Generalized Framework and Analysis for Bandwidth Scheduling in GPONs and NGPONs—The K-out-of-N Approach

Tamal Das, Ashwin Gumaste, Akhil Lodha, Ashish Mathew, and Nasir Ghani, *Senior Member, IEEE*

***Abstract*—Dynamic Bandwidth Allocation (DBA) is an important problem for upstream transmission in Fiber-to-the-Home (FTTH) systems. We propose a generalized scheduling mechanism for bandwidth allocation with a view to dissolution of the paradox between efficiency (utilization) and dynamism. Our scheme is shown to work for both TDM PONs as well as hybrid TDM/WDM PONs and pure WDM PONs as well as Next Generation PONs (NGPONs). While conventional bandwidth scheduling schemes pose efficiency as well as fairness issues, our proposed algorithm overcomes these. Three extensions as part of our scheduling technique include: 1) a K-out-of-N scheme to increase efficiency, with a general choice of K being a performance driven parameter; 2) strategic scaling to promote dynamism and reduce bandwidth starvation; and 3) a valuation based strategy that is uniquely tailored to reflect different service requirements. A thorough stochastic analysis based on a Markov-model is presented to compute the network-wide parameters such as delay, optimality and throughput. A detailed simulation model measures the performance of our scheme for latency, dynamism, efficiency and blocking comparing the analytical results with other techniques for dynamic bandwidth allocation in PONs.**

***Index Terms*—Dynamic bandwidth allocation, NGPON.**

I. Introduction

FIBER-TO-THE-HOME/curb/premises (generalized as FTTx) is soon becoming the best way of creating a high-speed access network. Even countries and regions deploying 3G/4G and other wireless technologies, are in parallel also deploying FTTx as a way to provide coarse bandwidth pipes to end-users. FTTx deployments though slow, due to infrastructural costs have been steadily increasing in the past few years. End-users in homes, offices or premises are connected to a service provider owned Central Office (CO) through a fiber network. The network can be inherently passive—meaning no switching within the network (optical) transport layer [1], [3], [4] or as shown in some of the NGPON [31] initiatives, can have active components such as wavelength selectable switches. A typical FTTx topology is tree-shaped, with end-users connected to the arms of the tree (at the leaves) and the CO situated at its center. Communication from the CO (also called the OLT or Optical Line Terminal), to the end-users (also called ONUs or Optical Network Units) can be of broadcast nature in case of TDM PONs, or can be of unicast nature, in the case of WDM PONs, and is termed *downstream communication*. Conversely, *upstream communication* from each of the many ONUs to the OLT, is typically unicast. For bidirectional communication, an FTTx system uses wavelength diversity—downstream wavelength(s) and an upstream wavelength(s) for communication. The upstream communication is a challenge as several (typically 32–128) ONUs time-share the same upstream wavelength. The time-sharing mechanism in the upstream, requires the centralized OLT to control and allocate bandwidth to each of the ONUs. Allocation of bandwidth (time-slots) leads to complex bandwidth allocation considerations [5], [9], [11]–[17]. Bandwidth assignment schemes in the upstream have received significant attention in the past few years [7], [9], [10], [14], [18], [20], and protocols have been proposed that enable dynamic (on-demand) bandwidth assignment to ONUs [13], [15], [19]. While maximization of throughput (or efficiency) has been one of the benchmarks of these protocols, another aspect has been their ability to meet service guarantees of revenue-bearing services (like voice, IPTV, Video-on-demand, etc.). These service-guarantees often imply adhering to requirements of latency and jitter, with the former dominating the requirement—necessitating that the ONUs have access to bandwidth within the time-intervals specified by the provisioned service.

In a preliminary version of this work shown in [23], we had proposed a basic DBA method (Section II) based on bandwidth scheduling between ONUs. The objective of the method was threefold: (1) to provide efficiency, (2) to meet service guarantees and (3) to ensure a better likelihood for an ONU to get bandwidth access in a well-loaded system.

The proposed method—called K-out-of-N scheduling—is different from many of the works in the literature, in the sense that it provides a comprehensive and generalized framework for service provisioning over a scalable GPON and is also extendable to NGPON using a WDM+TDM overlay. In this paper, we analyze the method for stochastic parameters, computing delay at a node, delay that a packet experiences, the network efficiency as a result of solving the tradeoff between dynamism and efficiency.

We next discuss our system design in Section II. In Section III, timing diagrams and implementation aspects are showcased. Section IV delves into the detailed stochastic

Manuscript received March 25, 2011; revised June 24, 2011; accepted June 27, 2011. Date of publication July 12, 2011; date of current version September 07, 2011. This work is supported in part by the Ministry of Communications and Information Technology, India and Nokia Siemens Networks, Germany.

T. Das, A. Gumaste, A. Lodha, and A. Mathew are with the Department of Computer Science and Engineering, Indian Institute of Technology, Bombay 400 076, India (e-mail: ashwing@ieee.org).

N. Ghari is with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2011.2161458

analysis, whereas Section V computes the optimality parameters in addition to using the stochastic results to obtain efficiency. In Section VI, we showcase our simulation model that extensively analyzes our scheme while also comparing simulation results to the analytical model.

II. System Model

We now consider the system model that would be used for our proposed algorithm. As shall be seen, the model is extendable from TDM-PON to WDM/NGPON without loss of generality.

A. TDM-PON

In the case of TDM PON, we allocate a single downstream wavelength that is broadcast to all the ONUs from the OLT. A second wavelength is used for upstream communication, whereby the N ONUs share the bandwidth using time-slotted access. In the K-out-of-N scheme, bandwidth allocation is done every K-time-slots, with all the N ONUs requesting upstream access. The OLT runs a DBA algorithm, and grants access to the top-K of the N requesting ONUs. The advantage of this approach over traditional DBA schemes is the reduction in the number of times that the requests need to be sent to the OLT. This results in better utilization of the upstream bandwidth.

B. WDM-PON

To address WDM-PON, we consider the following 3 scenarios: (i) each ONU is allotted a dedicated upstream wavelength, (ii) K wavelengths are shared across N ONUs for the upstream communication, and (iii) a single wavelength shared across all the N ONUs for upstream communication.

Downstream communication in each of the above mentioned cases is achieved through dedicated channels for each ONU. In case (i) above, burst-mode receivers (at the OLT) are limited in number (to K) for the upstream communication. It is important to conserve OLT resources and hence by allocating K burst-mode receivers, whereby $K \ll N$ we are able to conserve OLT resources. For example by keeping K to 20, in a 1000 user (as defined by FSAN for NGPONs) network, the OLT becomes significantly manageable. Hence, during each time-slot, only K of the N ONUs can transmit simultaneously in the upstream direction. Similarly, in case (ii), for the upstream communication, only K of the N ONUs can transmit simultaneously over the K unique wavelengths. Finally, in case (iii), the single upstream wavelength is time-shared between the N ONUs (TDM).

Thus, in each of the above mentioned scenarios, if we can choose K of the N ONUs for upstream communication in an efficient and fair manner, then we can establish the upstream communication by allotting the K time-slots/wavelengths to these K ONUs at a given time. It is important to note the interchangeability between time-slots and wavelengths, and that our scheme supports both, thereby paving a seamless mechanism for both TDM and NGPONs.

C. System Model

Further to the discussion in Section II.B, we now define our system to be able to choose K out of N ONUs in each *round* (defined in Section III). A list of the important conventions used is defined in Table I in addition to the same being reproduced below.

TABLE I
Important Nomenclature Used in the Paper

N	Number of ONUs in the network
$S = \{S_1, S_2, \ldots, S_h, \ldots\}$	universal set of services
T_S	duration of each time-slot
N_i	i^{th} node in the network
B_{hi}	Input buffers of service type S_h at node N_i
t_o	Generic time instant
λ_{hi}	arrival rate of service requests of type S_h at node N_i
$B_{hi}(t_o)$	size of the buffer consisting of requests of type S_h at node N_i at time t_o
$B_i(t_o)$	$= \sum_{h=1}^{\lvert S\rvert} B_{hi}(t_o)$, viz., the cumulative buffer size at node N_i
$Buff_{max}$	Buffer Size
λ_i	$= \sum_{h=1}^{\lvert S\rvert} \lambda_{hi}$ viz., effective arrival rate at node N_i
Δ_h	maximum tolerable delay of service type S_h
δ_{ji}	time interval within which the packet corresponding to the j^{th} service request must be serviced after its arrival at N_i
H_{ji}	absolute arrival time of the j^{th} packet at N_i
x_{ji}	time elapsed since the arrival of the j^{th} service packet at N_i.
K	An integer $\leq N$
T_g	Guard band time
$r_i(t)$	$= \frac{B_i(t)}{Buff_{max}}$, denotes the buffer occupancy
$P_i(t)$	time elapsed since N_i was last allocated bandwidth
X_{ih}	Amount of time elapsed since the first packet of service type S_h entered the buffer at ONU N_i and awaiting transmission
$Q_i(t)$	$= \min\{\Delta_h - X_{ih} \mid h = 1,2,\ldots\}$; denotes the maximum amount of time that ONU N_i can wait before it must be serviced, else a packet in its buffer will be timed-out.
$a_i(t)$	$= \frac{1}{1+Q_i(t)/P_i(t)}$, viz., the service desperation exhibited by N_i at time t
$val_i(t)$	*valuation* that N_i sends to the OLT
$\eta_{K,N}$	Efficiency of a system using the K-out-of-N mechanism, for a given value of K and N
$\overline{val}_i(t)$	Strategically scaled *valuation* that N_i sends to the OLT

An N-node time-slotted system with various service classes from the universal set of services as defined by $S = \{S_1, S_2, \ldots, S_h, \ldots\}$ is considered. We define T_S as the duration of each data time-slot. We define two-types of transmission slots—regular data-slots of duration 150 μs (defined as T_S) and smaller overhead-slots of duration 0.064 μs (T_b). Each node N_i in the network comprises of $|S|$ input buffers, represented by $B_{hi}, \forall h : S_h \in S$. We are interested in the system dynamics at a node N_i at time t_o (eventually generalized to a steady state). Each service arrival is characterized as a Poisson-arrival process. Let λ_{hi} be the arrival rate of service requests of type S_h and $B_{hi}(t_o)$ be the size of the buffer consisting of requests of type S_h at node N_i.

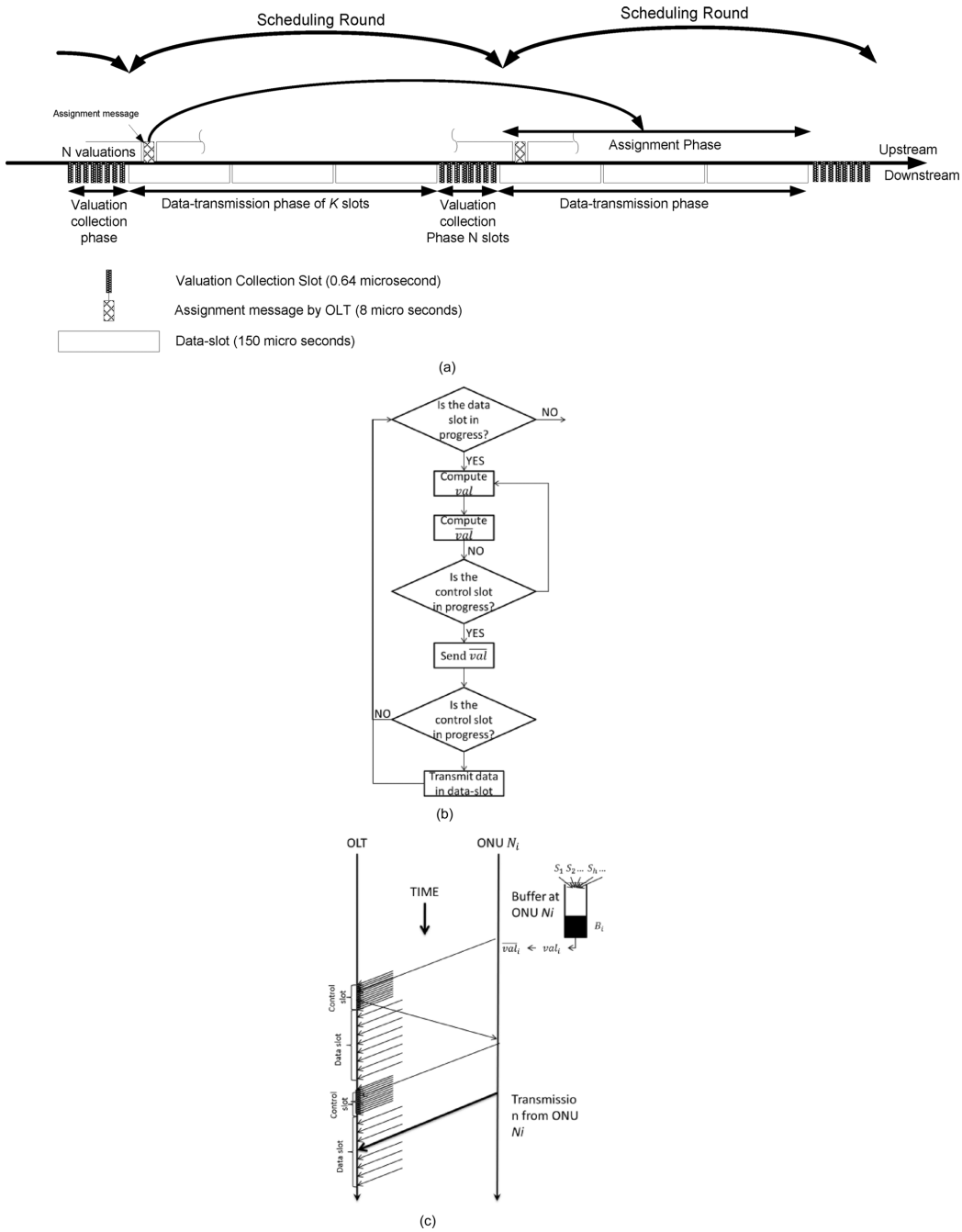

Fig. 1. (a) K-out-of-N protocol for $K = 3$ and $N = 8$. (b) Flowchart of the protocol at the ONU. (c) Time-wise description of the K-out-of-N protocol.

