FRR for Latency Reduction and QoS Provisioning in OBS Networks

Jingxuan Liu, *Student Member, IEEE*, Nirwan Ansari, *Senior Member, IEEE*, and Teunis J. Ott, *Member, IEEE*

Abstract—We propose a forward resource reservation (FRR) scheme to reduce the data burst delay at edge nodes in optical burst switching (OBS) systems. We also explore algorithms to implement the various intrinsic features of the FRR scheme. Linear predictive filter (LPF)-based methods are investigated and demonstrated to be effective for dynamic burst-length prediction. An aggressive resource reservation algorithm is proposed to deliver a significant performance improvement with controllable bandwidth cost. By reserving resources in an aggressive manner, an FRR system can reduce both the signaling retransmission probability and the bandwidth wastage as compared with a system without the aggressive reservation. An FRR-based QoS strategy is also proposed to achieve burst delay differentiation for different classes of traffic. Theoretical analysis and simulation results verify the feasibility of the proposed algorithms and show that our FRR scheme yields a significant delay reduction for time-critical traffic without incurring a deleterious bandwidth overhead.

Index Terms—Bandwidth overhead, latency reduction, linear predictive filter (LPF), optical burst switching (OBS), quality-of-service (QoS), resource reservation.

I. INTRODUCTION

OPTICAL burst switching (OBS) provides a feasible paradigm for Internet protocol (IP) over wavelength-division multiplexing (WDM) integration, which has been the focus of intense investigation owing to its flexibility in utilizing the terahertz bandwidth of a single optical fiber and its capability to support transparent data transmissions. However, with the emergence of multitype applications such as data, voice, and videoconferencing, the next-generation network must also be designed to provide a variety of quality-of-service (QoS) functionalities. One of the major challenges is the latency reduction issue [1], [2]. It has been widely recognized that, to date, the bandwidth is no longer the transmission bottleneck in many core networks, but it is the latency that dominates the transmission time and is becoming of paramount importance. The foregoing challenge has made it increasingly important for a network to support delay-sensitive applications and to facilitate QoS provisioning. This is also a critical consideration when designing an OBS system.

The basic ideas underlying an OBS system are twofold: the burstification of IP packets, and the decoupling of the transmission and switching of a control header and its data payload. A control header, also called a burst header packet (BHP), is transmitted in an earlier time window than its data payload. While a BHP is processed at each and every intermediate node in the core network to reserve resources and set up a switching path, the corresponding data payload is switched throughout the network transparently without the need to interpret the data format or bit rate. Such separation maintains the desirable property of data transparency and leads to a better synergy of both the mature electronic technologies and advanced optical technologies [3], [4].

One of the main advantages of an OBS approach lies in its switching granularity, i.e., a data burst. It is a solution to compensate for the time constraint of directly switching individual IP packets at optical routers due to the mismatch between the transmission capability of WDM fibers and the processing capability of the electronic control plane, thus alleviating the heavy burden of electronic devices for lightpath configuration. This advantage results from the particular procedure of burstification [4], whereby multiple IP packets are aggregated into a single data burst at the network ingress. A side effect imposed by such a burst-buildup process, however, is an artificial delay. The typical end-to-end delay of a data burst thus mainly consists of three components: burst assembly delay at edge routers, path setup delay caused by control headers, and the propagation delay in the core network.

We observe that the bandwidth at the core network (OC192 and beyond) is much higher than that in the edge network (OC3-OC48). The time for assembling a burst, which usually consists of hundreds of IP packets and is at the time scale of hundreds of microseconds, is comparable with the switching path setup time, which is also presumed to be in the range of microseconds [5]. The burst delay at network edges is substantial and has a significant impact on the end-to-end burst delay. This influence is especially detrimental to the real-time traffic, which has stringent delay constraints. Since the propagation time of a data burst, which is intrinsic, cannot be reduced, reducing burst delay at network ingresses will be greatly beneficial to latency reduction and QoS provisioning.

The latency reduction issue has been incorporated in many OBS system designs [4]–[7]. For example, the typical one-way signaling protocol of just-in-time (JIT) mechanism is particularly designed for ultra-low latency unidirectional transport of data bursts across an optical network. Xiong *et al.* [4] provided a thorough discussion on the architecture and design considerations of OBS networks. The delay reduction problem in that paper is considered in terms of the switching node

architectures, scheduling algorithms, and burst assembly duration. A common characteristic of these studies is that, however, the end-to-end burst delay reduction mechanism has been focused on the core network, from the perspectives of signaling protocols, hardware designs, and scheduling algorithms at network intermediate nodes.

In this paper, we propose a novel transmission mechanism at network ingresses, called forward resource reservation (FRR), to facilitate the end-to-end burst delay reduction functionality. We also propose algorithms to implement the intrinsic features of the FRR scheme. Our proposals have the following characteristics: First, we advocate the latency reduction mechanism at network edges, the interfaces between the metropolitan area network (MAN) and the WDM backbone, where multiple IP packets are assembled into a single burst. Second, the end-to-end burst delay at the application level is reduced by a parallel strategy, i.e., by masking the impact of multiple latency contributors. Third, we make use of the available bandwidth in core networks for QoS support. Fourth, the FRR scheme reduces the real-time data burst delay without increasing that of nonreal-time bursts. And last but not least, our algorithms are based on mature techniques and are simple to implement.

Our contributions include the following: 1) we propose an innovative transmission mechanism embedded at ingress nodes to efficiently reduce the end-to-end burst delay; 2) we justify the feasibility of a least mean square (LMS)-based linear predictive filter (LPF) for dynamic burst-length prediction; and 3) an efficient resource reservation algorithm is derived to achieve controllable FRR performance enhancement, including improved latency reduction capability, lowered average bandwidth cost and reduced signaling retransmission probability. Besides the theoretical analysis, extensive simulations have been conducted to evaluate the system performance. Three distinctive traffic scenarios have been considered: MPEG video traces, Poisson traffic, and self-similar traffic. We devote more efforts on the self-similar traffic which is essentially the *de facto* trend to characterize the multitimescale burstiness of the Internet traffic [8], [9]. Since measurement of such real networks has shown that the aggregated traffic exhibits long-range dependence, which leads to substantially different performance evaluations from those based on the Poisson traffic model, it is practical and necessary to factor in the self-similarity property when we discuss any system design and evaluate its performance.

The rest of the paper is organized as follows. Section II specifies the OBS system environment and our system model. Section III describes the basic transmission mechanism of an FRR scheme. Its intrinsic characteristics, in terms of burst-length prediction, aggressive resource reservation, and QoS provisioning, are presented in Section IV. In Section V, we investigate the system performance by theoretical analysis and simulation evaluations. Conclusions are drawn in Section VI.

II. SYSTEM DESCRIPTION AND PROBLEM STATEMENT

This section details the system model to which an FRR-based transmission strategy applies. A brief review of related work or approaches on different aspects of the OBS system is introduced. We also formulate the problem we will solve.
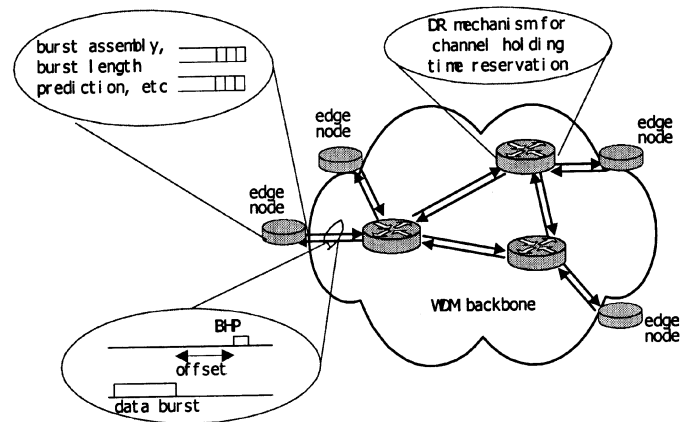

Fig. 1. System model.

