Therefore, a database of HMM parameters for the up and downlink for different radio channel models can be created. This database could be the basis for evaluation of an OFDM based real time emulator.

6. Conclusion

In this paper an adaptive HMM for indoor OFDM based physical layers have been presented. In such physical layers the fading is centered on one of the subcarriers. Due to slow nature of the indoor channel the fading remains on the same carrier for several OFDM blocks, producing a periodicity in the errors position. The merit of the proposed model is its ability to reproduce this periodicity easily, saving only some parameters.

The efficiency of the HMM has been verified by applying it to a typical indoor OFDM system. Off-line simulations

for the uplink have been carried out in order to obtain the statistics of the real sequences.

In order to train the HMM, three statistical parameters have been analyzed; the probability of the number of errors, the length of the errors run and the length of the error free intervals in a packet. From the results presented it can be concluded that the proposed HMM can approach the error distribution of an indoor OFDM physical layer with good accuracy.

Therefore, the proposed HMM is a method that reduces substantially the computational effort and allows real time emulation of indoor OFDM systems and fast simulation of higher layers protocols and algorithms on a realistic physical layer. As a consequence, a good model capable to characterize OFDM based indoor wireless communication systems can be obtained.

References

[1] J. Khun-Jush, P. Schramm, U. Wachsmann, and F. Wenger, "Structure and performance of the HIPERLAN/2 physical layer", in *Proc. IEEE Veh. Technol. Conf. Fall'99*, Amsterdam, Netherlands, 1999, pp. 2667–2671.

[2] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High Speed Physical Layer in the 5 GHz band", IEEE Std 802.11a/D7.0-1999.

[3] B. Classon, K. Blankenship, and V. Desai, "Channel coding for 4G system with adaptive modulation and coding", *IEEE Wirel. Commun.*, vol. 9, no. 2, pp. 8–13, 2002.

[4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[5] J. García-Frías and P. M. Crespo, "Hidden Markov models for burst error characterisation in indoor radio channels", *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 1006–1020, 1997.

[6] W. Turin and R. van Nobelen, "Hidden Markov modelling of flat fading channels", *IEEE J. Selec. Area Commun.*, vol. 16, no. 9, pp. 1809–1817, 1998.

[7] N. Nefedov, "Discrete channel models for wireless data services", in *Proc. IEEE Veh. Technol. Conf.'98*, Ottawa, Canada, 1998, pp. 683–687.

[8] J. J. Olmos, P. Diaz, O. Sallent, F. Casadevall, and I. Berberana, "Hidden Markov models used for modelling the physical layers of third generation radio interfaces", in *Proc. ACTS Summit 1997*, Aalborg, Denmark, 1997, pp. 138–143.

[9] S. Sivaprakasam and K. S. Shanmugan, "An equivalent Markov model for burst errors in digital channels", *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 282–286, 1995.

[10] C. M. Dearlove, "A channel error model for the simulation of error control coding in a digital communications system", RACE-II RMTP/RB/L177, MEC025, June 1991.

Christos V. Verikoukis received the M.Sc. in telecommunications from the Aristotle University of Thessaloniki (A.U.Th.), Greece, in 1998. He got the Ph.D. degree in telecommunications from the Universidad Politécnica de Cataluña (U.P.C.), Barcelona, Spain, in 2000. He has collaborated with the radio communications laboratory of A.U.Th. since 1995 and with the radio communications group of the U.P.C. since 1997. He is currently a member of the research staff of the Southeastern Europe Telecommunications and Informatics Research Institute, Greece, and research associate in the Telecommunications Technological Centre of Catalonia, Barcelona, Spain. His research interest focuses on wireless broadband communications systems; he works on multicarrier schemes, HMM, MAC protocols and scheduling algorithms.

e-mail: cveri@inatelecom.org
Southeastern Europe Telecommunications
and Informatics Research Institute (INA)
9th klm Thessaloniki—Thermi
P.O. Box 60080, 570 01-Thessaloniki, Greece

e-mail: cveri@cttc.es
Centre Tecnològic de Telecomunicacions
de Catalunya (CTTC)
c/ Gran Capità 2-4
08034, Barcelona, Spain

Regular paper

Testing of interworking between network terminals with FSK receivers and public exchanges providing display and related services

Wojciech Michalski

Abstract—Implementation of new services in the network requires appropriate methods and tools for checking correctness of interworking between terminals and exchanges. In this article the methodology and test procedures prepared for testing terminals handling FSK protocol transmitted over the local loop for display and related services was described. Methodology presented here is based on ETSI standards. Test procedures were developed for services offered in Polish network. Tests cover all levels of FSK protocol. For two lower layers, separate procedures for "on-hook" and "off-hook" loop states were prepared. The procedure for "on-hook" state contains tests related to data transmission "associated" and "not associated with ringing". These procedures cover normal cases with parameter values and sequence elements complying with standards and exceptional procedures with the extreme values of the parameters and with modified elements of the sequences.

Keywords—testing methodology, display and related services, CLIP service, MWI service, SMS service, UBS solution, NBS solution, FSK protocol, Call Setup message, Message Waiting Indicator message, transmission associated with ringing, transmission not associated with ringing, data transmission prior to ringing, on-hook data transmission, off-hook data transmission, physical layer, data link layer, presentation layer.

1. Introduction

In order to increase revenues, operators have to look for new strategies. One element of these strategies is providing new services, e.g., for analog subscribers. Generally, new services appear as result of development of new technologies for transfer of information. The scope of the services dedicated to analogue subscribers may be extended through use of frequency shift keying (FSK) protocol over analogue subscriber line.

Following worldwide trends, Polish operators try to adjust their service offers to ones of the most European operators. As result of this, the list of services is modified and new services, better suited to the present and future needs are implemented. One of such services is *Short Message Service* (SMS), which up to now has been available to GSM subscribers only. *Calling Line Identification Presentation during Call Waiting* (CLIP CW) and *Message Waiting Indication* (MWI) supplementary services, being standard services of digital networks are very attractive for analogue subscribers, too.

The present offer of the dominant Polish operator Telekomunikacja Polska SA (TP SA), for analogue subscribers, includes new services, such as: *Calling Line Identification Presentation* (CLIP), *Calling Line Identification Restriction* (CLIR), and services mentioned above: CLIP CW, MWI and SMS, called display and related services.

Introduction of these services gives many important functions for analogue subscribers, offered earlier exclusively to ISDN and GSM subscribers. The display and related services, especially MWI (concerning voice mail service) and SMS (which is very popular in GSM networks), will be probably widely used by analogue subscribers and likely as popular as in GSM networks.

2. New needs for subscriber terminals testing

Introduction of new services stimulates development of new test methods and procedures for evaluating correctness of interworking between terminals and public switches during provision of these services.

Services implemented in Polish network cover most data transmission cases, applied also in many other services utilizing the FSK protocol. Implementation of the CLIP CW service means that test procedures for data transmission in "off-hook" state must be prepared. Introduction of the MWI service into the network causes that the terminals have to be tested for data transmission in a "not associated mode". To assure the required quality of transmission for the short message (SM), the terminals should be tested for data transmission in an "associated mode".

From this reason, development of test method to verify handling FSK protocol, by terminals equipped with FSK receivers, during realization of the CLIP CW, MWI and SMS services was done in 2003 in the Switching Systems Department. It was one of the tasks of the work: *"Development of modern measurement method harmonized with the requirements of European Union, concerning network terminals connected to the PSTN network"*, performed last year at the National Institute of Telecommunications in Warsaw. As a result of this task, both the methodology

and testing procedures were developed. Final results have considerable practical use.
The article presents general description of results obtained.

3. CLIP CW, MWI and SMS services as examples of display and related services implementation

Overall description of the CLIP CW service, including controlling procedures, is presented in the ETSI standards [1–3] and in the document [16]. According to these documents, the *Call Waiting Tone* (CWT), followed by the *Dual Tone Alerting Signal* (DTAS), should be sent to the controlling user, when he is engaged in communication with the second subscriber and the other subscriber is attempting to obtain connection to his telephone number. The DTAS signal informs the terminal of the controlling user that data transmission will be started. The CLIP function shall apply to the line in "off-hook" state. The CLIP CW service is related to the *Call Setup* message. Parameters of this message are defined in [6].
The MWI supplementary service is offered together with the *Call Forwarding to Voice Mail* (CF-VM) service. This service is typically used between a voice mailbox service provider (controlling user) and a user of the voice mailbox service (receiving user). The MWI service enables the network, upon the request of controlling user, to indicate to the receiving user, that there is at least one message waiting in voice mail. The message may be sent in immediate or deferred modes. For analogue subscribers, a visual indicator and an informative message can be displayed on their terminals. Data transmission not associated with ringing is used to support MWI service. The MWI message is used to handle information related to messages in message system. Detailed description of this service is given in document [7]. Parameters of this message are defined in [6].
In accordance with the national requirements, the CLIP CW and MWI services should operate in standard modes described in the above mentioned documents.
The SMS supplementary service enables the originating user to send a SM of limited size to a destination user via short message service center (SM-SC). Messages may be input to the SM-SC by means of a suitable telecommunication service either from the fixed network, e.g., speech, telex, facsimile, etc. or from a mobile network.
The SMS service may operate as user based solution (UBS) or network based solution (NBS). In UBS solution, messages are transported via a SM-SC using normal voice band call through the network using in-band signalling. The UBS solution is supported by Protocol 1 and Protocol 2, specified in [15]. It is a network operator's option to choose which protocol is used. According to Polish requirements, Protocol 1 should be used. In UBS solution the exchange participates in data transmission in a limited scope. Data are transmitted transparently through the network between the terminals (of originating and terminating user) and the SM-SC. The first role of the exchange is to set up the call in order to send SM from originating user and in order to receive SM by terminating user. The second role is to support CLIP function during SM delivery from SM-SC to SM-TE. The NBS solution uses the SMS message, specified in [6].

4. FSK protocol features used for implementation of display and related services

Possibility of implementation of new services in the network depends on whether the terminals and the exchange can serve the FSK protocol in "on-hook" and "off-hook" states and whether they support data transmission (associated and not associated with ringing), in accordance with the requirements for particular service. The fundamental principles of interworking between terminals and public exchanges are described in ETSI standards [4, 9 10]. Full scope of the messages and parameters used in the display and related services is presented in [6]. In the Polish network, only basic parameters related to the CLIP and the MWI services are transmitted currently.
According to the national requirements, the following parameters should be used for CLIP CW service in the *Call Setup* message:

- date and time (M),
- calling line identity (M),
- reason for absence of calling line identity (M),
- calling party name (O),
- reason for absence of calling party name (O),
- called line identity (O),
- first called line identity (O),
- call type (O).

The following notes apply:

1. "M" means mandatory parameter, "O" optional parameter (for use in national network).
2. Generally, the ISUP 1 protocol is implemented in national network (ISUP 2 exist actually only in part of the network), so it is not possible to send the calling party name and reason for absence of calling party name parameters.

In accordance with the national requirements, the following parameters should be used for MWI service in MWI message:

- date and time (M),
- visual indicator (M),

– number of messages (M),
– calling line identity (O),
– reason for absence of calling line identity (O),
– calling party name (O),
– reason for absence of calling party name (O).

Notes listed above are valid also here.

When SMS service is supported by the UBS solution, the most important items for this service are specified in [14] (containing service description) and [15] (describing short message communication between a fixed network short message terminal equipment (SM-TE) and SM-SC). The document [6] is important only for the CLIP function. FSK protocol described there consists of different messages than protocol described in above mentioned standards.

According to the [14], to send and to receive short message a voice band communication path is established in the PSTN/ISDN between SM-TE and SM-SC using basic call control procedures. The SM transfer is split into two steps, the SM transmission (transfer of a SM from the sender to the SM-SC) and SM delivery (transfer of a SM from the SM-SC to the receiver).

In the first step (SM submission), SM-TE establishes a call to the SM-SC to submit the SM to the SM-SC, which acts following the store and forward principle. The network shall provide the caller ID (CLI) of the SM-TE to the SM-SC (SM-SC uses this information to identify the SM-TE). After the voice band connection between SM-TE and SM-SC has been established, the end-to-end SM data transfer phase is entered for short message transfer. After the SM has been transferred, the connection is released.

In the second step (SM delivery), the SM-SC establishes a call to the SM-TE to deliver the SM to this SM-TE. In this case, the network shall provide the CLI of the SM-SC to the SM-TE. The SM-TE uses this CLI information to identify and connect an incoming call from the SM-SC. As in the first step, the short message is transmitted after the voice band connection has been established. After the SM has been transferred, the connection between SM-SC and SM-TE is released.

In case of PSTN access, the CLI function is provided with FSK signalling according to documents [4] and [6], describing the end-to-end interworking and the protocol between SM-TE and the exchange. Than, from the exchange point of view, the FSK protocol is the same for SMS and CLIP services.

In accordance with the national requirements, the end-to-end interworking between SM-TE and SM-SC should be provided with the FSK protocol according to [14] and [15] and between SM-TE and the public exchange according to [6]. It is important, that the national requirements comprise only end-to-end interworking between SM-TE and the exchange. Subjects concerning interworking between SM-TE and SM-SC are beyond scope of these requirements.

5. Principles of testing of terminals equipped in FSK receivers

Till now, international standard bodies have not published appropriate documents containing detailed test procedures, which allow to test the display and related services (in particular CLIP CW, MWI and SMS services).

The testing process is currently covered by ETSI standards [11–13]. These documents contain some indications of the organization and design of tests, but do not include explicit requirements for testing. Although current versions of these documents do not comprise all information needed for testing, they are nevertheless important reference points for methodology.

Document [11] provides the PICS proforma for the subscriber line protocol for support of PSTN display services at local exchange in "on-hook" and "off-hook" states. The first state is defined in [6] and [9] in compliance with the relevant requirements and in accordance with the relevant guidance in ISO/IEC 9646-7[1]. The second state is defined in [6] and [10] in compliance with the relevant requirements and in accordance with the relevant guidance in ISO/IEC 9646-7[1]. It is a document, in form of a questionnaire, which should be fulfilled by product supplier. The PICS confirm conformance to a given protocol specification.

The standard [12] specifies the TSS&TP for both the "on-hook" and the "off-hook" data transmission over PSTN access for terminal equipment. In order to stay aligned with structure of the base standards, this document specifies test purposes for FSK protocol. This document contains items related to the naming convention, structure of tests, test strategy and principles of test design and execution. It comprises also items concerning general principles of testing particular layers of FSK protocol. The document does not cover interaction with other supplementary services.

