Q-Learning Based Energy-Efficient Network Planning in IP-over-EON

Pramit Biswas, Md Shahbaz Akhtar, Aneek Adhya, Sriparna Saha, and Sudhan Majhi

Abstract—During network planning phase, optimal network planning implemented through efficient resource allocation and static traffic demand provisioning in IP-over-elastic optical network (IP-over-EON) is significantly challenging compared with the fixed-grid wavelength division multiplexing (WDM) network due to increased flexibility in IP-over-EON. Mathematical optimization models used for this purpose may not provide solution for large networks due to large computational complexity. In this regard, a greedy heuristic may be used that intuitively selects traffic elements in sequence from static traffic demand matrix and attempts to find the best solution. However, in general, such greedy heuristics offer suboptimal solutions, since appropriate traffic sequence offering the optimal performance is rarely selected. In this regard, we propose a reinforcement learning technique (in particular a Q-learning method), combined with an auxiliary graph (AG)-based energy efficient greedy method to be used for large network planning. The Q-learning method is used to decide the suitable sequence of traffic allocation such that the overall power consumption in the network reduces. In the proposed heuristic, each traffic from the given static traffic demand matrix is successively selected using Q-learning technique and provisioned using the AG-based greedy method.

Index Terms—Elastic optical network, Reinforcement learning, Power consumption.

I. INTRODUCTION

With the increase in adoption of data-intensive services, such as ultra-high-definition video streaming, cloud gaming and virtual and augmented reality video streaming, global IP traffic is anticipated to increase 3-fold from 2017 to 2022 [1]. With the increase in communication network traffic, global energy consumption in the network is expected to grow to a staggering level of 21% of the global electricity consumption by 2030 [2]. Achieving energy-efficiency in communication networks is not practicable without paying due attention to the energy efficiency in optical backbone networks. In view of this, in this paper, we focus on energy efficiency in optical backbone networks.

We aim to minimize the power consumption (PC) in optical backbone network during network planning phase through efficient resource allocation and traffic provisioning taking into consideration the expected (static) traffic demands of node pairs. Elastic optical network (EON) architecture enabling the flexible grid (spectrum) allocation and use of flexible transponders [i.e., sliceable bandwidth variable transponders (SBVTs)], orthogonal sub-carriers and adaptive modulation schemes is considered to be prospective next-generation optical backbone network architecture [3, 4]. EON allows coexisting of multiple lightpaths with different capacities, spectrum, maximum transparent reaches (MTRs) and related PCs at SBVT in

each unidirectional fiber[1]. In IP-over-EON architecture, IP layer is integrated with EON, and access network traffic from multiple sources are groomed through electrical layer traffic grooming in IP-core routers for onward transmission through EON [5]. Energy efficiency in the network can be improved by exploiting the flexibility and reconfigurability of IP-over-EON. In the network planning phase, resource and traffic provisioning are implemented employing static traffic demand matrix, which is typically obtained from the long-term average traffic demands of node pairs or a predetermined percentile of the peak traffic demands between the nodes.

In IP-over-EON, intelligent traffic provisioning may be considered as one of the approaches to improve energy efficiency in the network. The complexity to obtain the optimal network planning increases many-fold for IP-over-EON compared with the fixed-grid wavelength division multiplexing (WDM) networks due to increased flexibility in IP-over-EON. To provision traffic demands (elements) from the static traffic demand matrix, a mathematical optimization model may be used for optimal network planning [6–8]. However, the optimization model may not provide solution for large networks in presence of all related constraints due to large computational complexity. Thus, a greedy heuristic may be used that intuitively selects one traffic demand at a time (i.e., at each step) and routes the traffic [7, 9]. The process is repeated for all traffic elements in the traffic demand matrix with the traffic elements selected in a sequence, as the heuristic attempts to obtain the best solution. However, the greedy heuristic has only one chance to select and route a traffic, and never reviews the decision taken at an earlier step. It may offer the best performance for specific traffic sequence(s) only, and any other traffic sequence will yield suboptimal solution. As for example, Zhao *et al.* show that different sequences of traffic demands, such as traffic demands with decreasing order of data-rate requirement or decreasing order of the number of fiber links along the shortest path between the source and destination (SD) nodes may change the network performance in terms of the maximum allocated sub-carrier indices on a fiber [10]. Thus, in the planning phase, there lies opportunity for optimization of network performance by deciding the best sequence of static traffic demands to be selected. If in a given network with N number of nodes, traffic demands exist among all nodes in the static traffic demand matrix, i.e., $N(N\text{-}1)$ number of traffic exist (excluding the diagonal elements), there can be $[N(N\text{-}1)]!$ numbers of possible sequences following

---

[1]In this paper, unless stated otherwise, a fiber represents a bidirectional fiber with two unidirectional fibers in the opposite directions.

which the traffic demands can be provisioned. Therefore, the search space is significantly large for moderate to large value of N. An analogy can be made between this problem and finding the shortest Hamiltonian path problem, which is considered to be NP-hard. A Hamiltonian path ensues that travelling along the path, all nodes in the graph are visited only once, similar to provisioning all traffic elements only once. However, one important difference between the two problems is that for the shortest Hamiltonian path, the distance between any two nodes (i.e., the related cost in traversing between the nodes) is fixed irrespective of the order (i.e., sequence) of traversal of the nodes, whereas in case of network planning problem, the cost value is not fixed, and it depends on the existing network condition, i.e., the existing lightpath status, resource availability etc.

Reinforcement learning (RL) is a category of machine learning (ML) technique that may be used to make a sequence of decisions. Watkins first propose Q-learning algorithm [11], which is an RL technique, and the convergence of the algorithm is proved in [12]. Gambardella *et al.* propose Q-learning based algorithms in a weighted complete graph to find the shortest Hamiltonian tours, analogous to determining the sequence of static traffic demands to be selected for optimal provisioning [13]. For efficient network planning, we explore a Q-learning based technique, which is a reward based trial and learn method and does not require any labeled data.