A. System Model

Fig. 1 highlights the architecture of an OBS network under investigation. Our focus is on the latency reduction mechanism at WDM network edges.

Edge nodes of such networks are divided into two nondisjoint sets: ingress (source) nodes and egress (destination) nodes. A node can be a source node as well as a destination node at the same time. Burstification is performed at ingresses, where a burstification control unit (BCU) [4] resides and coordinates the assignment and transmission of data channels and control channels. In our system, the BCU also performs functions particular to our algorithms. We employ the time-based burst assembly mechanisms [4], [10], whereby a new burst is aggregated and is ready to be sent into the core network when a predefined threshold is reached (e.g., a timer expires).

Signaling protocols for OBS systems are based on two alternative schemes: "Tell-and-wait" (TAW) and "tell-and-go" (TAG) [7]. While the former features a two-way reservation, the TAG scheme uses the one-way signaling, i.e., at the ingress node, a control packet is sent out and after a certain offset time, without waiting for confirmation from the network, its data burst is transmitted. Some popularly discussed protocols (e.g., the JIT [5], [6] and just-enough-time (JET) [7], [11]) are TAG in nature. For more details on the TAG-based signaling protocol and the related work, interested readers are referred to [6] and [7] and references therein. In our scenario, we adopt the TAG scheme because it incurs a shorter burst delay at a source node. The latency reduction strategy proposed in this paper supports both JIT and JET protocols.

Reservation schemes of an OBS system differentiate from each other depending on how an intermediate switch node is made aware of the beginning and the ending of a data burst. Four main reservation schemes are discussed in literature [6]: 1) explicit setup and explicit release; 2) explicit setup and estimated release; 3) estimated setup and explicit release; and 4) estimated setup and estimated release. These variants result in different complexity of hardware requirements and different amount of time that the switching elements are reserved for an individual burst. In this paper, we assume that the switching matrix is set up and reserved for a burst according to scheme II. In addition, we employ an improvement of this approach,

termed delayed reservation (DR) [11], i.e., the resources at intermediate nodes are reserved for the incoming data payload from its arrival time and are released (torn down or timed-out) at its departure time, determined from the arrival time of the BHP and the burst length. This approach enables a BHP to reserve resources for a more precise duration that corresponds to the burst length, and delivers efficient bandwidth utilization and high system throughput. A BHP in this scenario has the knowledge of its payload, including ingress/egress node identification and the data burst length.

The application streams in our WDM network are specified by two parameters. One is the traffic load which characterizes the incoming traffic intensity. The other is the delay allowance which indicates the time constraint. According to this parameter, we partition the traffic into M QoS classes, with the class-j traffic being more delay-sensitive than the class-k one when $0 \leq j < k \leq M - 1$. The QoS requirement considered in this paper is the delay constraint.

B. Problem Statement

Based on the above system model and service requirements, we can formulate the latency reduction issue at the edges of the core network as follows. Given the OBS system supporting multiple classes of applications, each with different delay constraints, the problem we are facing now is to design an ingress node transmission scheme to provide less end-to-end burst delay for the delay-sensitive traffic, while keeping the network cost within limits. We also investigate the enabling technologies which are indispensable for the proposed scheme to deliver satisfactory performance figures of merits.

III. FRR SCHEME

In this section, we present the novel FRR scheme to reduce the end-to-end burst delay based on our system model.

A. Motivations

In a typical OBS system, the transmission of a BHP is dependent on the burst assembly process [12], [13]. To acquire the necessary information of its data payload, including the data burst length, a BHP waits for the completion of the burst assembly before it is transmitted for signaling and resource reservation. To allow enough time for switching nodes to process the BHP and to set up the switching matrix, the data payload should be further delayed at the ingress node for an offset time before being launched into the core network. The data burst delay at an edge node has to account for these two factors, both considerable sources of delay.

Our intuitive idea on this observation is that, rather than performing the above two processes in sequence, the burst assembly procedure and the transmission of a BHP should be processed in parallel, and thereby minimize their impact on the total end-to-end burst delay. A brief summary of the design objectives of our FRR scheme includes the following.

- A BHP specifies, among other information, a reservation duration, which corresponds to the length of its data payload.

TABLE I
NOTATIONS FOR THE FRR SCHEME ($i = 0, \ldots, M - 1$)

Term	Explanation
T_b^i	The time when a new burst of class-i begins to assemble at an edge node
T_h^i	The time when a class-i BHP is sent into the core network
T_d^i	The time when a class-i data burst is sent into the core network
τ_a^i	The duration to assemble a burst of class-i traffic
τ_o^i	The offset between a class-i BHP and its data payload. τ_o^i can adopt a pre-existing protocol that is most efficient in the system

- While preserving the all-optical transparency advantage for its payload, a BHP should enable the data burst to be transmitted as early as possible, thus resulting in minimum latency at the source node.
- The system can behave differently for different classes of traffic and achieve service differentiation in terms of the burst delay.

Our FRR scheme meets the first two requirements by a parallel strategy, and it is further extended to facilitate the QoS-capable requirement.

B. FRR for Ingress Node Transmission

To explain the FRR scheme, we first define some notations that simplify our description (Table I). In part of this paper, when we discuss the behavior and performance of an individual traffic class to which the FRR scheme applies, for notational simplicity, the referencing of traffic class is omitted.

An FRR scheme involves a three-step procedure as follows.

- *Phase 1*: *Prediction*. As soon as the previous burstification is done and a new burst assembly begins at T_b, the BCU predicts the length of the next incoming data burst. This estimation is based on a linear prediction method, as will be discussed in the next section.
- *Phase 2*: *Pretransmission*. Instead of waiting for the burst assembly to complete, a control header is constructed instantly upon the completion of the prediction. The BCU injects into the BHP the information necessary for path setup, including a resource reservation length which is determined with an aggressive reservation algorithm. The BHP is then launched into the core network at time T_h ($T_h = \max\{T_b, T_b + \tau_a - \tau_o\}$).
- *Phase 3*: *Examination*. When the burst assembly is fully carried out, the actual burst length is compared with the reservation length in the pretransmitted BHP to ensure the prereserved duration is enough for the actual burst length. There are two cases of interest to consider:

  1) If the actual burst length is less than or equal to the prereserved duration, i.e., the BHP has reserved enough bandwidth for the data payload, the BHP pretransmission is deemed a success. In this case, the data burst is sent into the core network at $T_d = T_h + \tau_o$.
  2) If the actual burst length exceeds the prereserved duration, the BHP pretransmission is deemed a failure. The BHP has to be retransmitted for this burst at a
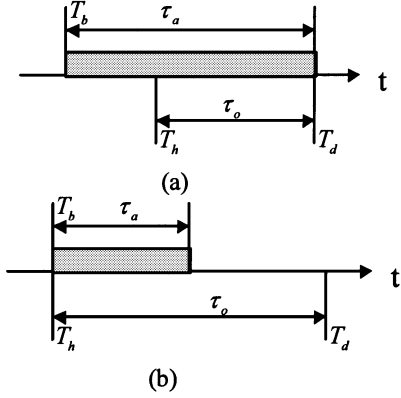
Fig. 2. FRR principle. (a) $\tau_a \geq \tau_o$. (b) $\tau_a < \tau_o$.

later time of $T_b + \tau_a$ with the actual burst size, and the data payload lags behind by the offset τ_o.

