Reinforcement Learning Based Routing in All-Optical Networks with Physical Impairments

Yvan Pointurier and Fariba Heidari
Department of Electrical and Computer Engineering
McGill University, Montreal, QC
Email: {yvan.pointurier,fariba.heidari}@mail.mcgill.ca

Abstract— **We present and evaluate a reinforcement learning-based RWA algorithm for all-optical networks subject to physical impairments. The technique is suitable for decentralized networks and is compared with other techniques with similar computational complexity.**

I. INTRODUCTION

All-optical networks are a promising technology to provide flexible network management and high rate data transmission by keeping signals in the optical domain from end-to-end without resorting to electronic conversion or regeneration at the network nodes. However, due to the absence of electronic regeneration, signals are subject to accumulating physical impairments during propagation and perfect physical layer is no longer a valid assumption – measuring Quality of Transmission (QoT) through bit-error rates (BER), this means that lightpaths' BERs (a lightpath is a combination of a route and a wavelength) can reach values beyond acceptable values as set by the network manager. Routing and Wavelength Assignment has emerged as a cross-layer technique to route calls in the network over lightpaths accounting for physical impairments, in order to decrease the probability a call is blocked because of the lack of available resources or because establishing the lightpath would cause this lightpath's or another lightpath's QoT beyond the acceptable threshold [1]. In this paper, we propose to use a reinforcement learning technique in a distributed fashion, to choose a tentative lightpath among a set of alternates based only on the past events seen locally. We compare by simulations the reinforcement-learning technique based algorithm with other RWA algorithms that have similar computational complexity.

Alternate routing is known to perform better than fixed shortest path routing in circuit switched routing. Moreover, it was shown in [2] that the time to compute a lightpath and establish a call in an all-optical network with is critical; if the call establishment procedure takes too long then the network state may change during the lightpath computation/call establishment and additional blockings due to outdated information can result. Reinforcement learning has been used in the past to perform routing in a decentralized fashion in telephone [3], circuit-switched (e.g., MPLS) [4], and all-optical networks [5], but QoS impairments have never been accounted for. We propose a routing procedure derived from alternate routing (wavelength assignment is assumed to be "first fit", arbitrarily) where the chosen route for each call is drawn probabilistically from a set of alternates and the probability with which a route is chosen is updated using a reinforcement learning technique based only on past, local information. For each call, our technique selects only one route to minimize the computations needed to verify the route meets the constraints (resource availability and acceptable QoT) and hence the time to establish the call. It is therefore crucial that the right route is chosen so as to minimize the overall call blocking probability.

With reinforcement learning-based routing, each node makes decisions to route calls based on past events (blocking or acceptance of prior calls) observed locally, instead of using the network state. The goal of reinforcement learning is here to decrease the network's call blocking probability by learning online what decisions cause blocking or establishment. Such event-based (as opposed to state-based) routing is desirable because state-based routing imposes high overhead on the network to flood the network status to each node at all time.

II. SYSTEM MODEL

We assume the absence of wavelength conversion and of any electrical regeneration, such that the network we consider is purely transparent. When a call arrives, it must be routed over a lightpath. Calls can be blocked for two reasons: (wavelength continuity constraint) if, on any route, no wavelength is available from end to end between source and destination then the call sustains *wavelength blocking*; (QoT constraint) if no lightpath that meets the wavelength continuity constraint yields an acceptable QoT then the call is rejected due to *QoT blocking*. The overall blocking probability of the network is the probability that a call is blocked due to inability to simultaneously meet both constraints.

