

Adaptive Bandwidth Allocation in Broadband Networks

G. Lauks

*Faculty of Electronics and Telecommunications, Riga Technical University,
Azene str. 12, Riga, Latvia, LV10-48, e-mail: lauks@rsf.rtu.lv*

Introduction

New broadband network applications require certain performance guarantees that can be provided through proper resource allocation. Allocation techniques are needed to provide these guarantees as efficiently as possible since resources are limited.

MPLS and DiffServ have been widely accepted as the service model to adopt for providing Quality of Services (QoS) over next generation IP networks. The combination of MPLS and DiffServ presents a very attractive strategy to broadband network providers.

DiffServ provides scalable and “better than best effort” QoS. DiffServ routers is stateless and do not keep track of individual flows, making it scalable to be deployed in the Internet. Generally, there are three classes of services: high priority (HP), normal priority (NP) and best effort priority (BE) classes. Current Bandwidth Allocation methods can be categorized as either off-line or on-line. Off-line methods predetermine allocation amounts before transmission begins. Such a method may allocate one resource level for the duration of the application, or may renegotiate the resource level at various times. In [2] and previous works of authors has shown that it is generally favorable for both customer and provider to allow renegotiation of bandwidth allocation. The customer saves costs during phases of low demand and provider can make better use of the capacity of the network. Disadvantage of off-line methods is static bandwidth allocation. For interactive applications, where the traffic source is not known nor directly controllable, these methods are not suitable.

On-line methods periodically renegotiate resource allocation based upon predicted traffic behavior. Predictions of traffic are derived from measurements and QoS observations. There are lot of methods for solving of Bandwidth Allocation problem. In [5] was proposed an adaptive provisioning mechanism for Bandwidth Allocation that determines at regular intervals the amount of bandwidth to provision for each Per Hop Behavior (PHB) aggregate , based on traffic conditions and feedback received about the extend to which QoS is being met. The mechanism adjusts to minimize a penalty function that is

based on the QoS requirements agreed upon in the service level agreement (SLA). The idea is to use reinforcement learning (RL) algorithm to solve the Bandwidth Allocation Problem dynamically in discrete time intervals. Reinforcement learning provides the relationships between network state, set of actions and penalty function. Here Bandwidth Broker is used as Agent between Ingress, Core and Egress Routers. Relationships between amount of Bandwidth and traffic conditions are modeled with neural network (NN). To represent the relationship between state and action authors use a simple Multi-layer Perceptron NN with 2 layers. The Learning algorithm is based on iterative gradient-descend method. The exponential penalty function for each traffic class describes bounds of traffic losses and bounds of traffic delay as a arguments. The simulation results show that RL is able to adapt well to changing traffic conditions, as well as various QoS specifications. The proposed method utilizes only one objective and constraint of problem. The significant variable in Bandwidth Allocation Problem is a cost function. For correctness of problem statement it is necessary to define (as minimum) bandwidth utilization variable as a representative of cost function. The problem here is that without pre-knowledge or experience of the network dynamics, utility functions are difficult to establish. One of possible solutions can be use a fuzzy approach to model these functions. It is clear that Bandwidth provisioning problem is a hard problem.

IETF Differentiated Services Working Group in [1] presents functional specification and defines an experimental protocol for Bandwidth Allocation DiffServ traffic engineering (DS-TE). Today under discussion we have tree bandwidth constraints models: maximum allocation model (MAM), Russian Doll Model (RDM) [4] and Maximum Allocation with reservation [MAR]. Accordingly to simulation results [1] MAR bandwidth Allocation achieves greater efficiency in bandwidth sharing while still providing bandwidth isolation and protection against QoS degradation. The performance analysis of MAR Bandwidth Allocation methods is based on large, 135-node USA National Network simulation results. The objectives of this paper are to simulate and estimate the efficiency of MAR in small network cases. In

this case it is essential to estimate Bandwidth utilization.

Problem statement

Let us assume the following notations:
 c – class-type as a set of Traffic Trunks crossing a link that is governed by a specific set of bandwidth constraints. (Up to 8 classes are supported DiffServ).

$$1 \leq c \leq 8;$$

BC_c – Bandwidth Constraint for class c (Each class is assigned either a Bandwidth Constraint, or a set of Bandwidth Constraints. Up to 8 Bandwidth Constraints are supported);

$BW \max_i$ – maximum reserved bandwidth on link k specifies the maximum bandwidth that may be reserved;

$A_{HP,k}(n), A_{NP,k}(n), A_{BE,k}(n)$ – Traffic flow from high, normal and best effort priority classes, respectively.

