On the Synergy between Flexible Ethernet and Point-to-multipoint Optical Networks

Meihan Wu[1], Xiaoliang Chen[1], Francesco Musumeci[2], Ruoxing Li[1], Yuxiao Zhang[1], Qian Lv[1], Zuqing Zhu[1]

1. University of Science and Technology of China, Hefei, Anhui 230027, China, Email: zqzhu@ieee.org
2. Politecnico di Milano, Milan, Italy, Email: francesco.musumeci@polimi.it

Abstract: We present the first comparative study on the synergistic benefits between flexible Ethernet and point-to-multipoint optical networks. Simulations confirm that by combining the two techniques adaptively, service provisioning can be achieved with significantly higher cost-effectiveness.

OCIS codes: (060.1155) All-optical networks; (060.4251) Networks, assignment and routing algorithms.

1. Introduction

Flexible Ethernet (FlexE) resolves the mismatch between variable media access control (MAC) rates and the rigid capacity of Ethernet physical channels (PHYs), and thus enables agile aggregation of client streams. In particular, FlexE inserts a FlexE shim between the MAC and PHY layers, which divides PHY bandwidth into fine-grained transmission slots (TS') based on time-division multiplexing and maps client streams to the TS' adaptively [1]. Consequently, FlexE facilitates high link utilization and guarantees deterministic delay, making it promising for IP over optical transport networks (IPoOTNs) [2, 3]. Nevertheless, existing IPoOTNs employ point-to-point transceivers (P2P-TRXs), where the single-destination constraint can significantly restrict the flexibility of data aggregation by FlexE. Recently, with the prevalence of hub-and-spoke (H&S) traffic (produced by cloud applications, deep learning workloads, *etc.*), coherent point-to-multipoint transceivers (P2MP-TRXs) have emerged as critical building blocks for future IPoOTNs [4]. P2MP-TRXs utilize digital subcarrier multiplexing (DSCM) to transmit high-speed traffic to multiple destinations with low-rate Nyquist subcarriers (SCs), while offering comparable complexity and unit cost as those of P2P-TRXs [5]. Therefore, the combination of P2MP-TRX and FlexE can remove the single-destination constraint of P2P-TRXs and thereby assists in fully exploiting the fine-grained scheduling by FlexE for optimized service provisioning in IPoOTNs. However, to the best of our knowledge, such a combination has not been considered in the literature yet.

This paper studies the synergistic benefits achieved by combining FlexE and P2MP-TRXs (FlexE-P2MP). We first detail the principle of FlexE-P2MP and propose a dynamic programming-based cross-layer (DP-CL) algorithm for the service provisioning in FlexE-P2MP. We then consider two benchmark architectures, *i.e.*, FlexE with P2P-TRXs (FlexE-P2P) and link aggregation groups (LAGs) with P2MP-TRXs (LAG-P2MP), to highlight the benefits of FlexE-P2MP. Under various traffic scenarios, qualitative and quantitative comparisons reveal the superiority of FlexE-P2MP.

2. Operation Principle

We denote an IPoOTN by $G(\mathbf{V},\mathbf{E})$, where $\mathbf{V}$ and $\mathbf{E}$ are the node and link sets, respectively. Let $\mathbf{C}$ represent the collection of client streams, with each stream $c_i(s_i,d_i,x_i)$ from node s_i to node d_i demanding a data rate of x_i. Then, the service provisioning in such an IPoOTN can be stated as: accommodating all the streams in $\mathbf{C}$ by proper allocating

