Managing QoS in Multiservice Data Networks

A Rassaki

Abstract—**Next-generation networks require organized methods to offer Quality of Service (QoS) guaranteed IP network connectivity. This study suggests a solution for combined control of routing and flow problems, namely an algorithm based on flow deterministic network models. The algorithm solves the problem by identifying optimal routes and triggering the flow control law only for those paths. This experiment aims to assess how QoS and MPLS traffic engineering (TE) can advance Internet performance. It also aims to ascertain avenues for Internet improvement and to devise innovative mechanisms to ensure traffic engineering provision, and Class-of-Service (CoS) features in next-generation networks. The performances of the algorithm were evaluated on a fully connected six-node network, the data for which were extracted from a realistic network.**

***Index Terms*—MPLS, IP, QoS, multiservice networks, call admission control, call blocking probability, packet delay.**

I. INTRODUCTION

QUALITY OF SERVICE (QoS) is an important subject in contemporary multiservice broadband telecommunication networks, regardless of whether they have a basis in Internet Protocol (IP) or Asynchronous Transfer Mode (ATM) architecture. Actually, QoS should be addressed separately of network architecture. The aspect of "multiservices" raises a number of fundamental questions. The QoS concept must be studied on cell level, burst level and call level (see [1], [2]). Previous research studies principally concentrated on cell and burst levels. Published research on call level and mult-rate traffic are rare in the literature.

Network control in terms of flows is important to enable QoS to end-users. In QoS, the source must identify itself and the nature of its traffic, thereafter an "admission control" algorithm is run by the network resulting in the request being accepted or denied. This method is deemed inefficient and impractical for various reasons [3]. In terms of the nominal traffic level, sources may encounter a *rejection probability*. This raises to the issue of "large demands" which require a vast quantity of network resources are more likely to be rejected than small demands. This may seem an unfair situation.

Network managers may furthermore require close monitoring of rejection probabilities. "Equalization" is the simplest solution providing that all demands have the same rejection probability distinctly of the amount of resources needed.

The primary objective of call admission control in a broadband system is to concurrently sustain the QoS for various traffic pathways with diverse features.

A Rassaki is with the Department of Applied Information Systems University of Johannesburg, PO Box 524, Auckland Park, 2006 South Africa Email: arassaki@uj.ac.za

A large body of prevailing knowledge exists in this field (see [1], [4], [5], [6] and references therein). With the exception of [4] and [7], for example, existing research has focused on cell level QoS, such as cell loss ratio, cell delay and cell delay variation. The majority of call admission control (CAC) protocols considers the notion of effective bandwidth.

This concept has been researched at length (see [8], [9], [10], [11]). The formulation of effective bandwidth in published research is essentially based on cell level QoS. The rudimentary call admission principle is *service integration (SI).* This means that a new call will be allowed if the formerly established cell level QoS of this and any calls already currently underway is not disrupted. To benefit from the statistical multiplexing, however, it is required to carry traffic with diverse statistical features on the same network. In this case, SI can set off a substantial variation in call blocking probabilities between sources with different bandwidth needs. Therefore, to be able to meet the call level QoS condition for higher bandwidth calls under SI, facility size will lead to low bandwidth calls with call blocking probabilities that are far lower than necessary. This paper examines how to consider call level QoS when investigating call admission control. The routing problem, which minimizes the average delay, is introduced first. Secondly, the proposed algorithm is extended into multiservice data networks.

QoS network metrics diverge at different OSI layers. These can include call blocking probability, signal to interference ratio or bit error probability at the higher and physical layers respectively. QoS metrics can also be at packet level (explicitly delay and jitter, queue throughput, packet dropping probability) or at call level (call dropping probability). To guarantee QoS at different layer, call or packet level, cross-layer optimization is required. Cross-layer design has been introduced by network designers to bridge OSI structured design because it optimizes metrics at different levels and thereby enhances performance [12][13][14]. A cross-layer CAC with extended QoS metrics is therefore essential to maintain QoS in next-generation networks.