Generally the bandwidth allocation problem can be formulated as follows:

For all nodes,

Given,

$$A_{HP,k}(n), A_{NP,k}(n), A_{BE,k}(n),$$

$k \in K$: set of outgoing links from node,

C_k – Capacity of link k ;

Select,

$P_{HP,k}, P_{NP,k}, P_{BE,k}$ – weighted fair proportions of bandwidth for each class on link;

Subject to QoS constraint

$$L_{HP,k} \leq Const1, L_{NP,k} \leq Const2, L_{BE,k} \leq Const3,$$

where Const1, Const2, Const3 – constants accordingly to SLA.

Maximum Allocation with Reservation Model

In this model [1] the following terms are used:

BW_k - bandwidth-in-progress on the link k ;

BW_{avg_k} - minimum guaranteed bandwidth required for the link k to carry the average offered bandwidth load;

$BW \max_k$ - the bandwidth required for the link k to meet the blocking/delay probability grade of service objective for CRLSP bandwidth allocation requests;

DBW - delta bandwidth requirement for a bandwidth allocation request;

$Rthr_k$ - reservation bandwidth threshold for link k ;

φ_c - empirical factor that results some “over allocation” of the maximum reserved bandwidth;

A_c - forecast or measured traffic load bandwidth for c on link k ;

BW_r_c - reserved bandwidth-in-progress on c on link k ;

BW_c - total amount of the bandwidth reserved by all the established LSPs that belong to c ;

BW_{unr} - unreserved link bandwidth on link k specifies the amount of bandwidth not yet reserved for any c ;

BW_{unr_c} – unreserved link bandwidth on link k specifies the amount of bandwidth not yet reserved for c .

The expressions for variables are

$$BW_{unr_c} = BW_{unr} - \Delta * Rthr, \quad (1)$$

where values of Δ are defined in Table 1.

Table 1. Values for increment Δ

Conditions	Value of Δ
$IF(BW_r_c < BC_c)$	0
$IF(BW_r_c \geq BC_c)$	1

Unreserved Bandwidth is calculated as follows

$$BW_{unr} = BW \max - \sum_{c=1}^8 BW_r_c. \quad (2)$$

Rule for admitting LSP Bandwidth Request is expressed by (3).

$$IF(DBW_c \leq BW_{unr_c}) THEN (Admission = TRUE). \quad (3)$$

Bandwidth constraints for traffic classes are given as defined in Table 2.

Table 2. Calculation of Bandwidth Constraints

Class	BW_{avg_i}	φ_c
High priority	$BC(c, k) = \varphi_i p_c$	2
Normal priority	$BC(c, k) = \varphi_i p_c$	1
Best-effort priority	$BC(c, k) = \varphi_i p_c$	0

where weighted fair proportions of bandwidth for each class on link are calculated by (4).

$$p_c = \frac{A_c}{\sum_{j=1}^{MaxCT} A_j} * BW \max_c. \quad (4)$$

As described above MAR leaves out of consideration any variables of cost function.

Case of multiple objectives

In the case of multiple objectives a Bandwidth allocation problem is to find the set of the Pareto optimal solutions, and from this set select the optimum solution as a decision at time instant. Generally the bandwidth allocation as a multi-objective optimization problem can be written as:

$$\text{Min/max } f_m(X), m = 1, 2, \dots, M. \quad (5)$$

Subject to constraint

$$g_k(X) \leq c_k, k = 1, 2, \dots, K, \quad (6)$$

where $X = (x_1, x_2, \dots, x_N)$ is an N -tuple vector of variables; $F = (f_1, f_2, \dots, f_M)$ is an M -tuple vector of objectives.

Typically, the optimal solution in case of multiple objectives (Pareto front) is processed between pricing variables, QoS variables and resources as variables.

Pricing for dynamically allocated resources are studied in [8]. Previous research work [7] has shown that for multiple objective approaches optimization can be achieved using genetic algorithms. As shown in [8], for pricing of resources it is necessary to take in account set of undertaking specific variables.

In this paper for analysis of Bandwidth allocation methods is used only one parameter – Bandwidth utilization.

