Traffic engineering in telecommunication networks

Ernst Nordström
ernstn@bizopt.se

April 5, 2010

Abstract

The objective of traffic engineering (TE) in telecommunication networks is to maximize the profit, i.e. the difference between revenue from user charges and the total network cost. The constraints of TE include requirements on service performance, i.e. Quality of Service (QoS), Quality of Experience (QoE) and Grade of Service (GoS). TE relies on a relationship between three models: traffic model, network model, and performance model. The choice of performance model involves an accuracy-simplicity dilemma.

1 Fundamentals

1.1 Definition

Traffic Engineering (TE) for Internet is defined in RFC 3272 and involves both capacity management and traffic management. Capacity management includes capacity planning, routing control, and resource management. Traffic management includes (1) nodal traffic control functions such as traffic conditioning, queue management, scheduling, and (2) other functions that regulate traffic flow through the network or that arbitrate access to network resources.

1.2 Tradeoff between effectiveness and simplicity

TE is concerned with finding an efficient tradeoff between effectiveness (in terms of proximity to optimality) and simplicity (in terms of time and space complexity) for the TE solution. A highly effective TE solution is desired since it means that fewer call requests needs to be rejected leading to increased revenue. Moreover, expansion of the network capacity is driven by the increase in network demand. An effective TE solution allows longer time periods between capacity upgrades which means longer time periods for the depreciation of the capital expenditures.

However, increased effectiveness normally requires a more complex computer-based solution. A complex solution has larger capital expenditures (e.g. implementation costs) and operational expenditures (e.g. system maintenance costs). A complex solution is based on a more complex system model and/or computation algorithm that requires longer execution times and/or larger memory space. Note that the system response time requirements restrict the complexity of the TE algorithms. Therefore, the chosen TE solution should provide a good balance between the potential for large revenue (measured by TE effectiveness) against

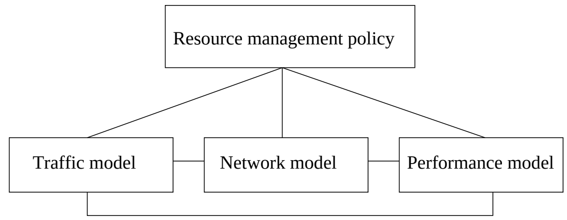

Figure 1: Resource management policy

the expected costs (measured by TE complexity). TE planning involves finding the set of TE algorithms with maximal effectiveness that provides the desired TE complexity.

1.3 Service guarantees

The network is offered flows from several service classes. Each service class should be given end-to-end performance guarantees in terms of QoS and GoS metrics on the packet/burst and call level respectively. QoS metrics include loss, delay, delay-jitter and throughput guarantees. Packet loss can occur due to buffer overflow or delay bound violation. Packet delay is composed of a fixed part due to packetization, propagation, transmission, reassembly, switching, and a variable part due to stochastic queueing effects. Packet delay-jitter is defined as the difference between delays of any two packets. A *statistical* service specifies QoS by loss probability, delay quantile, and delay-jitter quantile, while a *deterministic* service specifies QoS by zero loss, worst-case delay, and maximum difference between delays of any two packets. GoS metrics includes flow request blocking probability and flow setup delay specified as a quantile.

The service-level specification (SLS) is part of the service-level agreement (SLA) negotiated at flow set up. The contents of the SLS include the essential QoS-related parameters, including scope and flow identification, traffic conformance parameters, and service guarantees [189]. Specifically, the traffic conformance parameters include five parameters: the peak rate p (bytes/s), maximum burst size b (bytes), mean rate r (bytes/s), minimum policed unit m (bytes), and maximum packet size M (bytes).

1.4 Resource management policy

The resource management policy relies on a relationship between three models: traffic model, network model, and performance model, see Figure 1. To enable effective and efficient control of network resources (router CPU and buffer capacities, link capacities) all these models need to be sufficiently accurate but also simple enough to limit the delay for processing of flow requests.

1.4.1 Traffic model

A realistic traffic model is a prerequisite for accurate performance evaluation. The traffic model consists of five layers: physical network, virtual network, call, burst, and packet layers [87], see Figure 2. Internet traffic measurements and advancements in modeling the last two decades have revealed that traffic arrivals on all five layers could be self-similar. This means the statistical variability in the arrival process carries over from small to larger time scales. Likewise, the traditional models for call holding times used in telephone networks are not adequate for some significant Internet services.

1.4.2 Network model

We adopt the TE and QoS optimization network model outlined by Ash [12]:

- Physical network (PN) implemented by optical cross connects and fibers;
- Virtual networks (VNs) implemented by GMPLS;
- IP transport within each VN implemented by MPLS; and
- Differentiated services.

DiffServ manages aggregates of flows to achieve scalability. The per-hop-behavior (PHB) of the aggregate flow defines its scheduling treatment in the routers [1] and the traffic conditioning rule at ingress and egress boundary nodes.

We assume the network runs the IP/MPLS [172] or IP/ATM protocols [90, 7]. MPLS is a technology that integrates label-swapping paradigm with network-layer routing. The Label Switching Router (LSR) has the same function as the ATM switch. The LSP between two routers can be the same as the layer 3 hop-by-hop route, or the sender LSR can specify an explicit route for the LSP. LSPs in MPLS networks are similar to Virtual Channel Connections (VCCs) in ATM networks. Traffic engineered (TE) LSPs in MPLS networks correspond to Virtual Path Connections (VPCs) in ATM networks. TE-LSPs carry multiple LSPs in MPLS networks. VPCs carry multiple VCCs in ATM networks.

1.4.3 Performance model

The choice of performance model involves an accuracy-simplicity dilemma. The model needs to be both accurate and simple. High accuracy is desired to admit correct and efficient control/allocation decisions. High simplicity (low computational complexity) is required since decisions should be made fast enough to be acceptable by the users. To obtain a solution feasible in real-time, approximations are normally introduced. However, this will reduce the network utilization and thereby the revenue for the network operator.

2 Hierarchical resource management model

Hierarchical layers of dynamic resource management are performed at decreasing time scales, see Figure 3. Resource management at the higher layer aims at providing sufficient performance at the lower layer.

Resource management at each layer can be done at regular time intervals or be triggered by changes in traffic or capacity at the higher layer.

[1] In the literature, the term "router" or "gateway" is used in an internetworking environment, while "label switching router" and "switch" is used in the context of MPLS and ATM networks, respectively. In this project we refer to switching elements as "routers".

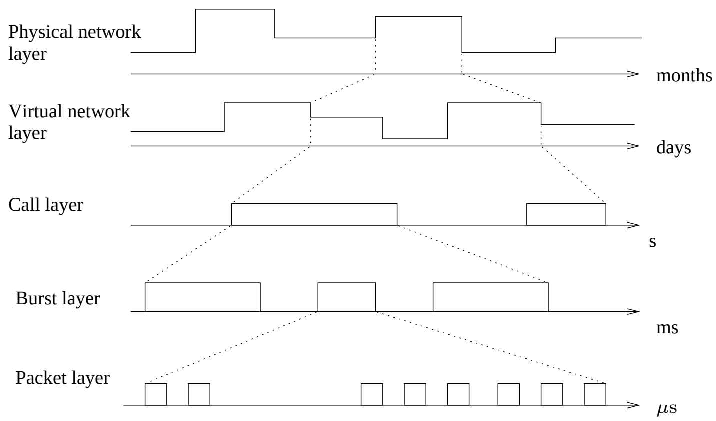

Figure 2: Hierarchical traffic model

2.1 Physical and virtual network layer

We assume a set of virtual networks (VNs) overlay the physical domain network. A VN consists of a set of VN nodes interconnected by a set of VN links. The topology of the VN may be different from the physical network topology. A VN link defines a path (consisting of one or more physical links) between two VN nodes.

The Virtual Path (VP) concept may be used in IP/MPLS and IP/ATM networks [130]. A VP provides transport of TE-LSP and carries multiple LSPs. A series of VN links defines the VP routing path.

On the packet layer, the division of physical link capacity among multiple VNs is efficiently handled by the WFQ packet scheduler due to its sharing and isolation capabilities. On the call layer, the division of physical link capacity can be implemented by complete sharing, complete partitioning or partial sharing.