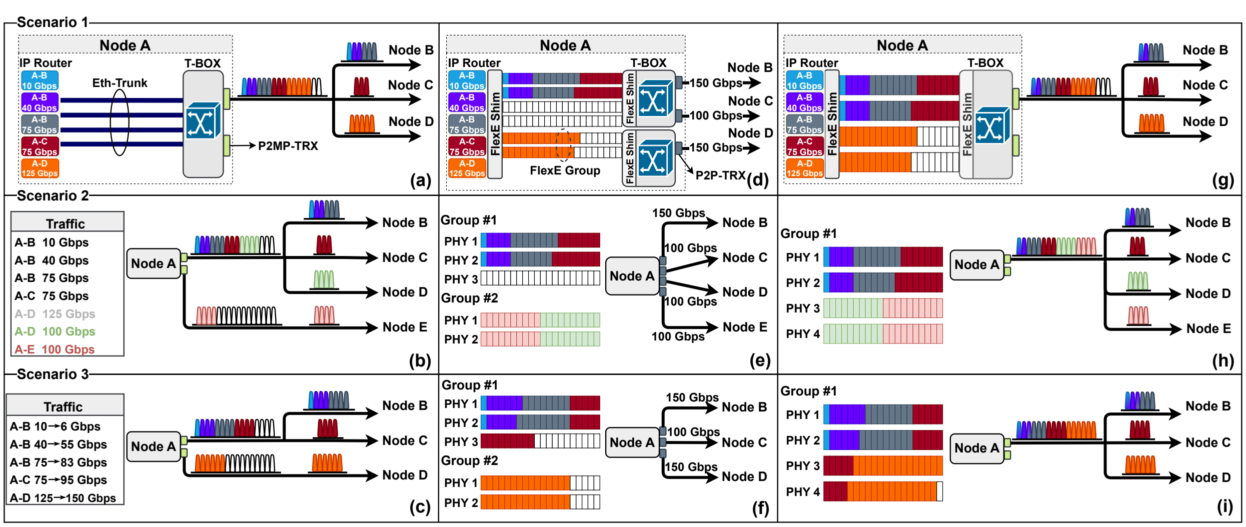

Fig. 1. Operation principles of architectures: (a)-(c) LAG-P2MP, (d)-(f) FlexE-P2P, and (g)-(i) FlexE-P2MP.

spectra and/or TS', while minimizing the overall resource usage (in number of transceivers, spectrum usage, *etc*.). In the following, we will explain the principle of FlexE-P2MP and underscore its synergistic benefits by comparing the service provisioning in it with those in two existing architectures under three traffic scenarios. In particular, *Scenario 1* involves five client streams, $c_1(A,B,10)$, $c_2(A,B,40)$, $c_3(A,B,75)$, $c_4(A,C,75)$, and $c_5(A,D,125)$ (data rate in Gbps), and both *Scenarios 2* and *3* exhibit traffic changes based on *Scenario 1*. *Scenario 2* removes c_5 and adds two new streams: $c_6(A,D,100)$ and $c_7(A,E,100)$, while *Scenario 3* decreases the rate of c_1 and increases the rates of others. For the sake of clarity, we ignore the routing schemes and their impact on the choice of modulation formats.

LAG-P2MP denotes the existing P2MP-TRX-based solution without FlexE. With LAG [1], LAG-P2MP can bundle a set of 100-Gbps PHYs into a single logical link to support higher-rate client streams. The aggregated traffic is then mapped to a P2MP-TRX using an OTN transport box (T-Box). We assume that the P2MP-TRX operates at 400 Gbps with 16 SCs, *i.e.*, each SC supports 25 Gbps. Figs. 1(a)-1(c) show the service provisioning with LAG-P2MP under the three traffic scenarios. Under *Scenario 1*, we respectively allocate 1, 2, 3, 3 and 5 SCs to serve the five streams. As c_1 and c_2 do not fully utilize the capacity of their SCs, the bandwidth efficiency is 84%. When c_5 is substituted by c_6 and c_7 in *Scenario 2*, we allocate *SCs* $[10,13]$ to c_6 and activate a new P2MP-TRX for c_7, because the residual capacity of the original P2MP-TRX is insufficient. The traffic change between *Scenarios 1* and *3* further degrades the bandwidth efficiency of the original P2MP-TRX to 68.8%, which squeezes out c_5 and calls for a second P2MP-TRX likewise. Therefore, LAG-P2MP can suffer from low bandwidth efficiency and thereby may consume more P2MP-TRXs.