In this paper, we explore energy-efficient network planning for IP-over-EON using static traffic demands by deciding the best sequence of traffic to be provisioned using Q-learning technique combined with a greedy heuristic. Each traffic from the given static traffic demand matrix is successively selected using Q-learning technique and provisioned using greedy heuristic. The process is repeated for the traffic demand matrix for multiple times, and the best network planning is identified. We use an auxiliary graph (AG)-based energy-efficient greedy heuristic to provision the selected traffic with the least increase in PC. For identification of traffic sequence, Q-learning based algorithm is used, where an agent predicts an action, i.e., the next traffic to be provisioned from the past experiences, and gains reward based on the PC needed for provisioning. From the received rewards, the agent tries to develop the optimal policy that helps to decide the best sequence for energy-efficient traffic provisioning. The performance of the proposed Q-learning and AG based energy-efficient network planning for IP-over-EON heuristic, referred to as QAG-ENP-IoE, is assessed with realistic network setting. Moreover, we also use virtualized elastic regenerators (VERs) to enhance flexibility, connectivity, and improving energy efficiency in the network.

In Section II, related literature are presented. Next, in Section III, a brief description of IP-over-EON architecture and PC model for network equipment are presented. In Section IV, the proposed heuristic is described, while in Section V the performance of the heuristic is studied. Finally, Section VI concludes the paper.

II. RELATED WORK

Klinkowski *et al.* propose an integer linear programming (ILP) based optimization model to solve routing and spectrum allocation (RSA) problem with the objective to minimize the use of spectrum resources, i.e., frequency slots (FSs) for optimal network planning in EON using static (offline) traffic demands [14]. Since the optimization model is NP-hard, a greedy heuristic is proposed for large problem size where traffic demands are allocated in the decreasing order of requested FSs. Zhang *et al.* propose a heuristic to maximize optical layer traffic grooming in IP-over-EON facilitating simultaneous generation/termination of multiple lightpaths of different capacities, FSs and MTRs by a single SBVT, so as to minimize the overall PC [7]. The proposed heuristic provisions traffic in descending order of the requested bandwidth. The authors in [9] propose an AG-based heuristic for energy efficient network planning in IP-over-EON, where the sequence of traffic to be processed is determined following the descending order of traffic bandwidth. Furthermore, a pruning strategy is used to further reduce the PC wherever possible. Ramaswami *et al.* propose an mixed ILP (MILP) model for network planning in fixed-grid WDM networks [15]. The MILP model is further decomposed into virtual topology design and traffic routing sub-problems, and the solution from the first sub-problem is provided as input to the second sub-problem so as to improve computational tractability for large networks. Zhao *et al.* explore an ILP formulation to solve nonlinear impairment-aware RSA problem for EON with the objective to reduce the maximum index number of FSs to be used on a fiber [10]. In this regard, simulated annealing based heuristic is also explored for large problems following three different traffic demand ordering policies, viz., decreasing order of traffic demands, decreasing order of the number of links along the shortest paths, and decreasing order of the product of shortest path length and traffic demand [10]. In [16], with reference to dynamic traffic provisioning in IP-over-EON, different job scheduling strategies (i.e., identifying sequence of traffic demands) are adopted based on the bandwidth and holding time criteria, and their impact on the overall energy efficiency is studied.

Musumeci *et al.* discuss different application areas in optical networking domain, such as path computation, dynamic traffic prediction, failure management, and quality of transmission (QoT) estimation, where ML techniques can be used [17]. Salani *et al.* present an ML classifier based QoT estimator to estimate parameters in transmission reach constraints of an ILP model, in order to solve the routing, modulation format and spectrum assignment problem [18]. Following iterative approach, the method excludes lightpaths with poor QoT as estimated by the ML classifier, until either a feasible solution for all lightpaths is found, or the upper limit of iteration count is reached. In [19], to achieve a fast network recovery from an IP node failure in IP-over-EON, a Q-learning based recovery algorithm is presented. Kiran *et al.* propose algorithms based on Q-learning to solve path selection and wavelength selection in optical burst switch (OBS) networks with objective to minimize the burst loss probability [20]. The algorithm is used to select path and wavelength from a set of pre-computed paths and a set of wavelengths, respectively. In case of IP node failure in 5G and beyond 5G (B5G) IP-over-optical network, Gu *et al.* use Q-learning based algorithm to re-configure the

optical layer in view of service recovery, without requiring to re-route the affected traffic flows individually [21]. The proposed method helps in mitigating exhaustive IP forwarding and routing requirement. To the best of our knowledge, identifying the sequence of traffic for network planning in IP-over-EON using any ML technique has not been studied till date.

