Received: June 6, 2022. Revised: August 19, 2022. 152

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Dynamic Virtual Network Embedding with Latency Constraint in Flex-Grid Optical Networks

Sumaya A. Ibraheem[1]* **Omar Y. Shaaban[1]**

[1]*Information and Communication Engineering Department, Al Khwarizmi Engineering College, University of Baghdad, Baghdad, Iraq*
* Corresponding author's Email: sumayaa.majeed@gmail.com

Abstract: Nowadays, latency factor in network virtualization takes a serious attention in 5G networks upgrading. Flex-grid/elastic optical network (EON) virtualization recently become known as the fifth-generation (5G) enabling technology for the optical network slicing efficiency. As a consequence, the solution of the virtual optical network embedding (VONE) problem over EON must be solved with the latency targets guarantee. In this work, a new loading factor was suggested after adding the latency effect to the existing EON constraints by identifying its contributing factors in the network. Then, a virtual network mapping algorithm was presented to solve the VONE problem including latency constraint. Network topologies (NSFNET, USNET) were used under the dynamic scenario and four optimization methods which are; alignment and consecutiveness-aware virtual network embedding (ACT-VNE), the local resource capacity and the shortest path first fit (LRC-SP-FF), greedy algorithm and the shortest path first fit (Greedy-SP-FF), load and resource-aware based on ant colony optimization (LRA-ACO) were used and results have been compared to get the optimal solution. The blocking probability on average was increased by approximately 50 % in both topologies affected by adding latency (between 20-50 ms) as compared to the non-latency cases of previous works. Also, during all simulations, the LRA-ACO gave the best embedding performance.

Keywords: Elastic optical network, Network virtualization, Virtual optical network embedding, Latency.

1. Introduction

Many of emerged applications have a specific latency demands as a quality of service (QoS) necessity [1, 2]. Online conferences, multi-player online gaming, and security applications need a strict requirement of latency. Additionally, health-care applications may have a threat on life if not meet the latency conditions [2]. Also, the significant impact of latency to the revenue which is generated by the services for instance the electronic commerce and trading of high-frequency [3]. In the finance sector, low-latency networks and high-performance data centers determine the success of the business, e.g., "a 1-*ms* advantage in trading applications can be worth 100 million US dollars a year to a major brokerage firm" [4]. As a result, latency-sensitive applications have emerged as one of the most important commercial motivations for 5G mobile networks development [5, 6]. Network virtualization is a key enabler for latency-sensitive applications deploying in 5G networks [7]. It enables creating one or more virtual networks (VNs) with allocated substrate resources to ensure the latency between application end nodes. Creating and mapping of VNs on the substrate network for applications that is sensitive to latency efficiently is a challenging issue with a virtual network embedding (VNE) problem [8]. Although previous studies focused on VNE problem, the influence of considering latency as a constraint with VNE problem does not take a place for EONs extensively [9].

Many works with the rapid growth of internet traffic were made. Different studies were made to improve the performance of the virtual optical network embedding (VONE) process by optimizing resources i.e., modulation format devices [10, 11], energy-aware, and resource utilization [12, 13].

Latency effects was first considered in conventional networks. Mapping process to meet a certain latency target, was modeled as in [14, 15]. A virtual network mapping algorithm that considers trading off between low application delay and efficient substrate utilization was presented by K. Ivaturi and T. Wolf [16].

Many sources of latency were discussed for conventional optical networks. J. A. Jay showed the importance of low latency in optical communications networks especially on time spend to respond to a message [17]. Also, V. Bobrovs et al. described the latency caused in fiber optical metro networks. Some available techniques and solutions for reducing latency were discussed [18].

After vast research done in Layer-2/3 networks, investigations begin on virtual optical networks (VONs) embedding like the work done in [19, 20]. The main challenges that the optical domain poses are derived from the analogue nature of the optical substrate in opposition to the electrical one, and, as a consequence, new constraints appear on the field. The physical optical infrastructures were the fixed-grid one also called WDM networks [21].

The shortcomings of fixed-grid WDM have triggered the evolution towards flex-grid optical networks. The more recent flex-grid approach divides the frequency spectrum into smaller (e.g., 12.5 Ghz) slices and allows allocation of an arbitrary number of slices to right size spectrum allocation to customer needs. This new generation of optical networks is usually referred to as elastic optical networks (EONs) [22].

R.Yang et al. proposed a VONE method that considered the load jointly with the continuity and contiguity constraints through the embedding process in an EON infrastructure [23]. Despite the good results obtained for embedding VN over EON, it did not take the latency constraint in consideration.

Although, EONs are considered to be the next-generation high-speed network, less works focused on the VNE problem with latency constraint. S. Taeb et al. demonstrated and identified the latency and its different contributing factors with VNE problem in EON with flexible transmission parameters. An ILP and a heuristic solution were presented for VNE problem with guaranteed virtual path latency. A real network topology besides realistic transmission configurations were used to evaluate and observe the proposed solution under different scenarios. The results for fixed-grid and flex-grid optical networks were compared showing additional cost added when latency was considered by 2.5 % and 0.8 % on average for both networks respectively. The latency constraint effect also increases the blocking probability in dynamic scenario [24].

This work represents a study of the latency effect as a constraint with EON constraints for VNE process. Four optimization methods were used to get the optimal results for embedding. The latency constraint was applied on the virtual links for each VON. Our model used a modified version of the load-balance factor in [23], after adding the latency effect of the EON parts. This factor was inserted through the embedding algorithm such that a comparison was made between conventional results (without latency) and that obtained with latency cases. This is the main contribution accomplished. Also, our work differs from [24] by considering this new load-balance factor, and comparing the different optimization methods to attain an optimal blocking probability value.

The paper is organized as follows; section 1 presents an introduction with a literature review for the previous works, section 2 presents the mathematical model of the VONE in an EON, in section 3 the results of the proposed method with appropriate discussion are presented. Finally, conclusions presented in section 4.

2. Problem statement and mathematical model

2.1 Substrate EON

The substrate elastic optical network was represented as an undirected graph weighted on both nodes and edges, $G^s = (V^s, E^s)$, where V^s is the set of substrate optical nodes, E^s is the set of optical fiber links, each link is connected to a pair of nodes e.g., $e^s = (s, d) \in E^s$, where s (source node) and d (destination node) $\in V^s$. Each link $e^s \in E^s$ has usable bandwidth, B. The bandwidth B is partitioned into multiple N equal-width slots represented by the set of frequency slots FS^s. To be used efficiently, all nodes have a capacity Cap^s of available computing resources (CPU) associated with each node, each physical link has a capacity represented by several frequency slots FS^s which are equal to N, the set of substrate paths denoted by PP. The number of physical optical links for a physical path pp is represented by $|pp|$ and the path physical length (measured by kilometers) is $length(pp)$. The set of the network delay associated with each substrate link is denoted by L^s. Figure. 1 is an illustration of a substrate network and a virtual optical network request (VONR) in (a), where the numbers inside the red rectangles are the available CPU of nodes, the numbers beside edges are the link available FS^s number, and the numbers inside green octagons are the link latency.

