

Dimensioning Hierarchical Access Modules

Balázs Karlócai

RESEARCH ARTICLE

Received 2011-12-02, accepted 2012-09-21

Abstract

Network dimensioning has been in the center of interest ever since the earliest networking architectures were put to use in the seventies. Since then service providers try to serve an ever increasing traffic with a given QoS. As a result, one of the main challenges of network dimensioning is how to design minimum capacity and minimum HW complexity network access modules which can carry a given amount of traffic with some predefined QoS parameters. The QoS is usually measured by cell loss probabilities and average cell delays.

In this paper we provide dimensioning algorithms for Hierarchical Access Modules (HAMs) built up by subracks in a hierarchical manner and developed to provide different types of user-access to internet. The main objective is to design minimum complexity HAMs, which can serve a given population of users with a pre-negotiated level of QoS. Since the complexity and cost are measured by the number of racks and the link capacities used in HAM, optimal design is defined here as constructing a HAM which contains the smallest possible number of racks (cards) and the lowest capacity links. However, there are plentiful various topologies and corresponding link capacity arrangements which can serve a given population of users with a predefined QoS. As a result - when HAM is described as a multidimensional vector representing the topology the number of subracks and the corresponding link capacities - the optimal solution is to be sought in a large vector space. This casts dimensioning as a combinatorial optimization problem and we refer to this task as Node and Capacity Arrangement Problem (NCAP). In this paper, we develop such algorithms and provided numerical results for typical traffic loads and QoS parameters which can occur while using HAMs.

Keywords

network dimensioning · HAM · quality of service · capacity arrangement

Balázs Karlócai

Faculty of Information Technology, Pázmány Péter Catholic University, H-1083 Budapest, Práter u. 50/A., Hungary
e-mail: karlocai.balazs@itk.ppke.hu

1 Introduction

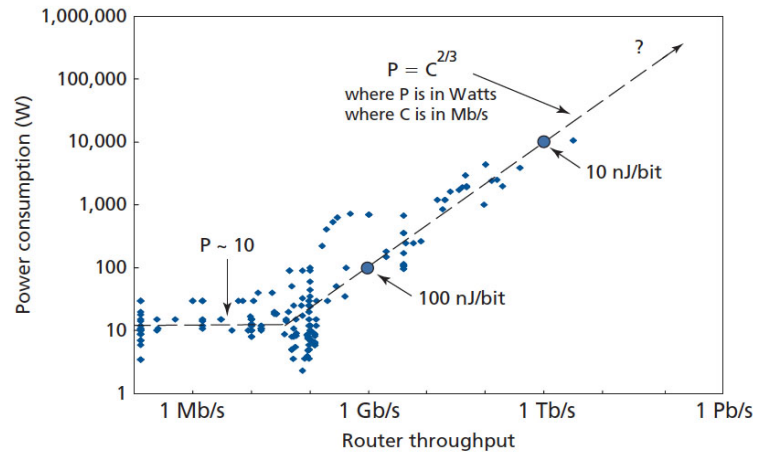
Ever since the introduction of the first packet switched networks teletraffic engineers have been engaged with the dimensioning problem. Due to the hunger for larger network traffic, users have been striving to have increased data speed which frequently surpasses the available network capacity. Therefore, dimensioning has become an extremely important problem [15]. Another angle of dimensioning has been presented by [20] when the relationship between power consumption (reflecting the complexity of networking architectures) and the router throughput has been investigated and the findings are described by Figure 1.

A Japanese research paper [17] claims that in 2015 the energy consumption of the network devices will take 10% of Japan total energy needs, and this number can reach 50% in 2020. This causes a huge cost for the network development and it conflicts with the the required energy-awareness and the sustainable development principles. This prompts the development of novel network planning methods which can serve the required traffic with minimum HW complexity devices and, as a result, with minimum power consumption.

However, up till now, there is not any standard procedure or metric for the dimensioning. There are many attempts for this, like IPCP (IP Capacity Planning), based on Frank Kelly's method, at Cambridge University, developed with IBM. This is a framework, which can be used to plan standard packet-switched networks (3G, ATM, MPLS, IP)[16]. Some other research papers [18, 19] focus on this problem by using real traffic traces, but no general method has emerged for network dimensioning, which also satisfies energy awareness.

When dimensioning HAM, one may think of a typical multi-node Access Network module which contains subracks for ADSL nodes where the number of subracks and the corresponding capacities are limited. Therefore, the main objective of the paper is to develop novel, multi-node dimensioning algorithms which are capable to handle these constrains. With these algorithms at hand, one can calculate what is the minimal number of nodes (subracks) and link capacities which are necessary to provide a given number of users with a given level of Cell Loss

Fig. 1. The energy consumption of the routers and the corresponding bandwidth



Source data: METI, 2006, Nordman, 2008

Probability (CLP).

