Reinforcement Learning for Provisioning OTN Leased Lines

Ashwin Gumaste[1], Joao Pedro[2] and Harald Bock[3]
[1]Infinera Corporation San Jose, CA USA, [2]Infinera Unipessoal Lda, Portugal, [3]Infinera GmbH Munich Germany
agumaste@infinera.com, jpedro@infinera.com, hbock@infinera.com

Abstract: We discuss reinforcement learning-based strategies for provisioning OTN leased-lines including all-optical provisioning and strategic aggregation subject to stochastic traffic in metro/regional networks. We show benefit in transceiver count and reduction of OTN cross-connect capacity. © 2022 The Author(s).

1. Introduction

Leased-line (LL) traffic that constitutes much of wholesale bandwidth and latency sensitive financial traffic has followed the growth trend of IP traffic. Much of LL traffic is provisioned as OTU-k lines, due to requirements of low-latency, high-availability, reliability, and maintaining separation from IP-traffic. As bandwidth requirements for LLs grow, they hit an interesting inflection point, where service bandwidth is now increasingly close to wavelength rates, especially in the metro and regional networks. Concurrently, one also observes that the ASICs that powered OTN cross-connects (XCs) [1,4] and used for provisioning OTN traffic, have not undergone the same type of evolution as compared to the corresponding process evolution in IP routers. While using routers to provision leased-lines could be a possibility, this approach is unlikely to gain immediate traction among providers and their customers, who have always expected deterministic delay, higher availability, etc. than what the IP network has offered. In contrast, utilizing the optical layer to provision LLs as wavelength services, especially those that are close to wavelength rates, could be an option. Naturally, questions such as restoration and provisioning need to be answered when one considers an optical layer provisioning scheme. Assuming we can provision and efficiently restore, then one idea can also be that we extend the optical provisioning scheme to not just those services that approach wavelength granularity, but also between strategic aggregation points in a network, thus alleviating the need for further enhancements in OTN cross-connects. The question as to which demands should be provisioned all-optically, which should be aggregated at the network edge, and which should be subject to strategic in-path aggregation under stochastic conditions leads to a reinforcement learning (RL) problem. Of interest is to compute a policy that allows strategic aggregation, optical provisioning and traditional OTN support, while reducing overall transceiver count and XC size. We propose RL-based schemes for provisioning traffic using a combination of pluggables (OpenZR/ZR+), muxponder optical engines (OE) and OTN XCs.

Consider the schemes in Fig. 1 below. In Fig. 1a (the vanilla scheme), LL traffic is aggregated by an OTN muxponder at the network edge, which is then sent to the core that uses OTN XCs for service provisioning. In Fig. 1b, (strategic aggregation case) the OTN traffic from various access nodes is groomed at strategic locations to form higher-rate interim OTN connections. When such aggregation is done at the edge or metro-core boundaries, then this model represents the present mode of OTN operation. However, if we are to select aggregation nodes based on traffic profiles or near-optimal aggregation requirements, then we have to define a good aggregation node selection strategy (policy) subject to stochastic traffic growth. Finally, there is the model in Fig. 1c, whereby the LL capacity of *some* of the demands is such that they can occupy a full wavelength (such as an OTU4) and is hence all-optically provisioned as a wavelength service. The fourth model is a combination of the second and third models, in which lower granularity demands are strategically aggregated, and higher bandwidth demands continue to be all-optically provisioned.

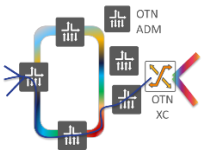

Fig. 1. (a) Vanilla aggregation model.

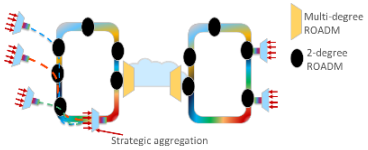

Fig. 1(b) Strategic aggregation.

