Adaptive Elastic Spectrum Allocation Based on Traffic Fluctuation Estimate under Time-Varying Traffic in Flexible OFDM-Based Optical Networks*

Mirai CHINO(✉1), Student Member, Misato KAMIO1, Nonmember, Jun MATSUMOTO1, Member, Eiji OKI1, Satoru OKAMOTO1, and Naoaki YAMANAKA1, Fellows

SUMMARY A flexible orthogonal frequency-division multiplexing optical network enables the bandwidth to be flexibly changed by changing the number of sub-carriers. We assume that users request to dynamically change the number of sub-carriers. Dynamic bandwidth changes allow the network resources to be used more efficiently but each change takes a significant amount of time to complete. Service centric resource allocation must be considered in terms of the waiting time needed to change the number of sub-carriers. If the user demands drastically increase such as just after a disaster, the waiting time due to a chain-change of bandwidth becomes excessive because disaster priority telephone services are time-critical. This paper proposes a Grouped-elastic spectrum allocation scheme to satisfy the tolerable waiting time of the service in an optical fiber link. Spectra are grouped to restrict a waiting time in the proposed scheme. In addition, the proposed scheme determines a bandwidth margin between neighbor spectra to prevent frequent reallocation by estimating real traffic behavior in each group. Numerical results show that the bandwidth requirements can be minimized while satisfying the waiting time constraints. Additionally, measurement granularity and channel alignment are discussed.

key words: elastic optical networks, spectrum allocation, optical OFDM, spectrum sharing, time-varying traffic

1. Introduction

The traffic in core networks continues to grow every year. The main reason for this growth is the increased popularity of various real-time applications, for example video streaming [1], [2]. For enhancing the core network, high-capacity and high-efficiency optical transmission technologies are being researched. A fixed channel spacing Wavelength-Division Multiplexing (WDM) network is one such technology and achieves high-capacity optical transmission with 50 GHz or 100 GHz fixed channel spacing. However, in a WDM network, if the bandwidth assigned to a channel is larger than the load of the channel, the frequency bandwidth efficiency decreases because one grid is assigned for one channel. For example, channel spacing 200 GHz may be required for 400 Gbps WDM transmission, but 25 GHz may be enough for 10 Gbps. Therefore, 175 GHz bandwidth is wasted by each 10 Gbps channel in mixed 10 Gbps/400 Gbps transmission.

Orthogonal Frequency-Division Multiplexing (OFDM)-based networks which can more flexibly assign channel spacing than WDM are being intensively studied [3]. In OFDM, several carriers (sub-carriers) are allocated for one channel. The number of subcarriers depends on user demand.

In this paper, we focus on elastic spectrum allocation (SA), which is defined as the ability to dynamically change allocated spectra in the OFDM-based networks. The OFDM-based networks enable to expand/reduce slot width of channel. Furthermore future OFDM-based networks will change slot width according to time varying traffic by changing the number of sub-carriers flexibly. Spectrum expansion/reduction denotes increase/decrease in the number of sub-carriers. For example, in a video-on-demand service, service providers require more bandwidth at night hours with high demands, while they can released the unused bandwidth for other network services at morning hours with low demand.

Several elastic SA schemes that change the bandwidth dynamically have been studied in [4]–[7]. The authors of [4] places two connections adjacent to share the optical spectra. A general policy to allocate sub-carriers to time-varying traffic was presented in [5]. Elastic SA schemes in the OFDM-based networks are categorized into three schemes: a Deterministic SA scheme, a Semi-elastic SA scheme, and an Elastic SA scheme [6], [7].

In the Deterministic SA scheme, both center frequency and sub-carriers are statically assigned. The Deterministic SA scheme accommodates time-varying traffic with only allocated bandwidth. Each channel performs stable communications because of guaranteed bandwidth. On the other hand, if spectrum usage is smaller than allocated bandwidth, bandwidth is not efficiently utilized.

In the Semi-elastic SA scheme, a center frequency is statistically assigned and sub-carriers are dynamically assigned. The Semi-elastic SA scheme provides statistical multiplexing effect by sharing bandwidth with the neighbor channels. Meanwhile, each spectrum has to keep space be-
tween neighbor spectra large enough to avoid blocking of an expansion request on time-varying sub-carriers.

In the Elastic SA scheme, both center frequency and sub-carriers are dynamically assigned. Furthermore, spectrum reallocation prevents blocking of the expansion request on time-varying sub-carriers in spite of smaller space. There are two spectrum reallocation techniques: Make-before-break and Push-pull [9]–[14]. Make-before-Break needs another transponder for source and destination nodes to generate a new center frequency. Push-pull needs to shift with a certain speed while synchronizing optical switches, transmitter and receiver. Hence Push-pull requires shifting-speed-dependent time for reallocation spectra. For instance, if Push-pull shifting speed is 100 ms/2.5 GHz [12], [13], the shifting time takes 2 seconds for 50 GHz shifting spectrum.