This objective will be accomplished in the following manner:

- in *Section 2*, the purpose of dimensioning is briefly stated and a brief technological description of the investigated architecture is presented
- in *Section 3*, the specific task at hand is summarized mentioning the input parameters (traffic profiles, number of possible users ... etc.)
- in *Section 2*, an abstract model inferred from the technological parameters is given (the dimensioning algorithm will operate on the data structures delineated in this section)
- in *Section 4*, the algorithms used for single and multi-node dimensioning are described
- in *Section 5*, the numerical results calculated by the dimensioning algorithms are given together with some conclusions
- in *Appendix A* the Chernoff bound is discussed
- in *Appendix B* the mathematical details of multi-node dimensioning algorithms are elaborated

2 Technical description of HAMS

In this section we give a brief summary of HAMS and about the services the module must provide.

The users are usually grouped into three traffic classes:

- Internet Access1;
- Internet Access2;
- Voice over ADSL.

From traffic point of view, they are regarded as On/Off sources with the following typical parameters:

On/Off sources are characterized as binary i.i.d.r.v.-s, with the probability distribution

$$P(X = h) = \frac{m}{h}; P(X = 0) = 1 - \frac{m}{h}, \quad (1)$$

Tab. 1. On/Off sources with typical parameters

	mean(kBit/s)	peak(kBit/s)
Internet Access1 (uplink)	2	64
Internet Access1 (downlink)	20	384
Internet Access2 (uplink)	2	512
Internet Access2 (downlink)	20	2048
Voice over DSL (uplink)	24	40
Voice over DSL (downlink)	24	40

where h is the peak rate and m is the mean rate, respectively. The maximum number of supported users is typically in the range of a couple of thousands (in the numerical example we set this number to be 3000). Based on the typical user requirements, we assume four possible user (input load) configurations:

- 1 100% Internet Access1 users (uplink+downlink): $\mathbf{n} = (3000, 3000, 0, 0, 0, 0)$;
- 2 70% Internet Access1 users (uplink+downlink)+ 30% Voice over DSL users (uplink+downlink): $\mathbf{n} = (2100, 2100, 0, 0, 900, 900)$;
- 3 50% Internet Access1 users (uplink+downlink) + 20% Internet Access2 users (uplink+downlink) + 30% Voice over DSL users (uplink+downlink): $\mathbf{n} = (1500, 1500, 600, 600, 900, 900)$
- 4 70% Internet Access2 users (uplink+downlink)+ 30% Voice over DSL users (uplink+downlink): $\mathbf{n} = (0, 0, 2100, 2100, 900, 900)$

As far as its structure is concerned, HAM is regarded to be a set of nodes (racks) arranged in a tree topology as an example is depicted by Figure 2.

As mentioned before, optimal design boils down to find the topology containing the minimum number of nodes and minimum link capacities.

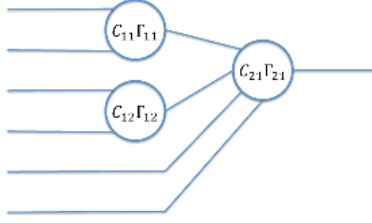


Fig. 2. Configuration of the nodes

The main QoS measure considered here is the Cell Loss Probability (CLP). In the case of internet access only 1 % of packets can be lost. The average size of a packet is 1500 bytes. The packets are split into ATM cells (48 bytes of payload is available in each cell). Based on this, an average TCP packet is carried by 32 ATM cells. Therefore, 1% TCP packet loss rate translates into $3.14 \cdot 10^{-4}$ cell loss probability at the ATM cell level, yielding $\gamma = -\ln 3.14 \cdot 10^{-4} = 8.067$ level of QoS. Instead of CLP we use parameter γ because it fits better to the notion of equivalent bandwidth.

3 Dimensioning as an algorithmic endeavour

Based on the discussion above, HAM is regarded as a set of nodes arranged in a tree topology, which can be represented as

$$HAM = \{V, E, \mathbf{C}, \mathbf{\Gamma}\}, \quad (2)$$

where V refers to the vertices, E denotes the edges, while the elements of matrices \mathbf{C} and $\mathbf{\Gamma}$ denote the capacities and QoS of a corresponding node in HAM as follows:

$$C_{kj} = C_j(k) \quad (3)$$

is the capacity associated to node j in layer k ;

$$\Gamma_{kj} = \gamma_j(k) \quad (4)$$

is the QoS requirement associated to node j in layer k .

One must note that when HAM is built up from several nodes then cells can get lost on each node. Therefore, there must be more stringent CLP requirement on a node-by-node basis than the one required from HAM as a whole. As a result, the overall CLP level should be "decomposed" into a CLP arrangements valid at the node level (each node has an associated CLP). Since the cells can be lost on any node along the path, the sum of the CLPs associated to the nodes should fulfill the overall CLP criterion. This will be discussed at a more formal level later. Furthermore, in reality capacities are associated with links instead of nodes. However, from the point of the computational model it does not make any difference associating capacities with nodes instead of links, on the basis of the node capacity taken to be equal with the output link capacity leaving that specific node.