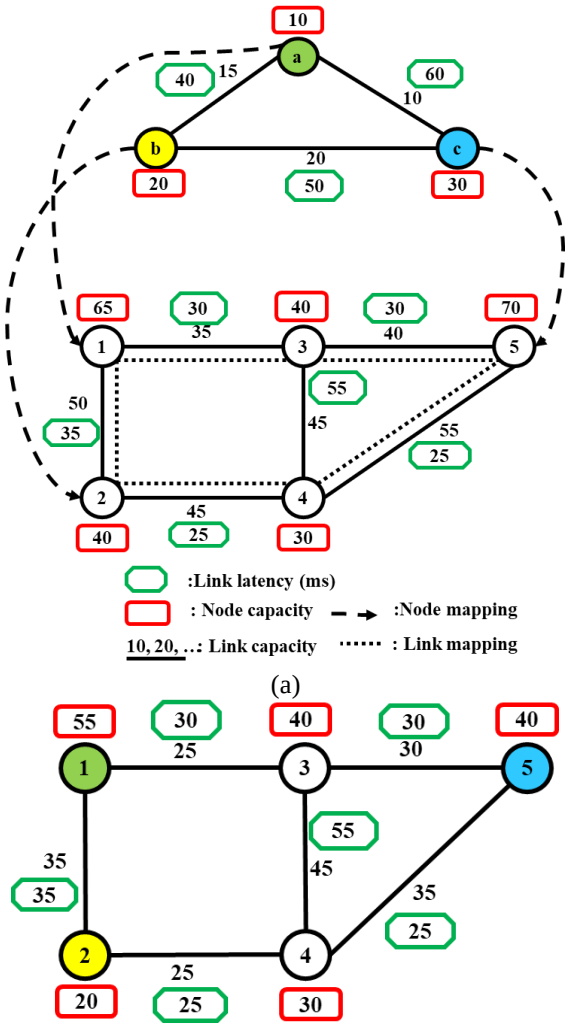


Figure. 1 Virtual optical network request mapping: (a) virtual optical network request and substrate optical network, and (b) a possible mapping of virtual optical network request and resources updating

Fig. 1. (a) shows the result for node mapping of a VONR, which is $\{a \rightarrow 1, b \rightarrow 2, and\ c \rightarrow 5\}$. After all link mapping constraints are satisfied, the result for link mapping of a VONR by dotted line graphically representation, which is {(ab) → (12), (bc) →(24,45), and (ca) →(53,31)}. Fig. 1. (b) represents the final result of mapping with resource updating.

2.2 Virtual optical network request

The virtual optical network request $VONR$ was represented as an undirected graph weighted on both nodes and edges, $G^v = (V^v, E^v)$, where V^v is the set of virtual nodes, E^v is the set of virtual links, and each link is connected to a pair of virtual nodes. Each $VONR$ has a specific demand for node and link mapping on the substrate optical network. The notations are mentioned in Table 1.

2.3 Virtual network mapping

Virtual network mapping VNM is defined as MP: $G^v \rightarrow G^s$, which embeds the virtual optical network to the substrate optical network with constraints satisfaction as shown in Fig. 1 illustrates an example of VONR consisting of virtual nodes and virtual links with mapping requests. The VNM could divided into two phases:

2.3.1. Node mapping phase

A mapping function provided by node mapping phase from a virtual node to a substrate node, MP_V: $V^v \rightarrow V^s$, such that: $\forall\, i, j \in V^v$. When a $VONR$ arrived, it consists of several virtual nodes that need to be mapped to a set of substrate nodes after satisfying constraints in Eq. (1), Eq. (2) and Eq. (3), where $\varepsilon_{v^s}^{v^v,vor}$ is a Boolean. If i^{th} virtual optical network request ($vonr$) with virtual node v^v mapped successfully to a substrate node v^s, $\varepsilon_{v^s}^{v^v,vonr} = 1$, otherwise $\varepsilon_{v^s}^{v^v,vonr} = 0$. A virtual node mapped on substrate node is represented in Eq. (1) with a capacity Cap^v less than the total capacity of computing resource for substrate node Cap^s. Each virtual node v^v mapped to only one unique substrate node v^s as in Eq. (2). In Eq. (3) a substrate node v^s can host only one virtual node v^v for the same $vonr$.

$$\sum_{vonr \in VONR} \sum_{v^v \in V^v} Cap^v \cdot \varepsilon_{v^s}^{v^v,vonr} \leq Cap^s \quad \forall v^s \in V^s \tag{1}$$

$$\sum_{v^v \in V^v} \varepsilon_{v^s}^{v^v,vonr} \leq 1 \quad \forall v^s \in V^s, \forall vonr \in VONR \tag{2}$$

$$\sum_{v^s \in V^s} \varepsilon_{v^s}^{v^v,vonr} = 1 \quad \forall v^v\ V^v, \forall vonr \in VONR \tag{3}$$

2.3.2. Link mapping phase

After node mapping is satisfied in the first phase, the link mapping MP_E: $E^v \rightarrow PP$ provides a virtual link mapping function to a shortest physical path PP which consist of one or more links between mapped nodes as in Eq. (4), such that: $\forall\, e^v = (i, j) \in E^v$ where e^v is a virtual link connecting virtual nodes i and j. Also, the virtual links are embedded on substrate paths that satisfy the constraint of bandwidth capacity B^s represented by a number N of frequency slots $|FS^s|$ and $|FS^v|$ is the requested frequency slots number as in Eq. (5) and Eq. (6) below:

$$pp \geq e^s \quad \forall\, e^s \in E^s \tag{4}$$

Table 1. Notations

Parameter	Description
G^s	Substrate network
V^s	Set of substrate nodes
E^s	Set of substrate links
s	Source node
d	Destination node
e^s	Substrate link between two nodes
B^s	Available bandwidth capacity
N	Number of slots in substrate link
FS^s	Substrate frequency slots
Cap^s	Substrate node capacity
CPU	Maximum node capacity
PP	Set of physical paths
L^s	Set of substrate link latency
L^v	Set of requested latency
VONR	Set of virtual optical network requests
G^v	Virtual network
V^v	Set of virtual nodes
E^v	Set of virtual links
MP_V	Node mapping function
MP_E	Link mapping function
Cap^v	Virtual node capacity
$\|FS^s\|$	Number of substrate frequency slots
$\|FS^v\|$	Number of requested frequency slots
pp	Physical path between s and d nodes
$\|pp\|$	Number of physical links in physical path
B^v	Requested virtual bandwidth
$S_{x,x'}$	Continuity vector
$s_i^{x,x'}$	Element of continuity vector
x, x'	Representation of substate fiber links
CNF	Continuity factor
SFL	Substrate fiber link
$CGF_{x,x'}$	Contiguity factor for substrate links
$\|E^s\|$	Number of substrate fiber links
L_{PP}	Physical path latency
VNE	Virtual network embedding
$n_{amp}(pp)$	Number of amplifiers on a lightpath
f_{span}	Typical distance between two amplifiers
Lat_{max}, Lat_{min}	Minimum and maximum value of substrate links latencies in EON
L_{norm}	Normalized latency
$Ls(s,d)$	Latency for substrate link between s and d nodes
CNF_{v^s}	Continuity factor for substrate node
CGF_{v^s}	Contiguity factor for substrate node
X^{v^s}	Set of substrate fiber links connected to substrate node
NA_{v^s}	Node availability factor
NPF_{pp_i}	Node path influence factor of path pp_i
SW	Slot width (GHz)

FL x	1	1	0	1	1	1	0	0	1	1
FL x'	1	0	1	1	1	0	0	1	1	1

Figure. 2 Logical representations of substate fiber links x, x'

$$B^v(e^v) \leq B^s(e^s) \tag{5}$$

$$|FS^v| \leq |FS^s| \tag{6}$$

2.4 EON constraints

To meet the bandwidth constraints, link mapping needs to ensure that the selected FS^s s must be contiguous and continuous along the selected physical path. Also, the link latency will be considered with previous constraints in a new node availability factor to be chosen for more efficient node mapping.