ATM networks have generally used delay parameter based call admission control schemes [15] which is established on the maximum delay bounds. Circuit-switched networks with fixed capacity also sometimes utilize measurements based call admission control [16].

Many approaches have been proposed for the problem of finding the routing which minimizes network delay. All are based on the intuitively appealing idea that we start with some feasible solution (i.e., one where the link flows are smaller than the link capacities), and then move requirements away from the more heavily travelled routes to less travelled routes. The way in which the flow is moved from one path to another distinguishes the algorithms from one another.

Previous research studies on network QoS have focused on various scheduling, queueing and buffer management protocols to assign a fixed capacity and delay (based on representative utilization) between flows at a statistically multiplexed resource [17]. Ciucu et al [18] suggested a strategy of provisioning based on characterization of the statistical service curve. They demonstrated how scheduling adds insignificant value to such provisioning.

Walingo and Takawira [19] devised a model for CAC next-generation networks with delay and signal to interference ratio as parameters. This CAC scheme utilizes delay and signal-to-interference ratio as user-specified QoS parameters to accept or reject calls. This also ensures a certain call blocking probability QoS metric.

The authors in [20] deliberated upon the CAC design for a single Markovian model of a multiservice statistical multiplexer. The main feature considered by the CAC is to address both the cell and call QoS concerns, which differs from the conventional focus on the cell level only.

This study focuses on this specific network engineering question of how to carry diverse traffic classes in MPLS networks through the construction of virtual paths (tunnels) in such a manner that the amount of tunnels on each MPLS router/link is minimized and load balanced.

In this paper the attention is limited to service integrated packet flows – the set of packets is not partitioned into forwarding equivalence classes (FECs) based on service class and each FEC contains packets belonging to any service class. The Multiprotocol Label Switching (MPLS) mechanisms introduced by the Internet Engineering Task Force (IETF) are used to minimize the expected packet delay in MPLS network with explicit routing.

The rest of the paper has the following structure: a definition of quality of service concept and a little introduction to the different architectures of providing QoS are presented in Section II. Section II discusses a mathematical model followed by an expression for the network expected delay experienced in multiservice networks. In Section IV, the experimental results for finding optimal paths in multiservice platform are displayed and the conclusion of the work is given in Section V.

II. BACKGROUND

There have been several service paradigms and methods for meeting QoS requirements proposed by the IETF. IETF concentrates on intrinsic QoS rather than perceived QoS. Intrinsic QoS refers to the aspects of service that relate to the technical parameters. Quality is realized via a number of factors: the optimal choice of transport protocols, the QoS assurance methods, and the associated utility of the technical constraints. Perceived QoS is defined by the clients' experience, and therefore perception, of utilizing a specific service. The perception of service is a combination of the expectation of service as compared to the actual level of service.

The IETF focus on intrinsic QoS is derived from the principal aims of IETF, which is Internet architecture, its development, reliability, and efficacy. As such, IETF defines QoS as “A set of service requirements to be met by the network while transporting a flow." [21] This definition is quite similar to the concept of network performance as classified by ITU/ETSI and is identified in terms of parameters.

Extensive work by the IETF has focused enormously on QoS assurance in IP networks. IETF has created various QoS mechanisms for the Internet. It recommended two important network architectures: Integrated Services (IntServ) [22] and Differentiated Services (DiffServ) [23]. In addition, it normalized the Resource reSerVation Protocol (RSVP) signalling protocol, which had initially been developed for IntServ model implementation and was subsequently expanded to other uses. It also created the idea of IP-QoS architecture as an inclusive model for QoS and made several recommendations for solutions.

Formally, QoS represents the architecture of dealing with several service level agreements (SLA[1]), which can be guaranteed. Service level agreements are comprised of issues such as QoS parameters or class of service offered, service dependability and accessibility, authentication concerns, and agreement end-dates. Service providers monitor, measure and assess service quality to determine whether the service is in compliance with the SLA.

Practically speaking, QoS uses a variety of attributes (e.g., classification, policing, queueing, shaping, scheduling) in the framework of the prevailing architecture (e.g., Integrated Service, Differentiated Services) to guarantee delivery of the SLA features by the network for the effective use of applications.

