Design and Provisioning of Optical Wireless Data Center Networks: A Traffic Grooming Approach

Abdulkadir Celik, Amer Al-Ghadhban, Basem Shihada, and Mohamed-Slim Alouini

***Abstract*—Traditional wired data center networks (DCNs) suffer from cabling complexity, lack flexibility, and are limited by the speed of digital switches. In this paper, we alternatively develop a top-down traffic grooming (TG) approach for design and provisioning of optical wireless DCNs. While switches are modeled as hybrid opto-electronic crossconnects, links are modeled as wavelength division multiplexing (WDM) capable free-space optic (FSO) channels. Using the standard TG terminology, we formulate the optimal mixed integer linear problem considering the virtual topology, flow conversation, connection topology, non-bifurcation, and capacity constraints. Thereafter, we develop a fast sub-optimal solution where mice flows (MFs) are groomed and forwarded on predetermined rack-to-rack (R2R) lightpaths. On the other hand, elephant flows (EFs) are forwarded over dedicated server-to-server (S2S) express lightpaths whose routes and capacity are dynamically determined based on wavelength and capacity availability. Emulation results show that proposed models and algorithms provide a significant throughput improvement upon traditional DCNs for both MFs and EFs.**

I. Introduction

Data centers (DCs) are an intrinsic part of the computing infrastructures for emerging technologies such as Big Data applications, social media, Internet-of-Things, bioinformatics, etc. Scalability of DC networks (DCNs) are expected to accommodate a large number of servers and supply adequate speeds and bandwidths. DCNs are typically constructed based upon a hierarchical topology where servers are arranged in racks as shown in Fig. 1 where intra-rack communication is realized by edge switches (ESs), a.k.a. top-of-rack (ToR) switches. On the other hand, intra-rack communications are fulfilled by connecting ToR switches over higher layer switches.

In today's DCNs, network equipments communicate over coaxial or fiber-optic wires, which induces *cabling complexity* and lacks *bandwidth flexibility and efficiency*. Fortunately, wired DCNs can be augmented with wireless technologies such as multi-gigabit mmWave [1] or multi-terabit free-space optic (FSO) [2]. While mmWave technology does not require line-of-sight (LoS) links and offers some degree of penetration, it suffers from interference and short ranges due to high attenuation. On the other hand, FSO necessiates the existence of LoS links, which naturally improves the physical layer security and yields an interference-free communication [3]. It is worth noting that indoor FSO links are slightly affected from outdoor FSO channel impairments such as pointing error and atmospheric turbulence. Therefore, indoor FSO links can offer high bandwidths, which can further be enhanced with wavelength division multiplexing (WDM) methods [4], [5].

Regardless of the potentially achievable multi-terabit link capacities by combining WDM and FSO, the DCN bottleneck is still determined by data processing capability of power hungry state-of-art switches which can handle 10-40 Gbps rate at each port. Alternatively, power consumption and processing limitations can be mitigated by optical or hybrid optoelectonic switches. As the bandwidth request of a flow can be much lower than the capacity offered by WDM channels, *traffic grooming* (TG) arises as a necessary operation which refers to the aggregation of subwavelength flows onto high speed lightpaths subject to equipment costs and capacity [6]. In the past decades, TG was extensively studied for synchronous optical networks for a variety of topologies (please refer to [6] and references therein). Hamza et. al. compare virtues and drawbacks of potential wireless technologies for DCNs under the classification of 60 GHz and FSO communications. Thereafter, they survey the recent efforts on wireless data center networks. However, to the best of our knowledge, its potentials and prospects are not studied in the realm of DCNs except in [7] where traffic is simply groomed into three classes of wavelengths which are confined to broadcasting within racks and higher layer switches.

Accordingly, this paper develops a top-down TG approach for design and provisioning of optical wireless DCNs. While switches are modeled as hybrid cross-connects which consists of digital and electronic switches, fiber cables of traditional DCNs are replaced with FSO link each consists of multiple wavelengths thanks to the WDM. By means of standard TG terminology, we formulate an optimal mixed integer programming problem which consists of three NP-Hard subproblems: 1)Virtual topology design, i.e., lightpath provisioning and routing over the physical topology; 2) Wavelength assignment to the lightpaths; and 3) Developing a grooming policy and routing the groomed traffic requests on the virtual topology. Since finding an optimal solution for a single traffic instance of a small size DCN has a very high time complexity, we propose a fast yet high performance sub-optimal solution where all R2R MF traffic are groomed and forwarded over pre-determined lightpaths. On the other hand, elephant flows (EFs) are forwarded over dedicated S2S express lightpaths whose routes and capacity are dynamically determined based on wavelength and capacity availability.