Note that our FRR scheme does not introduce any extra burst delay. Even a failed forward reservation causes the same latency with a transmission not using the FRR scheme. Fig. 2 depicts the principle of the basic FRR scheme when a pretransmission of the BHP succeeds.

Our FRR scheme features two prominent advantages: First, the advanced transmission of a BHP enables concurrent performing of burstification and resource reservation. This way, at least part of the delay of the two procedures is transparent to the higher layers and the total delay due to the sequential executions is reduced. The other significant merit is that our effort for delay reduction is focused at network edges, where sophisticated computation is affordable and a large amount of electronic buffer is available. This line of thought complies with the current trend that only simple and scalable control and management should be done in the high-speed WDM layer, while most of the intelligence of the network, such as traffic engineering and QoS provisioning, is implemented at the IP layer [11], [14].

In addition, we emphasize that to deliver considerable performance improvement in terms of latency reduction without deleterious system cost, the forward transmission of the BHP should be adopted in tandem with the aggressive reservation strategy (as will be investigated in Section IV and Section V). A strategy discussed in [15] also involves an early release of a signaling message, yet with different transmission principle, enabling technologies, and performance concerns.

IV. FRR BASIC FEATURES

Besides the pretransmission of a BHP, the proposed FRR scheme also features a LPF-based burst-length prediction, an aggressive resource reservation, and the ease for QoS provisioning.

A. Burst-Length Prediction

The FRR scheme benefits from the parallel execution of the BHP signaling and the burst assembly. The forward transmission of a BHP requires *a priori* knowledge of the burst length. We propose to make this possible by a prediction-based method. There is no doubt that to make a prediction algorithm practical for an OBS system, it should not only deliver good esti-

mation performance, but also be simple and fast. Norros [16] proposed to find the weight factor function of the predictive filter based on integration of a weakly singular integral equation. This method is proved to be accurate and effective, especially for the self-similar traffic. However, the calculation is done off-line, and it requires a significant computational complexity, that limits its application for high-speed networks. In our system, we employ an N-order LPF which works as follows [17].

Let $L_d(k)$ be the length (in the time scale) of the k th burst. The length of the next incoming burst is then predicted according to those of the previous N bursts by

$$\tilde{L}_d(k+1) = \sum_{i=1}^{N} h(i) \cdot L_d(k-i+1) \tag{1}$$

where $h(i), i \in \{1, \ldots, N\}$ are the coefficients of the predictive filter.

There are a variety of options to obtain the predictive filter coefficients. For our system, we examine two approaches. One is based on the Yule–Walker method, whereby the predictive filter coefficients can be expressed as $\mathbf{R}\mathbf{h} = \mathbf{r}$, where $\mathbf{R}$ and $\mathbf{r}$ are the autocorrelation matrix and the autocorrelation vector of the data burst lengths, respectively, and $\mathbf{h}$ is the coefficient vector [18].

An alternative is the N-order normalized LMS-based recursive LPF. We update the predictive filter coefficients by an efficient algorithm [17], where the coefficients for the $(k + 1)$th prediction are defined as $(\mathbf{h})^{k+1} = (\mathbf{h})^k + ((\mu \cdot e(k) \cdot (\mathbf{L_d})^k)/\|(\mathbf{L_d})^k\|^2)$, with μ being an adjustable parameter of the LPF, $e(k)$ the residual between the actual and the predicted length of the k th data burst and $(\mathbf{L_d})^k$ the vector of $L_d(j)$ ($j \in \{(k-1) \cdot N + 1, \ldots, k \cdot N\}$).

We verify by simulations (the results of which are partly reported in Section V) that the LMS-based method is more appropriate, within the context of burst-length prediction, to forecast the length of the next data burst. The LMS-based method achieves satisfactory prediction performance without knowing the autocorrelation of the input traffic stream in advance and, thus, can be used as an on-line algorithm for bandwidth forecast. Meanwhile, the LMS-based approach outperforms the other alternative in terms of computational simplicity. Its time complexity for the coefficient calculation is $O(N)$, which is much less than that of Yule–Walker equations ($O(N^2)$).

Therefore, in this paper, we will present the experimental results only for the LMS-based LPF. The impact of the prediction order on the forecast performance is presented in Section V.

B. Aggressive Resource Reservation

In an FRR-enabled OBS system, resources in the intermediate nodes are reserved for an incoming data burst according to the predicted burst length. The forward-reserved length, denoted as $L_r(k + 1)$, if optimal, should be equal to the actual burst length. Due to the imperfection of a predictor, however, an estimated length may turn out to be smaller or larger than the actual burst duration. Suppose the reservation length is set to be equal to the predicted value, a smaller prediction of burst length $(e(k+1) = (L_d(k+1) - \tilde{L}_d(k+1)) > 0)$ will result in
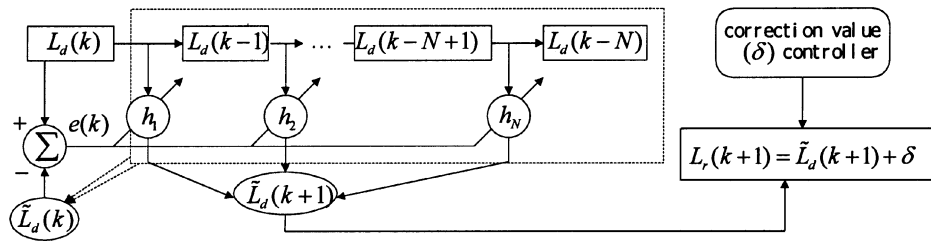
Fig. 3.   Prediction and aggressive reservation.

an insufficient reservation of the path holding time for the data burst. This requires the control header to be retransmitted after the burst assembly finishes, thus degrading the FRR latency reduction performance.

We compensate for this problem by an innovative aggressive resource reservation method. Instead of making $L_r(k + 1) = \tilde{L}_d(k + 1)$, we define the reservation length as $L_r(k + 1) = \tilde{L}_d(k+1)+\delta$, where δ is a small margin of correction. The value of δ has a significant impact on both the BHP pretransmission success probability—therefore the latency reduction capability of the FRR scheme and the system costs (e.g., the resource utilization and the signaling overhead). It should be carefully determined according to the tradeoff between these two performance metrics.

Conceptually, the probability that a BHP pretransmission succeeds is:

$$P_s = P\left(e(k) < \delta\right) = \int_{-\infty}^{\delta} f\left(e(k)\right) de(k) \tag{2}$$

where $f(e(k))$ is the distribution of the prediction residuals $(e(k))$. If we assume $e(k)$ resembles white noise, i.e., zero-mean Gaussian distribution with variance equal to σ^2, we get

$$P_s = \frac{1}{\sqrt{2\pi} \cdot \sigma} \int_{-\infty}^{\delta} e^{-\frac{e^2(k)}{2\sigma^2}} de(k) = 1 - Q\left(\frac{\delta}{\sigma}\right) \tag{3}$$

where $Q(\cdot)$ is the Q-function [18]. Equation (3) shows that the BHP pretransmission success probability is a function of the ratio of δ over σ. While a variety of methods are possible to determine the values for δ, we propose to choose δ to be a multiple of σ (i.e., $\delta = \alpha \cdot \sigma$). The rationale is that, this way, the explicit control on P_s can be achieved by simply choosing α which satisfies

$$\alpha = \frac{\delta}{\sigma} = Q^{-1}(1 - P_s). \tag{4}$$

Meanwhile, the inherent property of the Q-function in tandem with (3) implies that a too large α (e.g., $\alpha \geq 4$) will have little contribution to the performance improvement. Furthermore, we will show in Section V that the bandwidth overhead caused by the aggressive reservation algorithm can also be expressed as a function of the ratio of δ over σ, implying that by choosing a proper value of α, we can explicitly bound the bandwidth wastage and achieve the optimal tradeoff between the performance gain and the bandwidth cost.