In the forthcoming discussions we will use the following notations:

- traffic classes: $i = 1, \dots, M$;

- layers in the tree topology: $k = 1, \dots, K$;
- nodes in layer k : $l = 1, \dots, L_k$;
- admission vector of node j in layer k : $\mathbf{n}^j(k)$, where component $n_i^j(k)$ indicates the number of sources from class i
- the set of admissible vectors is termed as *Admission Set* (AS) containing traffic vectors which are associated with the corresponding nodes in the tree topology is defined as

$$AS = \{\mathbf{n}^l(k) \forall l = 1, \dots, L_k \forall k = 1, \dots, K\}; \quad (5)$$

- the input traffic state vector is given as

$$\mathbf{v}(1) = (\mathbf{n}^1(1), \mathbf{n}^2(1), \dots, \mathbf{n}^{L_1}(1)) \quad (6)$$

Note that there is relationship between the input state vector and AS, namely every input state vector can be decomposed into an AS by using the following definition:

Decomposition of the input state vector:

The decomposition of the input state vector $\mathbf{v}(1) = (\mathbf{n}^1(1), \mathbf{n}^2(1), \dots, \mathbf{n}^{L_1}(1))$ into an AS according to the flow of the graph is defined in the following way:

$$n_i^l(k) = \sum_{j \in A_l} n_i^j(k-1), \quad (7)$$

where A_l denotes the set of nodes in layer $k-1$ which are connected to node l in layer k .

It is clear that the decomposition defined above can be regarded as a mapping $V \rightarrow AS$, where the input is an input state vector $\mathbf{v}(1)$ and the output is an Admission Set denoted by $AS(\mathbf{v}(1))$.

The structure of HAM is represented by a topology matrix which can be described as follows:

$$G_{kl} = \begin{cases} 1 & \text{if there is a node at position } l \text{ in layer } k \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

One can arrange the topologies based on the number of nodes (i.e., number of 1s in matrix \mathbf{G}) according to some rule $\mathcal{G} = \{\mathbf{G}_{min}, \dots, \mathbf{G}_{max}\}$ (e.g., the matrices follow each other with respect to their binary weights and when two matrices have the same number of 1s, the matrix containing 1s with lower indices precedes the other). According to these conventions, \mathbf{G}_{min} is a minimal (one-node) configuration, whereas \mathbf{G}_{max} refers to the maximal topology (containing the largest number of nodes).

The QoS arrangement of HAM, in the case of a given topology \mathbf{G} , is denoted by a matrix $\mathbf{\Gamma}^G$ where element kl indicates the QoS parameter belonging to node l in layer k . If there is no node at position l in layer k then $\Gamma_{kl}^G = 0$, meaning that $G_{kl} = 0$ implies $\Gamma_{kl}^G = 0$. Furthermore, we assume that the possible QoS values form a discrete set $\gamma_1, \dots, \gamma_V$. Therefore, the set $\Gamma = \{\Gamma_{min}, \dots, \Gamma_{max}\}$ contains the possible QoS matrices. The matrix $\mathbf{\Gamma}_{min}^G$ is defined as $\Gamma_{kl} = \text{Min}$, while matrix

Γ_{max}^G is defined as $\Gamma_{kl} = \text{Max}$, where Min and Max are previously determined values. In this way searching for a proper QoS scheme, the dimensioning algorithm will sweep through the interval $G_{kl} \in (\text{Min}, \text{Max})$ for each node ($l = 1, \dots, L_k$ and $k = 1, \dots, K$).

The capacity arrangement of HAM is expressed by matrix \mathbf{C} . One must note that if $G_{kl} = 0$ then $C_{kl} = 0$, which means that capacity can only be allocated to existing nodes in the topology. A possible capacity matrix belonging to a topology \mathbf{G} is denoted by \mathbf{C}^G (where $C_{ij}^G \in \{C_1, \dots, C_R\}$). These matrices form a discrete space denoted by $\mathbf{C}^G = \{\mathbf{C}_{min}^G, \mathbf{C}_2^G, \dots, \mathbf{C}_{max}^G\}$. Where $\mathbf{C}_{min}^G : C_{ij} = C_1 G_{ij} \forall i, j$ is the network topology containing minimum capacity nodes and $\mathbf{C}_{max}^G : C_{ij} = C_R G_{ij} \forall i, j$ is the same topology but containing maximum capacity nodes. Since there is a finite number of possible capacities the programmer can order the set \mathbf{C}^G according to any arbitrary rules. (We adapted the ordering scheme which is based on the sum of the elements and on the rank of indices of the corresponding matrices.)

4 Algorithms for dimensioning

In order to come to grasp with dimensioning, first we describe the single-node dimensioning algorithm. Here the objective is to find the minimum capacity which can serve a given traffic mixture with a pre-defined CLP. The results are derived by using the Chernoff bound and the log-moment generating functions (see Appendix A). Further details can be found in the papers [1, 4–7, 9, 11, 12, 14].