Suppose that the Elastic SA scheme uses Push-pull reallocation. If each spectrum is allocated without any margin space between neighbor spectra, spectrum reallocation frequently occurs each time increasing the number of sub-carries is requested. In other words, when a spectrum expansion is requested, spectrum reallocation continuously occurs for reservation of adequate continuous bandwidth. As a result, communication services have to wait until the spectra are expanded. If the user demands drastically increase such as just after a disaster, the waiting time due to a chain-change in the spectrum becomes excessive because disaster priority telephone services are time-critical.

Conventional SA schemes cannot guarantee the requirement for the waiting time. The tolerable waiting time is different depending on a type of services. For this reason, the continuous reallocation that causes the unpredictable waiting time should be restricted.

This paper proposes a Grouped-elastic SA scheme as an elastic SA scheme to satisfy a different waiting time constraint for each service [15], [16]. The proposed scheme groups the spectra according to the tolerable waiting time, and restricts the range of Push-pull reallocation. A space between neighbor spectra is defined as a margin bandwidth. A margin bandwidth is determined by the traffic fluctuation estimate in order to prevent frequent spectrum reallocation. The feature of the proposed scheme is to manage both waiting time and bandwidth utilization rate by designing the suitable group size based on the constraint of waiting time of each service to meet the requirement. Our performance evaluations confirm that the proposed scheme reduces the required bandwidth per spectrum with satisfying the waiting time constraint.

Note that this paper considers a spectrum bandwidth management by monitoring traffic behaviors in an optical fiber link as the first step to clarify the basic characteristics of the management. The scenario of one optical fiber link in this work can be extended to that of the whole network topology.

This paper is organized as follows. Section 2 introduces the elastic spectrum allocation schemes. Section 3 presents the spectrum reallocation techniques. Section 4 presents the proposed scheme called Grouped-elastic SA. Section 5 presents performance evaluation of the proposed scheme compared with that of the conventional scheme. Finally, this work is summarized in Sect. 6.

2. Conventional Elastic Spectrum Allocation Schemes

The Deterministic SA and Semi-elastic SA schemes do not shift the center frequency. On the other hand, the Elastic SA scheme shifts the center frequency and their spectra. Shifting the center frequency and its spectrum is called as a spectrum reallocation. Spectrum reallocation prevents a increase of sub-carriers to be blocked. There are two spectrum reallocation techniques: Make-before-Break generates a new spectrum and performs wavelength switching using another transponder. Push-pull secures continuous bandwidth and synchronizes the transmitter and the receiver and performs switching optical wavelength.

This section introduces features and problems of conventional elastic spectrum allocation schemes, which are Semi-elastic SA and Elastic SA. Further Make-before-Break and Push-pull techniques related with the Elastic SA scheme are explained.

2.1 Semi Elastic Spectrum Allocation Scheme

The Semi-elastic SA scheme has time-flexible allocated bandwidth, where the center frequency is static. Figure 1 shows that when time $t$ changes to $t + 1$, channels 1 and 2 request channel reduction and expansion respectively according to each traffic fluctuation. Changing bandwidth for each request flexibly enables to share bandwidth with neighbor channels temporally. As a result, the Semi-elastic SA scheme obtains a statistical multiplexing gain. Each channel expands/reduces the spectrum slots with the static center frequency. There are several channel expansion/reduction policies such as using higher frequency than center frequency or using alternately higher and lower frequency [9]. At the same time, the Semi-elastic SA scheme needs to design spaces between neighbor channels large enough to prevent blocking of bandwidth change requests.

2.2 Elastic Spectrum Allocation Scheme

The Elastic SA scheme can change assigned bandwidth such
as the Semi-elastic SA scheme and center frequency is not static, as shown in Fig. 2. When time $t$ changes to $t+1$, channel 2 requests channel expansion because of increasing traffic. However, channel 2 cannot increase sub-carriers directly due to neighbor channels 1 and 3. Then channel 2 expands after channel 1 is reallocated.

As an advantage of the Elastic SA scheme, it is capable to avoid request blocking. Even if spaces between neighbor channels are small, the Elastic SA scheme accommodates time-varying traffic by reallocating channels. The simplest approach in the Elastic SA scheme is a Pure-elastic SA scheme. The Pure-elastic SA scheme aims to minimize required bandwidth by minimizing spaces between neighbor channels. The Pure-elastic SA scheme can utilize frequency resources without a small piece of unused bandwidth (fragment) thanks to reallocation.

3. Spectrum Reallocation Techniques

3.1 Make-before-Break

Figure 3 presents a Make-before-Break reallocation behavior. First, another connection is prepared for switching connection between sender and receiver switches. Next, the receiver switch switches connection from original frequency $f_0$ to the other frequency $f_1$. Finally, the original connection is torn down. Make-before-Break needs another transponder at both sender and receiver switches. That is, Make-before-Break requires additional transponders.