FlexE-P2P denotes the existing FlexE-terminal architecture with P2P-TRXs [2], which incorporates a FlexE shim in each T-Box to identify and switch FlexE clients. This allows for time-sharing of spectrum resources and enables efficient aggregation of diverse client streams. The spectrum efficiency can be further improved by adopting bandwidth-variable transponders (BV-Ts) [6] to operate on customized spectrum ranges [2]. As shown in Fig. 1(d), for *Scenario 1*, we allocate 2, 8, 15, 15 and 25 TS' to the five streams, respectively, and transmit their data with three BV-Ts, each of which targets to a different destination. In Fig. 1(e) (*Scenario 2*), an additional BV-T is activated for c_7 since its destination is new (*Node E*), suggesting a cost increase. FlexE-P2P accommodates all the traffic changes in *Scenario 3* by only reallocating TS', without adding BV-Ts. This exemplifies the flexibility of FlexE in aggregating streams.

FlexE-P2MP denotes the synergy between FlexE and P2MP. It can overcome the drawbacks of LAG-P2MP and FlexE-P2P, as evidenced by Figs. 1(g)-1(i). For *Scenario 1*, FlexE-P2MP allocates *SCs* $[1,2]$ to c_1 and c_2, allowing them to fully utilize *SC* 1 by transmitting data of two and three TS' alternately. By contrast, LAG-P2MP reserves three SCs for c_1 and c_2. Meanwhile, the introduction of P2MP-TRX improves the utilization of PHYs in each FlexE group and obviates the need for activating new T-Boxes or TRXs. The synergistic benefits are further demonstrated by Fig. 1(h) (*Scenario 2*), where *PHYs* 3 and 4 of the originally P2MP-TRX accommodate the new streams c_6 and c_7 successfully. Whereas LAG-P2MP needs to activate an additional P2MP-TRX in this scenario. Similarly, in Fig. 1(i) (*Scenario 3*), by adjusting the allocation of TS', FlexE-P2MP still serves all the streams with one P2MP-TRX.

3. Service Provisioning Algorithm

To fully explore the benefits of FlexE-P2MP for service provisioning, we propose a DP-CL algorithm. Given client streams $\mathbf{C}$ and a set of viable P2MP-TRXs (at 25/100/400 Gbps), DP-CL first clusters the streams into subsets $\{\mathbf{C}_s|_{s\in\mathbf{V}}\}$ based on their sources ***(Step 1)***. Then, for each $s \in \mathbf{V}$, we apply a knapsack-based DP algorithm to pack $\mathbf{C}_s$ into FlexE groups ***(Step 2)***. Specifically, we treat each FlexE group as a knapsack with a capacity of W, while the streams in $\mathbf{C}_s$ are taken as items. Each item $c_i \in \mathbf{C}_s$ has a weight w_i and a value f_i, both set to be its rate x_i. Then, DP-CL maximizes the total rate (value) of the streams (items) assigned to each FlexE group (knapsack), through recursive value updates. Specifically, for all c_i and $w \in [1,W]$: if $w < w_i$, we set $\mathbf{F}[i][w] \leftarrow \mathbf{F}[i-1][w]$, and $\mathbf{F}[i][w] \leftarrow \max\{\mathbf{F}[i-1][w], \mathbf{F}[i-1][w-w_i]+f_i\}$, otherwise. Next, DP-CL performs P2MP-TRX/SC/TS allocation for each FlexE group. It checks the path distance to each destination d in the group to get the corresponding modulation format of SCs [5], gets the number of required SCs n_d accordingly, and records the related streams in $\tilde{\mathbf{C}}_d$ ***(Step 3)***. If the SCs required by $\tilde{\mathbf{C}}_d$ exceed the maximum capacity of a P2MP-TRX (16 SCs), we activate a highest-rate P2MP-TRX and reuse the aforementioned knapsack-based DP algorithm to pack streams in $\{\tilde{\mathbf{C}}_d\}$ into the P2MP-TRX ***(Step 4)***. This time, the P2MP-TRX and $\{\tilde{\mathbf{C}}_d\}$ are treated as the knapsack and items, respectively, with the value and weight of each item set as n_d. Then, $\{\tilde{\mathbf{C}}_d\}$ is updated to only contain the remaining streams. ***Step 4*** is repeated until the total number of SCs required by $\tilde{\mathbf{C}}_d$ is within the maximum capacity of a P2MP-TRX, and then an appropriate P2MP-TRX is activated for the streams in $\{\tilde{\mathbf{C}}_d\}$ ***(Step 5)***. Finally, we update the allocation of P2MP-TRXs in destination d to set up the receiving end ***(Step 6)***.