Simulation Setup and details

The aim of this simulation is to acquire the data for analysis of MAR method in small network cases and estimate the bandwidth utilization for different priority services.

Traffic forecasting model

Internet data traffic exhibits burst feature at multiple timescales. Therefore, traffic prediction is one of significant bandwidth allocation problems. If the traffic predictor on a given sampling window underestimate the bandwidth requirement that varies at a shorter time-scale, resulting in possible violation of QoS guaranties. If the predicted bandwidth overestimates the actual bandwidth required, it results in inefficient resource utilization. Here is assumed that routers can measure mean and standard deviation of the aggregate priority traffic for different times.

When the number of aggregated traffic flows of the same class stands large, the aggregate arrival rate tends to have a Gaussian distribution under Central Limit Theorem. As recommended in [3], for instantaneous traffic load a variation, the load is typically modeled as a stationary random process over a given period characterized by a fixed mean and variance.

The good prediction of traffic descriptors can be achieved with multi-layer Perceptron NN with two layers. In simulated model the Learning algorithm is based on iterative gradient-descend method [5].

The values of traffic descriptors were selected on QoS bounds. Standard deviation was selected for large scale networks (typically $\sigma=1$) and small scale networks ($\sigma=2, 3$).

Simulation Results

In the first experiment there was set out the MAR Bandwidth Allocation with traffic prediction using NN and without prediction using only measured means. As shown in Fig. 1 significant role traffic predictor plays for NP and BE services, because for HP services factor $\varphi_c=2$ (it means overestimation). Range of Bandwidth utilization where traffic forecasting reduce non-admitted LSP is between 0.7- 0.9. In overload regime traffic losses growth very fast and traffic forecasting stands non-significant. Traffic flows were simulated as three independent

Gaussian processes. Simulation results of first experiment are shown in Fig. 1. For HP services traffic prediction is not significant.

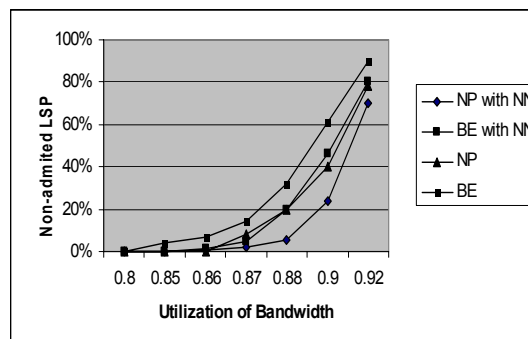


Fig. 1. Impact of traffic forecasting to QoS and bandwidth utilization

The second experiment is focused to study of influence of standard deviation on QoS and Bandwidth utilization using traffic predictors.

Simulation results for NP and BE services are presented in Fig. 2 and Fig. 3, respectively.

For NP services and standard deviation $\sigma=1$ the range of bandwidth utilization is widened. (It is traditional advantages of large scale networks)

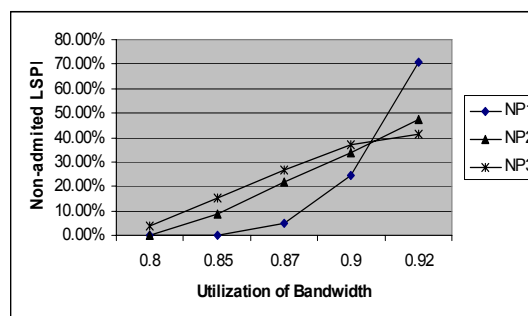


Fig. 2. Impact of standard deviation for normal priority (NP) services: for NP1 series $\sigma=1$, for NP2 series $\sigma=2$ and for NP3 series $\sigma=3$

It means that MAR works properly for large scale networks and small values of standard deviation ($\sigma=1$).

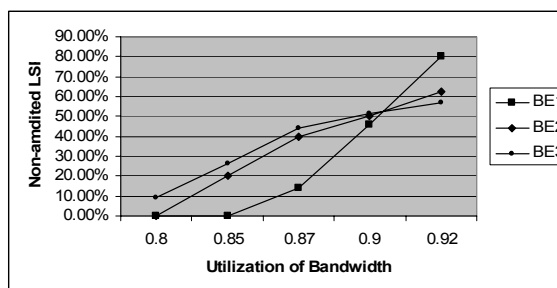


Fig. 3. Impact of standard deviation for best effort priority (BE) services: for BE1 series $\sigma=1$, for BE2 series $\sigma=2$ and for BE3 series $\sigma=3$

Conclusions

In this paper MAR Bandwidth Allocation methods are studied for large scale and small scale network applications. Adaptive bandwidth allocation means here a maximum allocated bandwidth with reservation [MAR] and adaptive traffic prediction under dynamically changed arrival processes.