3.2 Push-Pull

Figure 4 depicts a Push-pull behavior. The first step is the reservation of contiguous frequency from original frequency $f_0$ to new frequency $f_1$. The second step is the shifting original frequency $f_0$ to $f_1$ using a tunable laser. The last step is the reduction of the frequency bandwidth to new frequency $f_1$ by releasing contiguous frequency from $f_0$ up to but not including $f_1$. Push-pull guarantees no traffic disruption in principle, and does not require another transponder. However, Push-pull takes some time to shift frequencies. Currently, a demonstration experiment indicates 100 ms/2.5 GHz shifting speed [12], [13].

Suppose that Pure-elastic SA uses Push-pull as the reallocation technique for no disruption. In Pure-elastic SA, channel expansion request needs waiting time. As an example shown in Fig. 2, we consider the situation, where channel 2 requests channel expansion at time $t$, but there is no space for expansion without any reallocation. The expansion request cannot satisfy at time $t$. First, channel 1 is reallocated at time $t+1$. Second, channel 2 is reallocated and expanded at time $t+2$. If a frequency slot is 12.5 GHz/slot and Push-pull shifting time is 100 ms/2.5 GHz, spectrum reallocation takes 1.5 seconds corresponding to 3 slots, where channel expansion time is ignored. If channel 1’s location at time $t$ is the same as that at time $t+1$, Push-pull time reduces to 2 slot’s shifting time. The minimization of spaces between neighbor channels causes a lot of waiting time for time-varying requests. Meanwhile, tolerable waiting time differs depending on the services. The shifting range of spectrum reallocation that accounts for a large portion of total waiting time should be restricted with respect to each service.

4. Grouped Elastic Spectrum Allocation Scheme

We present our proposed scheme, called as Grouped-elastic SA scheme, to solve the waiting time problem in Pure-elastic SA. Each channel belongs to the group corresponding to a tolerable waiting time, as shown in Fig. 5. The spaces between neighbor channels are small for a large tolerable waiting time, while they are large for a small one. When reallocation is required, Push-pull is used only within the group in order to satisfy each services waiting time constraints. To adapt time-varying traffic, a margin bandwidth is designed for each space between neighbor channels. A margin bandwidth is determined by using probability density functions (PDFs) of link rate of neighbor channels. Reallocated channel goes back to an initially-allocated center
The channel has the margin bandwidth with neighbor channels to prevent frequent reallocation. The channel is grouped adequately to satisfy each service waiting time.

To achieve the above purposes, the flow chart of the proposed scheme is shown in Fig. 6. All traffic is constantly monitored by a traffic monitor. The traffic monitor counts the link rate of each of traffic and creates the histogram of the link rate. The histogram is utilized as PDF for grouping and margin calculation. Service providers determine in advance the tolerable waiting time of the service and the tolerable reallocation probability. The proposed scheme consists of two steps. In the grouping step, the PDF of traffic and tolerable waiting time are given as inputs. The grouping step calculates the adequate group size belonging to the channel. In the margin calculation step, a margin bandwidth is calculated for elastic spectrum allocation by using the tolerate reallocation probability, the group size and PDF of the traffic.

4.1 Grouping Step

Spectra are grouped based on the tolerable waiting time and the PDF of each channel. Each group size is designed with shifting speed and the tolerable waiting time. For example, the service whose tolerable waiting time is 5 seconds tolerates frequency shift up to 125 GHz (10 slots) given 100 ms/2.5 GHz shifting speed. In other words, the group of the service is designed for maximum 125 GHz group size. The frequency shift is performed only in the group so as not to violate the service waiting time constraint. All channels are allocated in the group that has a small size enough to satisfy the constraint.

4.2 Margin Calculation Step

The margin calculation with neighbor channels is performed in order to prevent spectrum reallocation of frequent occurrence. The selected group size, the reallocation probability user tolerates, and the PDF of the traffic are considered for margin calculations.

Figure 7 explains margin bandwidth calculation between channels $i$ and $j$. The left and right used bandwidth based on the center frequency are defined as $x_i$ and $x_j$ respectively. The used bandwidth between channels $i$ and $j$ is defined as $z_{ij} = x_i + x_j + \text{guardband}$. The guardband is assigned to avoid the interference between adjacent channels [5]. The space between channels $i$ and $j$ is defined as margin bandwidth $M_{ij}$. Spectrum reallocation occurs when $z_{ij}$ exceeds $M_{ij}$. This reallocation probability $P(z_{ij} \geq M_{ij})$ is expressed as below:

$$P(z_{ij} \geq M_{ij}) = \int_{M_{ij}}^{\infty} f_{ij}(z_{ij}) dz_{ij},$$

where $f_{ij}(z_{ij})$ is the PDF of $z_{ij}$, which is, the PDF of the sum of channels $i$ and $j$ bandwidth.

