Fragmentation-Aware Load-Balancing Virtual Optical Network Embedding (VONE) Over Elastic Optical Networks

Fariborz Mousavi Madani, Sheida Mokhtari
Faculty of Engineering
Alzahra University
Tehran, Iran
email: mosavif@alzahra.ac.ir, sh.mokhtari@student.alzahra.ac.ir

Abstract—Optical network virtualization has recently been studied extensively as a promising solution to share physical infrastructure resources among different users and applications. Virtual network embedding over elastic optical networks has recently attracted massive attentions as a promising solution to realize fine-grained flexibility in resource provisioning. Variation of bandwidth requirements associated with virtual network requests accompanied by nondeterministic nature in arrival and holding times of them result in fragmentation of spectral blocks along both frequency and spatial directions. Spectral fragmentation refrains commitment to spectrum contiguity and spectrum continuity constraints in lightpath establishment thereby severely exacerbate blocking probability, as well as spectrum utilization. Fragmentation-aware routing and spectrum assignment problem has been extensively studied in the context of elastic optical network provisioning. These schemes, however, cannot be applied to virtual network embedding since we cannot refer to routing and spectrum assignment problem to implement virtual link mapping without regard to virtual node mapping. Therefore, in this paper, an integer linear programming formulation for fragmentation-aware virtual network embedding was developed for the first time to our knowledge. Numerical simulations for three different scenarios were carried on and demonstrated that our proposed model outperforms the previous work over a wide range of offered traffic loads under the same conditions.

Keywords—elastic optical network; fragmentation-aware; load-balancing; VONE

I. INTRODUCTION

Optical networks are evolving from a fixed International Telegraph Union-Telecommunication (ITU-T) Dense Wavelength Division Multiplexing (DWDM) wavelength grid to a flexible grid in which the optical spectrum is divided into smaller Frequency Slots (FSs) with 6.25 or 12.5 GHz width each. In such a flexible grid or Elastic Optical Network (EON), resources are assigned by matching resource requirements of connection requests and the necessary number of FSs. The flexible use of the spectrum grid enabled by EONs promotes a more efficient and adaptable use of networks resources. However, the setup and release of connections in a dynamic network scenario can create gaps in the optical spectrum, which sizes in terms of the number of FSs may not be sufficient to accommodate an incoming connection request, causing what is known as the fragmentation problem. The spectrum fragmentation problem may lead to inefficient resource utilization and a high blocking probability [1]. A Virtual Optical Network Embedding (VONE), generally, consists of two stages, node mapping and link mapping, because of the complexity originated from correlation between these two stages, we brought it up in Integer Linear Programming (ILP) formulation. A Virtual Optical Network (VON) is composed of several Virtual Nodes (VNs) interconnected by Virtual Optical Links (VOLs). Typically, a network operator constructs VONs using optical network embedding, which allocates necessary resources in the physical infrastructure to each VON through node and link mapping stages. Due to the complex inter-dependence between these stages, VONE has become a major challenge for optical network virtualization [2]. Even though VONE over fixed-grid Wavelength-Division Multiplexing (WDM) networks, assuming that all the Substrate Nodes (SNs) were equipped with sufficient wavelength converters has been studied extensively [3]-[5], VONE over flexible-grid EONs has just started to attract research interests [2][6]-[8], which can potentially provide efficient support to emerging cloud services, especially for the highly distributed and data-intensive applications such as petabits-scale grid computing. In this paper, we develop and investigate an ILP model for Fragmentation-Aware Load-Balancing-VONE (FALB-VONE) over flexible-grid EONs for the first time to the best of our knowledge. The simulation results from three different scenarios demonstrate that our proposed ILP model outperforms previous work on minimization of resource utilization.

The rest of the paper is organized as follows. Section II surveys related work on VONE. The network models and problem descriptions of FALB-VONE over EONs are presented in Section III. Section IV discusses the ILP model and the performance evaluations of FALB-VONE. Finally, Section V concludes the paper.