From an implementation standpoint, for our LPF-based FRR system, we choose δ to be a multiple of the root mean square (RMS) of the sample residuals of the LPF, i.e.,

$$\delta = \alpha \cdot \sqrt{\frac{\sum_{i=1}^{N} e^2(k - i + 1)}{N}} \tag{5}$$

where α is a real value and may be determined by the network management. Fig. 3 depicts the principle of our traffic prediction and aggressive reservation algorithms.

Our solution is designed to take full advantage of the abundant bandwidth available in the core network to fulfill some QoS functionalities. Aiming at increasing the successful BHP pretransmission probability, the aggressive reservation method enhances the FRR scheme with an improved latency reduction capability. One may argue that the introduction of a small margin of correction will result in bandwidth wastage. It is well known, however, that in the future core network, bandwidth is no longer a limiting factor, while latency will be the major challenge to overcome [1], [2]. As reported, only 2% to 5% of the deployed fibers in the USA, i.e., the potential bandwidth, are lit, carrying just 10% utilization [19], [20]. Meanwhile, we will demonstrate that our aggressive reservation algorithm delivers a significant performance gain with a very limited reservation cost (see Section V). Furthermore, by properly choosing the correction values, the aggressive reservation method can actually reduce both signaling retransmission probability and the system bandwidth wastage as compared with a reservation method without a correction value (i.e., a zero-correction reservation method).

C. FRR for QoS Provisioning

Following the previous discussion on the delay constraints of different traffic classes and the FRR transmission scheme, we discuss an FRR-based strategy to facilitate QoS functionality at the ingress nodes and will refer to it as the FRR-based QoS provisioning. The aim is to reduce the burst delay for the real-time traffic and to achieve flexible QoS differentiation for different classes of applications.

The FRR-enabled OBS system facilitates the QoS provisioning by assigning each individual class-i traffic two system parameters: the interval (τ_p^i) to control when to launch the BHP into the core network prior to the burst assembly completion, and the real value α^i [defined in (4)] to achieve controllable BHP pretransmission success probability. It is precisely the flexibility of τ_p^i and α^i that enable us to implement the scalable delay reduction and QoS isolation degree [11], [23] between classes.
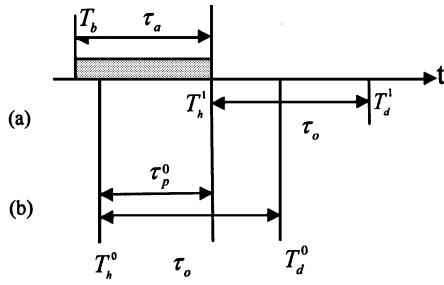
Fig. 4. FRR-based QoS strategy $(\tau_a > \tau_p^o)$. (a) Class-1 traffic (delay tolerant). (b) Class-0 traffic (delay sensitive).

TABLE II
NOTATIONS FOR THE FRR SCHEME PERFORMANCE

Term	Explanation
D_n	Average burst delay in an NFRR system
D_a	Average burst delay in an FRR system
D_f	Data burst delay when the FRR scheme fails
D_s	Data burst delay when the FRR scheme succeeds
P_s	The probability that the FRR scheme succeeds

Fig. 4 presents the discipline of our QoS strategy by illustrating the behaviors of BHPs belonging to two traffic classes (class-0: delay-sensitive; class-1: delay-tolerant) when $M = 2$. For simplicity, both classes are defined to have the same burst assembly time and offset time, denoted by τ_a and τ_o, respectively, and $\tau_p^1 = 0$. The advanced transmission of the class-0 data burst is achieved $(T_d^0 < T_d^1)$. Accordingly, the average delay that the time-critical traffic experiences at the ingress node can be decreased, taking into account of $\alpha^0 \geq \alpha^1$ (analyzed in Section V).

V. PERFORMANCE ANALYSIS AND SIMULATION RESULTS

The system performance is evaluated via theoretical analysis and simulation results. Performance metrics include: η—the latency reduction improvement of an FRR system, P_s—the BHP pretransmission success probability and γ—the bandwidth overhead. The referencing of traffic class will be omitted for notational simplicity when we conduct the performance evaluation for the individual traffic class to which the FRR scheme applies. We also investigate the prediction performance of an LMS-based LPF under a variety of traffic parameters and justify the predictability of the self-similar traffic. The order of the LPF is four, if not otherwise specified.

To focus on the effect of the FRR scheme on latency reduction, we do not consider the queuing delay due to the edge node scheduling. Table II summarizes the notations we will use in the analysis.

A. Latency Reduction Improvement

We first study the burst delay at network ingresses under the simple mode of the BHP signaling scheme (called NFRR for non-forward resource reservation) and that under the FRR alternative. Then, we analyze the latency improvement by the FRR scheme. The delay of a data burst is defined as the average delay of all the packets composed of this burst, due to the burst
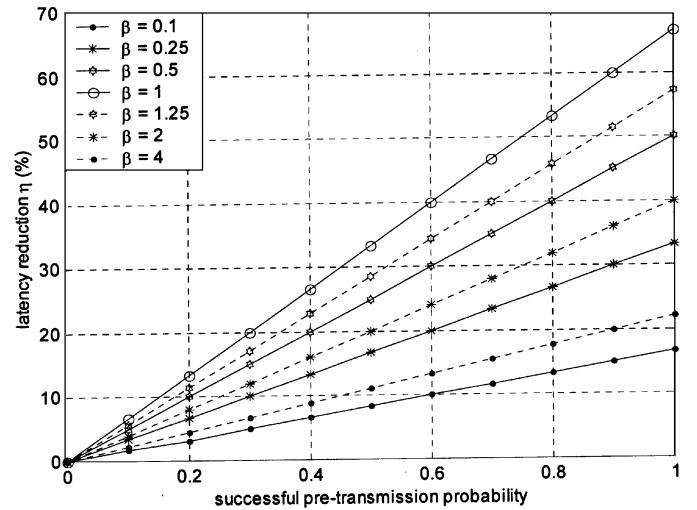

Fig. 5. Latency improvement versus BHP pretransmission success probability $(P_s)(\tau_p = \tau_a)$.

assembly and the basic offset time. Therefore, the burst delay caused by burst assembly is $1/2 \cdot \tau_a$.

1) Burst delay in an NFRR system.

In an NFRR transmission system, the burst delay at an ingress node is

$$D_n = \frac{1}{2} \cdot \tau_a + \tau_o. \tag{6}$$

2) Burst delay in the FRR system.