A different system often used in the work towards service quality is multiprotocol label switching (MPLS) [26]. IntServ and DiffServ network models are not reliant on Open Systems Interconnection (OSI) layer 2 techniques. These models generally define QoS architecture for IP networks, which can assimilate different transmission methods in one IP network. Like ATM and Frame Relay, MPLS is a networking method defined in layers 2 and 3. MPLS was originally designed to streamline packet forwarding in routers, and is not intended for service quality management. At this time, its primary function is traffic engineering and virtual private network support. Certain aspects of MPLS, though, assist in QoS guarantees. It may broaden the IntServ and DiffServ capacities to a more extensive range of platforms outside of the IP environment. It offers IP QoS services via Frame Relay and ATM networks. Additional MPLS attributes such as load balancing capacity, flow control, explicit routing, and tunnelling, are significant from the QoS perspective. The following sections summarize some of the IETF QoS architectures.

A. QoS Architectures

In this section we introduce the different architectures, service models and mechanisms of providing QoS. Among them are the *Integrated Services [22], Resource ReSerVation Protocol [24] Common Open Policy Services* model [25], the *Differentiated Services* [23] and the *MPLS* [26].

[1] An agreed upon arrangement between a customer and the service provider regarding service characteristics levels and related metrics.

Resource reservation protocol is the most common characteristic of the Integrated Services model. In real-time scenario, applications first establish routes and establish required resources before data are sent. This can be done with RSVP, which is a signaling protocol. The reservation setup protocol must address alterations in network topology, therefore when a link goes down, the reservation protocol should set up a new reservation and tear down the old reservation. Resource reservation often handles financial transactions; including issues of authorization, authentication, and billing. A reservation may have to be authenticated by the person paying for the reservation before it can start. The user requesting the reservation must be authenticated, and the reservation is documented for accounting purposes.

Common Open Policy Services (COPS) is a complementary practice for RSVP which specifies and enforces policies. A router responsible for policing utilizes COPS to contact policy server in order to understand the flow parameters. The RSVP is seldom used as it was devised to offer fine-grain, per-flow QoS. In Integrated Services, the RSVP protocol was established as the reservation setup protocol for the Internet.

The Differentiated Services mechanism was created to satisfy the need for fairly straightforward, coarse means of offering different levels of service for Internet traffic, to bolster assorted applications and for defined business needs.

In Differentiated Services, packets are labelled in different ways to establish numerous packet classes, which then receive the appropriate various services. The DiffServ model splits traffic into fewer classes and apportions resources per-class while the Integrated Services architecture allocates resource to separate flows. When looking at the operationalization and organization, the DiffServ model provides a simpler solution. The core of a DiffServ network differentiates a small number of forwarding classes instead of separate, discrete flows. No resource reservation setup is required.

Multi-Protocol Label Switching is a prominent technology that can increase routing efficiency [27][28], and it can also feature significantly in uniting ATM and IP architectures. MPLS, as a networking technology, can provide TE capability and QoS operation for communication networks.

Comer [29] describes MPLS as a connection-oriented communication mechanism built on top of IP. The connection-oriented routing model used by MPLS is originally from the ATM *virtual connection* prototype which sees traffic directed over bandwidth tunnels known as label switched paths (LSPs). To utilize MPLS, a network manager creates forwarding pathways via a group of MPLS-capable routers. At one of the pathway boundaries, each datagram is condensed in an MPLS header, then inserted into the MPLS pathway. At the opposite boundary, each datagram is removed, the MPLS header is detached, and the datagram is transmitted to its target destination. It can be useful to allocate traffic scheduling policy to an MPLS path, so that QoS parameters are established for datagrams that are added to a specific pathway. The ISP may, therefore, create an MPLS pathway for voice data, separate from the MPLS path used for other types of data.

TE via MPLS allows the traffic to be mapped efficiently to current network technologies. The potential of MPLS is the simplification of network design and management by way of integrating connections and predictability into IP networks.