II. RELATED WORKS

Most of the previous studies on VONE were targeted for fixed-grid WDM networks, but recent studies have suggested that optical network virtualization over flexible-grid EONs can potentially provide efficient support to emerging cloud services, especially for the highly distributed and data-intensive applications such as petabits-scale grid computing [2]. Chen et al. [9] investigated a spectrum allocation
approach through utilizing the relationship between accommodation capability of spectrum blocks and traffic bandwidth distribution. Based on this, a Fragmentation-Aware Spectrum Allocation (FSA) algorithm was presented. Zhang, Lu, Zhu, Yin, and Yoo [10] discussed two Routing and Spectrum Assignment (RSA) algorithms that can control the increase of fragmentation ratio and reduce necessary defragmentation operations in networks, the first one is the Minimum Network Fragmentation Ratio RSA (MNFR-RSA) algorithm, which makes resource allocation based on the Network Fragmentation Ratio (NFR) and the second one is the Maximum Local Utilization RSA (MLU-RSA) algorithm, which alleviated bandwidth fragmentation by utilizing the slots that have already been used the most in the network. In [11] Talebi, et al. examined and categorized solution approaches to Fragmentation-Aware RSA (FA-RSA) including Proactive FA-RSA, which attempts to prevent or minimize spectrum fragmentation at the time a new request is admitted and Reactive FA-RSA that employs defragmentation techniques to accommodate high-rate and long-path connections. The objective of defragmentation is to rearrange the spectrum allocation of existing traffic demands so as to consolidate available slots into large contiguous and continuous blocks. Yin, Zhu, and Yoo in [12] analyzed the fragmentation problem in detail for service provisioning in dynamic EONs, and proposed three fragmentation-aware algorithms for Routing, Modulation and Spectrum Assignment (RMSA) in EONs. Moura, Fonsenca, and Scaraficci [13] introduced a multi-graph Shortest Path algorithm, which represented the spectrum occupancy by a multi-graph, where allocation decisions were based on cost functions, which tried to capture the potentiality of spectrum fragments to allocate incoming requests. In [14], Yin et al. have investigated the spectrum fragmentation problem in EON in both the spectral and the spatial dimensions and have proposed two fragmentation aware RSA algorithms to proactively prevent fragmentation during RSA to the incoming lightpath request.

Among the published papers on VONE, [2][8][15] have taken care of minimizing spectral utilization of substrate fiber links for mapping the requested virtual link bandwidth, possibly with the cost of imposing spectral fragmentation and intensifying load unbalancing. In the following section, we will develop an ILP formulation that encompass appropriate fragmentation-awareness and load-balancing expressions within the objective function.

III. MATHEMATICAL MODEL

We model the substrate EON as an undirected graph, denoted as $G^S(V^S, E^S)$, where $V^S$ is the set of SNs, and $E^S$ represents the set of Substrate Fiber Links (SFLs). We assume that each SN $v^S \in V^S$ has a computing capacity of $c_{v^S}$ and each SFL $e^S \in E^S$ can accommodate $B^S$ Frequency Slots (FS').

Similar to the substrate topology, a VON request can also modeled as an undirected graph $G^V(V^E, E^V)$, where $V^E$ is the set of virtual nodes (VNs), and $E^V$ represents the set of VOLS. Each VN $v^E \in V^E$ is associated with a computing resource requirement $c_{v^E}$, while the Band Width (BW) requirement of each VOL $e^V \in E^V$ is $bw_{e^V}$ (in terms of Gb/s). This work considers both node mapping and link mapping in the VONE process. In the course of node mapping, each VN from the VON request is mapped onto a unique SN that has sufficient computing capacity. The link mapping course is essentially a special RSA operation. Specifically, the RSA sets up a lightpath in the physical infrastructure for each VOL to satisfy its BW requirement, under the spectrum non-overlapping, continuity and contiguity constraints [2].