4.1 Single node dimensioning algorithm

Let us assume that source j from traffic class i presents a random traffic load denoted by $X_j^{(i)}$. This bound on the tail of the aggregate traffic is given as

$$P\left(\sum_{i=1}^M \sum_{j=1}^{n_i} X_j^{(i)} > C\right) \leq e^{\sum_{i=1}^M n_i \mu_i(s) - sC}, \quad (9)$$

where

$$\mu_i(s) = \log\left(1 - \frac{m_i}{h_i} + \frac{m_i}{h_i} e^{sh_i}\right) \quad (10)$$

is the logarithmic moment generating function of an On/Off source belonging to class i , and n_i denotes the number of sources present from class i . Therefore, guaranteeing a γ level of QoS means to enforce

$$e^{\sum_{i=1}^M n_i \mu_i(s^*) - s^*C} < e^{-\gamma},$$

or consequently

$$\sum_{i=1}^M n_i \mu_i(s^*) < s^*C - \gamma, \quad (11)$$

where

$$s^* : \min_s \sum_{i=1}^M n_i \mu_i(s) - sC$$

As a result, the single-node dimensioning algorithm based on the notion of effective bandwidth is given as follows:

Single-node dimensioning algorithm

Given a set of discrete capacities $\mathbf{C} = \{C_1, \dots, C_R\}$ $C_1 < C_2 < \dots < C_R$, an input load expressed by traffic configuration vector $\mathbf{n} = (n_1, \dots, n_M)$, and a CLP level γ as the QoS parameter.

Set $C := C_1$ and $r := 1$.

Calculate the logarithmic moment generating functions $\mu_i(s)$ $i = 1, \dots, M$.

- 1 Determine $s_{opt} : \inf_s \sum_{i=1}^M n_i \mu_i(s) - sC$
- 2 Check whether $\sum_{i=1}^M n_i \mu_i(s_{opt}) < s_{opt}C - \gamma$ holds.
- 3 If YES then return with C if NOT then set $r := r + 1$ and go back to Step 1.

With this algorithm one finds the minimal capacity C_{min} which is sufficient enough to accommodate the load vector \mathbf{n} at a γ level of QoS.

4.2 Multinode dimensioning algorithms

In the case of dimensioning the designer's task is to find a topology of HAM with the corresponding capacities which fulfill a given overall QoS parameters for a given load vector. Therefore, we seek a mapping from the input load vector $\mathbf{v}(1)$ and end-to-end QoS requirement to a $G_{opt}(V, E, \mathbf{C}, \Gamma)$. This optimization problem, referred to as NCAP, can be formally defined as

$$G_{opt}\{V, E, \mathbf{C}, \Gamma\} = \Psi(\mathbf{v}(1), \gamma); \quad (12)$$

where

$$G_{opt}(V, E, \mathbf{C}, \Gamma) : \min_{G(V, E, \mathbf{C}, \Gamma)} \sum_{k=1}^K L_k. \quad (13)$$

However, one must pay attention to the fact that there are many different Γ matrices which can fulfill an overall QoS requirement γ . Therefore, not only the capacity arrangement but also a QoS arrangement must be given, which identifies how the overall QoS parameter is "distributed" among the nodes of a given topology.

To yield a solution, we take a recursive approach to the problem. Namely, we start with a minimal configuration $G\{V, E, \mathbf{C}, \Gamma\}$ (containing the smallest number of nodes) then we check whether the required QoS level is met or not. If not, we enlarge this configuration by adding nodes and check the QoS requirement until for the given input configuration $\mathbf{v}(1)$ the overall CLP is met. Since we start with the smallest capacity arrangement and continuously enlarge it this algorithm will find the optimal solution for a given input configuration. In order to put this plan at work, the following procedure must be carried out:

Pick a minimal topology and a corresponding capacity scheme and QoS arrangement.

- decompose the input state vector into an AS;
- check the capacity constrains given by the current capacities node by node.

- If the present capacity arrangement fulfills the QoS requirement then NCAP is solved. If not then enlarge the topology and change the capacity and QoS arrangement, and return to the calculation again.

This gives rise to the following procedure:

Given an input state vector $\mathbf{v}(1) = (\mathbf{n}^1(1), \mathbf{n}^2(1), \dots, \mathbf{n}^{L_1}(1))$ a Capacity Arrangement $C_l(k)$, $l = 1, \dots, L_k$, $k = 1, \dots, K$ with minimal topology, and a matrix of logical variables \mathbf{T} the T_{kl} element of which indicates whether the QoS criterion on node l in layer k is met or not.

1 Decompose $\mathbf{v}(1)$ into a corresponding AS($\mathbf{v}(1)$).

2 Based on the

$$AS = \{\mathbf{n}^l(k) \forall l = 1, \dots, L_k \forall k = 1, \dots, K\},$$

obtained in the first step, check the following set of inequalities

$$T_{lk} = \begin{cases} TRUE \text{ if} \\ P\left(\sum_{i=1}^M \sum_{j=1}^{n_i^l(k)} X_i > C_l(k)\right) < e^{-\gamma_l(k)} \\ FALSE \text{ if} \\ P\left(\sum_{i=1}^M \sum_{j=1}^{n_i^l(k)} X_i > C_l(k)\right) \geq e^{-\gamma_l(k)} \end{cases} \quad (14)$$

for each $l = 1, \dots, L_k$ and $k = 1, \dots, K$.