Let the buffer size (in terms of number of packets) at N_i be denoted as $B_i(t_o) = \sum_{h=1}^{|S|} B_{hi}(t_o)$. Further, let each of the N-ONUs have a buffer of maximum value $\mathrm{Buff}_{\max}$. Assuming independence, $\lambda_i = \sum_{h=1}^{|S|} \lambda_{hi}$ denotes the effective arrival rate at N_i.

Let $\{\Delta_h : h = 1, 2, \ldots\}$ denote the maximum tolerable delay of each service class, whereas $\{\delta_{ji} : j = 1, 2, \ldots\}$ denotes the time interval within which the packet corresponding to the jth service request must be serviced after its arrival at N_i. Let H_{ji} denote the absolute arrival time of the jth packet at N_i. Further, x_{ji} denotes the time elapsed since the arrival of the jth service packet at N_i.

For a node, we address the interval between its two consecutive wins as a *scheduling round*. Given the recursive behavior of a generic scheduling round the complexity involved in the stochastic computation is immense. We hence break the scheduling rounds into two parts—the first round and then based on the first round, the generic round, both of whose parameters are computed in Section IV. We assume that if a node has won a slot then at the end of this slot its buffer is empty, *except* for the requests that arrived within this slot.

III. The K-out-of-N Technique

To provision services and maximize efficiency we propose a K-out-of-N scheduling technique for bandwidth assignment in PONs. In this technique, time is divided into *scheduling rounds*, each of which has a data period and a control period (see Fig. 1(a)–(c)). In the data period, there are K-upstream

transmissions, followed by N control slots in which nodes send their requests as *valuations*. The N-ONUs send *valuations* in these N control slots to the OLT. The OLT acts as an arbiter, and selects the top K valuations. The corresponding K nodes (that sent the top-K valuations) transmit in the next K slots. *The novelty of this technique is in the engineering aspect of the scheduling process, computation of valuations, enabling service provisioning thus resulting in an efficient allocation algorithm* (Fig. 1(c)).

The first factor of consideration is that of service provisioning. Our contribution is in the *linear* transformation of bandwidth requests (both quantitatively and qualitatively) into numerical valuations that enable the network to allocate bandwidth in an efficient manner—maximizing network-wide utilization and meeting the needs of an ONU.

For an N-node network, assigning bandwidth based on the classical DBA schemes, is significantly inefficient, as the OLT has to receive N valuations and then allocate bandwidth to the ONU that sends the highest valuation. Here we propose the K-out-of-N method to increase the efficiency while maintaining service guarantees. The choice of K represents a trade-off—a higher value indicating a better utilized (efficient) system, while a lower value indicates lesser delay.

A second factor that we consider is that of fairness [6] with the notion of node-starvation, i.e., facilitating nodes with services that are lower quality to reserve bandwidth. We address this issue by extending the bandwidth scheduling method using a *strategic scaling* scheme. A node submits a *valuation* which is defined as its bandwidth request in addition to a *scaling* factor that increases the success of a starved node (either itself or someone else). The scaling factor *attempts* to smoothen the difference between the node's valuation and a *threshold* valuation—corresponding to the lowest winning node (that was awarded bandwidth in the previous iteration) amongst the K successful nodes.

Protocol Working: Communication in the upstream direction is done in *rounds*, where each round has two phases—a valuation-collection phase (control period) followed by a data-transmission phase (data period) (as shown in Fig. 1(a)–(b)). The OLT sends a request message just prior to the beginning of a valuation-collection phase, and in response, each ONU (in a pre-defined sequence, decided during ONU discovery phase) sends its valuation (in the valuation-collection phase) to the OLT. From the N-valuations that the OLT receives, it selects a group of K ONUs (corresponding to the K-highest valuations) and informs (all the nodes) of the selected ONUs (Fig. 1(c)). The selected ONUs transmit their data in time-slots within the data-transmission phase. For efficient performance, we assume that valuation-collection and data-transmission are interleaved (see Fig. 1(a)–(c)), as well as assume that the valuation-collection phase translates to data-transmission in the next round of communication (Fig. 1(a)). Valuation computation reflects the requirement of the node's bandwidth in terms of provisioning services with different arrival rates and different service parameters. The choice of K determines the efficiency of the protocol.

As a node is allotted a time-slot for transmission, its buffer is emptied based on the following scheme: If the valuation of the winning node was dominated by *buffer occupancy*, then its buffer is emptied in the decreasing order of its occupancies with respect to the different services, whereas, if the winning utility value was dominated by the *service criticality* at the node, the node's buffer is emptied according to an Earliest-Deadline-First (EDF) discipline [28].

We now discuss the three sub-methods for dynamic allocation of bandwidth in the upstream direction:

A. Scheduling Mechanism

To meet the critical requirements of services it is important that the valuation sent by an ONU to the OLT, be a *true* representation of the requirement of the bandwidth for that ONU. The valuations are computed based on the packets that arrive at the ONU from the end-users. For appropriate service guarantees, valuations must reflect two kinds of traffic parameters—the average rate at which the packets arrive (from clients/end-users) into the ONUs (buffer occupancies) and the maximum allowable latencies of packets (of different services) that can wait in the ONU buffer. While the former is called *buffer occupancy ratio* (fraction of the buffer occupied), the latter is termed as *service desperation* defined as a parameter that denotes how critical is it for an ONU to receive a transmission slot before its packets are dropped because its corresponding maximum tolerable delay limit was reached [2].

B. K-out-of-N Auctioning

The problem with generic scheduling mechanisms is the associated overhead i.e., for every data-slot in the upstream direction, each of the N ONUs must submit their requests (with each request being separated by a guard-band time T_g), resulting in poor efficiency. To increase the efficiency of our scheduling algorithm, we propose a modification—the K-out-of-N scheme—in which N nodes submit their valuations, and the OLT allocates bandwidth to K ($\leq N$) ONUs for transmission in the next K consecutive time-slots. By effectively choosing K, we argue that the efficiency is improved while the services are appropriately provisioned (within respective delay bounds). Note that as the value of K increases towards N, the latency suffers.

Buffer occupancy is denoted by

$$r_i(t) = \frac{B_i(t)}{\mathrm{Buff}_{\max}}. \tag{1}$$

Note that the buffer occupancy ratio is bounded in [0, 1].

Next, we compute the service-desperation component. We set a timer at ONU N_i, whose value $P_i(t)$ indicates the time elapsed since the ONU was last allocated bandwidth. Let X_{ih} represent the value of time elapsed since the first packet of service type S_h entered the buffer at ONU N_i and awaiting transmission. We now define delay-tolerance at a node as

$$Q_i(t) = \min\{\Delta_h - X_{ih} \mid h = 1, 2, \ldots\}. \tag{2}$$

Delay tolerance $Q_i(t)$ represents the maximum amount of time that ONU N_i can wait before it must be serviced, else a packet in its buffer will be timed-out. We compute ONU *service-desperation* as

$$a_i(t) = \frac{1}{1 + Q_i(t)/P_i(t).} \tag{3}$$

Note, that (3) is bounded in [0, 1]. From (1) and (3), we compute *the valuation* that an ONU sends to the OLT as

$$\mathrm{val}_i(t) = \max\{a_i(t), r_i(t)\}. \tag{4}$$

The value of $\mathrm{val}_i(t)$ is bounded in the interval [0, 1]. Of the two quantities, buffer occupancy ratio and service desperation, the one with a larger value is *passed* as the valuation. However, the method has two associated problems. First, for every data-slot, we would require N ONUs to transmit their valuation to the OLT resulting in an upstream efficiency of

$$\eta_{1,N} = T_S/(T_S + N(T_g + T_b)). \tag{5}$$

As the value of N increases, the efficiency decreases substantially, resulting in poor throughput. In contrast, a larger choice of T_S results in higher experienced latency.

Second, though the valuation mechanism enables nodes to truly reflect their bandwidth needs, the average delay between successive transmissions of ONUs is proportional to the service requirement. This means that ONUs with heavier flows of a particular service type will have a lesser average latency (for that service), than for ONUs with finer flows of the same provisioned service leading to issues of fairness.

The issue of efficiency is solved by the K-out-of-N scheme. For the chosen K nodes to transmit in the upstream, the data-transmission phase constitutes of K data-slots (with interleaved guard bands). The efficiency of such a system is therefore

$$\eta_{K,N} = KT_S/(K(T_S + T_g) + N(T_g + T_b)). \tag{6}$$

Remark: As K increases, the value of efficiency in (6) increases more than that in (5) for similar values of T_S.

Now consider the choice of K. While choosing a large value for K the efficiency of such a system would be high, while the corresponding *average* latency experienced is also high (more time between two consecutive transmissions for a particular node). Conversely, a smaller value of K means that the *average* latency experienced would be low, though the efficiency of the system would also be low. An optimal K analysis is presented in Section IV.

C. Strategic Scaling

Adhering to service provisioning (catering to delay bounds) and maximizing efficiency have an effect on the fairness in PONs. The effect is severe for nodes with lower quality services or lower arrival rates. The group of K can be viewed as an elite group whose membership can be privileged to selective bandwidth-hungry nodes, especially those with large (bursty) traffic needs. Also nodes with lower arrival rates may not be able to be a part of this group leading to bandwidth *starvation*. To avoid bandwidth starvation, we further enhance the K-out-of-N scheme by strategic scaling.

In the system, for any given round we have two groups of ONUs—K ONUs that have been allocated bandwidth, and $(N - K)$ ONUs that have not been allocated bandwidth (refer Fig. 2). The strategy is to *positively scale* those valuations of the ONUs, which were in the group of $(N - K)$ in the previous round, and *negatively scale* the valuations of the ONUs which were in the group of K in the previous round. The amount of

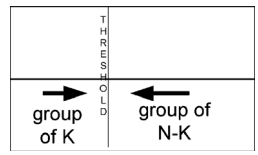

Fig. 2. Strategic Scaling.

scaling is determined by the *distance* of the ONU's valuation from the threshold (Kth maximum valuation) in the previous round. This implies that the node with the highest valuation (in the previous round) witnesses maximum negative scaling, while a node whose valuation was the lowest in the previous round, receives maximum (positive) benefit from this scaling. This circulatory behavior of nodes between the two groups act as a protection against bandwidth starvation resulting in nodes with lower bandwidth needs still being able to have lower latencies.

We introduce a scaling strategy that facilitates dynamism and also prevents bandwidth starvation. Namely, an ONU scales its valuation by a factor (that either enhances or lowers its valuation) depending on its past valuation history. As an example, an ONU that is part of the group of K in the tth round will scale its valuation in the $(t + 1)$th round by a lowering factor (i.e., multiply its valuation by $(1 - Z)$, where Z is a fraction). Likewise, an ONU in the group of $(N - K)$ in the tth round would enhance its valuation in the $(t + 1)$th round by multiplying its valuation by $(1 + Z)$.

The scaling of valuations based on their position *viz-a-viz* the group-of-K, is a strategy that the ONUs use to improve their expected average delay as well as make the system more dynamic (without compromising efficiency of the K-out-of-N technique).

For the $(t-1)$th round let us denote the lowest valuation in the group of K as $\mathrm{thresval}\ (t-1)$ which implies that all the ONUs corresponding to valuations lesser than $\mathrm{thresval}\ (t-1)$ were then in the group of $(N - K)$. We call $\mathrm{thresval}\ (t-1)$ as the *threshold*. In the tth scheduling round, an ONU N_i strategically scales its *true* valuation with the threshold in the previous round and this is given by:

$$\overline{\mathrm{val}}_i(t) = \begin{cases} \mathrm{val}_i(t), & \text{for } t = 1 \\ \mathrm{val}_i(t)[1 + \mathrm{thresval}\ (t-1) - \overline{\mathrm{val}}_i(t-1)], & \text{for } t > 1. \end{cases} \tag{7}$$

Through simulation and analysis, we see that the dynamism exhibited by the system results in an overall improvement of latency for ONUs.