The next subsection introduces a path-based ILP formulation for the FALB-VONE over EONs following the recent work presented in [2]. In-line with modulation-level adaptability provided by Optical Orthogonal Frequency Division Multiplexing (OOFDM) scheme, we assume that substrate lightpath can select one of Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), 8 Quadrature Amplitude Modulation (8QAM), and 16QAM modulation-levels adaptively according to the transmission distance. If we take the BW of one FS to be 12.5 GHz, it can carry 12.5 Gb/s signal when the modulation-level is BPSK.

Based on the experimental results in [16], we set the transmission reach of BPSK, QPSK, 8QAM, and 16QAM signals as 3000, 1500, 750, and 375 km, respectively. For all pre-calculated substrate shortest paths, the highest possible modulation-levels are predetermined and stored. BW requirement of VOL is normalized as multiples of one subcarrier BPSK signal operating at 12.5 Gbs.

A. Definitions
1) $Y \triangleq |E^r|$ Total number of VOLs in a VON request.
2) $M \triangleq |E^v|$ Total number of links in substrate EON.

B. Notations
1) $G^*(V^*, E^*)/G^r(V^r, E^r)$ Graph of physical infrastructure/ VON request.
2) $c_{v^r}/c_{e^r}$ Computing capacity/resource requirement of each SN $v^r \in V^r$/VN $v^E \in V^E$.
3) $B$ Total number of FSs on each SFL $e^s \in E^s$.
4) $w_e^{k} / s_e^{k}$ Starting/Ending FS index of $k$th Maximal Contiguous Slot-Block (MCSB), which is a Contiguous Slot-Block (CSB) that includes all the available and contiguous FS(s) at a spectral location on SFL $e^s$.
5) $s_e^{r} / d_e^{r}$ End-nodes of the VOL $e^r \in E^r$.
6) $P^S$ Set of all pre-calculated routing paths in the substrate network $G^S(V^S, E^S)$.
7) $s_p^e / d_p^e$ Source/ Destination node of the path $p^S \in P^S$.
8) $P^{e^s}$ Set of routing paths that use the SFL $e^s$, obviously, $P^{e^s} \subset P^S$.
9) $m_{p}^e$ Highest modulation-level for path $p^S$.
10) $bw_{e^r}$ Normalized BW requirement of the VOL $e^V$ in the VON request $G^r(V^r, E^r)$.
11) \( n_s^s \) Number of MCSBs on SFL \( e^s \) before FS assignment.
12) \( \eta_L^{\min} / \eta_L^{\max} \) Minimum/Maximum value of \( \eta_l \); weighting coefficient for link load balancing expression in the objective function.
13) \( \eta_N^{\min} / \eta_N^{\max} \) Minimum/Maximum value of \( \eta_N \); weighting coefficient for node load balancing expression in the objective function.
14) \( \Delta C_L^e / \Delta C_N^e \) Difference between maximum and minimum available Link/Node capacities before FS assignment.

C. Objective Function

Minimize

\[
\alpha \cdot \frac{1}{M} \sum_{e^r \in E^r} \left( \Psi_{e^r, \cdot} / w^s_{e^r, \cdot} \right) + \beta \cdot \frac{1}{M} \sum_{e^r \in E^r} \left( \sum_{l=1}^{n_e^r+Y} f_{e^r, \cdot, l} / n_{e^r}^s \right) + \frac{\gamma}{M} \sum_{e^r \in E^r} \left( \sum_{p^r \in P^r} \sum_{p^{\cdot, r} \in P^{\cdot, r}} \zeta_{e^r, p^r} \cdot \left| p^{\cdot, r} \right| \left[ \frac{bw_{e^r}}{m_{e^r}} \right] \right)
\]

\[
+ \eta_L \cdot \frac{C_L^{\max} - C_L^{\min}}{\Delta C_L^e} + \eta_N \cdot \frac{C_N^{\max} - C_N^{\min}}{\Delta C_N^e}
\]