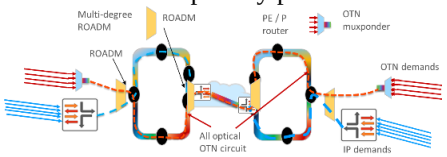

Fig. 1(c). All optical provisioning.

2. The Reinforcement Learning Scheme Description

The network is defined as a tuple of transponders, pluggables, muxponders, demands, routers and OTN XCs denoted by $S_t = \{Tr, P, D, R, X\}$ at each node in a graph $G(V, E, C)$, of vertices, edges and spectral capacity (C). We consider the arrival of a batch of traffic demands and its provisioning in a network (including selecting the model from the 4 aforementioned ones), as an episode. A traffic demand d_{ij} between source i and destination j, of capacity c_{ij} is routed

generally on a shortest path, with a node and edge disjoint protection path. LL demands are OTU0,2,4 and c4 and are muxponded as and when required. Muxponders are interfaced either with pluggables or optical engines both of which have a reach according to the reach tables presented in [2]. We assume that there is an agent, that develops a policy which learns a set of actions based on data from history and its inference of the system. We model both off-policy and on-policy learning. Each time there is either a new request, or a change in bandwidth, the agent must take an action. The action taken is from the set $A = \{prov, drop, aggregate, node_{agg}\}$, indicating either to provision the demand as is (a wavelength service), or drop the demand altogether/partially, or aggregate this demand/change in demand on to an existing connection/muxponder and if so, then choose the optimal node(s) at which such aggregation can happen. The system moves from its current state to its next state after it takes action from among the list in A at state s, denoted by A_s. We associate the state-action pair $q(s, a)$ to denote the action at a state s, and our goal is to compute an optimal policy π^*, with optimal value function $v_\pi^*(s)$ which denotes the value of a state s under policy π^*. The terminal state for a demand is when it cannot find any path due to wavelength exhaustion on a link of that path (even after XCs are placed for wavelength translation). For each provisioned demand using a wavelength (no aggregation), we give a reward of 2 units (as we avoid interim XCs), while a demand that is aggregated gets a reward of 1 unit (reward for using existing infrastructure). If it does happen that a demand which is provisioned on a wavelength must be torn down, and aggregated into a higher capacity channel, for efficient spectral usage, we give a negative reward of -4 units. An example of this is a OTU4 provisioned as a wavelength, but due to spectral exhaustion, must be torn down and muxponded into an existing 400Gb/s channel, thereby freeing up spectrum (the whole act of tearing down and re-provisioning the service being highly undesirable and hence the high penalty). Optical provisioning is possible only for wavelength granular demands (100Gb/s and multiples thereof). We use three temporal difference tools for comparison: off policy Q-learning, on-policy SARSA and Expected SARSA (E-SARSA) [3]. In each of the three methods we follow principle of Generalized Policy Improvement (GPI), in which we first evaluate a policy and then improve it, in a repetitive manner till the optimal policy v_π^* is found. In the off-policy model, we play a large number of episodes (on the same network data as actual) and let the agent learn different scenarios. The agent is further aided by training data generated by an ILP [2] at discrete time-intervals. In the on-policy model, we select a greedy policy and with a small probability explore other policies but without having to worry about specific starting states.

From a learning perspective, the goal of our model is to minimize mean-squared value error (VE) resulting from the computation of approximate value function $\hat{v}(S_t, \boldsymbol{w_t})$; where the parameterized weight vector $\boldsymbol{w}$ consists of the tuple that describes the network state S_t, and in addition the OTN XC locations, their granularities, muxponder locations and their granularities. The weight vector $\boldsymbol{w}$ is updated via the stochastic gradient descent (SGD), using the following relationship:

$$\boldsymbol{w_{t+1}} = \boldsymbol{w_t} + \alpha[v_\pi(S_t) - \hat{v}(S_t, \boldsymbol{w_t})]\nabla\hat{v}(S_t, \boldsymbol{w_t}), \text{ where, } \nabla f(\boldsymbol{w}) = \left(\frac{\delta f(\boldsymbol{w})}{\delta w_1}, \frac{\delta f(\boldsymbol{w})}{\delta w_2}, \ldots, \frac{\delta f(\boldsymbol{w})}{\delta w_d}\right)^T \quad (1)$$