Margin bandwidth is designed to satisfy the probability of spectrum reallocation $T$ (e.g. $T=0.01$). The shaded area in Fig. 8 represents the probability of spectrum reallocation $T$. The margin bandwidth $M_{ij}$ is calculated so that the shaded area is less than the spectrum reallocation probability $T$. The PDF $f_{ij}(z_{ij})$ is expressed by convolving the PDFs for channels $i$ and $j$: 
where \( f_i(x) \) and \( f_j(x) \) are the PDF of bandwidth of the channel \( i \) and \( j \). Figure 9 shows the PDF of bandwidth calculated with core network traffic obtained by CAIDA traces [17]: the link data of OC192 backbone link (9953 Mbps) of a Tier1 ISP between Chicago, IL and Seattle, WA at March 20th, 2014 from 13:01:00 UTC~13:01:59 UTC, average 1.4 Gbps. For simplicity, modulation method assumes Binary Phase-Shift Keying (BPSK), and spectral efficiency 1 b/s/Hz. Both channels \( i \) and \( j \) utilize the same traffic data. For example, when \( T=0.01 \) is given, \( M_{ij} \) is designed so that the area in Fig. 8 is less than 10\%, and thus the required margin bandwidth becomes 3.04 GHz. When \( T \approx 0.0 \) is given, the required margin bandwidth is 3.54 GHz. Grouped-elastic SA reduces the required bandwidth to coordinate the margin bandwidth.

4.3 Channel Alignment Policies

A channel alignment policy determines the alignment sequence of channels depending on their characteristics. A channel alignment affects the bandwidth utilization, since the total margin bandwidth in the group is different depending on the alignment. We investigate how channel alignment policies affect the performance of the proposed scheme.

Figure 10 depicts the alignment example of the channel 1, channel 2, and channel 3. Channel \( i \) has the PDF \( f_i(x_i) \) of its traffic in Fig. 10. \( f_i(x_i) \) mean \( \mu_i \) and standard deviation \( \sigma_i \). The margin bandwidth with channels \( i \) and \( j \) is represented by \( M_{ij} \) in Eq. (1). \( M_k \) is the margin bandwidth of the edge channel \( k \) in the group. The margin bandwidth of the edge channel \( M_k \) is calculated in the same manner as Eq. (1). The reallocation probability \( P(x_k \geq M_k) \) is expressed as below:

\[
P(x_k \geq M_k) = \int_{M_k}^{\infty} f_k(x_k) dx_k \leq T,
\]

(3)

where \( T \) is the probability of spectrum reallocation. We assume that \( \sigma_1 < \sigma_2 < \sigma_3 \). The channel alignment in Fig. 10 has three patterns, \{channel 1, channel 2, channel 3\}, \{channel 1, channel 3, channel 2\}, and \{channel 2, channel 1, channel 3\}. To take permutation into consideration, there are six patterns. Note that each of the six patters belongs to one of the above three patterns from the point of view of the total margin bandwidth.

We model the channel alignment as a complete graph \( G(V, E) \), where the set of channels sorted by standard deviation \( \sigma \) is denoted as \( V \), and the set of adjacent channel relationships is denoted as \( E \). The link cost corresponds to the margin bandwidth \( M_{ij} \) for adjacent channels. The minimization problem of the channel alignment can be converted to the shortest hamiltonian path problem (e.g. Fig. 11) to minimize the following objective function:

\[
\min_H \sum_{(i,j) \in H} (M_{ij} + M_r + M_l),
\]

(4)

where the Hamiltonian path is denoted as \( H \), \( r \) is the right edge channel, \( l \) is the left edge channel in the group. The starting vertex of the hamilton path corresponds to the left edge channel in the group. The ending vertex is the right edge channel. Note that the traveling salesman problem, which is known as NP-Complete [18], is polynomially reducible to this minimum-cost Hamiltonian path problem in a full-mesh network. Therefore, the the minimum-cost Hamiltonian path problem is also NP-complete.

We introduce three policies to investigate to find a suitable channel alignment: ascending order, middle min, and middle max. In the ascending-order policy, the channel of the left edge has the minimum deviation. The neighbor has the second minimum standard deviation. The channel of the right edge has the maximum one. In the middle-min policy, the channel that has the minimum standard deviation is allocated at the middle of the group. Then, the channel is allocated at the left and right next to the middle channel alternately in ascending order of standard deviations. The inner channel has smaller standard deviation than the outer one. The middle-max policy is the reverse alignment to the middle-min policy. The inner spectrum has larger standard deviation than the outer one.