According to this standard, test specifications should be divided into three parts applicable to three FSK protocol layers. The physical layer and data link layer signals should be generated in "on-hook" and "off-hook" states. The presentation layer tests should comprise sequences with both correct and incorrect elements.

In scope of the test's design, the standard [12] specifies the naming convention and structure of tests. According to the convention, test name consists of the following elements: name of layer, name of service, group number and sequential number. Groups are organized according to the TSS and sequential number starts with "001", within each group. The structure of a single test consists of the following general elements: header, stimulus (e.g., pre-test conditions), reaction (action, conditions) and message structure (message containing parameters).

[1] "Information technology—Open Systems Interconnection—Conformance testing methodology and framework—Part 7: Implementation Conformance Statements", ISO/IEC 9646-7:1995.

Test strategy of test design and execution should be based on assumptions, that:

- the tests should check the correctness of transfer of each FSK protocol element,
- all messages should contain at least the mandatory parameters and the parameters should have correct values,
- neither message nor a parameter, which can lead to a "fail" or "inconclusive" verdict, should be used.

To ensure the correct reception of the message by terminal equipment (TE), the test operator should observe the TE after test execution.
Indications and the notes concerning design and execution of the tests included in standard [12] inform, that the conformance of lower layers (the physical (PH) and data link layer (DL)) of the terminal equipment (TE) under test should be diagnosed by either proper reception or no reception of messages by the TE at the presentation layer, sent through the PH and DL layers. Absence of reception of a message may result from:

- no support of the implementation under test for that particular service or parameter,
- non-conformance of the physical layer of the TE.

In case of reception of a valid message, through a valid DL layer, the implementation under test shall activate the corresponding indicators. This assumes the proper reception of the physical signal. The test operator will evaluate the correct reception of the message, and consequently of the physical signal, by observing the reaction of the TE. The TE should react to message reception by activating indicators (e.g., a LED) or displaying the received information (for example calling line ID).
Standard [12] contains also the test purposes (TP) in outline form, which are intended to check that particular layers of the FSK protocol are correct. The test purpose consists of elements like: name of the test, references to the base standards and expected result. Each layer is covered by one group of test cases.
Standard [13] describes abstract test method and specifies the PIXIT for both the "on-hook" and "off-hook" data transmission over PSTN access for terminal equipment. Based on the abstract test method, different types of abstract service primitives (ASP) are presented, which enable to send or receive a protocol data unit (PDU), using parameters transmitted in ASPs. The structure of the ASPs fit the type of PDUs or signals to be send or received. Some ASPs contain a duration parameter. This means that by sending this ASP the corresponding signal is maintained within this duration. In the test case, the next event can only start after the signal is completely sent, i.e., at the end of its duration. Generally, the document mentioned illustrates the particular behaviour during creation of the signals (for physical layer), messages for data link and presentation layer and the full sequences, using the FSK signal features as defined by corresponding parameters (i.e., mark and space frequency, level and noise).

6. Methodology of subscriber terminal testing

The methodology was developed after analysis of ETSI standards: [6, 9–12] and Polish national requirements [17] and [18].
According to these documents, methodology contains descriptions of principles for testing each service, test configurations and selected instruments used for testing. Because CLIP CW, MWI and SMS services operate in different environments (transmission modes, and loop states), it is assumed that lower layers tests will be dedicated to each layer, each transmission mode and each loop state, as separated test procedures. For tests of presentation layer separated procedures, dedicated to each service, will be used.
The testing procedure of the CLIP CW service comprises group of tests in "off-hook" state. It consists of tests specified in standard [12]. These tests should be used to check the following items:

- response of the terminal to receipt of *Call Setup* message containing correctly and incorrectly coded mandatory and optional parameters,
- response of the terminal to receipt of extremely valued signals,
- timing functions concerning signals transmission in the subscriber loop.

This procedure includes also many additional tests, not specified in standard [12], which give the possibility to check the following items:

- response of the terminal to receipt of incorrect sequences containing incorrect codes of the signal transmitted before data transmission has started (*Mark Signal*),
- response of the terminal to receipt of incorrect elements of sequences sending in presentation layer, which are not specified in [12],
- response of the terminal to the loop state change (from busy to idle state),
- electrical parameters of the terminal equipment acknowledgement (TE-ACK) signal.

Test procedure for MWI service contains group of tests executed in idle state. The testing process comprises checking data transmission not associated with ringing. It consists of tests containing the MWI message with the visual indicator parameter and other mandatory and optional parameters.

These tests allow to check the response of the terminal to receipt of:

- MWI message containing correctly and incorrectly coded mandatory and optional parameters,
- extremely valued signals and timing signals for transmission in the loop.

The procedure contains standard tests which are developed in relation to tests specified in [12]. This procedure contains also non-standard tests, e.g., additional tests being out of the scope of the above mentioned standard and tests related to national requirements. Additional tests may be used to check the following items:

- response of the terminal to receipt of incorrect sequences containing incorrect codes of the signals transmitted before data transmission has started (channel seizure signal and mark signal),
- response of the terminal to receipt of incorrect elements of sequences transmitted in presentation layer, which are not specified in [12],
- response of the terminal to the loop state change (from idle to busy state).

The presentation layer testing procedure, concerning SMS service, is completely different from the proposal presented in standard [12] for this service. According to the [14], the SMS can be implemented in two ways, either as a NBS or as a UBS.

In the NBS solution a supplementary service is offered as a part of a function within the public network. In UBS solution the service is offered as a part of a function within end user equipment, which does not require any specific short message function inside the public network.

The document [12] assumes that the SMS service operates as a NBS solution and the SM message is used. Procedure developed contains tests concerning only the UBS solution, because this application will be used (according to the [14] and [15]) in the Polish network. In accordance with these documents, during realization of SMS service, except for the interworking between a TE and an exchange, direct interworking between TE (calling and called) and SM-SC and different messages (than in the UBS solution) is needed.

In the UBS solution, the outgoing message from the originating TE shall be sent to the SM-SC and shall contain the address of the receiver user. The incoming message from SM-SC to the terminating TE shall include the CLI function.

Detailed specifications related to TE and SM-SC interworking are out of scope of the national requirements. In this situation, the SMS testing procedure contains only tests related to CLI function and the methodology of SMS testing covers part of methodology of the CLIP CW service.

7. Test procedures

Test procedures were prepared on the base of methodology and standards concerning display and relating services, FSK protocol and testing of terminals. Tests comply with national requirements for services and FSK protocol. Test documentation consists of group of detailed tests, which extend the cases specified in document [12] and of group of additional tests not specified in this document.

In accordance with the principles specified in [12] each layer has separate testing procedures. Moreover, separated procedures in "on-hook" and "off-hook" loop state for two lower layers have been prepared. The procedure for "on-hook" state contains tests related to data transmission used in Polish network (associated and not associated with ringing). These are procedures concerning normal cases in which parameter values and sequence elements comply with requirements. These are also the exceptional procedures concerning cases with the extreme values of the parameters and with modified elements of the sequences (not complying with requirements).

Physical layer testing procedure contains tests dedicated to check the response of the terminal to receipt of Ringing Pulse Alerting Signal (RPAS), DTAS and to verify timers concerning data transmission (T_2, T_3 for idle state and T_U, T_F for busy state).

Data link layer testing procedure contains tests dedicated to check the function of recognition of signals transmitted before data transmission and the message codes. This procedure allows to check the response of the terminal to receipt of correct and incorrect sequences of FSK protocol, in particular in case of the signal transmitted before data transmission is started (channel seizure signal and mark signal) and messages with incorrect codes.

Presentation layer testing procedure contains tests dedicated to check the function of displaying the mandatory and optional parameters concerning realization of each display service. These tests consist of:

- one or two mandatory parameters,
- one mandatory and one optional parameter,
- all mandatory and all optional parameters not excluding one another.

The above mentioned procedures give possibility to check the response of the terminal to receipt of incorrect sequences of FSK protocol, in particular the message with unknown parameter, without parameter, with two equal parameters, two parameters excluding one another and others.

Test documentation was prepared according to recommendations described in [12] but it comprises wider scope of tests than this standard and tests are also more detailed. Certain tests were modified according to the national requirements.

Table 1
Normal procedure—subscriber line in busy state

IDENTIFIER: PHY_02_001	
TITLE: Receipt of a DTAS signal	
SUBTITLE: Receipt of a DTAS signal, return of a valid TE-ACK signal and mute voice path in T_A	
REFERENCE: ETSI ES 200 778-2 p. 4.3.2 and ETSI ES 200 778-3 PICS: MC.4	
PURPOSE: Verification of a DTAS signal recognition within T_A	
PRE-TEST CONFIGURATIONS: Subscriber line in busy state	
CONFIGURATION: Fig. 2	
EXPECTED TEST SEQUENCE: Simulator ⟷ Network terminal DTAS signal ⟶ Timer T_A ⟵ TE-ACK Timer T_A expired	
	TEST DESCRIPTION:
1	Set $T_A = 85$ ms in simulator's data base
2	Make a call (handset off-hook)
3	Send a DTAS signal with nominal values
4	Check that the terminal receiving a DTAS signal correctly, mutes the voice path and returns a valid TE-ACK signal within T_A
EXPECTED RESULTS: Correct reception of a DTAS signal Muting of the voice path	

8. Example of the test description

Here is a sample of detailed test description included in our set procedures. Full set includes about 100 detailed test descriptions. An example of the physical level test description concerning receipt of a DTAS signal is presented in Table 1.

9. Conclusion

Testing methodology and test procedures were prepared, based on the most recent ETSI standards. This guarantees that NIT's solution is true, fair and in compliance with UE requirements. The detailed tests, developed according to the assumptions given in the methodology and in [12], allow to check wide scope of functional and electrical parameters of terminals handling display and related services. Physical and data link layer tests (as lower layer tests, common to all services based on FSK protocol) may also be used to test another services. Based on this methodology, presentation layer test procedures may be easily extended in order to test wider group of services. This methodology may be used for testing of subscriber terminals in various phases of implementation and operation in public network. Tests give the possibility to estimate, in wide scope, the conformity of the testing implementation to the national requirements and European standards.

Tests developed may be executed using commercially available test equipment.

It should be underlined, that elaborated tests and methodology were based on experience gained during testing of telecommunication services and other signalling systems and any other tests have not been used for testing of FSK protocol.

The methodology was presented on 19th November 2003 during AT-F Working Group meeting in Sophia Antipolis.

Methodology described in this paper will be used at National Institute of Telecommunications (NIT) in the near future in order to perform extended scope of tests in NIT's Laboratory.

References

[1] "Integrated Services Digital Network (ISDN); Call Waiting (CW) supplementary service; Service description", ETS 300 056 (October 1991).

[2] "Public Switched Telephone Network (PSTN); Calling Line Identification Presentation supplementary service (CLIP); Service description", ETS 300 648 (March 1997).

[3] "Public Switched Telephone Network (PSTN); Calling Line Identification Restriction supplementary service (CLIR); Service description", ETS 300 649 (March 1997).

[4] "Access and Terminals (AT); Analogue acess to the Public Switched Telephone Network (PSTN); Subscriber line protocol over the local loop for display (and related) services; Part 1: On-hook data transmission", ETSI EN 300 659-1 V1.3.1 (2001-01).

[5] "Access and Terminals (AT); Analogue acess to the Public Switched Telephone Network (PSTN); Subscriber line protocol over the local loop for display (and related) services; Part 2: Off-hook data transmission", ETSI EN 300 659-2 V1.3.1 (2001-01).

[6] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Subscriber line protocol over the local loop for display (and related) services; Part 3: Data link message and parameter codings", ETSI EN 300 659-3 V1.3.1 (2001-01).

[7] "Integrated Services Digital Network (ISDN); Message Waiting Indication (MWI) supplementary service; Service description", ETSI EN 300 650 V1.2.1 (2001-05).

[8] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Subscriber line protocol over the local loop for display (and related) services; Part 3: Data link message and parameter codings", ETSI TS 100 659-3 V1.1.1 (2001-11) (corrections needed to EN 300 659-3 V1.3.1).

[9] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Protocol over the local loop for display and related services; Terminal equipment requirements; Part 1: On-hook data transmission", ETSI ES 200 778-1 V1.2.2 (2002-11).

[10] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Protocol over the local loop for display and related services; Terminal equipment requirements; Part 2: Off-hook data transmission", ETSI ES 200 778-2 V1.2.2 (2002-11).

[11] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Protocol over the local loop for display and related services; Terminal equipment requirements; Part 3: Protocol Implementation Conformance Statement (PICS) proforma specification; On-hook and Off-hook", ETSI ES 200 778-3 V1.1.2 (2002-11).

[12] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Protocol over the local loop for display and related services; Terminal equipment requirements; Part 4: Test Suite Structure and Test Purposes (TSS&TP); On-hook and Off-hook", ETSI ES 200 778-4 V1.1.2 (2002-11).

[13] "Access and Terminals (AT); Analogue access to the Public Switched Telephone Network (PSTN); Protocol over the local loop for display and related services; Terminal equipment requirements; Part 5: Abstract Test Suite (ATS) and partial Protocol Implementation eXtra Information for Testing (PIXIT) proforma specification for the user; On-hook and Off-hook", ETSI ES 200 778-5 V1.1.2 (2002-11).

[14] "Services and Protocols for Advanced Networks (SPAN); Short Message Service (SMS) for PSTN/ISDN; Service description", ETSI ES 201 986 V1.1.2 (2002-01).

[15] "Access and Terminals (AT); Short Message Service (SMS) for PSTN/ISDN; Short Message Communication between a fixed network Short Message Terminal Equipment and a Short Message Service Centre", ETSI ES 201 912 V1.1.1 (2002-01).

[16] "Handbook on services and facilities offered to subscriber in modern systems; Section I & II: Services and Facilities within the Public Network", European Conference of Postal and Telecommunication Administrations (CEPT), 1992.

[17] "Operator's technical requirements for display and related services realization in analogue line", Warszawa: Telekomunikacja Polska SA, 2002.

[18] "Technical requirements for digital switching systems in the Polish national telecommunication network", Warszawa: Instytut Łączności, 1991.