III. NETWORK ARCHITECTURE AND PC MODEL FOR NETWORK EQUIPMENT

The IP-over-EON architecture has a partial mesh topology where nodes are connected with fibers, and fibers have inline optical amplifiers installed with a fixed span (regular interval) of distance L. As shown in Fig. 1, at each node, a post-amplifier is connected with the outbound unidirectional fiber, whereas a pre-amplifier is connected with the inbound uni-directional fiber. An optical amplifier placed at a location is composed of two unidirectional amplifiers operating at opposite directions, along with the related electronic circuit. Thus, optical amplifiers are treated as bidirectional. We consider the following PC model for an optical amplifier that each optical amplifier has a constant overhead PC of P_{A_O} along with a fixed PC of P_{A_d} for each unidirectional amplifier [9, 22]. As shown in Fig. 1, each node has one IP core router where each port has a fixed capacity of C_R and a fixed PC of P_R [9, 23]. Each router port is connected to a SBVT that is equipped with multiple sub-transponders with each sub-transponder supporting transmission/reception of one lightpath. Each lightpath is characterized by multiple attributes, such as the capacity, the MTR, the number of data-slots and the required PC at SBVT. We consider that different lightpath transmission options are available at SBVT following Table I where each data-slot is considered to be of 12.5 GHz [9, 24, 25]. PC to transmit/receive a lightpath at SBVT is considered to be half of the PC of an active sub-transponder [7, 9]. Each SBVT is connected to an add-drop port of a bandwidth variable optical cross-connect (BV-OXC). A BV-OXC has the same PC model as that of the OXC used in fixed-grid WDM networks. The PC in a BV-OXC is considered to be $(135.d+150)$ W for 100% add-drop facility, where d is the physical degree of the node accommodating the BV-OXC. Some of the nodes may also have one or more VERs [26, 27] connected to BV-OXC for 4R-regeneration (re-shaping, re-timing, re-amplifying, and re-modulation) of lightpaths and merging of FSs when routes of multiple lightpaths after regeneration are the same. A VER consists of a splitter, a coupler and an array of spectrum selective regenerators (SSRs) [26, 27]. We consider that PC in a VER depends on the lightpaths being regenerated through SSRs, where PC for lightpath regeneration is same as the summation of the PC for termination and generation of the same type of lightpath through sub-transponders in SBVT (as shown in Table I) [24]. We consider that overhead PC for each SSR is P_{V_S}, and the number of SSRs used in a VER depends on the number and type of lightpaths regenerated. The overhead PC for each VER is considered to be P_{V_O} [8]. The total PC in a VER is the summation of the PC for the SSRs used and the overhead PC [8].
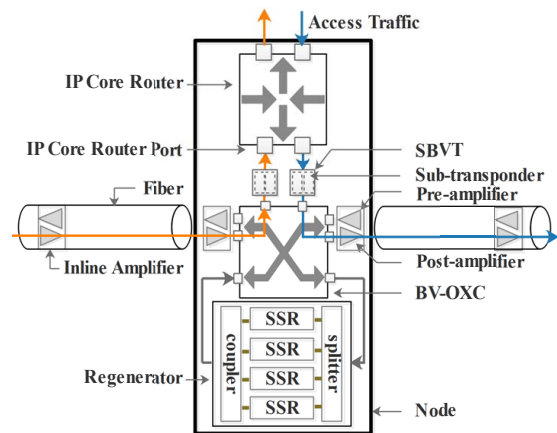

Fig. 1: An IP-over-EON node architecture.

TABLE I: Multiple transmission options for sub-transponder of an SBVT and VER [9, 24, 25]

Capacity (Gbps)	MTR (km)	Data Slot	PC (W)	Capacity (Gbps)	MTR (km)	Data Slot	PC (W)
40	600	1	154.8	100	600	1	198
	1900	1	183.6		1900	1	270
	2500	2	183.6		2500	2	270
	3000	3	183.6		3000	3	270
	4000	4	183.6		3500	4	270
200	500	1	333	400	500	4	432
	600	2	333		600	6	432
	750	3	333		750	8	432
	1900	4	432		1900	10	630
	2200	5	432		2200	12	630
	2500	6	432		2500	14	630

IV. PROPOSED HEURISTIC: QAG-ENP-IoE

A. Problem Description

In the proposed heuristic, our objective is to minimize the overall PC in an IP-over-EON with the given network topology and static traffic demands of node pairs. The outcome (solution) of the heuristic is to determine the network resources (viz., IP core router ports, SBVTs and VERs) to be provisioned at the appropriate node locations, the setting up of lightpaths and allocation of traffic demands through lightpaths (i.e., traffic routing). The PC in the network is also computed. Typically, for this purpose, mathematical optimization models are used for small networks, whereas greedy heuristics are used for large networks. In this study, Q-learning technique integrated with a greedy heuristic is explored for large networks so as to reduce the overall PC as estimated using any greedy heuristic. The greedy heuristic we use for this purpose is discussed in the following.

B. Greedy Heuristic

We use modified AG-based energy-efficient network planning for IP-over-EON (mAG-ENP-IoE) heuristic (Algorithm 1) based on [9] as the greedy heuristic. Using the heuristic we provision the selected traffic demands, one at a

time, in an energy-efficient manner. A brief description of the heuristic is presented in the following.

For a given traffic demand, first, we determine the capacity of the lightpath (that may required to be set up) as the closest (equal or higher) transmission rate supported by an SBVT with reference to the traffic demand. Thereafter, with reference to the traffic demand we construct an AG (Fig. 2), as also described in [9]. In this regard, we consider a physical node to be composed of an electrical layer auxiliary node (AN) A_E, and multiple optical layer ANs, viz., A_1, A_2, etc., corresponding to different available transmission options in SBVTs. AG (Fig. 2) is constructed using different types of edges, viz., transmission edges, Tx/Rx edges and lightpath edges. Transmission edge is set up between two similar type optical layer ANs located at two different physical nodes if the required FSs for possible lightpath set up are available in the connecting fiber. The transmission edge weight is represented by the PC of optical amplifiers to be used on the fiber. Tx/Rx edge is set up from (to) an electrical layer AN to (from) an optical layer AN within the same physical node if the sub-transponder with required capacity is available at SBVT. The related edge weight is represented by the PC of IP core router and SBVT to originate (terminate) the lightpath. Lightpath edge is set up between two electrical layer ANs located at two different physical nodes if free capacity in the existing lightpath between the two physical nodes is available to provision the traffic demand. Lightpath edge weight is represented by the PC to accommodate the traffic demand to an existing lightpath. As for an example, in Fig. 2, we show three physical nodes A, B and C, with distances between A and B, and B and C of 1500 km and 900 km, respectively [9, 24]. We consider availability of only two transmission options in the network with MTRs of 1000 km and 1500 km. We show several AG edges: Tx/Rx edges (viz., A_E-A_1, A_E-A_2 etc.), lightpath edge (viz., A_E-B_E) and transmission edges (viz., A_2-B_2, B_2-C_2 etc.)