**Fig. 9** Probability of bandwidth calculated with CAIDA traffic traces.

**Fig. 10** An alignment example with three channels.

**Fig. 11** Hamiltonian path problem with five vertices.
5. Numerical Results

Performances of the Semi-elastic SA scheme, the Pure-elastic SA scheme and the proposed Grouped-elastic SA scheme are evaluated in this section. Resource utilization, measurement granularity, waiting time of services, and performance of the channel alignment are measured.

5.1 Resource Utilization

The required bandwidth per channel is evaluated as resource utilization in Fig. 12 and Fig. 13. The group size indicates the number of channels belonging to the group. The group size is assumed from 1 to 16.

First, we investigate the basic benefit of the grouping in Fig. 12. In this evaluation, each channel utilizes the average 1.4 GHz traffic by CAIDA traces [17], and every channel is assumed to be the same traffic. The spaces of the channels are determined by calculating the margin bandwidth to satisfy the given reallocation probabilities. The spectrum reallocation is assumed to be Push-pull technique. When the group size is 2 in Fig. 12, the plots represent the required bandwidth in the Semi-elastic SA. When the group size is 16, the plots describe the required bandwidth in the Pure-elastic SA, since the Pure-elastic SA may convolve all channels (in this evaluation, the maximum number of spectra is 16.) in the reallocation.

In Semi-elastic SA, the required bandwidth increases for the average 1.4 GHz traffic, since Semi-elastic SA shares the bandwidth with only the neighbor channels. The number of the involved channel increases with an increase of group size in the reallocation. As a result, Pure-elastic SA depicts the lowest required bandwidth. At the same time, the reduction effect of the required bandwidth is decreased with an increasing of group size.

Under the actual traffic situation, the traffic distribution depends on services, locations, and the like. Second, we analyze the benefit under different traffic distributions. The average 1.4 GHz and $\sigma = 98$ MHz traffic by CAIDA traces is utilized [17]. The standard deviation $\sigma'$ of the traffic distribution of each channel is set to be equally distributed in the range from $\sigma$ to $\sigma_{\text{max}}$, where $\sigma_{\text{max}}$ is set to $\sigma$, 2$\sigma$, 3$\sigma$, 4$\sigma$, and 5$\sigma$. Let $n$ be the number of channels accommodated in the link. We set $\sigma' = (\sigma, \sigma + \frac{1}{n}(\sigma_{\text{max}} - \sigma), \sigma + \frac{1}{n}(\sigma_{\text{max}} - \sigma), \cdots, \sigma_{\text{max}})$. The spaces of the channels are determined by calculating the margin bandwidth to satisfy the given reallocation probability of 0.01.

Figure 13 shows the relation between group size and required bandwidth per channel with the reallocation probability 0.01 under different traffic distributions. We observe how much bandwidth each channel requires by using Fig. 13. The traffic distribution, group size and reallocation probability are used for calculating the required bandwidth.

The larger range of $\sigma'$ is, the more benefit of grouping is. In other words, grouping and margin bandwidth are effective in the situation with the various types of traffic flows. In $\sigma' = [\sigma, \cdots,\sigma]$, the required bandwidth per channel is about 1.7 GHz when the group size is one. The required bandwidth per channel is about 1.5 GHz when the group size is 16. The reduction of required bandwidth is about 0.2 GHz. In $\sigma' = [\sigma, \cdots, 5\sigma]$, the required bandwidth per channel is about 2.85 GHz when the group size is one. The required bandwidth per channel is about 1.6 GHz when the group size is 16. The reduction of required bandwidth is about 1.25 GHz. The reduction of required bandwidth gets larger as $\sigma_{\text{max}}$ increases.

When group size is relatively small, the larger the range of $\sigma'$ is, the more the required bandwidth per channel is. When the group size is one, the required bandwidth per channel in $\sigma' = [\sigma, \cdots, 5\sigma]$ is about 2.85 GHz. The required bandwidth per channel in $\sigma' = [\sigma, \cdots, \sigma]$ is about 1.7 GHz. The required bandwidth per channel in $\sigma' = [\sigma, \cdots, 5\sigma]$. The required bandwidth per channel in $\sigma' = [\sigma, \cdots, 5\sigma]$ is about 70% larger than that of $\sigma' = [\sigma, \cdots, \sigma]$.

On the other hand, when the group size is 16, the required bandwidth per channel in $\sigma' = [\sigma, \cdots, 5\sigma]$ is about 1.8 GHz. The required bandwidth per channel in
Fig. 14  Group size and required bandwidth per channel with the reallocation probability 0.1 under the different traffic distribution.

Fig. 15  Group size and required bandwidth per channel with the reallocation probability 0.001 under the different traffic distribution.

Fig. 16  Group size and required bandwidth per channel with the reallocation probability 0.0001 under the different traffic distribution.

Fig. 17  Measurement granularity of link rate.

\[ \sigma' = [\sigma, \cdots, \sigma] \] is about 1.5 GHz. The required bandwidth per channel in \( \sigma' = [\sigma, \cdots, 5\sigma] \) about 20% larger than that of \( \sigma' = [\sigma, \cdots, \sigma] \).