Network design is a crucial function in our TE framework. The design is carried out on the physical network, virtual network and call layer. Design on the physical and virtual network layers determines the topology and the set of link capacities. Physical network design is carried out on the long-term time scale. Virtual network design is carried out on multiple time scales [68]:

- Successive capacity reallocation redistributes capacity on a fixed virtual network topology;
- Successive topology reconfiguration establishes and/or tears down TE-LSPs, within an existing virtual network topology;
- Global reconfiguration consists of both *global capacity reconfiguration* and *global topology reconfiguration*. This activity potentially affects all the TE-LSPs in the network;
- Long-term planning derives a static (or general) set of TE-LSPs and initial or minimum capacity assignments for them.

Successive capacity reallocation relies on policies for set up and teardown of TE-LSPs, and the policy for routing of the new TE-LSPs [6]. Constraint-based routing selects a network

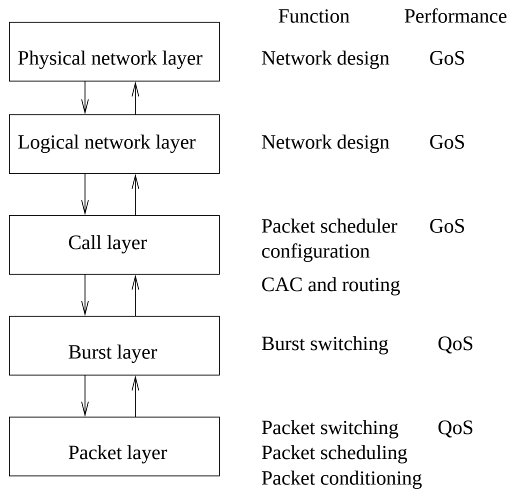

Figure 3: Hierarchical resource management model

path in the physical network for the TE-LSP subject to a set of constraints [17]. Constraint-based routing generalizes QoS routing by finding routes for traffic trunks instead of micro flows.

2.2 Call layer

2.2.1 Multi-constrained optimal path problem

QoS routing can be modeled as a multi-contrained optimal path (MCOP) problem. We formulate the MCOP problem as follows [113]. Let $G(\mathbf{V}, \mathbf{E})$ denote a network topology, where $\mathbf{V}$ is the set of nodes and $\mathbf{E}$ is the set of links. The origin and destination nodes are denoted s and d, respectively. The number of QoS measures are denoted by m. Each link is characterized by an m-dimensional link weight vector, consisting of m nonnegative QoS weights as components. Stochastic QoS weights are represented by a probability distribution function. Deterministic QoS weights are represented by a constant value. QoS measures can be classified into *additive* (e.g. cost, delay), *concave* (e.g. bandwidth, policy flags) or *multiplicative* (e.g. loss). In case of an additive deterministic measure, the QoS value of a path is equal to the sum of the corresponding weights of the links along the path. In case of an additive stochastic measure, the QoS value of a path is equal to the mathematical convolution of per-link QoS distributions. For a concave measure, the QoS value of the path is the minimum (or maximum) link weight along that path. A multiplicative measure can be transformed into an additive measure by taking the logarithm. In general, concave measures can easily be dealt with by pruning from the graph all links that do not satisfy the requested QoS constraint. Additive measures cause more difficulties.

Definition 1: Multiconstrained path (MCP) problem: Consider a network $G(\mathbf{V}, \mathbf{E})$. Each link $(u, v) \in \mathbf{E}$ is associated with m additive weights $w_i(u, v) \geq 0, i = 1, \ldots, m$. Given m constraints $L_i, i = 1, \ldots, m$, the problem is to find a path P from s to d such that:

$$w_i(P) \stackrel{def}{=} \sum_{(u,v) \in P} w_i(u, v) \leq L_i$$

for $i = 1, \ldots, m$. A path obeying the above condition is said to be *feasible*. Note that there may be multiple feasible paths between s and d. A modified (and more difficult) version of the MCP problem is to retrieve the shortest "length" path among the set of feasible paths. This problem is known as the multi-constrained optimal path problem, and is attained by adding a second condition on the path P definition 1: $l(P) \leq l(Q)$ for any feasible path Q between s and d, where $l(\dot{})$ is a path length (or cost) function.

2.3 Burst and packet layer

2.3.1 Packet scheduling

The routers are assumed to be non-blocking, i.e when packets arrive at an input link, they can be routed directly to the appropriate output links without switching conflicts. Packet destined for different output links do not interfere with each other, and queueing occurs only at the output ports of the router. The service disciplines deployed at the output links allocate three types of resources: *bandwidth* (which packets get transmitted), *promptness* (when those packets gets transmitted) and buffer space (which packets gets discarded). The bandwidth, promptness and buffer allocation policy affects, in turn, the QoS parameters loss, delay, jitter and throughput.

The packet scheduling schemes deployed at the output links at the routers must be able to support multiple service classes. Possible schemes include First-Come-First-Served (FCFS), Strict Priority (SP), Weighted Fair Queueing (WFQ), and Earliest Deadline First (EDF) [213].

FCFS, SP, WFQ, and EDF provide long-term bandwidth guarantees. However, the WFQ scheme has the advantage of also giving a short-term bandwidth guarantee, given by the class weight, to each class. The isolation means that less jitter is introduced in the output packet stream.

FCFS, SP and WFQ are known to be suboptimal in comparison to EDF scheduling, for both deterministic and statistical end-to-end guarantees [180]. To implement EDF a complex sorting mechanism is required [181]. PGPS is often preferred over EDF due to its simplicity. However, optimal GPS scheduling requires dynamic re-synchronization of bandwidth weights which is considered costly in switches of today [180].

The ideal GPS scheme assumes packets are infinitesimally divisible and that the server can serve multiple packets simultaneously. The WFQ scheme has a server that serves the packets from the backlogged sessions in the order of service completion under the GPS scheme, or equivalently, bit-by-bit round robin. For the WF^2Q discipline, the server does not consider all the backlogged packets; rather it considers only those packets that have already started service, and possibly finished, under GPS. WF^2Q is the most fair packet-by-packet scheduler known.

In the WFQ and WF^2Q disciplines, computing the tag for selecting the next packet to be transmitted may be too complex for high speed networks. The Self-Clocked Fair Queueing (SCFQ) and Start-Time Fair Queueing (SFQ) are approximate methods computing such a tag in an simple but less fair manner [213].

In Class Based Queueing (CBO) the priority level of each packet selects a dedicated scheduler which can be of any type. Floyd and Van Jacobsen used a modified deficit version of Weighted Round Robin (WRR) [65], while Millet and Mameri used WFQ [120].

Parekh and Gallager analyzed the worst case delay in an network with PGPS schedulers and leaky-bucket constrained session flows [161, 162]. The worst case end-to-end delay was found to be a sum of per-link delay metrics. Quantiles of the end-to-end delay distribution, for statistical delay and jitter guarantees, can be computed by convolution of per-link queueing delay distributions.

2.3.2 Buffer allocation

There are basically three types of buffer allocation schemes [84]; *complete partitioning* (CP), *complete sharing* (CS), and *partial sharing* (PS). It has been shown that, e.g. [104], under a relatively balanced input condition, the CS scheme can acheive a lower blocking probability than the CP scheme. When the inputs are unbalanced, however, the buffer space may not be efficiently used by the users. The PS scheme provides a good trade-off between buffer utilization and loss probabilities among the users.

2.4 Modeling of burst traffic

Recent studies of high-quality, high-resolution traffic measurements have revealed a new phenomenon with potentially important ramifications to the modeling, design, and control of multi-service networks. These include an analysis of hundreds of millions observed packets over an Ethernet LAN in a R & D environment at Bellcore [123], an analysis of few millions of observed frame data generated by VBR video services [23]. In these studies, the packet arrival process appears to be statistically self-similar.

A self-similar (or fractal) phenomenon exhibits structural similarities across a wide range of time scales. In the case of packet traffic, self-similarity is manifested in the absence of natural length of a burst: at every time scale ranging from a few milliseconds to minutes to hours, similar-looking traffic bursts are evident.

Taqqu, Willinger and Sherman showed in [186] that the superposition of many ON/OFF sources whose ON periods and OFF periods exhibit the Noah effect (i.e. have high variability or infinite variance) produces aggregate network traffic that features the Joseph effect (i.e. is self-similar or long-range dependent). The superposition converges after scaling to fractional brownian motion (FBM), as the number of users tends to infinity.

