Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Computer Networks 41 (2003) 761–777

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

A new scheme for traffic estimation and resource allocation for bandwidth brokers

T. Anjali [a,*], C. Scoglio [a], G. Uhl [b]

[a] *Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

[b] *Swales Aerospace & NASA Goddard, Beltsville, MD 20705, USA*

Received 1 May 2002; received in revised form 16 October 2002; accepted 14 November 2002

Responsible Editor: C. Douligeris

Abstract

This paper is motivated by the concern of a multi-service network provider who plans to offer quality of service guarantees to users. A bandwidth broker acts as the resource manager for each network provider. Neighboring bandwidth brokers communicate with each other to establish inter-domain resource reservation agreements. Conventional approaches for resource allocation rely on pre-determined traffic characteristics. If allocation follows the traffic demand very tightly, the resource usage is efficient but leads to frequent modifications of the reservations. This would lead to increased inter-bandwidth-broker signaling in order to propagate the changes to all the concerned networks. Contrarily, if large cushions are allowed in the reservations, the modifications are far spaced in time but the resource usage becomes highly inefficient. In this paper, a new scheme for estimating the traffic on an inter-domain link and forecasting its capacity requirement, based on a measurement of the current usage, is proposed. The method allows an efficient resource utilization while keeping the number of reservation modifications to low values.

Keywords: Bandwidth broker; Resource allocation; Resource provisioning; Capacity reservation; Traffic estimation; Bandwidth usage

1. Introduction

There is a growing demand for guaranteed bandwidth and latency by users of the present day Internet. These guarantees, called as quality of service (QoS), are made up of three basic components from a router's point, namely

- defining packet treatment classes,
- specifying the amount of resources for each class,
- sorting incoming packets into respective classes.

Differentiated services concept [1] has been proposed as an architecture to provide these QoS guarantees in the Internet. It specifies traffic classes (per hop behaviors, PHBs) and a mechanism to classify incoming packets (based on the differentiated services code point in each packet). It thus removes maintenance of state in the core routers in

* Corresponding author.
E-mail addresses: tricha@ece.gatech.edu (T. Anjali), caterina@ece.gatech.edu (C. Scoglio), uhl@rattler.gsfc.nasa.gov (G. Uhl).

1389-1286/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S1389-1286(02)00438-3

a network, and outsources "expensive" tasks to leaf routers. This helps in attaining different service levels without increasing the overhead in the core routers of the network.

The bandwidth broker (BB) [2] is an agent responsible for allocating preferred service to users as requested, and for configuring the network routers with the correct forwarding behavior for the defined service for each class. It helps in dynamic resource management for the DiffServ classes. In a centralized broker environment, one BB is associated with a particular domain/autonomous system (AS). On the other hand, there can be multiple BBs in a domain to distribute the functionality. A BB has a policy database that keeps the information on who can do what, when and a method of using that database to authenticate requesters. Only a BB can configure the leaf routers to deliver a particular service to flows, crucial for deploying a secure system. To ensure robustness, a network can have a primary bandwidth broker supported by backups. These multiple agents can introduce the problems of consistency and increased signaling. The tasks of a bandwidth broker are split into two main categories: *intra-domain* and *inter-domain*. Intra-domain tasks include resource management and traffic control within the domain of the BB. Inter-domain tasks cover the specification of bilateral service level agreements (SLAs) with neighboring domains and managing the boundary routers to police/shape the incoming/outgoing traffic to adhere to the SLAs. The BB of a transit domain has to reserve resources between the ingress and egress points of the domain. End-to-end QoS can then be achieved by concatenation of the intra- and inter-domain reservations. Several protocols have been suggested for inter-domain resource management signaling, such as RSVP [3], SIBBS [4] etc. Intra-domain resource management and signaling protocols are left to the administrator of the domain. Examples of such would be COPS-PR [5], IntServ/RSVP and DiffServ. Currently, there are several implementations of bandwidth broker under way. An architecture of a bandwidth broker for scalable support of guaranteed services is given in [6,7]. It considers mainly the admission control issues for a bandwidth broker. Most of the implementations are concerned with the issues of inter-BB communications and BB–router communications. The release 1.0 of the bandwidth broker from BCIT [8] provides static provisioning based on the service level specification (SLS). The Globus Architecture for Reservation and Allocation (GARA) [9] and the QoS architecture [10] present a broker architecture that supports advance reservations for resources. The architecture presents the three steps of control: sensors, decision procedures and actuators. The architecture includes per-application modification of reservations.

Let us exemplify the operations of a bandwidth broker, as defined in RFC 2638 [2] where the concept of the bandwidth broker was first introduced. When an allocation is desired for a particular flow, a request is sent to the bandwidth broker of the concerned AS. The request specifies the service type, target rate, maximum burst, and the time period when service is required. Note that, various other DiffServ/BB designs do not impose this requirement. But, in this paper we will follow the original specifications for the bandwidth broker. In general, the request can be originating from an end-user or a neighboring region's BB. The BB has to authenticate the credentials of the requester. If the request is valid, the BB finds the route along which request will be forwarded. The BB then verifies the existence of sufficient unallocated bandwidth on the link with the next AS to satisfy the requested QoS. If the request passes these tests, the network resources are correspondingly provisioned. In the case of a transit AS, the BB has to verify sufficient resources within the network and on the downstream link. Under the DiffServ architecture, user flows are aggregated on the boundary nodes. Consequently, only aggregate information is maintained in the BB and resource allocations are made on an aggregate basis in the core. Provisioning on the edge routers (ERs) can be easily determined based on the SLS in place with the customer devices. To guarantee the end-to-end QoS requirements of a request, the BB makes bilateral agreements with its neighboring BBs, rather than multilateral agreements with all possible destination domains. An important requirement of the inter-domain agreements is that the changes involved should be less frequent and

should be on a time-scale larger than the individual flow variations. If not satisfied, the scalability of the provisioning scheme is compensated. A proposed scheme for resource provisioning is to have a bandwidth "cushion", wherein extra bandwidth is reserved over the current usage. As proposed in [11], if the traffic volume on a link exceeds a certain percentage of the agreement level, it leads to a multiplicative increase in the agreement. A similar strategy is proposed in case the traffic load falls below a considerable fraction of the reservation. This scheme satisfies the scalability requirement but leads to an inefficient resource usage. This drawback can become increasingly significant once the bandwidth requirements of the users become considerable.

Conventional approaches for resource allocation rely on pre-determined traffic characteristics. Network traffic can be divided into elastic (e.g. TCP) and non-elastic streaming (e.g. UDP) traffic [12]. These two types differ in their requirements from the network. Packet level characteristics of elastic traffic are controlled by the transport protocol and its interactions with the network, whereas the non-elastic flows have inherent rate characteristics that must be preserved in the network to avoid losses. The source characteristics may not be known ahead of time, specified parameters may not characterize the source adequately or a large number of parameters may be required to define traffic characteristics, thus making the pre-determined traffic characteristic-based resource allocation inefficient.

Current resource allocation methods can be either *off-line* or *on-line*. Off-line, or static, methods determine the allocation amount before the transmission begins. These approaches (e.g. [13]) are simple and predictable but lead to resource wastage. On-line, or dynamic, methods (e.g. [14–16]) periodically renegotiate resource allocation based on predicted traffic behavior. These methods undergo a large number of renegotiations to satisfy the QoS.

In this paper we propose an on-line scheme called estimation and prediction algorithm for bandwidth brokers (EPABB) to forecast the bandwidth utilization of inter-domain links. The scheme is designed to be simple, yet effective, when compared to more advanced prediction algorithms because it is intended to be used by bandwidth brokers and so the design goal is simplicity. The first step of the scheme is to perform an optimal estimate of the amount of traffic utilizing an inter-domain link based on a measurement of the instantaneous traffic load. This estimate is then used to forecast the traffic bandwidth requests so that resources can be provisioned between the two domains to satisfy the QoS of the requests. The estimation is performed by the use of Kalman filter [17] theory while the forecast procedure is based on deriving the transient probabilities of the possible system states. As we will show later, this scheme outperforms the current resource reservation mechanism ("cushion-based" allocation [11,18]) employed by bandwidth brokers and also some other prediction schemes based on Gaussian [19, 20] as well as local maximum [21] predictor. Our scheme reduces bandwidth wastage without introducing per-flow modifications in the resource reservation. Kalman filters have been previously applied to flow control in high-speed networks. In [22], Kalman filter was given for state estimation in a packet-pair flow control mechanism. In [23], Kalman filter was used to predict traffic in a collection of VC sources in one VP of an ATM network. Our work distinguishes itself from previous work as the Kalman filter is used as an optimal estimation algorithm, instead of filtering or smoothing and it is an input to the capacity forecast step.

The rest of the paper is organized as follows. In Section 2, we give the mathematical setup and formulation of the EPABB for estimating and forecasting the capacity allocation for the traffic. The experimental results for the performance evaluation of EPABB are described in Section 3 with examples and comparisons. Finally, Section 4 concludes the paper.

