Iterative configuration in elastic optical networks [Invited]

Citation for the original published paper (version of record):
Hadi, M., Agrell, E. (2020). Iterative configuration in elastic optical networks [Invited]. 2020 24th International Conference on Optical Network Design and Modeling, ONDM 2020.
http://dx.doi.org/10.23919/ONDM48393.2020.9133016

(article starts on next page)

Iterative Configuration in Elastic Optical Networks

(Invited Paper)

Mohammad Hadi and Erik Agrell
Electrical Engineering Department, Chalmers University of Technology, Gothenburg, Sweden
mohadi@chalmers.se, agrell@chalmers.se

***Abstract*—A general dynamic resource allocation scheme is introduced, which iteratively reconfigures an elastic optical network according to traffic dynamism and service diversity. As an application example, the scheme is employed for impairment-aware power-efficient resource allocation in a short-haul metro elastic optical network to demonstrate a typical 5% power plus service penalty for compensating bandwidth limitations compared with the corresponding limitless counterpart.**

***Index Terms*—elastic optical networks, bandwidth limitations, resource allocation, stochastic iterative optimization**

I. Introduction

The promising advent of elastic optical network (EON) and software-defined networking (SDN) has paved the way for extending optical technologies to metro and even access networks with more diverse quality of service (QoS) and dynamic traffic paradigms [1], [2]. In fact, high volume, fast dynamism, and service diversity are the main properties of metro networks, which cannot be addressed by the rigid and slow resource allocation strategies used in conventional optical networks [3], [4]. To properly support these features and exploit the fast controllability of the SDN for iterative reconfiguration of the flexible network elements in an EON, a dynamic resource allocation scheme is required. Such a scheme is usually formulated as a large-scale optimization problem, whose optimal solution imposes a huge computational complexity [5].

Offline resource allocation algorithms usually waste the available resources by a long-term fixed resource provisioning beyond the actual needs of the requests [6]. On the other hand, online resource allocation methods provide a better resource utilization, but can lead to severe spectrum fragmentation and consequently, service blockage [6], [7]. In iterative configuration, we involve both the average and instantaneous network behavior in the resource allocation to better handle unexpected conditions and avoid unnecessary reconfigurations that fragment the spectrum [5].

Stochastic iterative optimization is a powerful tool for iterative reconfiguration in EONs, which allows to simultaneously target average network behavior, handle instantaneous unexpected conditions, avoid inessential reconfigurations, and consider various physical constraints [5]. Inspired by the Lyapunov technique [8] as a main iterative optimization method, a general scheme for iterative configuration in an SDN-based EON is introduced. The scheme includes a sequence of simple (usually linear) optimizations followed by few recursive update equations, can stably and causally be implemented using available software optimizers, is mathematically supported in many operational conditions, and iteratively approaches the optimal solution of the large-scale stochastic resource allocation formulation that governs the EON. The generality of the scheme allows to cover different reconfiguration criteria, operational limitations, service requirements, and queuing policies described by an arbitrary set of instantaneous or time-averaged mathematical constraints. The proposed scheme is developed based on a cross-layer perspective and can reconfigure tunable elements of the different networking layers. As an example, the application of the proposed scheme for an impairment-aware power-efficient resource allocation is demonstrated and the effect of bandwidth limitations on its power efficiency and service quality is investigated.

II. Mathematical Representation

Consider a general EON, which is controlled by a central SDN controller and operates in discrete time intervals $n \in \mathbb{Z}_0^\infty$, where $\mathbb{Z}_a^b$ includes the integers between a and b inclusively. There are I connections, each having a queue with backlog $q_i[n]$, which is filled and emptied by $a_i[n]$ arrived and $s_i[n]$ departed bits in every interval n. The served bits $s_i[n]$ are determined according to the configurable parameters of the EON (such as modulation type), which are collected in a resource vector $\boldsymbol{\alpha}[n]$. The resource vector $\boldsymbol{\alpha}[n]$ is periodically reconfigured in each interval n such that a targeted performance metric (such as power consumption) is optimized, while some QoS requirements (such as average transmission delay) are guaranteed and operational constraints (such as spectrum non-overlapping) are satisfied.