2.4.1. Continuity and contiguity constraints

The continuity vector can be defined as:

$$S_{x,x'} = vect_{i=1\ldots,N}\{\delta_{x,i}^{v^s} \,\&\, \delta_{x',i}^{v^s}\} \tag{7}$$

where N represents the total frequency slot number of each substrate fiber link (FL), $i = (1 \ldots, N)$, $x = (1 \ldots, |E^s|)$, δ^{v^s} considered as the logical link connected to substrate node v^s. When the i^{th} frequency slot in link x^{th} is available, $\delta_{x,i}^{v^s} = 1$, otherwise, $\delta_{x,i}^{v^s} = 0$. After that, a vector with N length will obtained. The continuity vector obtained from Fig. 2 below, and the result is $S_{x,x'} =$ (1,0,0,1,1,0,0,0,1,1). The occupied frequency slots are represented with 0, while 1 represents the available frequency slots.

A continuity factor (CNF) can be defined as:

$$CNF_{x,x'} = \sum\{S_{x,x'}\} \tag{8}$$

CNF is a representation of summing all elements in vector $S_{x,x'}$. The spectrum continuity for the substrate fiber links (SFLs) could be measured by using continuity factor. Though, the good spectrum continuity for the substrate link does not mean having sufficient spectrum resources, so the second constraint which is contiguity must be considered, which is defined as:

$$CGF_{x,x'} = \frac{1}{\sum_{i=2}^{N} |s_i^{x,x'} - s_{i-1}^{x,x'}|} \tag{9}$$

where $s_i^{x,x'}$ represents the vector $S_{x,x'}$ element. The degree of contiguous frequency slots factor for not occupied slots represented by Eq. (9). The denominator of Eq. (9) represents the summation of difference between the adjacent elements of the continuity vector. The larger amount of the denominator means less contiguous frequency slots. Although in Fig. 3. (a) the continuity factor is the same as in Fig. 3. (b) (this implies that continuity resources of the spectrum are equal), contiguity factor in Fig.3. (a) is far superior to that in Fig.3. (b)., that means Fig.3. (b) has less consecutiveness slots than Fig.3. (a).

FLx	1	1	1	1	0	0	1	1	1	0
FLx'	1	1	0	0	1	1	1	1	1	0

(a)

FLx	1	0	1	1	1	1	0	1	1	1
FLx'	1	1	0	1	0	1	0	1	0	1

(b)

Figure. 3 The varying of aggregation degrees for free slots in a fiber link (a) and (b)

Table 2. Node latency contribution in EON.

Latency contributor	Value
OTN/FlexE elements (L_{otn})	Neglected
Transponders (L_{txp})	30 ns
FEC (L_{fec})	10 µs for std., 150 µs for super

Table 3. Physical lightpath latency contribution in EON.

Latency contributor	Value
Fiber propagation (L_{prop})	4.9 µs/km
Re-generators (L_{rgn})	Not considered
Amplifiers (L_{amp})	150 ns per unit
CD compensation (L_{dcf})	Not considered
ROADMs and BV-OXCs (L_{roadm})	30 ns

2.4.2. Latency constraint

The other important constraint which we take into consideration is the latency constraint, which must be satisfied in addition to the continuity and contiguity constraints during the embedding process. The latency contribution in the elastic optical network (EON) could be found in nodes and lightpaths as shown in Table 2, and Table 3.

The total latency along a lightpath is calculated as:

$$L_{PP} = L_V + L_{path} \tag{10}$$

where L_V is the latency contribution from a node given by:

$$L_V = 2(L_{txp} + L_{fec}) \tag{11}$$

and the path latency is given by:

$$L_{path} = length(pp)L_{prop} + n_{amp}L_{amp} + (|pp| + 1)L_{roadm} \quad (12)$$

The total latency along the lightpath in Eq. (10) including both node and lightpath latencies will aid in determining the lightpaths that may be used for VNE. L_{PP} represents the physical path latency, $|pp|$ is the number of physical link(s) along the lightpath, n_{amp} represents the number of amplifiers on a lightpath pp, $length(pp)$ is the physical length of pp in kilometer. The number of amplifiers on a lightpath pp, $n_{amp}(pp)$ is $\approx \lceil length(pp)/\, f_{span} \rceil$, where f_{span} stands for typical distance between two amplifiers, which is called fiber span (in the vicinity of 80 kilometers) [25].

$$L^v(e^v) \geq \sum_{e^s \in pp\,(MP_E(e^v))} L^s(e^s) \quad (13)$$

For every substrate fiber link in the physical path, it is essential to have no less than the same amount of virtual link resources. Furthermore, the total amount of latency for every physical path have to meet the latency requested by a virtual link. For a coming virtual optical network request ($VONR$), meeting the spectrum continuity, contiguity, and link latency constraints is an essential requirement [1], otherwise, the request will be blocked. The latency for each substrate FL is calculated using Eq. (12), the value of delay for each virtual link in a virtual request is between the range of minimum and maximum value of substrate links latencies in EON which is denoted by $[Lat_{max}, Lat_{min}]$.

2.5 Node-load resource balancing availability

After the request arrived with specific demands of resources (CPU, bandwidth, latency), the first embedding phase performed is the node availability resources (Node Mapping phase). A suggested method to measure the best node to be mapped is to determine the influence of all ink resources connected with the node besides CPUs as follows:

The availability of optical link connected to the node as load factor is given by:

$$Load_{v^s} = \frac{\sum_{x=1}^{X^{v^s}} (N - \sum_{i=1}^{N} \delta_{x,i}^{v^s})}{N} \quad (14)$$

where X^{v^s} are the fiber links connected to the substrate node, the link with heavy load has a large load factor. In another word more requests received and mapped to a FL means more load on the link. The link with a low load will accept more new requests which are good to reduce the number of blocked requests.

The effect of latency is given by:

$$L_{norm} = \frac{(Lat_{max} - Ls(s,d))}{(Lat_{max} - Lat_{min})} \quad (15)$$

where L_{norm} normalized latency, $Ls(s,d)$ is the latency of the substrate fiber link connected between source s and destination d substrate nodes. The normalized latency factor L_{norm} is between the range [0, 1], where 1 is a representation of minimum latency for links and 0 represents the maximum latency of links [14].