Problem statement of MAR is enhanced with bandwidth utilization variables. The simulation results demonstrate the effectiveness of traffic predictor based on neural network approach for all types of services and under-load regime. For small networks with high standard deviation of incoming traffic the efficiency of Bandwidth allocation with MAR grows down.

References

1. **Ash J.** Max Allocation with Reservation Bandwidth Constraints Model for DiffServ-aware MPLS Traffic Engineering // IETF Network Working Group, RFC 4126. – June 2005.
2. **Heckmann O., Schmitt J., Steinmetz R.**, Robust Bandwidth Allocation Strategies // Proceedings of IWOoS, 2002. – P.138–147.
3. **ITU-T Rec. E.360.1, 2, 3** (05/2002)
4. **Le Faucher F.** Russian Dolls Bandwidth Constraint Model for DiffServ-aware MPLS Traffic Engineering // IETF Network Working Group, RFC 4127. – June 2005.
5. **Chee-Kin Hui, Chen-Khong Tham.** Reinforcement Learning-based Dynamic Bandwidth Provisioning for Quality of Service in Differentiated Services Networks // Computer Communications, Elsevier Science, 2004.
6. **Atlatis A. F., Vasilikos A.** The Use of Reinforcement Learning Algorithms in Traffic Control of High Speed Networks // Proceedings of ESIT 2000, European Symposium on Intelligent Techniques, Sept 2000.
7. **G. Lauks.** Multi-objective optimization of Bandwidth Broker with Evolutionary algorithm // Electronics and Electrical Engineering. – Kaunas: Technologija, 2005. – No. 3(59). – P. 14–17.
8. **Fulp E.** Resource Allocation and Pricing for QoS Management in Computer Networks // PhD thesis, North Carolina State University, 1999.

Submitted for publication 2006 03 02

G. Lauks. Adaptive Bandwidth Allocation in Broadband Networks // Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. – No. 5(69). – P 33–36.

MPLS and DiffServ have been widely accepted as the service model to adopt for providing Quality of Services (QoS) over next generation IP networks. The combination of MPLS and DiffServ presents a very attractive strategy to broadband network providers. Bandwidth Allocation mechanism in DiffServ guaranties to meet the QoS requirements for all types of services. The objectives of this paper are to simulate and estimate the efficiency of Maximum Allocation with Reservation (MAR) method in small network cases. Problem statement of MAR is enhanced with bandwidth utilization variables. Simulation results illustrate the efficiency of traffic predictors and MAR in small network cases. III. 3, bibl. 8 (in English; summaries in English, Russian and Lithuanian).

Г. Лаукс. Адаптивное присвоение полосы передачи в широкополосных сетях // Электроника и электротехника. – Каунас: Технология, 2006. – № 5(69). – P 33–36.

MPLS и DiffServ технологии сегодня являются общепризнанными для обеспечения необходимого качества в сетях IP нового поколения. Целью этой статьи является моделирование и оценка эффективности работы алгоритма присвоения полосы передачи MAR в случае сетей малых размеров. Постановка задачи по сравнению с известными расширена новым переменным – использованием полосы. Результаты симуляций демонстрируют эффективность работы предсказателей трафика и MAR в случае малых сетей. Ил. 3, библи. 8 (на английском языке; рефераты на английском, русском и литовском яз.).

G. Lauks. Adaptyvus perdavimo juostos paskirstymas plačiajuosčiuose tinkluose // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2006. – Nr. 5(69). – P. 33–36.

MPLS ir DiffServ šiandien yra visuotinai pripažintos technologijos, naudojamos būtinai kokybei užtikrinti naujos kartos IP tinkluose. Šio darbo tikslas – imituoti ir įvertinti MAR metodo efektyvumą esant mažiems tinklams. Šios užduoties formuluotė papildyta pralaidumo juostos kintamaisiais. Imitavimo rezultatai parodo duomenų srauto prognozavimo rodiklį ir MAR efektyvumą, esant mažiems (vietiniams) tinklams. II. 3, bibl. 8 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).