Next, using the AG, the shortest path, i.e., the path with the minimum PC between the electrical layer ANs of the SD nodes of the traffic demand is determined. If any path is not available, the traffic demand is split into two traffic elements between the same SD pair, such that at the least one traffic element becomes the closest transmission rate supported by an SBVT, and the algorithm starts again from Line 1 for possible provisioning of each split traffic one by one. If a path is available and the path consists of one or more lightpath
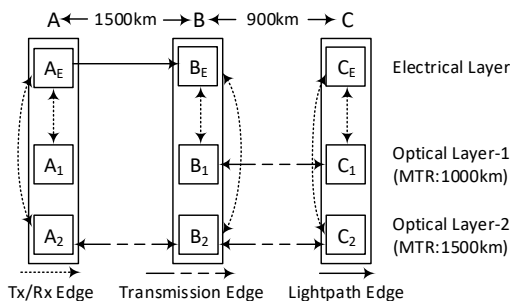
edges, the traffic is groomed with the current traffic in the existing lightpath(s). On the other hand, if the path consists of one or more Tx/Rx edges, new lightpath(s) are set up with due consideration of possible utilization of VER in the path to improve energy-efficiency. Finally, the traffic demand is accommodated and the increase in PC due to provisioning of the traffic demand through the path is computed. If the traffic demand is provisioned fully, $true$ is returned; else, $false$ is returned.

Algorithm 1: Modified AG-based energy-efficient network planning for IP-over-EON (mAG-ENP-IoE)

---

1 Set the capacity of lightpath that may need to be set up between SD node pair.
2 Construct AG.
3 Find shortest path between SD node pair.
4 **if** *no path available* **then**
5   Split traffic demand into two traffic elements and start again from Line 1 to provision all split traffic.
6 **else**
7   **if** *Lightpath edge(s) appear in the path* **then**
8   Groom traffic element in the existing lightpath(s).
9   **end**
10   **if** *Tx/Rx edge(s) appear in the path* **then**
11   Set up new lightpath(s).
12   **end**
13   Accommodate traffic.
14 **end**
15 Compute PC due to traffic flow through the path.
16 **if** *traffic provisioned fully* **then**
17   return $true$
18 **else**
19   return $false$
20 **end**

C. Q-Learning

The main idea of Q-learning is to develop a policy to take certain actions based on the state of an environment or system. In Q-learning based algorithms, agents learn the policy based on rewards received by taking actions in different states, and the objective is to maximize the overall reward. Thus, to maximize the overall reward, agent focuses on long-term rewards rather than focusing only on the immediate reward. Next, we discuss Q-learning [28] in the following in brief.

There are three sets in Q-learning, viz., the set of states **S** in the environment, the set of actions **A** from which an agent can take an action, and the set of rewards **R** which comprises the reward received by the agent for each action. At each step t, agent is in a representative state $s \in \mathbf{S}$ in the environment where the agent takes an action $a \in \mathbf{A}$ forming a state-action pair (s, a). For action a taken in the state s at step t, the agent in the next step (i.e., $t+1$) reaches to state $s' \in \mathbf{S}$ and receives reward $R_{t+1} \in \mathbf{R}$. At step t the expected overall reward (G_t) is the summation of all future rewards:

Fig. 2: An example auxiliary graph [9, 24].

$$G_t = R_{t+1} + R_{t+2} + R_{t+3} + ... \quad (1)$$

However, it may happen that in order to gain immediate high reward the agent may fail to win distant future high reward. Thus, rather than optimizing Eq. (1), discounted summation is optimized [Eq. (2)]. Here, $\gamma \in [0, 1]$ represents the discount rate; having value close to 0 shows the focus of agent is on short term rewards and value close to 1 shows focus on long term rewards.

$$\begin{aligned} G_t &= R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... \\ &= R_{t+1} + \gamma G_{t+1} \end{aligned} \quad (2)$$

The probability of selecting an action at a particular state is determined by the policy function, i.e., following policy π at step t on state s, the probability of selecting action a is $\pi(a|s)$. The merit of any state for policy π, i.e., the value of a state if policy π is followed, can be determined by state-value function v_π using Eq. (3), where $\mathbb{E}_\pi$ denotes the expected value of random variable when agent follows policy π and S_t denotes the state at step t.

$$v_\pi(s) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s \right] \quad (3)$$

Similarly, merit of any action at a state for policy π, i.e., the value of an action under policy π can be determined by action-value function q_π using Eq. (4), where A_t denotes the action at step t. This shows the quality of taking an action at a state, also known as the Q-function, and the obtained value is referred to as Q-value for the state-action pair.

$$q_\pi(s, a) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a \right] \quad (4)$$

A policy π can be said to be better than any other policy π' if and only if $v_\pi(s) \geq v_{\pi'}(s), \forall s \in \mathbf{S}$. The optimal state-value function $v_*(s)$, where $v_*(s) = \max_\pi v_\pi(s), \forall s \in \mathbf{S}$ shows the highest expected reward by any policy at each state. Similarly, the optimal action-value function, or optimal Q-function $q_*(s, a)$, where $q_*(s, a) = \max_\pi q_\pi(s, a), \forall s \in \mathbf{S}, a \in \mathbf{A}$ shows the highest expected reward by any policy for each possible state-action pair.

Hence, for any state-action pair (s, a) at step t, the expected reward for selecting an action a at state s following optimal policy is the summation of expected reward for taking action a at state s (i.e., R_{t+1}) and the maximum expected discounted reward that can be achieved from any possible next state-action pair (s', a').

$$\text{i.e., } q_*(s, a) = \mathbb{E} \left[R_{t+1} + \gamma \max_{a'} q_*(s', a') \right] \quad (5)$$

Eq. (5) is known as Bellman equation for q_*. In this study, Q-learning method is used to learn the optimal Q-value for each state-action pair in order to find the optimal policy. In Q-learning, iteratively Q-values are updated for each state-action pair using Bellman equation to converge Q-function towards the optimal Q-function (q_*). This approach is known as value iteration. Thus, from optimal q_*, optimal policy can

be determined, as at any state s action a can be found using Q-learning algorithm that maximizes $q_*(s, a)$.