IV. Stochastic Model for Scheduling Algorithm

A. Preamble

The *K-out-of-N* technique presented in the previous section solves the paradox between delay and efficiency. From an analytical perspective, our interest is to compute (1) the maximum efficiency achieved while meeting the permissible latency requirements for a particular value of K and (2) to compute the value of K that provides the maximum efficiency while subscribing to the service delay parameters. To this end, a stochastic model is now developed and presented.

The model first involves computing delay which is further dependent on the computation of the probability of success for a node, with the assumption of services (like voice, video, data, etc. along with their associated latency requirements). The computation of probability of success is complex as it involves a recursive (in time domain) computation of the probabilities of dependent random variables due to the reliance of the success probability on the past state of the system. Recall that the valuation a node sends in round t (from (7)) depends on: (1) its position (amongst the N nodes) in round $(t-1)$, and (2) the valuation of other nodes in the previous round, as well as in the current round. Once we compute the success probability, we are able to apply the same for delay computation. Our desire is to compute: (1) the delay that a node experiences, given the traffic profiles (arrival rates, latency needs) and (2) the delay that a packet experiences once it enters a node. The first delay computation is an extension of the probability of success of a node summed over the round time, while the second delay computation involves a more detailed analysis. We identify six mutually exclusive and exhaustive scenarios of a packet arriving at an ONU before being granted transmission rights. To this end, we begin with the computation of the probability of success of an ONU, given the system model in Section III.

B. Probability of Success of an ONU in the K-out-of-N Scheme

We analyze the algorithm by first considering the computation of probability of success for an ONU. This process is complex given the multiple dependencies of the associated random variables. Hence, this is simplified as: (1) the computation at the startup of the system and then (2) a generic state. For the generic state, a Markov model is developed whose steady state probabilities are of interest to us, leading to the computation of the success probabilities at an ONU. As an ONU wins a transmission slot, we define the interval between its two consecutive wins as a *scheduling round* for that ONU.

Note: The analysis in this paper is a generalization of our work in [30]. In particular, the bandwidth allocation algorithm in [30] is executed for one slot at a time. However, in this paper, we run the scheduling algorithm once every K slots. Hence, the analysis in [30] and this paper though similar for the first scheduling round, differ significantly for the generic round. Only for $K=1$, our analysis in this paper reduces to that of [30], however in practice, this is never the case and hence this analysis is unique to the *K-out-of-N* scheme.

Analysis of the First Scheduling Round: Referring to the system model in Section III, we define, $V_i(t_o) \triangleq \{j \,|\, H_{ji} + \delta_{ji} \geq t_o\}$ comprising of all valid requests at ONU N_i at time t_o. Thus

$$P_i(t) = \max_{j \in V_i(t_o)} \{x_{ji}\} \quad \text{and}$$
$$Q_i(t) = \max_{j \in V_i(t_o)} \{\delta_{ji} - x_{ji}\} \tag{8}$$

where $x_{ji} = t_o - H_{ji}$. Hence, the valuation of N_i for time-slot t_o reduces to the equation shown at the bottom of the page, where $\text{Buff}_{\max}$ is the buffer size. Amongst all the valid requests at N_i at time t_o, we define the earliest arrival time across all the services as $A_i(t) \triangleq \min\{H_{ji} \,|\, j \in V_i(t)\}$. In addition, the earliest instant that a packet is timed-out is $D_i(t) \triangleq \min\{H_{ji} + \delta_{ji} \,|\, j \in V_i(t)\}$. We require the joint distribution function of $P_i(t), Q_i(t)$, and $B_i(t)$ to compute success probability.

Using (8) and rearranging, we get

$$\begin{aligned} &P(P_i(t_o) < \alpha, Q_i(t_o) < \beta, B_i(t_o) < \gamma \,|\, V_i(t_o) \neq \phi) \\ &\quad = P(A_i(t_o) > t_o - \alpha, B_i(t_o) < \gamma \,|\, V_i(t_o) \neq \phi) \\ &\qquad - P(A_i(t_o) > t_o - \alpha, D_i(t_o) \\ &\qquad\quad > t_o + \beta, B_i(t_o) < \gamma \,|\, V_i(t_o) = \phi). \end{aligned} \tag{9}$$

First, consider the second term of (9). From the definition of $A_i(t)$ and $D_i(t)$, we state (10), shown at the bottom of the page.

The arrival time of the jth service request at N_i viz., H_{ji} varies in the interval $[t_o - \Delta_h, t_o]$, given that the request belongs to the service class S_h. We divide this interval into two non-overlapping sub-intervals, based on (10), corresponding to the jth service request arrival at N_i that belongs to S_h as follows. This is also illustrated in Fig. 3.

$$\begin{aligned} \mu_{ji} &= \left[\max\{0, t_o - \Delta_h\}, \right. \\ &\qquad \left. \min\left\{t_o, \max\left\{\begin{matrix} t_o + \beta - \delta_{ji}, \\ t_o - \alpha, t_o - \delta_{ji} \end{matrix}\right\}\right\}\right] \\ \theta_{ji} &= [\min\{t_o, \max\{t_o - \alpha, t_o + \beta - \delta_{ji}, t_o - \delta_{ji}\}\}, t_o]. \end{aligned}$$

$$\begin{aligned} \text{val}_i(t_o) &= \max\left\{\frac{B_i(t_o)/|S|}{\text{Buff}_{\max}}, \frac{\max_{j \in V_i(t_o)}\{x_{ji}\}}{\max_{j \in V_i(t_o)}\{x_{ji}\} + \max_{j \in V_i(t_o)}\{x_{ji}\}}\right\} \\ &= \max\left\{\frac{B_i(t_o)/|S|}{\text{Buff}_{\max}}, \frac{t_o - \max_{j \in V_i(t_o)}\{H_{ji}\}}{\max_{j \in V_i(t_o)}\{H_{ji} + \Delta_i\} - \min_{j \in V_i(t_o)}\{H_{ji}\}}\right\} \end{aligned}$$

$$\begin{aligned} &P(A_i(t_o) > t_o - \alpha, D_i(t_o) > t_o + \beta, B_i(t_o) < \gamma \,|\, V_i(t_o) \neq \phi) \\ &\qquad = P\left(\begin{matrix} H_{ji} > \max\{t_o - \alpha, t_o + \beta - \delta_{ji}, t_o - \delta_{ji}\}, \\ B_i(t_o) < \gamma \end{matrix} \,\middle|\, V_i(t_o) \neq \phi\right) \end{aligned} \tag{10}$$

Fig. 3. Illustrating definition of μ_{ji} and θ_{ji}.

Note that for all values of α and β, we get valid μ_{ji} and θ_{ji} intervals. Let us also consider the following events at N_i:

$$G_i(t_o) \triangleq \text{No request} \in V_i(t_o) \text{ arrived at } N_i \text{ in } \mu_{ji}$$
$$L_i(t_o) \triangleq \text{No request } \in V_i(t_o) \text{ arrived at } N_i \text{ in } \theta_{ji}.$$

The jth service request at N_i will not time-out till t_o, if it arrives in either one of these intervals: μ_{ji} or θ_{ji}. Hence, any request that arrived in either of these intervals *will* belong to $V_i(t_o)$. Thus, G_i and L_i simplifies to

$$G_i(t_o) = \text{No request arrived at } N_i \text{ in } \mu_{ji}$$
$$L_i(t_o) = \text{No request arrived at } N_i \text{ in } \theta_{ji}.$$

We define the event $M_i(t_o) \triangleq \{B_i(t_o) < \gamma\}$. Hence

$$G_i(t_o) \cap \bar{L}_i(t_o) \cap M_i(t_o) = \left\{ \begin{array}{c} A_i(t_o) > t_o - \alpha, D_i(t_o) > t_o + \beta, \\ B_i(t_o) < \gamma, \quad V_i(t_o) \neq \phi \end{array} \right\}.$$

Also $P(G_i(t_o) \cap \bar{L}_i(t_o) \cap M_i(t_o)) = P(G_i(t_o) \cap M_i(t_o)) - P(G_i(t_o) \cap L_i(t_o) \cap M_i(t_o))$. We now solve for the individual terms of this expression as follows:

$$\begin{aligned} &P(G_i(t_o) \cap M_i(t_o)) \\ &= P((\text{No request arrivals at } N_i \text{ in } \mu_{ji}) \cap (B_i(t_o) < \gamma)) \\ &= P(B_i(t_o) < \gamma \text{ with arrivals limited to } \theta_{ji}) \\ &= P\left(\cup_{n=0}^{\lfloor \gamma \rfloor} \cup_{Z_n} \cap_{h=1}^{|S|} B_{hi}(t_o) = n_h \right. \\ &\quad \left. \text{with arrivals only in } \theta_{ii}\right). \end{aligned}$$

where

$$Z_n \triangleq \left\{ (n_1, \ldots, n_{|S|}) \;\middle|\; \sum_{h=1}^{|S|} n_h = n \right\}$$
$$= \sum_{n=0}^{\lfloor \gamma \rfloor} \sum_{Z_n} \prod_{h=1}^{|S|} \frac{e^{-j_{hi}(|\mu_{ji}|+|\theta_{ji}|)} (\lambda_{hi}|\theta_{ji}|)^{n_h}}{n_h!}. \quad (11)$$

where $|\cdot|$ denotes the cardinality operator, i.e., $|\mu_{hi}| + |\theta_{hi} = \min\{t_o, \Delta_h\}$ and $|\theta_{ji}| = \max\{0, \min\{\alpha, \delta_{ji} - \beta, \delta_{ji}\}\}$. Further,

$$\begin{aligned} &P(G_i(t_o) \cap L_i(t_o) \cap M_i(t_o)) \\ &= P\left(\cap_{h=1}^{|S|} \text{No request of type } S_h \text{ arrived in } \mu_{ji} \cup \theta_{ji}\right) \\ &= \exp\left(-\sum_{h=1}^{|S|} \lambda_{hi} \times \min\{t_o, \Delta_h\}\right) \end{aligned} \quad (12)$$

Using (11) and (12), we solve for the second term in (9), shown in (13) at the bottom of the page.

Similarly, we solve for the first term in (9) as shown in (14) at the bottom of the page, where $\mu'_{ji} = [\max(0, t_o - \Delta_h), \min(t_o, \max\{t_o - \alpha, t_o - \delta_{ji}\})]$, $\theta'_{ji} = [\min(t_o, \max\{t_o - \alpha, t_o - \delta_{ji}\}), t_o]$ and thus $|\mu'_{ji}| + |\theta'_{ji}| = \min\{t_o, \Delta_h\}$, $|\theta'_{ji}| = \max\{0, \min\{\alpha, \delta_{ji}\}\}$. Further, substituting (13)–(14) in (9), we get (15), shown at the bottom of the next page.

Probability Distribution Function (pdf) of Valuation at an ONU: Using (15), we compute the pdf of the valuation of N_i as

$$\begin{aligned} F_{\text{val}_i(t_o)}(a) &\triangleq P(\text{val}_i(t_o) < a \,|\, V_i(t_o) \neq \phi) \\ &= \int_0^{\infty} \int_0^{\frac{(1-a)x}{a}} \frac{\delta^2}{\delta x \delta y} (F_{\sigma,\psi,B_i}(x, y, ahB_{\max}))\, dy\, dx. \end{aligned} \quad (16)$$

Probability of Success in the First Round: We denote the probability of an ONU N_i winning a time-slot in the next batch of K time-slots for transmission amongst all the competing nodes at steady state, as P^i_{succ}, i.e., probability of success of N_i at time t_o. For each ONU N_i, we define the maximum valuation across all nodes except N_i as

$$P(A_i(t_o) > t_o - \alpha, D_i(t_o) > t_o + \beta, B_i(t_o) < \gamma | V_i(t_o) \neq \phi) = 1 - \frac{1 - \sum_{n=0}^{\lfloor \gamma \rfloor} \sum_{Z_n} \prod_{h=1}^{|S|} \left(e^{-\lambda_{hi}(|\mu'_{ji}|+|\theta'_{ji}|)} (\lambda_{hi}|\theta'_{hi}|)^{n_h}/n_h!\right)}{1 - \exp\left(-\sum_{h=1}^{|S|} \lambda_{hi} \times \min\{t_o, \Delta_h\}\right)} \quad (13)$$

$$\begin{aligned} &P(A_i(t_o) > t_o - \alpha, B_i(t_o) < \gamma \,|\, V_i(t_o) \neq \phi) \\ &= P(H_{ji} > \max\{t_o - \alpha, t_o - \delta_{ji}\}, B_i(t_o) < \gamma \,|\, V_i(t_o) \neq \phi) \\ &= 1 - \frac{1 - \sum_{n=0}^{\lfloor \gamma \rfloor} \sum_{Z_n} \prod_{h=1}^{|S|} \left(e^{-\lambda_{hi}(|\mu'_{ji}|+|\theta'_{ji}|)} (\lambda_{hi}|\theta'_{hi}|)^{n_h}/n_h!\right)}{1 - \exp\left(-\sum_{h=1}^{|S|} \lambda_{hi} \times \min\{t_o, \Delta_h\}\right)} \end{aligned} \quad (14)$$

$O_i(t_o) \triangleq \max_{i'=1,i'\neq i}^{N}\{\mathrm{val}_{i'}(t_o)\}$. Given that the utilities of different nodes over the network are independent, the distribution function of $O_i(t_o)$ is

$$F_{O_i}(y) = \prod_{i'=1,i'\neq i}^{N} F_{\mathrm{val}_{i'}(t_o)}(y).$$

The probability of success of N_i at time t_o is given by

$$\begin{aligned} P^i_{\mathrm{succ}}(t_o) &= P(\mathrm{val}_i(t_o) \geq O_i(t_o)) \\ &= \int_0^1 \int_0^x f_{\mathrm{val}_i,O_i}(x,y)\,dy\,dx. \end{aligned}$$

The distributions of $\mathrm{val}_i(t_o)$ and $O_i(t_o)$ are independent since the computation of valuation at a node is affected only by *local* factors like arrival rates, buffer occupancies and delay stringencies and is not affected by *other* nodes in the network. Hence, $f_{\mathrm{val}_i,O_i}(x,y) = f_{\mathrm{val}_i}(x)\times f_{O_i}(y)$.. We thus obtain

$$P^i_{\mathrm{succ}}(t_o) = \int_0^1 f_{\mathrm{val}_i(t_o)}(x) \times F_{O_i(t_o)}(x)\,dx. \qquad (17)$$

Generic Scheduling Round: We now extend the above analysis to the generic scheduling round. We assume that if a node has won slot l, then at the end of this slot, its buffer is emptied *except* for the packets that arrive during slot l. This assumption allows us to say that if all slots have the same duration, then we need the probability distribution of the valuation at the penultimate slot in each *batch* (round) of K slots, given that the changes are only at these batch boundaries.