The continuity factor of node v^s is given by:

$$CNF_{v^s} = \frac{\sum_{x,x' \in X^{v^s}, x \neq x'}^{X^{v^s}} \{CNF_{x,x'}\}}{X^{v^s}.(X^{v^s} - 1)/2} \quad (16)$$

The contiguity factor of node v^s is given by:

$$CGF_{v^s} = \frac{\sum_{x,x' \in X^{v^s}, x \neq x'}^{X^{v^s}} \{CG_{k,k'}\}}{X^{v^s}.(X^{v^s} - 1)/2} \quad (17)$$

To determine the best nodes to be chosen for the node mapping phase, all four factors above are jointly measured in addition to the CPU resources and represented in a new factor called Node Availability factor (NA), given by:

$$NA_{v^s} = \frac{Cap(v^s)}{Cap(v_{max}^s)} . \frac{CNF_{v^s}}{N} . CGF_{v^s} . \frac{1}{Load_{v^s} + 1} . \frac{1}{L_{norm} + 1} \quad (18)$$

$Cap(v_{max}^s)$ and N are used to normalize the variables. To avoid the 0 of the denominators, 1 was added in both load and latency terms in Eq. (18). When virtual links are mapping, the changing in node availability rank was considered. By assuming the virtual link is mapped on the substrate path pp_i, which consist of one or more FLs and many substrate nodes which represented by $y_1, \dots, y_{n-1}$ and $z_1, \dots, z_n$, respectively, n is the number of substrate nodes along the path, j is an integer real number $j = 1 \dots, n$. The node path influence factor (NPF) is given by:

$$NPF_{pp_i} = \sum_{j=1}^{n} \frac{NA_{z_j} - NA_{z_j}^{pp^i}}{NA_{z_j}} \quad (19)$$

where NPF_{pp_i} is the node path influence factor of path pp_i. NA_{zj} represents the availability rank of the

node z_j and $NA_{zj}^{pp^i}$ is the availability rank for the node z_j if the node z_j is included in the path pp_i. Eq. (19) determines the different values of node availability rank. In link mapping phase k-shortest paths first found between the chosen substrate nodes, the node that has the lowest node path influence factor will be chosen for virtual link mapping. In other words, the link mapping phase has the minor impact on the substrate nodes resources. When the virtual network request arrives in the future, the substrate nodes will still have adequate resources to map.

The objective equation is given by:

$$L = min \sum_{e^v \in E^v} \sum_{e^s \in E^s} \sigma_{e^s} . \left(N \sum_{i=1}^{N} \delta_{e^s,i}^{e^v}\right) + \sum_{e^v \in E^v} \sum_{i=1}^{pp} NPF_{pp_i}^{e^v} \quad (20)$$

where σ_{e^s} is Boolean variable. σ_{e^s}= 1 when FL e^s is a part of the spectrum path pp_i, otherwise, σ_{e^s} = 0.

There are two main objectives for adding the node-load resource balancing factor: First, is to minimize the blocking probability during the embedding process because the node selection will obey a new search restrictions based on all resources added to the node (no. of connected links, link bandwidth, load of available bandwidth in each link). While the selection in other studies were based only in the node load (CPU resources) without considering the link bandwidth connected to it. Secondly, another additional factor was inserted in this selection mechanism, which is the latency of each attached link (after latency computation), together with node latency contribution. According to that, the blocking probability of future VONRs can be minimized by taking into account the node path influence factor.

3. Evaluation parameters

3.1.1. Blocking probability

The important factor to be tasted and compared between the two network topologies is the blocking probability. It defined as the ratio of rejected requests (RJ) summation over the total number of VONRs ($|VONR|$) during the unit time, which is given by [26]:

$$BP = \frac{\sum RJ}{|VONR|} \quad (21)$$

3.1.2. Node utilization

Node utilization (NU) is defined by the summation of the occupied CPU resource in the substrate node from embedded VON requests divided by the total amount of CPU resource capacity of the substrate optical network [27], and can be written as:

$$NU = \frac{\sum_{v^s \in V^s} \sum_{vor \in VOR} \sum_{v^v \in V^v} CAP^v . \varepsilon_{v^s}^{v^v, vor}}{\sum_{v^s \in V^s} CAP^s} \quad (22)$$

3.1.3. Spectrum utilization

Spectrum utilization (SU) is defined by the summation of the occupied frequency slots $FS^v(s)$ from substrate fiber link FL for the embedded VONRs divided by the total amount of FS^s of the substrate FL in the substrate network [27], which can be written as:

$$SU = \frac{\sum_{e^s \in E^s} \sum_{vor \in VOR} \sum_{e^v \in E^v} FS^v . \omega_{e^s}^{e^v, vor}}{\sum_{e^s \in E^s} FS^s} \quad (23)$$

where $\omega_{e^s}^{e^v, vor}$ is a binary variable. When a virtual link of a virtual optical request vor is mapped on substrate path, $\omega_{e^s}^{e^v, vor} = 1$, Otherwise equal 0.

3.1.4. Frequency slots occupation

The number of available frequency slots $|FS|$ for a bandwidth B^s with a certain capacity of optical fiber link is given by [28]:

$$|FS| = \frac{B^s(GHz)}{SW\ (GHz)} \quad (24)$$

where SW represents the FS width (equal to 12.5 GHz). The number of occupied frequency slots $|OFS|$ which achieves the requested bandwidth demands B^v is given by [28]:

$$|OFS| = \left\lceil \frac{B^v(GHz)}{SW\ (GHz)} \right\rceil \quad (25)$$

where $\lceil . \rceil$, is the next higher integer number. Eq. (26) below represents the vector which its elements are the ratio between the occupied and demanded frequency slots.

$$slots\ ratio = \frac{\sum FS^s . \epsilon}{\sum FS^v} \quad (26)$$

The occupied frequency slots ratio (OFSR) could be written as:

$$OFSR = \frac{\sum slots\ ratio}{|slots\ ratio|} \quad (27)$$

where $|slots\ ratio|$ represents the number of elements in $slots\ ratio$ vector.

Algorithm 1: Node mapping

Input: Virtual Optical Network Request $VONR\ (G^v)$, Substrat Network G^s;
Output: Node Mapping MP_V;
1: find NA for each v^s in G^s;
2: erase tabu list; // initiate tabu list;
3: $W \leftarrow$ sort virtual nodes randomly;
4: **for** $i = 0$ to $i = |W| - 1$ **do**
5: # virtual nodes mapping.
6: probability calculation according to NA;
7: after mapping virtual node on substrate node put the chosen substrate
8: node in tabu list;
9: **if** node mapping fails **then**
10: $G^v$is blocked;
11: **end if**
12: **end for**

4. Embedding algorithms

4.1 Node mapping algorithm

The node availability for all substrate nodes in the optical network is determined in node mapping phase. When the virtual node is mapped on a substrate node that fulfills the conditions, a list of reserved substrate nodes is initiated to avoid using the same node for another node mapping in the same VONR. The link mapping phase will be calculated if and only if the request meets all conditions for node mapping. Otherwise, the request will be blocked. The node mapping process is explained in algorithm 1.

4.2 Link mapping algorithm

In link mapping phase as explained in algorithm 2, the k-shortest path is computed including latency, between source and destination substrate nodes that are chosen for node mapping. For efficiency's sake, the substrate fiber links which have not enough frequency slots will be blocked before doing the calculation of the shortest path. If the process of finding the shortest path between two substrate nodes that satisfies the virtual link demand for bandwidth and latency constraints was impossible, then the $VONR$ is rejected. Simply in this phase the shortest path which has the minimum latency among the substrate nodes will be chosen, to reduce resource utilization.