The remainder of the paper is organized as follows: Section II presents the network model. Section III formulate the optimal problem and Section IV develops the proposed sub-optimal solution. Emulation results are presented in Section V and Section VI concludes the paper with a few remarks.

II. Network Model

A. Network Topology

In order to enable the use of FSO in DCNs, we first propose a physical topology to ensure LoS communication within DCNs. We consider a two-tier Clos architecture where every lower-tier switch (leaf layer) is connected to each of the top-tier switches (spine layer) in a full-mesh topology. The leaf layer consists of N ESs that connect to servers within a rack. The core switches (CSs) in the spine layer are responsible for interconnecting all leaf switches such that every ES connects to every CS. LoS links of DCN can be implemented using the physical topology shown in Fig. 1 where optical transceivers of ESs are connected to CS transceivers located at the top.

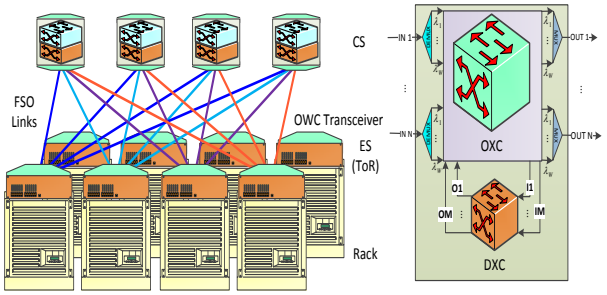

Fig. 1: Proposed topology for $N = 4$ and spine/leaf ratio of 1/2.

B. Hybrid Crossconnect Architecture

As shown in Fig. 1, switches are designed as a hybrid crossconnect (HXC) and have N input and output ports connected to receivers and transmitters, respectively. The signals at each of the N input ports are first demultiplexed into W individual wavelengths and then processed by optical cross-connect (OXC) or digital crossconnect (DXC) units. OXC is responsible for wavelength switching and routing operations. OXCs may also execute some grooming functions using optical couplers and decouplers. On the other hand, DXC provides flexibility and bandwidth efficiency by grooming several low-speed flows into a high-capacity lightpath which is simply a wavelength circuit/path between a node pair. It operates on O-E-O conversion principle and can handle M incoming signals at a given time, which determines grooming ratio of an HXC, i.e., $\frac{M}{N}$, $M \leqslant N$. We assume that DXC receivers and transmitters are tunable to any wavelength. Please also note that DXCs are also limited by processing speed limitation which is generally defined in Gbps.

C. Channel Model

Consider WDM-FSO links formed by directed laser-diode transmitters and photo-diodes receivers which employ intensity-modulation and direct-detection (IM-DD). Thanks to the WDM, FSO links can be treated as parallel channels where received signal at HXC_m from switch HXC_n on wavelength ω can be given as

$$r^{\omega}_{m,n} = h^{n}_{m} s^{\omega}_{m,n} + z^{n}_{m}, \quad 1 \leqslant \omega \leqslant W \tag{1}$$

where $s^{\omega}_{m,n} \in \mathbb{R}_+$ is the transmit intensity, $h^{n}_{m} \in \mathbb{R}_+$ represents the optical channel gain, z^{n}_{m} is the Gaussian noise with zero mean and unit variance, and $r^{\omega}_{m,n}$ is the received signal. Due to hardware and safety concerns, transmit signal intensity has to satisfy an individual and total average intensity constraint given by $\mathbb{E}[s^{\omega}_{m,n}] = \text{E}^{\omega}_{m,n} \leqslant \text{E}$ and $\sum_{\omega} \text{E}^{\omega}_{m,n} \leqslant \text{E}_T$, respectively. As optical channel variations are very slow in comparison to the symbol duration, h^{n}_{m} is further assumed to be constant throughout a transmission block and modeled as follows [8]

$$h^{n}_{m} = \rho h^{l}_{m,n} h^{a}_{m,n} h^{p}_{m,n} \tag{2}$$

where ρ is the detector responsivity, $h^{l}_{m,n}$ is the path loss, $h^{a}_{m,n}$ is the atmospheric turbulence, and $h^{p}_{m,n}$ is related to the pointing error. Due to the weak turbulence conditions, we assume log-normal atmospheric turbulence where $h^{a}_{m,n}$ is distributed according to [9]