III. MULTISERVICE NETWORK DIMENSIONING

In the previous papers [30][31][32], the focus has been on routing protocols supported by *single-service* networks communication, for example, networks where one call engages one circuit per link along the length of its routes. In this section, we turn our attention to *multiservice networks.*

Multiservice networks carry calls which belong to several call classes with different bandwidth requirements - a telephone call for example requires one unit of transmission capacity whereas a video call may require hundreds of units of capacity. With a telephone network, when the mandatory resources (end-to-end circuit) cannot be allotted to the call, it is blocked (that is, prohibited from joining the network) and the user receives a busy signal. No benefit can be realized by permitting a flow into a network where the resources are not available to allow adequate QoS to be serviceable. Costs are involved in permitting a flow that is not given adequate QoS, because network resources are utilized to accommodate the flow that ultimately gives no value to the user.

Networks can ensure that allowed flows will be able to receive required QoS by unambiguously allowing or denying flows depending on the resources required and the source requirements of acknowledged flows. The flow must implicitly declare its QoS requirements in order for it to be provided with guaranteed QoS. *Call admission* is the practice of a flow declaring its conditions for QoS, then allowing the network to accept the flow (offering the requisite QoS) or deny the flow. If adequate resources are not on supply, and QoS must be assured, a call admission process is required.

In the next sections, the FOA algorithm [30] is extended to investigate the performance of optimal routing in multiservice networks carrying several classes of traffic each with differing bandwidths and differing level of service requirements.

A. Analytic Techniques

A physical network consisting of a set of N nodes is represented by $\mathcal{N}$ and a collection of L physical links represented by $\mathcal{L}$. The nodes symbolize the routers in the MPLS-capable network. The traffic requirements are indicated by an $N \times N$ matrix $Re = r_{ij}$, called the requirement matrix, whose entries are non-negative. Let $C_{i,j}$ represent the capacity in bandwidth units of the physical link from an origin (ingress LSR) node i to a destination (egress LSR) node j. Every route consists of a non-cycling series of physical links. The goal of the design problem is to find optimal flows that would optimize the objective function.

$$T = \sum_{(i,j)} \frac{F_{ij}}{\gamma} T_{ij}$$

where $\gamma = \sum_{(ij)} \lambda_{ij}$ is the total message arrival rate from external sources (bits/sec), $F_{ij}$is the flow on the link (i,j) in message/sec and T_{ij} is the average delay experienced by a message on link (i,j) (sec) subject to:

$$0 \le F_{ij} \le C_{ij} \qquad \forall i,j \in \mathcal{N}$$

The original Flow Deviation Algorithm used an objective function based on the $M/M/1$ queue. This queue assumes that the packets arrive according to a Poisson process and that the packet lengths are exponentially distributed. In the single service network the total delay on the link, (i,j) with service time T_s and utilization U_{ij} is

$$T_{ij} = \frac{T_s}{1 - U_{ij}},$$

where T_s is the average message length of size M, divided by the capacity of the link C_{ij}, and $U_{ij}$is the flow in the link, F_{ij}divided by C_{ij}. Thus,

$$T_{ij} = \frac{M/C_{ij}}{1 - F_{ij}/C_{ij}} = \frac{M}{C_{ij} - F_{ij}}$$

The weighted network delay is therefore

$$T = \sum_{(i,j)} \frac{MF_{ij}}{C_{ij} - F_{ij}}$$

where M is the average message length. In a multiservice network, the inputs of the models correspond to those of single class models. The additional consideration is the specification of the link service discipline which is the rule for selecting the next customer to receive service. Each link in our multiservice problem will be modelled as a processor sharing queue in which the total service capacity is equally shared between the available customers.

For the Processor Sharing, the total average system response time for a class-k, where $k \in K$, is:

$$T = \sum_{(i,j)} \sum_{k} \frac{M_k F_{ijk}}{C_{ij} - \sum_k M_k F_{ijk}}$$

where M_k is the length of a class-k message in the system, and F_{ijk} is the class k flow on link (i,j). As the model is extended to the case where several paths connect each user to the system, let K represents such paths.