We assess the QoT of a lightpath via its BER, which depends on the physical impairments sustained by the lightpath during its propagation though the network. We present the lightpath model used throughout this paper in Fig. 1. At the source node, a laser transmits an On-Off Keyed (OOK) signal over a given wavelength. The signal then propagates in nodes (OXCs), fiber spans and optical amplifiers until it reaches the detector, modeled as a square-law device followed by a filter (filters are not depicted on the figure.) The BER of the signal can then be determined from the distributions of the received "0" and "1" samples. More specifically, designing by μ_0 and μ_1 the means of the received "0" and "1" samples, respectively,
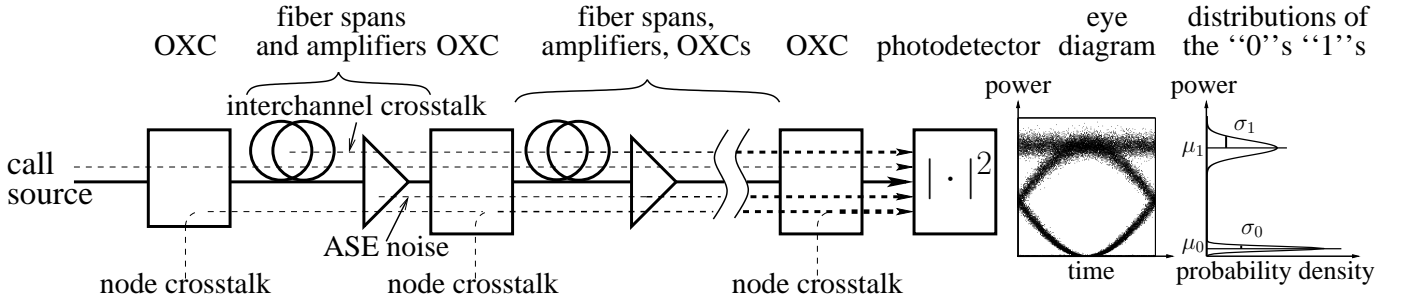
Fig. 1. Model for a lightpath in an all-optical network, and sources of physical impairments: ISI (fiber spans, photodetector), ASE noise (amplifiers), interchannel crosstalk (fiber spans), node crosstalk (OXCs).

Algorithm 1 RWA(s,d)

Initialization (offline):

$\mathcal{R}_{s,d}$: predetermined set of alternate routes between the source s and the destination d.

$P_j^{s,d} = 1/|\mathcal{R}^{s,d}|$: prob. route j is selected ($1 \leq j \leq |\mathcal{R}_{s,d}|$).

Output: Lightpath = pair (route, wavelength).

1: Select a route $R_i^{s,d} \in \mathcal{R}^{s,d}$ with probability $P_i^{s,d}$
2: LP $= \emptyset$, $D_i^{s,d} \leftarrow D_i^{s,d} + 1$
3: Reward $= 0$
4: **repeat**
5:    Select sequentially tentative wavelength λ
6:    Request establishment of LP$' = (R_i^{s,d}, \lambda)$
7:    **if** LP$'$ satisfies both wavelength continuity and QoT constraints **then**
8:       LP $=$ LP$'$
9:       Reward $= 1$
10:    **end if**
11: **until** $\lambda > C$ or LP $\neq \emptyset$
12: $P^{s,d} \leftarrow$ UPDATE_PMF($P^{s,d}, \mathcal{R}_{s,d}, i, $ Reward)
13: return LP

Algorithm 2 UPDATE_PMF

Inputs: P, $\mathcal{R}_{s,d}$: defined in Alg. RWA; i: index of the alternate route to update; Reward: binary reward value.

Parameters: G and B: learning rate parameters.

Output: updated distribution P.

1: **if** Reward $= 1$ (route i was accepted: increase P_i) **then**
2:    $P_i \leftarrow (1-G)P_i + G$
3:    $\forall j \neq i : P_j \leftarrow (1-G)P_j$
4: **else**
5:    (route i was rejected: decrease P_i)
6:    $P_i \leftarrow (1-B)P_i$
7:    $\forall j \neq i : P_j \leftarrow (1-B)P_j + \frac{B}{|\mathcal{R}_{s,d}|-1}$
8: **end if**

III. ROUTING ALGORITHM

We propose a decentralized routing algorithm where, for each call, a route is selected among a set of candidate routes using reinforcement learning techniques. Wavelength assignment is left out of the scope of this paper and is assumed to be "first fit", that is, for a given candidate route, wavelengths are selected in a fixed order. Here, we focus on the routing part of Routing and Wavelength Assignment.