In the FRR-based system, the burst delay at an ingress node differs according to the success or failure of the pretransmission of a BHP. Suppose the forward resource reservation succeeds with a probability of P_s, the average burst delay of a class-0 burst is

$$D_a = P_s \cdot D_s + (1 - P_s) \cdot D_f$$
$$= \begin{cases} \frac{1}{2} \cdot \tau_a + \tau_o - \tau_o \cdot P_s, & \tau_p \geq \tau_o \\ \frac{1}{2} \cdot \tau_a + \tau_o - \tau_p \cdot P_s, & \tau_p < \tau_o. \end{cases} \tag{7}$$

We assume that β is a real value that represents the ratio of τ_o over τ_a, i.e., $\tau_o = \beta \cdot \tau_a$. The latency improvement (η) of the FRR scheme over the NFRR scheme can be expressed by

$$\eta = 1 - \frac{D_a}{D_n} = \begin{cases} \frac{2 \cdot \beta \cdot P_s}{1 + 2 \cdot \beta}, & \tau_p \geq \tau_o \\ \frac{2 \cdot \tau_p \cdot P_s}{(1 + 2 \cdot \beta) \cdot \tau_a}, & \tau_p < \tau_0. \end{cases} \tag{8}$$

Therefore, the system performance improvement η depends on three parameters: the ratio of τ_o over $\tau_a(\beta)$, the advanced period τ_p, and the probability that the forward reservation of a BHP succeeds (P_s). Fig. 5 presents the latency reduction percentage versus P_s, as β varies, when $\tau_p = \tau_a$. It shows that η increases as τ_o approaches τ_a and reaches its maximum gain when the ratio is one. Specifically, if the burst length can be predicted precisely such that the pretransmission of the BHP succeeds with a high probability $(P_s \rightarrow 100\%)$, our FRR scheme can reduce the edge node latency for the class-0 traffic by 66% when $\tau_o = \tau_a$.

B. BHP Pretransmission Success Probability

Equation (8) indicates that the probability of successful BHP pretransmissions (P_s) has a substantial impact on the latency
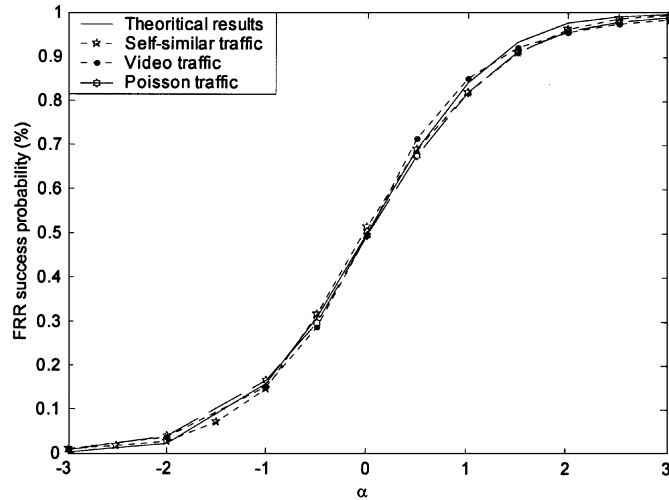
Fig. 6. BHP pretransmission success probability (P_s) versus α (where the aggressive reservation $\delta = \alpha \cdot \sigma$). The theoretical result is based on (3). The mean and variance of the self-similar traffic flow are 2 and 100 K, respectively.

improvement η. Since P_s depends largely on the difference between the prereserved duration and the actual burst length (in our system, $L_r(k) - L_d(k) = \delta - e(k)$), we study the effect of the correction margin (δ) on P_s.

Provided that the distribution of the prediction residuals of our LPF could be approximated by a zero-mean Gaussian function with variance equal to σ^2 (further justification of this assumption will be presented in the next section), we can obtain the P_s based on (3), i.e., $P_s = 1 - Q(\alpha)$.

We conducted a set of simulations tracing the cumulative density function (CDF) of the successful BHP pretransmissions under all three traffic scenarios, as shown in Fig. 6. It can be seen that for each input traffic scenario, the pretransmission of a BHP succeeds with a probability of about 50% in a system without aggressive reservation ($\alpha = 0$). However, the significant improvement on P_s is achieved with small values of correction (δ), and this performance gain slows down as α increases. For example, the BHP pretransmission succeeds with a probability of about 96% and 99%, if $\delta = 2 \cdot \sigma$ and $\delta = 3 \cdot \sigma$, at which point the latency improvement is about 64% and 66% ($\beta = 1$), respectively, as shown in Fig. 5. For comparison, we also plot the CDF of a standard Gaussian distribution. The experimental CDF curves match the theoretical curve very well, indicating that our white noise assumption for prediction residuals is valid.

C. Bandwidth Overhead

Our FRR strategy increases the BHP pretransmission success probability and improves the latency reduction performance for the delay-sensitive traffic by means of an aggressive bandwidth reservation. For the class-i traffic to which the FRR scheme applies, let γ^i represent the ratio of the average extra reservation length to the average actual burst length. γ^i can be referred to as the bandwidth overhead of this traffic class. Now, we consider the bandwidth overhead as a long-term system performance, and omit the index of the burst sequence number. This way, an advanced reservation length is simply denoted as L_r^i, which is equal to $\tilde{L}_d^i + \delta^i$, where $\tilde{L}_d^i$ is the estimated burst length and δ^i the correction margin. The actual burst length is referred to as

L_d^i. Let ε^i and ξ^i represent the difference between L_d^i and $\tilde{L}_d^i$, and that between L_r^i and L_d^i, respectively. Then, we have the relationships of $\varepsilon^i = L_d^i - \tilde{L}_d^i$, $\zeta^i = L_r^i - L_d^i$, and $L_r^i = \tilde{L}_d^i + \delta^i$.

The bandwidth overhead of our FRR scheme factors in both the successful and the unsuccessful pretransmission probabilities of a BHP. A BHP pretransmission succeeds when $\zeta^i > 0$, which implies $\varepsilon^i < \delta^i$. The average ε^i in this case, denoted as $\overline{\varepsilon}^i$, is given by

$$\overline{\varepsilon}^i = \int_{-\infty}^{\delta^i} \varepsilon^i \cdot f(\varepsilon^i) d\varepsilon^i \qquad (9)$$

where $f(\varepsilon^i)$ is the distribution function of ε^i.

The bandwidth overhead caused by a successful forward resource reservation is, thus

$$\gamma_s^i = \frac{\delta^i - \overline{\varepsilon}^i}{\tilde{L}_d^i + \overline{\varepsilon}^i} \cdot P_s^i. \qquad (10)$$

Meanwhile, the bandwidth overhead caused by an unsuccessful pretransmission of the BHP is 100%, i.e., $\gamma_f^i = P_f^i$.

Provided that the distribution of the residuals of our LPF is a zero-mean Gaussian function with variance σ^{i^2}, and that we have $\delta^i = \alpha^i \cdot \sigma^i$, the bandwidth overhead of class-i traffic can thus be expressed as (refer to [21] for a more detailed derivation)

$$\gamma^i = \gamma_s^i + \gamma_f^i = \frac{\alpha^i \cdot \sigma^i + \frac{\sigma^i}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{\alpha^{i^2}}{2}}}{\tilde{L}_d^i - \frac{\sigma^i}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{\alpha^{i^2}}{2}}} \cdot \left(1 - Q(\alpha^i)\right) + Q(\alpha^i). \qquad (11)$$

Of more interest is the system bandwidth overhead, i.e., the bandwidth overhead of the whole system where multiple traffic classes exist, defined as

$$\overline{\gamma} = \sum \gamma^i \cdot \rho^i$$
$$i = \text{the traffic class index} \qquad (12)$$

where ρ^i is the traffic load of class-i. For example, in a two-class QoS scenario where the FRR and the NFRR schemes are applied to the class-0 traffic and the class-1 traffic, respectively, suppose the traffic load distribution of the real-time traffic and the non-real-time traffic is 3:7, then the system bandwidth overhead is $\overline{\gamma} = 0.3 * \gamma^0$. Fig. 7 illustrates $\overline{\gamma}$ as a function of α^0. Both theoretical values (11) and simulation results are presented.