3 QoS evaluation

3.1 Link QoS models

The fluid flow queueing model captures the behavior of burst scale congestion. Hence, performance measures derived from the fluid flow queueing model are accurate in the case of large buffers. For smaller buffer sizes, the queue analysis should be performed assuming Markov-modulated Poisson process (MMPP) packet arrivals. The MMPP queueing model captures the behavior of both packet and burst scale congestion.

The original fluid flow model for the FCFS queue with constant service rate was proposed by Kosten [109] and further developed by Anick, Mitra and Sondhi [5]. The model has been extended to handle producers and consumers coupled by a buffer by Mitra [143], rate-based congestion control by Elwalid and Mitra [55], service priorities by Elwalid and Mitra [57], Kulkarni and Gautam [116], loss priorities by Elwalid and Mitra [56], and GPS scheduling by Presti, Zhang, and Towsley [169].

The tradional assumption in fluid flow models is exponentially distributed activity periods. Recently, a fluid flow model for the FCFS queue with heavy-tailed activity-period distributions was proposed Boxma and Dumas [29]. Borst *et al.* [28], Jelenkovic *et al.* [97], Pereira *et al.* [166] and Kotopoulos *et al.* [110] analyzed the GPS system fed by heavy-tailed ON/OFF fluid sources.

Nagaraja, Kurose and Towsley [118] and Baiocchi *et al* [18] approximated the superposed arrival process by a two-state MMPP. Nagaraja, Kurose and Towsley calculated the MMPP parameters by matching of statistical moments. Baiocchi *et al.* applied a method called asymptotic matching to obtain the MMPP parameters. Both papers analyzed the multiplexer performance using a MMPP/D/1/K queueing model.

3.1.1 End-to-end QoS models

Lelarge *et al.* derived results on the asymptotic tail distribution of end-to-end delay in networks of queues with self-similar (FBM) cross traffic [129].

Approximative results for the packet delay variation are outlined by Korpeoglu *et al.* [108].

Ying *et al.* analyzed the change in burstiness as the flow traverses multiple hops. The performance distortions at each node were found the be negibly small: around 1% for mean delay and 5 % for overflow probability [208].

3.2 Grade of Service models

3.2.1 End-to-end GoS models

End-to-end GoS models for loss networks operating under Least Loaded Routing (LLR) or Markov Decision Process (MDP) routing, with Poisson flow arrival processes to the OD pairs, and exponentially distributed flow service times, have been proposed in [40, 70, 170] and [52], respectively. The end-to-end GoS measures are obtained by solving a set of Erlang fixed-point equations, also called reduced load approximation. Besides the flow traffic model and the network capacity model, the routing algorithm strongly affects the end-to-end GoS model.

4 CAC and routing

4.1 MDP-based QoS routing

State-dependent link costs can be determined from a Markov decision process (MDP) model of the the flow-level behavior of each link. With MDP-derived costs, the network is able select hop-by-hop or explicitly routed paths that maximize the long-term operator revenue.

MDP routing has recently been applied to per-flow QoS routing with on-demand flow set up and/or delayed flow set up. Lea evaluated MDP-based QoS routing for on-demand flow set up for IntServ domain networks [124]. Chang studied MDP-based QoS routing for on-demand flow set up for ATM networks with three level PNNI hierarchy [35, 36]. We analyzed MDP-based QoS routing with mixed on-demand and delayed flow set up [154, 155].

In summary, MDP-based QoS routing works as follows. First, the CAC_{QoS} function finds the set of feasible paths that satisfies the end-to-end QoS and administrative constraints of the requested flow class. Second, the routing function selects a minimum MDP cost path for the new flow. Third, the CAC_{GoS} function accepts (rejects) this choice if the minimum path cost is smaller (larger) than the expected reward from serving this flow.

Numerical experiments carried out by several auhors show that MDP-based routing gives higher average revenue rate than Least Loaded Routing (LLR), Event Dependent Routing (EDR), and sequential routing [47].

MDP-based routing is not implemented in any real network around the world, which is explained by the modest improvement of average revenue rate, and the theoretical complexity of algorithm.

To implement MDP-based routing it makes sense to have a bandwidth broker (BB) distributed among the edge routers. The origin (edge) node is connected to a set of destination (edge) nodes via a routing network. The state of this routing network provides the basis for generic (preliminar) CAC and routing decisions for new calls between the OD-pairs. Signalling is then done along the chosen network path to check the actual state of each link (actual CAC).

It is desirable that the state of the routing network is known with resonable accuracy. The update of link states can be done periodically or driven by some event process. Paths which are lightly loaded can be evaluated as feasible without exact state information; as paths become more loaded the decisison becomes more critical and the state update frequency should increase.

5 Network design

5.1 Bandwidth and buffer management

Optimal weight selection for GPS schedulers under statistical loss, overflow and/or delay guarantees was studied by Kumaran *et al.* [117], Elwalid *et al.* [58], Lieshout *et al.* [131], Lapiotis *et al.* [121], and Lee *et al.* [128]. Optimal weight selection under deterministic delay guarantees was considered by Szabo *et al.* [185], Georgiadis *et al.* [69], Panagakis *et al.* [160]. GPS weight selection in CDMA networks taking fading and inter-cell interference into acount was studied by Xu *et al.* [199, 200].

Optimization of buffer management by complete sharing based on virtual partitioning has been studied by Lapioties *et al.* [121] for the GPS scheduler, and by Wu *et al.* [196] for the FCFS scheduler.

5.2 Network topology

The second function in network design is the design of the topological structure of the network, i.e. where to place the nodes and how to interconnect them.

Harms *et al.* [80] studies the global topological reconfiguration problem for physical networks. For ATM and MPLS physical networks in most cases the result of the topological design phase will lead to a partly or fully meshed backbone network structure.

Anjali *et al.* have proposed a successive topology reconfiguration policy for the VN MPLS network [6]. Srikitja *et al.* analyze the global topological reconfiguration problem for VNs over MPLS [182].

5.3 Network link capacities

Groskinkky *et al.* propose a method for successive capacity reallocation based on analysis of a time-dependent loss queueing system [73].

Global capacity reconfiguration determines, given the network topology, traffic demand and GoS requirements, the capacities of the physical and virtual network links. The objective used in design of link capacities in virtual and physical networks can be of several types. Two common examples are maximization of the average revenue rate and minimization of total network link cost. The GoS constraints for each flow class are expressed in an absolute or relative manner. The global capacity reconfiguration task can be formulated as an optimization problem with non-linear objective function subject to a set of non-linear constraints. The optimization task requires a model of the network GoS, which besides from the traffic and capacity model, also depends on the CAC and routing policy.

Results for global capacity reconfiguration have been derived for

- single-service networks under fixed routing and load sharing routing by Kelly [100];
- multi-service networks under fixed routing and load sharing routing by Farago [59] and Mitra [144, 145];
- single-service networks under least loaded routing by Huberman [86] and Girard [71];
- multi-service networks under MDP routing by Dziong [53];
- multi-service networks under multi-commodity flow routing model by Girard [72].

References

[1] M. Aissa, A.B. Mnaouer, “A new delay-constrained algorithm for multicast routing tree constriuction”, *International Journal of Communication Systems*, Vol. 17, No. 10, pp. 985-1000, 2004.

[2] T. Alrabiah, T. Znati, “Delay-constrained, low-cost multicast routing in multimedia networks”, *Journal of Parallel and Distributed Computing*, 61, pp. 1307-1336, 2001.

[3] A. Andersen, B. Nielsen, “A Markovian approach for modeling packet traffic with long-range dependence”, *Journal of Selected Areas in Communications*, Vol. 16, No. 5, Jun. 1998.

[4] S. Androulidakis, T. Doukolou, G. Patikis, D. Kagklis, ”Service differentiation and traffic engineering in IP over WDM networks”, *IEEE Communcations Magazine*, Vol. 46 , No. 5, 2008.

[5] D. Anick, D. Mitra, M. Sondhi, “Stochastic theory of a data handling system with multiple sources”, *Bell System Technical Journal*, Vol. 61, No. 8, pp. 1871-1894, 1982.