Routing (and wavelength assignment) is performed by Alg. 1 ("RWA"). Before network utilization, a set of alternate routes $\mathcal{R}^{s,d}$ (e.g., the K-shortest paths for some K) is predetermined for each pair of nodes (s,d). Furthermore, each route $R_j^{s,d} \in \mathcal{R}^{s,d}$ is associated to a probability $P_j^{s,d}$ ($\sum_j P_j^{s,d} = 1$). To achieve decentralization, Alg. 1 is instantiated for each (s,d) pair of nodes. During network operation, when a call (lightpath demand) arrives for pair (s,d), node s draws a route among the set $\mathcal{R}_{s,d}$ according to the distribution $(P^{s,d})$; that is, route $R_i^{s,d}$ for a given i is picked with probability $P_i^{s,d}$. Then, for each wavelength λ, the lightpath $(R_j^{s,d}, \lambda)$ is tentatively established. This can be done by means of an appropriate protocol such as RSVP-TE in MPLS networks. During tentative establishment, the wavelength continuity and the QoT constraint are checked (again, this can be done in a decentralized manner by the nodes on the tentative route). If the lightpath meets both constraints, it is accepted and the distribution $(P^{s,d})$ is updated according to Alg. 2 (UPDATE_PMF). If no wavelength λ permits lightpath $(R_j^{s,d}, \lambda)$ to meet the constraints, then the call is blocked.

and by σ_0 and σ_1 their respective standard deviations, we define the Q factor as $Q = (\mu_1 - \mu_0)/(\sigma_0 + \sigma_1)$. Then, using a Gaussian assumption [6], the BER of the signal is $BER = 1/2 \cdot \text{erfc}(Q/\sqrt{2})$. During its propagation, the signal sustains a number of physical impairments which contribute to decrease its BER. We consider four physical impairments: Inter-Symbol Interference, amplifier noise, interchannel crosstalk, node crosstalk. Inter-Symbol Interference (ISI) and amplifier noise (ASE noise) depend only on the hardware (receiver filter, number of amplifiers, ...) present over the lightpath. On the other hand, interchannel crosstalk (resulting from the nonlinear interaction within fiber spans of several signals co-propagating on different wavelengths) and node crosstalk (resulting from optical leaks in OXCs, for instance due to imperfect demultiplexing) depend on the network state. Indeed, crosstalk is generated by the interaction of several lightpaths. In this work, we use the physical layer model we presented in [1]. In particular, we refer the reader to [1] for efficient techniques that relate the impact of ISI, ASE noise, interchannel and node crosstalk to μ_0, μ_1, σ_0, σ_1 and hence BER.
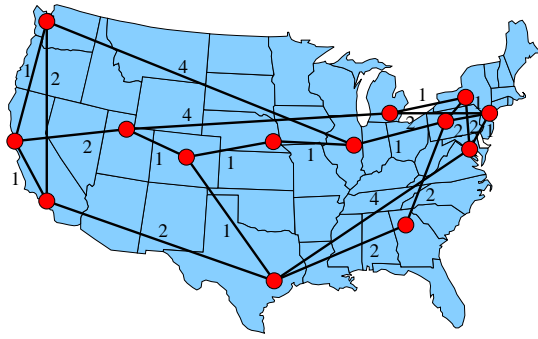
Fig. 2. NSF topology used for the simulations. The weights on the links are the number of 70km spans. We simulate the network operation for 5000 s for various offered loads and repeat each experiment 8 times to obtain 95% confidence intervals.

Notice here that, in order to reduce the amount of computations, we require that only one tentative route can be picked, thereby limiting the number of tentative lightpaths to the number of wavelengths. Other variants of this scheme, e.g, allowing drawing another route (again according to the distribution $P^{s,d}$) if all tentative lightpaths for the route originally picked are blocked, are possible and left for future work.

We now give more details about the reinforcement learning algorithm (Alg. 2) we use to update the distributions $P^{s,d}$: "Linear Reward-ϵ Penalty" (LRϵP) [7]. When a call is accepted on a route i, LRϵP rewards call establishment success by linearly increasing the associated probability $P_i^{s,d}$. When a call is rejected, LRϵP penalizes call establishment success by linearly decreasing the associated probability $P_i^{s,d}$. The algorithm is parametrized by two gains G and B (corresponding, respectively, to reward and penalty), as can be seen in Alg. 2. LRϵP is known to perform well in non-stationary environments and avoids remaining stuck in absorbing states, and is therefore adapted to the setup described in this paper.