Note that the value function is approximate, and always for our computation is linear in chosen weights by using a feature vector that represents a state s. To linearize, we compute feature vector, x_t, short for $x(S_t)$. Since the target output may not be a true value of $v_\pi(S_t)$, we approximate it with U_t as an unbiased estimate resulting from SGD with $\boldsymbol{w_{t+1}} = \boldsymbol{w_t} + \alpha[U_t - \hat{v}(S_t, \boldsymbol{w_t})]\nabla\hat{v}(S_t, \boldsymbol{w_t})$, leading to our learned vector.

$$\boldsymbol{w_{t+1}} = \boldsymbol{w_t} + \alpha[v_\pi(S_t) - \hat{v}(S_t, \boldsymbol{w_t})]\mathrm{x(S_t)} \quad (2)$$

For our analysis, we will assume step size α proportional to the traffic arrival change and the probability of being in a state proportional to the aggregation possibilities into and out of that state.

3. Results

We applied the reinforcement learning model to 5 networks: 2 regional metro+access, and 3 metro+access networks with 276-1652 nodes. The network specifics are shown in Table 1. Computations were done using the TensorFlow and Networkx libraries. LL traffic is modelled from 24 to 381 Tb/s in a 3-year period across the networks. Though we consider 3 years of traffic growth, the model assumes incremental growths with an average of 40-60% year on year growth. This allows us to build many episodes from which the model also learns to adapt a policy (aggregate/provision). The traffic is modelled as a ratio of 5:10:1 for OTU0s, OTU2s and OTU4s randomly distributed across nodes. XC sizes are in increments of 600Gb/s, up to a max of 25.6Tb/s. An ILP-based model is also used to derive instance-wise optimal results and also provides for starting states for the learning. Feature vectors are developed using learning across 10^4 episodes for each RL type (Q-learning, SARSA and E-SARSA). We apply the RL models to the 5 networks by considering the use of pluggable ZR+ or muxponders with optical engines (OEs). The RL models learn where to place aggregators, and which demands should be provisioned all-optically.

Network	Avg Path L (km)	Avg hop	Core nodes	Rings	Number of Metro nodes	Yr 1 traffic*	Yr 2 traffic*	Yr 3 traffic*
MN-1	892	3	52	50	276	24	35	56
MN-2	1299	3.2	75	60	310	21	28	37
MN-3	275	2.8	4	32	172	24	31	42
RN-1	320	3.1	24	276	1652	234	279	333
RN-2	355	3.3	40	312	1578	176	272	381
								* in Tb/s

Table 1: Network characteristics.

Shown in Fig. 3 is the number of transceivers for all techniques by using only ZR+ interfaces averaged over the 3-year period. This results in more regens, also implying more possible aggregation sites. The vanilla aggregation is the worst performing, followed by aggregation in selected sites (strategic aggregation), then the optical provisioning scheme and finally the best performing is the strategic aggregation with optical provisioning. The difference between the last 2 schemes is within 8%, while the difference between the vanilla aggregation and the strategic aggregation+optical provisioning scheme is 25%. In Fig. 4 is the comparison across the schemes when we consider both ZR+ and OE muxponders, which are particularly useful for x100Gb/s longer distance all-optical connections without regens. Note that the results in Fig. 4 better those that of Fig. 3 by an average of 14%. With OE muxponders, strategic aggregation+optical provisioning is on average 31% better than the vanilla aggregation scheme.