Wojciech Michalski was born in Bogate, in Poland, in 1952. He received the M.Sc. degree in telecommunications engineering from Technical University in Warsaw in 1977. He has been with the Switching Systems Department of National Institute of Telecommunications (NIT) since 1977, currently as a senior specialist. His research interests and work are related to PSTN backbone and access networks, GSM networks and IP networks. He is an autor and co-autor of technical requirements and many documents concerning telecommunication services, FSK protocol, charging and accounting, and network maintenance.

e-mail: W.Michalski@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Regular paper

Labeling of signals in optical networks and its applications

Krzysztof Borzycki

Abstract—The paper is a review and comparative analysis of most common techniques proposed to attach additional data or identification information to digital signals in optical fiber networks by purely optical means. Such "labels" or "headers" can be attached either to continuous bit streams, e.g., in SDH networks or to optical packets. They enable to monitor, route and identify signals in transparent optical networks, especially those with optical wavelength multiplexing, allow management and supervision of remote optical amplifiers and can be used in optical switching systems. Other applications of this relatively unknown technology include monitoring of optical path dispersion, equalization of channels in DWDM systems and detection of intrusion or jamming in highly secure networks.

Keywords—optical fiber transmission, transparent optical network, pilot tone, network management, overhead data channel, optical fiber dispersion, optical label.

1. Introduction

Advanced transport network must enable management, monitoring, switching and protection of all signals. This functionality depends on transmission of management data associated with specific channels, e.g., for channel origin, destination and content identification.
Synchronous digital hierarchy (SDH) and dense wavelength division multiplexing (DWDM) networks provide dedicated data channels for this purpose, implemented as:

- overhead bytes within SDH frame structure, constituting the data communication channels (DCC);
- bytes inside forward error correction (FEC) overhead added to SDH frame, known as digital wrapper, particularly at STM-64 level;
- separate wavelength reserved for management purposes—the optical supervisory channel (OSC).

Each of these solutions provides transmission capacity of at least 2 Mbit/s.
In packet networks, e.g., gigabit Ethernet (GbE), each packet has header with origin, content and routing data. Routers and other equipment read the header prior to packet handling and processing.
Access to management data associated with transmission channel requires O/E conversion and partial demultiplexing—extraction of DCC bytes, separation of packet header, etc. Electronic circuits performing such tasks must operate at line data rates, e.g., 10.7 Gbit/s or 42.7 Gbit/s. Specific circuitry usually accepts only one data rate and signal structure.
While reliable and standardized [1–3], such solutions are fairly expensive and not compatible with "all-optical network" approach, where costly and bandwidth-limiting O/E and E/O conversions and electronic signal processing are avoided, except for the network edge. The ultimate goal is a "transparent" network, where all signals remain in optical domain. Processing functions: amplification, filtering, 2R/3R regeneration, wavelength conversion, quality monitoring and switching shall be implemented with photonic devices only. Transparent optical network shall be able to handle all kinds of traffic with uniform set of features, like optical switching. Unfortunately, several standards of client signals do not support transmission of associated management data for higher order systems. While the 2R transponders used in DWDM and coarse wavelength division multiplexing (CWDM) equipment handle bit streams of any structure, adding channel associated management data requires an "overlay" solution.
Avoiding electronic signal processing is beneficial in terms of flexibility and future upgrades to higher bit rates or different signal types. Equipment cost, failure rates, size and power consumption are dramatically reduced—there are no repeater cards for all channels, with associated installation and maintenance costs. Unfortunately, access to signal overhead is lost.
Auxiliary data channels are necessary to support functions like:

1) network and equipment management;
2) order wire and auxiliary data channels;
3) identification of signal content, origin and destination;
4) connection verification and quality monitoring;
5) protection switching.

Functions (1) and (2) implemented in SDH/DWDM and optical cross-connect (OXC) systems require considerable bandwidth, in order of 2 Mbit/s, but can use shared, dedicated wavelength (OSC). Most DWDM systems transmit OSC at 1510 nm, outside standard C and L bands—1528–1565 nm and 1570–1610 nm, respectively [3]. This allows separation of OSC with low-cost optical filters.
Data associated with functions (3)–(5) shall be attached to each single channel in a transport network based on wavelength division multiplexing (WDM) technology. End-to-end optical path may transit several networks, e.g., originating metropolitan area network (MAN)—national core network—destination MAN, each run by different operator

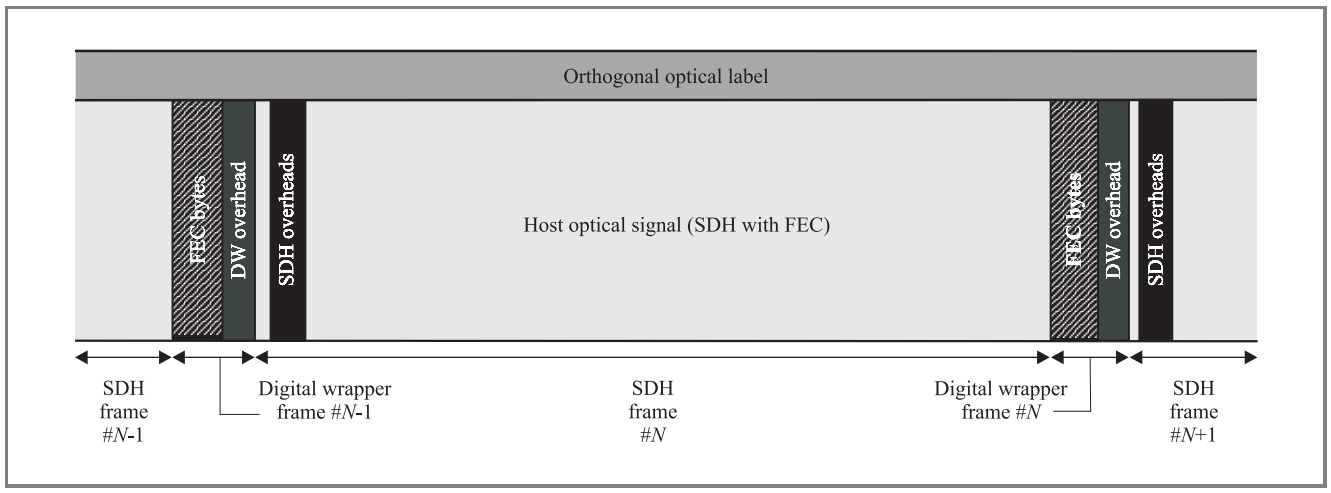


Fig. 1. Overheads provided by optical label, SDH frame and digital wrapper.

with separate management system, and be optically dropped or switched en route. Volume of data transmitted for this purpose is in order of few kbit/s. Capability to add such data to any optical channel regardless of its origin, bit rate and structure is highly desirable.

In a packed-switched network, optical header is always attached to individual data packet and must be read fast in order to avoid delays at optical routers. Otherwise, we need optical buffering of packets with fiber delay lines providing time for header processing, which is expensive and inflexible. Storage of 64-byte and 1024-byte packets at 10 Gbit/s requires approx. 10.5 m and 168 m of single mode fiber, respectively. Fiber sensitivity to bending dictates minimum coil diameter of approx. 50 mm and certain level of mechanical protection. While fiber cost and attenuation are negligible, its splicing and packaging is labor-intensive and expensive. Planar optical waveguides are not suitable for this purpose due to high loss ($\approx$ 100 dB/m) and limited lengths dictated by wafer size.

2. Optical labels

Overhead information associated with optical channel or data packet is known as "optical label". This is a generic term for variety of modulation and multiplexing techniques developed to attach extra information to host optical signal for many applications (see Section 3).

As signal labels need to be accessed in several locations along signal path, process of label readout should not be demanding in terms of hardware required: optical filters and other components, detector bandwidth, decoder complexity, etc. When network carries mixed traffic with different bit rates, line codes and frame structures, labeling shall be "orthogonal"—independent from payload modulation and coding. Label needs to be read without detection and decoding of host signal, otherwise optical labeling does not offer advantages over SDH overhead or digital wrapper. Adding label to signal does not preclude utilization of

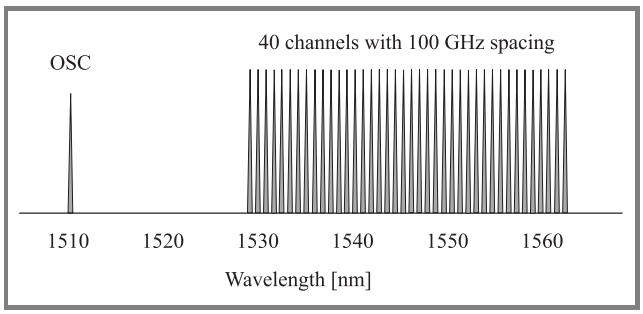


Fig. 2. Typical arrangement of wavelengths in C-band DWDM link.

data channels provided by SDH overhead, digital wrapper, OSC and other means. Figures 1 and 2 provide a comparison.

Figure 1 presents contents of each optical channel (wavelength) indicated in Fig. 2. Size of digital wrapper is smaller than shown in Fig. 1, where exaggeration was needed for clarity. FEC and proper overhead bytes [1] constitute approximately 6.27% and 0.40% of the total data stream, respectively. Most optical labels has bit rates lower than 1/1000th of host signal rate.

Development of optical label technologies is largely stimulated by factors not considered during development of SDH/SONET standards in the 1980s. This includes:

- Development of DWDM, being a dominant transmission technology in core networks today.
- Replacement of electronic signal regeneration with optical amplification.
- Coexistence of different signal formats in transport networks: SDH, synchronous optical network (SONET), GbE, fiber channel, asynchronous transfer mode (ATM), digital video (DV), etc.
- Interconnection of separate networks—preferably by optical means.

- Introduction of optical add-drop multiplexers (OADMs) and optical switching (planned).
- Strong drive to reduce investment and operating costs since 2001.

These conditions reveal limits of SDH—technology providing unparalleled reliability and management features, but developed for networks where signal is fully regenerated at every station, with access to its overhead provided at little extra cost. Many optical transport networks do not use SDH today, especially in the MAN and storage area network (SAN) segment. A need to provide some SDH-like functionality over non-SDH channels exist.

Ideally, optical label should:

– not degrade transmission of host channel, also with optical amplification using erbium doped fiber amplifiers (EDFA), Raman amplifiers or semiconductor optical amplifiers (SOA); acceptable power penalty usually ranges from 0.5 to 2 dB;

– require no modifications to optical components in the network: fibers, amplifiers, filters, etc.;

– be compatible with every type of signal expected in a given network;

– tolerate influence of chromatic dispersion (CD), polarization mode dispersion (PMD), amplifier spontaneous emission (ASE), nonlinear effects, crosstalk, etc., as specified for transmission of host signal, preferably with a power margin of 1–5 dB.

Unfortunately, labels based on subcarrier modulation above payload bandwidth are sensitive to CD and PMD.
Labels with high overhead rates, up to 2.5 Gbit/s on 10 Gbit/s host, tend to introduce significant power penalty and restrict choice of modulation for the host signal, e.g., limit permitted extinction ratio [19]. Such solution is better described as two optically multiplexed channels, jointly designed and optimized.
Certain properties of EDFAs limit choice of label type and parameters:

– amplitude modulation at frequencies below approximately 50 kHz and reuse of frequencies among several WDM channels passing through the same amplifier(s) shall be avoided;

– narrow-band amplitude modulated labels or pilot tones can interfere with EDFA control system;

– nonlinearity of amplifier or fiber, particularly stimulated Raman scattering (SRS) or cross-phase modulation (XPM) may result in transfer of labels between channels.

Labels based on phase modulation and spread spectrum (O-CDMA) techniques are more immune to such phenomena than solutions based on amplitude (intensity) modulation.

A simple label may be formed by specific pattern of unused bits (Section 4.1.5). This pattern creates spectral component in narrow frequency range, picked out by low-cost, narrow-band receiver or is detected by simple pulse sequence correlator implemented with optical delay lines. This technique is akin to frame alignment patterns in digital systems. Improperly selected label patterns may, however, interfere with clock extraction or frame alignment and consequently introduce excessive jitter to host signal.

3. Applications

Known applications for optical signal labels include:

1) optical channel management: supervision, identification by origin, destination or content, quality monitoring, etc.;

2) non-intrusive monitoring and automatic adjustment of WDM link: channel equalization, counting of active channels, automatic gain control, detection of signal failure;

3) optical channel switching: transfer of commands, provision of channel identification data (operator, channel number, etc.), verification of output signal;

4) protection switching: detection of signal failure/degrade, verification of output signal;

5) provision of extra data channel(s) for network operator;

6) packet switching (functions as in packet networks with electronic processing);

7) network security: detection of intrusion, unauthorized signal substitution or jamming.

In most cases, signal label serves one or two purposes only.

In general, one may divide applications into:

1) related to single channel or packet, used for broadly defined routing and supervision;

2) related to operation of network or network element, like optical cross-connect.

Labels of certain type, particularly pilot tones and subcarrier modulation with proper set of frequencies, can be recovered from mix of multiple signals and provide data on their relative power levels. Line signal in this case is tapped from a fiber into single detector and analyzed in frequency domain. This enables supervision and automatic adjustment of DWDM link without optical spectrum analyzer (OSA).
Highly secure networks, e.g., for military, government or banking applications need reliable means to detect optical jamming or insertion of foreign signals instead of legitimate ones. Signal substitution, its processing with inline

electronic device to modify content or injection of powerful jamming signal into optical line amplifier results in erasure of label or significant level deviation. Optical labels, except for modified bit sequences, are removed when signal is regenerated, e.g., when a repeater (without labeling support) is inserted to modify payload or management data. Label absence indicates malicious activity, even when the host signal itself appears unaffected. Labeling scheme for security applications shall remain confidential and be regularly changed to prevent reverse engineering and emulation.

4. Review of labeling techniques

Label technology and content strongly depend on particular application. Label functionality ranges from carriage of simple static message to advanced solutions providing data channel up to 2.5 Gbit/s, enough to support full network management and extra traffic on the same wavelength.
Industry standard have not emerged so far. Several technologies were proposed and discussed within ITU-T Study Group 15 (SG15) since 1997, but remained in unpublished contributions only. ITU-T Recommendation G.709 [1] recognizes the need for optical channel labeling in general, but no specific method is described, recommended or forbidden. Lack of standard prevents commercial introduction despite fairly extensive research work, mostly in Japan, USA, Korea, China and the Netherlands.
Optical labeling is covered by two European Community IST-OPTIMIST projects:

- STOLAS IST-2000-28557: *Switching Technologies for Optically Labeled Signals*. The subject is to develop technologies for very high capacity optical packet switched networks, with: orthogonal optical packet labeling, wavelength conversion, purely optical switching and WDM transport. Labeling methods currently include frequency shift keying (FSK) and differential phase shift keying (DPSK), also with high overhead rates—up to 622 Mbit/s on 10 Gbit/s host channel. The project began in December 2001, with duration of 36 months.