The network model includes the below steps to determine the network delay:

Step 1: Initialization Allocate the link lengths using as a basis the first derivative of delay in connection with flow starting with zero flows.

$$d_p = \sum_{(i,j)\in p} D'_{ij}(F_{ij})$$

Step 2: Compute the least-cost paths using Bellman's algorithm for every O-D pair.

Step 3: Apply the shortest path for every pair of requirements.

Steps (4) through (9) below execute the iterations:

Step 4: Modify the link capacities where required to ensure that the path flows are feasible.

Step 5: Allocate new link lengths using as basis the first derivative of delay in connection with current flow. The new flow is an enhancement on the previous flow when applied to the same link capacities.

Step 6: Determine the shortest paths for every O-D pair.

Step 7: Include a new path to the path set and compute how much flow must be moved to it. The amount of flow δ to move off of path p is computed as

$$\delta = \alpha(d_p - d_{\bar{p}_w})/H_p$$

Step 8: For every O-D pair, move the flow from all other paths to the least cost paths.

Step 9: Compute the new QoS network average delay.

Step 10: Stopping rule. Should the network delay decrease ends, then discontinue. If not, return to step 5.

The repetition halts when the calculated current delay is no longer considerably less than the previous delay. In order to avoid endless repetition, the algorithm also stops as soon as the new factor of capacity adjustment is not considerably less than the preceding factor.

B. Service integration

In this most recurrently used technique, the transmission link can be assigned to any call type or class. For a call to be accepted, the condition is as follows:

An arriving call of class i will be accepted if and only if the available link capacity Cr is greater than or equal to the bit rate requirement Di.

The multiservice traffic is formed as follows: The class k requirement λ_{ij}^k between two given nodes (i,j) is equal to the base traffic intensity λ_{ij} multiplied by a class-dependent traffic intensity factor γ_k multiplied by the bandwidth requirement $b_k$for this service.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the algorithm, we consider the network topology shown in Figure 1 presented in [33]. The network consists of 8 nodes and each link carries traffic in one direction. The transmission capacity of each uni-directional link is 2812 bandwidth units. The double lines indicate two-unidirectional links each having a transmission capacity of 5624 bandwidth units.

The objective is to find minimum delay routes using the algorithm in a multiservice network.

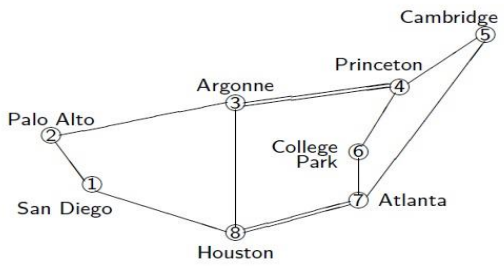

Fig. 1. The NSF Network Infrastructure

The network carries six traffic classes: the bandwidth requirement of the first service is 1 unit and the bandwidth of services 2 through 6 are 3, 4, 6, 24, and 40 respectively. Table 1 presents the different message lengths per service while the base traffic intensity matrix is shown in Table 2. The class dependent traffic intensity is $p_{ij}^s = p_{ij}\gamma_s b_s$. The optimal

flows per service class are shown in Table 3. Figures 2 and 3 represent these optimal link flows.

TABLE 1
CLASS-DEPENDENT FACTOR AND SLOTS PER SERVICE

	class 1	class 2	class 3	class 4	class 5	class 6
γ_s	0.4	0.4	1.0	0.5	0.5	0.1
b_s	1	3	4	6	24	40
M_s	1	3	2	3	1	1

TABLE 2
TRAFFIC INTENSITY MATRIX

nodes	1	2	3	4	5	6	7	8
1	-	6	7	1	9	5	2	3
2	7	-	24	3	31	15	6	9
3	8	25	-	4	37	18	7	11
4	1	3	3	-	4	7	1	1
5	11	33	39	5	-	24	9	15
6	5	14	16	2	21	-	4	6
7	2	5	6	1	8	4	-	2
8	3	8	10	1	12	6	2	-

Experiments show that links with larger bandwidths have larger flows, and therefore smaller service times. For example, the 3-4 and 7-8 links have 5624 bandwidth units while the other links 2812 bandwidth units.