In each interval n, the resource vector $\boldsymbol{\alpha}[n]$ can be reconfigured by solving the optimization problem

$$\min_{\boldsymbol{\alpha}[n]} \quad \Gamma f_0[n] + \textstyle\sum_{i\in\mathbb{Z}_1^I} q_i[n](a_i[n]-s_i[n]) + \sum_{l\in\mathbb{Z}_1^{L_1}} p_{l,1}[n] f_{l,1}[n] + \sum_{l\in\mathbb{Z}_1^{L_2}} p_{l,2}[n] f_{l,2}[n] \quad \text{s.t.} \tag{1a}$$

$$f_{l,3}[n] \leqslant 0, \quad \forall l \in \mathbb{Z}_1^{L_3} \tag{1b}$$

$$f_{l,4}[n] = 0, \quad \forall l \in \mathbb{Z}_1^{L_4} \tag{1c}$$

followed by the recursive update equations

$$q_i[n+1] = \max\{q_i[n] + a_i[n] - s_i[n], 0\}, \quad i \in \mathbb{Z}_1^I, \tag{2a}$$

$$p_{l,1}[n+1] = \max\{p_{l,1}[n] + f_{l,1}[n], 0\}, \quad l \in \mathbb{Z}_1^{L_1}, \tag{2b}$$

$$p_{l,2}[n+1] = p_{l,2}[n] + f_{l,2}[n], \quad l \in \mathbb{Z}_1^{L_2}, \tag{2c}$$

with initialization $q_i[0] = 0$, and $p_{l,m}[0] = 0$, $m \in \mathbb{Z}_1^2$. $f_0[n]$, $f_{l,m}[n]$, and $s_i[n]$ are arbitrary functions of $\boldsymbol{\alpha}[n]$, while Γ and $p_{l,m}[n]$ are called Lyapunov penalty coefficient and virtual queue length, respectively. Optimization (1) approaches the minimized time-averaged objective $\overline{f_0}$ with a proxim-

This work was supported in part by Vinnova under grant no. 2017-05228 and the Knut and Alice Wallenberg Foundation under grant no. 2013.0021.

ity controlled by Γ, where the time average is defined as $\overline{x} = \lim_{n\to\infty} \frac{1}{n} \sum_{n'\in\mathbb{Z}_0^{n-1}} \mathcal{E}\{x[n']\}$ with an expectation over random arrival events. As proven in [8] using the Lyapunov analysis, when the arrival process $a_i[n]$ is independent and identically distributed (i.i.d.) over intervals and $q_i[n]$ does not appear in the functions $f_{l,m}[n]$, the first penalty term in (1a) and update equation (2a) assures the queue stability by imposing $\overline{a_i} \leqslant \overline{s_i}$ to keep queue lengths finite over time. Similarly, the second and third penalty terms along with their corresponding update equations (2b) and (2c) enforce $\overline{f_{l,1}} \leqslant 0$ and $\overline{f_{l,2}} = 0$, respectively, which can be used to describe various QoS and operational constraints using time-averaged inequality or equality expressions. Moreover, (1b) and (1c) define additional instantaneous inequality and equality constraints to be satisfied in each interval. The described scheme can handle unexpected conditions with the least reconfiguration efforts since the network status directly intervenes to optimization (1) by $p_{l,m}[n]$ and $q_i[n]$. Moreover, to choose a resource vector with the least necessary reconfigurations, a previously obtained $\boldsymbol{\alpha}[n]$ can be used as an initial seed for the next optimization, or a constraint may be defined using $\overline{f_{l,m}}$ to keep the most possible similarity between resource vectors.

When $a_i[n]$ is not i.i.d. or $q_i[n]$ appears in $f_{l,m}[n]$, the performance of (1) is not mathematically supported. However, even in more relaxed scenarios, the proposed general scheme yields acceptable results, as illustrated in [5] and [9], where we have employed the scheme for special cases of power- and spectrum-efficient QoS-aware resource allocation in EONs, respectively. To further show the generality and applicability of the scheme, we show in the next section how bandwidth limitations can also be considered.