- NEFERTITI IST-2001-32786: *Network of Excellence on Broadband Fiber Radio Techniques and its Integration Technologies*. This project is devoted to development of microwave-frequency optical fiber technologies and equipment, up to 1000 GHz and higher. Optical labeling plays minor role, but microwave frequency subcarrier-modulated labeling and optical channel switching techniques are included. The project began in 2001.

The idea of orthogonal optical labeling is not new. Several line systems belonging to the plesiochronous digital hierarchy (PDH) developed and deployed in the 1980s utilized amplitude modulation of main bit stream to transmit overhead data. Modulation depth and bit rate were not standardized, reaching 1–10% and 2–50 kbit/s, respectively. Payload and overhead were received by single detector and separated by low/high-pass electrical filters after amplification. Such mechanism was a *de facto* industry standard before 1992, competing with solutions based on digital multiplexing and controlled clock jitter. Digital multiplexing of overhead streams was adopted for SDH/SONET in 1988 and use of analog modulation has stopped.

Labeling methods reported in literature include:

1) pilot tones (Section 4.1) usually with amplitude modulation (AM) of host signal (Section 4.1.1);

2) low-frequency subcarrier-modulated data channels; modulation methods of the host signal include AM, FSK, and DPSK (Sections 4.1.2 and 4.1.4);

3) high frequency subcarrier-modulated data channels (AM), located above spectrum occupied by the host signal (Section 4.1.3);

4) optical code division multiplexed (OCDM) data channels—rarely used;

5) insertion of label data into fixed locations within frame of host signal (Section 4.1.5);

6) header added to data packet; can be at lower bit rate (Section 4.2.1);

7) solutions based on wavelength division multiplexing (Section 4.2.2).

Methods (1)–(5) were developed for labeling of continuous bit streams. Methods (6) and (7) are for packet switched networks only; solution (3) is also used. Solution (2) with shallow AM modulation (1–10%) has been included in few ITU-T SG15 contributions since 1997, but never included in any standard.

4.1. Labeling of continuous bit streams

4.1.1. Pilot tones

This the simplest form of labeling: host signal is amplitude-modulated with continuous sine wave of fixed frequency. In a multi-wavelength system, each optical channel is assigned a unique label frequency (Fig. 3). Modulation depth

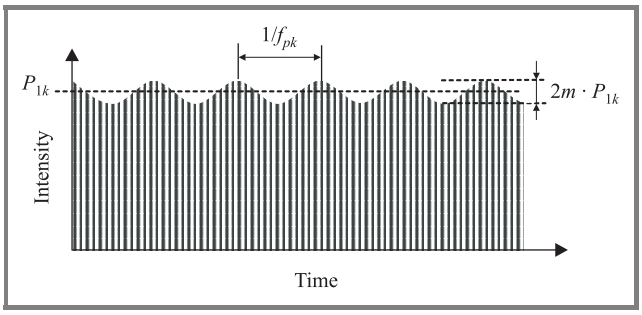


Fig. 3. Amplitude modulation of host digital signal with pilot tone.

is usually limited to 1–5%, keeping resultant power penalty below 0.5 dB and minimizing inter-channel interference during optical amplification.
Label is usually added to host signal at the source, e.g., a transponder inside DWDM terminal. Application of pilot tone to signal in transit, e.g., at switching node is possible by means of optical modulator or amplifier (Section 4.1.1.4).
Labeling is particularly useful for monitoring of WDM networks, where simple and reliable method to identify active channels and measure their relative power levels is needed. Important application for pilot tones is monitoring and equalization of channels in DWDM networks without using costly OSA or DWDM demultiplexer necessary in other methods.

4.1.1.1. DWDM link monitoring: limitations of optical spectrum analysis

As channels sent over WDM link have distinct wavelengths, an obvious and well established solution is to tap small portion of signal, typically 1–3% by means of in-line coupler and use an optical spectrum analyzer to measure spectrum of composite signal. The OSA resolves all channels, measures their wavelengths, relative levels and optical signal to noise ratio (OSNR). Test instrument vendors offer software for automated analysis of WDM spectrum. OSA delivers data for system supervision and adjustment, particularly channel equalization. Unfortunately, fully featured OSAs are expensive, require skilled operators and need periodic calibration. While simplified purpose-built OSA cards for integration with DWDM equipment have appeared, with no moving parts, fixed wavelength range and interface to network management system, hardware cost remains high.

Additionally, optical spectrum analysis cannot establish origin of particular signal or its content. Signal identification needs to be provided by source, e.g., multiplexer. Advances in DWDM technology have reduced channel spacing in installed systems to 0.4 nm (50 GHz); equipment with 0.2 nm (25 GHz) and 0.1 nm (12.5 GHz) spacing is being tested. Typical OSAs have resolution of 0.05–0.10 nm and may not be suitable for testing advanced DWDM networks.

4.1.1.2. Pilot tone principle

Each digital signal in DWDM link is modulated as follows:

$$P_k(t) = P_{1k} \cdot (1 + m_k \sin 2\pi f_{pk} t) \cdot a(t)\,, \qquad (1)$$

where: $P_k(t)$—instantaneous optical power of kth channel; t—time; P_{1k}—average power of kth channel in 1 (ON) state; m_k—modulation index of pilot tone of kth channel; f_{pk}—pilot tone frequency of kth channel; $a(t)$—waveform of digital stream carrying payload information: $a = 1$ for 1 (ON) state of digital signal. For 0 (OFF) state the value of a is ideally a zero. In real systems, value of a in the 0 state is an inverse of signal extinction ratio, ranging from 0.03 to 0.15.

An example of digital transmitter with labeling function is shown in Fig. 4.

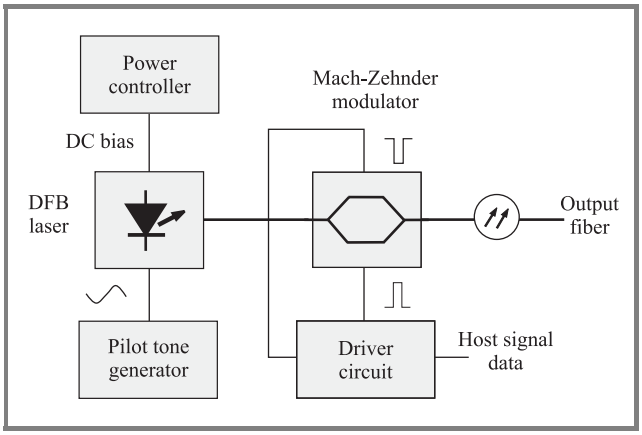


Fig. 4. Transmitter arrangement for adding AM pilot tone.

Monitoring device (Fig. 5) usually comprises optical coupler to tap a small portion, typically about 1% of line signal to narrowband receiver with PIN photodiode, whose output is fed to (electrical) spectrum analysis device. Integrated monitoring modules comprising coupler and InGaAsP photodiode are commercially available.
Following detection by photodiode, amplification and low-pass filtering to remove high-speed digital signal, the resultant voltage for n channels carrying digital signals (NRZ code with 50% mark ratio) is:

$$U(t) = \sum_{k=1}^{n} R(\lambda_k) \cdot G \cdot T \cdot P_k \cdot m_k \ \sin 2\pi f_{pk} t\,, \qquad (2)$$

where: $U(t)$—instantaneous voltage at amplifier output; $R(\lambda_k)$—photodiode sensitivity at wavelength of kth channel; G—amplifier trans-conductance; T—tap ratio of optical coupler used for line or amplifier monitoring; P_k—average power of kth channel.
Over a narrow wavelength range, e.g., 30 nm (whole C-band or L-band), the sensitivity of InGaAs photodiode is constant within ± 0.05 dB and fixed R value can be assumed without degradation of measurement accuracy. Wider range, e.g., in CWDM link or in C+L band system, necessitates correction using spectral sensitivity table.
Formula (2) allows the following general conclusions:

- monitoring works independently of channel spacing;
- each channel is represented by single frequency component in receiver output;
- accuracy of power monitoring depends on equalized, stable pilot modulation index (m);
- any phenomena changing pilot modulation index degrade monitoring accuracy;
- order of pilot frequencies does not necessarily follow order of optical carriers.

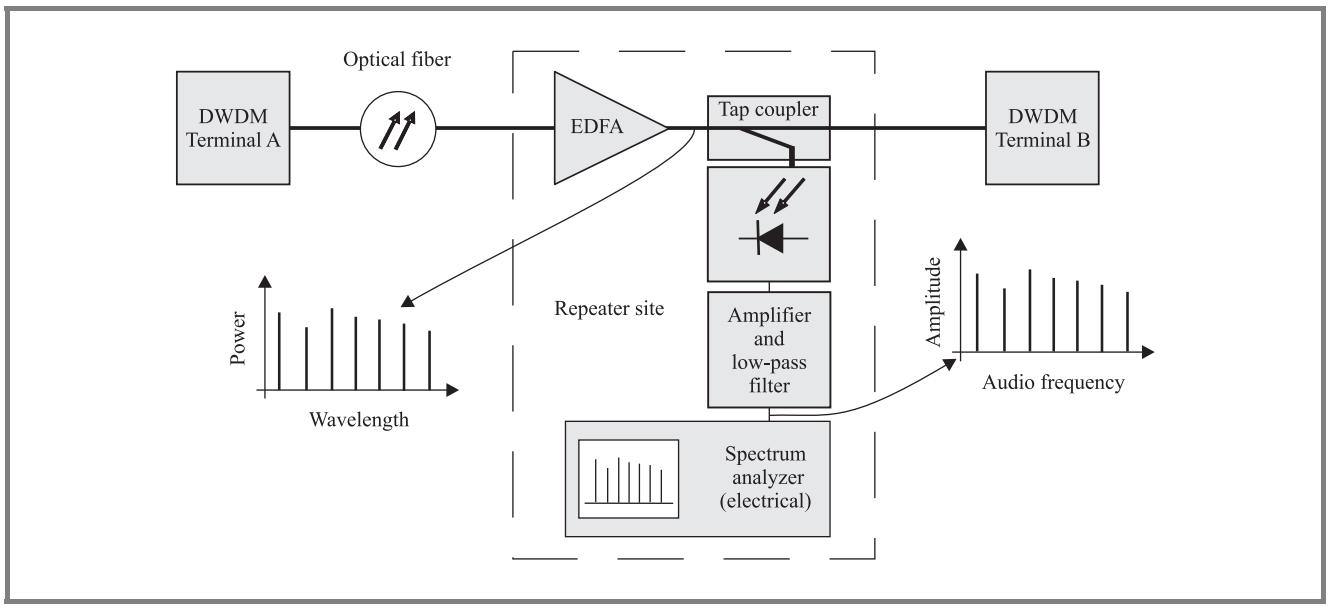


Fig. 5. Pilot tone monitoring in a DWDM link.

Fast Fourier transform (FFT) processing of detected signal yields its spectrum. Peaks correspond to active channels. Measurement data from multiple locations are sent to network management system, being used to adjust channel powers and report system faults.

Selection of pilot frequency is important, because of possible interference with host signal components and the need to minimize negative effects of fiber and amplifier characteristics on pilot tone transmission. In particular, pilot tones shall not coincide with peaks of host signal spectral distribution, to obtain adequate signal-to-noise ratio with limited modulation index.

Choice of label frequencies depends on optical amplifiers used and as type of host signal:

- All telephone-related digital signals (PDH, SONET, SDH) have frames based on 8 kHz sampling, so exhibit peaks of spectral density at multiplies of this frequency. Frequencies like 8 kHz, 16 kHz, 24 kHz, etc., shall be avoided to ensure best label to noise ratio. Other standards, like digital TV with 50 Hz field rate and 16.25 kHz line rate (Europe) impose different restrictions in this respect.

- Systems without optical amplifiers: lowest range, typically 0.1–5 kHz. This range ensures best signal-to-noise ratio due to very low spectral density of virtually all host signals, usually NRZ or RZ-modulated. DWDM transponders do not support bit rates below 100 Mbit/s. Frequencies below 100 Hz are avoided due to mains frequency interference and increased density of receiver shot noise.

- Systems with EDFA amplifiers: 50 kHz–1 MHz due to limited lifetime of Er^{3+} ions in excited state [13, 14]. This phenomenon causes attenuation of signal amplitude components below approx. 50 kHz and increased inter-modulation at low frequencies when signal transits saturated EDFA, resulting in transfer of label (with inverted phase) from one signal to all other passing through the same EDFA. Situation gets worse in long-distance link, as high-pass frequency characteristics and inter-modulation of several EDFAs accumulate. Amplitude-modulated label can also interfere with EDFA power control system, which may interpret low-frequency label as signal power variation and try to compensate for it, resulting in label attenuation and gain instability. Labels above 1 MHz are likely to pass "transparently" through EDFA with any type of power regulation, working both in saturated and unsaturated mode.

- Systems with semiconductor optical amplifiers: modulation index shall be kept low to minimize inter-modulation and undesirable label transfer. These phenomena are fairly independent of label frequency. Considering host signal spectral density, low frequency (≈ 1 kHz) labels are best, provided other factors like noisy power supply do not impose restrictions here.

- Systems with Raman amplifiers: nonlinearity and "competition" for pump power occur in fiber section ≈ 50 km away from pump source, where the pump signal is heavily depleted. Effects on amplitude modulated labels resemble those in SOA. Labels with frequencies below 8 kHz and minimized modulation index should work well.

- Absence of amplifiers gives more freedom, but system designer shall analyze possibility of future upgrades and amplification being added.