where, the first term is introduced to keep the assigned CSBs as close as possible to the starting index of first MCSB. It comprises sum of ratios of highest starting slot index among CSBs allocated to \( e^r \)'s that uses \( e^s \) to the starting FS index of the first MCSB of that \( e^s \) divided by \( M \) to get the normalized value. \( \alpha \) is the weighting coefficient to adjust the relative contribution of this term. The second term is included to minimize extra spectral fragmentation induced by VOL mapping and consists of normalized sum of total number of free slot blocks in \( e^r \) after allocating CSBs to \( e^s \)'s divided by number of MCSBs of that \( e^s \). \( \beta \) is the weighting coefficient to tune its impact. The third term accounts for minimizing BW utilization, likewise presented in [2], but here, it is divided by a lower band of BW utilization for normalization purpose and \( \gamma \) is the corresponding weight coefficient. Finally, the last two terms are responsible to maintain link and node load balancing in substrate EON. \( \eta_l \) and \( \eta_N \) are their weighting coefficients, respectively, which are computed from the given parameters \( \eta_L^{\min} / \eta_L^{\max} \) and \( \eta_N^{\min} / \eta_N^{\max} \) by the following equations:

\[
\eta_L = (\eta_L^{\max} - \eta_L^{\min}) \cdot \exp(-\min C_L^e / \Delta C_L^e) + \eta_L^{\min}
\]

\[
\eta_N = (\eta_N^{\max} - \eta_N^{\min}) \cdot \exp(-\min C_N^e / \Delta C_N^e) + \eta_N^{\min}
\]

Under excessive load unbalancing conditions, some links are highly loaded \((\min C_L^e \approx 0)\) while some others are lightly loaded \((\Delta C_L^e \gg 0)\) so the ratio inside exponent becomes nearly zero to make \( \eta_L \) become close to \( \eta_L^{\max} \). However, when all links are lightly loaded \((\min C_L^e \approx 0)\) and/or slightly unbalanced \((\Delta C_L^e \approx 0)\), the exponent goes to zero to make \( \eta_L \) become close to \( \eta_L^{\min} \). The same pattern applies to \( \eta_N \).

D. Constraints

Equations (3)–(18) formulates general requirements for node and link mapping, as articulated in [2], and the remaining equations are properly defined to characterize additional variables and expressions.

\[
\sum_{v^r \in V^r} \xi_{v^r, v^s} = 1, \forall v^r \in V^r
\]

\[
\sum_{v^s \in V^s} \xi_{v^r, v^s} \leq 1, \forall v^s \in V^s
\]

Equations (3), (4) ensure that each VN in the VON request is mapped onto a unique SN,

\[
\sum_{v^r \in V^r} \xi_{v^r, v^s} \cdot c_{v^r} \geq c_{v^s}, \forall v^r \in V^r
\]

Equation (5) ensures that the embedding SN has enough computing capacity to accommodate the VN.

\[
\zeta_{e^r, p^r} \leq \zeta_{e^r, p^s}, \forall e^r \in E^r, \forall p^s \in P^s
\]

\[
\zeta_{e^r, p^r} \leq \zeta_{e^r, p^s}, \forall e^r \in E^r, \forall p^s \in P^s
\]

\[
\sum_{p^r \in P^r} \zeta_{e^r, p^r} = 1, \forall e^r \in E^r
\]

\[
\sum_{e^r \in E^r} \zeta_{e^r, p^r} \leq 1, \forall p^s \in P^s
\]

Equations (6)–(9) ensure that each VOL is mapped onto a single substrate lightpath, and the end nodes of the lightpath are the SNs that the corresponding VNs are mapped onto.

