A Novel Connection Setup Management Approach for Optical WDM Networks

Wissam Fawaz, Ken Chen, and Chadi Abou-Rjeily

***Abstract*— Wavelength Division Multiplexing (WDM) has significantly increased the transmission capacity of today's optical networks. An increase in the available bandwidth is promoting at the same time the introduction of new services, each having different Quality of Service (QoS) requirements. In this regard, QoS parameters applicable to connection setup, namely connection blocking probability and connection setup time, are expected to greatly influence the procedure of lightpath setup. In view of this, we propose to consider the connection setup time requirement as a (timely increasing) priority indicator during optical connection provisioning. In other words, we envisage in this letter the assignment of the highest setup priority to those optical connection requests having the shortest setup time requirements. To achieve this purpose, we adapt the well known Earliest Deadline First (EDF) scheduling discipline to the particular case of optical connection setup management. In order to gauge the impact of our proposal on the QoS perceived by optical clients, we introduce in this letter a computaional method. This latter is used for the assessment of the percentage of optical connections that are successfully established under the proposed setup management approach.**

***Index Terms*— Optical networks, connection setup management, Earliest Deadline First scheduling, performance analysis.**

I. Introduction

THE continuing growth in terms of data traffic is creating a situation where the need for higher and higher bandwidth becomes inevitable. Wavelength Division Multiplexing (WDM) is arising in this regard as a key technology that is able to boost the transmission capacity of optical networks, allowing optical operators to accomodate the continuing expansion of traffic demand. WDM divides the tremendous bandwidth of a fiber into many non-overlapping wavelengths, each of them operated at the peak electronic speed of several gigabits per second.

This increase in the available bandwidth is however coupled with the advent of new services, each having different Quality of Service (QoS) requirements. A great deal of effort has been made in this context to provide a predictable quality of transport service, which is defined by every parameter affecting data flow after the connection is established (for further information see [1]). Nonetheless, most of the previous studies left out two important parameters when dealing with the subject of lightpath setup management, namely *the connection setup time* and *the connection blocking probability* [2]. Therefore, we turn our attention in this letter to the connection setup time parameter that is likely to become a potential service differentiator in Service Level Agreements (SLA) between optical operators and their customers [3]. The agreed upon connection setup time reflects the time needed from the moment a lightpath (i.e., a service) request is generated and the moment a lightpath is set up. The main purpose of this work is to investigate the effect of considering the connection setup time during optical connection setup management. Building on this observation, this letter proposes a novel connection setup management approach. The rationale behind our proposal lies in considering the optical connection setup time as a timely increasing priority indicator during the setup process. In this manner, it can be assimilated to a *deadline*. As such, we adapt the well known *Earliest Deadline First (EDF)* scheduling discipline to the particular case of optical connection setup management. In other words, the connection requests which are blocked due to lack of optical resources at a certain optical source node A are not immediately dropped. But they are queued instead at A's level within an EDF queue according to an increasing order of their required connection setup times (deadlines). The first customer to be served will be the one having the smallest connection setup time. The wider idea behind this proposal is to use the *connection setup time* as an indicator of the urgency of the request (a kind of competition oriented parameter which may be linked to pricing), in order to achieve a better service differentiation. The model we will present below yield a way to forecast the success probability of a request at the same instant it is queued.

Manuscript received August 29, 2007. The associate editor coordinating the review of this letter and approving it for publication was Dr. Wayne Grover.

W. Fawaz and C. Abou-Rjeily are with Lebanese American University (LAU), Byblos, Lebanon (e-mail: {wissam.fawaz, chadi.abourjeily}@lau.edu.lb).

Ken Chen is with the University of Paris 13 - L2TI Lab, 99, Avenue Jean-Baptiste Clement, 93430 Villetaneuse, France (e-mail: chen@galilee.univ-paris13.fr).

Digital Object Identifier 10.1109/LCOMM.2007.071437.

II. Description of the Proposed Scheme

Consider the sample scenario presented in Figure 1(a) where two Wavelength Routers (WRs) A and B are connected through a fiber link that holds W wavelengths (WLs). Note that in a previous work [4], we handled the simplified case of one wavelength between the two WRs. As a key distinguishing feature from this previous work, we take in this letter one step further by considering the general case of multiple (W) WLs per fiber link. When W connections are established between A and B (occupying all W WLs), future connection requests between A and B will be blocked. However, if the proposed setup management approach is deployed, the blocked connections are guaranteed to be queued at A's level in an EDF queue according to an increasing order of their required setup times. A queuing representation of this scenario is depicted in Figure 1(b). The *service* here is the offer of an optical connection (i.e., the occupation of a WL). The W WLs are thus represented in Figure 1(b) by W servers. This scenario will be considered in the next section where a performance evaluation study of the proposed setup scheme is presented.