When one considers the utilization of the links 2-3, 4-5 and 5-7, one can see that it is much higher compared with the utilization of the other links. However their flows are not large compared with the other links (see Figures 4 and 5). Although links 3-4 and 7-8 have large flows, the utilizations are moderate. The reason is that they have large capacities, 5624 bandwidth units as opposed to 2812.

The fact is that the total utilisations of the different links do not differ much, indicate that the flow deviation algorithm tends to spread the flow equally in the network.

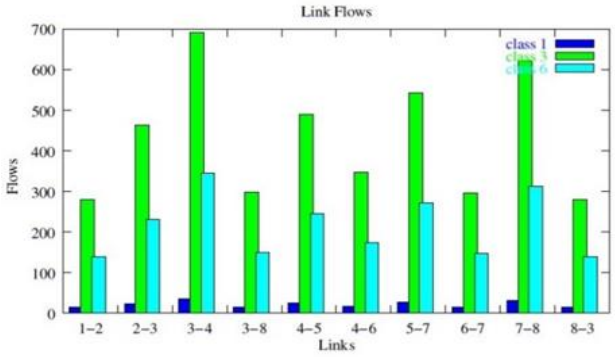

Fig 2. Optimal link flows per class

TABLE 3
NFS NETWORK: OPTIMAL FLOWS PER SERVICE CLASS

Links	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1-2	13.970	125.735	279.412	314.338	419.117	139.706
1-8	22.889	206.002	457.782	515.005	686.673	228.891
2-1	14.889	134.002	297.782	335.005	446.673	148.891
2-3	23.110	207.998	462.218	519.995	693.327	231.109
3-2	23.629	212.665	472.588	531.662	708.883	236.294
3-4	34.591	311.322	691.826	778.305	1037.74	345.913
3-8	14.919	134.276	298.392	335.691	447.587	149.196
4-3	34.062	306.540	681.199	766.349	1021.80	340.600
4-5	24.497	220.473	489.941	551.183	734.911	244.970
4-6	17.320	155.888	346.418	389.721	519.628	173.209
5-4	27.268	245.429	545.377	613.550	818.066	272.689
5-7	27.131	244.180	542.623	610.450	813.934	271.311
6-4	14.817	133.360	296.355	333.399	444.532	148.177
6-7	14.782	133.040	295.645	332.601	443.468	147.823
7-5	24.303	218.727	486.059	546.817	729.089	243.030
7-6	14.679	132.112	293.582	330.279	440.372	146.791
7-8	31.140	280.260	622.801	700.651	934.201	311.400
8-1	23.570	212.135	471.412	530.338	707.117	235.706
8-3	13.969	125.725	279.389	314.313	419.084	139.695
8-7	0	0	0	0	0	0