3 The input traffic vector $\mathbf{v}(1) = (\mathbf{n}^1(1), \mathbf{n}^2(1), \dots, \mathbf{n}^{L_1}(1))$ is accepted if $\bigcap_{k=1}^K \bigcap_{l=1}^{L_k} T_{lk} = TRUE$ then accept the given Capacity arrangement otherwise enlarge the topology and go back to Step 1.

This algorithm defines a mapping over the input state vector space to the $G_{opt}\{V, E, \mathbf{C}, \mathbf{\Gamma}\}$. When using the Chernoff inequality to calculate the tail distribution, the exact dimensioning algorithm is given as follows:

Multi-node dimensioning algorithm

Given a traffic configuration at the input of all nodes in the first layer

$$\mathbf{v}(1) = (\mathbf{n}^1(1), \mathbf{n}^2(1), \dots, \mathbf{n}^{L_1}(1)) \quad (15)$$

and a matrix of logical variables \mathbf{T} the T_{kl} element of which indicates whether the local QoS criterion on node l in layer k is met or not.

An overall logical variable U indicating whether the overall QoS requirement is met or not.

1 Set $\mathbf{G} = \mathbf{G}_{min}$;

2 Decompose $\mathbf{v}(1)$ into an admissible set $AS(\mathbf{v}(1)) = \{\mathbf{n}^l(k), l = 1, \dots, L_k, k = 1, \dots, K\}$;

3 Set $\mathbf{C}^G := \mathbf{C}_{max}^G$;

4 Set $\mathbf{Q}^G := \mathbf{Q}_{max}^G$;

5 Calculate $s_{l\ opt}(k)$ by solving

$$s_{l\ opt}(k) : \sum_{i=1}^M n_i^l(k) \frac{d\mu_i(s)}{ds} = C_l(k)$$

$$\forall l = 1, \dots, L_k \quad k = 1, \dots, K$$

6 If

$$T_{lk} =$$

$$\begin{cases} TRUE, \text{ if} \\ \sum_{i=1}^M n_i^l(k) \mu_i(s_{l\ opt}(k)) < s_{l\ opt}(k) C_l(k) - \gamma_l(k) \\ FALSE, \text{ if} \\ \sum_{i=1}^M n_i^l(k) \mu_i(s_{l\ opt}(k)) > s_{l\ opt}(k) C_l(k) - \gamma_l(k) \end{cases} \quad (16)$$

7 Calculate $U := \bigcap_{k=1}^K \bigcap_{l=1}^{L_k} T_{lk}$.

8 If $U = FALSE$ then enlarge the topology by setting $\mathbf{G} := \mathbf{G}_2$ and go back to Step 3 and repeat this loop until $U = TRUE$

9 If $U = TRUE$ then topology is found but relax QoS requirements by choosing another $\mathbf{Q}^G \in \mathbf{Q}^G$ by setting $\mathbf{Q}^G := \mathbf{Q}_2$ and go back to Step 5 and repeat this loop until $U = FALSE$ or $\sum_{k=1}^K \sum_{l=1}^{L_k} e^{-\gamma_l(k)} > e^{-\gamma}$ then return to the previous value of \mathbf{Q}^G

10 If $U = TRUE$ then topology is found, but decrease capacity by choosing another $\mathbf{C}^G \in \mathbf{C}^G$ by setting $\mathbf{C}^G := \mathbf{C}_2$ and go back to Step 4 and repeat this loop until $U = FALSE$ then return to the previous value of \mathbf{C}^G

11 Return with matrices $\mathbf{G}, \mathbf{C}^G, \mathbf{Q}^G$ which completely determine $G_{opt}\{V, E, \mathbf{C}, \mathbf{\Gamma}\}$

One can see that this algorithm returns the optimal HAM indeed, as not only the number nodes are minimized (finding the smallest size topology) but, at the same token, the corresponding capacity and QoS arrangements, as well. In this way we find the least stringent conditions in which the minimal topology network can serve the balanced traffic load vector with the given overall QoS.

5 Numerical results

In this section the results of the dimensioning algorithms are discussed and a minimum complexity HAM is designed for the specific traffic case elaborated in Section 3. Here we present the optimal topology with the associated capacities and QoS parameters.

Since we have six classes (internet Access1, Internet Access2, Voice over DSL) the input load configuration (or traffic configuration) is a six dimensional vector $\mathbf{n} = (n_1, n_2, n_3, n_4, n_5, n_6)$ where component n_i denotes the number of users from the corresponding traffic class.

It also useful to recall that the input traffic volume is expressed as a number of input traffic configuration vectors for the given QoS requirement is met. More precisely, the input traffic volume is simply the size of the admissible input vector set. Table 2 shows the optimal topologies found for the traffic load configurations (each the traffic load configuration is expressed by the number of users from the particular traffic class):

One must not forget that only 10 % of the users are actively engaged with transmitting or receiving cells.