5. Results and discussion

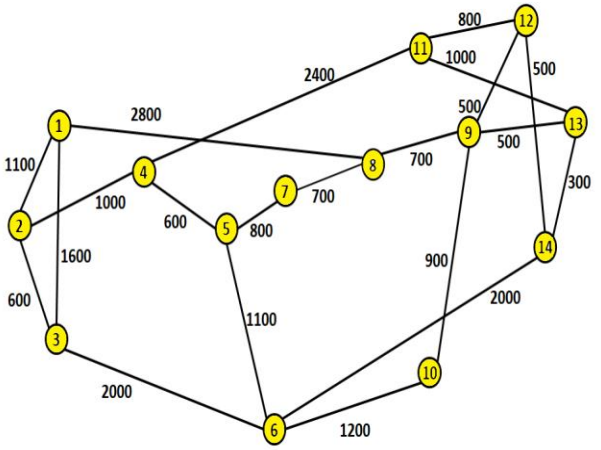
The performance has been extensively evaluated for the virtual optical network embedding through four optimization methods. The performance was evaluated and compared to find the best optimization

Algorithm 2: Link mapping

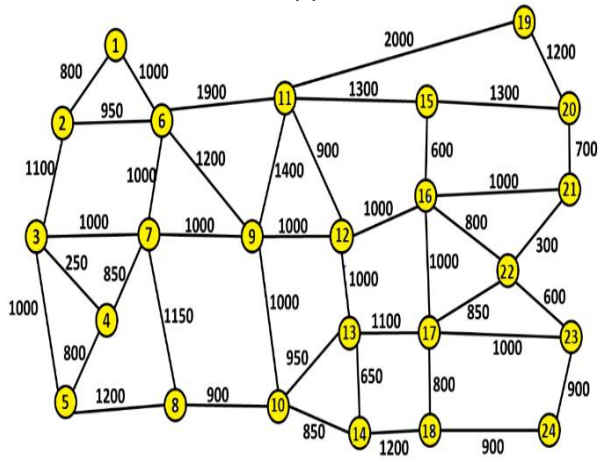
Input: Virtual Optical Network Request $VNOR\ (G^v)$ Substrate Network G^s;
Output: Link Mapping MP_E;
1: $MP_E \leftarrow 0$;
2: $link\_mapping \leftarrow true$;
2: **for** all virtual links $e^v = (i, j) \in E^v$
3: # reject links in G^s that do have not enough
4: # bandwidth
5: **for all** $physical\ links\ e^s \in E^s$ **do**
6: **if** $B^v(e^v) > B^s(e^s)$ **then**
7: reject links $e^s \in E^s$;
8: **end if**
9: **end for**
10: determine the shortest path $pp\ (MP_V(i), MP_V$
11: between the two nodes $MP_V(i)$ and $MP_V(j$ in G^s with latency consideration;
12: # if the shortest path including latency calcula
13: # was found and matched the VOR for latenc
14: # constraint demands
15: **if** $|pp\left(MP_V(i), MP_V(j)\right)| \neq 0$ and $pp\left(MP_V(i), MP_V(j)\right).\ sum\_L^s \leq L^v(e^v)$ **then**
16: # mapping process
17: $MP_E(i\,j) \leftarrow pp\ (MP_V(i), MP_V(j))$;
18: update the bandwidth B^s of the substrate links (e^s)s involved in G^s;
19: **else**
20: $link\_mapping \leftarrow false$;
21: the $VONR$ is marked as blocked;
22: **break**;
23: **end if**
24: **end for**
25: **if** $link\_mapping \leftarrow false$ **then**
26: undo bandwidth B^s and CPU resources update in G^s;
27: **end if**

method result for virtual optical network embedding (VONE). Two optical network topologies national science foundation network (NSFNET) 14-nodes 21-links for small scale simulation, and 24- nodes 43-links USNET for large scale simulations in Fig. 4, were used as the substrate physical platform. The simulations were done by using MATLAB R2019a under Windows10 operating system for P.C. with processor Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz 1.99 GHz, RAM 16.0 GB.

The optimization methods are; alignment and consecutiveness-aware virtual network embedding (ACT-VNE), the local resource capacity and the shortest path first fit (LRC-SP-FF), greedy algorithm and the shortest path first fit (Greedy-SP-FF), load and resource-aware based on ant colony optimization



(a)



(b)

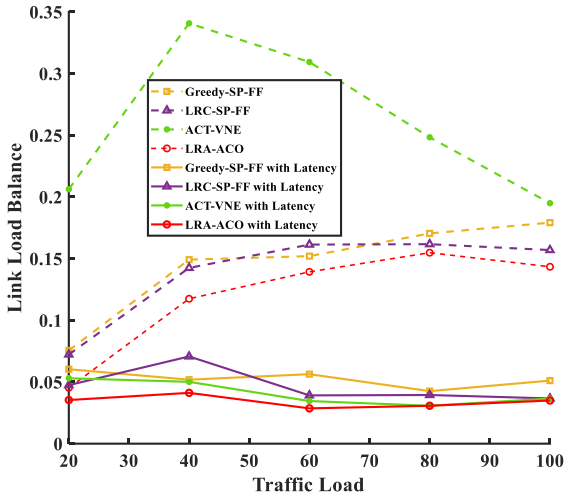
Figure 4. Network topologies: (a) 14 nodes, 21 link NSFNET topology and (b) 24 nodes, 43 link USNET topology

Table 3. Simulation parameters

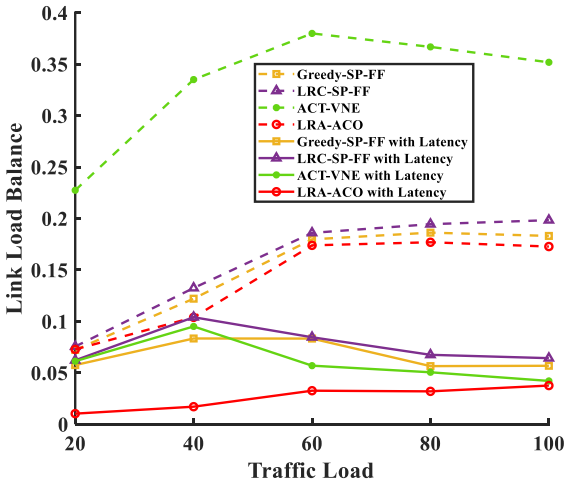
Parameter	NSFNET	USNET
Substrate fiber link number	21	43
Substrate node number	14	24
Computing unit capacity in a physical node (CPU)	400 units	400 units
Frequency slots no. for each SFL (BW capacity)	200	200
Computing unit capacity in a virtual node (CPU)	[1-3] units	[1-3] units
Frequency slots no. for each virtual link	[1-5]	[1-5]
Number of virtual nodes in virtual optical network	[2-5]	[2-5]
Latency for a virtual link (ms)	[20-50]	[20-50]

(LRA-ACO). The simulation parameters used are shown in Table 3.