The results with reallocation probability of 0.1, 0.001 and 0.0001 are showed in Figs. 14, 15, and 16, respectively. All parameters except the reallocation probabilities are the same as those of Fig. 13. We observe that the results of Figs. 14, 15, and 16 in terms of the dependencies of the group size and the range of \( \sigma' \) have the same tendency as those of Fig. 13. The reduction of required bandwidth per channel tends to increase as the reallocation probability that must be satisfied decreases.

5.2 Measurement Granularity

The performance of resource utilization is examined by changing measurement granularity of the link rate. Figure 17 shows the required bandwidth per channel for each group when the measurement granularity changes from 0.001 seconds to 10 seconds. Each channel utilizes the average 1.4 GHz traffic by CAIDA traces [17], and every channel is assumed to be the same traffic. The number of channels is assumed to be 16 for one link. The required bandwidth for each group converges in coarse-grained measurement, since coarse-grained measurement ignores the fine-grained behavior of the traffic. The gradient of required bandwidth per channel between 0.1 seconds and 1 seconds is smaller than that between 1 seconds and 10 seconds when the horizontal axis is a logarithmic scale, while the vertical axis is a linear scale. We use the semilogarithmic graph so that our results can be visualized clearly with various measurement granularities from 0.0001 to 10. If we use a linear scale for the horizontal axis, the gradient of required bandwidth per channel between 0.1 seconds and 1 seconds is larger than that between 1 seconds and 10 seconds. In other words, the second derivative of required bandwidth per channel is positive. In this context, fine-grained behavior means that traffic amount changes at a short period of time (for example 0.001 seconds or 0.01 seconds). The measurement granularity depends on a type of applications; real-time applications re-
quire fine-grained measurement. Fine-grained measurement is used to apply for real-time services, but it is also used for non-real-time services. As the measurement granularity becomes finer, the required bandwidth increases. The effect is strong when the group size $g$ is small. Consequently, fine-grained measurement receives resource reduction benefits from the grouping of channels.

5.3 Waiting Time of Services

We analyze a waiting time of services in two ways: numerical calculation and simulation.

5.3.1 Numerical Result

In order to analyze basic characteristics of waiting time of the proposed scheme, M/M/1 model is used to evaluate the queuing behavior of each group, since reallocations in different groups are performed independently. We adopt the following assumptions. Each group allows at most one reallocation at the same time, which corresponds to one server. Reallocation request arrivals follow a Poisson arrival process with the average arrival rate of $\lambda$. A reallocation time follows an exponential distribution with the average service rate of $\mu$. While one reallocation request is being served, other requests wait in a queue whose length is infinite.

Note that we consider a spectrum bandwidth management on the scenario of one optical fiber link in this paper. If we expand the one link model to a network-wide model, multiple servers for reallocations can be considered. If two channels are link-disjoint, they can be reallocated independently at the same time. Otherwise, their reallocations may be affected each other. The work in [22], [23] addressed spectrum reallocation to reduce the latency by considering parallelization.

Assuming that the number of channel per one link is $n$, Pure-elastic SA has the average service rate $\mu$ and the average arrival rate $\lambda$ because every channel belongs the same group. In Semi-elastic SA, the number of group is $n/2$. Since Semi-elastic SA configures a group with only the neighbor channel. Thus, the average service rate is $n/2$-times as often as Pure-elastic SA. The average arrival rate is $n/2$-part of Pure-elastic SA. In Grouped-elastic SA, the number of group is $n/g$, where $g$ is the group size. The average service rate in the grouped elastic are $n/g$-times as often as Pure-elastic SA. The average arrival rate is also the same; $n/g$-part of Pure-elastic SA. Besides, the average service utilization in the entire system is defined as $\rho = \lambda/\mu$. The following Table 1 summarizes the parameters for each scheme in M/M/1 model.

The number of connections per one link is assumed $n = 128$ by reference to dense WDM networks [19], [20]. In order to examine the effect of arrival rate, the average arrival rate is normalized by the average service rate in the system: $\mu = 1.0$ [s]. The average service utilization and the waiting time of services are shown in Fig. 18.

The plots whose group size is $g = 2$ shows Semi-elastic SA. The plots whose group size is $g = 128$ show Pure-elastic SA. Grouped-elastic SA corresponds every plot (from $g = 2$ to $g = 128$) thanks to design the group size to each service. The waiting time in Semi-elastic SA is smaller than those of Pure-elastic SA and Grouped-elastic SA. This is caused by no reallocation. In contrast, Pure-elastic SA reallocates channels frequently by minimizing the spaces between neighbor channels. Consequently, the waiting time increases explosively with an increase of service utilization.

A waiting time depends on a group size. Grouped-elastic SA is able to balance the between resource utilization and waiting time by configuring a group size for each service.