$$f_a(h^{a}_{m,n}) = \frac{1}{h^{a}_{m,n}\sqrt{2\pi\sigma^2_R}} \exp\left\{-\frac{\left(\log h^{a}_{m,n} + \sigma^2_R/2\right)^2}{2\sigma^2_R}\right\}$$

where σ^2_R is the Rytov variance. The pointing error is accounted by $h^{p}_{m,n} = A_0 \exp\left\{-\frac{2d^2_{mn}}{\omega^2_{zeq}}\right\}$ where d_{mn} is the distance from the beam center to the receiver aperture, $A_0 = (\text{erf}(\nu))^2$, $\omega^2_{zeq} = \omega^2_z \frac{\sqrt{\pi}\,\text{erf}(\nu)}{2\nu \exp(-\nu^2)}$, $\nu = \frac{\sqrt{\pi}a}{\sqrt{2}\omega_z}$, a is the receiver aperture radius, and ω_z is the beam radius at the receiver. We assume that d^{n}_{m} follows Rayleigh distribution with the jitter variance σ^2_s at the receiver. Optical intensity channel capacities are studied in [10], where upper and lower bounds are shown to converge at high-SNRs as follows

$$C^{\omega}_{mn} = B\frac{1}{2}\log\left(1 + \frac{e(h^{n}_{m})^2 (E^{\omega}_{m,n})^2}{2\pi}\right) \tag{3}$$

where B is defined as the bandwidth of a wavelength.

III. TG Problem Statement and Formulation

TG is usually split into three joint subproblems: 1) designing the virtual topology, i.e., lightpath provisioning and routing over the physical topology; 2) assigning wavelength to the lightpaths; and 3) developing a grooming policy and routing the groomed traffic requests on the virtual topology. Noting that each of these sub-problems are NP-hard [11], TG is also an NP-hard problem which falls into mixed-integer linear programming (MILP) class. Using standardized TG formulations [11], [12], we formulate this MILP problem based on parameters/variables in Table I and following assumptions and constraints:

AS_1 : HXCs are not capable of wavelength conversion. Thus, a lightpath should be routed on the same wavelength through to its destination.

AS_2 : Bifurcation of flows are not allowed. That is, a connection requests cannot be divided and routed separately.

AS_3 : DXCs can groom as many flows into a lightpath as needed, as long as DXC processing capability and wavelength capacity is not exceeded.

AS_4 : I/O ports of DXCs are tunable to any wavelength.

TABLE I: Notations, parameters, and variables

Notations and given parameters:

M_i	Number of DXC I/O ports of node i.
$\bar{S}$	Total number of servers.
(m,n)	Originating and terminating points of a physical (i.e., FSO) link.
(i,j)	Originating and terminating points of a lightpath which may traverse multiple FSO links.
t	A flow among a total number of T traffic requests from all sources. Each t corresponds f^{th} flow for a source and destination server pair, i.e., $t \triangleq (s,d,f)$. Note that t may traverse through single or multiple lightpaths, e.g., $t: s \rightarrow (i,j) \rightarrow d$ or $t: s \rightarrow (i,j) \rightarrow (k,l) \rightarrow d$.
F_{mn}	Number of FSO links from m to n, $F_{mn} \in \{0,1\}$.
W	Number of wavelengths on (m,n) iff $F_{m,n}=1$.
C_{mn}^{ω}	Capacity of (m,n) on wavelength $\omega \in [1,W]$.
$\mathcal{D}$	Bandwidth demand matrix, $\mathcal{D} \in \mathbb{R}_{+}^{\bar{S}\times\bar{S}}$, where each entry is a vector of flow requests, i.e., $\mathcal{D}[s,d] = \{D_{sd}^{f} \vert 1 \leqslant f \leqslant T_s^d\}$ and T_s^d is the total number of flows from s to d, $\sum_s T_s^d = T$. Note that $\mathcal{D}$ is necessarily not a symmetric matrix.
DXC_i	Digital processing capacity of i.