For a particular node N_i we define ζ_i as the time interval from an arbitrary point in the past to the present slot (t_o) with the following properties.

1) At time $(t_o - \zeta_i)$, the buffer at N_i was empty.
2) In $[t_o - \zeta_i, t_o]$, N_i has not been granted any slots.

We note that the scaled valuation—$\overline{\mathrm{val}}_i(t)$—is an unbounded real number, implying that a Markov chain would lead to a denumerable state space. To avoid such an occurrence, we restrict the value of our scaled valuation to between 0 and 1 *till its 5th decimal place*. To this end, in addition to the above-mentioned assumptions, we make the following assumptions.

1) $\overline{\mathrm{val}}_i(t)$ is constrained to values between 0 and 1. If it crosses either bound its value is taken to be that bound, i.e.,

$$\overline{\mathrm{val}}_i(t) = \begin{cases} 0, & \text{if } \overline{\mathrm{val}}_i(t) < 0 \\ \overline{\mathrm{val}}_i(t), & \text{if } 0 \leq \overline{\mathrm{val}}_i(t) \leq 1 \\ 1, & \text{if } 1 \leq \overline{\mathrm{val}}_i(t). \end{cases}$$

2) In case of a tie, where all valuations are non-zero, the slot is allotted randomly.
3) If all the valuations are zero, the first $(K-1)$ slots are granted to nodes $(K-1)$ to 1 in that order and the Kth slot is granted to the Kth node.

Given the assumption regarding the empty buffers after each transmission, this process is modeled as a Markov chain on a non-denumerable state space. The state space of the system $\mho$ thus consists of (a) the successful transmitting node in the current time-slot, (b) information about the duration passed since each node received a transmission slot in the past, and (c) the knowledge of the valuations by all nodes for the next batch of K slots. A state in this chain is thus a 3-tuple $\{v, a, b\}$, where $v = \{v_1, v_2, \ldots, v_N\}$, $a \in \{1, 2, \ldots, N\}$ and $b = \{b_1, b_2, \ldots, b_N\}$ and defined *over the last slot* out of each batch of K slots, as follows.

1) $\forall j : 0 \leq j \leq K-2, \exists! i'_j \in \{1, 2, \ldots, N\}$, such that $v_{i'_j} = j$ and $a \neq i'_j$ i.e., for each of the last $(K-1)$ slots in the current batch of K slots, a unique node has won each slot.
2) Remaining $v'_i s$ may be any integer between $(K-1)$ and $M = \max_{\forall h: S_h \in S}\{\triangle_h\}$, i.e., their last winning slot was not part of the current round.
3) $b_i \triangleq \overline{\mathrm{val}}_i$ at the beginning of this slot.

Let $\{v = \{v_1\ v_2, \ldots, v_N\}, a, b = \{b_1, b_2, \ldots, b_N\}\}$ be a state of the system belonging to the state space $\mho$. We define the transformation $\Im : \mho \rightarrow 2^{|\mho|}$ as $\Im(\{v, a, b\}) = \{v' = \{v'_1, v'_2, \ldots, v'_N\}, a', b' = \{b'_1, b'_2, \ldots, b'_N\}\}$, where we define the following *properties:*

(*p1*) $\forall j \in \{1, 2, \ldots, K-1\}$, $v'_{i_j} = K - 1 - j$, where $i_j \triangleq$ index of the jth largest component of b
(*p2*) If $a \neq i_j$, $\forall j \in \mathbb{Z}, 1 \leq j \leq K-1$, $v_a = K-1$.
(*p3*) For all other indexes i, $v'_i = \min\{v_i + K, M\}$.
(*p4*) $a' = i_k \triangleq K$th largest component in b.
(*p5*) $\forall l, 0 \leq b_l \leq 1 + b_{i_K} - b_{l''}$ which signifies that the $\mathrm{val}_i \in (0, 1)$.

We next derive the one-step transition probabilities from $\{v, a, b\}$ to $\{v', a', b'\}$. In the trivial case, where these two states are not connected by the transformation $\Im$, the transition probability is 0, i.e., $P(\{v, a, b\} \rightarrow \{v', a', b'\}) = 0$, whereas, for the non-trivial case, using (16) for the pdf of val_i the transition probabilities are given by

$$\begin{aligned} &P(\{v, a, b\} \xrightarrow{\Im} \{v', a', b'\}) \\ &\quad = P(\{v', a', b'\}|\{v, a, b\}) = P(b'|\{v, a, b\}) \\ &\quad = \prod_{i=1}^{N} \left(\int_{L_i}^{U_i} f^I_{i(v_i)}(y_i) dy_i \right) \end{aligned}$$

$$\begin{aligned} F_{\sigma,\psi,B_i}(\alpha,\beta,\gamma) &\triangleq P(\sigma_i(t_o) < \alpha, \psi_i(t_o) < \beta, B_i(t_o) < \gamma | V_i(t_o) \neq \phi) \\ &= \sum_{n=0}^{\lfloor \gamma \rfloor} \sum_{Z_n} \prod_{h=1}^{|S|} \frac{e^{-\lambda_{hi}(\min\{t_o, \Delta_h\})} \times (\lambda_{hi} \times (|\theta'_{ji}| - |\theta_{ji}|))^{n_h}}{n_h! \times \left(1 - e^{-\sum_{h=1}^{|S|} \lambda_{hi} \times \min\{t_o, \Delta_h\}}\right)} \end{aligned} \qquad (15)$$

where

$$L_i = \begin{cases} -\infty \text{ if } x_i = 0 \\ \frac{x_i \cdot (1-5\times 10^{-6})}{1+b_{i_k}-b_i} o/w \end{cases}$$

$$U_i = \begin{cases} +\infty \text{ if } x_i = 1 \\ \frac{x_i \times (1+5\times 10^{-6})}{1+b_{i_k}-b_i} \ o/w \end{cases}.$$

Steady State Analysis of Markov Chain: Consider the state $q \equiv \{\{0, 1, 2 \ldots K-2, K-1, M, \ldots, M\}, K, \{0, 0, \ldots, 0\}\}$, which denotes the state reached when the valuations at all the nodes remain 0 for M consecutive rounds. Given the construction of q, it is trivially reachable from all states in $\mho$, and is hence recurrent. Further, starting from q, as there is a non-zero probability that the system will return to the same state in a batch of K time-slots, q is aperiodic. Denote by $\wp$ the set of all states that are reachable from q. Since q is recurrent, $\wp$ will be an irreducible subset of $\mho$. Also, since one state in $\wp$ is recurrent and aperiodic, all its states are recurrent and aperiodic [25]. Further, because $\wp$ is finite, all its states are recurrent, non-null aperiodic [29]. To study the state of the system in the limit as $t \to \infty$, we only need to consider the set of all recurrent states, which we claim is $\wp$.

Lemma 1: A state in this system (as described above) is recurrent if and only if it belongs to $\wp$.

Proof: For any state $a \in \overline{\wp}$, there is a non-zero probability that it will reach q i.e., $p_{a,q} > 0$. From q, the chances of returning to a are 0 as $a \notin \wp$. Hence, the probability of eventually returning to a from a is not 1, i.e., $p_{a,a} \neq 1$. ■

This means that in the transition matrix we only need to consider the states from $\wp$ and evaluate its steady state probabilities. All the states in $\overline{\wp}$ are transient and thus ignored for a steady state analysis. Note that $\wp$ is recurrent, non-null, aperiodic, and that every state in $\wp$ is mutually reachable from every other state via q. Hence, the stationary state probabilities are the limiting probabilities and they are independent of the initial state [29]. Below, we state an explicit characterization of the states in $\wp$.

Lemma 2: Any state in $\wp$ is of the form $\{v = \{v_1, v_2, \ldots, v_N\}, a, b = \{b_1, b_2, \ldots, b_N\}\}$, where in addition to the properties *p1* to *p5* (stated while defining the transformation $\Im$), the following property also holds true:

(*p6*) $\forall y = 1, 2, \ldots, M-1$, $|\{i | i \in \{1, \ldots, N\}, v_i = y\}| \leq 1$

Proof: By induction on the number of steps (transformations) to reach a state from q.

Base Case: One step from q. By the definition of the transformation $\Im$, because none of the v_i less than M are repeated in q, there will be no such repetitions in all the states reachable from q by a single transformation.

Induction Hypothesis: All states that are reachable from q by less than $(s+1)$ steps satisfy the above property.

Induction Step: Consider a state f which is reachable from q after $(s+1)$ steps. In the sequence of states from s to f the state which immediately precedes f will satisfy the above property, by the hypothesis. Hence, by the definition of the transformation, f will also satisfy the same property.

Lemma 3: All states in $\mho$ satisfying the property (*p6*) (as stated in Lemma 2) are reachable from q.

Proof: Let $f = \{\boldsymbol{v}, a, \boldsymbol{b}\}$ be a state in $\mho$ satisfying properties *p1–p6*. We will prove by induction (on the number of components of $\boldsymbol{v}$ greater than $(K-1)$ and less than M) that all the states of the form $\{\boldsymbol{v}, a, *\}$ are reachable from q.

Base Case: $\boldsymbol{v}$ has exactly one component that is greater than $(K-1)$ and less than M. Let this component be v_j. Without loss of generality, we assume $v_{K+1} = 0, v_{K+2} = 1, \ldots, v_{2K} = K-1$, and $j \in \{2K+1, \ldots, N\}$. Thus, we can reach f from q using the following sequence of transformations as shown at the bottom of the page.

Induction Hypothesis: If f is a state with exactly n of the components of its first vector less than M and greater than $(K-1)$ then f is reachable from q.

Induction step: Let $f = \{v, a, b\}$ be a state in $\mho$ satisfying properties *p1–p6* and exactly $(n+1)$ components of v are greater than $(K-1)$ and less than M. Let these components be $v_{j_1}, v_{j_2}, \ldots, v_{j_{n+1}}$. Without loss of generality, we assume $v_{K+1} = 0$, $v_{K+2} = 1, \ldots, v_{2K} = K-1$, and $\forall i, j_i \in \{2K+1, \ldots, N\}$. Let $v_{j^*} = l \cdot K + m$ be the smallest

$$q \equiv \{\{0, 1, 2 \ldots K-2, K-1, M, \ldots, M\}, K, 0\}$$

$$\xrightarrow{N_{K+1}, N_{K+2}, \ldots, N_{v_j(\text{mod } K)-1}, N_j, N_{v_j(\text{mod } K)+1}, \ldots, N_{2K} \text{ win}}$$

$$\{\{K, K+1, \ldots, 2K-1, 0, 1, \ldots, v_j(\text{mod}K)-1, M, v_j(\text{mod } K) + 1, \ldots, K-1, \ldots, v_j(\text{mod } K), \ldots\}, *, *\}$$

$$N_{K+1}, N_{K+2}, \ldots, N_{v_j(\text{mod } K)-1}, N_j, N_{v_j(\text{mod } K)+1}, \ldots, N_{2K} \text{ win repeatedly for other } v_i'\text{s to saturate to reach } M$$

$$\{\{M, M, \ldots, M, 0, 1, \ldots, v_j(\text{mod}K)-1, M, v_j(\text{mod } K)+1, \ldots, K-1, \ldots, v_j(\text{mod } K), \ldots\}, *, *\}$$

$$\xrightarrow{N_{K+1}, N_{K+2}, \ldots, N_{2K} \text{ win repeatedly for the next } \lceil v_j/k \rceil \text{ cycles}}$$

$$\{\{M, M, \ldots, M, 0, 1, \ldots, v_j(\text{mod}K), \ldots, K-1, \ldots, v_j(\text{mod } K) + K\lceil v_j/k \rceil \equiv v_j, \ldots\}, *, *\} \equiv f$$

such component. Then by the induction hypothesis, the state $f' = \{v', a', *\}$ is reachable from q, where

$$v_i' = \begin{cases} v_{j^*} - l \cdot K, & \text{if } K-1 < v_i < M, \\ v_i, & \text{otherwise.} \end{cases}$$

Further, if the K nodes that transmit in the round, in f' keep being allocated the next l rounds, in the same relative order, we shall reach f. Thus

$$q \rightarrow f' \rightarrow f$$

■

Once we have the stationary state probabilities by solving the linear system, we can find the probability of success i.e., the probability that a node N_j will be one of K nodes that is allotted a slot in the next auction by summing up the stationary-state probabilities across all states, where $|\{i|b_i > b_j\}| < K$. The probability of success would then be given by

$$\begin{aligned} P_{\text{succ}}^j &= P(a = N_j) + \sum_{k=0}^{K-2} \sum_{v \in \{v|v_j = k\}} P(v, a \neq N_j) \\ &= \sum_{v \in V} \sum_{b \in B} P(v, a, b) \\ &\quad + \sum_{k=0}^{K-2} \sum_{v \in \{v|v_j=k\}} \sum_{\substack{a=1 \\ a \neq N_j}}^{N} \sum_{b \in B} P(v, a, b) \end{aligned} \tag{18}$$

where V and B denote the set of all valid vectors v and b respectively.