Each frequency slot has a bandwidth granularity of $12.5GHz$, so the total bandwidth for each SFL is $2.5THz$. BPSK modulation was assumed to be used in the lightpath. The virtual optical network requests



(a)

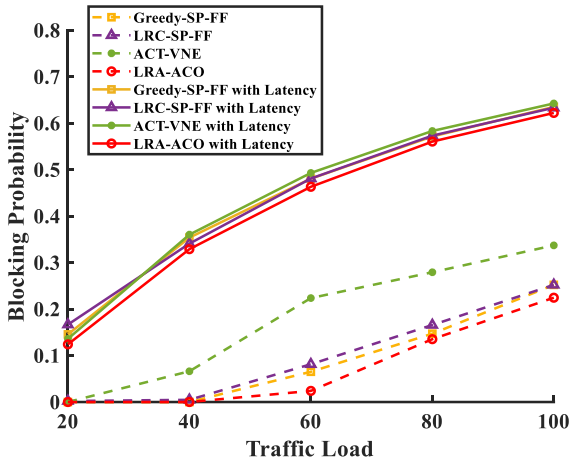


(b)

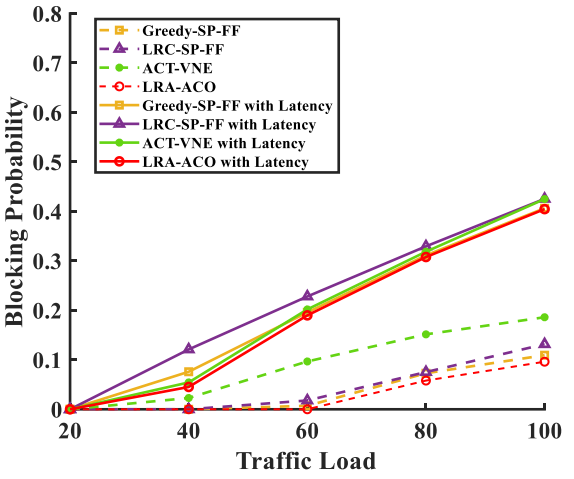
Figure. 5 Load Balancing: (a) NSFNET and (b) USNET

(VONRs) arrival process is assumed to obey a Poisson process with rate λ, each request has an exponential lifetime with an average μ. The VONRs are generated with random demands for the virtual nodes and their CPUs in each one, FSs number, and latency of each virtual link. Virtual nodes are connected randomly to form a virtual optical network with a probability of p. The comparison of results was made between NSFNET and USNET topologies with and without latency limitation, and for existing work in NSFNET without latency in [23]. Many metrics were used to determine the effect of EONs constraints on both topologies. After running the simulation program, the average number of arrived requests was around 7500 requests during a time units. More than 50 runs for each network topology were performed.

Fig. 5, represents the link-load balance variation for both NSFNET (Fig. 5.a) and USNET (Fig. 5.b) topologies. NSFNET results are almost similar to



(a)



(b)

Figure. 6 Blocking probability: (a) NSFNET and (b) USNET

Table 4. comparison of optimization methods for NSFNET

Optimization method	Success requests no. without latency	Success requests no. with latency
LRA-ACO	5700	2766
LRC-SP-FF	5544	2757
Greedy-SP-FF	5500	2744
ACT-VNE	4914	2688

Table 5. Comparison of optimization methods for USNET.

Optimization method	Success requests no. without latency	Success requests no. with latency
LRA-ACO	6665	4429
Greedy-SP-FF	6568	4358
LRC-SP-FF	6404	4231
ACT-VNE	6002	4299

[23] for non-latency case, but the latency case experiences a distinguishable reduction of about 30 % (from 0.15 to around 0.05). When comparing the optimization methods, the LRA-ACO performance gave mostly the same results in [23] for non-latency NSFNET case. This performance comes from considering the load balance in the virtual node and virtual link mapping. The ACT-VNE has the worst performance exceeding a 0.35 value. Including latency alters the difference between the optimization methods performance, but LRA-ACO still has the lowest link-load balance value.

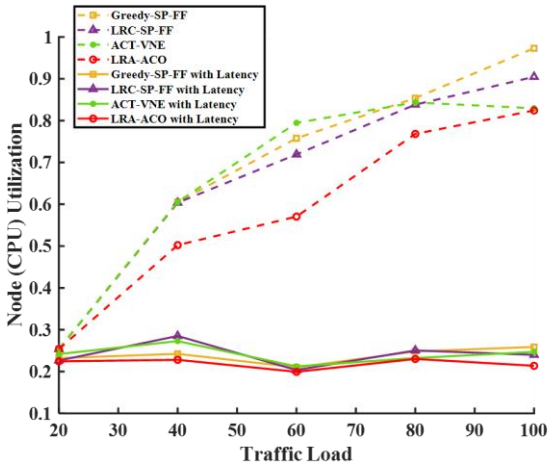
Results for USNET (Fig. 5.b) showed approximately similar behavior for non-latency case as NSFNET (0.1-0.4). LRA-ACO also showed the best link-load balance in both cases. Adding latency influenced both networks by decreasing the load-link balance along with all traffic loads. In this approach, the FL with the minimum load balance factor will be chosen for link mapping, to minimize the occupied FSs.

Fig. 6.a, compares the *BP* in NSFNET. Also, LRA-ACO revealed the best *BP*, noticing values approximately like [23] for the non-latency case. For latency case the *BP* for all methods increased by approximately 50 %. The difference in the efficiency of the four optimization methods is less distinguishable for the latency, maintaining an advantageous *BP* values for LRA-ACO.

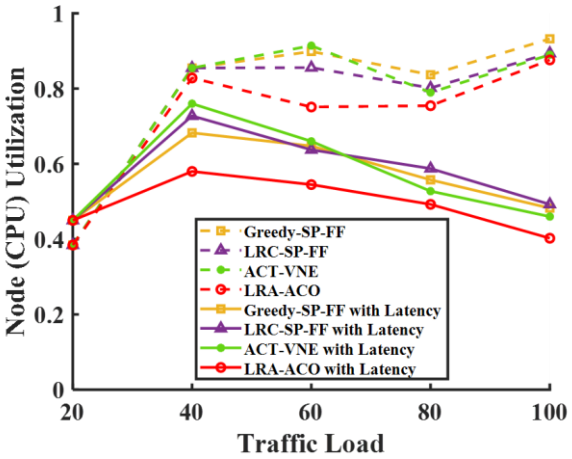
There is a distinguishable decrease in the *BP* range in USNET for both cases (with and without latency) compared to NSFNET e.g., in Fig. 6.b at the traffic load (60) the decrease in *BP* average value is more than 50 % compared to Fig. 6.a (from 0.5 in NSFNET to 0.2 in USNET) for the latency case. It is clear that the abundance of resources in USNET lead to an increase in accepted requests at the expense of rejected ones which as a result reduced the *BP*.

Also, for both network topologies, there is approximately 50 % increase in *BP* values between the results with latency as compared to the non-latency case. This increase is due to adding additional constraint represented by latency besides the existing ones (contiguity and continuity). Constraints in EONs are already difficult to achieve in embedding process. When the complexity increased by adding another condition, this leads to limiting the choices for mapping. Table 4, and Table 5 represent a comparison for accepted requests number between the optimization methods to both topologies in both cases.