Over the time the agent goes through multiple episodes (iterations). In each episode, agent takes action at every step based on the highest Q-value of the present state (i.e., performs exploitation), and updates state and Q-value accordingly until the stopping criteria of the episode is met. Initially, Q-values for all state-action pair are maintained at zero value, and thus agent explores different states (i.e., performs exploration). To balance between exploration and exploitation, initially exploration rate (ϵ) is typically set to 1 and gradually decayed (having decay rate ϵ_Δ) with increase in episodes. The exploration rate may be determined following Eq. (6), where $\epsilon_{\min}$ and $\epsilon_{\max}$ are set to 0 and 1 respectively. At each step, a random number $[0, 1]$ is generated and compared with ϵ to select between the exploration or exploitation modes. In exploration mode, an action is selected randomly and this strategy is referred to as epsilon greedy strategy.

$$\epsilon = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min})e^{-\epsilon_\Delta \cdot episode} \quad (6)$$

How quickly agent discards the already learned Q-value is controlled by the learning rate $\alpha \in [0, 1]$. Lower the learning rate, more slowly the agent updates the Q-value. Thus, from Eq. (5), for state-action pair (s, a) at step t, the new Q-value, as shown in Eq. (7) is computed as the weighted sum of the previous Q-value and the learned value.

$$\begin{aligned} q^{new}(s, a) &= (1 - \alpha)q(s, a) \\ &+ \alpha \left[R_{t+1} + \gamma \max_{a'} q(s', a') \right] \end{aligned} \quad (7)$$

D. QAG-ENP-IoE

In Algorithm 2, we present QAG-ENP-IoE heuristic where we sequentially identify a traffic using Q-learning technique and provision it using the AG-based greedy heuristic (i.e., mAG-ENP-IoE). Here, at a given episode, a state signifies the number of traffic that already have been provisioned, and action signifies taking attempt to provision the next traffic. First, all parameters are initialized: all Q-values are set to zero, the total number of actions and the maximum number of steps per episode are made equal to the total number of traffic to be provisioned (i.e., all non-zero elements in the traffic demand matrix), the total number of states is made equal to the total number of traffic to be provisioned added with two (representing the initial state when no traffic has been provisioned and the final state when all traffic have been provisioned), and the total number of episodes, α, γ, $\epsilon_{\min}$, $\epsilon_{\max}$ and ϵ_Δ values are provided.

For each episode, the agent first initializes the starting state to zero and $doneFlag$ to false. The execution of the current episode is terminated (stopped) in case all traffic have been provisioned or a given traffic cannot be provisioned. If any of these two conditions is met, $doneFlag$ is set to true. Next, for each step, first, the action based on exploration or exploitation is decided. A random number [0,1] is generated and compared with ϵ. In case the random number is greater than ϵ, an unprovisioned traffic is selected based on the maximum Q-value at the current state. Otherwise, an unprovisioned traffic

Algorithm 2: Q-learning and AG-based energy-efficient network planning for IP-over-EON (QAG-ENP-IoE)

---

1 Initialize all parameters.
2 for $episode \leftarrow 0$ **to** $totalEpisodes$-1 **do**
3   state = 0, doneFlag = False
4   **for** $step \leftarrow 0$ **to** $maxStepsPerEpisode - 1$ **do**
5   explorationRateThreshold = random.uniform(0, 1)
6   **if** $explorationRateThreshold > \epsilon$ **then**
7   action = action having max Q-value in current state except already taken actions.
8   **else**
9   action = select random action except already taken actions.
10   **end**
11   Use Algorithm 1 to provision traffic.
12   **if** $trafficProvisioned$ **then**
13   reward = -(increment in overall PC)
14   **if** $allTrafficProvisioned$ **then**
15   doneFlag = True
16   reward += R
17   **end**
18   **else**
19   doneFlag = True
20   reward = $-P$
21   **end**
22   update Q-value using Eq. (7)
23   state += 1
24   **if** $doneFlag$ **then**
25   break
26   **end**
27   **end**
28   update ϵ using Eq. (6)
29 end

---

is selected through uniform random distribution. Next, attempt is made to provision the selected traffic using mAG-ENP-IoE (Algorithm 1). If the traffic is provisioned successfully, the incurred reward is the negative of the increase in PC, else reward is $-P$, where, P represents a very large number used as penalty. Reward is considered to be negative as our problem to reduce PC is a minimization problem. Now, for unsuccessful traffic provisioning or in case all traffic have been provisioned, $doneFlag$ is set to true. Additionally, in case all traffic can be provisioned, a high reward R is added to the existing reward. Next, corresponding Q-value is updated following Eq. (7) and state value is increased by one. If $doneFlag$ is $true$ then the current episode ends. After completion of each episode, ϵ value is updated following Eq. (6).

V. Performance Analysis of Proposed Heuristic

To assess the performance of QAG-ENP-IoE heuristic, we consider a realistic large network topology, the National Knowledge Network (NKN) of India with 31 nodes and 81 fibers (Fig. 3) [9]. We consider that inline optical amplifiers are

placed on fibers 80 km apart, and each IP router port supports 400 Gbps. Each sub-transponder of an SBVT and each VER supports all possible transmission options as presented in Table I. The PC for each IP core router port, SBVT, BV-OXC, VER and amplifier are shown in Table II. The maximum numbers of SBVTs and VERs that can be installed at any node are considered to be 64 and 3, respectively. Sliceability of an SBVT and the number of SSRs in a VER are fixed at 3 and 16, respectively. It is considered that VERs can be placed only at the top 30% of the nodes with high degrees. Static traffic demands for all SD node pairs exist, with the exception of the traffic demands within the same nodes (i.e., the diagonal elements of the traffic demand matrix). Traffic demands are considered to be uniformly distributed within [5, $2X$-5] Gbps, where X represents the average traffic demand (ATD). The following input parameters: α, γ, ϵ, $\epsilon_{\min}$, $\epsilon_{\max}$, and ϵ_Δ to be used with QAG-ENP-IoE are given in Table III. For each ATD, we execute simulation using QAG-ENP-IoE for ten thousand episodes with a given traffic demand set (matrix), and the best result is taken. For a given value of ATD, we show result averaged over ten different traffic demand sets.
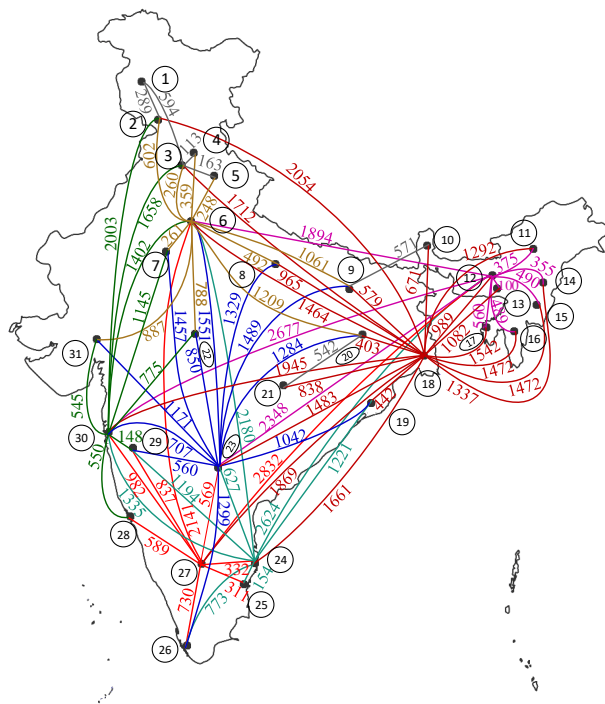