[6] T. Anjali, C. Scoglio, J. de Oliveira, I. Akyildiz, G. Uhl, “Optimal policy for labeled switched path setup in MPLS networks”, *Computer networks*, Vol. 39, No. 2, 2002.

[7] ATM Forum, “Traffic management specification”, Version 4.0, 1996.

[8] ATM Forum, “PNNI specification Ver 1.0, 1996.

[9] G. Apostolopoulos, R. Guérin, S. Kamat, S. Tripathi, “Quality of service based routing: a performance perspective”, In *Proc. ACM Sigcomm’98*, Vancouver, B.C., 1998.

[10] G. Apostolopoulos, D. Williams, S. Kamat, R. Guérin, A. Orda, T. Przygienda, “QoS routing mechanisms and OSPF extensions”, IETF RFC 2676, IETF RFC 2702, Available from ¡http://www.ietf.org/rfc/rfc2676¿, Aug. 1999.

[11] G. Apostplopoulus, R. Guerin, ”Intradomain QoS routing in IP networks: a feasibility and cost/benefit analysis”, *IEEE Network*, Vol. 13, No. 5, 1999.

[12] G. Ash, *Traffic Enigneering and QoS Optimization of Integrated Voice & Data Networks*, Morgan Kaufmann, 2007.

[13] P. Ashwood, L. Berger, “Generalized multi-protocol label switching (GMPLS) constraint-based routed label distribution protocol (CR-LDP) extensions”, IETF RFC 3472, Available from ¡http://www.ietf.org/rfc/rfc3472¿, Jan. 2003.

[14] I. Atov, R.J. Harris, A. Cassin, ”Capacity planning tool for multiservice IP networks”, In *Proc. of Networks*, 2004.

[15] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow,”RSVP-TE: extensions to RSVP for LSP tunnels”, IETF 3209, Available from ¡http://www.ietf.org/rfc/rfc3209¿, Dec. 2001.

[16] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao, “Overview and principles of Internet traffic engineering”, IETF RFC 3272, Available from ¡http://www.ietf.org/rfc/rfc3272¿, May 2002.

[17] D. Awduche, J. Malmcom, J. Agogbua, M. O’Dell, J. McManus, “Requirements of traffic engineering over MPLS”, IETF RFC 2702, Available from ¡http://www.ietf.org/rfc/rfc2702¿, Sep. 1999.

[18] A. Baiocchi, N. Melazzi, M. Listanti, A. Roveri, R. Winkler, “Loss performance analysis of an ATM multiplexer loaded with high-speed on/off sources”,*Journal of Selected Areas in Communications*, Vol. 9, No. 3, pp. 388-393, 1991.

[19] F. Baker, C. Iturralde, F. Faucheur, B. Davie, “Aggregation of RSVP for IPv4 and IPv6 reservations”, IETF RFC 3175, Available from ¡http://www.ietf.org/rfc/rfc3175¿, Sep. 2001.

[20] S.M. Banik, S. Radhakrishnan, C.N. Sekharan, “Multicast routing with delay and delay variation constraints for collaborative applications on overlay networks”, *IEEE Transactions on Parallel & Distributed Systems*, Vol. 18, No. 3, pp. 421-432, 2007.

[21] A.R. Bashandy, E.K.P. Chong, A. Ghafoor, “Generalized quality-of-service routing with resource allocation”, *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 2, pp. 450-463, 2005.

[22] P. Belotti, A. Capone, G. Carello, F. Malucelli, ”Multi-layer MPLS network design: the impact of statistical multiplexing”, *Computer Networks*, Vol. 52, 2008.

[23] Beran J., Sherman R., Taqqu M., Willinger W., Long-range dependence in variable-bit-rate video traffic, *IEEE Transactions on Communications*, Vol. 43, pp. 1566-1579, 1995.

[24] L. Berger, “Generalized multi-protocol label switching (GMPLS) signaling functional description”, IETF RFC 3471, Available from ¡http://www.ietf.org/rfc/rfc3471¿, Jan 2003.

[25] S. Bhulai, “On the value function of the M/Cox(r)/1 queue”, *J. Appl. Prob.*, Vol. 43, No. 2, 2006.

[26] E. Biton, A. Orda, “QoS provision and routing with stochastic guarantees”, In *Proc. QSHINE*, 2004.

[27] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, “An architecture for differentiated services”, IETF RFC 2475, Available from ¡http://www.ietf.org/rfc/rfc2475¿, Dec. 1998.

[28] S. Borst, M. Mandjes, M. van Uitert, “Generalized processor sharing with light-tailed and heavy-tailed input”, *IEEE/ACM Transactions on Networking*, Vol. 11, No., 5, pp. 821-834, 2003.

[29] O. Boxma, V. Dumas,”Fluid queues with long-tailed activity period durations”, Special issue on Stochastic Analysis and Optimization of Communication Systems, *Journal of Computer Communications*, 1998.

[30] R. Braden, L. Zhang. S. Berson, S. Herzog, S. Jamin, “Resource ReSerVation Protocol (RSVP) – version 1 functional specification”, IETF RFC 2205, Available from ¡http://www.ietf.org/rfc/rfc2205¿, Mar. 2002.

[31] T. Braun, M. Guenter, I. Khalil,”Management of Quality of Service enabled VPNs”, *IEEE Communications Magazine*, Vol. 39, No. 5, pp. 90-98, May 2001.

[32] K. Calvert, M. Doar, E. Zegura, “Modelling Internet topology”, *IEEE Communication Magazine*, Vol. 35, No. 6, pp. 160-164, 1997.

[33] D. Chakraborty, G. Chakraborty, N. Shiratori, “A dynamic multicast routing satisfying multiple QoS constraints”, *International Journal of Network Management*, Vol. 13, No.5, pp.321-335, 2003.

[34] B. Carpenter, K. Nichols, “Differentiated services in the Internet”, *Proceedings of the IEEE*, Vol. 90, No. 9, pp. 1479-1494, Sep. 2002.

[35] B. Chang, R. Hwang, “Hierarchical QoS routing in ATM networks based on MDP cost function”, In *Proc. International Conference on Networks*, pp. 147-151, 2000.

[36] B. Chang, R. Hwang, “Efficient hierarchical QoS routing in ATM networks”, *Computer Communications*, Vol. 24, No. 15-16, pp. 1648-1660, 2001.

[37] R. de Cassia, R. Rodriques, S. Carvalho, “Optimization of a PH/PH/1/N queue with controllable service time and partial information”, *International Transactions on Operations Research*, Vol. 8, No. 2, pp. 235-248, 2001.

[38] S. Chen, K. Nahrstedt, "An overview of quality-of-service routing for next generation high-speed networks: problems and solutions", *IEEE Network*, Vol. 12, No. 6, 1998.

[39] S. Chen, M. Song, S. Sahni, "Two techniques for fast computation of constrained shortest paths", *IEEE/ACM Transactions on Networking*, Vol. 16, No. 1, 2008.

[40] S. Chung, A. Kashper, K. Ross, "Computing approximate blocking probabilities for large loss networks with state-dependent routing", *IEEE/ACM Transactions on Networking*, Vol.1, No. 1, pp. 105-115, 1993.

[41] C. Chaudhary, K. Drishnan, S. Vishnubhatla, "Technology-trial of dynamic traffic routing with 5 minutes updates", In *Proc. Global Telecommunications Conference*, Vol. 3, pp. 1696-1700, 1991.

[42] J. Chu, C. Lea, A. Wong, "Cost-based QoS routing", In *Proc. 12th International Conference on Computer Communications and Networks*, pp. 485-490, 2003

[43] C. Chi, D. Huang, D. Lee, X. Sun, "Lazy flooding: a new technique for information dissemination in distributed network systems", In *Proc. International Conference on Info-tch and Info-net*, Vol. 2, pp. 1-23, 2001.

[44] M. Crovella, A. Bestavros, "Self-similarity in world wide web traffic – evidence and possible causes",*IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, pp. 835-846, 1997.

[45] A. Demers, S. Keshav, S. Shenker, "Analysis and simulation of a fair queueing algorithm", *Journal of Internetworking: Research and Experience*, pp. 3-26, Oct. 1990, Also in *Proc. ACM SIGCOMM'89*, pp. 3-12.

[46] A. Diwan, J.Kuri, "Routing and rate allocation in rate-based multi-class networks", *Performance Evaluation*, Vol. 59, No.4, 2005.