C. Packet Delay

We now compute the expected delay experienced by a packet as it arrives at a node. We define various events that affect the delay that are instructive in our analysis:

Succ:	Event of a node *winning* a slot for transmission in the next round.
Fl:	Event of a node *losing* all the K slots in the next transmission round.
EB:	Event of a buffer at a node being *empty*.
NEB:	Event of a buffer at a node being *non-empty*.
A_k:	Event of a packet arriving within the kth slot in a round, where $k \in \{1, 2, \ldots, K\}$.
W_m:	Event of a node winning mth slot for transmission in a round, given Succ, where $m \in \{1, 2, \ldots, K\}$.
Tr:	Event of a packet being transmitted in the *same* round in which it arrived.

μ_i denotes the average number of rounds that N_i needs to win a slot. μ_i is computed using an *arithmetic-geometric progression* (AGP) from the probability of success (in (18)) and the associated probability of failure to win a slot multiplied by the round number under consideration. This is solved as follows:

$$\begin{aligned} \mu_i &= P_i^{\text{succ}}.1 + (1 - P_i^{\text{succ}}) \cdot P_i^{\text{succ}} \cdot 2 \\ &\quad + (1 - P_i^{\text{succ}})^2 \cdot P_i^{\text{succ}} \cdot 3 + \cdots \\ &= 1/P_i^{\text{succ}}. \end{aligned} \tag{19}$$

A packet that just arrived finds the buffer at N_i empty, if (1) no packets arrived in its buffer since the start of the system or (2) the buffer size of N_i was less than CT_S/L, just before its last transmission and no packet arrived in its buffer since then

$$\begin{aligned} &P(\text{EB}(t)) \\ &= P(\text{no packet arrival in } (0,t)) \\ &\quad + \sum_{t_1=0}^{t} P(N_i \text{ won a slot at } t_1 (< t)) \\ &\quad \times P(B_i(t) < CT_S/L) \times P(\text{no packet arrival in } (t_1, t)) \\ &= e^{-\lambda_i (t-t_1)} \left(e^{-\lambda_i t_1} + P_i^{\text{succ}} \right. \\ &\quad \left. \sum_{t_1=0}^{t} \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right) \right). \end{aligned} \tag{20}$$

where $\Gamma(\,\cdot\,)$ denotes the incomplete gamma function [27], and $\lfloor \cdot \rfloor$ denotes the *floor* operator. Further, the probability that the arriving packet finds the buffer at the node non-empty is given by

$$\begin{aligned} P(\text{NEB}(t)) &= 1 - P(\text{EB}(t)) \\ &= 1 - e^{-\lambda_i (t-t_1)} \left(e^{-\lambda_i t_1} + P_i^{\text{succ}} \sum_{t_1=0}^{t} \right. \\ &\quad \left. \times \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right) \right) .. \end{aligned} \tag{21}$$

The tree diagram in Fig. 4 classifies the various possible scenarios depending on the aforementioned events. The first event Succ has an associated probability that is given by

$$P(\text{Succ}) = P_i^{\text{succ}}. \tag{22}$$

Now, with exponential arrivals in a slotted departure system, an arrival is equally likely to occur in any of the K slots in a round [28]. Hence

$$P(A_k) = 1/K. \tag{23}$$

To calculate the probability of the event W_m, we note that for a node that wins a slot, it is equally likely to be assigned any of the K slots in the round

$$P(W_m) = 1/K. \tag{24}$$

We next compute the average delay involved in each of the cases listed in Fig. 4 and their respective probabilities of occurrence.

Arrivals at a Non-Empty Queue: We first consider the cases when a packet arriving at a node finds the buffer non-empty. Note that the maximum number of packets that can be transmitted in a slot is CT_S/L.

Case (c_1)*:* Shown in Fig. 5 is the case, where a packet of service type S_h arrives in a non-empty buffer and the node has not been allocated any slots in the present round. This event is denoted as c_1 and is represented as

$$c_1 = \text{NEB}(t) \cap \bigcup_{k=1}^{K} (A_k \cap Fl). \tag{25}$$

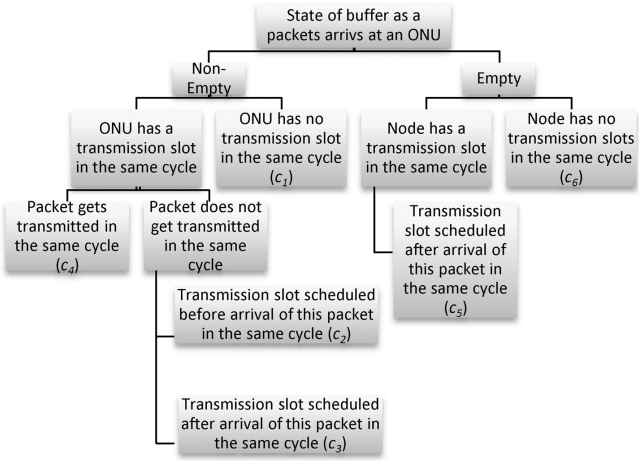

Fig. 4. Tree diagram listing the various scenarios.

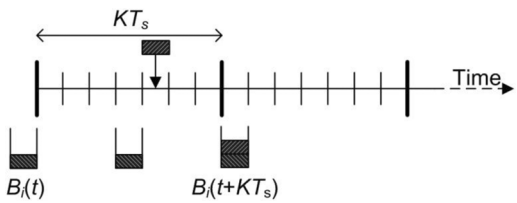

Fig. 5. Case c_1.

Hence, the probability of this case is

$$P_1 = P(c_1) = P(\mathrm{NEB}(t) \cap \bigcup_{k=1}^{K} (A_k \cap Fl)) = P(\mathrm{NEB}(t)) \sum_{n=1}^{K} P(A_n) \cdot P(Fl) = (1 - P_i^{\mathrm{succ}}) P(\mathrm{NEB}(t)). \quad (26)$$

After the arrival of the new packet, the total number of packets in the corresponding buffer is $(B_i + 1)$. Hence, the packet under consideration will be transmitted in the $\lceil (B_i+1)L/CT_S \rceil^{\mathrm{th}}$ successful slot for N_i. For simplicity, let $\lceil (B_i+1)L/CT_S \rceil = (B_i+1)L/CT_S = q$. Since, the node gets one slot in μ_i rounds, this packet will be transmitted in the $\mu_i \cdot q$th round.

Further, as the packet does not get transmitted in the current round, it has to wait for an average amount of time $KT_S/2$ before the next round begins. Additionally, it has to wait for a period of KT_S in each of the intermediate $(\mu_i q - 1)$ rounds. Hence, the total expected delay is

$$D_1 = KT_S/2 + (\mu_i q - 1)KT_S + KT_S/2 = \mu_i q KT_S.$$

Further, $E[q] = (L/(CT_S))E[B_i + 1] = (\lambda L)/(CT_S)$.. Thus, the total delay for a given packet is $\mu_i KT_S E[q] = K\lambda L/CP_i^{\mathrm{succ}}$. However, considering the time-out values of various different services and averaging these over all the services types, the expected delay is

$$E[D_1] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K\lambda L}{CP_i^{\mathrm{succ}}}, \Delta_h \right\}. \quad (27)$$

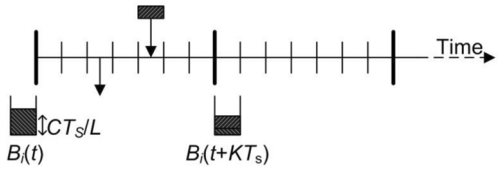

Fig. 6. Case c_2.

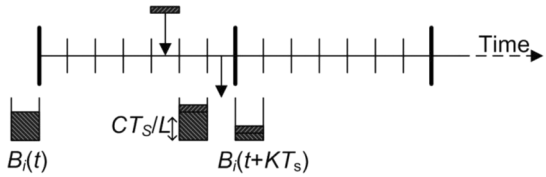

Fig. 7. Case c_3.

Case c_2: In this case, a packet arrives at node N_i with non-empty buffer as shown in Fig. 6. The node wins a slot in the current round and the packet is not transmitted because the winning slot was scheduled before the arrival of the packet in the same round. The event for case c_2 is shown in Fig. 6 and this is denoted as

$$c_2 = \mathrm{NEB}(t) \cap \bigcup_{k=2}^{K} \left(A_k \cap \bigcup_{m=1}^{k-1} W_m \cap \mathrm{Tr}' \right). \quad (28)$$

where Tr' signifies the event that the packet was not transmitted in the current round. This event probability is

$$P_2 = P(c_2) = P\left(\mathrm{NEB}(t) \cap \bigcup_{k=2}^{K} \left(A_k \cap \bigcup_{m=1}^{k-1} W_m \cap \mathrm{Tr}' \right) \right) = P(\mathrm{NEB}(t)) \times (2K^2 - 1) \times (K-1)/12K^2. \quad (29)$$

Note, for this case, $P(\mathrm{Tr}') = 1$. Since the transmission slot of the node has already passed before the arrival of the packet, this case reduces to c_1, whereby the average delay is

$$E[D_2] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K\lambda L}{CP_i^{\mathrm{succ}}}, \Delta_h \right\}. \quad (30)$$

Case c_3: In this case, a packet arriving at N_i finds a non-empty buffer. The node wins a slot scheduled after the arrival of the packet, while the packet is not transmitted in the same round due to large number of packets queued ahead of it, in the buffer. This is shown in Fig. 7 and represented in (30)

$$c_3 = \mathrm{NEB} \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \mathrm{Tr}' \right). \quad (31)$$

This can occur only if $B_i \geq CT_S/L$. Thus, $P(\mathrm{Tr}) = P(B_i \geq CT_S/L)$, which is evaluated using the *cdf* of B_i. Hence, probability of occurrence of this case is

$$P_3 = P(c_3) = P\left(\mathrm{NEB} \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \mathrm{Tr}' \right) \right) = \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right) \left(\frac{K+1}{2} \right) \times P(\mathrm{NEB}(t)). \quad (32)$$

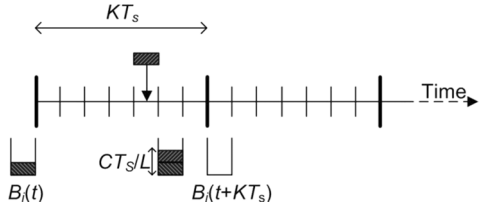

Fig. 8. Case c_4.

This case is similar to c_1, with the exception that in this case the node transmits the first CT_S/L packets in its buffer in the current round itself. Thereby reducing the delay for the current packet by the waiting time to get one slot, which is $\mu_i KT_S = KT_S/P_i^{\text{succ}}$. Hence, in a similar manner as in (27), the expected delay is

$$E[D_3] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K}{P_i^{\text{succ}}} \left(\frac{\lambda L}{C} - T_S \right), \Delta_h \right\}. \quad (33)$$

Case c_4: In this case, a packet arrives at node N_i with a non-empty buffer. The node wins a slot and the packet gets transmitted in the same round. This is shown in Fig. 8 and represented in (34)

$$c_4 = \text{NEB} \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \text{Tr} \right). \quad (34)$$

This represents that a packet arrives in the kth slot, and the node has won one of the remaining slots in the current round. The packet is now transmitted in this winning slot (as show in Fig. 8). For the packet under consideration to get transmitted in the current round, $(B_i + 1) \leq CT_S/L$. Hence

$$\begin{aligned} P_4 = P(c_4) &= P\left(\text{NEB} \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \text{Tr} \right) \right) \\ &= P(\text{NEB}) \times \sum_{k=1}^{K} \left(P(A_k) \times \left(\sum_{m=k}^{K} P(W_m) \right) \times P(\text{Tr}) \right) \\ &= P(\text{NEB}(t)) \times \frac{K(K+1)}{2} \\ &\quad \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right). \end{aligned} \quad (35)$$

Let the packet arrive in slot k and get transmitted in slot m ($\geq k$). Thus, the delay is $d \times T_S$, where $d = m - k$, which follows a uniform distribution in $[0, K-1]$ i.e., $U[0, K-1]$. Hence, the delay will be $\sum_{x=0}^{K-1} xT_S P(d = x) = ((K-1)/(2))T_S$. Thus

$$E[D_4] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \left(\frac{K-1}{2} \right) T_S, \Delta_h \right\}. \quad (36)$$

Arrivals at an Empty Queue: Now we consider the case when a packet arrives at a node and the buffer is empty.