Fig. 3: 31-node NKN, India topology. Links are shown with different colors to improve readability [9].

TABLE II: PC for different network equipment

Equipment	PC
IP core router port	P_R = 560W, C_R = 400 Gbps [23]
SBVT	As shown in Table I [9, 24, 25]
BV-OXC	135·d+150, d is physical degree of node [23]
VER	Lightpath transmit/receive: Table I P_{V_S} = 10W, P_{V_O} = 25W [24]
Amplifier	P_{A_d} = 30W, P_{A_O} = 140W [22]

TABLE III: Different parameters for QAG-ENP-IoE

$\alpha = 0.1$	$\epsilon = 1$	$\epsilon_{min} = 0.01$
$\gamma = 0.99$	$\epsilon_\Delta = 0.001$	$\epsilon_{max} = 1$

Fig. 4 shows the average overall PC and the PC distribution among different network equipment (SBVT, IP core router, amplifier, VER and OXC) using QAG-ENP-IoE for different ATDs. The overall PC increases with increase in ATD. Fig. 4 also shows that average execution time increases with increase in ATD. Table IV shows the standard deviations of PC of different equipment computed for different ATDs. The relative PC (%) for different equipment are also tabulated in Table V. In general, PC due to SBVTs predominates the overall PC of network, followed by PC due to IP core routers, amplifiers, OXCs, and VERs. Relative PC due to amplifiers decreases with increase in ATD, since once all amplifiers on fibers are activated, further increase in PC due to amplifiers cannot take place. Relative PC due to IP core routers increases with increase in ATD, since requirement of IP core router ports increases with increase in ATD.
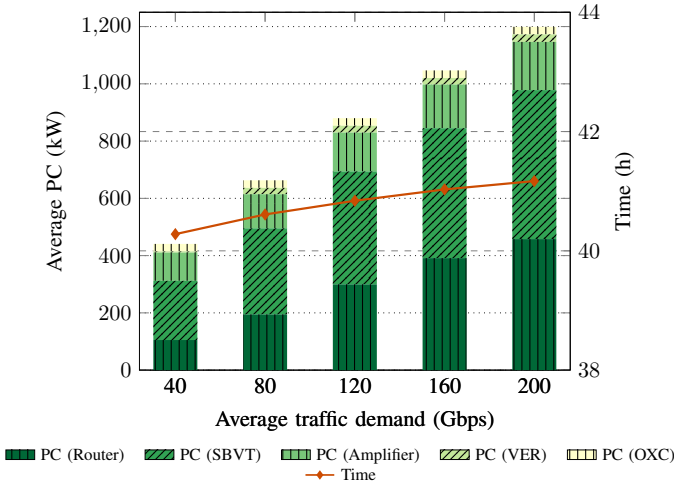

Fig. 4: Average PC for different equipment and average execution time for NKN designed with QAG-ENP-IoE.
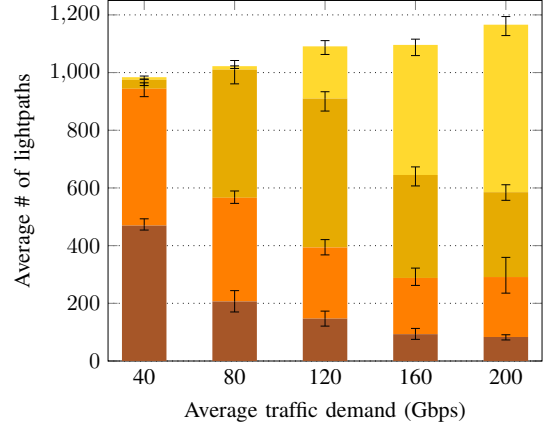
TABLE IV: Standard deviation in PC of different equipment for NKN

X (Gbps)	40	80	120	160	200
Router	1.02	3.2	6.48	12.75	10.42
SBVT	1.01	6.11	6.69	12.11	13.79
Amplifier	0.83	6.48	5.21	5.25	6.07
VER	1.22	1.80	1.45	2.04	1.36
OXC	0	0	0	0	0
Total PC	2.79	11.03	16.81	24.28	21.38

Fig. 5 shows the average number of total lightpaths and average number of lightpaths with different capacities set up with different ATDs. With the increase in ATD, more high capacity (e.g., 400 Gbps) lightpaths are required to be set up to efficiently accommodate the increasing traffic demands. Moreover, in general, higher number of lightpaths are set up with increase in ATD. Since PC at SBVT is higher for

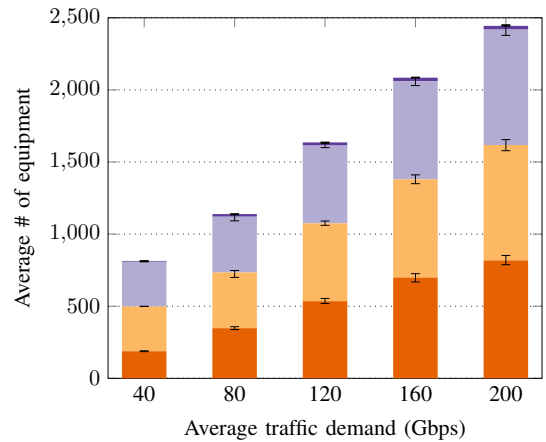TABLE V: Relative PC (%) for NKN designed with QAG-ENP-IoE