IV. SIMULATION RESULTS

We evaluate the RWA technique on a scaled version NSF topology depicted in Fig. 2 (in the original NSF network, distances are too long to allow transmission with acceptable quality between all pairs of nodes even when only ISI and ASE noise are present.) All links are bidirectional and can carry up to $C = 8$ wavelengths per direction. The physical parameters are chosen to model large-scale metropolitan or medium-scale regional networks (SMF fiber, full post-dispersion compensation, 6 dB noise factor amplifiers, 2 mW super-Gaussian NRZ pulses modulated at 10 Gbps, 25 GHz grid spacing) and the maximum acceptable BER is set to 10^{-9}. Calls between each pair of nodes (s, d) arrive according to a Poisson process with rate $\ell^{s,d}$ and call durations are exponentially distributed with unit mean, such that $\ell^{s,d}$ is the offered load in Erlang from s to d. The set of alternate routes $\mathcal{R}^{s,d}$ is fixed to the set of the 4 shortest routes (in distance) between s and d. The learning parameters G and B are set to 0.01 and 0.001, respectively.

In Fig. 3, we compare the algorithm presented in Section III (diamonds) with two other RWA algorithms: shortest path (the only alternate path is the shortest path between any two nodes;
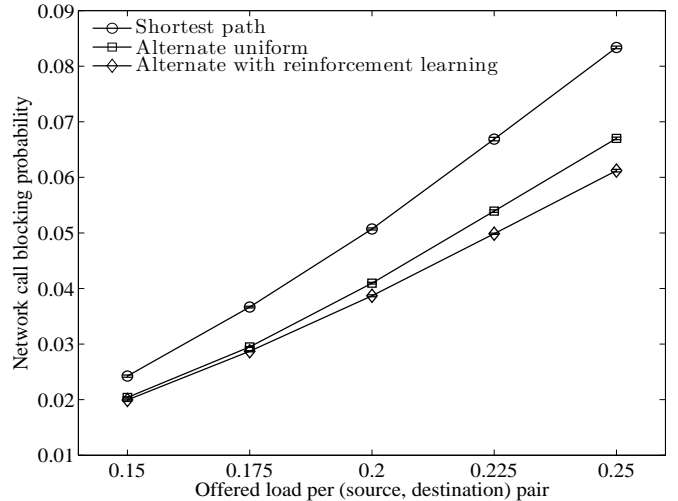

Fig. 3. Blocking probability for the downscaled NSF topology.

circles), and alternate-uniform where an alternate route is picked randomly uniformly (instead of being picked according to the load sharing factors; squares). Alternate-random thus does not implement any reinforcement learning technique. All three algorithms have the same "computational complexity" in the sense that they put the same burden on the network when a call arrives to check whether the two constraints are met; indeed for each algorithm a single route is chosen for each call arrival, hence the constraints for at most C lightpaths are checked per call arrival. Alternate-uniform yields lower call blocking probabilities than fixed shortest path event if a single alternate is picked randomly. Our proposed algorithm performs better than alternate-uniform for the higher loads, and similar to alternate-uniform for the lower loads.

V. CONCLUSIONS

We applied a reinforcement learning technique, and more specifically, the LRϵP algorithm, to the problem of low-complexity, distributed route selection in all-optical networks subject to physical impairments. The technique is shown by simulation to perform well. This paves the way for more research, including studying the convergence behavior of the technique to changes in the network conditions such as link failure or modifications in the offered loads.

REFERENCES

[1] Y. Pointurier, M. Brandt-Pearce, T. Deng, and S. Subramaniam, "Fair QoS-aware adaptive Routing and Wavelength Assignment in all-optical networks," in *Proc. IEEE ICC*, 2006.

[2] J. He, M. Brandt-Pearce, and S. Subramaniam, "QoS-aware wavelength assignment with BER and latency guarantees for crosstalk limited networks," in *Proc. IEEE ICC*, 2007.

[3] K. Narendra, E. Wright, and L. Mason, "Application of learning automata to telephone traffic routing and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 785–792, 1977.

[4] F. Heidari, S. Mannor, and L. Mason, "Reinforcement learning-based load shared sequential routing," in *Proc. IFIP-Networking*, 2007.

[5] K. Lukac, Z. Lukac, and M. Tkalic, "Wavelength routing in future WDM networks," in *Proc. IASTED Artificial Intelligence and Soft Computing*, 2005.

[6] G. Agrawal, *Fiber-Optic Communications Systems*. John Wiley & Sons, Inc., 2002.

[7] K. Narendra and M. Thathachar, *Learning Automata*. Prentice-Hall, 1989.