Congestion Control for Highly Loaded DIFFSERV/MPLS Networks

Srećko Krile

University of Dubrovnik
Department of Electrical Engineering and Computing
Cira Carica 4, 20000 Dubrovnik, Croatia
Tel. +385 20 445-739, Fax. +385 20 435-590, e-mail: srecko.krile@unidu.hr

Dario Krešić
University of Zagreb
Faculty of Organization and Informatics (FOI)
Pavlinska 2, 42000 Varazdin, Croatia, e-mail: dario.kresic@foi.hr

Abstract - **Optimal QoS path provisioning of coexisted and aggregated traffic in networks is still demanding problem. All traffic flows in a domain are distributed among LSPs (Label Switching Path) related to service classes, but the congestion problem of concurrent flows (traversing the network simultaneously) can appear. For LSP creation the IGP (Interior Getaway Protocol) uses simple on-line routing algorithms (e.g. OSPFS, IS-IS) based on shortest path methodology. In the presence of premium traffic where some links may be reserved for certain traffic classes or for particular set of users it becomes insufficient technique. On other hand, constraint based explicit routing (CR) based on IGP metric ensures traffic engineering (TE) capabilities. It may find a longer but lightly loaded path better than the heavily loaded shortest path. In this paper a new approach to explicit constraint-based routing is proposed. LSP can be pre-computed much earlier, possibly during SLA (Service Level Agreement) negotiation process. It could be a very good solution for congestion avoidance and for better load-balancing purpose where links are running close to capacity. To be acceptable for real applications such complicated routing algorithm can be significantly improved. Further improvements through heuristic approach are made and comparisons of results are discussed.**

Keywords - **intra-domain QoS routing, traffic engineering in DiffServ/MPLS networks, constraint-based routing**

1. Introduction

MPLS uses extensions to Resource Reservation Protocol (TE-RSVP) and the MPLS forwarding paradigm to provide explicit routing; see [9], [10] and [12]. With OSPF (*Open Shortest Path First*), widely-used IGP routing protocol, some paths may become congested while others are underutilized. Such implicit routing can be appropriate only for under loaded networks. For highly loaded networks we need prediction of congestion probability and it has to be done much before the moment of service utilization. Constraint-based routing (CR) as a extension of explicit routing allows an originating (ingress) router to compute a path (LSP) to egress router (sequence of intermediate LSRs), taking care of constraints such as bandwidth, delay and administrative policy; see [7]. With constraint-based label distribution protocol (CR-LDP) we can ensure the bandwidth provisioning directives and other information (list of router's neighbors, attached networks, actual resource availability and other relevant information). It can be distributed for each service class at each link along the path (LSP); see [13]. CR process can be incorporated into each ingress router and co-exists with the conventional routing tecnique.

As we need firm correlation with bandwidth management and traffic engineering (TE) the initial (pro-active) routing can be pre-computed in the context of all priority traffic flows (former contracted SLAs) traversing the network simultaneously; see fig. 1.a. It could be a very good solution for congestion avoidance and for better load-balancing purpose where links are running close to capacity; see [11]. If we want to obtain quantitative end-to-end guarantees the QoS provisioning has to be in firm correlation with bandwidth management; see [1].

Detail explanation of new constraint-based routing approach is given in section 2. CR routing technique can be seen as the capacity expansion problem (CEP) The mathematical model explanation and CEP algorithm development are given in the section 3. The comparison of results for different algorithm options we can see in the section 4.

2. LSP Creation During SLA Negotiation

The service provider in domain (e.g. ISP) wants to accept new SLA that results with priority traffic flow between edge routers. A *traffic trunk* is defined as a logical pipeline within an LSP, with reservation of certain amount of capacity to serve the traffic associated with a certain SLA. So it is clear that LSP between an ingress/egress pair may carry multiple traffic trunks associated with different SLAs; see [6]. In fig. 1.b. we have situation on the path for the example of simultaneous SLA flows from fig. 1.a. All traffic flows on the path are participating possibly in the same time (the worst case). In that sense the network operator (e.g ISP) has to find the optimal LSPs for aggregated flows without any possible congestion in the core network; see [5]. Each traffic demand can be satisfied on appropriate or higher QoS level. The main condition is: the sufficient network resources must be available for the priority traffic at any moment.