[47] Z. Dziong, *ATM Network Resource Management*. McGraw-Hill (ISBN 0-07-018546-8), 1997.

[48] Z. Dziong, L. Mason, "Control of multi-service loss networks", in *Proc. of 28-th IEEE Conference on decision and control*, Tampa, Florida, USA, Dec. 1989.

[49] Z. Dziong Z., K-Q Liao, L.G. Mason, "Flow control models for multi-service networks with delayed flow set up", In *Proc. of Infocom'90*, pp. 39-46, IEEE Computer Society Press, 1990.

[50] Z. Dziong, L. Mason, "Call admission and routing in multi-service loss networks", *IEEE Transactions on Communications*, Vol. 42, No. 2, 1994.

[51] Z. Dziong, M. Jia, L.G. Mason, "Analysis of multicast routing algorithms for broadband networks", ' In *Proc. ATM Workshop*, pp. 186-194, 1998.

[52] Z. Dziong, J. Mignault, and C. Rosenberg, "Blocking Evaluation for Networks with Reward Maximization Routing", In *Proc. of Infocom'93*, San Fransisco, USA, 1993.

[53] Z. Dziong, Y. Xiong, L. Mason,"Virtual network concept and its application for resource management in ATM based networks", In *Proc. IFIP/IEEE Broadband Communications'96*, pp. 223-234, 1996.

[54] T. Engel, H. Granzer, B. Koch, M. Winter, P. Sampatakos, I. Venieris, H. Hussmann, F. Ricciato, S. Salsano,"AQUILA- adaptive resource control for QoS using an IP-based layered architecture", *Communications Magazine*, Vol. 41, No. 1, pp. 46-53, Jan. 2003

[55] A.I. Elwalid, D. Mitra, "Analysis and design of rate-based congestion control in high-speed networks, I: stochastic fluid models, access regulation", *Queueing Systems (Special Issue on Communication Systems)*, Vol. 9, pp. 29-64, 1991.

[56] A. Elwalid, D. Mitra,"Fluid models for the analysis and design of statistical multiplexing with loss priorites on multiple classes of bursty traffic", In *Proc. IEEE Infocom'92*, pp. 415-425, 1992.

[57] A. Elwalid, D. Mitra,"Analysis, approximations and admission control of a multi-service multiplexig system with priorities", In *Proc. IEEE Infocom'95*, pp. 463-472, 1995.

[58] A. Elwalid, D. Mitra,"Design of Generalized Processor Sharing schedulers which statistically multiplex heterogeneous QoS classes", In *Proc. IEEE Infocom'99*, pp. 1220-1230, 1999.

[59] A. Farago, S. Blaabjerg, G. Gordos, T. Henk,"A new degree of freedom in ATM network dimensioning: optimizing the local configuration", *IEEE Journal of Selected Areas in Communications*, Vol. 13, No. 7, Sept. 1995.

[60] A. Feldmann, "Characteristics of TCP connection arrivals", Tech. Rep., AT&T Labs Research, 1998.

[61] A. Feldmann, A. Gilbert, A. Willinger, T. Kurtz, “The Changing Nature of Network Traffic: Scaling Phenomena”, *Computer Communication Review*, Vol. 28, No. 2, pp. 5-29, April 1998.

[62] A. Feldmann, W. Whitt, “Fitting mixtures of exponentials to long-tail distributions to analyze network performance models”, *Performance Evaluation*, Vol. 31, No. 3-4, pp. 245-279, 1998.

[63] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, ”NetScope: traffic engineering for IP networks”, *IEEE Network*, Vol. 14, No. 2, 2000.

[64] G. Feng, “Tightest constraint first: an efficient delay sensitive multicast routing algorithm”, *International Journal of Communication Systems*, Vol. 18, o. 7, pp. 671-685, 2005.

[65] S. Floyd, V. Jacobsen, “Link-sharing and resource management models for packet networks”, *IEEE/ACM Transactions on Networking*, Vol. 3, No. 4, Aug. 1995.

[66] D. R. Figueiredo, B. Liu, Y. Guo, J. Kurose, D. Towsley, ”On the efficiency of fluid simulation of networks”, Computer Networks, Vol. 50, 2006.

[67] W. Fisher, K. Meier-Hellstern, ”The Markov-modulated Poisson process (MMPP) cookbook”, Performance Evaluation, 18, North-Holland, 1992.

[68] V. Frisen, J. Harms,”Resource management with virtual paths in ATM networks”, *IEEE Network*, Vol. 10, No. 5, pp. 10-21, 1996.

[69] L. Georgiadis, R. Guérin, V. Peris, K. Sivarajan, “Efficient network qos provisioning based on per node traffic shaping”, *IEEE/ACM Transactions on Networking*, Vol. 4, No. 4, pp. 482-501, Aug. 1996.

[70] A. Girard, M.A. Bell,”Blocking evaluation for networks with residual capacity adaptive routing”, *IEEE Transactions on Communications*, Vol. 37, pp. 1372-1380,l 1989.

[71] A. Girard, N. Liau,”Dimensioning of adaptively routed networks”, *IEEE/ACM Transactions on Networking*, Vol. 1, pp. 460-468, Aug. 1993.

[72] A. Girard, B. Sansó, “Multicommodity flow models, failure propagation, and reliable loss network design”, *IEEE/ACM Transactions on Networking*, Vol. 6, No. 1, Feb. 1998.

[73] B. Grosksinsky, D. Mehdi, D. Tipper, “An investigation of adaptive capacity control schemes in a dynamic traffic environment”, *IEICE Transactions on Communications*, Vol. E00-A, No. 13, 2001.

[74] D. Ghosh, V. Sarangan, R. Acharya, ”Quality-of-service routing in IP networks”, *IEEE Transactions on Multimedia*, Vol. 3, No. 2, Jun. 2001.

[75] R. Guérin, A. Orda,”QoS routing with inaccurate information”, *Transactions on Networking*, Vol. 7, No. 3, pp. 350-364, 1999.

[76] V.Gurbani, L. Jagadeesan, V. Mediratta, “Characterizing session initiation protocols network performance and reliability”, In *Proc. of International Conference on Software Engineering* (CICSE), pp 472-482, Shanghai, China, 2006

[77] R. Gusella, ”Characterizing the variability of arrival processes with indexes of dispersion”, IEEE Journal on Selected Areas in Communications, Vol. 9, No. 2, 1991.

[78] F. Hao, E.W. Zegura, S. Bhatt, “Performance of the PNNI protocol in large networks”, In *Proc. ATM Workshop*, pp. 315-323, 1998.

[79] H.A. Harira, S. Pierre, “Joint routing and admission control under statistical delay and jitter constraints in MPLS networks”, *Computer Communications*, Vol. 31, 2008.

[80] J. Harms, C. Hu, “Physical network design to facilitate capacity reallocation”, *Discrete Applied Mathematics*, Vol. 129, No. 1, pp. 129.153, Elsevier, 2003.

[81] H. Heffes, D.M Lucantoni, ”A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance”, IEEE Journal on Selected Areas in Communications, Vol. SAC-4, No. 6, 1986.

[82] P. Heidelberger, ”Fast simulation of rare events in queueing and reliability models”, *ACM Transactions on Modelling and Computer Simulation (TOMACS)*, Vol. 5, No. 1, 1995.

[83] S. Herzog, “RSVP extensions for policy control”, IETF RFC 2750, Available for ¡http://www.ietf.org/rfc/rfc2750, Jan. 2000.

[84] M.G. Hluchyj, M.J. Karol,”Queueing in high performance packet switching”, *IEEE Journal of Selected Areas in Communications*, Vol. SAC-6, No. 9, Dec. 1988.

[85] T.-L. Huang, D.T. Lee, "A distributed multicast routing ' algorithm for real-time applications in wide-area networks", *Journal of Parallel and Distributed Computing*, Vol. 67, pp. 516-530, 2007.

[86] R. Huberman, S. Hurtubise, S.A. Le Nir, T. Drwiega, "Multihour dimensioning for a dynamically routed network", In *Proc.International Teletraffic Congress*, pp. 4.3A-5-1–4.3A-5-7, Sept. 1985.

[87] J. Hui, M. Gursoy, N. Moayeri, Y. Yates, "A layered broadband switching architecture with physical or virtual path configurations", *Journal of Selected Areas in Communications*, Vol. 9, No. 9, pp. 1416-1426, 1991.