It is interesting to see that by properly choosing a small margin of correction in addition to the predicted burst length, our aggressive resource reservation-enhanced FRR system actually reduces the bandwidth overhead as compared to a system with a zero-correction reservation algorithm. Provided $\alpha^0 \in [0, 3.0]$ the upper bound of the bandwidth overhead corresponds to the one with $\alpha^0 = 0$. The reason is that the correction value, which is much smaller than the length of a data burst, dramatically increases the BHP forward signaling success probability, and reduces the wasted resource reservation due to insufficient burst-length prediction, which will otherwise contribute a greater bandwidth overhead. Correction values larger than some threshold (e.g., $\delta^0 \geq 2 \cdot \sigma^0$), however, result in a slightly higher pretransmission success probability at the cost of a larger system bandwidth overhead (see Figs. 6 and 7).
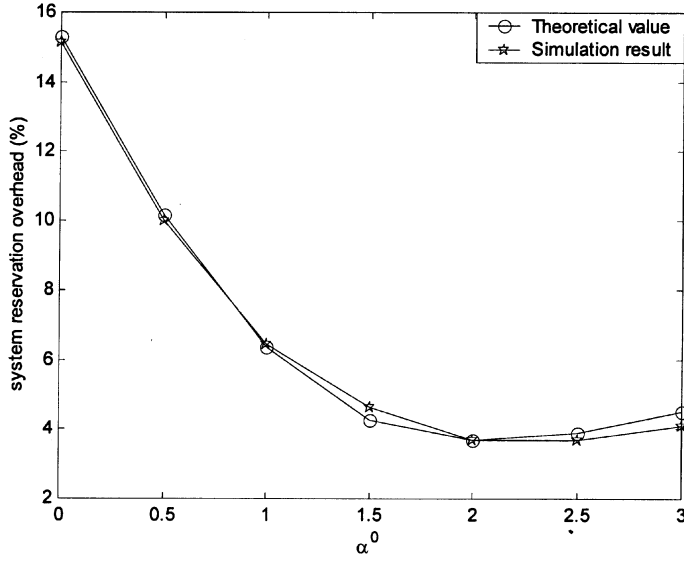
Fig. 7. System reservation overhead versus α^0 (where the aggressive reservation $\delta^0 = \alpha^0 \cdot \sigma^0$). The theoretical result is based on (13). The mean and variance of the traffic flow are 4 and 100 K, respectively. $H = 0.8$.
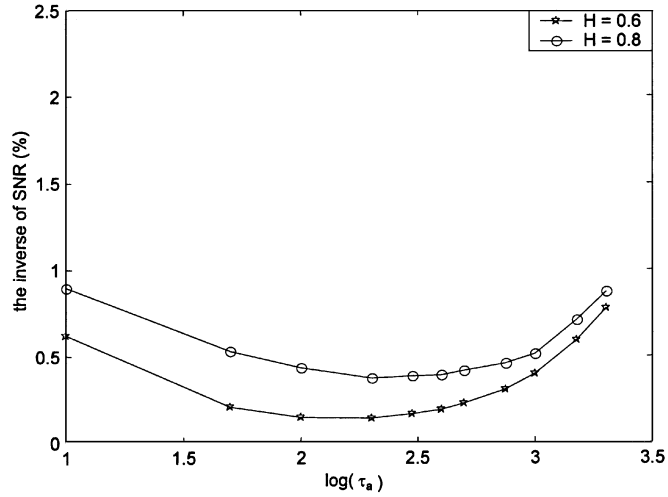

Fig. 9. SNR^{-1} versus Hurst parameter H. M and V represent the mean and variance of the input traffic flow, respectively.

Fig. 8. SNR^{-1} versus burst assembly time τ_a. The mean and variance of the traffic flow are 2 and 100 K, respectively.

Fig. 10. SNR^{-1} versus traffic load ρ. The input traffic is generated from 1024 ON-OFF sources. $H = 0.8$.

Our FRR scheme gains a significant latency reduction at the cost of a very small system bandwidth overhead, as can be seen from Figs. 5–7, which reinforce our aforementioned conclusion that the FRR scheme should be applied in tandem with the aggressive reservation algorithm to achieve satisfactory performance figures of merits with minor operation overhead.

D. LPF Performance and Traffic Predictability

The accuracy of an LMS-based LPF is assessed by two parameters: $\mathrm{SNR}^{-1} = ((\sum e^2(k))/(\sum L^2(k)))$ which is the inverse of the signal-to-noise ratio (SNR) and the autocorrelation of the residuals after the forecast. Special attention has been paid to the self-similar traffic scenario generated from the FFT-FGN model [22], if not otherwise specified.

The first set of simulations are conducted by tracing the dependence of SNR^{-1} on the parameters of burst assembly duration τ_a, Hurst parameter H (the traffic bursty degree), and traffic load ρ, r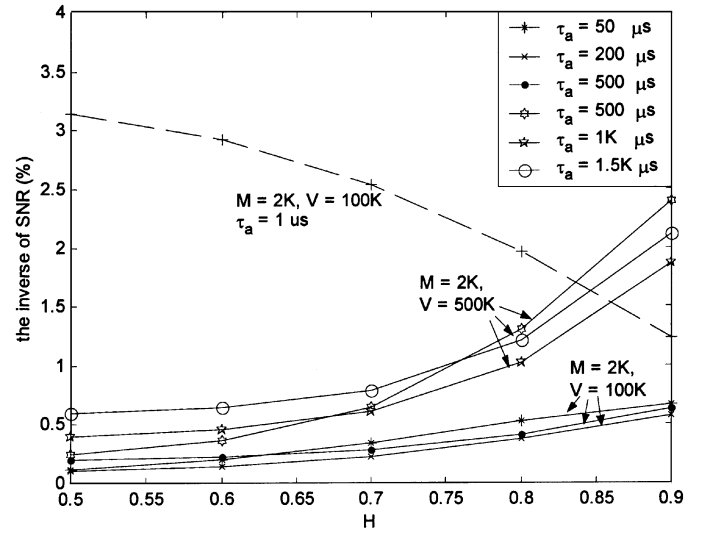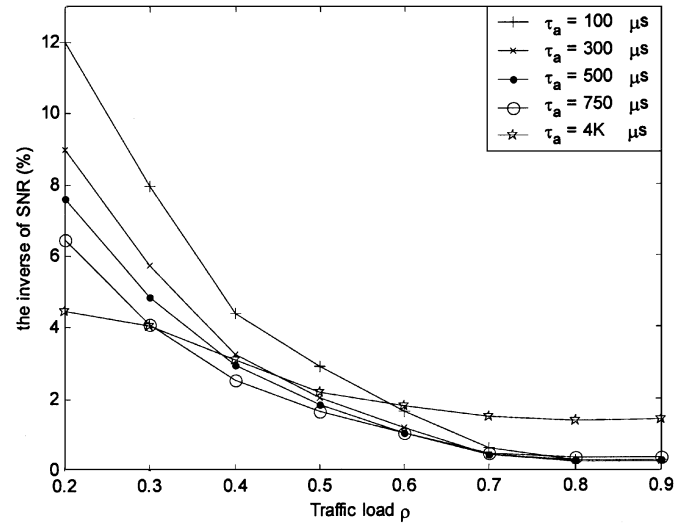espectively (Figs. 8–10). The performance of an LPF is influenced by all the three variables. Fig. 8 shows the effect of the burst assembly duration. While smaller SNR^{-1} values are achieved when the burst assembly time (τ_a) is between 100–1000 μs, a shorter or longer assembly time results in worse performance (i.e., larger SNR^{-1}). Meanwhile, it shows that the optimal τ_a, i.e., the burstification interval that delivers smaller SNR^{-1}, shifts as the H value varies. This can also be seen from Fig. 9, which shows that the prediction filter performance degrades slightly as the H value becomes larger. However, the LMS-based LPF presents acceptable prediction throughout the range from $H = 0.5$ to $H = 0.9$. For example, given that the mean value of the input traffic flow is 2000 bytes/μs and variance 10^5, when the burst assembly time is 200 μs, the SNR^{-1} is 0.22% and 0.37% for H of 0.7 and 0.8, respectively. Note that the burstification interval changes the performance of an LPF on the self-similar traffic. For $\tau_a = 1 \mu s$ (i.e., no further assembly on the input trace), the prediction performance is improved as H gets larger. This phenomenon is consistent with the conclusion given in [16] and [24]. However,