2. Traffic estimation and resource forecast

Consider two ASs, AS1 and AS2 as shown in Fig. 1. There exists a link $l(1,2)$ between the two domains. Each domain has a BB (BB1 and BB2, respectively, as shown) associated with it. We estimate the level of traffic between the two domains,

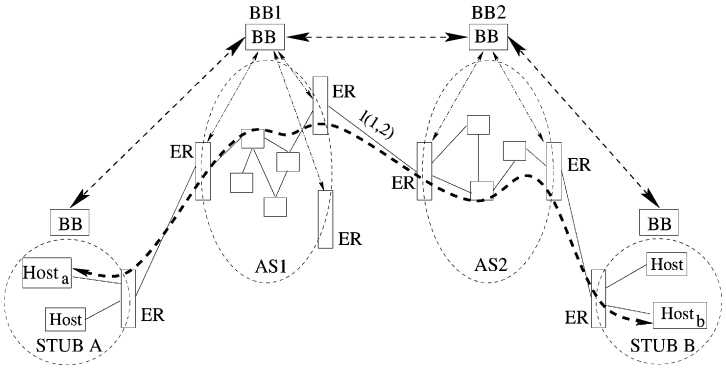

Fig. 1. Network topology.

for a given traffic class, based on a periodic measurement of the aggregate traffic on $l(1,2)$. We assume that the traffic measurements are performed at discrete time-points mT, $m = 1, 2, \ldots, M$ for a given value of T. The value of T is a measure of the granularity of the estimation process and denotes the renegotiation instants. Larger values imply less frequent estimation which can result in larger estimation errors. In EPABB, we have assumed periodic renegotiation instants like other approaches. At the time instant m (corresponding to mT), the aggregate traffic on $l(1,2)$ for a given traffic class in the direction AS1 to AS2 is denoted by $y(m)$. We also assume that for the duration $(0, MT]$, the number of established sessions that use $l(1,2)$ is N. For each session, flows are defined as the active periods. So, each session has a sequence of flows separated by periods of inactivity. For a given traffic class, we denote by $x(m)$ the number of flows at the instant m and by $x(mT + t),\ t \in (0, T]$ the number of flows in the time interval $(mT, (m+1)T]$, without notational conflict. Clearly, $x(m) \leqslant N$ and is not known/measurable. We assume that each flow within the traffic class has a constant rate of b bits per second. So, nominally, for a traffic class

$$y(m) = bx(m). \tag{1}$$

We concentrate, without loss of generality, on a single traffic class and its associated resource utilization forecasting. To consider a scenario with different classes of traffic with different bandwidth requirements, the same analysis can be extended and applied for each class. In this paper, we deal with the resource allocation for the DiffServ Expedited Forwarding (EF) classes, for which the assumption of constant resource requirement is valid. The underlying model for the flows is assumed to be Poisson with exponentially distributed inter-arrival times (parameter λ) and durations (parameter μ). Characteristics of IP traffic at packet level are notoriously complex (self-similar). However, this complexity derives from much simpler flow level characteristics. When the user population is large, and each user contributes a small portion of the overall traffic, independence naturally leads to a Poisson arrival process for flows [12,24]. The following analysis has been carried out using this assumption and then the experimental results show that the capacity forecast is very close to the actual traffic.

The EPABB scheme for resource allocation is split into two steps. In the first step, a rough measure of the aggregate traffic $y(m)$ is taken and it is used to evaluate the number of flows through the Kalman filter estimation process. In the second step, we reserve the resources $R(m)$ on the link $l(1,2)$ for the time $t \in (mT, (m+1)T]$ based on the forecast of the evolution of $x(m)$. First, we give a brief introduction to the discrete Kalman filter theory.

2.1. Discrete Kalman filter

The Kalman filter [17] is a set of mathematical equations that provides an efficient computational (recursive) solution of the least-squares method. It implements a predictor–corrector type estimator

that is optimal in the sense that it minimizes the estimated error covariance—when some presumed conditions are met. It estimates a process by using feedback control. It supports estimation of past, present and future states, even if the knowledge of the precise nature of the modeled system is lacking. The Kalman filter tries to estimate the state x of a discrete-time controlled process. Let the system be described by the state vector x that is observed at times $m = 0, 1, \ldots$ which is governed by the linear stochastic difference equation

$$x(m+1) = Ax(m) + Bu(m) + w(m). \tag{2}$$

Each of the observations is, however, corrupted by noise and thus, the actual measurement $\bar{y}$ at times $m = 0, 1, \ldots$ is given by

$$\bar{y}(m) = Cx(m) + v(m). \tag{3}$$

The random variables $w(m)$ and $v(m)$ represent the process and measurement noise, respectively. They are assumed to be independent zero-mean Gaussian white noise processes [25].

$$E[w(m)w(k)] = \begin{cases} \sigma_w^2, & k = m, \\ 0, & \text{otherwise}, \end{cases}$$

$$E[v(m)v(k)] = \begin{cases} \sigma_v^2, & k = m, \\ 0, & \text{otherwise}, \end{cases}$$

$$E[v(m)w(k)] = 0.$$

The parameter A in Eq. (2) relates the states at previous and current time-steps, in the absence of either a driving function or process noise. We assume that A stays constant over the analysis. The parameter B relates the optional control input u to state x whereas the parameter C in Eq. (3) relates the state to the measured value. The objective of the Kalman filter is to obtain a "best" estimate, in some suitable sense, of $x(m)$ based on the observed values $\bar{y}(m)$ and knowledge of the statistical properties of the process and measurement noise. According to the Kalman filter mechanism, the best estimate for $x(m)$ can be obtained recursively from the previous best estimate and its covariance matrix. Thus, the Kalman filter can be divided into two steps: *prediction* and *correction*. The first step is responsible for projecting forward in time the current state to obtain a priori estimates of the states and covariance in the next time step.

State Equation : $\boldsymbol{x(m) = A(m-1)x(m-1) + Bu(m-1) + w(m-1)}$
Observation Equation : $\boldsymbol{y(m) = C(m)x(m) + v(m)}$
Step 1: Initialization
$\hat{x}(0|0) = E[x(0)]$
$P(0|0) = E[x(0)x(0)^T] = \sigma^2_{x(0)}$
Step 2: Computation
for m = 1,2, …
prediction step:
$\hat{x}(m|m-1) = A(m-1)\hat{x}(m-1|m-1) + Bu(m)$
$P(m|m-1) = A(m-1)P(m-1|m-1)A^T(m-1) + \sigma_w^2$
correction step:
$k(m) = P(m|m-1)C^T(m)[C(m)P(m|m-1)C^T(m) + \sigma_v^2]^{-1}$
$\hat{x}(m|m) = \hat{x}(m|m-1) + k(m)[y(m) - C(m)\hat{x}(m|m-1)]$
$P(m|m) = \{I - k(m)C(m)\}P(m|m-1)$

Fig. 2. Discrete Kalman filter.

The second step is responsible for the feedback to obtain an improved a posteriori estimate. The Kalman filter is summarized in Fig. 2.

2.2. Traffic estimation

The only measurable variable in the system is $\bar{y}(m)$ which is a measure, corrupted by noise, of the aggregate traffic on the link. Nominally, $x(m) = y(m)/b$, but we do not have access to the correct measurements of $y(m)$, even though b is a known quantity for a particular traffic class. Thus, we propose to use the Kalman filter setup to evaluate $\hat{x}(m)$, an estimate of the actual $x(m)$, using $\bar{y}(m)$ the noisy measurements. Real measurement noise makes $\bar{y}(m)$ noisy. In other words, the measurements obtained from the network for the instantaneous traffic on the link can be noisy due to miscalculation, misalignment of timings, etc. To use the Kalman filter setup, we need to find a relation between $x(m)$ and $y(m)$, and $x(m)$ and $x(m+1)$. To this purpose, we define $p_k(t)$, $t \in (mT, (m+1)T]$ to be the probability that the number of active flows at time t is k i.e. for $t \in (mT, (m+1)T]$

$$p_k(t) \triangleq \text{prob}\{x(t) = k\}. \tag{4}$$

The state-transition-rate diagram is shown in Fig. 3. The diagram depicts transitions among the

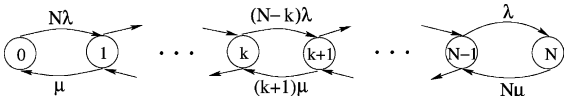

Fig. 3. State-transition-rate diagram.

states. From the diagram and by using queuing theory [26], we can write the following differential equations (5)–(7) for the probabilities $p_k(t)$.