A human-tolerable waiting time is assumed to be one second by analogy of a connection setup delay of telephone networks: Integrated Services Digital Network (ISDN) circuit switching service [21]. Pure-elastic SA exceeds the tolerable waiting time when $\rho$ is higher than 0.5. Semi-elastic SA and Grouped-elastic SA ($g=8$ and $g=16$) do not exceed the tolerable waiting time when $\rho=0.9$. Grouped-elastic SA ($g=32$) satisfies the tolerable waiting time when $\rho$ is lower than 0.8.

5.3.2 Simulation

We assumed that there is an infinite frequency resource in the numerical results in Sect. 5.3.1. An infinite resource means that all arrival requests are accepted without any request blocking. In order to investigate a waiting time under the finite frequency resource with request blocking, we perform simulation for elastic spectrum allocation. A new request

| Table 1 Parameters of each spectrum allocation in M/M/1 model. |
|-----------------|----------------|-------------------|
|                 | Average service rate $\mu$ | Average arrival rate $\lambda$ | Average service utilization $\rho = \lambda/\mu$ |
| Pure-elastic SA | $\mu$            | $\lambda$           | $\frac{\lambda}{\mu}$ |
| Semi-elastic SA | $\frac{\mu}{2}$ | $\frac{\lambda}{2\mu}$ | $\frac{\mu(n/2)}{\mu(n/2)}$ |
| Grouped-elastic SA | $\frac{n}{g}\mu$ | $\frac{\lambda}{n/g}$ | $\frac{\lambda}{\mu(n/g)^2}$ |

Fig. 18 Average service utilization and waiting time of services.
is blocked if there is no available frequency resource for the request after waiting requests are served. The average service rate $\mu$ denotes that the average number of requests that are processed. The average service rate $\mu$ is obtained by,

$$\mu = \frac{1}{\bar{y}}.$$  \hfill (5)

where $\bar{y}$ [s/slot] is the Push-pull shifting time per slot, $\frac{1}{y}$ [slot/request] is the average number of shifting slots per request. $\bar{y}$ is expressed by,

$$\bar{y} = \int_{0}^{\infty} yg(y)dy,$$  \hfill (6)

where $g(y)$ is a distribution of the number of slots shifted by Push-pull. $y$ is the number of shifting slots per request. $g(y)$ means that $y$ slots are shifted after the channel has $x$ slots. In other words, $g(y)$ is the probability that slots width of the channel changes to $x+y$ or $x-y$ from $x$. $g(y)$ is represented by,

$$g(y) = \int_{-\infty}^{\infty} [f(x)f(x+y) + f(x)f(x-y)]dx,$$  \hfill (7)

where $f(x)$ is the PDF of requests of slot and $x$ is the requested slot. $f(x)$ is assumed to be a normal distribution with mean $m$ and standard deviation $\sigma$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(x-m)^2}{2\sigma^2}}$$  \hfill (8)

We get a distribution of the number of shifted slots as the following equation by using Eq. (7) as below (see Appendix A).

$$g(y) = \frac{1}{\sigma\sqrt{\pi}}e^{-\frac{y^2}{4\sigma^2}}$$  \hfill (9)

By using Eq. (6), we get the average number of shifting slots per request as below.

$$\bar{y} = \frac{2}{\sqrt{\pi}\sigma}$$  \hfill (10)

$\bar{y}$ is proportional to $\sigma$. Thus, the average number of slots shifted by Push-pull is determined by the standard deviation of the PDF of arrival bandwidth requests. When the range of requested slots is large, the average number of shifted slots is also large. By using Eq. (5), $\mu$ is expressed by,

$$\mu = \frac{\sqrt{\pi}}{2\rho\sigma}.$$  \hfill (11)

We set a several $\sigma$ and $\lambda$ in the simulation. Then we calculate the average service utilization $\rho = \lambda / \mu$. A waiting time under the finite frequency resource and the average service utilization is presented in Fig. 19. The waiting time with 300 slots is compared to that of 10000 (\textasciitilde infinite slots) slots. 1 slot has 12.5 GHz frequency bandwidth. In order to analyze the relationship between resource and waiting time, Pure-elastic SA is adopted for elastic spectrum allocation. Push-pull shifting time is assumed to be 500 ms/slot on the ground of a demonstration experiment [12], [13]. $\mu=10$ and $\sigma=5$ are set as parameters of the PDF of slot requests.

The waiting time in 300 slots are almost the same as that of 10000 slots in $\rho \leq 0.5$. When $\rho$ becomes higher than 0.5, the waiting time is saturated approximately at the value of 70 s. For Pure-Elastic SA in 300 slots, since the request blocking condition is often reached with $\rho > 0.5$, the request blocking occurs, while, for Pure-Elastic SA in 10000 slots, the request blocking condition, which is described at the begging of section 5.3.2, is rarely reached when $\rho$ increases. This causes the saturation of waiting time with $\rho$ for Pure-Elastic SA in 300 slots. The blocking probability increases when $\rho$ is higher than 0.5, while the waiting time remains at a certain value.