III. Bandwidth Limitations

Targeting power efficiency, flexible transponders of an EON with I connections can be reconfigured in each interval with duration T according to the general scheme described in Sec. II. Assume that modulation format and working bandwidth of the flexible transponder can be selected from K available modulation types and N equal-bandwidth frequency slots with granularity W, respectively. The maximum selectable bandwidth for each modulation is limited by physical constraints such as signal-to-noise ratio (SNR) commitments and finite lasing bandwidth. Let the resource vector $\boldsymbol{\alpha}[n]$ include the binary variables $x_{i,k}[n]$ and $t_{i,i'}[n]$, and nonnegative integer variables $d_i[n]$, $h_i[n]$, and $b_{i,k}[n]$ for $i \in \mathbb{Z}_1^I$ and $k \in \mathbb{Z}_1^K$. Optimization (1) is converted to an integer linear program for this special case using the substitutions $\Gamma = 1$, $L_1 = L_2 = 0$, $L_3 = 3I + IK + I^2$, $L_4 = I^2$,

$$s_i[n] = TW\sum_{k=1}^K C_k b_{i,k}[n] + d_i[n], \quad i \in \mathbb{Z}_1^I \tag{3a}$$

$$f_0[n] = \sum_{i=1}^I \left[\sum_{k=1}^K b_{i,k}[n](E + FC_k) + Vd_i[n]\right] \tag{3b}$$

$$f_{l,3}[n] = a_i[n] - s_i[n], \quad i \in \mathbb{Z}_1^I \tag{3c}$$

$$f_{l,3}[n] = h_i[n] + \sum_{k=1}^K b_{i,k}[n] - N, \quad i \in \mathbb{Z}_1^I \tag{3d}$$

$$f_{l,3}[n] = \sum_{k=1}^K x_{i,k}[n] - 1, \quad i \in \mathbb{Z}_1^I \tag{3e}$$

$$f_{l,3}[n] = b_{i,k}[n] - U_{i,k}x_{i,k}[n]/W, \quad i \in \mathbb{Z}_1^I, k \in \mathbb{Z}_1^K \tag{3f}$$

$$f_{l,3}[n] = h_i[n] + \sum_{k=1}^K b_{i,k}[n] + G - h_{i'}[n] - (N+G) \Big(3 - t_{i,i'}[n] - \sum_{k=1}^K x_{i,k}[n] - \sum_{k=1}^K x_{i',k}[n]\Big), \quad i, i' \in \mathbb{Z}_1^I : \mathbf{P}_i \cap \mathbf{P}_{i'} \neq \varnothing \tag{3g}$$

$$f_{l,3}[n] = h_i[n] + \sum_{k=1}^K b_{i,k}[n] + G - h_{i'}[n] - (N+G), \quad i, i' \in \mathbb{Z}_1^I : \mathbf{P}_i \cap \mathbf{P}_{i'} = \varnothing \tag{3h}$$

$$f_{l,4}[n] = t_{i,i'}[n] + t_{i',i}[n] - 1, \quad i, i' \in \mathbb{Z}_1^I : i \neq i', \tag{3i}$$

$$f_{l,4}[n] = t_{i,i'}[n] + t_{i',i}[n], \quad i, i' \in \mathbb{Z}_1^I : i = i' \tag{3j}$$

where $l = i$, $l = I + i$, $l = 2I + i$, and $l = 3I + (i-1)K + k$ in (3c)–(3f), respectively, $l = 3I + IK + (i-1)I + i'$ in (3g) and (3h), and $l = (i-1)I + i'$ in (3i) and (3j). According to (3a), $TWC_k b_{i,k}[n]$ and $d_i[n]$ bits are transmitted or dropped, respectively, from the feeding queue of the transmit transponder, where $b_{i,k}[n]$ is the number of assigned frequency slots to connection i having modulation k with spectral efficiency C_k. The transponders of connection i with modulation k consumes the power $b_{i,k}[n](E + FC_k)$, where E and F are constant [5]. The objective function $f_0[n]$ in (3b) is a mixture of the transponder power consumption and number of dropped bits with a weighting coefficient V to adjust optimization priority. A simple buffering policy is given in (3c), where the entire $a_i[n]$ arrived bits in an interval are completely served in that interval. Therefore, no bit is buffered in the queue to be served in the next intervals and the buffer capacity should be enough to store the maximum arrived bits during an interval. The assigned slots settle in the working fiber bandwidth by (3d), where $h_i[n]$ is the first slot number, and $x_{i,k}[n]$ equals 1 if connection i uses path $\mathbf{P}_i$ and modulation k, and 0 otherwise. Each connection selects at most one modulation by (3e). Bandwidth limitations are considered in (3f) by keeping the number of slots $b_{i,k}[n]$ below a permitted upper bound $U_{i,k}/W$. Expressions (3g) and (3h) avoid spectrum overlapping by keeping G guard slots between adjacent spectrum slots of intersecting paths, where the relative location of the spectrum slots $t_{i,i'}[n]$ is set to 1 by (3i) and (3j) if $h_i[n] \leqslant h_{i'}[n]$, and 0 otherwise. Since $a_i[n] \leqslant s_i[n]$ by (3c) and consequently $q_i[n] = 0$, and since $L_1 = L_2 = 0$, the recursive equations (2) and penalty terms in (1a) are omitted.