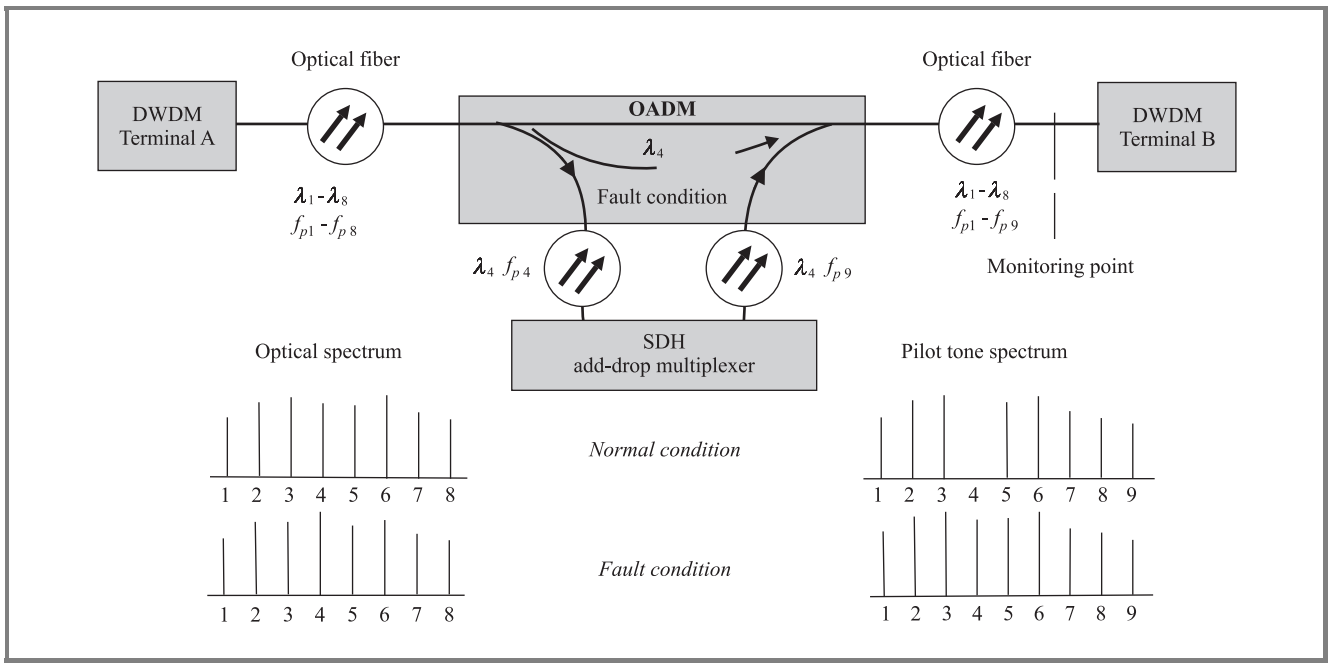


Fig. 6. Collision of two signals at same wavelength and its detection.

- Analog signals, e.g., in optical fiber cable TV networks may require special solutions, utilizing "gaps" in signal spectrum.

Selection of label frequencies is not trivial, because their number in DWDM network may exceed 50. All shall guarantee similar dynamic range of measurements, dictated by signal-to-noise ratio (SNR) and resolution of spectrum analysis system. Most often, a set of equally spaced label frequencies is adopted, e.g., 2.0 kHz, 2.1 kHz ... 5.9 kHz for a 40-channel link. Due to limited finesse (Q) of spectrum analysis devices or filter banks and flat spectral density characteristics of NRZ host signal (SDH, GbE), achievable SNR improves significantly with lower label frequency. Normalized power density of NRZ signal is given by formula:

$$P(f) = \left(\frac{\sin(\pi f / f_{clk})}{\pi f / f_{clk}} \right)^2 , \qquad (3)$$

where: $P(f)$—spectral density of signal power, normalized to 1 at $f = 0$; f—frequency; f_{clk}—clock frequency of digital signal (equal to bit rate).
This density is almost constant for frequencies up to $\approx 30\%$ of f_{clk}.
Assuming fixed Q of pilot tone detection system, resulting ratio of received pilot tone to interference from host signal is inversely proportional to pilot tone frequency. With constant bandwidth of pilot tone receiver, it remains constant for all frequencies of interest (10 Hz to 10 MHz), even with 100 Mbit/s host. Interference will be significantly lower in case of host with RZ modulation, often proposed for 40 Gbit/s networks.
The key advantage of pilot tone method is its simplicity and low cost of associated hardware. Functionality is restricted to monitoring of signal presence and its amplitude. No other information is carried.

4.1.1.3. Limitations of pilot tone method

Pilot tone method has, unfortunately, several limitations and problems:

- does not support transmission of management data;
- wavelength deviations are not detected;
- optical noise or FWM products cannot be measured;
- dynamic range limited to 20–30 dB due to interference from low frequency components of host signal;
- optical amplification can degrade measurement accuracy due to label transfer;
- generation of "ghost tones" due to stimulated Raman scattering (SRS) in long amplified links [9].

Pilot tone monitoring has unique capability to distinguish two or more signals of the same wavelength. This feature is useful for finding faults in optical add-drop multiplexers, networks with wavelength reuse or protection switching devices. Optical spectrum analysis cannot easily detect faults of this nature, as the only indication is increase of signal power, overlooked in system with significant channel level variations (Fig. 6).
Low frequency pilot tones are **immune to effects of chromatic and polarization mode dispersion**.

4.1.1.4. Modulation methods

Amplitude modulation with pilot tone is preferably done at the optical transmitter, transponder or repeater, by adding

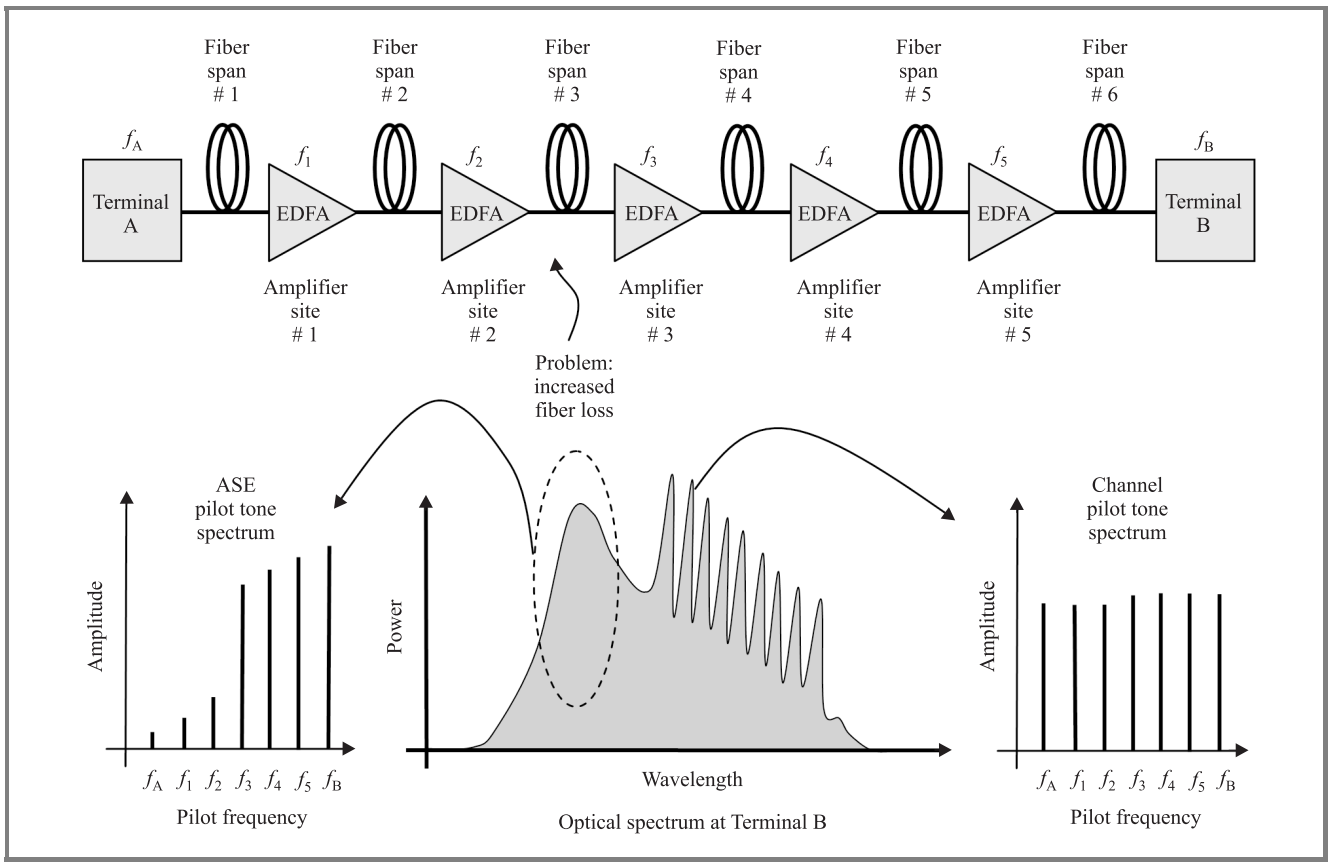


Fig. 7. Labels added at EDFAs in terminals and in-line repeaters enable to locate noisy amplifier span with degraded fiber.

low frequency sine wave to bias or drive current of semiconductor laser (Fig. 4). Alternative methods include passing host signal through:

- optical modulator driven with sine wave of low amplitude;
- semiconductor optical amplifier; sine wave added to DC bias current changes SOA gain;
- EDFA amplifier; sine wave is added to DC bias current of pump laser(s), modulating EDFA gain;
- vibrating reflective component, e.g., mirror in a MEMS switching matrix.

Modulating amplifier gain in WDM line system with pilot tone adds the same label to all signals, so cannot help with their selective identification. However, spurious signals like ASE noise and FWM products are modulated, too. Adding unique pilot tone at each amplifier allows to trace noise contribution of every span (Fig. 7).
Label of this type can be erased, e.g., by passing signal through semiconductor amplifier operating in saturated mode or O/E/O repeater and than replaced with new label if necessary. Re-labeling is of particular interest when signals enter another transport network with separate management system.

Alternatively, pilot tones may be imposed on host signals using phase modulation (PM) or frequency modulation (FM), rather than intensity modulation (AM). Despite advantages like no interference with amplifiers and more flexible selection of label frequencies, this method is less commonly tried, because:

- PM and FM labels require complicated and costly receivers;
- there is no simple method to read multiple labels from mix of several optical signals;
- host signal must exhibit high spectral purity and frequency stability.

However, PM and FM labels are not affected by EDFA dynamics or transferred due to optical nonlinearities, and better choice for optically amplified, non-repeatered DWDM links spanning large distances.

4.1.1.5. Receiver power penalty resulting from pilot tone label

Imposing pilot tone or subcarrier-modulated label (Section 4.1.2) by means of amplitude modulation adds noise to 1s of host stream (effects on 0s are negligible) and increases bit error ratio (BER). Power penalty measured as

increase in receiver input power necessary to maintain constant BER depends on modulation index (Fig. 8) and is independent of label frequency, as long as its lower than host clock frequency [13, 15]. Curve shown below is an estimate of upper limit of label penalty; experiments give penalty values up to 50% lower.

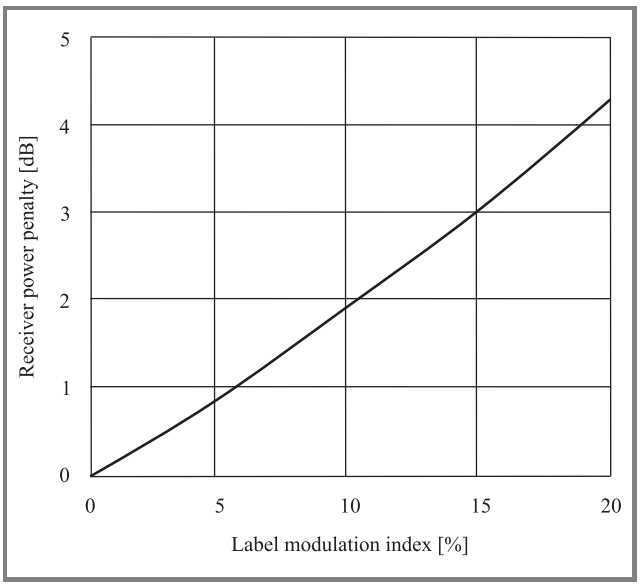


Fig. 8. Estimated label penalty introduced to NRZ receiver at BER $= 10^{-9}$ [15].

Label penalty is a kind of optical path penalty. The standard requirement for path penalties in SDH networks [4] is 1 dB, dictating modulation index not greater than 7%.

4.1.1.6. Transfer of amplitude-modulated sine-wave labels resulting from stimulated Raman scattering

Dense wavelength division multiplexing transport systems in backbone networks usually comprise several fiber spans, each 70–120 km long and EDFA line amplifiers with output powers between +13 dBm and +24 dBm; total length of unrepeatered link often exceeds 500 km. Stimulated Raman scattering occurring primarily in the fiber lengths directly after each EDFA causes cross-modulation of all WDM channels.

In effect, signals are amplitude modulated with new "ghost" label frequencies being products of intermodulation between original set of labels [5, 9]. Amplitude of "ghost" label increases proportionally to system length (number of EDFAs) and signal power; does not noticeably depend on frequency of pilot tones.

The problem is of particular importance when pilot tones are arranged with fixed spacing, as "ghosts" overlap with genuine labels, resulting in false channel identification and power measurement errors.

Investigation by team of Korean researchers [9] led to conclusion, that a DWDM system with 32 channels, +3 dBm channel power, and 100 GHz channel spacing, working on G.652 fiber in C band has a maximum length of 400 km (5×80 km), if a 10 dB ghost-to-label ratio is to be guaranteed.

Attempts were made to eliminate the problem by adding "saturation" or "control" channel, whose power is quickly adjusted to maintain fixed level of EDFA output; such solutions are already in use in certain commercial systems to maintain desired EDFA gain and spectral characteristics despite changing number of active channels. The solution is effective in a single span. In a multi-span system, however, uneven spectral gain of EDFAs and adding/dropping of channels results in loss of balance and appearance of "ghosts"—unless each amplifier site is equipped with its own saturation channel controller, which is expensive.