For resource utilization in both networks with both cases (non-latency and latency) Eq. (22) and Eq. (23) were used for node and link utilization respectively. Fig. 7.a shows the node utilization for NSFNET. The utilization for all optimization methods lies within the boundary of (0.3-1) for non-

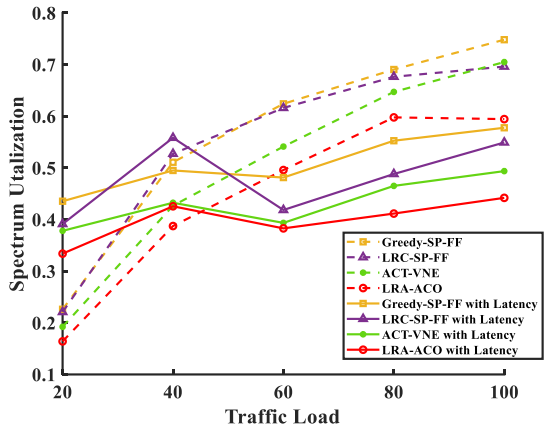


(a)

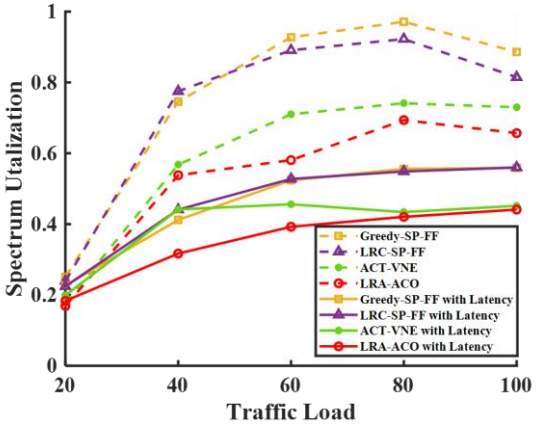


(b)

Figure. 7 Node Utilization: (a) NSFNET and (b) USNET



(a)



(b)

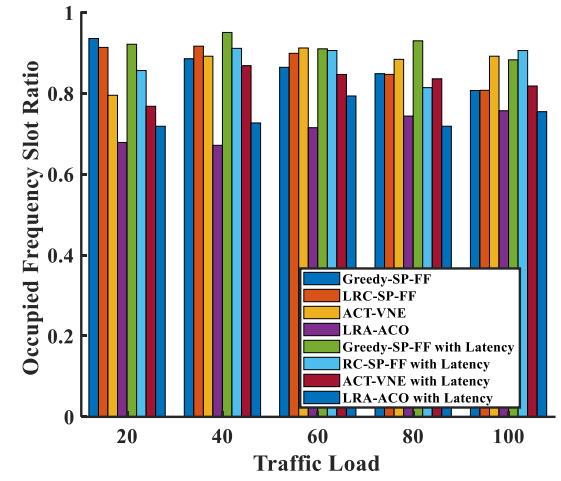
Figure. 8 Spectrum utilization: (a) NSFNET and (b) USNET

latency case. LRA-ACO have different response along the traffic load recording the lowest node utilization. For latency case, the node utilization average value was markedly reduced for all optimization methods ($\leq$0.3). The effect of adding latency was approximately of equal influence on all traffic loads. For USNET (Fig. 7.b), Adding latency reduced clearly the node utilization. Relating to *BP* (Fig. 6), the impact of that behavior can be distinguished (specially after the 70 traffic load point) causing an increase in the BP.

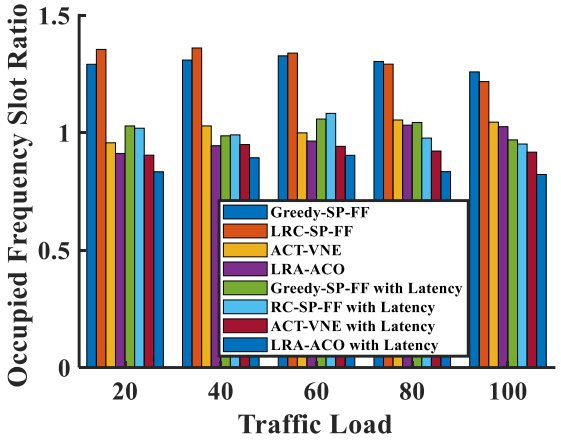
Fig. 8 illustrates the spectrum utilization for both scenarios in both topologies. A better spectrum utilization is obtained with the LRA-ACO for both cases of latency, with the fact that for latency case the spectrum utilization range is reduced compared to non-latency for all optimization methods in both network topologies. In both Fig. 8.a and Fig. 8.b, the effect of latency causes less spectrum utilization in both networks for all optimization methods, which means increasing in BP. Ideally the resources are all used when the NU and SU values equal to 1, that means all VONRs are accepted. So, when NU and SU equal or less than 1, leads to reduce BP value. Including latency is clear on reducing values of NU and SU for both NSFNET and USNET topologies, but it is clearer in USNET which it peaks values reduced more. The resource utilization peak values in USNET are higher in values means more accepted requests.

In Fig. 9.a and Fig. 9.b for NSFNET and USNET respectively, the first four bars represent the occupied frequency slots ratio (OFSR) without latency constraint and the second four bars represent the ratio with latency for both topologies. The difference between optimization methods performance was clear in NSFNET. In comparison with literature (as in [23]), there is difference in highest value reached for OFSR for all optimization methods, but still LRA-ACO leads to better performance. The difference in results comes from difference in resource capacities. When latency considered, OFSR for Greedy-SP-FF was the highest (except at 60 and 100 load, slightly better than LRC-SP-FF).

Adding latency effect diminished the OFSR using all optimization methods. The maximum OFSR value in USNET (Fig. 9.b) on average was higher than



(a)



(b)

Figure. 9 Ratio for occupied frequency slots: (a) NSFNET and (b) USNET

recognized in NSFNET for latency and non-latency cases; while the maximum OFSR does not exceed 0.8 in NSFNET for both cases. In USNET it was between (0.8 – 0.9) for LRA-ACO method.

6. Conclusions

Due to the increased demand for low-latency service, an analytical model was presented to add latency effect in the VONE over EONs. A new extended link-load balance formulation was obtained after adding the latency contribution.

Latency is evaluated in two network topologies, the results are obtained and compared using four optimization methods. Adding latency constraint overload both network topologies (NSFNET and USNET), affecting different metrics i.e., BP, resource utilization, frequency slots occupation, and load balance, showing a distinguishable decrease in load balance as compared to other researches that have not considered latency constraint. Also, latency adds more complexity to the embedding process in addition to the complexity of spectrum allocation process. The reduction in success requests was clear from the results.

USNET topology had more flexible BP as a result of doubling resources (nodes, links) which leads to more acceptance rate of requests compared to the NSFNET network. It was clear that its OFSR was better for both cases (with and without latency effect). Minimizing the occupied frequency slots leads to reduce BP. This reduction is caused by the abundance of available FSs along the fiber links to accept new incoming VONRs.

Finally, LRA-ACO proved to be the most efficient approach in accommodating the additional load along with the traffic load for both topologies. It ameliorated all analyzed metrics used to study the performance of both topologies.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, O. Y. Shaaban; methodology, O. Y. Shaaban, and S. A. Ibraheem; software, S. A. Ibraheem; formal analysis, O. Y. Shaaban; writing—original draft preparation, review and editing, O. Y. Shaaban, and S. A. Ibraheem; supervision, O. Y. Shaaban.