During SLA negotiation process the RM (Resource Manager) module has to determine the main parameters that characterize the required flow (i.e., bandwidth, QoS class, ingress and egress IP router addresses). At first RM can apply any shortest path-based routing algorithm (e.g. OSPF - Open Shortest Path First) to get initial LSP. The BB (Bandwidth Broker) will therefore check if there are enough resources on the calculated path to satisfy the requested service class, taking care of all existing flows in the same time (caused by former SLAs); see [4].

With such congestion control algorithm the RM can predict sufficient link resources to satisfy all traffic demands. If the optimal routing sequence has any link that exceeds allowed capacity limits (maximal bandwidth) congestion exists; see [2]. It means that link capacity on the path cannot be sufficient for such traffic. Such congested link has to be eliminated from further calculation and procedure starts again. Alternatively, adding capacity arrangement (short-term) is possible but can produce significant extra cost for network operator.

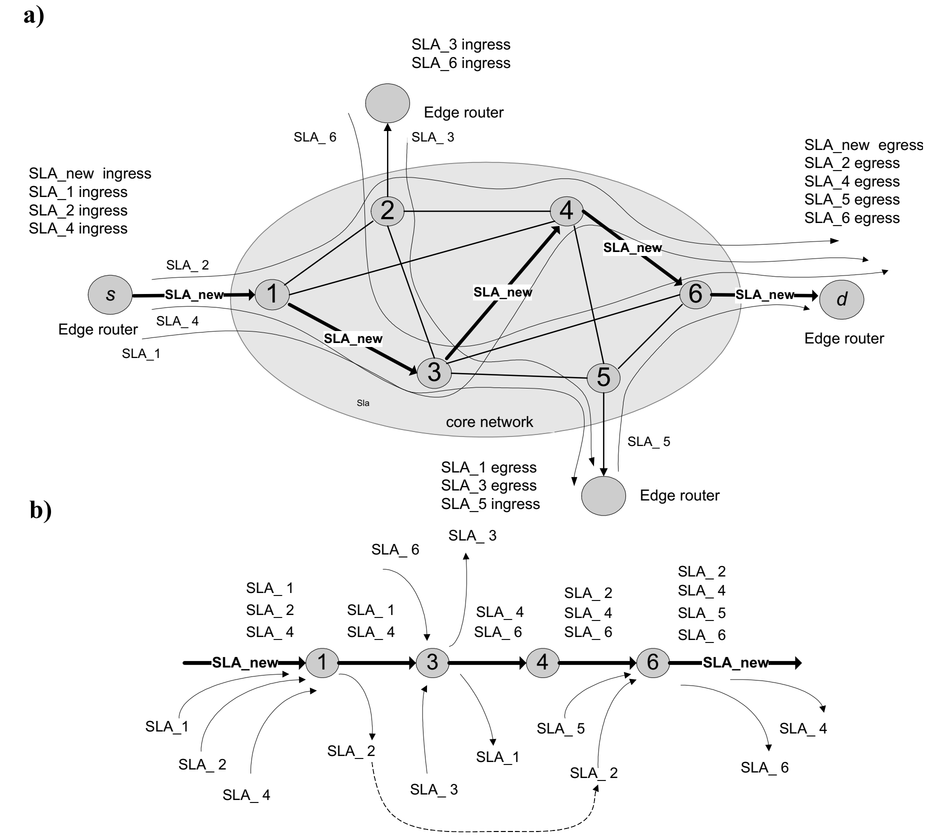

Fig. 1.a. An example of number of SLAs in the context of new SLA creation. In figure 1. b. we can see simultaneous flows with possibly congestion on the path.