[88] R. Hwang, J. Kurose, D. Towsley, "State-dependent routing for multirate loss networks", In *Proc. Globecom'92*, pp. 740-747, Orlando, 1992.

[89] R.H. Hwang, "Adaptive multicast routing in multirate loss networks", *Telecommunication Systems*, Vol. 12, No. 4, pp. 283-313, 1999.

[90] ITU-T, "Traffic control and congestion control in B-ISDN", I.371, 2004.

[91] ITU-T, "International emergency preference scheme (IEPS) for distaster and relied operations", E.106, 2003.

[92] ITU-T, '"Packet based multimedia communication systems", H.232, 2006.

[93] X. Jia, "A distributed algorithm of delay-bunded multicast routing for multimedia applications in wide area networks", *IEEE/ACM Transactions on Networking*, Vol. 6, No. 6, pp. 828-837, 1998.

[94] Y. Jia, J. Nikoolaidis, P. Gburzynski, "On the effectiveness of alternative paths in QoS routing", *International Journal of Communication Systems*, Vol. 17, pp. 1-26, 2004.

[95] Z. Jia, P. Varaiya, "Heuristic methods for delay constrained least cost routing using k-shortest-paths", *IEEE Transactions on Automatic Control*, Vol. 51, No. 4, Apr. 2006.

[96] J.M. Jaffe,"Algorithms for finding paths with multiple constraints", *Networks*, Vol. 14, pp. 95-116, 1984.

[97] P. Jelenkovic, P. Momcilovic,"Finite buffer queue with generalized processor sharing and heavy-tailed input processes", *Computer Networks*, Vol. 40, No. 3, pp. 433-444, 2002.

[98] S.H. Kang, D.K. Sung, "A CAC scheme based on real-time cell loss estimation for ATM multiplexers", *IEEE Transactions on Communications*, Vol. 48, No. 2, 2000.

[99] P. Khadivi, S. Samavi, T.D. Todd, "Multi-constraint QoS routing using a new single mixed metric", *Journal of Network and Computer Applications*, Vol. 31, 2008.

[100] F. Kelly, "Routing in circuit-switched networks: optimization, shadow-prices and decentralization", *Adv. Appl. Prob.*, No. 20, 1988.

[101] D. G. Kendall, "Some problems in the theorem of queues", *J. Royal. Statistical Soc.*, Vol. 13, pp. 151-185, 1951.

[102] G. de Veciana, G. Kesidis, "Bandwidth allocation for multiple Qualities if Service using Generalized Processor Sharing", *IEEE Transactions on Information Theory*, Vol. 42, No. 1, pp. 268-272, Jan. 1996.

[103] G. Kesidis, A Singh, D. Cheung, W.W. Kwok, "Feasibility of fluid event-driven simulation for ATM network", In *Proc. GLOBECOM*, Vol. 3, pp. 2013-2017, 1996.

[104] P. Kamoun, L. Kleinrock, "Queueing in shared finite storage in a computer network node environment under general traffic condition", *IEEE Transac- tions on Communications*, vol. COM-28, No. 7, July 1980.

[105] C. Kiddle, R. Simmonds, C. Willianson, B. Unger, "Hybrid packet/fluid flow network simulation", In *Proc. 7th Workshop on Parallel and Distributed Simulation*, pp. 143-152, 2003.

[106] A. Kolarov, J. Hui, "On computing Markov decision theory-based cost for routing in circuit-switched broadband networks", *Journal of Network and Systems Management*, Vol. 3, No. 3, pp. 405-426, 1995.

[107] V. Kompella, J.C. Pasquale, G.C. Polyszos, "Multicast routing for multimedia communcations", *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp. 286-292, 1993.

[108] I. Korpeoglu, M. Ebersprächer, H. Theimer, P.Khün, U. Briem,"Approximate analysis of the end-to-end delay in ATM networks", In *Proc. of Infocom'92*, pp. 978-986, 1992.

[109] L. Kosten, "Stochastic theory of a multi-entry buffer", *Delft Progress Report Series F*, Vol. 1, pp. 10-18, 1974.

[110] C. Kotopoulos, N. Likhanov, R. Mazumdar, "Asymptotic analysis of GPS systems fed by heterogeneous long-tailed sources", In *Proc. of Infocom*, 2001.

[111] K. Krishnan, F. Hübner, "Admission control and routing for multirate circuit-switched traffic", In *Proc. of 15-th International Teletraffic Congress*, ITC'15, Washington, USA, 1997.

[112] H. Kröner, M. Eberspрächer, H. Theimer, P. Kühn, U. Briem, "Approximate analysis of the end-to-end delay in ATM networks", In *Proc. of Infocom'92*, pp. 978-986, 1992.

[113] F. Kuipers, P. Mieghem, T. Korkmaz, M. Krunz, "An overview of constraint-based path selection algorithms for QoS routing", *IEEE Communications Magazine*, Vol. 40, No. 12, pp. 50-56, Dec. 2002.

[114] F. Kuipers, T. Korkmaz, M. Krunz, P. Mieghem, "Performance evaluation of constraint-based path selection algorithms", *IEEE Network*, Sep./Oct. 2004.

[115] F. A. Kuipers, P.F.A Van Mieghem, "Conditions that impact the complexity of QoS routing", *IEEE/ACM Transactions on Networking*, Vol. 13, No. 4, pp. 717-730, 2005.

[116] V. G. Kulkarni, N. Gautum, "Admission control of multi-class traffic with service priorities in high-speed networks", *Queueing Systems*, Vol. 1997, No. 1, pp. 79-97, 1997.

[117] K. Kumuran, G. Margave, D. Mitra, K. Stanley, "Novel techniques for the design and control of generalized processor sharing schedulers for multiple QoS classes, In *Proc. of Infocom'00*, Vol. 2, pp. 932-941, 2000.

[118] R. Nagarajan, J. Kurose, D. Towsley, "Approximation techniques for computing packet loss in finite-buffered voice multiplexers", *Journal of Selected Areas in Communications*, Vol. 9, No. 3, pp. 368-377, 1991.

[119] D. Medhi, S. Guptan,"Network dimensioning and performance of multiservice, multirate loss networks with dynamic routing", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, Dec. 1997.

[120] A. Millet, Z. Mameri, "Delay bound guarantees with WFQ-based CBQ discipline", In *Proc. IWQOS*, 2004.

[121] G. Lapiotis, S. Panwar,"Quality of service analysis of shared buffer management policies combined with generalized processor sharing", In *Proc. Globecom'99*, Vol. 1a, pp. 37-43, 1999.

[122] W. Lazarev, S. Strabinets, "The use of dynamic programming for optimisation of control in networks of computations of channels", *Engineering Cybernetics (Academy of Sciences, USSR)*, No. 3, 1977.

[123] Leland W, Taqqu M., Willinger W. and Wilson D., On the self-similar nature of Ethernet traffic (extended version), *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, pp 1-15, 1994.

[124] C. Lea, K. Ke, "Quantization and cost computation of MDP-based admission and routing", In *Proc. Infocom'96*, Vol. 3, pp. 1004-1011, 1996.

[125] J. Y. Lee, S. Kim, D. Kim, D. Sung, "Bandwidth optimization for Internet traffic for generalized processor sharing servers", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 16, No. 4, pp. 324-334, 2005.

[126] J. Y. Lee, S. Kim, D. Sung, "Fluid simulation for self.-similar traffic in generalized processor sharing servers", *IEEE Communication Letters*, Vol. 7, No. 6, pp. 272-274, 2003.

[127] J.W. Lee, B.G. Lee, "Performance analysis of an ATM multiplexer with MMPP input", *IEICE Transactions on Communications*, Vol. E75-B, No. 8, 1992.

[128] Y.-J. Lee, M. Atiquzzaman, "Least cost heuristic for the delay-contstrained capacited minimum spanning tree problem", *Computer Communications*, 28, pp. 1371-1379, 2005.

[129] M. Lelarge, Z. Liu, C. Xia, "Asymptotic tail distribution of end-to-end delay in networks of queues with self-similar cross traffic", In *Proc. of Infocom'04*, Vol. 4, pp. 2352-2363, 2004.