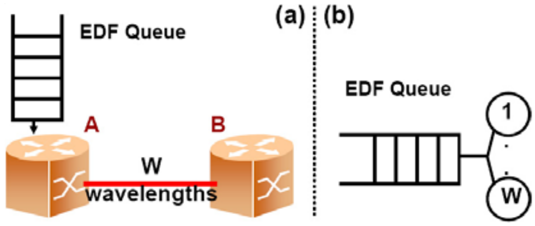

Fig. 1. A sample scenario.

III. Performance Evaluation: the Model

As our management approach relies on the EDF discipline (proposed by Jackson in 1955 [5]), it benefits and inherits the properties of the latter. In particular, EDF minimizes the deadline missing probability. Many efforts have been done for characterizing the EDF queue and assessing its main performance metrics, in particular the death probability. However, the general characterization of EDF (queue size distribution, stability condition, etc.) is still an open research issue. In particular, to the best of the authors' knowledge, the evaluation of the deadline mismatch ratio remains an open problem. This latter parameter is a major performance metric in our connection setup management approach, since it leads directly to the probability of setting up a connection in time. In this letter, we present our approach that tackles the aforementioned open problem by taking the initial position of the customer as a parameter of the problem. This additional initial condition allowed us to develop a Markovian model, from which we were able to derive key performance metrics. In the following subsections, we will use the vocabulary of customer instead of *connection setup request*, server instead of *wavelength*, and laxity instead of *setup time*.

A. Definitions and Assumptions

- The time axis is slotted, *i.e.* , Time is divided into equal-length unit-slots. Customers arriving during one slot are considered at the beginning of the next slot.
- To gain insight into the model, the service times are constant and assumed to be equal to one time slot each. As such, any customer in service at the beginning of a given slot will have left the system at the beginning of the next slot.
- Each customer comes with an initial *laxity* (denoted by L), which is the relative margin to its deadline expiration. We assume that the initial laxities of customers are independent and identically distributed (i.i.d) integer random variables (r.v.). The CDF (Cumulative Distribution Function) of the initial laxity will be denoted by $F_L(.)$. To simplify numerical computation, we assume that L is upper bounded by Λ. Thus, we have $F_L(0) = 0$, and $F_L(l) = 1 \ \forall l \geq \Lambda$. The residual laxity of a customer decreases as time advances. For the sake of simplicity of formulation and without loss of generality, we consider a customer *alive* at a slot if in the *beginning* of this slot its residual laxity is *strictly* positive. As we consider only the deadline mismatch probability, when the laxity hits zero, it is maintained at this value to serve as an indicator of deadline mismatch.
- Services are non-preemptive and work conserving. This means in particular that *dead* customers (those having 0 residual laxity) are served as well.
- The queue has $(K-1)$ waiting places. These Waiting places are numbered from 1 to $K-1$, the (virtual) position 0 refers to a service completion situation, and the (virtual) position K refers to a rejection situation. The first W customers are served on the next slot.
- The initial position of each customer is known. Customers move in the queue until being served (position 0) or pushed out of queue by customers with tighter laxity (position K).
- The overall arrival process is Poisson, denoted $\mathcal{P}$, with arrival rate λ_o. As the initial laxities are i.i.d integer variables, customers coming with an initial laxity strictly smaller than a particular laxity value, say l, form also a Poisson process $\mathcal{P}(l)$ with arrival rate $\lambda(l)$ given by $\lambda(l) = F_L(l-1)\lambda_o$ where $F_L(.)$ is the CDF of the initial laxity law. For a target customer $\mathcal{C}$ with residual laxity equal to l at the beginning of a slot, only those coming within the slot and having an initial laxity strictly smaller than l are to be inserted prior to $\mathcal{C}$. Let $A(l)$ denote the number of customers arriving in a slot with initial laxity $d < l$. As such, $a(l,i)$, the probability of having i customers inserted before a target customer with residual laxity l, is given by

$$a(l,i) = Pr\{A(l) = i\} = e^{-\lambda(l)} \frac{[\lambda(l)]^i}{i!}. \qquad (1)$$