4.1.2. Low frequency subcarrier-modulated label with data channel

This method is an extension of pilot tone label. Carrier wave of relatively low frequency (10 kHz–10 MHz) (see Fig. 9) is modulated with data using ASK, FSK or DPSK modulation, then superimposed on host signal by modulating its amplitude (AM/ASK), phase (PM/PSK/QPSK) or frequency (FM/FSK). Data rates are limited to 1–150 kbit/s, as the label is located within region of high spectral density of the host. STM-256 (40 Gbit/s) signals are expected to be RZ-coded, so reduced spectral density at low frequencies will allow for higher label rates. All modulation methods and transmitter design (Fig. 4) presented in Section 4.1 are applicable to subcarrier-modulated (SCM) labels as well.

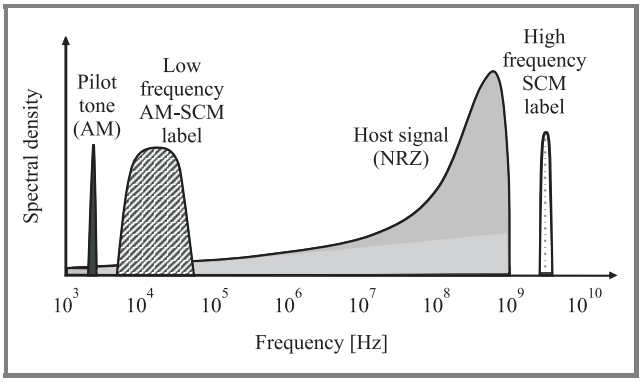


Fig. 9. Label signals with respect to spectrum of STM-16 host (density per octave shown). Only one label option is usually implemented.

Main advantages of this labeling method include:

- low-speed and inexpensive electronics for label processing;
- immunity to effects of fiber dispersion;
- for AM or ASK modulation: little or no changes to optical receivers.

For a given host structure and line code (e.g., STM-1/4/16/64—NRZ), available label bit rate grows with host

signal rate and allowable power penalty to host signal resulting from eye closure. The latter is preferably kept below 1 dB. For NRZ modulated SDH host and DPSK modulation, overhead bit rate is limited to $1–2 \cdot 10^{-5}$ of host rate. In a "transparent" system adding and reading labels in transit, label capacity must be suited to least supportive host type.
Ability to transmit user information makes this method versatile and suitable for complex networks. SCM labels can convey information like:

- channel identification: source, destination, content, protection priority, encryption, etc.;
- physical channel characteristics: wavelength, bit rate, use of FEC, etc.;
- management data: alarms, telemetry, protection switching commands, housekeeping, etc.;
- operator data associated with given channel;
- extra payload channel(s).

Communication with amplifier sites is also possible: modulation of EDFA or SOA gain allows to apply a new label and transmit overhead information [13] without adding dedicated service channel transponders and associated optical couplers. This solution was proposed for telemetry and remote control in long-distance submarine and terrestial systems, including transoceanic links with up to 300 EDFA amplifiers. Unfortunately, large variety of solutions proposed makes adoption of common standard unlikely.
Choice of operating frequency with AM modulation is restricted exactly as for pilot tones. Because of wider bandwidth occupied and therefore lower signal to noise ratio compared to unmodulated pilot tones, AM labels of this kind are generally unsuitable for signal power monitoring.

Application of label by frequency or phase keying (FSK/QPSK) of host signal avoids interference with optical amplifiers. Label modulation is accomplished by external Mach-Zehnder interferometer driven with signals of opposite phase. Another method is to modulate laser drive current, which results in changes of both intensity and frequency of emitted light. If the signal subsequently passes through a booster amplifier operating in saturation mode, unwanted amplitude modulation is suppressed.
Such label is "invisible" to ordinary intensity detectors, and must be detected by separate optical subsystem (e.g., interferometric), adding to equipment cost and complexity.

4.1.3. High frequency subcarrier-modulated data channel

Location of label above band occupied by the host signal (Figs. 9 and 10) has several advantages:

- labeled signals pass through multiple EDFAs without attenuation of labels;
- labels cause no interference to power controllers of EDFA amplifiers;
- fewer problems with label transfer caused by fiber or amplifier nonlinearity;
- possibility of dispersion monitoring in certain labeling schemes;
- very high label bit rates are possible.

With proper frequency separation, label-host interference is eliminated and label data rate may easily exceed 10% of host rate. Very high overhead capacity is possible, but reduces host power budget accordingly, as signal power is divided between two frequency multiplexed components: the host and the label. AM label can be separated from host signal by means of simple band-pass electrical filter in the receiver.

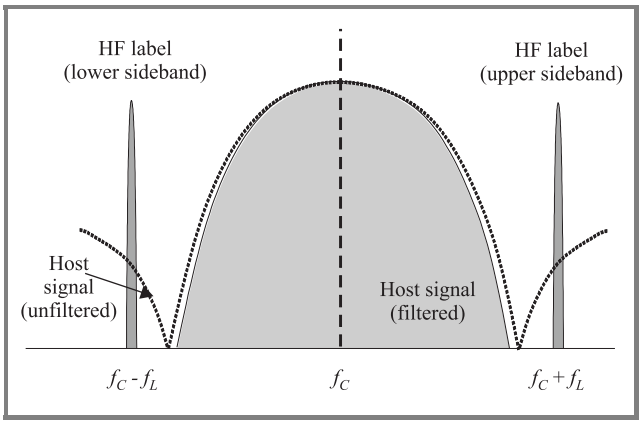


Fig. 10. Optical spectrum of host signal with high frequency SCM label (horizontal axis: optical frequency). Explanations: f_C—host carrier frequency; f_L—label frequency.

This solution has its problems, namely:

- label is more sensitive to fiber dispersion (CD, PMD) than host signal; dispersion induced fading can even erase the label completely (see Section 4.1.3.1);
- wideband, more expensive (and non-standard) receiver is needed;
- AM label modulation often requires use of external optical modulator instead of simpler direct modulation by changing the bias current of transmit or pump laser;
- monitoring scheme shown in Fig. 5 cannot be implemented.

Host receiver does not need modification, as long as its bandwidth is tailored to given host signal rate, which ensures attenuation of label before decision circuit. Use of "transparent" 2R transponders operating below specified maximum bit rate, e.g., receiving 1.25 Gbit/s GbE stream instead of 2.5 Gbit/s or 2.7 Gbit/s (STM-16/OC-48) violates this condition and may result in considerable power penalty or appearance of error floor.

With large frequency separation between host and label, one may apply optical filtering, e.g., removing host signal with matched fiber Bragg grating (FBG) filter (Fig. 11). This technique is expensive due to need for matched filter for each channel, but allows to use narrow-band, low-cost receiver for the label.

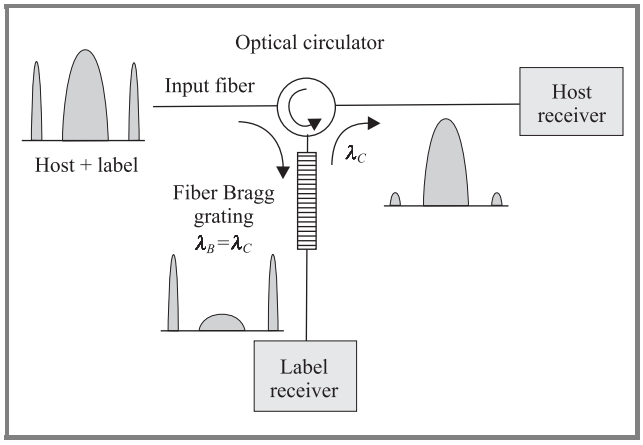


Fig. 11. Separation of host and high frequency label using optical filter [28].

Separation of label by filtering requires that overlap of host and label spectra be minimized. As the label occupies less bandwidth and carries less energy than host, the main issues when labeling NRZ-coded signal are:

- Removal of host components lying above clock frequency before modulation of optical carrier (see Fig. 10). A low-pass filter with 3 dB bandwidth equal to about 0.8 of clock frequency and linear modulator are necessary to support subcarrier-modulated label, while still meeting optical waveform specifications and avoiding noticeable power penalty. Pilot-tone solutions are considerably more tolerant, as label receiver bandwidth can be in principle be made as narrow as necessary, reducing host crosstalk accordingly.
- Selection of label frequency. Location of high bit rate label, carrying a considerable fraction of total signal power, e.g., 5% or 10% sets a rigid limit to symbol rate of host signal.
- Limiting extinction ratio of host signal. Certain level of carrier wave transmitted in 0s must be guaranteed, otherwise SCM label received will be severely distorted.

The last requirement directly compromises quality and the Q-factor of host signal. Minimum extinction ratio specified for SDH equipment [4] is 8.2 dB or 10.0 dB, and 9.0 dB for 1 Gbit/s gigabit Ethernet [35] in order to maximize receiver performance. Experiments with high frequency SCM–DPSK labels [19] involved reduction of host extinction ratio to as low as 3 dB, causing severe power penalties. Such arrangements do not, however, directly violate 10 Gbit/s gigabit Ethernet standard [36], specifying minimum extinction ratios of 3.0 dB or 3.5 dB only to allow use of low cost lasers and modulators.

4.1.3.1. Dispersion fading

Transmission of high frequency pilot tone or subcarrier-modulated data stream by means of amplitude (intensity) modulation is unfortunately subject to severe impairments caused by fiber dispersion. Chromatic dispersion (CD) produces a phase shift between two sidebands, increasing in proportion to CD coefficient and fiber length:

$$\Delta\phi = 4\pi L D \lambda_C^2 f_L^2/\mathrm{c}, \tag{4}$$

where: $\Delta\phi$—phase shift [rad]; λ_C—host carrier wavelength [nm]; f_L—label frequency [GHz]; L—fiber length [km]; D—fiber chromatic dispersion coefficient [ps/nm·km]; LD—link chromatic dispersion [ps/nm]; c—speed of light [approx. $2.99792 \cdot 10^8$ m/s].
As the sidebands are of equal amplitude, when their relative phase shift reaches 180 degrees ($\Delta\phi = \pi$), respective photocurrents generated in non-coherent, wideband photo detector cancel each other and the label is lost. This border condition is defined by equation [31]:

$$2LD\lambda_C^2 f_L^2 = \mathrm{c}. \tag{5}$$

Corresponding length of fiber link is:

$$L_{LOSS} = \mathrm{c}/2D\lambda_C^2 f_L^2. \tag{6}$$

A 3 dB ($1/\sqrt{2}$) reduction of receiver output voltage occurs at phase shift of 90° ($\pi/2$), which can be assumed as practical limit of label transmission. This happens at link length equal to:

$$L_{\max} = \mathrm{c}/4D\lambda_C^2 f_L^2. \tag{7}$$

This phenomenon is known as "RF dispersion fading" and constitutes important performance limit in frequency-multiplexed (SCM) fiber transmission systems.
For the following, typical parameters of STM-16 channel in C-band system implemented with G.652 fiber: $\lambda_C = 1550$ nm, $D = 17$ ps/nm · km, $f_L = 3$ GHz, we get $L_{LOSS} = 1836$ km and $L_{\max} = 918$ km. The latter value is lower than dispersion-limited length of 2.5 Gbit/s link on G.652 fiber, estimated at 1000 km. This is not surprising, because bandwidth of typical STM-16 receiver is only 1.8–2.0 GHz.
With proper choice of label frequencies and adequate power margin over host signal, say 5 dB, label and host are compatible in terms of CD and PMD tolerance. Label frequency significantly higher than host bandwidth is not feasible without additional dispersion compensation. As choice of label frequency depends on spectrum of particular host signal, network transparency is restricted or lost.
Use of amplitude- (AM) and phase-modulated (PM) labels is proposed for monitoring of dispersion in systems with

adaptive CD compensation [31, 32]. PM labels provide better accuracy in presence of PMD and fiber nonlinearity.

Experiments with FSK, PSK and DPSK modulation of optical carrier reported better performance and overhead bit rates up to 2.5 Gbit/s [19]. This method requires single mode laser with good frequency stability, so is predominantly applicable to DWDM equipment for core networks.

4.1.4. Phase and frequency modulation of optical host carrier

Researchers participating in STOLAS project have proposed several methods for adding overhead of considerable capacity (155 Mbit/s) to 10 Gbit/s host, by differential phase shit keying (DPSK) [24] or frequency shift keying (FSK) of the optical carrier wave carrying the amplitude keyed (NRZ) scrambled STM-64 host. The expected penalty in case of DPSK label was an increase of host jitter due to interaction with fiber dispersion, limited to about 0.02 unit intervals (UI) at maximum allowable path dispersion of 1500 ps/nm. Laser used here must emit radiation with very narrow spectral width, well below 50% of label clock frequency. For label detection, a DPSK–ASK conversion is provided by Mach-Zehnder interferometer with fiber delay line placed in one arm.
Frequency shift keying modulation can be imposed by directly adding label pulses to laser bias current.

4.1.5. Insertion of label bits into fixed locations within frame of the host signal

This is a kind of time division multiplexing utilizing spare capacity in frame structure, where specific, easily recognizable bit patterns are formed. Repetitive patterns can create low-frequency component in signal spectrum, detectable by narrow-band receiver without decoding host signal or recognized by simple (but fast) logic circuits. For example, alternatively inserting a string of 0s in 4 frames and equally long string of 1s in the next 4 frames creates AM component at 1/8th of frame repetition frequency. The first case is an alternative implementation of pilot tone method (Section 4.1.1).
Bit pattern label is not truly orthogonal, as it requires host signal of specific frame structure. It also prevents use of spare bits for other purposes, like in-band FEC or user data channels. Host signals with little or no spare bits cannot support this technique.
Bit insertion labels have unique advantages:

- can transit through properly configured O/E/O repeaters, transponders and other active devices;
- do not degrade receiver sensitivity;
- can be read very quickly;
- are generated by standard digital circuits handling host signal;
- can be accessed by multiplexers, routers or transmission analyzers with upgraded software.

Fast readout makes bit pattern labels attractive for routing purposes in networks carrying one type of traffic from diverse sources. Problems include:

- difficulty with adding a label to signal in transit without using costly O/E/O conversion;
- limited choice of label frequencies, usually tied to frame repetition frequency of host signal;
- possible interference to optical amplifiers with closed-loop gain control.