Optimization Variables:

$P_{mn}^{ij,\omega}$	A binary variable for routing on physical topology; equals 1 iff a lightpath between node pair (i,j) is routed on FSO link (m,n) on wavelength ω.
L_{ij}^{ω}	Number of lightpath from i to j on wavelength ω.
L_{ij}	Total number of lightpath from i to j, $L_{ij} = \sum_{\omega} L_{ij}^{\omega}$. Note that L_{ij} and L_{ji} are different variables.
$R_{mn}^{ij,\omega\ell}$	A binary variable to define different physical routes of lightpaths between the same pair of nodes and on the same wavelengths; equals 1 iff a ℓ^{th}, $1 \leqslant \ell \leqslant L_{ij}^{\omega}$ lightpath between node pair (i,j) is routed on FSO link (m,n) on wavelength ω
$X_{ij}^{t,\omega\ell}$	A binary variable, $X_{ij}^{t,\omega\ell} = 1$ iff $t = (s,d,f)$ employs ℓ^{th} lightpath from i to j on wavelength ω as an intermediate virtual link.
Y_{ij}^{t}	Real valued capacity exploited by t on lightpath(s) between from i to j, $Y_{ij}^{t} \geqslant 0$.
$G_{ij}^{tt',\omega\ell}$	A binary indicator for TG; is 1 iff t and t' are groomed into lightpath ℓ on wavelength ω from i to j.

♦ *Lightpath routing (flow conservation) constraints:*

$$\sum_{m} P_{mi}^{ij,\omega} F_{mi} = \sum_{n} P_{jn}^{ij,\omega} F_{jn} = 0, \forall i,j,\omega \tag{4}$$

$$\sum_{m} P_{mj}^{ij,\omega} F_{mj} = \sum_{n} P_{in}^{ij,\omega} F_{in} = L_{ij}^{\omega}, \forall i,j,\omega \tag{5}$$

$$\sum_{m} P_{mk}^{ij,\omega} F_{mk} - \sum_{n} P_{kn}^{ij,\omega} F_{kn} = 0, \forall i,j,k, k \notin \{i,j\} \tag{6}$$

$$\sum_{i,j} P_{mn}^{ij,\omega} \leqslant F_{mn}, \forall m,n,\omega \tag{7}$$

$$\sum_{i,j} R_{mn}^{ij,\omega\ell} \leqslant P_{mn}^{ij,\omega}, \forall m,n,\omega,\ell \tag{8}$$

$$\sum_{m,n,i,j,\omega} P_{mn}^{ij,\omega} \sum_{\ell=1}^{L_{ij}^{\omega}} R_{mn}^{ij,\omega\ell} = 1. \tag{9}$$

where (4) assures that there are no incoming (outgoing) flows for originating (terminating) node i (j) of lightpath (i,j) on wavelength ω. Supported by the underlying physical topology, the total number of lightpaths on wavelength ω from i to j is given in (5). Constraints in (6) and (7) ensure the wavelength continuity and protection against lightpath collisions, respectively. Finally, (8) and (9) permits at most one lightpath to be routed on (m,n) among all lightpaths.

♦ *Connection topology constraints:*

$$X_{ij}^{t} = \sum_{\omega} \sum_{\ell=1}^{L_{ij}^{\omega}} X_{ij}^{t,\omega\ell}, \forall i,j,t \tag{10}$$

$$\sum_{i} X_{is}^{t} = \sum_{j} X_{dj}^{t} = 0, \forall t \tag{11}$$

$$\sum_{i,i\neq k} X_{ik}^{t} = \sum_{j,j\neq k} X_{kj}^{t}, \forall t,k, k \notin \{s,d\} \tag{12}$$

where (10) defines the total number of connections established on all wavelengths and lightpaths from i to j. That is some portion of traffic t can be split to some different lightpath and wavelength combinations. Constraint in (11) guarantees that there is no incoming and outgoing traffic to the source and destination, respectively. On the other hand, (12) preserves the continuity of the flows on single or multiple lightpaths. Even though different flows between a source destination pair are allowed to be split to different lightpaths, wavelengths, or routes, non-bifurcation keeps a certain flow intact and exploit only one lightpath, wavelength, and physical route tuple.

♦ *Virtual topology constraints:*

$$\sum_{\omega} \sum_{j,j\neq i} L_{ij}^{\omega} \leqslant M_i, \forall i \tag{13}$$

$$\sum_{\omega} \sum_{j,j\neq i} L_{ji}^{\omega} \leqslant M_i, \forall i \tag{14}$$

$$\sum_{\omega} L_{ij}^{\omega} = L_{ij}, \forall i,j \tag{15}$$

where (13) and (14) ensure that number of originating and terminating lightpaths at HXC_i do not exceed the number of DXC I/O ports, respectively.