X (Gbps)	40	80	120	160	200
Router	23.85	29.37	34.04	37.37	38.17
SBVT	46.89	45.33	44.82	43.34	43.4
Amplifier	22.29	17.97	15.4	14.56	14.03
OXC	6.02	4	3.01	2.53	2.21
VER	0.95	3.33	2.72	2.2	2.18

Fig. 5: Average number of total lightpaths and average number of lightpaths with different capacities set up for NKN designed with QAG-ENP-IoE.

high capacity lightpaths compared with low capacity lightpaths (Table I), and higher number of lightpaths are set up with increase in ATD, more PC occurs with increase in ATD (as shown in Fig. 4). Similarly, with increase in ATD, requirement of different equipment also increases, as reflected in Fig. 6. As VERs are used only for regeneration and used only at the top 30% of the nodes with high degrees, the average number of VERs used in the network is very less compared with other equipment.

Fig. 6: Average number of different equipment installed for NKN designed with QAG-ENP-IoE.

The performance of QAG-ENP-IoE is compared with the following different heuristic methods:

(i) Shortest path network planning with traffic provisioning in descending order (SP).
(ii) Network planning using mAG-ENP-IoE with traffic provisioning in descending order (D-GH).
(iii) Network planning using mAG-ENP-IoE with traffic provisioning in ascending order (A-GH).
(iv) Network planning using mAG-ENP-IoE with traffic provisioning following node indices [starting from the first element to the last element of traffic demand matrix] (I-GH).

In Fig. 7, we show the average PC for network designed with SP, D-GH, A-GH, I-GH, and QAG-ENP-IoE. SP heuristic first explores the existing lightpaths with free capacity for provisioning the selected traffic. In case, such lightpaths are not available, a new lightpath is set up following the shortest distance path between SD node pair of the traffic, and the traffic is provisioned through it. In all cases (i.e., all ATD values), PC for network designed using SP is the highest as the method does not consider PC during network designing. With 40-120 Gbps ATD, D-GH, A-GH and I-GH offer almost 4-15%, 0-15% and 5-12% improvement in average PC compared with SP, respectively. Network designing is not feasible using D-GH, A-GH and I-GH for none of the considered traffic demand matrices with ATD of 200 Gbps, since resources are exhausted at one or more locations. These three methods can provide solution for only one traffic demand matrix with ATD of 160 Gbps, due to exhaustion of resources. QAG-ENP-IoE learns the best sequence for energy-efficient network planning. Even though the heuristic may fail to provide feasible solution for a given traffic demand matrix in some cases (out of the given ten thousand episodes) due to paucity of resources, the heuristic eventually can always successfully provision the entire traffic demand matrix. The heuristic comes up with an optimal sequence, i.e., a policy that provisions traffic for energy-efficient IP-over-EON. QAG-ENP-IoE offers 1-23%, 5-7%, 1-7% and 1-9% improvement in PC values compared with SP, D-GH, A-GH and I-GH methods, respectively. In
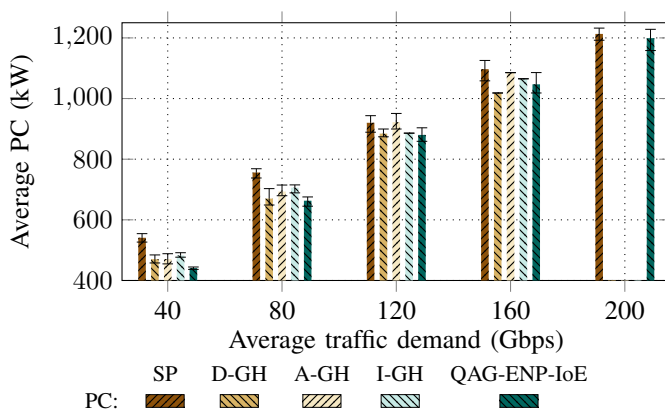
Fig. 8, the average number of feasible solutions, i.e., the average number of success per 1000 episodes are plotted. Initially, the number of success is low, and in general, the number of success increases with increase in the index of 1000 episodes.
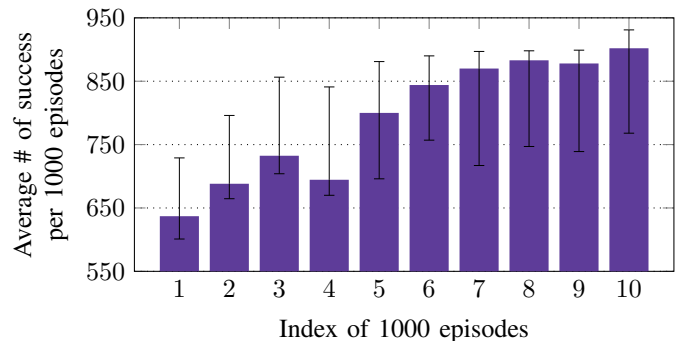

Fig. 8: Average number of success per 1000 episodes.

VI. Conclusion

In this paper, network planning implemented through efficient resource and static traffic demand provisioning for energy-efficient VER-assisted IP-over-EON is explored. We propose RL (in particular, Q-learning) and AG-based heuristic for energy-efficient resource allocation and traffic provisioning. Even though the heuristic may fail to provide feasible solution for a given traffic demand matrix in some cases (out of the given ten thousand episodes) due to paucity of resources, the heuristic eventually can always successfully provision the entire traffic demand matrix. Simulation results show the increase in PC with the increase in ATD (i.e., the traffic volume). The PC in SBVTs and IP core routers is found to predominate over the PC in other equipment. The proposed heuristic offers up to 23% and up to 9% improvement in PC compared with SP and AG-based greedy heuristics (viz., D-GH, A-GH and I-GH) respectively.