4.1.5.1. Bit pattern labeling of SDH streams

ITU-T Recommendation G.707 [2] reserves up to 30 bytes within STM-1 frame for various uses, so labeling does not violate existing standards. Frame repetition frequency is 8 kHz, so available pilot tone frequencies are 4 kHz (8 kHz/2), 2.6667 kHz (8 kHz/3), etc. STM-1 frame consists of 2430 bytes and has average mark ratio of 50%. Setting content of N bytes in alternate frames to 0s and 1s causes frame-to-frame relative power variation of:

$$N/(2430/2) = 8.23 \cdot 10^{-4} N \text{ or } 0.0823\% \text{ per byte.} \quad (8)$$

Filtering the envelope to pass pilot tone only results in modulation index m of approx. 0.05% per byte. Utilization of 4 bytes gives modulation index $m = 0.2\%$; reserving more bytes is not realistic due to multiple demands for this limited resource. This value shall be adequate for reliable detection of static pilot tone, and give at least 20 dB of dynamic range for amplitude measurements with typical filter bandwidth of 20 Hz or less.
Time multiplexing does not remove labels, but modulation index is reduced in proportion to bit rate. Therefore, an STM-1 stream labeled with pilot tone generated by 4 bytes and subsequently multiplexed to STM-64 level will have to be identified with $m = (0.2/64)\% = 0.0031\%$ only. Setting all 64 label patterns identical restores original m value, but individual tributaries are no longer traceable.
As 28 out of 30 reserved bytes within STM-1 frame are subject to scrambling, the label pattern must be pre-coded to result in desired sequence after scrambling.
Overhead capacity can be increased dramatically by taking away a part of payload capacity to create a strong, easy to read label, e.g., replacing one or more of STM-1 tributaries with streams of alternating fixed bit sequences, to create a modulated subcarrier. Modulation scheme must minimize components below ≈ 100 kHz, to avoid interference to EDFA controllers and inter-modulation between channels.
Such approach results in impressive performance [34], but is incompatible with existing standards and involves considerable loss of transmission capacity. Marking of SDH streams in transit is not possible.

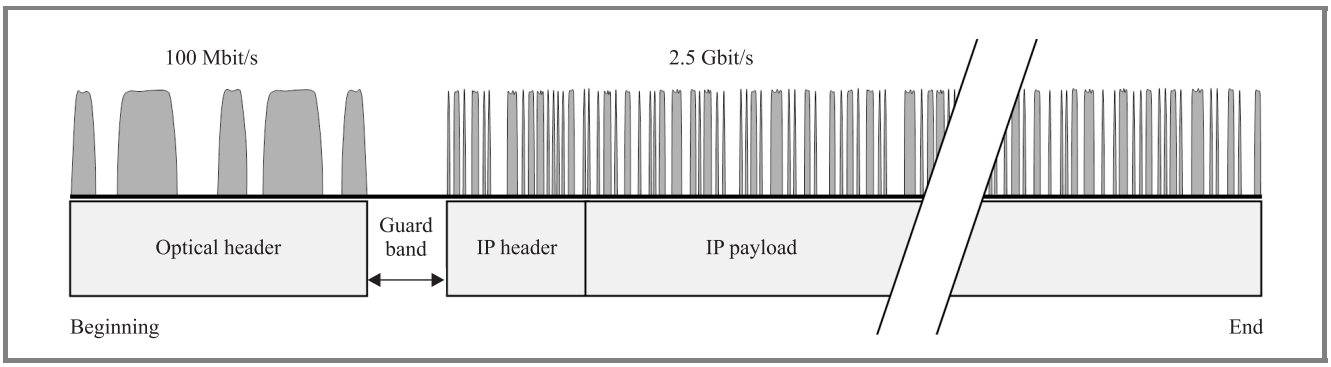


Fig. 12. Structure of optical packet.

4.2. Labeling in optical packet networks

Due to nature of packet networks, where individual packets are handled be several nodes in random order and generally do not form a continuous bit stream while in transit, there is a strong demand for "label swapping"—removal of existing label (or part of it) and replacing it with new one generated locally. Packet can undergo such operation in several nodes. Equivalent operations are routinely performed by electronic packet switching equipment in use today.

In optical packet networks, a label is added to every single data packet. Labels must be read fast, in order to effect packet routing without delay. There is no optical equivalent of random access memory (RAM) used for buffering data in electronic switches, except for fixed length, costly and bulky fiber delay lines. Readout time specified is measured as fraction of packet length. For example, allowing 20% of duration of 1024-byte, 10 Gbit/s packet gives 164 ns for label detection. Packets have standardized bit rate and structure, so label orthogonality is not essential.

Label handling devices are expected to be part of complex core routers and switches, so their cost and complexity are of lesser importance, but reliability requirements are high. The field is open for advanced components, initially manufactured in limited quantities.

4.2.1. Header (bit sequence) labeling of optical packets

This scheme is equivalent to solution used in electronic networks: the main (payload) part of packet is preceded by a header carrying routing and identification data (Fig. 12), whose contents is accessible to switching nodes, test equipment, etc.

The drive to transmit data optically at highest rates possible (40 Gbit/s and beyond) has resulted in "electronic bottleneck": existing digital circuits cannot directly handle them, or are prohibitively expensive and power consuming. Those are exactly the reasons for introduction of optical switching.

To allow readout of label (header) by inexpensive electronic receivers, label bit rate is often lower than payload rate, e.g., 100 Mbit/s versus 2.5 Gbit/s or 40 Gbit/s, as shown in Fig. 12. The payload itself comprises its own header, utilized by lower network layers. To allow for timing errors and enable proper receiver synchronization, both parts are separated by a guard band. This is unlike electronic systems where both rates are identical.

This header-payload split has a far-reaching consequence: one can envision a network, where payloads are of different bit rates, formats and lengths, but headers are all standardized and the optical core is "transparent" and upgradeable.

Lower bit rate means proportionally decreases receiver noise and input power required, so tapping a small portion of signal passing given node, e.g., 10% provides enough input power for header detection. "Slow" header is substantially more resistant to dispersion than payload, and pulse quality requirements (extinction ratio, rise/fall time, chirp, etc.) can be relaxed, with margin for degradation and added noise resulting from label swapping process. Dispersion management becomes more flexible, as "slow" headers are readable at any location along optical path, without accurate dispersion compensation required for payload.

4.2.1.1. Label swapping

Old header can be removed using gated semiconductor optical amplifier, as this device is compact, fast and potentially inexpensive. New label is added using DFB laser transmitter and optical coupler. A simple arrangement of optical header processor is shown (Fig. 13); several other variants are known from literature [10, 24, 37]. All active components are suitable for integration as single monolithic or hybrid IC. More complex processors [37] containing additional Mach-Zehnder interferometers, tunable laser and SOA are able to perform flexible wavelength conversion as well.

Implementation of scheme shown in Fig. 13 raises several design issues:

- label receiver must work with bursty signals, quickly synchronizing after idle period of random duration;
- total delay of lower and upper signal paths (including processing time) must be equal;
- levels and wavelengths of new label and payload must be precisely matched.

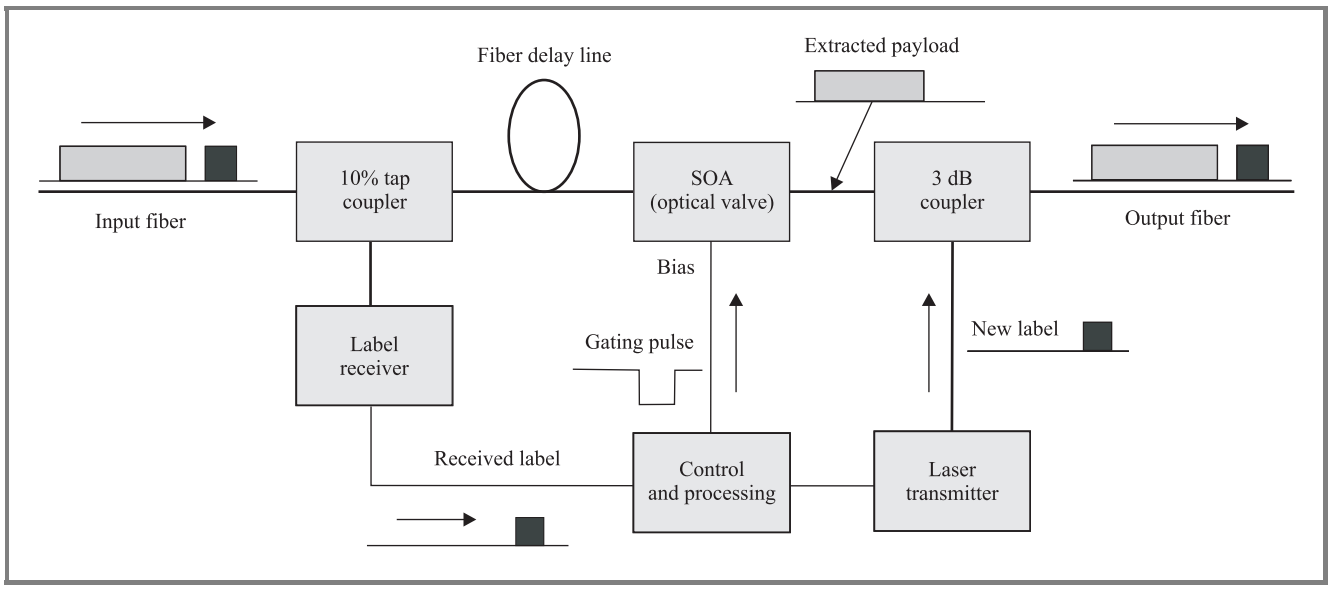


Fig. 13. Simple header processor for optical packet network.

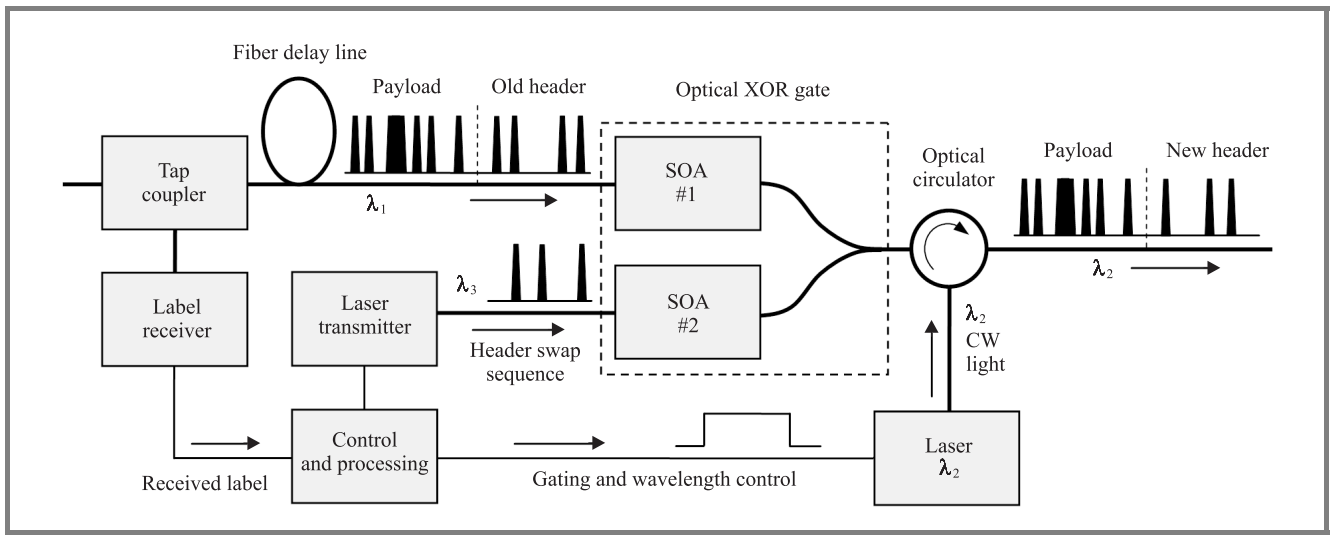


Fig. 14. Header processor with wavelength conversion.

Without wavelength match, transmission over longer distances results in relative shifting of label and payload due to fiber chromatic dispersion. More complex label processing schemes include transfer of both label and payload to new wavelength, using interferometric wavelength converter, also known as optical XOR gate (Fig. 14). Such a device has already been made as single monolithic circuit.

Tests with 10 Gbit/s packets and headers, both of them RZ-coded, converted from 1538 nm to 1543 nm revealed satisfactory performance: 0.4 dB power penalty due to wavelength conversion and 13 dB extinction ratio of the output signal [19].

4.2.2. WDM labeling of optical packets

The header can be sent on wavelength(s) separate from the payload. This technique has advantages:

- routing decisions can be generated by optical correlation of pulses, resulting in **purely optical** switching technology;
- reliable separation of label and payload, which can overlap in time; there are no limits to header size;
- use of techniques and components already developed for WDM networks.

Unfortunately, it is wasteful in DWDM networks, as header occupies at least one more wavelength. DWDM lasers and multiplexers are expensive, although adoption of CWDM technology will reduce costs. There are numerous research papers on this subject, most of them from Japan [23, 25, 26].

The primary reason for researching this complicated technology is its potential for purely optical and extremely fast core switches, with no electronics handling packet control

functions. In several experiments, a set of wavelengths served as "signaling channel", with routing data encoded as combinations of present and absent pulses. Complete header is then read using optical correlators based on set of fiber or waveguide delay lines and couplers.

Very fast routing in core network (between limited number of nodes) is accomplished by use of lookup tables, containing predefined new headers and sets of switching and wavelength conversion instructions. A purely optical technology to implement this function is not yet available.

4.2.3. SCM labeling of optical packets

High frequency SCM labeling is often proposed for optical packet networks. Carrier frequencies are usually higher than clock rate of payload bit stream (≥ 3 GHz) and efficient transmission of label often requires restricted extinction ratio of the host signal. The most common method of label carrier modulation is DPSK.

Label bit rates are in order of 100 Mbit/s and higher, therefore even a short packet carries adequate amount of overhead. For example, a 155 Mbit/s label imposed on 256-byte long 2.5 Gbit/s packet consists of 16 bytes, although considerable part of it must be sacrificed to allow for burst synchronization of receiver at the beginning of packet.

The method is robust, effective and can be implemented with existing components. Semiconductor amplifiers may perform amplitude modulation in transit nodes, working in non-saturated regime, with label signal added to bias current. Amplitude-modulated label may be erased by the same SOA working in saturated regime. Unfortunately, dispersion fading problems discussed in Section 4.1.3.1 may constitute a serious limit on transmission distance or fiber type used and impose more strict dispersion compensation requirements.