References

[1] M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, J. Sachs, J. Markendahl, and M. Dohler, “Business case and technology analysis for 5G low latency applications”, *IEEE Access*, Vol. 5, pp. 5917–5935, 2017.

[2] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, “A survey on low latency towards 5G: Ran, core network and caching solutions”, *IEEE Communications Surveys & Tutorials*, Vol. 20, No. 4, pp. 3098–3130, 2018.

[3] A. Technologies, *State of Online Retail Performance – 2017 Holiday Retrospective*, White paper, 2017.

[4] R. Martin, *Wall Street’s Quest to Process Data at the Speed of Light*, InformationWeek, 2007.

[5] Global Mobile Suppliers Association, *The Road to 5G: Drivers, Applications, Requirements and Technical Development*, White paper, 2015.

[6] Huawei, *5G: A Technology Vision*, White paper, 29 Jan 2014.

[7] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5G: Survey and

challenges", *IEEE Communications Magazine*, Vol. 55, No. 5, pp. 94–100, 2017.

[8] N. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization", *Computer Networks*, Vol. 54, No. 5, pp. 862–876, 2010.

[9] A. Fischer, J. F. Botero, M. T. Beck, H. D. Meer, and X. Hesselbach, "Virtual network embedding: A survey", *IEEE Communications Surveys & Tutorial*s, Vol. 15, No. 4, pp. 1888–1906, 2013.

[10] O. Y. Shaaban and O. A. Athab, "Hybrid compensation of polarization-multiplexed QPSK optical format for high bit rate networks", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 20, No. 3, pp. 1325-133, 2020.

[11] R. Lin, S. Luo, J. Zhou, S. Wang, A. Cai, W. Zhong, and M. Zukerman, "Virtual Network Embedding with Adaptive Modulation in Flexi-Grid Networks", *J. Light. Technol*, Vol. 36, No. 17, pp. 3551–3563, 2018.

[12] M. Zhu, Q. Sun, S. Zhang, P. Gao, B. Chen, and J. Gu, "Energy-Aware Virtual Optical Network Embedding in Sliceable-Transponder-Enabled Elastic Optical Networks", *IEEE Access*, Vol. 7, pp. 41897–41912, 2019.

[13] W. Wei, H. Gu, A. Pattavina, J. Wang, and Y. Zeng, "Optimizing energy and spectrum efficiency of virtual optical network embedding in elastic optical networks", *Opt. Switch. Netw.*, Vol. 37, 2020.

[14] F. Bianchi and F. L. Presti, "A Markov reward-based resource-latency aware heuristic for the virtual network embedding problem", *ACM SIGMETRICS Performance Evaluation Review*, Vol. 44, No. 4, pp. 57– 68, 2017.

[15] K. Hejja and X. Hesselbach, "Online power-aware coordinated virtual network embedding with 5G delay constraint", *Elsevier Journal of Network and Computer Applications*, Vol. 124, pp. 121–136, 2018.

[16] K. Ivaturi and T. Wolf, "Mapping of delay-sensitive virtual networks", In: *Proc. of IEEE International Conference on Computing, Networking and Communications (ICNC)*, pp. 341–347, 2014.

[17] J. A. Jay, "Low Signal Latency in Optical Fiber Networks", In: *Proc. of the 60th International Wire & Cable Symposium (IWCS)*, NC, USA, pp. 429-437, 2011.

[18] V. Bobrovs, S. Spolitis, and G. Ivanovs, "Latency causes and reduction in optical metro networks", *Opt. Metro Networks Short-Haul Syst. VI*, Vol. 9008, pp. 90080C1-90080C11, 2013.

[19] R. Nejabati, E. Escalona, S. Peng, and D. Simeonidou, "Optical network virtualization", In: *Proc. of 15th International Conference on Optical Networking Design and Modeling*, Bologna, Italy, 2011.

[20] S. Peng, R. Nejabati, S. Azodolmolky, E. Escalona, and D. Simeonidou, "An impairment-aware virtual optical network composition mechanism for future Internet", *Optics Express*, Vol. 19, No. 26, 2011.

[21] A. Pagès, J. Perelló, and S. Spadaro, "Virtual Network Embedding in Optical Infrastructures", In: *Proc. of 14th International Conference on Transparent Optical Networks (ICTON)***,** United Kingdom, 2012.

[22] N. Shahriar, S. Taeb, S. R. Chowdhury, M. Tornatorey, R. Boutaba, J. Mitraz, and M. Hemmati, "Achieving a Fully-Flexible Virtual Network Embedding in Elastic Optical Networks", In: *Proc. IEEE Conference on Computer Communications IEEE INFOCOM*, Paris, France, 2019.

[23] R. Yang, K. Wang, B. Wang, W. Wei1, X. Liu, Y. Guo, and H. Gu, "Resource and load aware mapping algorithm for elastic optical network", *IEICE Electronics Express*, Vol. 13, No. 14, 2016.

[24] S. Taeb, N. Shahriar, S. R. Chowdhury, M. Tornatore, R. Boutaba, J. Mitra, and M. Hemmati, "Virtual network embedding with path-based latency guarantees in elastic optical networks", In: *Proc. 2019 IEEE 27th Int. Conf. Netw. Protoc. ICNP*, pp. 1-12, 2019.

[25] Y. Zhu and M. Ammar, "Algorithms for Assigning Substrate Network Resources to Virtual Network Components", *INFOCOM*, 2006.

[26] T. Chen, J. Liu, Q. Tang, T. Huang, and R. Huo, "Virtual Network Embedding Algorithm for Location-Based Identifier Allocation", *IEEE Access*, Vol. 7, pp. 31159–31169, 2019.

[27] M. Zhu, Q. Sun, S. Zhang, P. Gao, B. Chen, and J. Gu, "Energy-Aware Virtual Optical Network Embedding in Sliceable- Transponder- Enabled Elastic Optical Networks", *IEEE Access*, Vol. 7, pp. 41897–41912, 2019.

[28] M. Gharbaoui, B. Martini, G. Cecchetti, and P. Castoldi, "Resource Orchestration Strategies with Retrials for Latency-Sensitive Network Slicing Over Distributed Telco Clouds", *IEEE Access*, Vol. 9, 2021.

[29] P. Afsharlar, A. Deylamsalehi, J. M. Plante, J. Zhao, and V. M. Vokkarane, "Routing and Spectrum Assignment with Delayed Allocation in Elastic Optical Networks", *Optical Society of*

America, Vol. 9, No. 3, 2017.
[30] Q. Zhu, X. Yu, Y. Zhao, A. Nag, and J. Zhang, "Auxiliary - Graph - Based Energy-Efficient Traffic Grooming in IP-Over-Fixed/Flex-Grid Optical Networks", *Journal of Lightwave Technology*, Vol. 39, No. 10, 2021.
[31] E. Rodrigues, E. Cerqueira, D. Rosário, and H. Oliveira, "Hybrid Routing, Modulation, Spectrum and Core Allocation Based on Mapping Scheme", *Sensors*, Vol. 20, No. 21, 2020.
[32] S. Shakya, N. Pradhan, X. Cao, Z. Ye, and C. Qiao, "Virtual Network Embedding and Reconfiguration in Elastic Optical Networks", In: *Proc. IEEE Global Communications Conference-Austin*, TX, USA, 2014.