Tab. 2. Optimal topologies

100% Internet Access1 users (uplink+downlink):	$\mathbf{n} = (3000, 3000, 0, 0, 0, 0)$
70% Internet Access1 users (uplink+downlink)+ 30% Voice over DSL users (uplink+downlink):	$\mathbf{n} = (2100, 2100, 0, 0, 900, 900)$
50% Internet Access1 users (uplink+downlink) + 20% Internet Access2 users (uplink+downlink) + 30% Voice over DSL users (uplink+downlink):	$\mathbf{n} = (1500, 1500, 600, 600, 900, 900)$
70% Internet Access2 users (uplink+downlink)+ 30% Voice over DSL users (uplink+downlink):	$\mathbf{n} = (0, 0, 2100, 2100, 900, 900)$

Tab. 3. optimal HAM

	number of subracks	QoS parameter((γ_1, γ_2))	capacity(C_1, C_2)	Traffic Volume
1.(100%)	1st layer 12	1st layer 7.895	5.23 Mbit/s	$2.50 * 10^2$
	2nd layer 1	2nd layer 9.215	12.21 Mbit/s	
2.(70-30)	1st layer 12	2nd layer 7.895	5.23 Mbit/s	$3.63 * 10^4$
	2nd layer 1	2nd layer 9.215	12.21 Mbit/s	
3.(50-20-30)	1st layer 12	1st layer 7.895	6.98 Mbits/s	$4.04 * 10^6$
	2nd layer 1	2nd layer 9.215	19.2 Mbit/s	
4.(70-30)	1st layer 12	1st layer 7.895	8.72 Mbit/s	$6.01 * 10^4$
	2nd layer 1	2nd layer 9.215	22.69 Mbit/s	

Running the dimensioning algorithm given in the previous section, the optimal HAM is obtained as given in the next table (Table 3):

The corresponding topology is depicted by Figure 3.

One can see that for different output vectors only the capacity and QoS arrangements are different. The next figure indicates the average capacity need ($\frac{1}{\sum_{k=1}^K L_k} \sum_{k=1}^K \sum_{l=1}^{L_k} C_l(k)$) of the four different input scenarios. One can also see that the fourth one needs the largest average capacity. From the barchart it is clear that configuration 4 represents the most stringent capacity requirements.

6 Conclusions

In the paper a novel algorithm was developed for network access module dimensioning to optimize the topology and to achieve minimum capacity design. After the formalizing the task, and solving the node Node Capacity Arrangement Problem a directed search algorithm were as used to find the optimal HAM. The algorithm has been tried out for typical traffic scenarios where the limits of subracks architecture were also taken into account. The methods developed in the paper can be applied to other dimensioning problems required by QoS communication in packet switched networks.

A The Chernoff bound

Chernoff bound is one of the basic tools of large deviation theory [5, 8]. This bound estimates the tail of a positive valued random variable denoted by Y in the following form [5, 9, 10]:

$$P(Y > C) \leq e^{\mu_Y(s) - sC}. \quad (17)$$

Here $\mu_Y(s) = \log E(e^{sY})$ is the logarithmic moment generating function of random variable Y [5, 8], which is often referred to as "effective bandwidth" (if divided by s) and s is an arbitrary positive value. The sharpest bound is obtained by choosing

$$s_{opt} : \inf_s \mu_Y(s) - sC. \quad (18)$$

In order to fully elaborate on the use of Chernoff bound, let us assume that $Y = \sum_{i=1}^M \sum_{j=1}^{n_i} X_j$ is the aggregated load. Then the Chernoff bound takes the form of

$$P\left(\sum_{i=1}^M \sum_{j=1}^{n_i} X_j > C\right) \leq e^{\sum_{i=1}^M n_i \mu_i(s) - sC}, \quad (19)$$

where

$$\mu_i(s) = \log\left(1 - \frac{m_i}{h_i} + \frac{m_i}{h_i} e^{sh_i}\right) \quad (20)$$

is the logarithmic moment generating function of an On/Off source belonging to class i . As a result, one can calculate the number of admissible load vectors (which are provided with a γ level of QoS) by using the following algorithm:

For each possible load vectors \mathbf{n} set $s_{opt} : \sum_{i=1}^M n_i \frac{d\mu_i(s)}{ds} = C$ then check whether

$$e^{\sum_{i=1}^M n_i \mu_i(s_{opt}) - s_{opt}C} < e^{-\gamma},$$

or consequently

$$\sum_{i=1}^M n_i \mu_i(s_{opt}) < s_{opt}C - \gamma. \quad (21)$$

Then count the number of admissible load vectors which yield the admissible load "volume".