♦ *Non-bifurcation and capacity constraints:*

$$X_{ij}^{t} \leqslant 1, \forall i,j,t \tag{16}$$

$$\sum_{i,j} X_{ij}^{t} \leqslant 1, \forall t \tag{17}$$

$$G_{ij}^{tt} = X_{ij}^{t}, \forall i,j,t \tag{18}$$

$$G_{ij}^{tt'} = \sum_{\omega} \sum_{\ell=1}^{L_{ij}^{\omega}} G_{ij}^{tt',\omega\ell}, \forall i,j,t \neq t' \tag{19}$$

$$G_{ij}^{tt'} \leqslant {}^{1}\!/_{2} \left(X_{ij}^{t} + X_{ij}^{t'} \right), \forall i,j,t \neq t' \tag{20}$$

$$L_{ij} = G_{ij}^{tt} + \sum_{t',t'\neq t} \left(G_{ij}^{t't'} - \bigvee_{x=1}^{t'-1} G_{ij}^{t'x} \right) \tag{21}$$

$$C_{mn}^{\omega} \geqslant \sum_{i,j,t} Y_{ij}^{t} X_{ij}^{t,\omega\ell} R_{mn}^{ij,\omega\ell}, \forall m,n \tag{22}$$

$$\text{DXC}_i \geqslant \sum_{t} \left(\sum_{j,j\neq i} X_{ij}^{t} Y_{ij}^{t} + \sum_{j,j\neq i} X_{ji}^{t} Y_{ji}^{t} \right), \forall i \tag{23}$$

where (16)-(21) satisfy non-bifurcation of traffic among lightpaths between different nodes, among wavelengths of a lightpath between the same pair of nodes, and among different

physical routes a lightpath between the same pair of nodes and on the same wavelength. Capacity constraint in (22) insures that total traffic request of set of flows, which are groomed on the same physical route on a certain lightpath and wavelength pair, must comply with the capacity of that route, i.e. the lowest capacity along the physical route. Finally, total amount of incoming and outgoing traffic to be processed is limited by processing capability of nodes as in (22).

IV. TG Policy Design for DCNs

In this section, we develop a suboptimal TG policy based on EF and MF classifications based on following rule sets

1) We assume that flows can be classified in a timely and efficient manner. While MFs are groomed into larger traffic, EFs are treated separately.
2) Lightpaths are first provisioned for groomed MFs. The residual wavelengths and links are then used to provision EF lightpaths.
3) Bifurcation at HXC's are not allowed as splitting and combining EFs can consume a significant portion of DXC processing capability at origin and terminal points of lightpaths, respectively. On the other hand, splitting a MF may not be necessarily efficient.
4) Aside from W wavelengths, there exists a broadcasting wavelength for control signaling. Current energy and wavelength availability state of DCN is formulated in graphs $\mathcal{G}_e\left(\mathcal{V},\mathcal{E}\right)$ and $\mathcal{G}_\omega\left(\mathcal{V},\mathcal{W}\right)$, respectively, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ presents available light intensity, and $\mathcal{W}$ presents edge weights for available wavelengths. This graphs are always kept updated over the control wavelength.

A. TG and Lightpath provisioning for MFs

MF grooming is designed to take place at source and switches in three steps as follows:

1) *S2S Step:* Servers groom all flow arrival destined to a certain destination server.
2) *Server-to-Rack (S2R) Step:* Servers further groom S2S flows according to destination rack and transfer S2S flows to ESs.
3) *R2R Step:* Received S2S flows are then groomed according to their destination rack to obtain R2R flows.

Thanks to WDM, a large set of route and wavelength possibility is already available. Hence, each R2R flow is provided by a dedicated always active lightpath defined on a certain route-wavelength pair. We derive the minimum number of wavelengths to set dedicated R2R lightpaths as $W \geqslant \lceil \frac{N/\chi-1}{N} \rceil$ where χ is the spine/leaf ratio. Proposed approach has low complexity, incurs less delay, and simplify the CS hardware design as they do not need DXCs. Since R2R flow size is limited by DXC port speed, capacity of MF wavelengths must also be upperbounded by DXC speed, which naturally open some room for extra intensity required by some EFs as explained next.