Case c_5: In this case, a packet arrives at node N_i with an empty buffer. The node wins a slot and the packet gets transmitted in the same round. This case can only happen when all the nodes have empty buffers i.e., at start up. This is shown in Fig. 9 and represented as

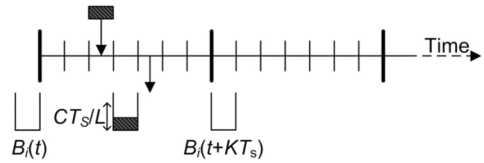

Fig. 9. Case c_5.

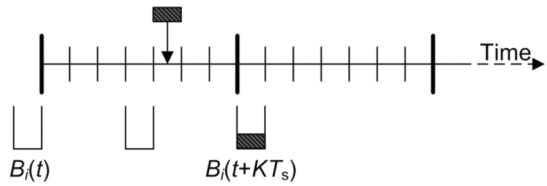

Fig. 10. Case c_6.

$$c_5 = \text{EB} \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \text{Tr} \right). \quad (37)$$

Here note that $P(\text{Tr}) = 1$. The probability of this case is

$$\begin{aligned} P_5 = P(c_5) &= P\left(\text{EB}(t) \cap \bigcup_{k=1}^{K} \left(A_k \cap \bigcup_{m=k}^{K} W_m \cap \text{Tr} \right) \right) \\ &= \left(\frac{K+1}{2K} \right) P(\text{EB}(t)). \end{aligned} \quad (38)$$

Since the average delay computation in this case is same as c_4

$$E[D_5] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \left(\frac{K-1}{2} \right) T_S, \Delta_h \right\}. \quad (39)$$

Case c_6: In this case, a packet arrives at node N_i with an empty buffer and the node does not win any slots in the same round. This is shown in Fig. 10 and represented in (40)

$$c_6 = \text{EB}(t) \cap \bigcup_{k=1}^{K} (A_k \cap Fl). \quad (40)$$

Hence, the probability of this case is

$$\begin{aligned} P_5 = P(c_5) &= P\left(\text{EB}(t) \cap \bigcup_{k=1}^{K} (A_k \cap Fl) \right) \\ &= (1 - P_i^{\text{succ}}) \times P(\text{EB}(t)). \end{aligned} \quad (41)$$

To get a winning slot, it has to wait for μ_i rounds. The delay in this case is $\mu_i KT_S = (KT_S)/(P_i^{\text{succ}})$ and the expected delay will be:

$$E[D_6] = \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{KT_S}{P_i^{\text{succ}}}, \Delta_h \right\}. \quad (42)$$

Hence, using (25)–(42), the expression for the expected delay at a node N_i is given by

$$\begin{aligned}
D_i &= \sum_{j=1}^{6} P_j \times E[D_j] \\
&= P(\mathrm{NEB}(t)) \times (1 - P_i^{\mathrm{succ}}) \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K\lambda L}{C P_i^{\mathrm{succ}}}, \Delta_h \right\} \\
&\quad + P(\mathrm{NEB}(t)) \times \frac{(2K^2-1)(K-1)}{12K^2} \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K\lambda L}{C P_i^{\mathrm{succ}}}, \Delta_h \right\} \\
&\quad + P(\mathrm{NEB}(t)) \times \left(\frac{K+1}{2} \right) \\
&\quad \times \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right) \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{K}{P_i^{\mathrm{succ}}} \left(\frac{\lambda L}{C} - T_S \right), \Delta_h \right\} \\
&\quad + P(\mathrm{NEB}(t)) \times \frac{K(K+1)}{2} \\
&\quad \times \left(1 - \frac{\Gamma(\lfloor CT_S/L \rfloor, \lambda_i P_i(t_1))}{\lfloor CT_S/L \rfloor !} \right) \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \left(\frac{K-1}{2} \right) T_S, \Delta_h \right\} \\
&\quad + P(\mathrm{EB}(t)) \times \left(\frac{K+1}{2K} \right) \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \left(\frac{K-1}{2} \right) T_S, \Delta_h \right\} \\
&\quad + P(\mathrm{EB}(t)) \times (1 - P_i^{\mathrm{succ}}) \\
&\quad \times \sum_h \frac{\lambda_h}{\lambda} \min\left\{ \frac{KT_S}{P_i^{\mathrm{succ}}}, \Delta_h \right\}. \qquad (43)
\end{aligned}$$

V. Efficiency and Optimal K Analysis

In this Section, we compute the network-wide efficiency and optimal value for K for a given network. The network utilization in each round is determined by the fraction of time during which packets are transmitted over the optical fiber in the upstream direction. Let τ_i denote the time duration for which a *winning* node N_i transmits packets in its given time slot. Then

$$\tau_i = \begin{cases} T_S, & \text{if } B_i L/C > T_S \\ B_i L/C, & \text{otherwise} \end{cases} = \min\left\{ T_S, \frac{B_i L}{C} \right\} \qquad (44)$$

$$\begin{aligned}
\overline{\tau_i} &= E[\tau_i] = T_S \times P\left(T_S < \frac{B_i L}{C} \right) \\
&\quad + \sum_{y=1}^{\lfloor CT_S/L \rfloor - 1} \frac{yL}{C} \times P\left(T_S > \frac{B_i L}{C} \right) \\
&= T_S \times P\left(T_S < \frac{B_i L}{C} \right) \\
&\quad + \frac{L}{2C} \left(\left\lfloor \frac{CT_S}{L} \right\rfloor - 1 \right) \left\lfloor \frac{CT_S}{L} \right\rfloor \times P\left(T_S > \frac{B_i L}{C} \right). \qquad (45)
\end{aligned}$$

Equation (44) is evaluated using the pdf of $B_i(t)$ from (15). Note that τ_i is a function of the random variable B_i. Using (6), network efficiency is thus

$$\eta(K, N, \lambda) = \frac{K\overline{\tau_i}}{N(T_b + T_g) + K(T_S + T_g)} \qquad (46)$$

where $\overline{\tau_i}$ is the expected value of slot utilization.

The choice of K provides a tradeoff between network-wide efficiency and the average delay experienced by any node in the network. The computation of optimal K can be approximated to a non-convex optimization process as one of the two parameters (delay) is concave. Since this sort of optimization is hard, we introduce an approximation. Instead of considering parameters of a large number of services at a node, we approximate by considering only one service type at every node—the one whose delay requirement is most stringent.

Approximating (45), the efficiency of the network is given by

$$\eta = \frac{K\overline{\tau_i}}{NT_b + KT_S}. \qquad (47)$$

Ensuring that the latency of the network does not exceed some upper bound δ, it results in the following inequality:

$$\delta_i \le \triangle \times \rho \Rightarrow \triangle \times \rho \ge \frac{T_S}{P_i^{\mathrm{succ}}}. \qquad (48)$$

Further, considering the network-wide behavior we substitute P_i^{succ} by its expected value, $\overline{P_i^{\mathrm{succ}}} = K/N$, resulting in

$$\delta_i \ge NT_S. \qquad (49)$$

Eliminating T_S between (47) and (49), we have

$$K \le \left(\frac{T_b}{\overline{\tau_i}/\eta + (\triangle \times \rho)/N} \right). \qquad (50)$$

Thus the optimal value of K is a function of the maximum achievable network-wide efficiency and is given by

$$K_{\mathrm{opt}} = \frac{NT_b}{\overline{\tau_i}/\eta_{\max} + (\triangle \times \rho)/N}. \qquad (51)$$

VI. Performance and Simulation

To evaluate the performance of our protocol, we developed a GE-PON discrete event simulation model with $N(N = 16, 32, 64, 128)$ ONUs. Each ONU is assumed to have a buffer of maximum capacity 5 MB for upstream transmission as per contemporary ONU technology [24].

The traffic arriving at an ONU is assumed to be a mixture of voice, video and data services in the ratio 30:10:60. Video and data traffic is generated by separate Pareto sources with Hurst parameter 0.8 and voice traffic is assumed to be generated by a Poisson source. The protocol data unit (PDU) being measured is that of Ethernet frames, with exponentially distributed sizes between 64–1500 bytes. The duration of the guard-band is assumed as 5 μs and that of the data-slot to be 150 μs. If a frame is larger than the data-slot, then it is made to "fit" into the data slot by fragmenting the same into multiple Ethernet frames and adding a corresponding header. The size of the valuation frame is fixed to $80\ \mathrm{bytes} = 0.64\ \mu$s. The physical layer uses 8b/10b

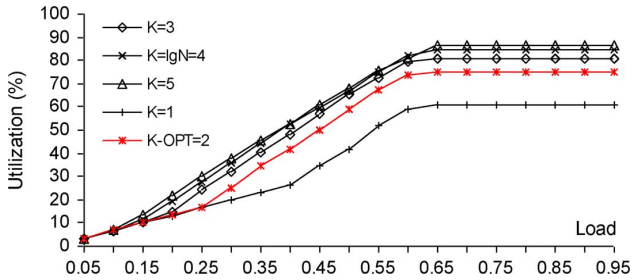

Fig. 11. Bandwidth utilization at different network loads for different values of K with $N = 16$.

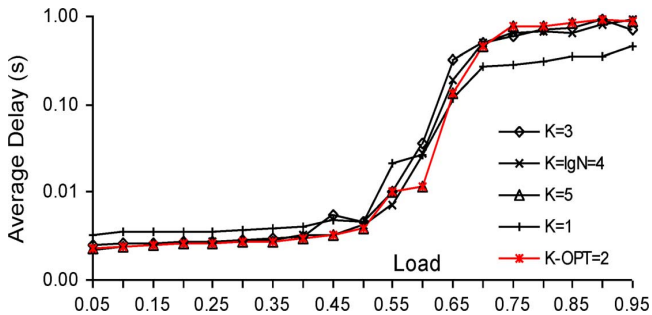

Fig. 12. Average Delay (in sec) for $N = 16$ and different values of K.

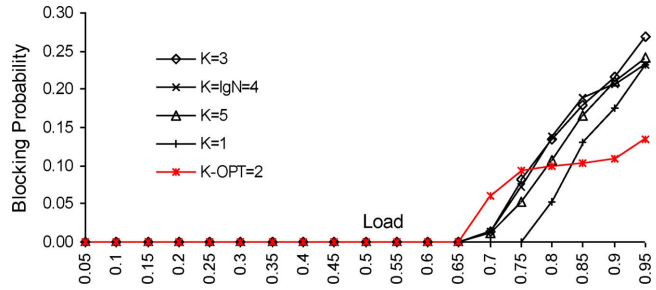

Fig. 13. Blocking probability for $N = 16$ and different values of K.

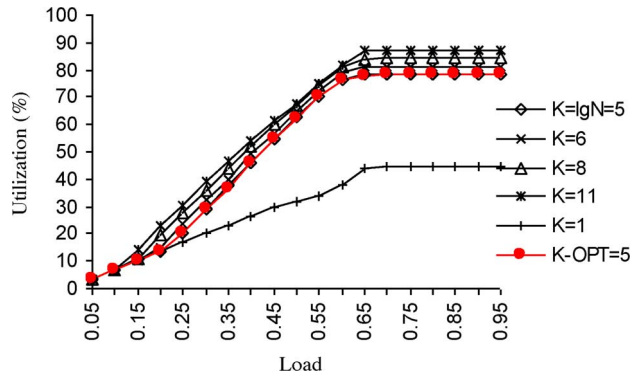

Fig. 14. Bandwidth utilization at different network loads for different values of K with $N = 32$.

encoding and hence reaches 1.25 Gbps capacity for GEPON, and 64/66b encoding for 10GE-PON.

To study the performance of the K-out-of-N scheduling technique with strategic scaling, we conducted simulation for different values of N and K. The idea was to compute utilization and delay as a function of load for the different values of N and K. We also compute blocking probability as a function of load. We compare the performance of our system with a well-known algorithm—IPACT [5], [14]. Finally, we compare our system for 10GE-PON using our proposed algorithm with 10GE-PON (both upstream and downstream) as well as showcase results for NGPON with WDM PON architecture.