B Multi-node dimensioning in a tree topology

As was mentioned before multi-node dimensioning differs from single node dimensioning as nodes in higher layers introduce additional capacity constrains. Therefore, the application

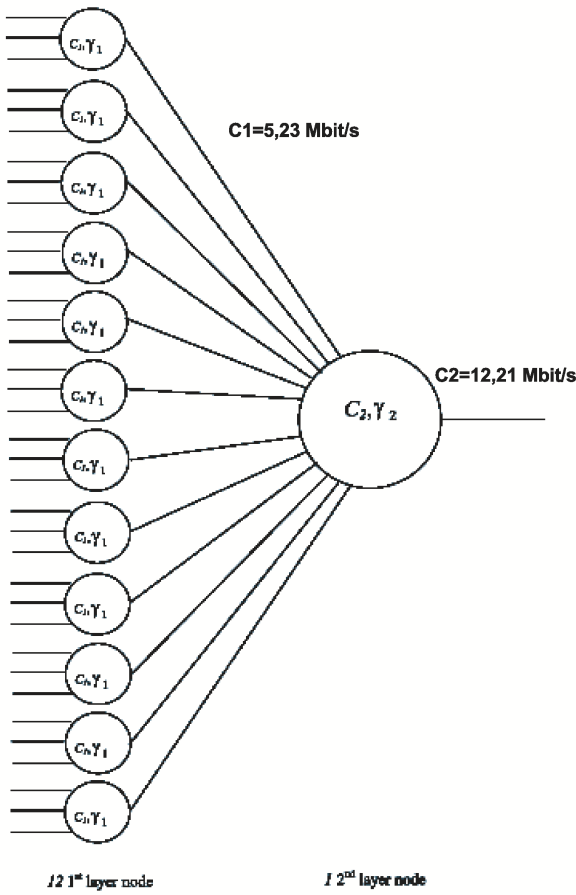


Fig. 3. Optimal network topology obtained by algorithm for balanced load according to Method 1

of Chernoff bound needs further considerations. In order to get an insight into these problems, first we investigate dimensioning on a "tree-primitive" containing only three nodes, two in the lower layer and one in the upper layer. The overall input load vector is $\mathbf{v}(1)$. The decomposed traffic state vectors are $\mathbf{n}^1(1), \mathbf{n}^2(1), \mathbf{n}(2)$, respectively. (Note that since there is a single node in layer 2, there is no need for upperindex in $\mathbf{n}(2)$). There are several different ways how the overall QoS can be distributed among the nodes given as $\gamma_1(1), \gamma_2(1), \gamma(2)$. In order to fulfill the overall QoS, the following equations must hold:

$$e^{-\gamma_1(1)} + e^{-\gamma_2(1)} \leq e^{-\gamma}$$

and

$$e^{-\gamma_2(1)} + e^{-\gamma(2)} \leq e^{-\gamma}.$$

If one only aims at finding the minimal capacities to accommodate the balanced load $\mathbf{v}_b(1)$ then the following algorithm is to be carried out:

Given $\mathbf{v}_b(1)$ and γ

- 1 Decompose AS $\{\mathbf{n}_b^1(1), \mathbf{n}_b^2(1), \mathbf{n}_b(2)\}$ from $\mathbf{v}_b(1)$.
- 2 Choose a capacity arrangement and QoS arrangement $C_1(1), C_2(1), C(2)$ and $\gamma_1(1), \gamma_2(1), \gamma(2)$ respectively, subject to

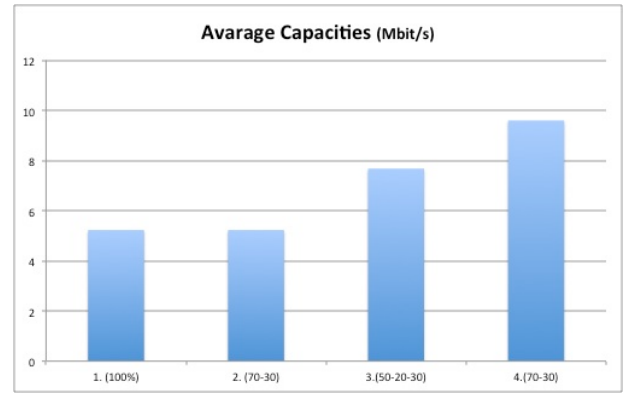


Fig. 4. Average capacities for the different traffic mixes

the following constrains

$$e^{-\gamma_1(1)} + e^{-\gamma_2(1)} \leq e^{-\gamma}$$

$$e^{-\gamma_2(1)} + e^{-\gamma(2)} \leq e^{-\gamma}$$

- 3 Calculate $s_{1\text{opt}}(1), s_{2\text{opt}}(1), s_{\text{opt}}(2)$ by solving the equations

$$s_{1\text{opt}}(1) : \sum_{i=1}^M n_{ib}^1(1) \frac{d\mu_i(s)}{ds} = C_1(1)$$

$$s_{2\text{opt}}(1) : \sum_{i=1}^M n_{ib}^2(1) \frac{d\mu_i(s)}{ds} = C_2(1)$$

$$s_{\text{opt}}(2) : \sum_{i=1}^M n_{ib}(2) \frac{d\mu_i(s)}{ds} = C(2)$$

- 4 check the following inequalities

$$\sum_{i=1}^M n_{ib}^1(1) \mu_i(s_{1\text{opt}}(1)) \leq s_{1\text{opt}}(1) C_1(1) - \gamma_1(1)$$

$$\sum_{i=1}^M n_{ib}^2(1) \mu_i(s_{2\text{opt}}(1)) \leq s_{2\text{opt}}(1) C_2(1) - \gamma_2(1)$$

$$\sum_{i=1}^M n_{ib}(2) \mu_i(s_{\text{opt}}(2)) \leq s_{\text{opt}}(2) C(2) - \gamma(2)$$

- 5 if the inequalities are satisfied then the capacities are found if not, then increase the capacities $C_1(1), C_2(1), C(2)$ and the QoS values $\gamma_1(1), \gamma_2(1), \gamma(2)$.