Unlike [5], several definitions are used for $U_{i,k}$ in (3f) to describe various bandwidth limitations. As a benchmark, when the whole fiber bandwidth can be freely occupied, $U_{i,k} = NW$. Other possible definitions are

$$U_{i,k} = \min\{2/\sqrt{27\chi\Omega_i^2\Theta^3\Psi_k^3}, NW\}, \tag{4a}$$

$$U_{i,k} = \min\{\Delta, 2/\sqrt{27\chi\Omega_i^2\Theta^3\Psi_k^3}, NW\}. \tag{4b}$$

The main SNR limitations are included in (4a) [5], which determines the maximum spectrum bandwidth for which the SNR expression in [10, Sec. IV-A] is greater than $\Theta\Psi_k$, where χ, Θ, Ψ_k, and Ω_i denote the power coefficient, SNR margin, SNR threshold of modulation k, and accumulated spontaneous noise per unit of spectrum along path $\mathbf{P}_i$, respectively. Definition (4b) is used if the assigned spectrum is further limited by the finite lasing bandwidth Δ.

The impact of the different bandwidth limitations in (4)

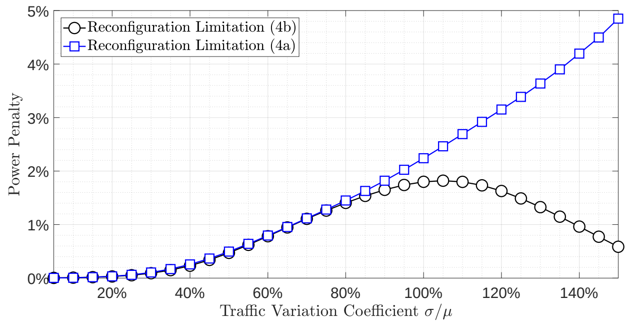

Fig. 1: Impact of bandwidth limitations on power consumption.

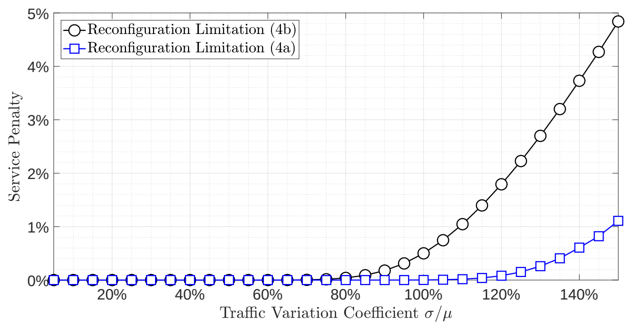

Fig. 2: Impact of bandwidth limitations on resource shortage.

on the formulated scheme (3) is numerically evaluated for a commercial metro network topology covering Stockholm [5]. The constants are $K = 5$, $T = 5$ s, $N = 640$, $W = 6.25$ GHz, $G = 1$, $V = 1000$, $E = 75.6$ W, and $F = 18.75$ W [5]. The first quadrature amplitude modulations with the spectral efficiency $C_k = 2k$ bits per dual polarization symbol and required SNR threshold Ψ_k given in [11] are employed. The arrival processes $a_i[n]$ are i.i.d. over all intervals and connections, with a log-normal distribution of an average $\mu = 100T$ Gbit and a variable standard deviation σ to change traffic dynamism. Thanks to the short links of the benchmark topology, each fiber link has only a pair of input and output amplifiers, whose gains equal the switching and fiber loss, respectively. The attenuation coefficient, dispersion coefficient, nonlinear constant, optical frequency, spontaneous emission factor, switch loss, SNR margin, and lasing bandwidth are set to 0.22 dB/km, 20393 fs^2/m, 1.3 1/W/km, 193.55 THz, 1.58, 15 dB, $\Theta = 3$ dB, and $\Delta = 50$ GHz, respectively, to approximate χ and compute Ω_i and $U_{i,k}$ according to [5], [10].