$$\frac{\mathrm{d}p_0(t)}{\mathrm{d}t} = \mu p_1(t) - N\lambda p_0(t), \tag{5}$$

$$\frac{\mathrm{d}p_k(t)}{\mathrm{d}t} = (N-k+1)\lambda p_{k-1}(t) + (k+1)\mu p_{k+1}(t) - (k\mu + (N-k)\lambda)p_k(t), \quad 1 \leqslant k < N, \tag{6}$$

$$\frac{\mathrm{d}p_N(t)}{\mathrm{d}t} = \lambda p_{N-1}(t) - N\mu p_N(t). \tag{7}$$

The generating function $G(z,t)$ is defined as the z-transform of the probability distribution function. It aids in the computation of the mean and variance of the probability distribution. Next, we calculate $\partial G(z,t)/\partial t$ using Eqs. (5)–(7) as

$$G(z,t) \triangleq \sum_{j=0}^{N} p_j(t) z^j,$$

$$\frac{\partial G(z,t)}{\partial t} = G(z,t)N\lambda(z-1) - \frac{\partial G(z,t)}{\partial z}(z-1)(\lambda z + \mu).$$

Utilizing the initial condition $G(z,mT) = z^{x(m)}$, i.e., the number of active flows at time mT is $x(m)$, we arrive at the following solution for $G(z,t)$: for $t \in (mT,(m+1)T]$,

$$G(z,t) = C(z,t)^{x(m)}\left(\frac{\lambda z + \mu}{\lambda C(z,t) + \mu}\right)^N,$$

where

$$C(z,t) = \frac{\lambda z + \mu - \mu(z-1)\mathrm{e}^{-t(\lambda+\mu)}}{\lambda z + \mu - \lambda(z-1)\mathrm{e}^{-t(\lambda+\mu)}}.$$

By the definition of the generating function and the special properties of the z-transform, we get

$$E[x(m+1)|x(m)] = \left.\frac{\partial G(z,T)}{\partial z}\right|_{z=1} = x(m)\mathrm{e}^{-T(\lambda+\mu)} + \frac{N\lambda}{\lambda+\mu}[1-\mathrm{e}^{-T(\lambda+\mu)}]. \tag{8}$$

Comparing the equation with Kalman filter (Eqs. (2) and (3)) formulation, we get

$$\begin{aligned} &A = \mathrm{e}^{-T(\lambda+\mu)},\\ &u(m) = 1,\\ &B = \frac{N\lambda}{\lambda+\mu}[1-\mathrm{e}^{-T(\lambda+\mu)}],\\ &C = b. \end{aligned} \tag{9}$$

Thus, from the Kalman filter setup, we get

$$\hat{x}(m) = A\hat{x}(m-1) + B + k(m)[\bar{y}(m) - CA\hat{x}(m-1) - CB], \tag{10}$$

where $k(m)$ is Kalman filter gain as defined in Fig. 2. This gives an estimate of the traffic on the link currently. This estimate will be used to forecast the traffic for the purpose of resource reservation.

2.3. Bandwidth request forecasting

The optimal estimate $\hat{x}(m)$ of the number of active flows can now be used to forecast $R(m+1)$, the resource requirement on the link $l(1,2)$ between AS1 and AS2. To this purpose, for $mT < t < (m+1)T$, we define

$$p_i(t) \triangleq \mathrm{prob}\{x(t) = i\}, \tag{11}$$

same as Eq. (4) in Section 2.2. We also define $\underline{P}$ and $\mathbf{Q}$ as

$$\underline{P} = \begin{bmatrix} p_0(t) \\ p_1(t) \\ \vdots \\ p_N(t) \end{bmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} -N\lambda & \mu & 0 & 0 & \\ N\lambda & -[(N-1)\lambda+\mu] & 2\mu & 0 & \cdots \\ 0 & (N-1)\lambda & -[(N-2)\lambda+2\mu] & 3\mu & \\ & \vdots & & & \ddots \end{pmatrix}$$

from Eqs. (5)–(7). It can be seen that $\dot{\underline{P}} = \mathbf{Q}\underline{P}$. It is easy to demonstrate that $\mathbf{Q}$ is similar to a real symmetric matrix and reducible to a diagonal form $\mathbf{Q} = \mathbf{Y}\mathbf{\Gamma}\mathbf{Y}^{-1}$ where $\mathbf{\Gamma}$ is a diagonal matrix with eigenvalues of $\mathbf{Q}$, and $\mathbf{Y}$ is the matrix of corresponding right eigenvectors y. The eigenvalues can be found to be $\gamma_k = -k(\lambda+\mu)$ for $k = 0,\ldots,N$ which proves that $\mathbf{Q}$ is non-positive definite,

guaranteeing the existence of a solution for $\dot{\underline{P}} = \mathbf{Q}\underline{P}$. The solution, for $t \in [mT, (m+1)T]$, is given by

$$\underline{P} = \mathbf{Y}\mathrm{e}^{\mathbf{\Gamma}t}\underline{C}, \tag{12}$$

where $\underline{C}$ is a constant vector determined from the initial condition ($x = x(m)$ at instant mT) as

$$\underline{C} = (\mathrm{e}^{\mathbf{\Gamma}mT})^{-1}\mathbf{Y}^{-1}\underline{P}_{mT}, \tag{13}$$

where $\underline{P}_{mT}$ is a vector with all 0's except the $x(m)$th element which is 1. Also we define

$$\widetilde{\underline{P}} \triangleq \frac{1}{T}\int_{mT}^{(m+1)T} \underline{P}\,\mathrm{d}t = \frac{1}{T}\mathbf{Y}\left(\int_{mT}^{(m+1)T} \mathrm{e}^{\mathbf{\Gamma}t}\,\mathrm{d}t\right)\underline{C} = \frac{1}{T}\begin{bmatrix}\tilde{p}_0 & \tilde{p}_1 & \cdots & \tilde{p}_N\end{bmatrix}^{\mathrm{T}} \tag{14}$$

using the notation that integral of a matrix is the integral of each element of the matrix. The elements $\tilde{p}_i$ of the vector $\widetilde{\underline{P}}$ denote the probabilities of transitioning to state i at instant $(m+1)T$. Now we define $\tilde{x}(m)$ as

$$\tilde{x}(m) \triangleq \min_{x \in [\hat{x}(m), N]} x \quad \text{s.t. } \tilde{p}_x < \beta. \tag{15}$$

In other words, we determine the minimum $\tilde{x}$ greater than or equal to $\hat{x}(m)$ such that the probability to be in state $\tilde{x}$ during the interval $(mT, (m+1)T]$ is less than a given threshold β, in effect choosing a state greater than the current utilization estimate such that the transition probability to the state is low. Then the resource requirement is forecasted to be

$$R(m+1) = b\tilde{x}(m). \tag{16}$$

3. Experimental results

The purpose of the simulations is to verify the accuracy of the proposed mechanisms for traffic estimation and bandwidth request forecast. Also we will compare the resource requirement forecasted using EPABB with results of other schemes. The results show how inter-domain agreements are adjusted depending on the traffic load and how closely they follow the load.

3.1. Estimation performance

In the simulation, the number of established sessions, N, is assumed to be 20 with λ and μ (parameters of exponential distributions for inter-arrival time and durations of flows, respectively) of 0.005 and 0.005. In other words, for each established session, the average inter-arrival time between flows and their average duration are 200 s each. Shown in Fig. 4 is the estimate $\hat{x}$ of the number of flows x, for a typical simulation run. The estimate is derived using the Kalman filter setup given in Eq. (10), from the noisy measurement $\bar{y}$. For this simulation, the measurement interval T, as defined in Section 2, was set at 1. The value of b, the constant rate for each flow of a traffic class, was set to 1 Mbps. The values for $E[x(0)]$ and $P(0|0) = E[x(0)x(0)^{\mathrm{T}}]$ in the initialization step of the Kalman filter were chosen to be 1 unit. The initial choice of $E[x(0)x(0)^{\mathrm{T}}]$ is not critical as long as it is non-zero because the filter will converge in any case. In the computation step, the previous values are used to compute the next values. Appropriate values were selected for process and measurement noise standard deviations as 1 and 1.5 units, respectively. If different initial seeds were chosen for $E[x(0)]$ and $P(0|0)$, the convergence time of the Kalman filter would vary, but the performance obtained would be similar, once the convergence is obtained. On the other hand,

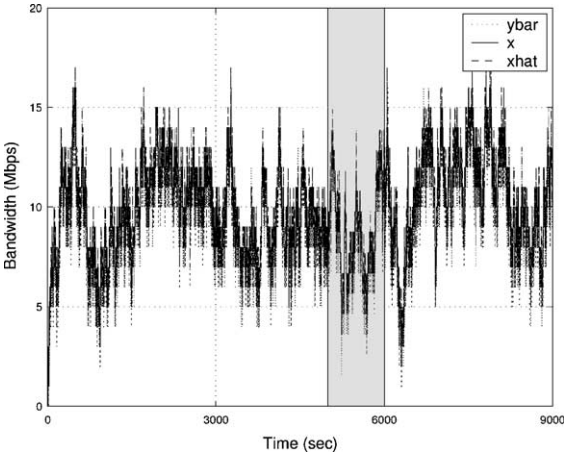

Fig. 4. Estimation performance.