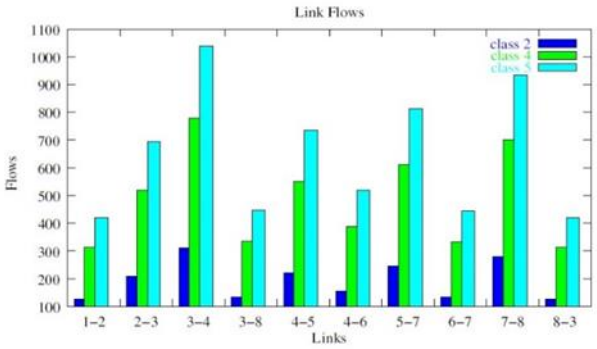

Fig 3. Optimal link flows per class

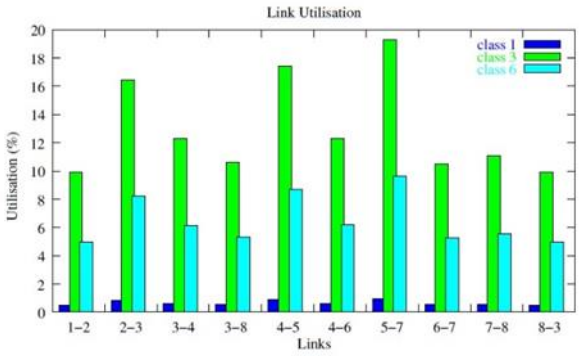

Fig 4. Link utilization per class

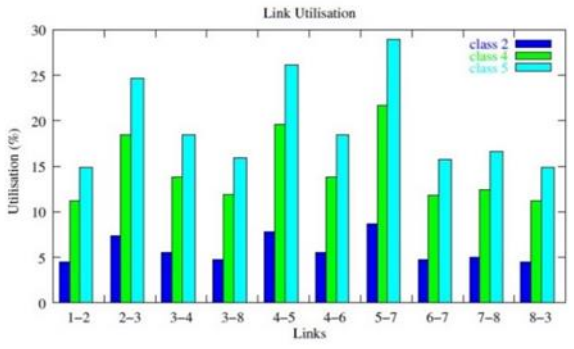

Fig 5. Link utilization per class

V. CONCLUSION

MPLS attains control over packet flows (and therefore performs traffic engineering in a flexible manner) by means of label switching. Different tunnels can be established for different types of classes that use the concept of label switching between label-edge routers (LER) where ingresses and egresses traffic.

This paper analyzed call level in the presence of multiservice sources. Service integration is the most extensively researched call admission policy type [34]. While it is not technically complex to put it into place, it routinely gives preference to calls with smaller bandwidth capacity needs. It can, therefore, increase the potential of blocking calls requiring a larger bandwidth when the arrival rates of class 1 blocking probability surpasses the QoS conditions, while the potential for blocking class 2 calls is much lower than its constraint. Using service integration, irrespective of which class of calls exceeds its engineered load, class 1 will experience a higher instance of call blocking probability.

REFERENCES

[1] J.W. Roberts (Ed.) COST 224: *Performance evaluation and design of multiservice networks.* ECSC-EEC-EAEC, Brussels, 1992.

[2] J. Roberts, U Mocci, J. Virtamo COST 242: *Broadband Network Teletraffic Performance evaluation and design of multiservice networks.* Springer-Verlag. July 1996.

[3] D Pompili, C Scoglio and CA Shoniregun. "Virtual-Flow Multipath Algorithms for MPLS" Internet Technology and Secured Transactions, Vol.1, No. ½, 2007.

[4] J.M. Hyman, A.A Lazar, and G. Pacifici. "A Separation Principle between Scheduling and Admission Control for Broadband Switching" *IEEE Journal on Selected Areas in Communications* 11:605-616, 1993.

[5] *Special Issue* on Advances in the Fundamentals of Networking – Part 1, *IEEE Journal of Selected Areas in Communications*, 13, 2015.

[6] H. Saito. "Call Admission Control in an ATM Network using upper bound of cell loss probability" *IEEE Transaction on Communications*, 40:1512-1521, 1992.

[7] M. Beshai, R. Kositpaiboon, and J. Yan. "Interaction of call blocking and cell loss in an ATM Network" *IEEE Journal on Selected Areas in Communications* 12:1051-1058, 2014.

[8] N.G. Bean. "Effective Bandwidths with different Quality of Service requirements" In *Integrated Broadband Communication Networks and Services.* V.B. Iverson (Ed.), IFIP, 2013.

[9] A.I. Elwalid and D. Mitra. "Effective bandwidth of general Markovian Traffic sources and Admission Control of High speed Networks" *IEEE/ACM Transactions on Networking*, 1:329-343, 1993.

[10] J.Y. Hui. "Resource Allocation for Broadband Networks" *IEEE Journal of Selected Areas in Communications*, 6, 2008.

[11] F.P. Kelly. Effective bandwidths at multiclass queues. *Queueing Systems*, 9:5-16, 1991.

[12] F. Foukalas, V. Gazis and N. Alonistioti: "Cross-Layer Design Proposals for Wireless Mobile Networks: A survey and Taxonomy" *IEEE Communication Surveys & Tutorials*, Vol. 43, No 1, pp. 70-85, 2014.