B. Intensity Allocation

WDM based FSO links provide two great advantages: channel diversity and capacity flexibility. While assigning wavelengths with the fixed intensities/capacities ignores the flexibility provided by optical wireless technology (which is referred to as ECMP-FSO in Section V), allocating unnecessarily high intensities to a certain wavelength destroys the wavelength availability for future flows. Therefore, a fast yet efficient intensity allocation methods are necessary to maximize the benefit. Each traffic request is related to a certain size κ_t and maximum service duration τ_t. Thus, if a traffic is routed over an FSO link between nodes m and n, required transmission intensity on can be calculated from (3) as follows

$$E^t_{mn} \geqslant \sqrt{\frac{2\pi\left(2^{\frac{2\kappa_t}{B\tau_t}}-1\right)}{e(h^n_m)^2}} \tag{24}$$

We first allocate required intensity for groomed MFs based on long term traffic arrival statistics. Residual intensity is shared among EFs based on a fair-share policy as follows

$$E^{\omega t}_{mn} = \min\left\{\sqrt{\frac{2\pi\left(2^{\frac{2\kappa_t}{B\tau_t}}-1\right)}{e(h^n_m)^2}}, \frac{E_T}{W} + E^a_{mn} - W^a_{mn}\frac{E_T}{W}\right\} \tag{25}$$

where E^a_{mn} is the available intensity on (m,n) after MFs allocations, $W^a_{mn} = |\mathcal{W}^a_{mn}|$ is available number of wavelengths for EFs on (m,n), and $\mathcal{W}^a_{mn}$ is the set of these wavelengths. In (25), a power demand exceeding equalized power policy can be obtained from the room opened by less demanding flows.

C. Lightpath provisioning for EFs

As aforementioned, EFs are treated separately from MFs and are not subject to traffic grooming as they already have significant traffic requests. Once an EF is detected, a S2S lightpath is required to be established between source and destination, which is terminated after the session completion. That is, EFs are sent express through OXCs on a certain wavelength and route pair. Based on $\mathcal{G}_e$ and $\mathcal{G}_\omega$, each server maintains shortest-paths lists toward all destinations, available total light intensity on each hop of shortest-paths, and indices of available wavelengths.

Taking the wavelength continuity into consideration, a route is feasible only if there exists a wavelength which is available at each hop. Denoting the set of such routes from server s to server d as $\mathcal{R}^d_s$, achievable capacity of the routes are determined as follows

$$C^r_t = \min_{(m,n)\in r} C^\omega_{mn}(E^{\omega t}_{mn}), \quad r \in \mathcal{R}^d_s, (s,d) \in t \tag{26}$$

Based on calculated route capacities and available number of wavelengths, route is determined based on a *best-fit* policy

$$r^\star_t = \text{argmax}_{r\in\mathcal{R}^d_s}\left(W^r_t C^r_t\right), \quad (s,d) \in t \tag{27}$$

where $W^r_t = |\bigcap_{(m,n)\in r}\mathcal{W}^a_{mn}|, r \in \mathcal{R}^d_s, (s,d) \in t$. The next step is assigning a wavelength to the selected route, which

can be done by a variety of methods, e.g., random, first-fit, least/most used, etc. Since they are shown to perform very close to each other [13], we employ the first-fit scheme and assign the lowest index of available wavelengths. Please note that route selection in (27) is desirable due to its low computational complexity and suitability for low EF loads.

If there exists many EF requests, however, servers may compete for potential routes. When a route is selected by multiple flows by (27), priority of a competing EF for such a route is defined as

$$\psi_t^r = \frac{[W_t^r C_t^r]^\alpha}{[\mathcal{K}_s]^\beta}, s \in t \tag{28}$$

where $\mathcal{K}_s$ is the average of service provided for the source of t within a certain time frame. Accordingly, the route is assigned to the flow with the highest priority as follows $t_r^\star = \text{argmax}_t(\psi_t^r)$ which reduces to round-robin, greedy, and proportional fair scheduling for $(\alpha = 0, \beta = 1)$, $(\alpha = 1, \beta = 0)$, $(\alpha = 1, \beta = 1)$, respectively. If a flow is rejected on a certain route, it competes for the second best-fit route calculated in (27), and so forth.

V. Numerical Results

We now present our evaluation of the proposed solution. We conducted our evaluation by using Mininet emulator [14] which uses real virtual hosts, POX-eel controllers [15], and real software switches, i.e., OVS switches [16]. The system characteristics of the used machine are Ubuntu 14.04 LTS installed on 16 x (2.5GHz-intel Xeon CPU E5-2680v3), and the memory is 128 GiB. Iperf is used to generate mice and EFs of sizes 100KB and 1GB, respectively. Emulated DCN topology has six spine switches and twelve leaf each with has 25 hosts, i.e., 300 hosts in total. Since the links in Mininet are limited by the processing capacity of the host machine, we configured the FSO-links with 10 Gbps and the non-FSO links with 1 Gbps which means FSO links are 10× faster than normal DCN links. Each FSO link consists of 4 wavelengths, which are realized as virtual links in Mininet. Since the emulator is limited in DCN size, optical channel gains are not distinguishably different, thus, assumed to be identical.