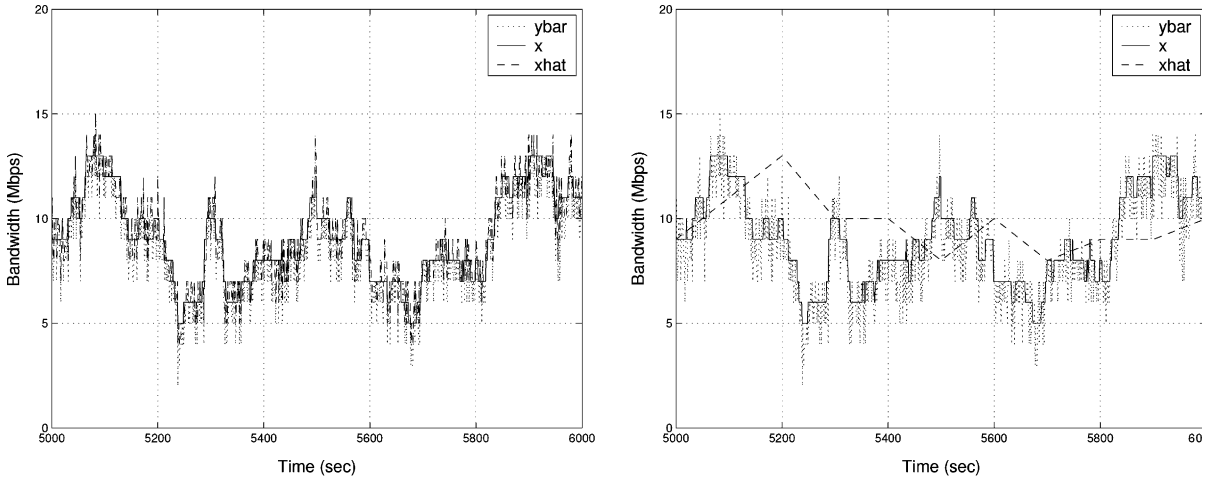

Fig. 5. Estimation performance (enlarged). (a) $T = 1$ s, (b) $T = 100$ s.

the variations in measurement noise standard deviation reflect in the performance of EPABB. If the standard deviation is higher, the filter is "slower" to believe the measurements and so is sluggish. If the standard deviation is smaller, the filter is "quicker" to believe the noisy measurements and follows the measurements more closely. Due to the small granularity of the estimation process, the estimated sequence has a very jagged profile in Fig. 4. In Fig. 5(a), the highlighted part of Fig. 4 has been enlarged to show that the estimated sequence is very close to the actual traffic, despite the noise in the measured sequence. If the granularity of the estimation procedure is increased, the estimation sequence becomes more smooth (as seen in Fig. 5(b)) but worse at the estimation, i.e., introduces estimation errors.

3.2. EPABB performance and comparison

Using the estimated sequence with high granularity (Fig. 4), we now forecast the resource requirement using the formulation given in Section 2.3. The forecast procedure computes the probabilities $p_i(t)$ of transitioning to all possible states from the current state (Eqs. (12) and (13)) and chooses the state whose transition probability is less than a threshold β as shown in Eq. (15). In the simulations, β was fixed at 1%, i.e., the forecasted state is chosen such that the system has less than 1% chance of exceeding that state. Using the memory-less property of the queuing model (Poisson arrivals and exponential durations), we can conclude that the probabilities $p_i(t)$, $t \in (mT, (m+1)T])$, as defined in Eq. (11) are independent of m. This simplification helps the simulation by reducing the computation effort and time. The calculation of $\underline{C}$ and $\widetilde{\underline{P}}$ (in Eqs. (13) and (14), respectively) can be performed off-line for all possible initial states and the results stored. Then the forecast process only involves a table-lookup to determine the next state at each instant for a current state based on Eq. (15).

If the forecast interval is small, there will be frequent changes in the forecasted value but less bandwidth wastage. On the other hand, if it is large, the forecasted sequence will be fairly stable at the expense of increased bandwidth wastage. Shown in Fig. 6(a) and (b) are the forecasted sequences ($\tilde{x}$) for small and large forecast intervals, respectively. As can be seen, the forecasted sequence follows the actual traffic more closely in the first case, at the expense of the amount of signaling effort required. In the second case, the forecasted sequence has a more stable profile (i.e. less signaling control is needed). The mean and standard

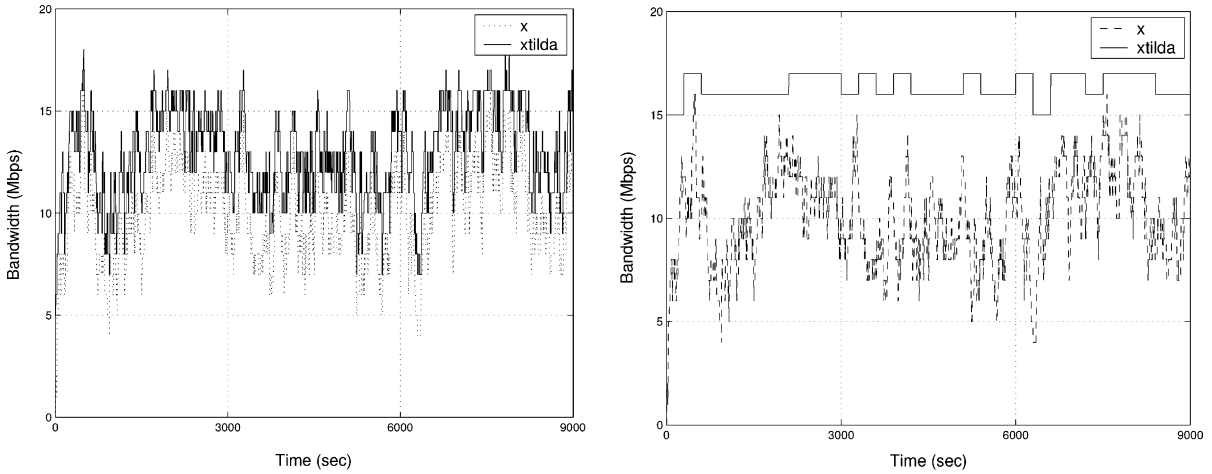

Fig. 6. Forecast performance. (a) $T = 10$ s, (b) $T = 300$ s.

deviation of the difference between the actual traffic and forecasted capacity requirement are 2.9 Mbps and 1.0, respectively, for small forecast interval in Fig. 6(a). The corresponding values are 6.35 Mbps and 2.19, respectively, for the large forecast interval in Fig. 6(b). These values reflect that the variation of the forecast error (about its mean) is small but the mean forecast error increases for large forecast intervals, which is expected.

In this paper, we have presented a scheme to forecast the resource requirement of an inter-domain link, as predicted by the bandwidth broker of the domain. Since, bandwidth brokers are still a subject of study and the number of available implementations can be counted on fingers, the prediction scheme that is currently used is very simple. It is the well-known "cushion"-based method [11] of over-provisioning where an estimate of the traffic utilizing the link is derived by measurement and then the resource requirement is forecasted to be the estimate plus a cushion to accommodate any fluctuations/measurement errors. Even though this scheme causes bandwidth wastage, it is the current method to determine resource requirements for an inter-domain link due to its simplicity. The prediction scheme utilized in bandwidth brokers should be simple and scalable as the bandwidth broker is a centralized server for the whole domain and thus has to perform similar computations for each peer domain. We can also compare the performance of EPABB with other well-known prediction schemes. The goal of the prediction scheme utilized by a bandwidth broker should be not to derive a near-perfect prediction, but to obtain an upper bound on the resource requirement which is not too conservative. This is because the resources on the links will be provisioned based on the predicted values and if the prediction is near-perfect, it can lead to blocking of new requests or degradation of service. With this aim in mind, the minimum mean square error linear predictor in [14] cannot be employed because it tries to predict the actual value of the measured sequence.