[130] H. Levy, T. Mendelson, G. Goren, "Dynamic allocation of resources to virtual path agents", *IEEE/ACM Transactions on Networking*, Vol. 12, No. 4, pp. 746-758, 2004.

[131] P.M.D. Lieshout, M. Mandjes, S.C. Borst, "GPS scheduling: selction of optimal weights and comparison with strict priorities", *Submitted for publication*, 2005.

[132] X. Lin, N. Shroff, "Utility maximization for communication networks with multipath routing", *IEEE Transactions on Automatic Control*, Vol. 51, No. 5, May 2006.

[133] X. Lin, N. Shroff, "An optimization-based approach for QoS routing in high-bandwidth networks", *IEEE/ACM Transactions on Networking*, Vol. 14, No. 6, Dec. 2006.

[134] D. Lucantoni, K. Meier-Hellstern, M. Neuts, "A single-server queue with server vacations and a class of non-renewal arrival processes", *Adv. Appl. Prob.*, Vol. 22, pp. 676-705, 1990.

[135] D. Logothetis, K. Trivedi, A. Puliafito, "Markov generative models", *Computer Performance and Dependability Symposium*, pp. 134-142, 1995.

[136] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, J.D. Robbins, "Performance models of statistical multiplexing in packet video communications", *IEEE Transactions on Communications*, Vol. 36, No. 7, pp. 834-844, 1988.

[137] P. Marbach, O. Mihatsch, J. Tsitsiklis,"Call admission control and routing in integrated services networks using neuro-dynamic programming", *IEEE Transactions on Communications*, Vol. 18, No. 2, pp. 197-208, 2000.

[138] K. Meier-Hellstern, "The analysis of a queue arising in overflow models", *IEEE Transactions on Communications*, Vol. 37, No. 4, pp. 367-372, 1989.

[139] M. Menth, "Analysis and design of advanced multiservice networks supporting mobility, multimedia and internetworking", *COST-279*, Midterm report, Jan. 2004.

[140] H. Michiel, K. Laevens, "Teletraffic engineering in a broadband era", *Proceedings of the IEEE*, Vol. 85, No. 12, pp. 2007-2033, Dec. 1997.

[141] P. Van Mieghem, (ed), F. Kuipers, T. Kormaz, M. Krunz, M. Curado, E. Monteiro, X. Masip-Bruin, J. Sole-Pereta, S. Sanchez-Lopez, "Quality of service routing", Chapter 3 of Future Internet Services, *EU-COST 263 Final Report*, edited by Smirnow et al in Springer LNCS 2856, pp. 80-117, 2003.

[142] P. Mieghem, F. Kuipers, "Concept of exact QoS routing algorithms", *IEEE/ACM Transactions on Networking*, Vol. 12, No. 5, Oct. 2004.

[143] D. Mitra,"Stochastic theory of a fluid model of producers and consumers coupled by a buffer", *Adv. Applied Probability*, Vol. 20, pp. 646-676, 1988.

[144] D. Mitra, A. Morrison, K. Ramkrishnan, "ATM network design and optimization: a multirate loss network framework", *IEEE/ACM Transactions on Networking*, Vol. 4, No. 4, 1996

[145] D. Mitra, J. Morrison, K. Ramakrishnan, "Virtual private networks: joint resource allocation and routing design" In *Proc. Infocom'99*, Vol. 2, pp. 480-490, 1999.

[146] S. Molnar, I. Maricza, "Source characterization in broadband networks", *Cost 257 Mid-Term Seminar*, Interim report, 2000.

[147] R. Nagarajan, J.F. Kurose, D. Towsley, "Approximation techniques for computing packet loss in finite-buffered voice multiplexers", *IEEE Journal on Selected Areas in Communications*,' Vol. 9, No. 3, 1991.

[148] R. Nagarajan, J.F. Kurose, D. Towsley, "Approximation techniques for computing packet loss in finite-buffered voice multiplexers",*Technical report*

[149] U. T. Nguyen, J. Xu, "Multicast routing in wireless mesh networks: minimum cost trees or shortest path threes?", *IEEE Communications Magazine*, Vol. 45, No. 11, pp 72-77, 2007.

[150] D. Nicol, M. Goldsby, M. Johnsson, "Fluid-based simulation of communication networks using SSF", In *Proc. European Simulation Symposium*, 1999.

[151] S. Norden,"Inter-domain routing: algorithms for QoS guarantees", *Computer Networks*, Vol. 49, No. 4, pp. 493-620, 2005.

[152] E. Nordström, J. Carlström, "A reinforcement learning scheme for adaptive link allocation in ATM networks", IWANNT*95 workshop, pp 88-95,Lawrence Erlbaum, Stockholm, Sweden, 1995.

[153] E. Nordström, J. Carlström, "A new reward model for MDP state aggregation with application to CAC and routing", *European Transactions on Telecommunications*,16:495-508, 2005.

[154] E. Nordström, Z. Dziong, "CAC and routing in multi-service networks with blocked wide-band calls delayed, Part I: exact link MDP framework",*European Transactions on Telecommunications*, 17:21-36, 2006.

[155] E. Nordström, Z. Dziong, "CAC and routing in multi-service networks with blocked wide-band calls delayed, Part II: approximate link MDP framework", To appear in *European Transactions on Telecommunications*, 2006.

[156] E. Nordström, "QoS analysis of GPS systems", working paper, 2008.

[157] R. Onvural, *Asynchronous Transfer Mode networks*, second edition, Artech House, 1995.

[158] T. Ott, K. Krishnan, "State-dependent routing of telephone traffic and the use of separable routing schemes", In *Proc. of 11-th International Teletraffic Congress*, ITC'11, 1985.

[159] C. Oottamakorn, S. Mao, S.S. Panwar, "On generalized processor sharing with regulated multimedia traffic flows", IEEE Transactions on Multimedia, Vol. 8, No. 6, Dec. 2006.

[160] A. Panagakis, N. Dukkipati, I. Stavrakakis, J. Kuri, "Optimal call admission control a single link with a GPS scheduler", *IEEE/ACM Transactions on Networking*, Vol. 12, No. 5, pp. 865-878, 2004.

[161] A. Parekh, R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single node case", *Transactions on Networking*, Vol. 1, No. 3, pp. 344-357, 1993.

[162] A. Parekh, R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the multiple node case", *Transactions on Networking*, Vol. 2, No. 2, pp. 137-150, 1994.

[163] W. Park, J.Y. Lee, D.K. Sung, "Bandwidth optimization algorithm based on weight vector adjustment in generalized processor sharing servers", *IEICE Transactions on Communications*, Vol. E90-B, No. 1, 2007.

[164] M. Parsa, Q. Zhu, J.J. Garcia-Luna-Aceves, "An iterative algorithm for delay-constrained minimumcost multicasting", *IEEE/ACM Transactions on Networking*, Vol. 6, No. 4, pp. 461-474, 1998.

[165] V. Paxson, S. Floyd, "Wide-area traffic: the failure of Poisson modeling", *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226-244, June 1995.

[166] F. de M Pereira, N. Fonseca, D. Arantes, "On the performance of generalized processor sharing under long-range dependent traffic", *Computer Networks*, Vol. 40, No. 3, pp. 413-431, 2002.

[167] M. Peyravian, R. Onvural, "Algorithm for efficient generation of link-state updates in ATM networks", *Computer Networks and ISDN Systems*, Vol. 29, pp. 237-247, 1997.

[168] M. Pioro, D. Medhi, "Routing, flow, and capacity design in communication and computer networks", Morgan Kaufmann, 2004.

[169] F. Presti, Z. Zhang, D. Towsley, "Bounds, approximations and applications for a two-queue GPS system", In *Proc. Infocom'96*, Vol. 3, pp. 1310-1317, 1996.

[170] J. Rénier, P. Blondeau, W.H. Cameron, "Grade of service of a dynamic call routing system", In *Proc. 10th International Teletraffic Congress*, Session 3.2,1983.

[171] N.-E. Rikli, "Modeling techniques for VBR video: feasibility and limitations", *Performance Evaluation*, Vol. 57, No. 1, pp. 57-68, 2004.

[172] E. Rosen, A. Viswanathan, R. Flowon, "Multiprotocol label switching architecture", IETF RFC 3031, Available for ¡http://www.ietf.org/rfc/rf3031¿, Jan. 2001.