\[
\sigma_{e^1, e^2} \geq \zeta_{e^1, p^1} + \zeta_{e^2, p^2} - 1, \forall e^1, e^2 \in E^r, \forall p^1, p^2 \in P^s
\]

\[
\rho_{e^1, e^2} + \rho_{e^1, e^2} = 1, \forall e^1, e^2 \in E^r
\]

\[
z_{e^r} - w_{e^r} + 1 \leq \{1 + \rho_{e^1, e^2} - \sigma_{e^1, e^2} \}, \forall e^1, e^2 \in E^r
\]

\[
z_{e^r} - w_{e^r} + 1 \leq B^r \cdot \{2 - \rho_{e^1, e^2} - \sigma_{e^1, e^2} \}, \forall e^1, e^2 \in E^r
\]

Equations (10)–(13) together with (15)–(18) ensure that the spectrum assigned to any VOL is contiguous and the spectrum assigned to any two VOLs, whose associated substrate lightpaths have common SFL(s) do not overlap,

\[
z_{e^r} - w_{e^r} + 1 = \sum_{p^r \in P^r} \zeta_{e^r, p^r} \left[ \frac{bw_{e^r}}{m_{e^r}} \right], \forall e^r \in E^r
\]

Equation (14) ensures that the number of FS' assigned to each VOL can just satisfy its BW requirement.

\[
\pi_{e^r, e^s} = \sum_{p^r \in P^r} \zeta_{e^r, p^s}, \forall e^r \in E^r, \forall e^s \in E^s
\]

\[
\sum_{k} \xi^{(k)}_{e^r, e^s} = \pi_{e^r, e^s}, \forall e^r \in E^r, \forall e^s \in E^s
\]

\[
w_{e^r,k} \geq w_{e^r,k} \cdot \left[ \xi^{(k)}_{e^r, e^s} + \pi_{e^r, e^s} - 1 \right], \forall e^r \in E^r, \forall e^s \in E^s, \forall k
\]
Equations (15)–(18) ensure that the assigned CSB for each VOL locates in a single MCSB on SFL $e_s^1$, and the MCSB’s size is not smaller than that of the assigned CSB.

$$\theta'_{e,m,n} \leq B^s \cdot \theta_2 \cdot \theta_{e,m,n}, \forall e^s \in E^s, 1 \leq m, n \leq n_e^s + Y$$

Equations (19)–(21) construct the equality $\theta'_{e,m,n} = w_{e,m,n} \cdot \theta_{e,m,n}$, likewise (22)–(24) construct the equality $\phi'_{e,m,n} = z_{e,m,n} \cdot \phi_{e,m,n}$.

$$\sum_{m=1}^{n_e^s + Y} \theta_{e,m,n} = 1, \forall e^s \in E^s, 1 \leq n \leq n_e^s + Y$$

$$\sum_{m=1}^{n_e^s + Y} \phi_{e,m,n} = 1, \forall e^s \in E^s, 1 \leq n \leq n_e^s + Y$$

Equations (25)–(28) together with (29)–(30) ensure one-to-one correspondence between the elements of $w_{e,m} / z_{e,m}$ and $\hat{w}_{e,m} / \hat{z}_{e,m}$, respectively.

$$\hat{w}_{e,m} = \sum_{m=1}^{n_e^s + Y} \theta_{e,m,n}, \forall e^s \in E^s, 1 \leq m \leq n_e^s + Y$$

$$\hat{z}_{e,m} = \sum_{m=1}^{n_e^s + Y} \phi_{e,m,n}, \forall e^s \in E^s, 1 \leq m \leq n_e^s + Y$$

Equations (31)–(32) make sure that elements of $\hat{w}_{e,m} / \hat{z}_{e,m}$ will be sorted in descending order.

$$l_{e,m} = \hat{w}_{e,m} - \hat{z}_{e,m}, \forall e^s \in E^s, 1 \leq m \leq n_e^s + Y$$

Equation (33) constructs pattern of lengths of free slot blocks for every substrate link. Equations (34)–(35) construct $\bar{f}_{e,m}$:

$$\Psi_{e^s} \leq B^s \cdot \pi_{e^s,e^s} + w_{e^s}, \forall e^s \in E^s, \forall e^s \in E^r$$