In the following, we present a comparison of the performance of EPABB with three different schemes for resource prediction. First is the cushion-based scheme [11,18]. Next is the prediction based on Gaussian assumption [19,20], from the central limit theorem, that the aggregate traffic resembles a Gaussian distribution. The last one is the auto-bandwidth allocator for MPLS from Cisco [21]. We can use the allocator concept without obtaining the tunnel reserved bandwidth but instead the link utilization estimate. The comparisons are provided in Fig. 7(a)–(c). The value of the forecast interval T is kept as 300 s for

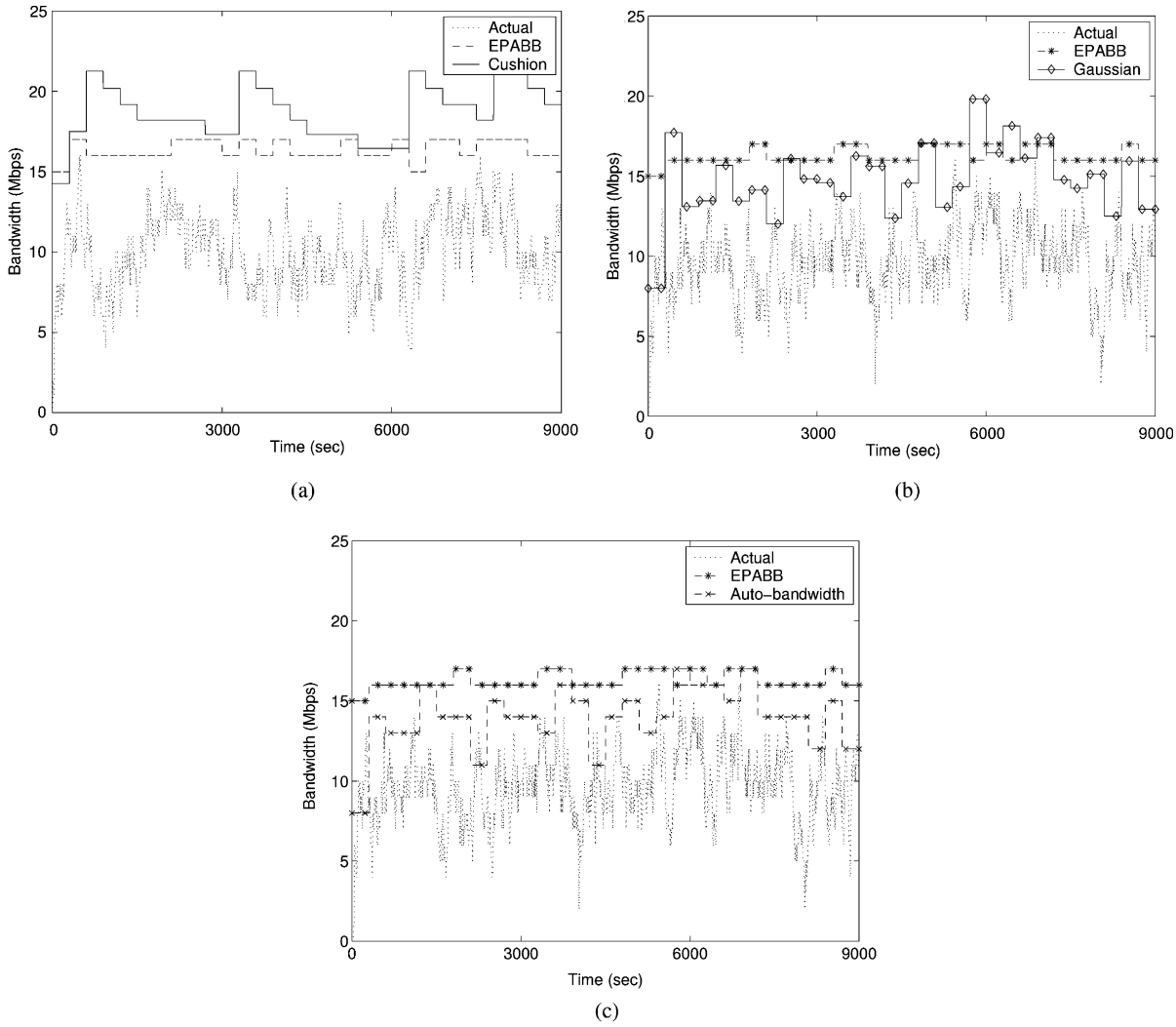

Fig. 7. Comparison with other methods; (a) cushion method, (b) Gaussian method and (c) auto-bandwidth method.

all the cases. To obtain the resource reservation for the cushion-based method, the traffic was first estimated using a time window measurement process with unity sampling period. This estimate was then used to calculate the resource reservation by over-provisioning a cushion of bandwidth. Whenever the traffic estimate is close to the current reservation level (watermark of 90%), the reservation level is increased in a multiplicative manner (by 25%). When the estimated traffic falls below a certain percentage of the current reservation level (80%), and stays there consistently for sometime, the reservation is reduced in an additive manner ($\gamma = 1/8$, $\beta = 2$, $\zeta = 4$). As can be seen from Fig. 7(a), the reservations achieved using this scheme are much higher than ours. This scheme thus leads to over-reservation of resources.

For the auto-bandwidth allocator scheme, we obtain the resource reservation by first obtaining the traffic measurement using a time window with unity sampling period. The reservation for a time window is determined to be the maximum link utilization value obtained from the previous time window. For the Gaussian assumption-based al-

location method, the measurements from the previous time window are obtained as before; but the allocation is determined to be the mean and 3σ from the previous window. This scheme allows for a cushion which reflects in the 3σ factor which can be reduced for a lower cushion at the expense of higher degraded QoS, as defined later. As can be seen, both these schemes infer the reservation for a time window based on the measurements from the previous time window. This introduces a lag in the reservation profile compared to the utilization. As can be seen from Fig. 7(b) and (c), both schemes have low values of the over-allocation but a very high variability in the reservation profile.

We propose the use of three parameters to measure and compare the performance of a resource allocation scheme, namely *switching rate*, *bandwidth wastage* and *degraded QoS factor*. Switching rate defines the rate at which the level of the reserved resources needs to be switched to realize the desired allocation profile. It includes the increments as well as decrements in the allocation profile. Bandwidth wastage is a measure of the over-allocation of the bandwidth resources. The mean and

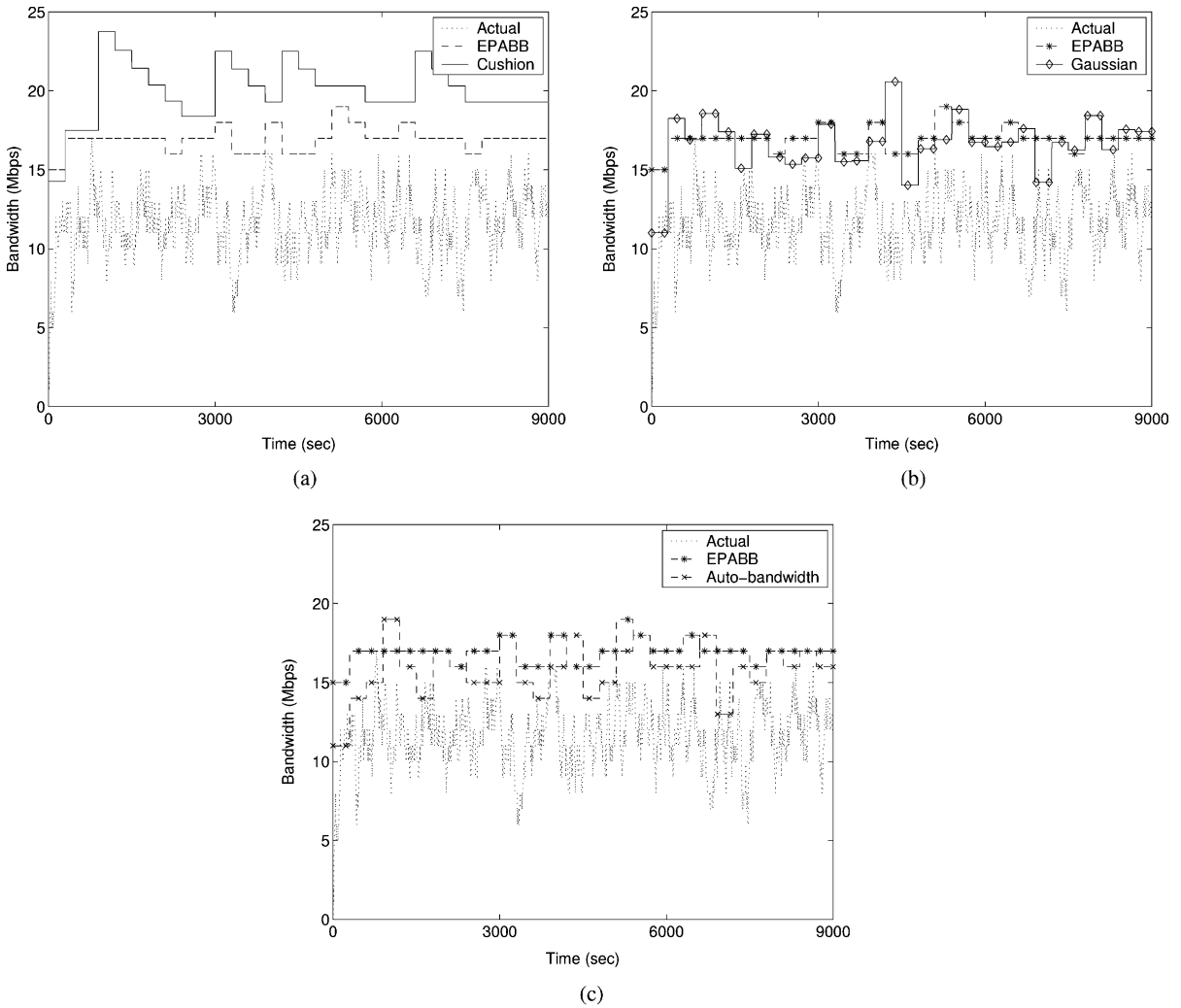

Fig. 8. Increased λ: (a) cushion method, (b) Gaussian method, and (c) auto-bandwidth method.

standard deviation of the over-allocation can be used to represent its stochastic properties. Degraded QoS factor measures the percentage of bandwidth requests that will receive a degraded QoS because there is not ample reservation for them on the inter-domain link.