If calculation finds the path without any congestion the new SLA can be accepted and related LSP is assigned to that flow and stored in database of BB. In opposite the new SLA cannot be accepted or must be re-negotiated. In the moment of service invocation such calculated and stored LSP can be easily distributed from BB to the MPLS network to support explicit routing, leveraging bandwidth reservation and prioritization; see [3].

In that way the LSP creation should be in co-relation with SLA, to enable better load-balancing and congestion avoidance in domain. In such CR approach we can observe the main difference from usual on-line routing techniques (e.g. OSPF): the LSP need not to be necessarily the shortest path solution.

3. CEP for congestion control and Load Balancing Purpose

The congestion control technique explained above can be seen as the capacity expansion problem (CEP) with or without shortages. For full traffic satisfaction we talk about CEP without shortages. Transmission link is capable to serve traffic demands for *N* different QoS levels (service class) for *i* = 1, 2, ..., *N*. For each load we need appropriate bandwidth amount, so it looks like bandwidth expansion. Bandwidth portions on the link can be assigned to appropriate service class up to the given limit (maximal capacity). Used capacity can be increased in two forms: by expansion or by conversion. Expansions can be done separately for each service class or through conversion (redirected amount) to lower quality class. It means that it can be reused under special conditions to serve the traffic of lover quality level. Bandwidth usage for each service class can be a part of resource reservation strategy. Fig. 2 gives an example of network flow representation for multiple QoS levels (*N*) and *M* core routers (LSR) on the path. In the CEP model the following notation is used:

Fig. 2 gives an example of network flow representation for multiple QoS levels (*N*) and *M* core routers (LSR) on the path. In the CEP model the following notation is used:

i, *j* and *k* = QoS level. We differentiate *n* service classes (QoS levels).The *N* levels are ranked from *i* = 1, 2,..., *N*, from higher to lower.

m = the order number of the link on the path, connecting two successive routers, *m* = 1, …., *M*+1.

u,*v* = the order number of capacity points in the sub-problem, 1 ≤ *u*, ..., *v* ≤ *M*+1.

$r_{i,m}$ = traffic demand increment for additional capacity for each router on the path. Any traffic demand can also be satisfied by converted capacity from any capacity type *k* with higher quality level.

For convenience, the $r_{i,m}$ is assumed to be integer. The sum of traffic demand for capacity type *i* between two routers:

$$R_i(m_1, m_2) = \sum_{m=m_1}^{m_2} r_{i,m} \qquad (3.1)$$

The sum of demands for whole path and for all capacity types has to be positive or zero: $\sum_{i=1}^{N} R_i(1, M) \geq 0$ (3.2)

It means that we don't expect reduction of total capacity on the path toward egress router, in other words we presume the increase of capacity. Traffic demand can also be satisfied by converted capacity from one capacity type to another, partially or in total amount.

$I_{i,m}$ = the relative amount of idle capacity on the link m, connecting two routers. Ee have positive and negative values. $I_{i1} = 0$, $I_{i,M+1} = 0$ that means: no adding capacity is necessary on the links toward edge routers. Those links are not the mater of optimization.

$x_{i,m}$ = the amount of adding capacity for each service class on the link *m*. Possible negative values (decrease).

$L_{i,m}$ = bandwidth constraints for link capacity values on the link *m* and for appropriate service class *i* ($L_{1,m}$, $L_{2,m}$, … $L_{N,m}$).

$y_{i,j,m}$ = the amount of capacity for quality level *i* on the link *m*, redirected to satisfy the traffic of lower quality level *j*.

$w_{i,m}$= weight for the link *m* and appropriate service class *i* (QoS level).

$del_{i,m}$= delay on the link *m* for appropriate service class *i*. Maximal delay on the path is denoted with DEL_i.