[173] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Parks, M. Handley, E. Schooler, "SIP: session initiation protocol", IETF RFC 3261, Available for ¡http://www.ietf.org/rfc/rf3261¿, Jun. 2002.

[174] H. Rummukainen, J. Virtamo, "Polynomial cost approximations in Markov decision theory based call admission control", *IEEE/ACM Transactions on Networking*, Vol. 9, No. 6, Dec. 2001.

[175] H.F. Salama, D.S. Reeves, Y. Viniotis, "Evaluation of multicast routing algorithms for real-time communication on high-speed networks", *IEEE Journal of Selected Areas in Communications*, Vol. 15, No. 3, 1997.

[176] S. Shah-Heydari, T. Le-Ngoc, "MMPP models for multimedia traffic", *Telecommunications Systems*, Vol. 15, No. 3, 2000.

[177] A. Shaikh, J. Rexford, K.G. Shin, "Evaluating the impact of stale link state on quality-of-service routing", *IEEE/ACM Transactions on Networking*, Vol. 9, No. 2, pp. 162-176, Apr. 2001.

[178] Y. Serres, L. Mason, "A multi-server queue, with narrow- and wide-band customers and WB restricted access", *IEEE Transactions on Communications*, Vol. 36, No. 6, June 1988.

[179] H. Schulzrinne, J. Polk, "Communication resource priority for the session initiation protocol (SIP)", Internet-draft, Jul. 2005.

[180] V. Sivaraman, F.M. Chiussi, M. Gerla, "End-to-end statistical delay service under GPS and EDF scheduling: a comparison study", In *Proc. Infocom'01*, Vol. 2, pp. 1113-1122, 2001.

[181] V. Sivaraman, F.M. Chiussi, M. Gerla, "Deterministic end-to-end delay guarantees with rate controlled EDF scheduling", *Performance Evaluation*, Vol. 63, No. 4-5, pp. 509-519, 2006.

[182] A. Srikitja, D. Tipper, "Topological design of multiple VPNs over MPLS network", In *Proc. Globecom'02*, Vol. 3, pp. 2195-2199, 2002.

[183] R. Sriram, G. Manimaran, C.S.R. Murthy, "Algorithms for delay-constrained low-cost multicast tree construction", *Computer Communications*, Vol. 21, No. 18, pp. 1693-1706, 1998.

[184] T. Stern, A.I. Elwalid, "Analysis of separable Markov modulated rate models for information-handling systems", *Advances in Applied Probability*, March, 1991.

[185] R. Szabo, P. Barta, F. Nemeth, J. Biro, "Call admission control in generalized processor sharing (gps) schedulers using non-rate proportinal weighting of sessions", In *Proc. of IEEE Infocom'99*, pp. 1220-1230, Mar. 1999.

[186] Taqqu M., Willinger W., Sherman R., Proof of a fundamental result in self-similar traffic modeling, *ACM/Sigcomm Computer Communications Review*, Vol. 27, pp. 5-23, 1997.

[187] W. Tian, "An exact and efficient analytical model for the ON/OFF burst traffic", In *Proc. 40th Annual Conference on Information Sciences and Systems*, 2006.

[188] H. Tijms,*A first course in stochastic models*, Wiley, 2003.

[189] P. Trimintzios et. al.,"A management and control architecture for providing IP differentiated services in MPLS-based networks", *IEEE Communications Magazine*, Vol. 39, No. 5, 2001.

[190] K.-C. Tsai, C. Chen, "Two algorithms for multi-constrained optimal multicast routing", *International Journal of Communication Systems*, Vol. 16, No. 10, pp. 951-973, 2003.

[191] K.-C. Tsai, C. Chen, "Algorithms for delay constrained ' group multicast routing", *International Journal of Communications Systems*, 18, pp. 905-933, 2005.

[192] M.J. Tunnicliffe, D.J. Parish, "A hybrid simulator for an ATM network", *International Journal of Communication Systems*, 13, pp. 179-184, 2000.

[193] Z. Wang, J. Crowcroft, "Quality-of-service routing for supporting multimedia applications", *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 7, Sep. 1996.

[194] M. Villn-Altamirano, J. Villn-Altamirano, "RESTART: a straightforward method for fast simulation of rare events", In *Proc. 1994 Winter Simulation Conference*, Lake Buena Vista, FL, USA, 1994.

[195] J.-J. Wu, R-H Hwang, H-I Lu, "Multicast routing with multiple 'QoS constraints in ATM networks", *Information Sciences*, 124, pp. 29-57, 2000.

[196] G-L Wu, J. W. Mark, "A buffer allocation scheme for ATM networks: complete sharing based on virtual partitioning", *IEEE/ACM Transactions on Networking*, Vol. 3, No. 6, 1995.

[197] G.-L. Wu, J.W. Mark, "Computational methods for performance evaluation of a statistical multiplexer supporting bursty traffic", *IEEE/ACM Transactions on Networking*, Vol. 4, No. 3, 1996.

[198] C. Wille, M. Mellia, E. Leonardi, M. Marsan, "Algorithms for IP design with end-to-end QoS constraints", *Computer Networks*, Vol. 50, No. 8, pp. 1086-1103, 2006.

[199] L. Xi, X. Shen, J. Mark, "Fair resource allocation with statistical QoS support for mulltimedia traffic in wideband CDMA cellular network", In *Proc. of IEEW WCNC'04*, Atlanta, Georgia, USA, March, 2004.

[200] L. Xi, X. Shen, J. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks", *IEEE Transactions on Wireless Communications*, Vol. 3, No. 1, 2004.

[201] X. Xiao, L. Ni, "Internet QoS: a big picture", *IEEE Network*, Vol. 13, No. 2, pp. 8-18, Mar./Apr. 1999.

[202] L. Xu, X. Shen, J.W. Mark, "Fair resource allocation with statistical QoS support for multimedia traffic in a wideband CDMA network", *Wireless Communications and Networking Conference*, 2004.

[203] W. Xu, J. Rexford, "MIRO: multi-path interdomain routing", In *Proc. Sigcomm'06*, Pisa, Italy, Sep. 2006.

[204] G. Xue, "Minimum-cost QoS multicast and unicast routing in communication networks", *IEEE Transactions on Communications*, Vol. 51, No. 5, 2003.

[205] G. Xue, A. Sen, W. Zhang, J. Tang, K. Thulasiraman, ”Finding a path subject to many additive QoS contraints”, *IEEE/ACM Transactions on Networking*, Vol. 15, No. 1, Feb. 2007.

[206] G. Xue, W. Zhang, J. Tang, K. Thulasiraman, “Polynomial time approximation algorithms for multi-constrained QoS routing”, *IEEE/ACM Transactions on Networking*, Vol. 16, No. 3, 2008.

[207] H. Yamada, S. Sumita, ”A traffic measurement method and its application for cell loss probability estimation in ATM networks”, *IEEE Selected Areas in Communications*, Vol. 9, No. 3, 1991.

[208] Y. Ying, R. Mazumdar,C. Rosenberg, F. Guillemin, “The burstiness of regulated flows in Internet”, In *Proc. of Networking* , Waterloo, Canada, 2005.

[209] O. Younis, S. Fahmy, ”Constrained-based routing in the Internet: basic principles and recent research”,*IEEE Communications Surveys*, Vol. 5, No. 1, 2003.

[210] X. Yuan, W. Zheng, S. Ding, “A comparative study of QoS routing schemes that tolerate imprecise state information”, In *Proc. Computer Communications and Networks*, pp. 230-235, 2002.

[211] X. Yuan, ”Heuristic algorithms for multiconstrained quality-of-service routing”, *IEEE/ACM Transactions on Networking*, Vol. 16, No. 2, Apr. 2002.

[212] E. Zegura, K. Calvert, S. Bhattacharjee, “How to model an internetwork”, In *Proc. of Infocom'96*, Vol. 2, pp. 594-602, Mar. 1996.

[213] H. Zhang, “Service disciplines for guaranteed performance service in packet-switching networks”, *Proceedings of the IEEE*, Vol. 83, No. 10, Oct. 1995.

[214] W. Zhengying, S. Bingxin, Z. Ling, “A delay-constrained least-cost multicast routing heuristic for dynamic multicast groups”, *Electronic Commerce Research*, Vol. 2, pp. 323-335, 2002.