Comparing the reservations obtained from EPABB and cushion-based scheme for the example given in Fig. 7(a), we see that the switching rates are $1.7e-3$ and $2.1e-3$ ms^{-1}, respectively. The mean and standard deviation of the over-allocation by EPABB are 6.35 Mbps and 2.19, respectively whereas for cushion scheme, 8.65 Mbps and 2.84, respectively. The proposed scheme reduces the bandwidth wastage by 42% compared to the cushion-based scheme. The degraded QoS factor for both cases are 0. So, in this case EPABB reduces the over-allocation of resources without increasing the switching rate and no degraded QoS. From Fig. 7(b) and (c), we see for the auto-bandwidth allocator and the Gaussian-based allocator, the switching rates are much higher ($2.67e-3$ and $3.22e-3$ ms^{-1}, respectively) and the over-allocation is considerably lower, but the degraded QoS

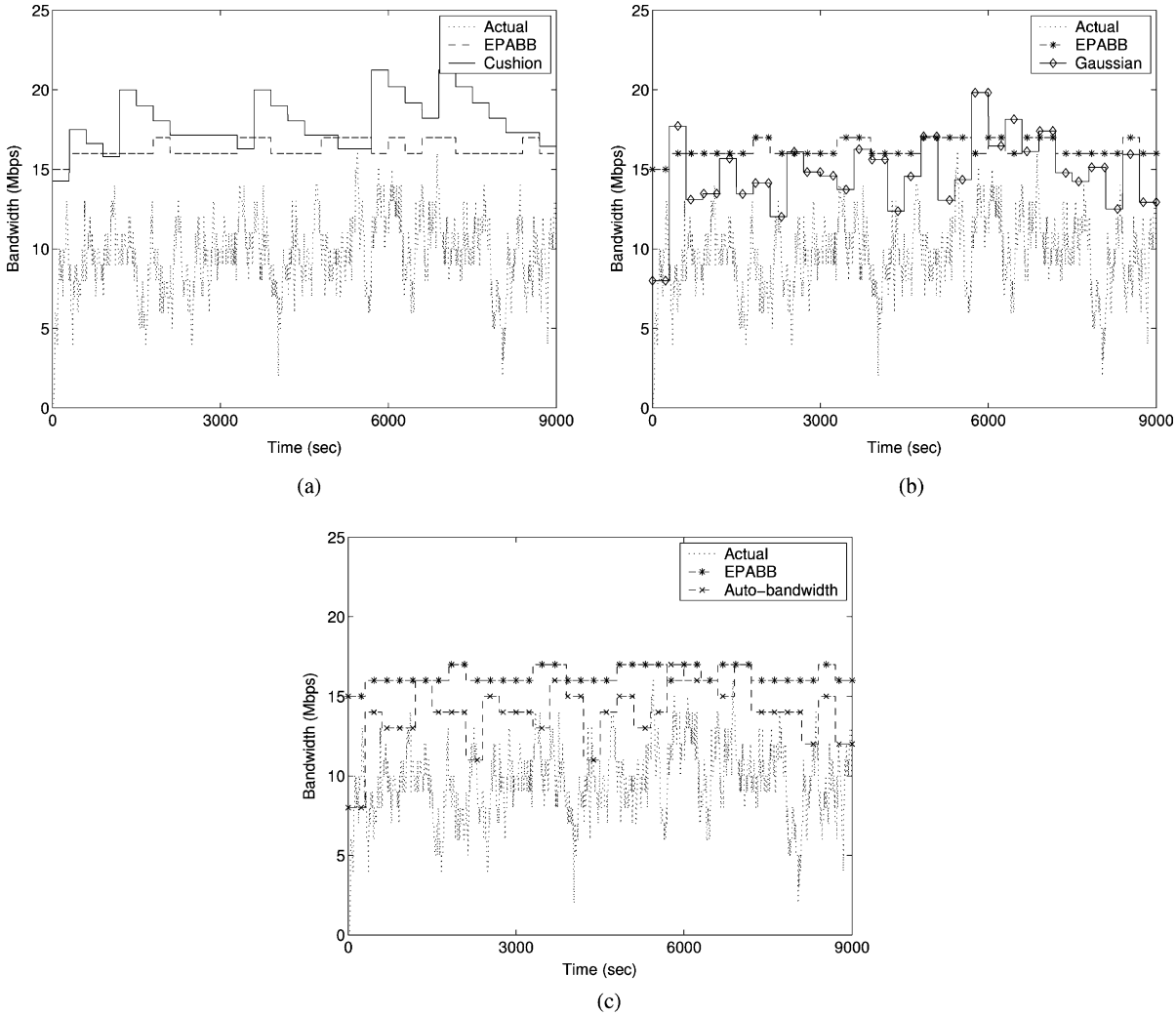

Fig. 9. Increased λ and μ: (a) cushion method, (b) Gaussian method, and (c) auto-bandwidth method.

factor is increased (1.95% and 1.43%, respectively) due to the phase lag involved. For high QoS demanding traffic, the degraded QoS can be a problem.

3.3. Robustness and sensitivity analysis

Next, we verify the robustness of EPABB. Two tests are performed to analyze how the proposed scheme fares if the assumptions made for the analysis are removed. For the first test, the simulated input traffic is modified such that it retains its Markovian properties but the rate of arrival and durations of the bandwidth requests (i.e., the parameters λ and μ) are varied while the Kalman filter estimation still uses the old values. If the Kalman filter estimation procedure overestimates the associated parameters of the traffic, the estimation will contain error in the form of increased overestimation of resources, which is not as harmful as underestimation or increased degraded QoS factor. On the other hand, the estimation procedure should be robust to the underestimation of the parameters. In Figs. 8 and 9, we demonstrate two such cases. From the results given in Fig. 8(a), we can infer that for a 40% underestimation, in the estimation procedure, of λ used for traffic generation, the switching rate of EPABB is still lower than the cushion-based scheme (compare $1.67e-3$ and $2.11e-3$ ms^{-1}) while the degraded QoS factor remains at 0% for both. The bandwidth wastage by EPABB is now 36% lower than the cushion-based method. The auto-bandwidth allocator and the Gaussian-based allocator have relatively higher switching rate and degraded QoS factor with a lower wastage of bandwidth. This test shows the tolerance of the proposed approach to misjudgment of traffic characteristics. We also test a scenario where both λ and μ for the traffic are underestimated in the Kalman filter estimation. From the results given in Fig. 9(a), we see that for a 40% variation in both λ and μ, the degraded QoS factor is still 0% for both EPABB and cushion scheme but the switching rate is lower for EPABB (compare $1.44e-3$ and $2.56e-3$ ms^{-1}). The results for the auto-bandwidth allocator and the Gaussian-based allocator are similar to the previous case. Also, we check the sensitivity of EPABB to variations in N, the number of established sessions. Theoretically, N is known to the bandwidth broker as each session is reserved after the broker provisions resources for the session. The effects on the performance of EPABB due to variations in N are shown in Fig. 10. From Fig. 10(a), we see that for a 25% underestimation of N, EPABB performs fairly well, with low switching rate and marginal degraded QoS factor. On the other hand, for a 25% overestimation of N (in

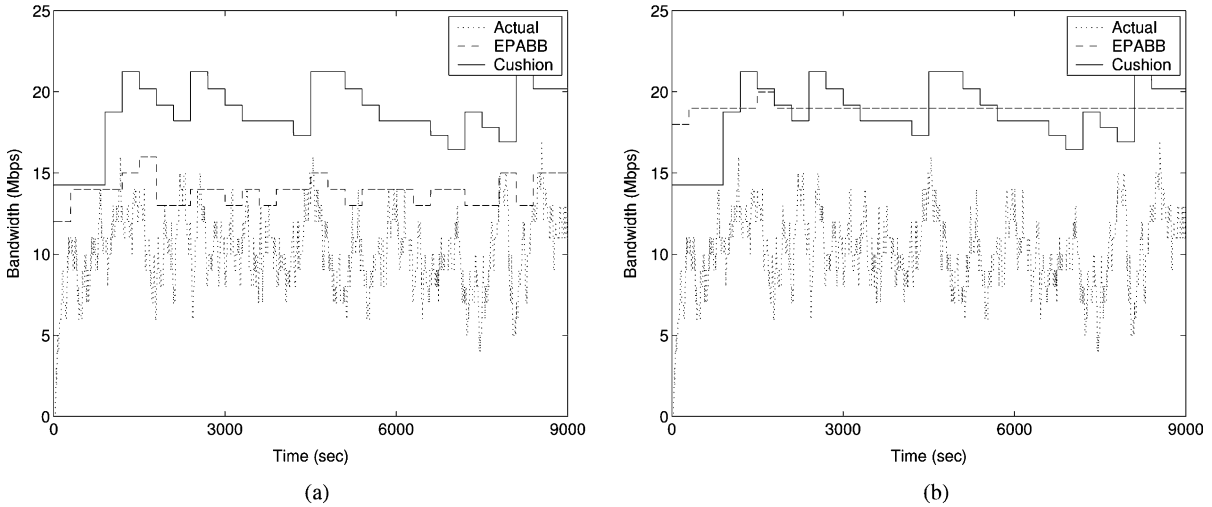

Fig. 10. Sensitivity of EPABB to N: (a) $N = 15$, (b) $N = 25$.