As we have nonlinear expansion functions (showing the economy of scale) the CEP can be solved by any

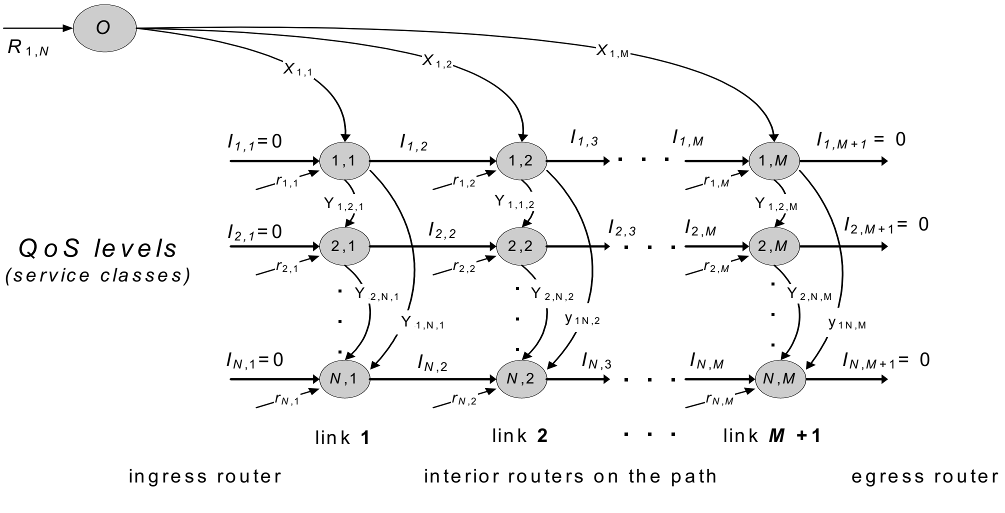

Fig. 2. - The network flow representation of the CEP model applied for congestion control purposes.

nonlinear optimization technique. Instead of a nonlinear convex optimization, that can be very complicated, the network optimization methodology is efficiently applied; see [9]. The main reason on such approach is the possibility of discrete capacity values for limited number of QoS classes, so the optimization process can be significantly improved. The problem can be formulated as Minimum Cost Multi-Commodity Flow Problem (MCMCF). Such problem (NP-complete) can be easily represented by multi-commodity the single (common) source multiple destination network; see fig. 2.

Let G (V, E) denote a network topology, where V is the set of vertices/nodes, representing link capacity states and A, the set of arcs representing traffic flows between routers. Each link on the path is characterized by z-dimensional link weight vector, consisting of z-nonnegative QoS weights. The number of QoS measures (e.g. bandwidth, delay) is denoted by z. In general we have multi-constrained problem (MCP) but in this paper we talk about one-dimensional link weight vectors for M+1 links on the path $\{w_{i,m}, m \in A, i = 1, \ldots, N\}$. E.g. the capacity constraint for each link on the path is denoted with $L_{i,m}$ ($L_{1,m}$ $\mathrm{L}_{2,m}$, … $L_{N,m}$). For a non-additive measure (e.g. bandwidth) definition of the single-constrained problem is to find a path from ingress to egress node with minimal link weight along the path.

In the context of MCP we can introduce easily the adding constraint of max. delay on the path (end-to-end). As it is an additive measure (more links on the path cause higher delay) it can be used as criteria to eliminate any unacceptable routing solution from calculation.

The flow situation on the link depends of expansion and conversion values ($x_{i,m}$, $y_{i,j,m}$). It means that the link weight (cost) is the function of used capacity: lower amount of used capacity (capacity utilization) gives lower weight. If the link expansion cost corresponds to the amount of used capacity, the objective is to find the optimal routing policy that minimizes the total cost on the path.

Definition of the single-constrained problem is to find a path P from ingress to egress node such that:

$$w(P) = \min \sum_{m=1}^{M+1} \sum_{i=1}^{N} w_{i,m}(I_{i,m}, x_{i,m}, y_{i,j,m}) \qquad (3.3)$$

where: $I_{i,m} \leq L_{i,m}$ (3.4)

$$\sum_{m_1}^{m_2} del_{i,m} \leq DEL_i \qquad (3.5)$$

satisfying condition: max. delay of $P \leq DEL_i$ (3.6)

for $i = 1, \ldots, N$; $m = 1, \ldots, M$

A path obeying the above conditions is said to be feasible. Note that there may be multiple feasible paths between ingress and egress node. Generalizing the concept of the capacity states for each quality level of transmission link m between LSRs in which the capacity states for each service class (QoS level) are known within defined limits we define *a capacity point* - α_m.