References

- 1 **de Prycker M.**, *Asynchronous Transfer Mode Solution for Broadband ISDN*, Ellis Harwood Ltd., 1991.
- 2 **Sohrabry K.**, *Flow Admission Control of ON-OFF Sources in High Speed Networks*, GLOBECOM, (1994).
- 3 **Eckberg A., Doshi BT, Zoccolillo R.**, *Controlling congestion in B-ISDN/ATM: issues and strategies*, IEEE Communications Magazine, **29**(9), (September 1991), 64–70, DOI 10.1109/35.90495.
- 4 **Hui JY.**, *Resource allocation for broadband networks*, IEEE Journal on Selected Areas in Communications, **6**(9), (December 1988), 1598–1608, DOI 10.1109/49.12887.

- 5 **Hui JY**, *Switching and Traffic Theory for Integrated Broadband Networks*, The Kluwer International Series in Engineering and Computer Science, Vol. 91, Kluwer Academic Publishers, 1990, DOI 10.1007/978-1-4615-3264-4.
- 6 **Yasuhiro M**, *A Dimensioning Scheme in ATM networks*, Networks'92, (May 1992), 171–176.
- 7 **Hiramatsu AYuhas B, Ansari N** (eds.), *ATM Traffic Control Using Neural Networks*, Neural Networks in Telecommunications, Kluwer Academic Publishers, 1994, DOI 10.1007/978-1-4615-2734-3_4.
- 8 **Kelly F**, *Notes on effective bandwidth*, INFORMS Telecommunications Conference, (March, 1995).
- 9 **Levendovszky J, Imre S**, *Comparative analysis of CAC algorithms for ATM networks*. Scientific Progress Report, COPP579-SPR-1, EU.
- 10 **Levendovszky J, van der Meulen EC**, *Tail Distribution Estimation for Call Admission in ATM Networks*, Proceedings of IFIP, Third Workshop on Performance Modelling and Evaluation of ATM Networks, (2-6th July 1995).
- 11 **Levendovszky J, Elek Z, Vegso C**, *Validating novel CAC algorithms on ATM testbeds*, 1999 2nd International Conference on ATM. ICATM'99 (Cat. No.99EX284), (1999), 195–211, DOI 10.1109/ICATM.1999.786803.
- 12 **Levendovszky J**, *Call admission control of ATM networks based on modulated Markov chains*, Journal on Communication dedicated to ATM Networks, XLVII, (March 1995), 19–24.
- 13 **Levendovszky J**, *Neuron based penalty function classifiers*, 19th WIC Conference, (May, 1998).
- 14 *Nonparametric decision algorithms for CAC in ATM networks*, Performance Evaluation, 41(2-3), (2000), 133–147, DOI 10.1016/S0166-5316(00)00013-4.
- 15 **Bache A, Buillou L, Layec H, Lorig B, Matras Y**, *RCP, The Experimental Packet-Switched Data Transmission Service of the French PTT: History, Connections, Control*, Proceedings of the Third International Conference on Computer Communication (ICCC), (1976), 3–6.
- 16 **Davies G, Hardt M, Kelly F**, *Come the revolution—network dimensioning, service costing and pricing in a packet switched environment*, Telecommunications Policy, 28(5-6), (2004), 391–412, DOI 10.1016/S0308-5961(04)00044-8.
- 17 **Imaizumi H, Morikawa H**, *Directions Towards Future Green Internet*, Towards Green Ict, 9, (2010), 37.
- 18 **Sharafeddine S, Dawy Z**, *Robust network dimensioning for realtime services over IP networks with traffic deviation*, Computer Communications, 33(8), (2010), 976–983, DOI 10.1016/j.comcom.2010.01.020.
- 19 **Soldatos J, Vayias E, Stathopoulos P, Mitrou N**, *Enforcing Effective Rates for Packet-Level QoS Control in IP Networks: Theory and Validation Based on Real Traffic Data*, Telecommunication Systems, 27(1), (2004), 9–31, DOI 10.1023/B:TELS.0000032941.38060.01.
- 20 **Tucker RS, Baliga J, Ayre R, Hinton K, Sorin WV**, *Energy consumption in IP networks*, 2008 34th European Conference on Optical Communication, (2008), 1, DOI 10.1109/ECOC.2008.4729102.