[13] V. Srivastava and M. Motani: "Cross-Layer Design: a survey and the road ahead" *IEEE Communications Magazine*, Vol. 43, No. 12, pp. 112-119, 2015.

[14] Fei: "Cross-Layer Optimal Connection Admission Control for Variable Bit Rate Multimedia Traffic in Packet Wireless CDMA Networks" *IEEE Transactions on Signal Processing*, Vol. 54, No. 2, February 2016.

[15] J.M. Hah and M.C. Yuang: "Estimation-based Call Admission Control with Delay and loss guarantees in ATM Networks" *IEEE Proceedings on Communications*, Vol. 144, No.2, April 1997.

[16] S. Jamin *et al*: "A Measurement-based Admission Control Algorithm for Integrated Services Packet Networks" *IEEE/ACM Transactions on Networking*, Vol. 5, No.1, February 1997.

[17] G. Cortese, R. Fiutem, P. Cremonese, S. D'Antonio, M. Esposito, S.P. Romano, and A. Diaconescu, "CADENUS: Creation and Deployment of end-user services in premium IP-networks" *IEEE Communications Magazine*, Vol. 41, No. 1 pp. 54-60, Jan 2013.

[18] F. Ciucu, A. Burcharti, and J. Liebeherr, "A Network Service Curve Approach for the Stochastic Analysis of Networks" In *Proceedings* of *ACM SIGMETRICS*, New York, NY, USA, pp. 279-290, 2005.

[19] T Walingo and F. Takawira. "Cross-Layer Extended Parameter Call Admission Control for Future Networks" *SAIEE* Vol. 104, No. 1, March 2013.

[20] D. Mitra, M.I. Reiman and J. Wang. "Robust Dynamic Admission Control for United Cell and Call QoS in Statistical Multiplexers" *IEEE/ACM Transactions on Networking*, 1:672-688, 2008.

[21] E. Crawley et al., "A framework for QoS-based routing in the Internet" *RFC 2386*, Aug. 1998.

[22] R. Braden, D. Clark, and S. Shenker "Integrated service in the Internet Architecture: an overview" *RFC 1633* Jun 1994.

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

[24] L. Zhang, R. Braden, S. Berson, S. Herzog, and S. Jamin. "Resource Reservation Protocol (RSVP) – Version 1 Functional specification" RFC 2205 Sep 1997.

[25] J. Boyle, R. Cohen, D. Durham, S. Herzog, R. Rajan and A. Sastry. "The COPS (Common Open Policy Service) Protocol" RFC 2748, Jan 2010.

[26] E. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture", IETF RFC 3031, January 2001.

[27] D.O Awduche, and B. Jabbari "Internet traffic engineering using MultiProtocol Label Switching (MPLS)", IEEE Computer Networks Vol. 40, No. 1, pp. 111-129, September 2012.

[28] X. Xiao, A. Hannan, B. Bailey and L. Ni "Traffic engineering with MPLS in the Internet" IEEE Network Magazine, Vol. 14, No. 2, pp. 28-33 2000.

[29] D. E. Comer, *Computer Networks and Internet* Pearson International Edition, 2010.

[30] A. Rassaki, and A.L. Nel "Quality of Service in MPLS Networks" in *Proceedings of the Fifteenth IASTED on Control and Applications*, Honolulu, USA, pp.67-74, Aug. 2013.

[31] D. Katabi, M. Handley, and C. Rohrs, "Congestion Control for High Bandwidth-delay product Networks" *ACM/SIGCOMM Conference*, Aug. 2012.

[32] D. Pompili, C. Scoglio, and C.A. Shoniregun,"Virtual-flow Multipath algorithms for MPLS" *International Journal: Internet Technology and Secured Transactions*, Vol. 1 No1/2, Dec. 2007.

[33] A. Rassaki and A.L. Nel, "Optimal Capacity Assignment in IP Networks" *Fifth International Conference on Digital Information Processing & Communications*, Switzerland, pp 7-9, Oct. 2015.

[34] A. Rassaki, and A.L. Nel "Optimizing Capacity Assignment in Multiservice MPLS Networks" *South Africa Computer Journal*, 29(1) June 2017.