5. Conclusions

There are several options for labeling and attaching extra information to signals transmitted in transparent optical networks. Few technologies developed for this purpose have found other applications, like monitoring of path dispersion and detecting network intrusions.

Unfortunately, international standards have not emerged despite extensive, but extremely divergent research activity. Interest in building high performance all-optical core networks is currently missing due to worldwide downturn in telecom sector, so most optical labeling technologies have to wait for future applications.

Exceptions include use of simple pilot tones and subcarrier modulation techniques in transoceanic systems and for protection of high security networks. Both applications are distanced from mainstream activities of most networks operators in Central Europe.

References

[1] "Interfaces for the optical transport network (OTN)", ITU-T Rec. G.709/Y.1331 (12-2003).

[2] "Network node interface for the synchronous digital hierarchy (SDH)", ITU-T Rec. G.707/Y.1322 (12-2003).

[3] "Optical interfaces for multichannel systems with optical amplifiers", ITU-T Rec. G.692 (06-2002).

[4] "Optical interfaces for equipments and systems relating to the synchronous digital hierarchy", ITU-T Rec. G.957 (12-2003).

[5] H. C. Ji, K. J. Park, J. K. Kim, and Y. C. Chung, "Optical path and crosstalk monitoring technique using pilot tones in all-optical WDM transport network", in *Proc. SPIE, APOC 2001*, Beijing, China, 2001, vol. 4584.

[6] K. U. Chu, C. H. Lee, and S. Y. Shin, "Optical path monitoring based on the identification of optical cross-connect input ports", in *Opt. Fib. Commun. Conf. OFC'99*, San Diego, USA, 1999, Paper FJ5.

[7] H. C. Ji, K. J. Park, J. K. Kim, S. K. Shin, M. J. Jang, and Y. C. Chung, "Demonstration of optical path monitoring technique using dual pilot tones in all-optical WDM transport network", in *OptoElectron. Commun. Conf. OECC*, Chiba, Japan, 2000 (published in *Tech. Dig.*, 2000, Ser. 5).

[8] K. J. Park, S. K. Shin, and Y. C. Chung, "A simple monitoring technique for WDM networks", in *Conf. OFC'99*, San Diego, USA, 1999, Paper FJ3.

[9] H. S. Chung, S. K. Shin, H. G. Woo, and Y. C. Chung, "Effects of stimulated Raman scattering on pilot-tone based WDM supervisory technique", in *Conf. OFC'2000*, Baltimore, USA, 2000, Paper WK7-1.

[10] Y. G. Wen, Y. Zhang, and L. K. Chen, "On architecture and limitation of optical multiprotocol label switching (MPLS) networks using optical-orthogonal-code (OOC)/wavelength label", *Opt. Fib. Technol.*, vol. 8, no. 1, pp. 43–70, 2002.

[11] R. Gaudino and D. J. Blumenthal, "WDM channel equalization based on subcarrier signal monitoring", in *Conf. OFC'98*, San Jose, USA, 1998, Paper WJ6.

[12] E. Kong, F. Tong, K. P. Ho, L. K. Chen, and C. K. Chan, "An optical path supervisory scheme for optical cross-connects using pilot tones", in *CLEO Pacific Rim 1999*, Seoul, Korea, 1999, Paper FT4.

[13] M. Murakami, T. Imai, and M. Aoyama, "A remote supervisory system based on subcarrier overmodulation for submarine optical amplifier systems", *J. Lightw. Technol.*, vol. 14, no. 5, pp. 671–677, 1996.

[14] N. Suzuki, K. Shimizu, T. Kogure, J. Nakagawa, and K. Motoshima, "Optical fiber amplifiers employing novel high-speed AGC and tone-signal ALC functions for WDM transmission systems", in *Conf. ECOC 2000*, Munich, Germany, 2000, vol. 2, p. 179.

[15] Y. Hamazumi and M. Koga, "Transmission capacity of optical path overhead transfer scheme using pilot tone for optical path network", *J. Lightw. Technol.*, vol. 15, no. 12, pp. 2197–2205, 1997.

[16] Y. Sun, A. A. M. Saleh, J. L. Zyskind, D. L. Wilson, A. K. Srivastava, and J. W. Sulhoff, "Time dependent perturbation theory and tones in cascaded erbium-doped fiber amplifier systems", *J. Lightw. Technol.*, vol. 15, no. 7, pp. 1083–1087, 1997.

[17] Y. C. Chung, "Wavelength control and monitoring in WDM networks", in *All Opt. Netw. ComForum IEC*, Orlando, USA, 1999.

[18] M. C. R. Medeiros, I. Darwazeh, L. Moura, A. L. Barradas, A. Teixeira, P. André, M. Lima, and J. da Rocha, "Dynamically allocated wavelength WDM network demonstrator", in *Proc. ConfTele 2001*, Figuera da Foz, Portugal, 2001, pp. 169–173.

[19] N. Chi, B. Carlsson, Z. Jianfeng, P. Holm-Nielsen, Ch. Peucheret, and P. Jeppesen, "Transmission properties for two-level optically labeled signals with amplitude-shift keying and differential phase-shift keying orthogonal modulation in IP-over-WDM networks", *J. Opt. Netw.*, vol. 2, no. 2, pp. 46–54, 2003.

[20] T. Koonen, S. Sulur, I. T. Monroy, J. Jennen, and H. de Waardt, "Optical labeling of packets in IP-over-WDM networks", in *Proc. URSI Gen. Assem. 2002*, Maastricht, The Netherlands, 2002, Oral Paper no. 1526.

[21] S. Sulur, T. Koonen, I. T. Monroy, H. de Waardt, and J. Jennen, "IP over DWDM networks supported by GMPLS-based LOBS deploying combined modulation format", in *Opticomm 2001*, Denver, USA, 2001.

[22] T. Koonen, G. Morthier, J. Jennen, H. de Waardt, and P. Demeester, "Optical packet routing in IP-over-WDM networks deploying two-level optical labeling", in *Eur. Conf. Opt. Commun. ECOC-2001*, Amsterdam, The Netherlands, 2001, Paper ThL.2.1.

[23] H. Harai and N. Wada, "Photonic packet forwarding in a multi-wavelength label switching node", in *IEEE ICC Works. Next Gen. Switch./Rout.: Opt. Role*, Helsinki, Finland, 2001.

[24] T. Fjelde, "Novel scheme for efficient label-swapping using simple XOR gate", in *Conf. ECOC-2000*, Munich, Germany, 2000, Paper 10.4.2.

[25] J. McGeehan, M. C. Hauer, A. B. Sahin, and A. E. Willner, "Reconfigurable multi-wavelength optical correlator for header-based switching and routing", in *Conf. OFC'2002*, Anaheim, USA, 2002, Paper WM4.

[26] A. Okada, "All-optical packet routing in AWG-based wavelength routing network using out-of-band optical label", in *Conf. OFC'2002*, Anaheim, USA, 2002, Paper WG1.

[27] V. J. Hernandez, Z. Pan, J. Cao, V. K. Tsui, Y. Bansai, S. K. H. Fong, Y. Zhang, M. Y. Jeon, and S. J. B. Yoo, "First field trial of optical label-based switching and packet drop on a 477 km NTON/sprint link", in *Conf. OFC'2002*, Anaheim, USA, 2002, Paper TuY4.

[28] H. J. Lee, S. J. B. Yoo, V. Tsui, and S. K. H. Fong, "A simple all-optical label detection and swapping technique incorporating a fiber Bragg grating filter", *IEEE Photon. Technol. Lett.*, vol. 13, no. 6, pp. 635–637, 2001.

[29] G. Bendelli, C. Cavazzoni, R. Girardi, and R. Lano, "Optical performance monitoring techniques", in *Conf. ECOC-2000*, Munich, Germany, 2000, vol. 4, p. 113.

[30] K. Kitayama, N. Wada, and H. Sotobayashi, "Architectural considerations for photonic IP router based upon optical code correlation", *J. Lightw. Technol.*, vol. 18, no. 12, pp. 1834–1844, 2000.

[31] G. Rossi, T. E. Dimmick, and D. J. Blumenthal, "Optical performance monitoring in reconfigurable WDM optical networks using subcarrier multiplexing", *IEEE J. Lightw. Technol.*, vol. 18, no. 12, pp. 1639–1648, 2000.

[32] K. J. Park, C. J. Youn, J. H. Lee, and Y. C. Chung, "Performance comparisons of chromatic dispersion-monitoring techniques using pilot tones", *IEEE Photon. Technol. Lett.*, vol. 15, no. 6, pp. 873–875, 2003.

[33] "National Communications System—Technical Information Bulletin 00-7: All-Optical Networks (AON)", Office of the Manager, National Communications System, Arlington, USA, 2000.

[34] Y. Horiuchi and M. Suzuki, "Ultra-robust optical routing by digitally encoded SCM label", in *Conf. OFC'2002*, Anaheim, USA, 2002, Paper TuV3.

[35] "IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements". Part 3: "Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications", IEEE Std 802.3-2002.

[36] "IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements". Part 3: "Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications". Amendment: "Media Access Control (MAC) Parameters, Physical Layers, and Management Parameters for 10 Gb/s Operation", IEEE Std 802.3ae-2002.

[37] D. Blumenthal, J. Bowers, and L. Coldren, "Integrated photonic modules for all-optical label swapping in WDM/TDM networks", in *DARPA NGI PI Rev.*, Santa Barbara: University of California, 2000.

Krzysztof Borzycki was born in Warsaw, Poland, in 1959. He received the M.Sc. degree in electrical engineering from Warsaw University of Technology in 1982. He has been with National Institute of Telecommunications (NIT) since 1982, working on optical fiber fusion splicing, design of optical fiber transmission systems, measurement methods and test equipment for optical networks, as well as standardization and conformance testing of optical fiber cables, SDH equipment and DWDM systems. Other activities included being a lecturer and instructor in fiber optics and technical advisor to Polish fiber cable industry. He has also been with Ericsson AB research laboratories in Stockholm, Sweden, working on development and testing of DWDM systems for long-distance and metropolitan networks between 2001 and 2002 while on leave from NIT. He is currently working toward Ph.D. degree, focusing on adding overhead information to optical signals for network management and switching purposes. Other important research activity is investigation of aging and temperature effects on polarization mode dispersion (PMD) in optical fibers and cables. He is an author or co-author of 1 book, 2 Polish patents and more than 30 scientific papers in the field of optical fiber communications, optical fibers and measurements in optical networks, as well as one of *Journal of Telecommunications and Information Technology* editors.
e-mail: k.borzycki@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

INFORMATION FOR AUTHORS

The *Journal of Telecommunications and Information Technology* is published quarterly. It comprises original contributions, both regular papers and letters, dealing with a broad range of topics related to telecommunications and information technology. Items included in the journal report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

The *Journal* is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, speech devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology. Suitable research-related manuscripts should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published by invitation only.

Papers published by invitation and regular papers should contain up to 15 and 8 printed pages respectively (one printed page corresponds approximately to 3 double-space pages of manuscript, where one page contains approximately 2000 characters).

Manuscript: An original and two copies of the manuscript must be submitted, each completed with all illustrations and tables attached at the end of the papers. Tables and figures have to be numbered consecutively with Arabic numerals. The manuscript must include an abstract limited to approximately 100 words. The abstract should contain four points: statement of the problem, assumptions and methodology, results and conclusion, or discussion, of the importance of the results. The manuscript should be double-spaced on only one side of each A4 sheet (210 × 297 mm). Computer notation such as Fortran, Matlab. Mathematica etc., for formulae, indices, etc., is not acceptable and will result in automatic rejection of the manuscript. The style of references, abbreviations, etc., should follow the standard IEEE format.

References should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. The **reference entry** (correctly punctuated according to the following rules and examples) **has to contain**.

From journals and other serial publications: initial(s) and second name(s) of the author(s), full title of publication (transliterated into Latin characters in case it is in Russian, possibly preceded by the title in Russian characters), appropriately abbreviated title of periodical, volume number, first and last page number, year. E.g.:

[1] Y. Namihira, "Relationship between nonlinear effective area and modefield diameter for dispersion shifted fibres", *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

From non-periodical, collective publications: as above, but after title – the name(s) of editor(s), title of volume and/or edition number, publisher(s) name(s) and place of edition, inclusive pages of article, year. E.g.:

[2] S. Demri, E. Orłowska, "Informational representability: Abstract models versus concrete models" in *Fuzzy Sets, Logics and Reasoning about Knowledge,* D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

From books: initial(s) and name(s) of the author(s), place of edition, title, publisher(s), year. E.g.:

[3] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

Figure captions should be started on separate sheet of papers and must be double-spaced.

Illustration: Original illustrations should be submitted. All line drawings should be prepared on white drawing paper in black India ink. Drawings in Corel Draw and Postscript formats are preferred. Colour illustrations are accepted only in exceptional circumstances. Lettering should be large enough to be readily legible when drawing is reduced to two- or one-column width – as much as 4:1 reduction from the original. Photographs should be used sparingly. All photographs must be gloss prints. All materials, including drawings and photographs, should be no larger than 175 × 260 mm.

Page number: Number all pages, including tables and illustrations (which should be grouped at the end), in a single series, with no omitted numbers.

Electronic form: A floppy disk together with the hard copy of the manuscript should be submitted. It is important to ensure that the diskette version and the printed version are identical. The diskette should be labelled with the following information: a) the operating system and word-processing software used, b) in case of UNIX media, the method of extraction (i.e. tar) applied, c) file name(s) related to manuscript. The diskette should be properly packed in order to avoid possible damage during transit**.**

Among various acceptable word processor formats, TEX and LATEX are preferable. The *Journal's* style file is available to authors.

Galley proofs: Proofs should be returned by authors as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within two weeks after receiving the offprints.

The copy of the **"***Journal***"** shall be provided to each author of papers.

Copyright: Manuscript submitted to this journal may not have been published and will not be simultaneously submitted or published elsewhere. Submitting a manuscript, the authors agree to automatically transfer the copyright for their article to the publisher if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and also all translation rights. No part of the present journal may be reproduced in any form nor transmitted or translated into a machine language without permission in written form from the publisher.

Biographies and photographs of authors are printed with each paper. Send a brief professional biography not exceeding 100 words and a gloss photo of each author with the manuscript.

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 2/2004

Regular papers

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Editorial Office

tel. +48(22) 872 43 88
tel./fax: +48(22) 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.itl.waw.pl/jtit