B. A Markov Chain Model

Let us consider a target customer $\mathcal{C}$ arriving in a slot taken as time origin ($t = 0$) with an initial laxity L and initial queueing position N. The state of $\mathcal{C}$ at the i-th slot (S_i) can be described by a pair of random variables (r.v.) $S_i = (n_i, m_i)$ where

- the meaning of n_i depends on its value
  - For $1 \leq n_i \leq K-1$, n_i gives the position occupied by $\mathcal{C}$in the waiting space of the queue.
  - By convention, $n_i = 0$ means $\mathcal{C}$ gets eventually served. Thus, once $\mathcal{C}$ enters a state with $n_i = 0$, it stays there forever.
  - $n_i = K$ means $\mathcal{C}$ is rejected without service. Once $\mathcal{C}$ enters a state with $n_i = K$, it stays there forever.
- m_i gives the residual laxity. By convention, $m_i = 0$ means that there is no more residual laxity, thus $m_i = 0, \ldots, \Lambda$.

The $\{S_i\}_{i\geq 0}$ form a Markov chain. The evolution of (n_i, m_i) is determined by the following relation:

$$m_{i+1} = \max(0, m_i - 1) \qquad (2)$$

$$n_{i+1} = \operatorname{minmax}\ (0, n_i - W + A(m_i), K) \qquad (3)$$

where the function minmax (m, x, M) is a double limitation function

$$\operatorname{minmax}\ (m,x,M) = \begin{cases} m, & x \leq m \\ x, & m < x < M \\ M, & x \geq M \end{cases}$$

TABLE I
CUSTOMERS WITH INITIAL LAXITY 12 AND INITIAL POSITION N

N	P_{cs}	P_{ls}	P_{sr}	P_{rr}
24	1.0000	0.0000	0.0000	0
25	0.9999	0.0001	0.0000	0
26	0.9997	0.0003	0.0000	0
27	0.9992	0.0008	0.0000	0
28	0.9980	0.0019	0.0001	0
29	0.9940	0.0050	0.0010	0

and $A(m_i)$ is the number of customers arriving during the i-th slot with initial laxity strictly smaller than m_i, the law of $A(m_i)$ is given by Eqn. 1.

The evolution of S_i is thus totally forecastable from its current position. In addition, $P(A(m_i) = l)$ depends only on the values of the residual laxity m_i, and does not depend on the particular time position i. We obtain thus a transient homogeneous Markov Chain with absorbing states. Actually, the states $(0, m)$ and (K, m), $m = 0, \ldots, \Lambda$, are *absorbing* states by our convention. The states $(0, m)$ are those representing a customer eventually being served with laxity m, and the states (K, m) are those representing a customer eventually being pushed out of queue with laxity m.Theses absorbing states are of particular interest. With adequate interpretation (semantic), they provide estimation of useful performance metrics.

Consider a homogeneous Markov chain $X = \{X_i\}_{i\in\mathbb{N}}$ with a finite state space Ω, and transition probabilities $p_{ij} = Pr\{X_{n+1} = j/X_n = i\}$, $(i, j) \in \Omega^2$. Ω contains a subset of absorbing states $\mathcal{G}$, $\forall i \in \mathcal{G}, \forall j \in \Omega - \mathcal{G}, p_{ij} = 0$. Let $\mathcal{M} = \Omega - \mathcal{G}$, $\mathcal{M}$ is the non-absorbing subset. $(\mathcal{G}, \mathcal{M})$ forms a partition of Ω. We are interested in the probability of entering these absorbing states, *i.e.* , the probability of the event $G(i)$ that a chain which started at $X_0 = i$ enters eventually in $\mathcal{G}$.

$$G(i) = \{\exists n \geq 0, \exists j \in \mathcal{G}, X_n = j/X_0 = i\}$$

Let $g(i) = Pr\{G(i)\}$. Before drawing down the computation of $g(i)$, let us first notice that for $i \in \mathcal{G}$, $g(i) = 1$. Thus, we only have to find formula for the cases $X_0 = i, i \in \mathcal{M}$. Let us first notice that, due to the memoryless property, and by assuming that $\{X_1 = j, X_0 = i\}$, $j \in \Omega$, does take place, we have

$$Pr\{G(i)/X_1 = j\} = Pr\{G(j)\}p_{ij} = g(j)p_{ij}.$$

We have

$$\begin{aligned} g(i) &= \sum_{j\in\Omega} Pr\{G(i)/X_1 = j\} \\ &= \sum_{j\in\mathcal{M}} g(j)p_{ij} + \sum_{j\in\mathcal{G}} 1 \times p_{ij} + \sum_{j\in\mathcal{B}} 0 \times p_{ij} \end{aligned}$$