Fig. 10(b)), the bandwidth wastage increases but it is still lower than the cushion-scheme.

The second test for verifying the robustness of the proposed scheme involves applying the scheme to an actual traffic profile obtained from measuring the traffic on a link. In this way, we verify the effect of removal of the Markovian assumption on the traffic. For this purpose, we have used a traffic profile obtained from the publicly available traffic archives of NLANR, an organization that provides technical, engineering, and traffic analysis support to high performance connections sites. This obtained profile is used as input to the Kalman filter estimator and subsequently the capacity predictor. The values for N, λ and μ can be derived from observing the traffic for some time in the past. We assume that the value of N is known from the SLA for the established sessions. The values of λ and μ can be derived by averaging the inter-arrival times and the inter-departure times. As the system is modeled as $M/M/N$, the average inter-arrival time is approximately $N\lambda\mu/(\lambda+\mu)$ and the

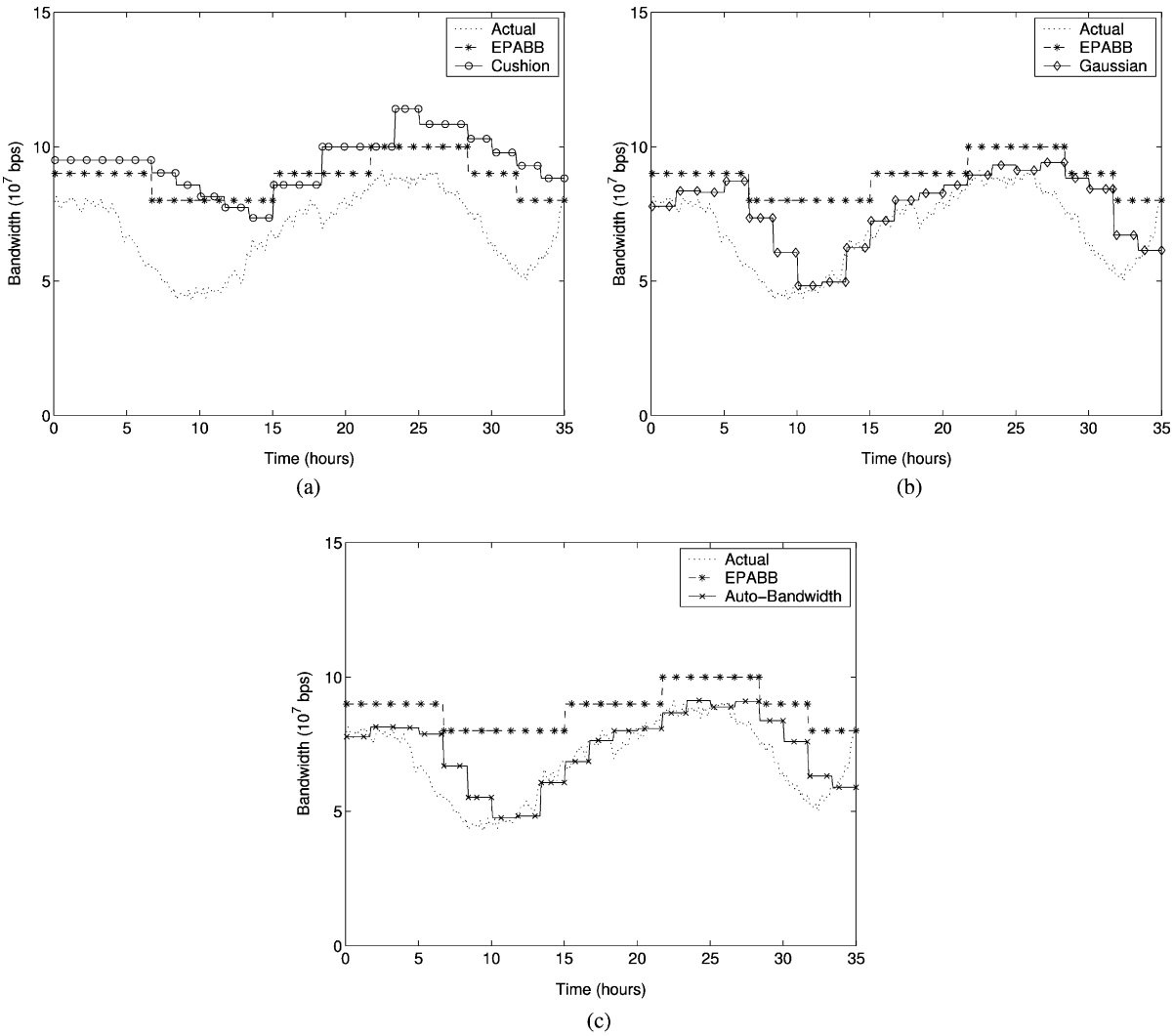

Fig. 11. NLANR traffic capacity prediction: (a) cushion method, (b) Gaussian method, and (c) auto-bandwidth method.

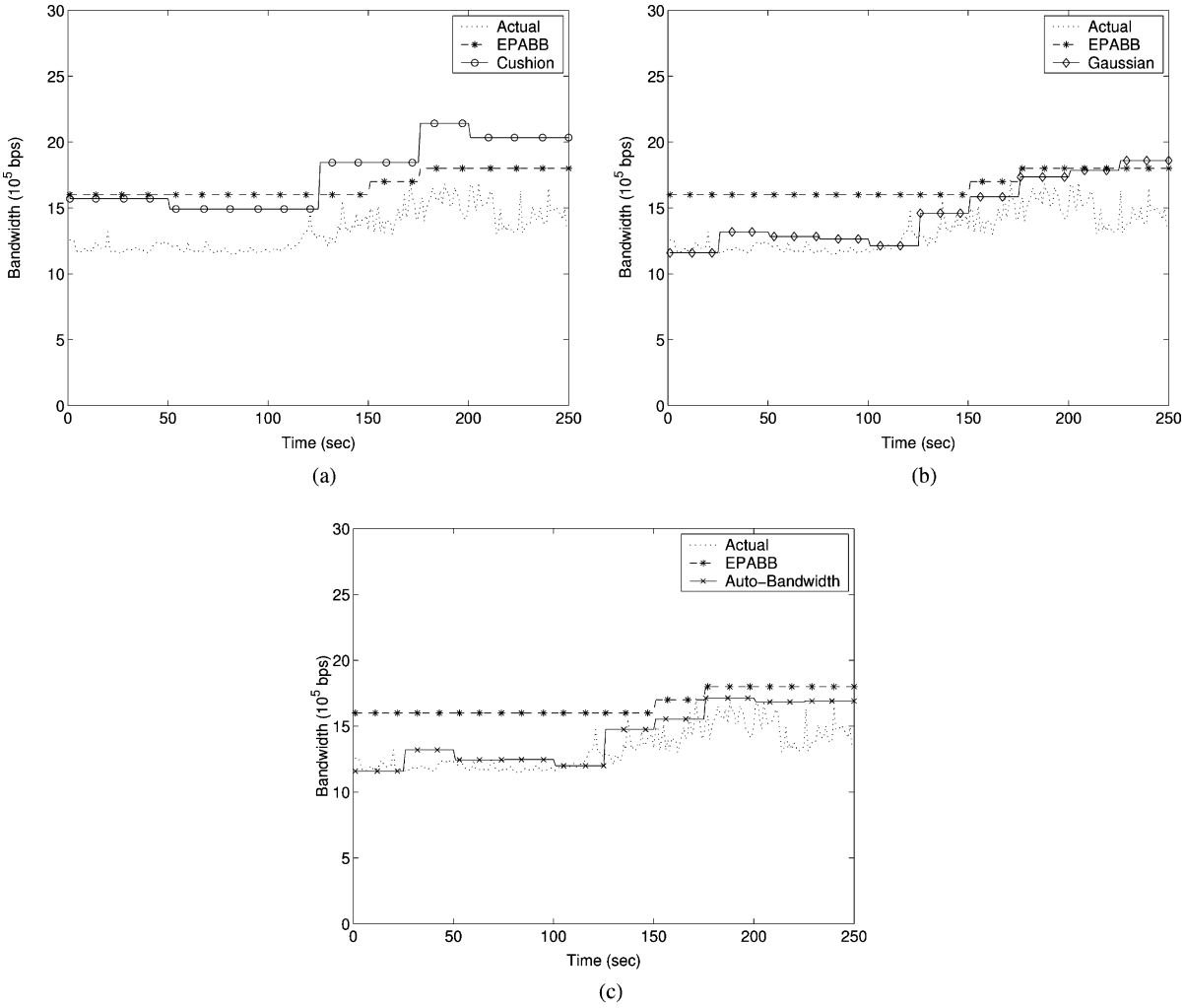

Fig. 12. "Georgia Tech" traffic capacity prediction: (a) cushion method, (b) Gaussian method, and (c) auto-bandwidth method.