Load is computed as the ratio of the total arrival rate at all ONUs in unit time to the total (net) capacity of the system (1 Gbps/10 Gbps) [32]. Hence, load is normalized to values in the range [0, 1]. Efficiency is computed as the ratio of the net time the system transmits data to the total time required for this transmission (inclusive of overhead guard-bands and time required for sending valuations). Blocking probability is computed as the number of packets dropped to those totally generated by the ONUs. The simulation was conducted till steady-state values of blocking probability and efficiency were obtained.

Shown in Fig. 11 through 13 are the delay, throughput and blocking probability plots for our proposed algorithm in a 16-node network.

The primary difference between the results shown in [23] and this paper is the addition of the K-OPT values as well as the inclusion of the NGPON simulation. We obtained the K-OPT values (optimal K values) by using the analytical results shown earlier in Section IV. *The key feature to note is that the K-OPT values lead to an excellent trade-off between utilization, delay and blocking*. It is hence correct to summarize that by keeping $K = K$-OPT we are able to achieve good utilization at an acceptable delay and with low-blocking.

Shown in Fig. 14–16 are the corresponding plots for a 32 node network. The utilization for the case of $N = 32$ and different values of K are shown in Fig. 14. From the graph it is evident that utilization increases with K for a given N and as a function of network load. This is in accordance with the K-out-of-N solution proposed in Section II to improve the efficiency as compared to a simplistic auction with $K = 1$. Further, Fig. 15 shows average delay for $N = 32$ for different values of K. The average delay increases with an increase in K at higher loads. This is because a higher value of K implies a larger time for a round and hence a losing ONU has to wait for longer duration to transmit in the next successful round. However, this analysis does not hold true for the case of $K = 1$ as the number of packets that timeout is more, and these do not add to delay computation. The K-OPT value for $N = 32$ is 5 nodes. The optimal value here gives a high utilization while maintaining acceptable delay and giving an exceptionally low blocking probability.

We provide average delay comparison of our scheduling technique with the constant credit variant of IPACT protocol for 16 ONUs in Fig. 17. The average delay measurements were taken from [5]. At higher loads our protocol outperforms IPACT. Our protocol has better efficiency (for all loads) due to K-out-of-N, as opposed to $K = 1$ as seen in the IPACT protocol. Our results are also significantly better than those for IPACT for larger network sizes, but not shown here since IPACT performance was measured for 16 ONUs as shown in [5].

Shown in Fig. 18 is a comparison of IPACT with our protocol as a function of throughput. Naturally the K-*by*-N scheme outperforms the IPACT scheme as we have better utilization. This is primarily because, in IPACT, the scheduling round has a single ONU transmission, while in our scheme, the there are K transmissions per cycle, implying that we make best use of the

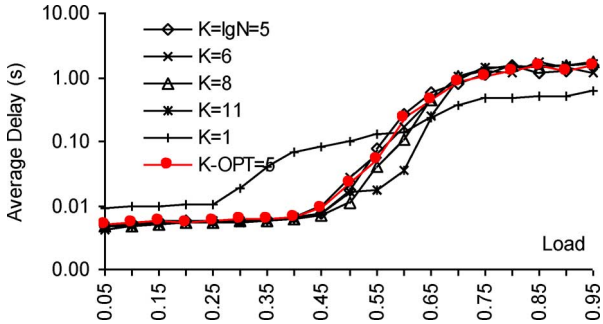

Fig. 15. Average Delay (in sec) for $N = 32$ and different values of K.

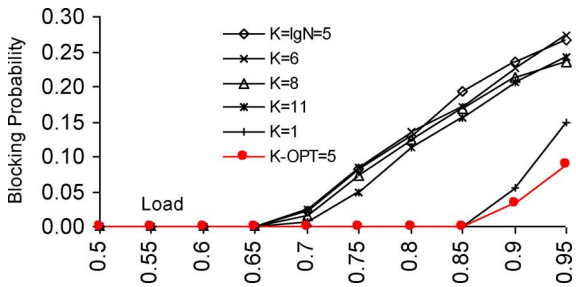

Fig. 16. Blocking probability for $N = 32$ and different values of K.

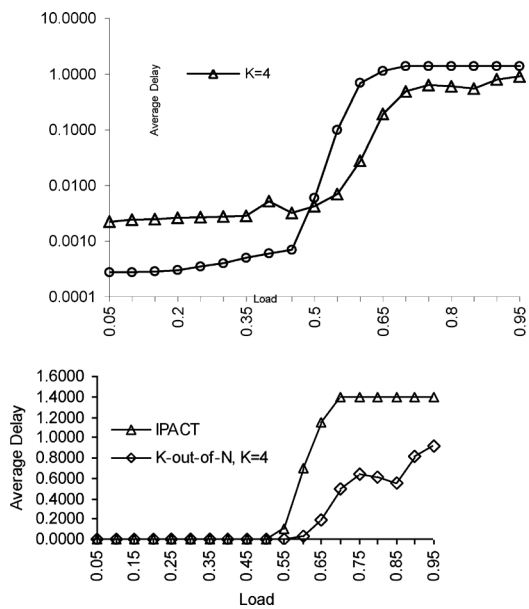

Fig. 17. Average delay comparison of K-out-of-N and IPACT for 16-ONUs for logarithmic scale (above) and linear scale (below).

cycle in transmitting actual data. Secondly, IPACT does not take into consideration latency sensitive services. While one may argue that tokens in the IPACT scheme can be allocated based on latency needs—this allocation is at best only static. IPACT does not propose any latency sensitive features, while the principle premise of our scheme is based on latency sensitivity. We compared in Fig. 16 both the optimal values of K for 32 nodes and 64 nodes. IPACT was compared for 32 nodes and not shown for 64 nodes as the performance betterment was only marginal. It should be noted that our protocol outperforms IPACT by at least 20% on an average and by 32% at the maximum separation point. It is also interesting to note that at high-loads (above 0.7), performance saturates thus making it more predictable.

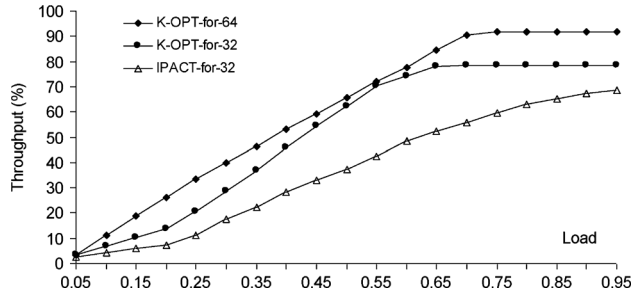

Fig. 18. Comparison of IPACT with K_{opt} for 32 and 64 ONUs—throughput comparison.

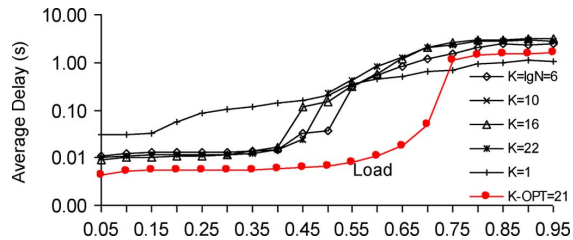

Fig. 19. $N = 64$ case, Average delay.

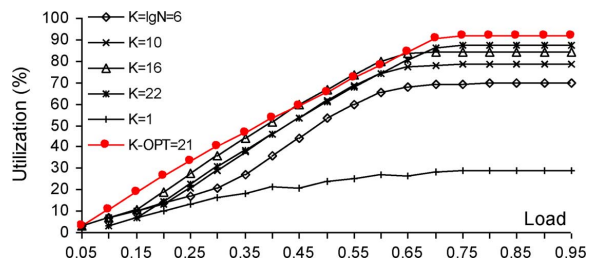

Fig. 20. $N = 64$ case, Utilization.

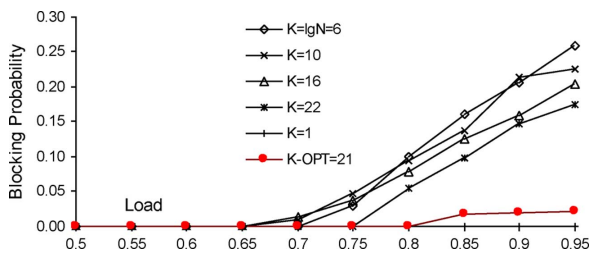

Fig. 21. $N = 64$ case. Blocking Probability.

Shown in Fig. 19–through–22 are average delay, utilization and blocking probability results for K-out-of-N protocol for $N = 64$, and $N = 128$. The protocol performs as expected within 10–20% simulation error. There is a worst case deviation of 10% from the above statement—which is within experimental error. Many DBA algorithms in PONs begin to give worse results as N increases. However, our technique does quite the reverse. In fact, our results (utilization wise) are best for $N = 128$ followed by $N = 64$ and so on. This shows stability and of the scheduling technique.

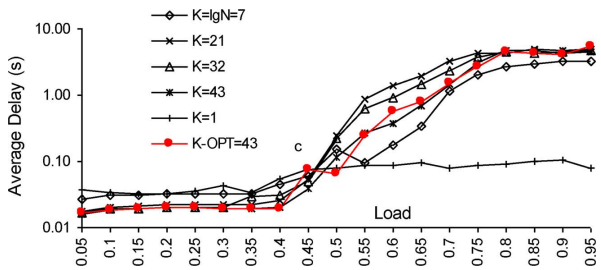

Fig. 22. $N = 128$ case, Average delay.

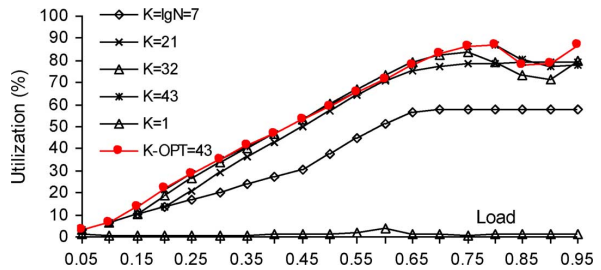

Fig. 23. $N = 128$ case, Utilization.

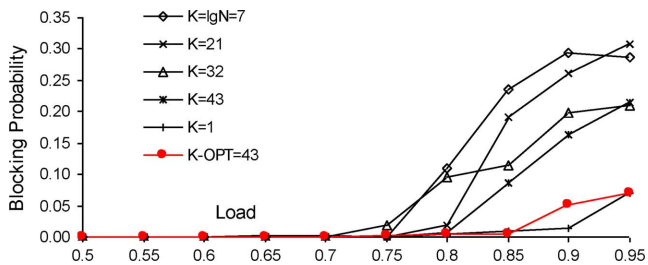

Fig. 24. $N = 128$ case. Blocking Probability.

Finally, shown in Fig. 25 is the delay and utilization (throughput) profiles for 32, 64, and 128 node system in a 10GEPON network. This network was simulated in accordance with the standard [24]. We demonstrate in Fig. 25 how our proposed algorithm scales for the 10GE-PON network from both a delay, as well as throughput perspective. In fact, the throughput in the 10GE-PON system is much better than the 1GE-PON system using our algorithm. Further, the 128-node case has better results than the 64 and 32 node cases, as expected for our technique. Delay at high-loads is however quite high—goes almost to 1 second and this is because of the large buffer—almost 50 Mb chosen leading to less packets being dropped—as the cost of delay.

Simulations for NGPON-2: In the NGPON2 models, we assume K upstream wavelengths to be time-shared between N users ($N = 128$). Our focus is to find the performance of such a system from delay and packet drop perspectives. We also investigate if the optimal K value in GPON is valid for NGPON2. All the simulation settings are assumed to be identical as in the GEPON case. The ONUs are assumed to have tunable lasers and the time to tune a laser is assumed to be an integral multiple of the guard-band time-slot-width [33]. Shown in Fig. 26 are the delay computations for NGPON2. Naturally the NGPON2 delays are much lesser than GEPON—however even with wavelength sharing (K out of N wavelengths being shared), the

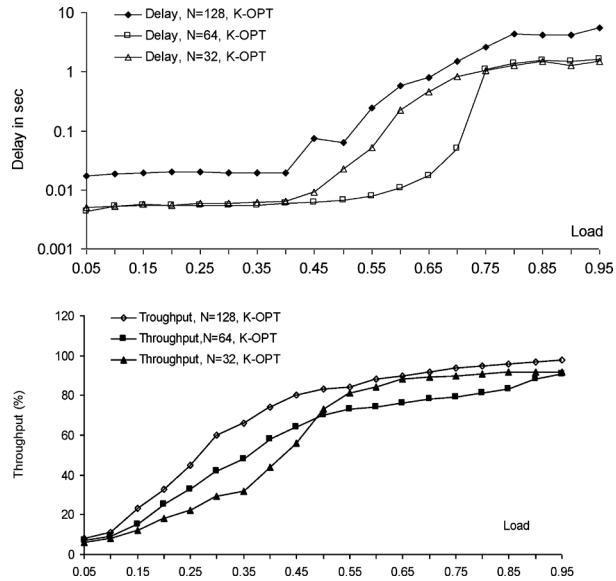

Fig. 25. Delay (above) and throughput (below) for 10GE-PON for 32, 64, and 128 nodes with K_{opt} values.