4. Simulation Results

We compared the performance of LAG-P2MP, FlexE-P2P and FlexE-P2MP with simulations conducted under the 24-node US Backbone topology [3]. We assumed that each FlexE group can bind four 100-Gbps PHYs with a T-Box,

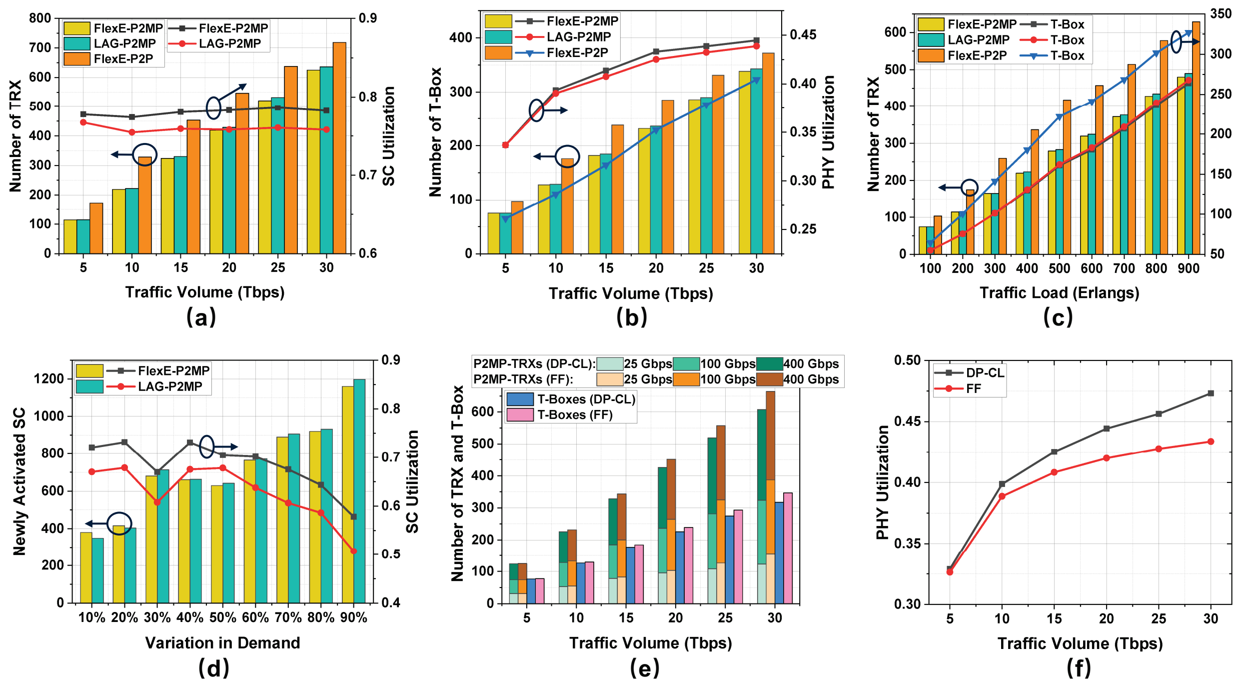

Fig. 2. Simulation results on architectures' performance to serve (a)-(b) different volumes of traffic, (c) dynamic Poisson traffic, (d) streams with time-varying demands, and (e)-(f) provisioning performance of DP-CL and FF in FlexE-P2MP.