Shown in Fig. 5 is the spectral availability post provisioning (how much bandwidth on average per link in the network is available) across the schemes. Strategic aggregation performs best, followed by vanilla aggregation, then optical provisioning with strategic aggregation and finally only optical provisioning. Even then, the impact of optical provisioning is only on average 11% worse than more expensive aggregation. Shown in Fig. 6 is the average XC size averaged across all nodes in the 5 networks for the 4 schemes. With optical provisioning schemes we need smaller XCs (by 36.8% over vanilla aggregation), and when strategic aggregation is added (it results in a further decrease in XC size by 40.2%). Finally in Fig. 7 is the learning error between the ILP target and our schemes also compared to vanilla aggregation (heuristic) scheme. Error is computed across 19 data-points corresponding to increase in load.

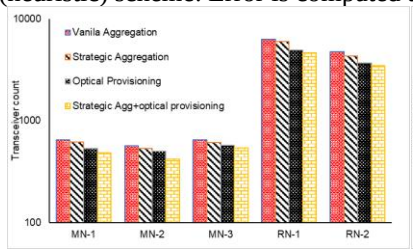

Fig. 3. Transceiver count (ZR+).

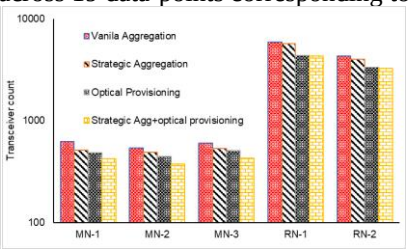

Fig. 4. Transceiver count (OE).

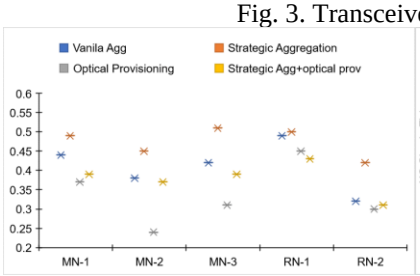

Fig. 5. Spectral availability.

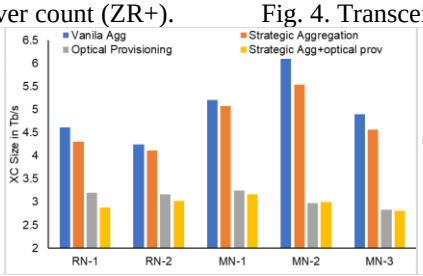

Fig. 6. XC size

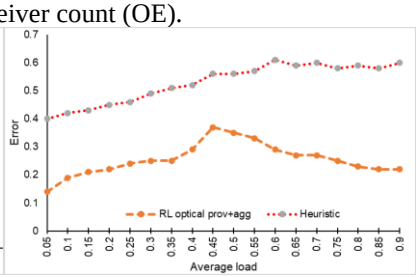

Fig. 7. Learning error.

4. Conclusion

We have compared RL-based approaches towards provisioning OTN leased-lines, and conclude that a combination of strategic aggregation with optical layer provisioning is effective from reducing transceiver count (by 25%) and XC size by 36% across 5 different metro and regional networks with stochastic traffic demands.

References

[1] W. Bigos, B. Cousin, S. Goddlin, M. Foll and H. Nakajima, "Survivable MPLS over optical transport networks: cost and resource usage analysis," IEEE Journal of Select. Areas in Commun. Vol 25 No 5, June 2007.

[2] A. Gumaste, M. Sosa, H. Bock, P. Kandappan, "Optimized IP-over-WDM core networks using ZR+ and flexible muxponders for 400 Gb/s and beyond" IEEE OSA Joun. of Optical Commun. Networks, Vol. 14, No 3, March 2022.

[3] R. Sutton and A. Barto, "Reinforcement Learning: An Introduction," 2nd Ed. MIT Press 1998.

[4] J. Santos, et al., "Optimized provisioning of 100-GbE services over OTN based on shared protection with collocated signal regeneration and differential delay compensation," Proc of OFC 2012. OTh4B, Los Angeles March 2012.