$$\alpha_m = (I_{1,m}, I_{2,m}, \ldots , I_{N,m}) \qquad (3.7)$$

$$\alpha_1 = \alpha_{M+1} = (0, 0, \ldots , 0) \qquad (3.8)$$

In formulation (3.7) α_m denotes the vector of capacities $I_{i,m}$ for each service class on link m, and we call it capacity point. On the flow diagrams (fig. 2.) each column represents a capacity point of the node, consisting of N capacity state values (for i-th QoS level). Link capacity is capable to serve different service classes. Capacity amount labeled with i is primarily used to serve traffic demands of that service class but it can be used to satisfy traffic of lower QoS level j ($j > i$).

Formulation (3.8) implies that idle capacities or capacity shortages are not allowed on the beginning and on the end of optimization. It means that process is starting with new SLA flow that must be fully satisfied through the network (to egress node).

The objective function for CEP problem can be formulated as follows:

$$\min\left(\sum_{m=1}^{M+1}\left\{\sum_{i=1}^{N} c_{i,m}\left(x_{i,m}\right) + h_{i,m}\left(I_{i,m+1}\right) + g_{i,j,m}\left(y_{i,j,m}\right)\right\}\right) \qquad (3.9)$$

so that we have:

$$I_{i,m+1} = I_{i,m} + x_{i,m} - \sum_{j=i+1}^{N} y_{i,j,m} - r_{i,m} \qquad (3.10)$$

$$I_{i,1} = I_{i,M+1} = 0 \qquad (3.11)$$

for $m = 1, 2, \ldots, M+1$; $i = 1, 2, \ldots , N$; $j = \mathrm{i} + 1, \ldots , N$.

In the objective function the total cost (weight) includes some different costs. Expansion cost (adding capacity) is denoted with $c_{i,m}$ ($x_{i,m}$). For the link expansion in allowed limits we can set the expansion cost to zero. We can differentiate expansion cost for each service class. We can take in account the idle capacity cost $h_{i,m}$ ($I_{i,m+1}$), but only as a penalty cost to force the usage of the minimum link capacity (prevention of unused/idle capacity). Also we can introduce facility conversion cost $g_{i,j,m}$ ($y_{i,j,m}$) that can control non-effective usage of link capacity (e.g. usage of higher service class capacity instead). Costs are often represented by the fix-charge cost or with constant value. We assume that all cost functions are concave and non-decreasing (reflecting economies of scale) and they differ from link to link. The objective function is necessarily non-linear cost. With different cost parameters we can influence on the optimization process, looking for benefits of the most appropriate expansion solution.

A. Algorithm Development

The network optimization can be divided in two steps. At first step we are calculating the minimal expansion weights $d_{u,v}$ for all pairs of capacity points in neighbor links on the path. The calculation of weight value between capacity points we call: capacity expansion sub-problem (CES); see (3.9). It requires solving repeatedly a certain single location expansion problem (SLEP). At second step we are looking for the shortest path in the network with former calculated weights between node pairs (capacity points). On that network optimization level problem can be seen as a shortest path problem for an acyclic network in which the nodes represent all possible values of capacity points. Then Dijkstra's algorithm or any similar algorithm can be applied. It is obvious that the optimal routing sequence need not to be the shortest path solution.

B. The improvement of CEP algorithm

The key for this very effective approach is in fact that extreme flow theory enables separation of these extreme flows which can be included in optimal expansion solution from those which cannot be; see Luss [8]. Most of the computational effort is spent on computing the sub-problem values. Any of them, if it cannot be a part of the optimal sequence, is set to infinity. It can be shown that a feasible flow in the network given in fig. 2. corresponds to an extreme point solution of CEP if and only if it is not the part of any cycle (loop) with positive flows, in which all flows satisfy given properties. One may observe that the absence of cycles with positive flows implies that each node has at most one incoming flow from the source node. This result holds for all single source networks. That means that optimal solution of $d_{u,v}$ has at most one expansion (or reduction) for each facility. Using a network flow theory properties of extreme point solution are identified. These properties are used to develop an efficient search for the link costs $d_{u,v}$.