which delivers a total capacity of 400 Gbps with two TRXs. A P2MP-TRX has a capacity of 25/100/400 Gbps, which operates on 1/4/16 SCs, respectively [4]. The routing paths and P2MP trees were computed using the shortest path routing. Based on the path distance, a TRX adopts either QP-16QAM ($\leq$ 500 km) or DP-QPSK as modulation format.

We first compare the service provisioning performance of the three architectures, and to ensure fair comparisons, we serve streams in all the architectures with the first-fit (FF) based TRX/SC/TS allocation and shortest-path routing (FF). Figs. 2(a) and 2(b) show the results on SC utilization, PHY utilization, and numbers of activated TRXs and T-Boxes with the three architectures under various traffic volumes. Wherein, each trace follows a uniform distribution, with the data rate of each stream randomly chosen from $\{10, 40, 25 \cdot n\}$ Gbps ($n \in [1, 8]$). FlexE-P2MP achieves the highest SC and PHY utilizations, and consequently, consumes the fewest TRXs and T-Boxes, which is in line with the observations drawn from Fig. 1. As expected, FlexE-P2P requires the most TRXs and T-Boxes. Note that, LAG can only utilize 70%-80% of link capacity as reported by [1], and thus the actual performance of practical LAG-P2MP could be even worse. Next, we assessed the three architectures under dynamic traffic scenarios where client streams arrive and terminate following the Poisson and exponential processes, respectively. The results on used TRXs and T-Boxes in Fig. 2(c) further confirm the advantage of FlexE-P2MP. We also studied the impact of traffic fluctuations on the performance of different architectures. Specifically, we generated a basic trace and introduced perturbations of 10% to 90% on it. Fig. 2(d) shows that FlexE-P2MP still consistently achieves higher SC utilization, demonstrating better adaptability to time-varying demands. Here, the sharp decline of SC utilization at 30% of perturbation is attributed to demands slightly exceeding a single SC's capacity, which leads to allocation of additional SCs for trivial traffic.

Finally, we evaluated the performance of our proposed DP-CL in FlexE-P2MP by comparing it with FF. Fig. 2(e) suggests that DP-CL reduces the numbers of used P2MP-TRXs and T-Boxes by 8% over FF, and the results in Fig. 2(f) confirm that DP-CL also improves PHY utilization effectively over FF, especially when the traffic volume increases.

5. Summary

We explored the synergistic benefits of FlexE and P2MP-TRX through comparative studies. Simulation results showed that FlexE-P2MP outperforms two existing benchmarks in terms of resource utilization and cost-effectiveness.

References

[1] OIF FlexE White Paper. [Online]. Available: https://www.oiforum.com/wp-content/uploads/2019/01/OIF_FlexE_White_Paper.pdf.
[2] W. Lu *et al.*, "How much can flexible Ethernet and elastic optical networking benefit mutually?" *ICC 2019*.
[3] M. Wu *et al.*, "Dynamic cross-Layer restoration to resolve packet layer outages in FlexE-over-EONs," *IEEE Trans. Netw. Serv. Manag.*, pp. 2600-2611, Sept. 2022.
[4] D. Welch *et al.*, "Point-to-multipoint optical networks using coherent digital subcarriers," *J. Lightw. Technol.*, vol. 39, pp. 5232-5247, Aug. 2021.
[5] R. Li *et al.*, "On the network planning of wavelength switched optical networks with P2MP transceivers," *J. Lightw. Technol.*, vol. 42, pp. 24-36, Jan. 2024.
[6] N. Sambo *et al.*, "Next generation sliceable bandwidth variable transponders," *IEEE Commun. Mag.*, vol. 53, pp. 163-171, Feb. 2015.