A. Workloads

We use MapReduce to mimic workloads of real DCNs, which shuffles phase communication pattern where k servers from every rack communicate with k servers in another rack. For instance, the hosts in R_i is divided into five sets and every set has k servers, let's say four mice and one elephant. Each k is communicating with k servers of rack j, $j \neq i$, different than other ks of the same rack.

B. Routing Algorithms

Equal-Cost-Multi-Path routing (ECMP) [17] is a widely used DCN routing method which uses the packet header information, such as the IP/MAC addresses and TCP port numbers, as a key for a hash function. The outgoing path is the output hash value modulo the number of outgoing paths. This strategy splits the flows among available paths. Since the header information for an individual flow is the same during the session, the packets of the same flow are always forwarded via the same path; maintaining the packet orders. We used OVS `bundle` command with `symmetric_4` hash function in ECMP algorithm.

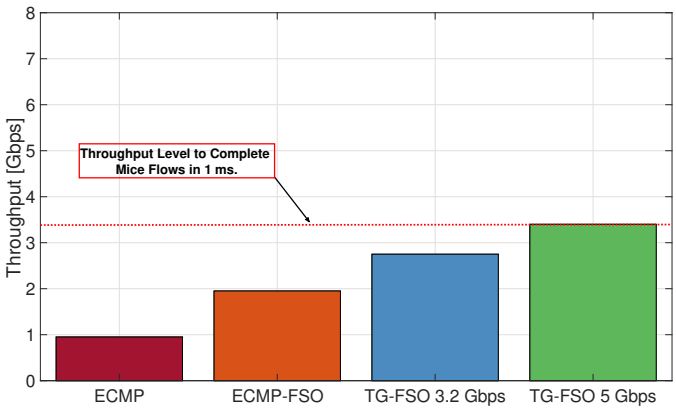

Fig. 2: Comparison of MF throughputs for different algorithms.

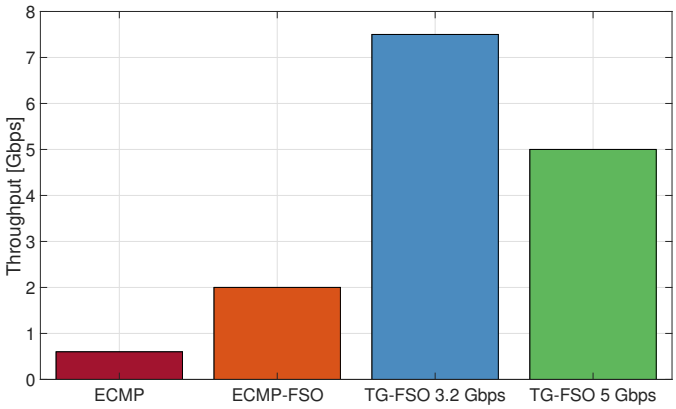

Fig. 3: Comparison of EF throughputs for different algorithms.

ECMP-FSO: is an ECMP routing algorithm supported by the FSO technology. In FSO, the speeds of the links are in order of magnitude faster than non-FSO links. In this evaluation, we use a factor of ten which means the FSO link is 10× faster than normal data center link. In this routing method, the link capacity is equally divided between the wavelength which means the capacity of every wavelength is fixed to 2.5 Gbps. Each flow was assigned to a single wavelength. However, when flows are more than the available number of lightpaths, the packets of the waiting flows are enqueued until a lightpath is available for transmission.

TG-FSO: is the proposed algorithm where the intensity is first allocated to the R2R mice-flow-wavelengths to meet MF demands. The remaining intensity is used by EFs as explained in Section IV-B.

C. Network Throughput Results

Since $k = 4$ is set to four, we have 4×100 KB MFs in every R2R communication, hence, the needed capacity by R2R groomed MFs is 3.2 Gbps with a $\tau = 1$ ms service duration. Due to the TCP behavior; a window of packets per RTT, and the waiting time in the link buffer, some of the transmitted flows finished after τ. In average 86% of the flows complete before τ. However, when the wavelength capacity increased to 5 Gbps, 100% of the flows satisfied the time constraint. The throughput results of MFs are displayed in Fig. 2. On the other hand, the throughput results of EFs are displayed in Fig. 3 where the service time demand of EFs, $\tau = 1$ s, are satisfied before τ in all link configurations.