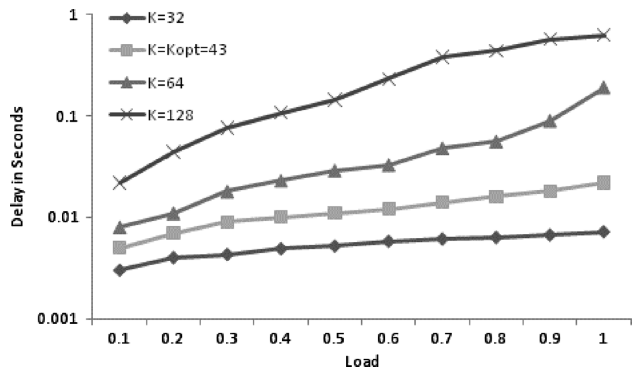

Fig. 26. Delay for NGPON2 as a function of Load.

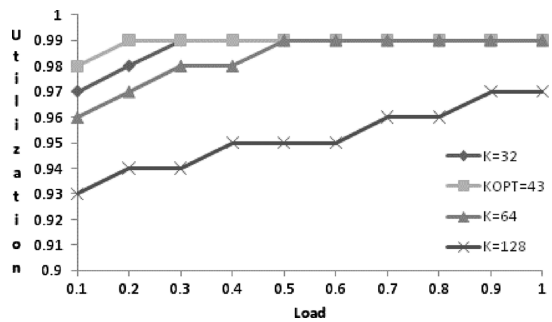

Fig. 27. Utilization for NGPON2 as a function of Load.

delay does not increase exponentially—as would be intuitively expected. Shown in Fig. 27 is the corresponding utilization for NGPON2 as a function of load for different values of K. Note that the utilization is best for optimal K. This is counter-intuitive and can be explained as follows: for optimal K—there is a slot/wavelength available exactly at the time when the buffer is getting filled to the brim, or the services are being timed out—resulting in very good utilization of the slot/wavelength.

VII. Conclusion

In this paper, we proposed and investigated a novel K-out-of-N protocol for dynamic bandwidth allocation in Next Generation PON networks. The proposed protocol is novel on three accounts: it provides higher efficiency, provides intuitive fairness and enables a logical (true) representation of bandwidth allocation in terms of ONU service requests. The proposed protocol hence is fair, and lowers average latency while enabling excellent network utilization. The protocol is evaluated through a rigorous stochastic model. The protocol provides for a generic framework for any DBA algorithms in GPON and NGPON systems. The stochastic model leads to delay and optimal K bounds and these are key to evaluating the performance of the proposed protocol. The optimal K value is also computed. To further understand the protocol a simulation model for different sized networks is built and rigorous simulation results are presented. We see that the choice of K leads to a trade-off; larger the value of K, better the network efficiency, but worse the average latency expected on a network-wide basis.

References

[1] G. Kramer, *Ethernet Passive Optical Networks*. New York: McGraw-Hill, 2005.

[2] A. Gumaste and S. Q. Zheng, "Dual auction (and recourse) opportunistic protocol for light-trail network design," in *Proc. IEEE Wireless Opt. Commun. Conf.*, Bangalore, India, 2006.

[3] A. Gumaste and T. Antony, *First-Mile Access Networks and Enabling Technologies*. New York: Cisco Press, 2004.

[4] I. Chlamtac, A. Gumaste, and C. Szabo, *Broadband Services: Business Models and Technologies for Community Networks*. New York: Wiley, 2005.

[5] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: A Dynamic Protocol for an Ethernet PON (EPON)," *IEEE Communications Magazine*, vol. 40, pp. 74–80, Feb. 2002.

[6] P. Klemperer, *Auctions: Theory and Practice*. Englewood Cliffs, NJ: Prentice Hall.

[7] A. Gumaste and I. Chlamtac, "A protocol to implement ethernet over PON," in *Proc. IEEE Int. Conf. Comm.(ICC)*, Anchorage, AK, May 2003, vol. 2, pp. 1345–1349.

[8] *Ethernet in the First Mile*, IEEE EFM Standard 802.3ah.

[9] M. Hajduczenia *et al.*, "On EPON security issues," *IEEE Commun. Surveys and Tutorials*, vol. 9, no. 1, pp. 68–83, 1st quarter 2007.

[10] M. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: A survey of dynamic bandwidth allocation (DBA) algorithms," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. S8–S15, Aug. 2004.

[11] S. Sherif *et al.*, "A novel decentralized ethernet-based PON access architecture for provisioning differentiated QoS," *J. Lightw. Technol.*, vol. 22, no. 11, pp. 2483–2497, Nov. 2004.

[12] A. Shami *et al.*, "QoS control schemes for two-stage ethernet passive optical access networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 8, pp. 1467–1478, Aug. 2005.

[13] T. Holmberg, "Analysis of EPONs under the static priority scheduling scheme with fixed transmission times," in *Proc. IEEE Conf. Next Generation Internet Design and Engineering (NGI)*, Apr. 2006, pp. 192–199.

[14] G. Kramer *et al.*, "Fair queuing with service envelopes (FQSE): A cousin-fair hierarchical scheduler for subscriber access networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1497–1513, Oct. 2004.

[15] C. Assi *et al.*, "Dynamic bandwidth allocation for quality-of-service over ethernet PONs," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 9, Nov. 2003.

[16] A. Dhaini *et al.*, "Dynamic wavelength and bandwidth allocation in hybrid TDM/WDM EPON networks," *J. Lightw. Technol.*, vol. 25, no. 1, pp. 277–286, Jan. 2007.

[17] A. Shami *et al.*, "Jitter performance in ethernet passive optical networks," *J. Lightw. Technol.*, vol. 23, no. 4, pp. 1745–1753, Apr. 2004.

[18] H.-J. Byun, J.-M. Nho, and J.-T. Lim, "Dynamic bandwidth allocation algorithm in ethernet passive optical networks," *Electron. Lett.*, vol. 39, no. 13, pp. 1001–1002, Jun. 2003.

[19] Y. Zhu, M. Ma, and T. Cheng, "Hierarchical scheduling to support differentiated services in ethernet passive optical networks," *Comput. Netw.*, vol. 50, no. 3, pp. 350–366, Feb. 2006.

[20] J. Zheng, "Efficient bandwidth allocation algorithm for ethernet passive optical networks," *IEE Proc. Commun.*, vol. 153, no. 3, pp. 464–68, Jun. 2006.

[21] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Trans. Net.*, vol. 1, no. 3, pp. 344–57, Jun. 1993.

[22] A. Banerjee, G. Kramer, and B. Mukherjee, "Fair sharing using dual service-level agreements to achieve open access in a passive optical network," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 32–44, Aug. 2006.

[23] K. Nagaraj, A. Gudhe, A. Gumaste, and N. Ghani, "A novel K-out-of-N auction mechanism and strategic scaling for dynamic bandwidth allocation in GE-PON," presented at the 49th IEEE Globecom Conf., San Francisco, CA, 2006.

[24] *Standard of the 10G-PON Working Group*, IEEE 802.3av 10GEPON DRAFT D2.IEEE P 2 802.3av.

[25] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1968, vol. I, II.

[26] J. Sun, E. Modiano, and L. Zhang, "Competitive fairness: Wireless channel allocation using an auction algorithm," *IEEE J. Sel. Areas Commun.*, May 2006.

[27] [Online]. Available: www.wikipedia.org/incomplete_gamma_function

[28] D. Bertsekas and R. Gallagher, *Data Networks*, 2nd ed. Cambridge, MA: MIT Press, 1994.

[29] K. S. Trivedi, *Probability and Statistics With Reliability, Queuing, and Computer Science Applications*. New York: Wiley, 2001.

[30] A. Gumaste, T. Das, A. Mathew, and A. Somani, "An autonomic virtual topology design and two-stage scheduling algorithm for light-trail WDM networks," *J. Opt. Commun. Netw.*, vol. 3, no. 4, Apr. 2011.

[31] FSAN Online. [Online]. Available: www.fsanweb.com

[32] M. P. McGarry, M. Reisslein, and M. Maier, "Ethernet passive optical network architectures and dynamic bandwidth allocation algorithms," *IEEE Communications Surveys and Tutorials 10*, vol. 1–4, no. 46–60, 2008.

[33] H. Rohde, "Coherent optical access networks," presented at the IEEE/OSA OFC 2011, Paper OTuB1.

Tamal Das received the B.Tech. and M.Tech. degrees in mathematics and computing from IIT Delhi, India. He is currently pursuing the Ph.D. degree in the Department of Computer Science and Engineering at IIT Bombay, India.

His research interests are in stochastic analysis of high-speed networks.

Mr. Das was awarded with IEEE ANTS 2010 Best Paper Award, and was nominated for Corning Outstanding Student Paper in IEEE/OSA OFC/NFOEC 2010.

Ashwin Gumaste is currently the James R. Isaac Chair and faculty member in the Department of Computer Science and Engineering at the Indian Institute of Technology (IIT) Bombay.

He is currently also a consultant to Nokia Siemens Networks, Munich where he works on optical access standardization efforts. He was a Visiting Scientist with the Massachusetts Institute of Technology (MIT), Cambridge, in the Research Laboratory for Electronics from 2008 to 2010. He was previously with Fujitsu Laboratories Inc in the Photonics Networking Laboratory (2001–2005). He has also worked in Fujitsu Network Communications R&D, Richardson, TX, and prior to that with Cisco Systems in the Optical Networking Group (ONG). His work on light-trails has been widely referred, deployed and recognized by both industry and academia. His recent work on Omnipresent Ethernet has been adopted by tier-1 service providers. He has 20 granted US patents and over 30 pending patent applications and has published about 130 papers in referred conferences and journals. He has also authored three books in broadband networks called *DWDM Network Designs and Engineering Solutions* (Wiley) (a networking bestseller), *First-Mile Access Networks and Enabling Technologies* (Wiley), and *Broadband Services: User Needs, Business Models and Technologies* (Wiley). He has been with IIT Bombay since 2005 where he convenes the Gigabit Networking Laboratory (GNL). The Gigabit Networking Laboratory has secured over 8 million USD in funding since its inception and has been involved in 4 major technology transfers to the industry.

Dr. Gumaste was awarded the Government of India's DAE-SRC Outstanding Research Investigator Award in 2010 as well as the Indian National Academy of Engineering's (INAE) Young Engineer Award (2010). He has served Program Chair, Co-chair, Publicity chair and workshop chair for IEEE conferences and as Program Committee member for IEEE ICC, Globecom, OFC, ICCCN, Gridnets etc. He is also a Guest Editor for *IEEE Communications Magazine* and *IEEE Network*. He is the Secretary of the IEEE Communication Society's Technical Committee on High Speed Networks (TCHSN) 2008–2010.

Akhil Lodha received a Dual Degree (Integrated B.Tech. and M.Tech.) in computer science from IIT Bombay, India, and the M.S. degree in computational finance from Carnegie Mellon University, Pittsburgh, PA.

He works as a quantitative trader focusing on low latency trading. Previously, he did research in different domains of networking like communication networks, mobile computation, and optical and access networks. With Dr. A. Gumaste, he worked on the design of novel network architectures for metro and access.

Ashish Mathew is pursuing the B.Tech. degree in computer science at IIT Bombay, India.

He has performed summer internships at Gigabit Networking Laboratories, IIT Bombay and Google, India. Research interests include stochastic analysis, networks and robust optimization.

Nasir Ghani (SM'06) received the B.S. degree in computer engineering from the University of Waterloo, ON, Canada, the M.S. degree in electrical engineering from McMaster University, ON, Canada, and the Ph.D. degree in computer engineering from the University of Waterloo, ON, Canada.

He is an Associate Professor in the Department of Electrical and Computer Engineering at the University of New Mexico, where he is actively involved in a range of research and teaching activities. His current interests include network cyber-infrastructure design, survivability, services/applications, and knowledge-based systems. He has published over 40 journal and 80 conference papers, several book chapters, and has also co-authored various industry standards contributions. Prior to joining academia, Dr. Ghani spent over 8 years working in industry and held key technical positions at several large companies (Nokia, Motorola, IBM) as well as some start-up organizations (Sorrento Networks, Array Systems Computing). In addition, he has actively been involved in a range of outreach and technical community service roles. For example, he has served as chair of the IEEE ComSoc Technical Committee on High Speed Networks (TCHSN) from 2008–2010 and has also been a symposium co-chair for IEEE Globecom (2006, 2010) and IEEE ICC (2006, 2011). He has also built and established a successful high-speed networking workshop series for the flagship IEEE INFOCOM conference and served as a panelist for numerous NSF, DOE, and international panels.

Dr. Ghani is an Associate Editor of IEEE COMMUNICATIONS LETTERS and has also guest-edited special issues of the *IEEE Network, IEEE Communications Magazine*, and Cluster Computing journals. He is has also been faculty advisor for the Etta Kappa Nu honor society. Overall, his work has been funded by several U.S. government agencies (including the National Science Foundation and Department of Energy) as well as some industry sponsors. In particular, he received the NSF CAREER award in 2005 to support his work in the area of multi-layer network design.