4. Testing Results and Comparison of Different Algorithm Options

The proposed algorithm is tested on many numerical test-examples, looking for optimal routing sequence on the path. Between edge routers there are maximum M core routers (LSR) and the path consists of maximum M+1links. Traffic demands (former contracted SLAs) are given in relative amount for each interior router on the path. Demands are overlapping in time and are defined for each capacity type (service class). Results obtained by improved algorithm (reduction of unacceptable expansion solutions) are compared with results obtained by referent algorithm that is calculating all possible expansion solutions for each CES.

For all numerical test-examples with improved algorithm (denoted with Basic_A) the best possible result (near-optimal expansion sequence) can be obtained, as same as with referent algorithm. For each test-example we know the total number of capacity points. The number of possible CES is well-known, so it is the measure of the complexity for the CEP-problem. Also, for each test-example we calculate the number of acceptable sub-problems, satisfying basic and additional properties of optimal flow. Savings in percents are on average over 20 % that is proportionally reflected on computation time savings.

The number of all possible CES values depends on the total number of capacity points as resultant of traffic demands. So CEP requires the computation effort of $O(NMC_p^2)$ with linear influence of N. In real application we normally apply definite granularity of capacity values through discrete values (only integer) of traffic demands R_i. It reduces the number of the capacity points significantly. Also, the lowest step ($step\_I_i$) of possible capacity change has strong influence on the complexity. For all numerical test-examples the improved algorithm (denoted with Basic_A) can obtain the best possible result (near-optimal expansion sequence), same as referent algorithm (without reduction of unacceptable expansion solutions). The number of sub-problems (CES) is well-known, that is the measure of the complexity of the CEP-problem. Also, for each test-

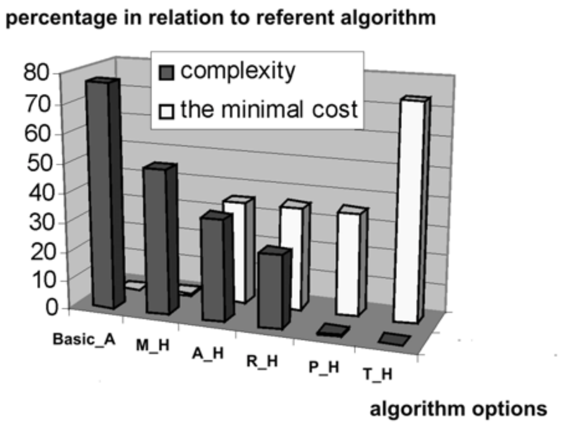

Fig. 3 - Trends of algorithm complexity and comparison of results (minimal cost)

example we can see the number of acceptable subproblems, satisfying basic and additional properties of optimal flow. For *N*=3 and *M*=6 savings in percents are on average near 20 % that is proportionally reflected on computation time savings.

In real situation we can introduce some limitations on the capacity state values, talking about heuristic algorithm options:

a) Only one negative capacity value in the capacity point. Such option is denoted with M_H *(Minimal-shortage Heuristic option*);

b) Total sum of the link capacity values (for all quality levels) is positive A_H *(Acceptable Heuristic option*);

c) Total sum is positive but only one value can be negative. Such option is denoted with R_H (*Real Heuristic option*);

d) Algorithm option that allows only non-negative capacity state values is denoted with P_H (*Positive Heuristic option*);

e) Only null capacity values are allowed. A trivial heuristic option (denoted with T_H) allows only zero values in capacity point (only one capacity point).