Thus, we get the solution of all of the $g(i), i \in \mathcal{M}$ by solving a system of Card$\{\mathcal{M}\}$ linear equations with Card$\{\mathcal{M}\}$ unknowns, in the form of

$$\sum_{j\in\mathcal{M}, j\neq i} p_{ij}g(j) + (p_{ii} - 1)g(i) = -[\sum_{l\in\mathcal{G}} 1 \times p_{il}], \quad i \in \mathcal{M} \quad (4)$$

C. Performance Metric Evaluation

We apply now the proposed model to our specific EDF queueing system. We take into account the following metrics, which are conditioned by both the initial position N and initial laxity L of the target customer.

- Conformed setup completion ($m > 0$) probability, denoted by $P_{cs}(N, L)$
- Setup completion in late ($m = 0$) probability, $P_{ls}(N, L)$
- Setup rejection ($m > 0$) probability, $P_{sr}(N, L)$
- Reasonable rejection ($m = 0$) probability, $P_{rr}(N, L)$.

These metrics are obtained by computing $g(i(N, L))$ (cf. Eqn. 4) under an appropriate definition of $\mathcal{G}$.

- For $P_{cs}(N, L)$, $\mathcal{G} = \{j(0, m)/m = 1, \ldots, \Lambda\}$.
- For $P_{ls}(N, L)$, $\mathcal{G} = \{j(0, 0)\}$.
- For $P_{sr}(N, L)$, $\mathcal{G} = \{j(K, m)/m = 1, \ldots, \Lambda\}$.
- For $P_{rr}(N, L)$, $\mathcal{G} = \{j(K, 0)\}$.

Various *scenarios* have been tested. Due to space limitations, we choose to present one of them. Following the guidelines presented in [4], this scenario is defined as follows. The duration of each connection is 5 minutes. According to our model, it is normalized to 1 time slot. The number of servers (W) is set to 4. The waiting space size ($K - 1$) is set to 29, that is, when all W servers are busy, up to 29 connection requests may be queued. We consider two classes of customers. The first class is denoted by C_2 and is associated with an initial laxity of 2 and an arrival rate of 0.75 customer/slot. The second class is C_{12} and is assigned an initial laxity of $\Lambda = 12$. The arrival rate of C_{12} customers has no impact on the metrics we want to compute. We computed the various metrics for customers of class C_{12}, as function of their initial position. The results are given in Table I, only some sample values are given. We notice that the setup rejection ratio (P_{sr}) remains very low, whereas if the initial position is 25 or higher, then the probability of matching the setup time requirement (P_{cs}) drops slightly below 1.

IV. CONCLUDING REMARKS

In this letter, we presented a detailed mathematical model for a novel optical connection setup management approach that relies on the EDF scheduling algorithm. We used this model to find some of the major performance metrics pertaining to QoS-aware connection setup. Finally, we motivated the use of the proposed management scheme since it allows optical operators to reduce connection blocking probability. This is achieved by exploiting the connection setup time requirement of a connection to queue the blocked connection requests.

REFERENCES

[1] A. Jukan and G. Franzi, "Path selection methods with multiple constraints in service-guaranteed WDM networks," *IEEE/ACM Trans. Networking*, pp. 59–72 vol.12, Feb. 2004.
[2] A. Szymanski *et al.*, "Grade-of-service-based routing in optical networks," *IEEE Commun. Mag.*, pp. 82–87, Feb. 2007.
[3] C. Pinart and G. J. Giralt, "On managing optical services in future control-plane-enabled IP/WDM networks," *J. Lightwave Technol.*, pp. 2868–2876, Oct. 2005.
[4] W. Fawaz and K. Chen, "SSTF: a shortest setup time first optical connections setup management approach with quantifiable success rate," in *Proc. IEEE Globecom'06*, pp. 1–5, Nov. 2006.
[5] R. W. Conway, W. L. Maxwell, and L. W. Miller, *Theory of Scheduling*. Addison-Welsey, 1967.