Equations (36)–(39) construct the relations:

$$\Psi_{e^s} \geq \max_{e^s'} \frac{\sum_{v^s \in V^s} \xi_{v^s'}}{\sum_{v^s \in V^s} \xi_{v^s'_1}}, \Psi_{e^s} \in E^s, \forall e^s \in E^r$$

Equations (42)–(43) find the maximum and minimum available link capacities among all SFLs, respectively, after allocating all VOL requests.

$$C^N_{e^s} \geq \sum_{e^s \in V^s \in V^s} \xi_{e^s} \cdot c_{e^s}, \forall v^s \in V^s$$

Equations (44)–(45) find the maximum and minimum available node capacities among all SNs, respectively, after allocating all VON requests.

IV. PERFORMANCE EVALUATION

We implemented the proposed ILP model by ILP solver IBM Cplex [17] to evaluate the blocking performance of a simple six-node topology as substrate network, shown in Figure 1. We assumed that all SFLs have the same length of 100 km, the number of VNs in each VON request is uniformly distributed in a preset range, and the probability that a VN-pair is directly connected equals 0.5, which means that there would be $n(n - 1)/4$ VOLS on average for a VON request with $n$ VNs. VONE requests arrive randomly according to Poisson traffic distribution and their holding time follow negative exponential distribution with an average value of 10 min. We also assume that each incoming request should be served instantly or will be rejected forever due to insufficiency of available resources.
The blocking performance of the proposed model under different request loads were then investigated and compared with those of the reference work [2] for three different scenarios through extensive numerical simulations. Table I depicts common parameter settings and Table II illustrates details of scenario-specific parameters. In scenarios 1 and 3, the range of variations of VOL’s BW requirement and VN’s computing requirement, respectively, were deliberately increased to present the effectiveness of load-balancing and fragmentation-awareness contributions more clearly. To find the optimal values of weighting coefficients, they were initially set to 1 and then each of them was precisely tuned against the others.

Figure 2 shows the variation of blocking probability versus VON request load for the first scenario under four different situations. The "reference" graph exhibits results obtained by preserving the third term of the objective function and eliminating all other terms, as well as their associated constraints. This served us to benchmark the performance of the proposed model against the previous work. The graph labeled "FA" illustrates results obtained by considering only the impact of Fragmentation-Awareness in VONE (α = β = γ = 1, η_L = η_L = 0). Conversely, the graph labeled "LB" shows results gained by including only the impact of Load-Balancing VONE (α = β = 0, γ = η_L = η_L = 1). Finally, the graph labeled "FA+LB" represents that the combined impact of all terms having their optimal weighting coefficients achieves the best performance. This verifies the presumption that all FA and LB terms synergistically contribute to lower the blocking probability.

Variation in the range of VOL’s BW requirement or VN’s computing requirement, was studied in scenarios 2 and 3, respectively. Figures 3 and 4 show that the proposed scheme could still successfully control spectral fragmentation and simultaneously maintain load-balanced condition. It is worth noting that, under light request load, the proposed scheme fulfilled an order-of-magnitude reduction in the blocking probability in all scenarios compared with the reference work without imposing any additional cost.
V. CONCLUSION AND FUTURE WORK

The recent work on virtual network embedding presented ILP formulation to optimize resource utilization. In this paper, we extended the model so that fragmentation-awareness, as well as load-balancing features were taken in to account in the optimization process. Simulation results verified that our proposed model outperforms the recent work under three different scenarios. Parameter settings in scenario 2 were exactly replicated the reference work, Scenarios 1 and 3 were arranged to demonstrate more clearly the effectiveness of load-balancing and fragmentation-awareness contributions, respectively. Moreover, our model could achieve remarkable reduction in blocking probability under wide range of request traffic loads, particularly in light request traffic loads.

Our future plan is to develop efficient heuristic algorithm based on the notions embodied in ILP formulation.

REFERENCES