We compared the efficiency of algorithm in above mentioned options. In figure 3. we can see the average values of results for *N*=3 and *M*=6. Only for few test-examples any algorithm options can find the best expansion sequence, providing the minimal cost no matter of algorithm option we use. For the most examples algorithm option M_H can obtain the best result with average saving of 50 %. For other algorithm options the significant reduction of complexity is obvious but deterioration of result appears. We can say the final results are still in acceptable limits (see fig. 3). In the most cases for trivial algorithm option (T_H) the significant deterioration of result is obvious. A very good fact all algorithm options is that efficiency rises with increase of value *M.*

5. Conclusion

We can check congestion probabilities on the path with algorithm of very low complexity first (e.g. P_H algorithm option). It means that only if congestion possibility appears we need optimization with more complex algorithm (e.g. A_H). With the most complex algorithm option (Basic_A) we can get the best possible result, so we can be sure if congestion on the path could appear or not. In the case of congestion appearance new SLA cannot be accepted or adding capacity arrangement should be done. It means that SLA re-negotiation has to be done and customer has to change the service parameters: e.g. bandwidth (data speed), max. delay or period of service utilization.

The proposed algorithm (with different options) can be efficiently incorporated in explicit intra-domain routing in DiffServ/MPLS networks. So we can get firm correlation with bandwidth management and admission control. Routing process can start much earlier, possibly during SLA negotiation process.

References

[1] Haddadou, K., Ghamri-Doudane, & all: “Designing scalable on-demand policy-based resource allocation in IP networks”, IEEE Communications Mag., vol. 44, No. 3, 2006, pp. 142-149.

[2] Yu Cheng, R. Farha, A. Tizghadam and all: “Virtual Network Approach to Scalable IP Service Deployment and Efficient Resource Management”, IEEE Communication Mag., Vol. 43, No. 10, 2005, pp.76-84.

[3] Giordano, S., Salsano, S., Ventre, G.: “Advanced QoS Provisioning in IP Networks”: The European Premium IP Projects, IEEE Communication Mag., Vol. 41, No. 1, 2003, pp.30-36.

[4] Boutaba, R., Szeto, W. and Iraqi, Y.: “DORA: Efficient Routing for MPLS Traffic Engineering”, Journal of Network and Systems Management (JNSM), Vol. 10, No. 3, 2002, pp.309-325.

[5] Lima, S., Carvalho, P. and Freitas, V.: “Distributed Admission Control for QoS and SLS Management”, JNSM, Vol. 12, No. 3, 2004, pp.397-426.

[6] Bhatnagar, S.; Ganguly, S.; Nath, B.: “Creating multipoint-to-point LSPs for traffic engineering”, IEEE Communications Mag., Vol. 43., No. 1, 2005, pp.95- 100.

[7] Kagklis, D., Tsakiris, C., Liampotis, N.: “Quality of Service: A Mechanism for explicit activation of IP Services Based on RSVP”, Journal of Electrical Engineering, Vol. 54, No. 9-10, Bratislava, 2003, pp. 250-254.

[8] Luss, H.: “A Heuristic for Capacity Expansion Planning with Multiple Facility Types”, Naval Res. Log. Quart., Vol. 33 (04), 1986, pp.685-701.

[9] Degrande, N., Van Hoey, G., La Vallee-Poussin, P. and Van Busch, S.: ”Inter-area Traffic Engineering in a Differentiated Services Network”, JNSM, Vol. 11, No. 4, 2003.

[10] D’Arienzo, M., Pescape, A. and Ventre, G.: “Dynamic Service Management in Heterogeneous Networks”, JNSM, Vol. 12, No. 3, 2004. pp. 349-370.

[11] Morrow, M. and Sayeed, A.: "MPLS and Next-Generation Networks: Foundations for NGN and Enterprise Virtualization", Cisco Press, 2006.

[12] Guichard, J., Le Faucheur, F. and Vasseur, J.P.: "Definitive MPLS Designs", Cisco Press, 2005. pp.253 – 264.

[13] Younis O., Fahmy S.; “Constraint-Based Routing in the Internet: Basic Principles and Recent Research,” *IEEE Communications Surveys & Tutorials*, Volume 5, Issue 1, pp. 2-13, 3rd quarter 2003.