References

[1] *CISCO VNI forecast and trends, 2017-2022*, https://davidellis.ca/wp-content/uploads/2019/05/cisco-vni-feb2019.pdf, Last accessed: 2021-04-02. [Online]. Available at: https://davidellis.ca/wp-content/uploads/2019/05/cisco-vni-feb2019.pdf.
[2] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.
[3] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Communications Magazine*, vol. 50, no. 2, S12–S20, Feb. 2012.
[4] M. Jinno, Y. Sone, H. Takara, A. Hirano, K. Yonenaga, and S. Kawai, "IP traffic offloading to elastic optical layer using multi-flow optical transponder," *European Conference and Exposition on Optical Communications, OSA*, 2011, pp. 1–3.
[5] V. Gkamas, K. Christodoulopoulos, D. J. Vergados, and E. Varvarigos, "Energy-minimized design of IP over flexible optical networks," *International Journal of Communication Systems, Wiley*, vol. 30, no. 7, pp. 1–17, 2015.

Fig. 7: Average PC for network designed with SP, D-GH, A-GH, I-GH and QAG-ENP-IoE.

[6] R. M. Krishnaswamy and K. N. Sivarajan, "Design of logical topologies: A linear formulation for wavelength-routed optical networks with no wavelength changers," *IEEE/ACM Transactions On Networking*, vol. 9, no. 2, pp. 186–198, 2001.

[7] J. Zhang, Y. Zhao, X. Yu, J. Zhang, M. Song, Y. Ji, and B. Mukherjee, "Energy-efficient traffic grooming in sliceable-transponder-equipped IP-over-elastic optical networks [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 1, A142–A152, 2015.

[8] P. Biswas and A. Adhya, "Energy-efficient, EDFA lifetime-aware network planning along with virtualized elastic regenerator placement for IP-over-EON," *Photonic Network Communications, Springer*, pp. 1–17, 2020.

[9] ——, "Energy-efficient network planning and traffic provisioning in IP-over-elastic optical networks," *Optik-International Journal for Light and Electron Optics, Elsevier*, vol. 185, pp. 1115–1133, 2019.

[10] J. Zhao, H. Wymeersch, and E. Agrell, "Nonlinear impairment-aware static resource allocation in elastic optical networks," *IEEE Journal of Lightwave Technology*, vol. 33, no. 22, pp. 4554–4564, 2015.

[11] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge, 1989.

[12] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[13] L. M. Gambardella and M. Dorigo, "Ant-q: A reinforcement learning approach to the traveling salesman problem," *Machine Learning Proceedings*, Elsevier, 1995, pp. 252–260.

[14] M. Klinkowski and K. Walkowiak, "Routing and spectrum assignment in spectrum sliced elastic optical path network," *IEEE Communications Letters*, vol. 15, no. 8, pp. 884–886, 2011.

[15] R. Ramaswami and K. N. Sivarajan, "Design of logical topologies for wavelength-routed optical networks," *IEEE Journal on Selected areas in communications*, vol. 14, no. 5, pp. 840–851, 1996.

[16] P. Biswas, S. K. Dey, and A. Adhya, "Auxiliary graph based energy-efficient dynamic connection grooming for elastic optical networks," *IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Nov. 2016, pp. 1–3.

[17] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2018.

[18] M. Salani, C. Rottondi, and M. Tornatore, "Routing and spectrum assignment integrating machine-learning-based QoT estimation in elastic optical networks," *IEEE Conference on Computer Communications (INFOCOM)*, 2019, pp. 1738–1746.

[19] M. Lian, R. Gu, Y. Qu, Z. Wang, and Y. Ji, "Flexible optical network enabled hybrid recovery for edge network with reinforcement learning," *OSA Optical Fiber Communication Conference (OFC)*, 2020, M1A–2.

[20] Y. Kiran, T. Venkatesh, and C. S. R. Murthy, "A reinforcement learning framework for path selection and wavelength selection in optical burst switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 9, pp. 18–26, 2007.

[21] R. Gu, Y. Qu, M. Lian, H. Li, Z. Wang, Y. Zhu, Q. Guo, J. Yang, D. Wang, and Y. Ji, "Flexible optical network enabled proactive cross-layer restructuring for 5G/B5G backhaul network with machine learning engine," *IEEE Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[22] J. L. Vizcaíno, Y. Ye, and I. T. Monroy, "Energy efficiency analysis for flexible-grid OFDM-based optical networks," *Computer Networks, Elsevier*, vol. 56, no. 10, pp. 2400–2419, 2012.

[23] W. V. Heddeghem, F. Idzikowski, W. Vereecken, D. Colle, M. Pickavet, and P. Demeester, "Power consumption modeling in optical multilayer networks," *Photonic Network Communications, Springer*, vol. 24, no. 2, pp. 86–102, 2012.

[24] P. Biswas, A. Adhya, S. Akhtar, J. Gupta, and S. Majhi, "EDFA active-sleep transition frequency and EDFA occupancy aware dynamic traffic provisioning for energy-efficient IP-over-EON," *IEEE International Conferences on Signal Processing and Communication Systems (ICSPCS)*, Dec. 2019, pp. 1–7.

[25] P. Papanikolaou, P. Soumplis, K. Manousakis, G. Papadimitriou, G. Ellinas, K. Christodoulopoulos, and E. Varvarigos, "Minimizing energy and cost in fixed-grid and flex-grid networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 4, pp. 337–351, 2015.

[26] M. Jinno and H. Takara, "Elastic optical transponder and regenerator: Toward energy and spectrum efficient optical transport networks," *IEEE International Conference on Photonics in Switching (PS)*, Sep. 2012, pp. 1–3.

[27] M. Jinno, K. Yonenaga, H. Takara, K. Shibahara, S. Yamanaka, T. Ono, T. Kawai, M. Tomizawa, and Y. Miyamoto, "Demonstration of translucent elastic optical network based on virtualized elastic regenerator," *National Fiber Optic Engineers Conference, OSA*, 2012, pp. 1–3.

[28] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.