TABLE III
SNR^{-1} OF LMS-BASED AND YULE–WALKER-BASED LPF FOR SELF-SIMILAR TRAFFIC (%). THE MEAN AND VARIANCE
OF THE TRAFFIC FLOW ARE 2 AND 100 K, RESPECTIVELY. THE BURST ASSEMBLY TIME IS 100 μs

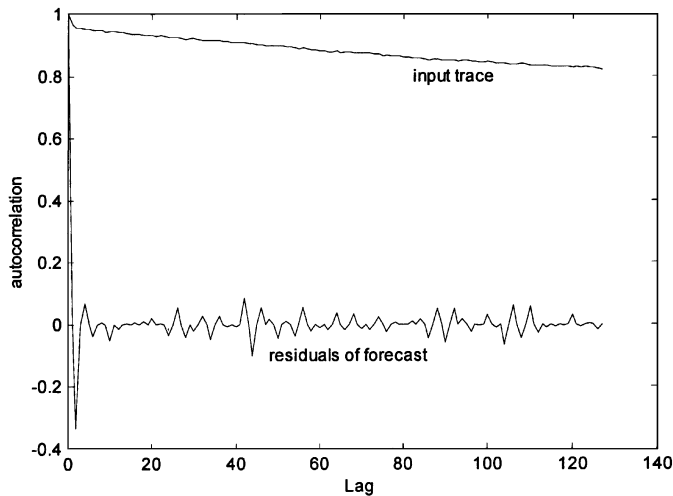|	$H = 0.6$		$H = 0.7$		$H = 0.8$	
	LMS	Yule-Walker	LMS	Yule-Walker	LMS	Yule-Walker
Order $= 4$	0.147709	8.373249	0.258466	8.541315	0.434525	8.849212
Order $= 8$	0.147787	3.657576	0.258708	3.827673	0.435030	4.167671
Order $= 12$	0.147822	2.272794	0.258836	2.433069	0.435378	2.764972
Order $= 16$	0.147844	1.725180	0.258923	1.894873	0.435647	2.252285
Order $= 24$	0.147866	1.507025	0.259010	1.676435	0.435923	2.046166

Fig. 11. Autocorrelation of the input traffic flow and the residuals of forecast under an LMS-based LPF. $H = 0.8$.

the effect of H on the prediction performance diminishes as the burst assembly interval grows. The traffic load also has substantial effect on the prediction performance (Fig. 10). Given the same traffic bursty degree and burst assembly time, the performance of an LMS-based LPF increases dramatically as the traffic load increases.

The effect of the order of the prediction filter on the forecasting performance is shown in Table III. It appears that the LMS-based LPF and the Yule–Walker-based[1] one behave differently as the order changes. For a self-similar traffic scenario, the LMS-based predictive filter emphasizes the most recent data burst lengths, i.e., an LPF can be made smaller to reduce the computation overhead.

Fig. 11 shows the autocorrelation of the input traffic and the prediction errors. Although the input trace presents the long-range dependence, the residuals of the LMS-based forecast resemble white noise.

The simulation results on both performance metrics imply the following conclusions. First, the LMS approach can deliver satisfactory prediction for the self-similar traffic. The length of the next incoming burst can be forecasted very well. Since the real Internet traffic can be best modeled by self-similar processes, our conclusion strongly verifies the viability of our LPF-based FRR mechanism. Second, the residuals of the LMS-based forecast are approximately Gaussian distributed, justifying our pre-

---

[1]To achieve the real-time forecast for the data burst length, our prediction is blind in that we estimate $\mathbf{R}$ and $\mathbf{r}$ in the Yule–Walker equation based on the length of the last N data bursts, where N is the order of the LPF.

vious derivations that are based on the white noise assumption. Third, in an LPF-based FRR system, a dynamic burst assembly interval is important to process the real-time traffic. The burst assembly time should be determined on-line, adaptive to the statistics derived from the previous traffic streams, i.e., $\tau_a = g(\rho, H)$. Meanwhile, we also propose that with the FRR mechanism, a burst assembly time should be no less than the burst offset time τ_o. The argument is that even though a burst assembly finishes earlier than the expiration of its offset time, the burst should wait for the end of the offset time and then be sent into the core network afterward. Algorithms to determine the optimal burst assembly duration τ_a, combining other constraints such as the number of data channels and control channels, are critically important and need further investigation.

VI. CONCLUSION

In this paper, a novel FRR scheme has been proposed and proved to be practical in reducing the data burst delay at network ingresses of an OBS system. The FRR scheme consists of three inherent features: a parallel execution of BHP signaling and burstification, an LMS-based LPF for burst-length prediction and an aggressive resource reservation. The FRR scheme has also been extended to facilitate QoS differentiation at network edges.

Theoretical analysis and simulations exhibit encouraging results. Our FRR mechanism leads to a significant latency reduction based on simple algorithms and mature techniques. QoS differentiation is facilitated at network edges. The aggressive reservation algorithm proves to be effective in increasing the successful pretransmission probability of a BHP. Furthermore, we have shown that the FRR scheme in tandem with this reservation algorithm results in less signaling retransmissions and bandwidth overhead as compared to a zero-correction system. The LMS-based LPF delivers excellent forecasting performance for the self-similar traffic which best models the Internet traffic. Optimal performance of the LPF has been found to depend on a variety of traffic parameters, including the traffic load, self-similar degree, and prediction interval. Such dependence on prediction interval implies the importance to devise algorithms that dynamically determine the burstification duration.

Several issues remain open and are worthy of further investigation to optimize our FRR scheme and to unleash the potential of the FRR-enabled OBS system. For example, the introduction of the aggressive reservation algorithm implies the importance to budget the correction value and to balance between the performance gain (in terms of the latency reduction

and QoS differentiation) and the operation cost (in terms of the resource utilization and the signaling overhead). The burst blocking probability and network resource utilization should be evaluated based on a variety of network topologies. Another issue is the optimization of the burst-length prediction algorithms and the comparison between the performance of an LMS-based LPF and those of other predictive filters under different traffic scenarios. These issues are the focus of our future research.