At all configurations, proposed TG-FSO algorithm outperforms the ECMP and ECMP-FSO cases. The achievable throughput of the proposed algorithm (3.2 Gbps) is about 3.4 times the ECMP algorithm and about 1.65 times the ECMP-FSO. We should emphasize that the sustained capacity for MFs are 10 Gbps and 3.2/5 Gbps for ECMP-FSO and TG-FSO, respectively. Therefore, the transmission power consumption of 3.2/5 Gbps TG-FSO are lower than ECMP-FSO by about 3.2/2 times, respectively. For the EFs, on the other hand, the highest achievable average throughput was 7.47 Gbps which is about 13 and 3.6 times the ECMP and ECMP-FSO throughputs, respectively.

Even though we had to set FSO link capacity to 10 Gbps due to the Mininet restrictions, Ciaramella et. al. recorded 40 Gbps wavelength capacity for a 32 wavelength outdoor WDM-FSO system over several hundreds meters [5]. Accordingly, the potential of the proposed design can be understood better when the 13 times performance enhancement is scaled up to higher number of wavelengths and capacities.

VI. Conclusions

In this paper, we addressed the design and provisioning of wireless DCNs from a TG perspective. In order to mitigate the system limitations of traditional wired DCNs, we considered a wireless approach by using hybrid opto-electronic switches and WDM capable FSO links. Contingent upon the optimal problem formulation, we developed a fast yet high performance sub-optimal solution which significantly improved the throughput of both MFs and EFs.

References

[1] D. Halperin *et al.*, "Augmenting data center networks with multi-gigabit wireless links," in *Proc. of the ACM SIGCOMM*, 2011, pp. 38–49.

[2] N. Hamedazimi, Z. Qazi, H. Gupta, V. Sekar, S. R. Das, J. P. Longtin, H. Shah, and A. Tanwer, "Firefly: A reconfigurable wireless data center fabric using free-space optics," in *Proc. the ACM SIGCOMM*, 2014, pp. 319–330.

[3] A. S. Hamza, J. S. Deogun, and D. R. Alexander, "Wireless communication in data centers: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1572–1595, thirdquarter 2016.

[4] A. O. Aladeloba, M. S. Woolfson, and A. J. Phillips, "Wdm fso network with turbulence-accentuated interchannel crosstalk," *J. Opt. Commun. Netw.*, vol. 5, no. 6, pp. 641–651, Jun. 2013.

[5] E. Ciaramella, Y. Arimoto, G. Contestabile, M. Presi, A. D'Errico, V. Guarino, and M. Matsumoto, "1.28 terabit/s (32x40 gbit/s) wdm transmission system for free space optical communications," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1639–1645, Dec. 2009.

[6] S. Huang and R. Dutta, "Dynamic traffic grooming: the changing role of traffic grooming," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 1, pp. 32–50, First 2007.

[7] G. C. Sankaran and K. M. Sivalingam, "Optical traffic grooming-based data center networks: Node architecture and comparison," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1618–1630, May 2016.

[8] A. Chaaban, Z. Rezki, and M. S. Alouini, "Fundamental limits of parallel optical wireless channels: Capacity results and outage formulation," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 296–311, Jan 2017.

[9] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2231–2258, Fourthquarter 2014.

[10] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct 2009.

[11] K. Zhu and B. Mukherjee, "Traffic grooming in an optical wdm mesh network," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 1, pp. 122–133, Jan 2002.

[12] R. Ul-Mustafa and A. E. Kamal, "Design and provisioning of wdm networks with multicast traffic grooming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 4, p. 53, 2006.

[13] H. Zang, J. P. Jue, B. Mukherjee *et al.*, "A review of routing and wavelength assignment approaches for wavelength-routed optical wdm networks," *Optical networks magazine*, vol. 1, no. 1, pp. 47–60, 2000.

[14] N. Handigol, B. Heller, V. Jeyakumar, B. Lantz, and N. McKeown, "Reproducible network experiments using container-based emulation," in *Proc. the 8th Intl. Conf. Emerg. Netw. Exper. Tech.* ACM, 2012, pp. 253–264.

[15] POX. [Online]. Available: http://www.noxrepo.org/pox/about-pox/.

[16] B. Pfaff, J. Pettit, K. Amidon, M. Casado, T. Koponen, and S. Shenker, "Extending networking into the virtualization layer." in *Hotnets*, 2009.

[17] C. Hopps, "Analysis of an equal-cost multi-path algorithm," *RFC 2992, Internet Engineering Task Force*, 2000.