We define the power penalty as the extra power consumed to compensate for the bandwidth limitation compared to the power consumption of the reference limitless case with $U_{i,k} = NW$. The service penalty is the ratio between the time-averaged total number of dropped and arrived bits, and measures the unserved portion of the traffic load. Figs. 1 and 2 show the power and service penalties versus traffic variation coefficient σ/μ for the bandwidth limitations (4). For both (4a) and (4b), the penalties increase with the traffic variation σ/μ. In fact, when traffic variation increases, there are more sudden traffic peaks and thus, more power is consumed to serve this high peak conditions. Sometimes, the arrived traffic peak is far from the offered capacity of the transponders and part of the arriving bits should be dropped due to resource shortage, as shown in Fig. 2. For $\sigma/\mu \leqslant 80\%$, both cases have almost the same behavior but for $\sigma/\mu > 80\%$, the limitation (4b) offers a less power penalty at the cost of a higher service penalty.

IV. Extensions and Developments

The generality of the proposed scheme allows to have a variety of extensions. Reconfiguration limitations such as service interruption or tuning speed can be included by adding more constraints to (1). Considering a cross-layer optimization perspective over networking layers, different buffering policies alongside elastic reconfiguration of the optical elements can be described using (1). QoS requirements can also be ensured by adding more time-averaged constraints to (1), as illustrated in [5]. Moreover, the optimization may aim at different objectives with arbitrary mathematical dependencies [9].

V. Conclusions

A generic mathematical platform for iterative reconfiguration of an SDN-based EON is developed and its performance is demonstrated through an application example. Some possible extensions and developments are also discussed. Simulation results confirm that the power efficiency and service quality can be undesirably affected by different bandwidth limitations, and this undesired effect is stronger for a resource allocation with more dynamic traffic loads.

References

[1] Z. Zhu, X. Chen, C. Chen, S. Ma, M. Zhang, L. Liu, and S. B. Yoo, "OpenFlow-assisted online defragmentation in single-/multi-domain software-defined elastic optical networks," *Journal of Optical Communications and Networking*, vol. 7, no. 1, pp. A7–A15, 2015.

[2] Z. Zhu, C. Chen, X. Chen, S. Ma, L. Liu, X. Feng, and S. B. Yoo, "Demonstration of cooperative resource allocation in an OpenFlow-controlled multidomain and multinational SD-EON testbed," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1508–1514, 2015.

[3] G. Shen, Y. Zhang, X. Zhou, Y. Sheng, N. Deng, Y. Ma, and A. Lord, "Ultra-dense wavelength switched network: A special EON paradigm for metro optical networks," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 189–195, 2018.

[4] P. Layec, A. Dupas, D. Verchère, K. Sparks, and S. Bigo, "Will metro networks be the playground for (true) elastic optical networks?" *Journal of Lightwave Technology*, vol. 35, no. 6, pp. 1260–1266, 2017.

[5] M. Hadi and E. Agrell, "Joint power-efficient traffic shaping and service provisioning for metro elastic optical networks," *Journal of Optical Communications and Networking*, vol. 11, no. 12, pp. 578–587, 2019.

[6] B. C. Chatterjee, S. Ba, and E. Oki, "Fragmentation problems and management approaches in elastic optical networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 183–210, 2017.

[7] Y. Xu, E. Agrell, and M. Brandt-Pearce, "Static resource allocation for dynamic traffic," in *European Conference on Optical Communication Dublin (ECOC)*, 2019.

[8] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[9] M. Hadi, M. R. Pakravan, and E. Agrell, "Dynamic resource allocation in metro elastic optical networks using Lyapunov drift optimization," *Journal of Optical Communications and Networking*, vol. 11, no. 6, pp. 250–259, 2019.

[10] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694–721, 2014.

[11] L. Yan, E. Agrell, H. Wymeersch, and M. Brandt-Pearce, "Resource allocation for flexible-grid optical networks with nonlinear channel model," *Journal of Optical Communications and Networking*, vol. 7, no. 11, pp. B101–B108, 2015.