REFERENCES

[1] D. Farber, "Predicting the unpredictable—the impact of all-optical networking," in *Proc. Keynote Speech IEEE LEOS Summer Topic Workshop on Bandwidth Networking*, Monterey, CA, July 1998.

[2] ——, "Communications technology and its impact between now and 2010," *Commun. ACM*, vol. 40, no. 2, Feb. 1997.

[3] J. S. Turner, "WDM burst switching for petabit data networks," in *Proc. Optical Fiber Communications Conf.*, vol. 2, 2000, pp. 47–49.

[4] Y. Xiong, M. Vandenhoute, and H. C. Cankaya, "Control architecture in optical burst-switched WDM networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1838–1851, Oct. 2000.

[5] J. Y. Wei and R. I. McFarland Jr., "Just-in-time signaling for WDM optical burst switching networks," *IEEE J. Lightwave Technol.*, vol. 18, pp. 2091–2037, Dec. 2000.

[6] I. Baldine, G. N. Rouskas, H. G. Perros, and D. Stevenson, "JumpStart: a just-in-time signaling architecture for WDM burst-switched networks," *IEEE Commun. Mag.*, vol. 40, pp. 82–89, Feb. 2002.

[7] M. Yoo and C. Qiao, "A high speed protocol for bursty traffic in optical networks," in *Proc. 1997 Digest IEEE/LEOS Summer Topical Meeting*, 1997, pp. 26–27.

[8] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.

[9] V. Paxson, "Empirically derived analytic models of wide-area TCP connections," *IEEE/ACM Trans. Networking*, vol. 2, pp. 316–336, Aug. 1994.

[10] A. Ge, F. Callegati, and L. S. Tamil, "On optical burst switching and self-similar traffic," *IEEE Commun. Lett.*, vol. 4, pp. 98–100, Mar. 2000.

[11] M. Yoo, C. Qiao, and S. Dixit, "Optical burst switching for service differentiation in the next-generation optical Internet," *IEEE Commun. Mag.*, vol. 39, pp. 98–104, Feb. 2001.

[12] Extension of GSMP for Optical Burst Switching. IETF Document, J. K. Choi, M. H. Kang, and J. Y. Choi *et al.*. (2002, June). *http://search.ietf.org/internet-drafts/draft-ietf-ippm-loss-pattern-07.txt* [Online]

[13] S. Verma, H. Chaskar, and R. Ravikanth, "Optical burst switching: a viable solution for terabit IP backbone," *IEEE Network*, vol. 14, pp. 48–53, Nov./Dec. 2000.

[14] M. Ma and M. Hamdi, "Providing deterministic quality-of-service guarantees on WDM optical networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2072–2083, Oct. 2000.

[15] D. Morato, J. Aracil, and L. A. Diez *et al.*, "On linear prediction of Internet traffic for packet and burst switching networks," in *Proc. 10th Int. Conf. Computer Communications Networks*, 2001, pp. 138–143.

[16] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE J. Select. Areas Commun.*, vol. 3, pp. 953–962, Aug. 1995.

[17] J. R. Treichler, C. R. Johnson Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*. New York: Wiley, 1987.

[18] A.Alberto Leon-Garcia, *Probability and Random Processes for Electrical Engineering*. Reading, MA: Addison-Wesley, 1994, pp. 431–432.

[19] Optical Networking Faces Dark Future—Low Bandwidth Utilization is a Big Problem, R. Jaques. (2001, Aug.). [Online]. Available: http://www.vnunet.com/News/1125932

[20] Surviving the Carrier Commoditization Wave: The ADVENTIS EBITDA Turbocharger (2000, June). [Online]. Available: http://www.adventis.com/views/pdfs/surviving.pdf

[21] J. Liu and N. Ansari, "Forward resource reservation for QoS provisioning in OBS systems," in *Proc. IEEE GLOBECOM 2002*, Taipei, Taiwan, Nov. 17–21, 2002.

[22] V. Paxson, "Fast Approximation of Self-Similar Network Traffic," Lawrence Berkeley National Lab., Tech. Rep. LBL-36 750, Apr. 1995.

[23] M. Yoo, C. Qiao, and S. Dixit, "QoS performance of optical burst switching in IP-over-WDM networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2062–2071, Oct. 2000.

[24] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman Hall, 1998.

Jingxuan Liu (S'98) received the B.S. and M.S. (*summa cum laude*) degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 1996 and 1999, respectively. She is currently working toward the Ph.D. degree in computer science from the New Jersey Institute of Technology (NJIT), Newark.

Her research interests include queuing theory and its applications to computer communications, network architecture, and design, control, and performance analysis of IP over optical networks.

Miss Liu received second place in the IEEE North Jersey Section Student Competition Contest.

Nirwan Ansari (S78'–M'83–SM'94) received the B.S.E.E. (*summa cum laude*) degree from New Jersey Institute of Technology (NJIT), Newark, the M.S.E.E. degree from University of Michigan, Ann Arbor, and the Ph.D. degree from Purdue University, West Lafayette IN, in 1982, 1983, and 1988, respectively.

He joined the Department of Electrical and Computer Engineering, NJIT, in 1988, and has been a Professor since 1997. He coauthored (with E.S.H. Ho) *Computational Intelligence for Optimization* (Norwell, MA: Kluwer, 1997) and it was translated into Chinese in 2000. He coedited *Neural Networks in Telecommunications* (Norwell, MA: Kluwer, 1994). He is a Technical Editor of the IEEE *Communications Magazine*, as well as the *Journal of Computing and Information Technology*. His current research focuses on various aspects of high-speed networks and multimedia communications.

Dr. Ansari is the General Chair of ITRE2003 and was instrumental, while serving as its Chapter Chair, in rejuvenating the North Jersey Chapter of the IEEE Communications Society, which received the 1996 Chapter of the Year Award. He served as the Chair of the IEEE North Jersey Section and in the IEEE Region 1 Board of Governors from 2001 to 2002 and currently serves in various IEEE committees. He was the 1998 recipient of the NJIT Excellence Teaching Award in Graduate Instruction and a 1999 IEEE Region 1 Award.

Teunis J. Ott (M'81) received the B.Sc. degree in mathematics, physics, and chemistry, in 1965, the Doctoral degree (*cum laude*) in mathematics in 1970 from the University of Amsterdam, North-Holland, The Netherlands, and the Ph.D. degree in operations research fom the University of Rochester, Rochester, NY, in 1974.

He has been a Professor in the Department of Computer Science, New Jersey Institute of Technology (NJIT), Newark, since September 2001. Before moving to NJIT, he spent 23 years in Bell Laboratories, Bellcore, and Telcordia as Member of Technical Staff, Supervisor, Director, and Senior Scientist. He performed operations research at the University of Rochester, Rochester, NY. He holds six patents on telephony routing and aspects of ATM and Internet protocol traffic. In the last few years, his research has focused on Internet routing and on modeling of transport protocol, in particular transmission control protocol (TCP) and aspects of ECN. His unpublished paper cited most frequently *The Stationary Behavior of Ideal TCP Congestion Avoidance* (with J. Kemperman and M. Mathis) August 1996 (see: http://web.njit.edu/~ott/papers/index.html). His published paper cited most frequently *The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm* (with M. Mathis, J. Semke, and J. Mahdavi) Computer Communications Review 27(3), July 1997. His work is often cited and available only on the web.