average inter-departure time is $N\mu^2/(\lambda+\mu)$. The values for λ and μ need not be changed frequently as we have already shown the insensitivity of the EPABB performance to the estimation error. Fig. 11(a)–(c) show that the proposed scheme is able to predict the capacity requirement well and has low switching rate, bandwidth wastage and degraded QoS factor. The figures compare the performance of EPABB to the other three schemes. In Fig. 12(a)–(c), we compare EPABB performance for a trace of traffic exiting Georgia Institute of Technology gateway to the ISP. The traffic profile used in both the figures have wide variations. Nevertheless, the EPABB scheme is able to predict the resource requirements very efficiently, without any changes in the prediction procedure or parameters.

4. Conclusions

A bandwidth broker is an agent associated with each domain and keeps track of the inter-domain SLAs with neighboring domains. The amount of resources reserved on the inter-domain links, for

each DiffServ class, is an important parameter as both over-reservation and under-reservation are not desirable. Over-reservation has the advantage of infrequent variations but leads to wastage of resources. Under-reservation, on the other hand, does not meet the QoS expectations of the user flows. A simple method for over-reservation is based on providing a cushion. In this paper, we have proposed a scheme to estimate the traffic on an inter-domain link by the use of Kalman filter and then forecast the capacity requirement at a future instant by the use of transient probabilities of the system states. Switching rate, bandwidth wastage and degraded QoS factor have been used as metrics to evaluate the performance of the proposed scheme. Also the robustness of the scheme has been verified by using an actual traffic pattern. The simulation results confirm the robustness of the scheme and show reduced wasted bandwidth.

Acknowledgements

This work was supported by NASA Goddard, Swales Aerospace under contract number S11201 (NAS5-01090) and NSF under award number 0219829.

References

[1] K. Nichols, S. Blake, F. Baker, D. Black, Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers, IETF RFC 2474, http://www.ietf.org/rfc/rfc2474, December 1998.

[2] K. Nichols, V. Jacobson, L. Zhang, A two-bit differentiated services architecture for the Internet, IETF RFC 2638, http://www.ietf.org/rfc/rfc2638, July 1999.

[3] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, Resource reSerVation protocol (RSVP) Version 1 Functional specification, IETF RFC 2205, http://www.ietf.org/rfc/rfc2205, September 1997.

[4] B. Teitelbaum, P.F. Chimento, QBone BB architecture, Internet2 Document, http://qbone.internet2.edu/bb/bboutline2.html, June 2000.

[5] K. Chan, J. Seligson, D. Durham, S. Gai, K. McCloghrie, S. Herzog, F. Reichmeyer, R. Yavatkar, A. Smith, COPS usage for policy provisioning (COPS-PR), IETF RFC 3084, http://www.ietf.org/rfc/rfc3084, March 2001.

[6] Z. Zhang, Z. Duan, L. Gao, Y.T. Hou, Decoupling QoS control from core routers: A novel bandwidth broker architecture for scalable support of guaranteed services, in: Proceedings of ACM SIGCOMM'00, Stockholm, Sweden, August 2000, pp. 71–83.

[7] Z.L. Zhang, Z. Duan, Y.T. Hou, L. Gao, Decoupling QoS control from core routers: A novel bandwidth broker architecture for scalable support of guaranteed services ACM SIGCOMM Computer Communications Review 30 (4) (2000).

[8] BCIT Projects Website, http://www.tc.bcit.ca/gait/projects/.

[9] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, A. Roy, A distributed resource management architecture that supports advance reservations and co-allocation, in: Proceedings of 7th International Workshop on Quality of Service, London, UK, June 1999, pp. 27–36.

[10] I. Foster, A. Roy, V. Sander, A quality of service architecture that combines resource reservation and application adaptation, in: Proceedings of 8th International Workshop on Quality of Service, Pittsburgh, USA, June 2000, pp. 181–188.

[11] A. Terzis, L. Wang, J. Ogawa, L. Zhang, A two-tier resource management model for the Internet, in: Proceedings of IEEE GLOBECOM'99, Rio de Janeiro, Brazil, December 1999, pp. 1779–1791.

[12] J.W. Roberts, Traffic theory and the Internet, IEEE Communications Magazine 39 (1) (2001) 94–99.

[13] W.C. Feng, F. Jahanian, S. Sechrest, An optimal bandwidth allocation strategy for the delivery of compressed prerecorded video, Multimedia Systems Journal 5 (5) (1997).

[14] A. Adas, Supporting real-time VBR video using dynamic reservation based on linear prediction, in: Proceedings of IEEE INFOCOM'96, San Francisco, USA, March 1996, pp. 1476–1483.

[15] H. Zhang, E.W. Knightly, RED-VBR: A renegotiation based approach to support delay sensitive VBR video, Multimedia Systems 5 (1997) 164–176.

[16] D. Reininger, G. Ramamurthy, D. Raychaudhuri, VBR MPEG video coding with dynamic bandwidth renegotiation, in: IEEE International Conference on Communications, Seattle, USA, June 1995, pp. 1773–1777.

[17] P.S. Maybeck, Stochastic Models, Estimation, and Control, Academic Press, New York, 1979.

[18] A. Terzis, A two-tier resource allocation framework for the Internet, Ph.D. thesis, University of California, Los Angeles, 2000.

[19] C.N. Chuah, L. Subramanian, R.H. Katz, A.D. Joseph, QoS provisioning using a clearing house architecture, in: Proceedings of 8th International Workshop on Quality of Service, Pittsburgh, USA, June 2000.

[20] N.G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K.K. Ramakrishnan, J.E. van der Merwe, A flexible model for resource management in virtual private networks, in: Proceedings of ACM SIGCOMM'99, Cambridge, USA, September 1999, pp. 95–108.

[21] White Paper, Cisco MPLS AutoBandwidth Allocator, http://www.cisco.com/warp/public/cc/pd/iosw/prodlit/mpatb_wp.htm, June 2001.
[22] S. Keshav, A control-theoretic approach to flow control, in: Proceedings of ACM SIGCOMM'91, Zurich, Switzerland, September 1991, pp. 3–15.
[23] A. Kolarov, A. Atai, J. Hui, Application of Kalman filter in high-speed networks, in: Proceedings of IEEE GLOBECOM'94, San Francisco, USA, November 1994, pp. 624–628.
[24] T. Bonald, S. Oueslati-Boulahia, J. Roberts, IP-traffic and QoS control: Towards a flow-aware architecture, in: Proceedings of 18th World Telecommunication Congress, Paris, France, September 2002.
[25] B. Cipra, Engineers look to Kalman filtering for guidance, SIAM News 26 (5) (1993).
[26] L. Kleinrock, Queueing Systems, Wiley, New York, 1975.

Tricha Anjali received the (Integrated) M.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1998. Currently, she is a research assistant in the Broadband and Wireless Networking Laboratory pursuing her Ph.D. degree. She is a student member of the IEEE Communications Society. Her interest is to investigate quality of service (QoS) issues in the next generation Internet (NGI).

Caterina Scoglio received the Dr. Ing. degree in electronics engineering from the University of Rome "La Sapienza", Italy (summa cum laude) in May 1987. She received a post-graduate degree in mathematical theory and methods for system analysis and control from the University of Rome "La Sapienza", Italy, in November 1988. From June 1987 to June 2000, she had been with Fondazione Ugo Bordoni, Roma, where she was a research scientist at the TLC Network Department—Network Planning Group. During the period of November 1991 to August 1992, she had been a visiting researcher at Georgia Institute of Technology "College of Computing" in Atlanta, GA, USA. Since September 2000, she has been with the Broadband and Wireless Networking Laboratory of the Georgia Institute of Technology as a research engineer. Her research interests include optimal design and management of multi-service networks.

George Uhl is the Lead Engineer at NASA's Earth Science Data and Information System (ESDIS) Network Prototyping Lab. He directs network research and protoyping activities for ESDIS. His current areas of research include network quality of service and end-to-end performance improvement.