Compared with a human-tolerable waiting time one second, Pure-elastic SA in 10000 slots and in 300 slots exceeds the value at any $\rho$. This is because Pure-elastic SA depends on the Push-pull shifting time, which depends on tuning technologies of wavelength tunable lasers. In order to reduce the waiting time close to 1 second in Pure-elastic SA, the push-pull shifting time presented in [12], [13] needs to be reduced to be 5 ms/slot.

### 5.4 Channel Alignment and Margin Bandwidth

We compare the required bandwidths of the three policies via numerical analysis. We assume that all the traffics follow the normal distributions. Their mean values are set to 2.8 [GHz] and the standard deviations are set to be distributively from 200 to 1200 [MHz] with incremental step of 200, i.e., 200, 400, \ldots, 1200. Each margin bandwidth is computed to satisfy the reallocation probability lower than 0.01.

Figure 20 shows the margin bandwidth per channel with the three policies about channel alignment. The middle-max policy requires the smallest bandwidth of the three, while the
middle-min policy requires the largest one. The middle-max policy requires 2.3 \% smaller bandwidth than the middle-min policy. This is because the middle-max policy is less affected by large standard deviations than the other policies. The channels of the right and left edges have a greater impact on the margin bandwidth than the middle channels. The required margin bandwidth of the ascending-order policy lies between those of the middle-min and middle-max policies.

6. Conclusion

This paper proposed an elastic spectrum allocation scheme, called as Grouped-elastic SA, which groups channel in order to restrict the range of Push-pull reallocation. Based on an assessment of real traffic behavior, the proposed scheme determines the bandwidth margin between neighboring channels to prevent frequent reallocation. The feature of the proposed scheme is to manage both waiting time and bandwidth utilization rate by designing the suitable group size based on the constraint of waiting time of each service to meet the requirement. Performance evaluations showed that the proposed scheme reduces the required bandwidth per channel while satisfying according to waiting time constraint. We observed that the grouping in the proposed scheme is useful for resource utilization in the fine-grained measurement. Especially, the edge channels have a greater impact on the bandwidth utilization than the middle channels.

Acknowledgment

This work is supported by “ACTION Project” funded by the National Institute of Information and Communications Technology (NICT) Japan.

References

Appendix: Derivation of Eq. (9)

We derive $g(y)$ in Eq. (9) from Eqs. (7) and (8). First, we substitute Eq. (8) for Eq. (7).

$$g(y) = \frac{1}{2 \pi \sigma^2} \int_{-\infty}^{\infty} \left[ e^{-\frac{(x-m)^2}{2\sigma^2}} + e^{-\frac{(x-m+y-m)^2}{2\sigma^2}} \right] dx$$

(A.1)

We expand the exponent portions in Eq. (A.1) and perform completing the square.

$$g(y) = \frac{1}{2 \pi \sigma^2} \int_{-\infty}^{\infty} \left[ e^{-\frac{(x-\frac{y-2m}{2})^2}{\frac{4\sigma^2}{3}}} + e^{-\frac{(x-\frac{y+2m}{2})^2}{\frac{4\sigma^2}{3}}} \right] dx$$

(A.2)

Transformation of $u = x + \frac{y-2m}{2}$ and $v = x - \frac{y+2m}{2}$, where $dx = du, dv = dv$ is adopted in Eq. (A.2).

$$g(y) = \frac{1}{2 \pi \sigma^2} e^{\frac{y^2}{4\sigma^2}} \int_{-\infty}^{\infty} \left[ e^{-\frac{u^2}{2\sigma^2}} \right] du + \frac{1}{2 \pi \sigma^2} e^{\frac{y^2}{4\sigma^2}} \int_{-\infty}^{\infty} \left[ e^{-\frac{v^2}{2\sigma^2}} \right] dv$$

(A.3)

We use the following relationship.

$$\int_{0}^{\infty} e^{-a^2 x^2} dx = \frac{\sqrt{\pi}}{2a} (a > 0)$$

(A.4)

By substituting Eq. (A.4) for Eq. (A.3), we get Eq. (9).
Satoru Okamoto is a Project Professor of the Keio University, Kanagawa, Japan. He received his B.E., M.E. and Ph.D. degrees in electronics engineering from Hokkaido University, Hokkaido, Japan, in 1986, 1988 and 1994. In 1988, he joined Nippon Telegraph and Telephone Corporation (NTT), Japan, where he conducted research on ATM cross-connect system architectures, photonic switching systems, optical path network architectures, photonic network management systems, and photonic network control technologies. In 1999, he investigated the HIKARI router ("photonic MPLS router") and the MPLambdaS principle. He is now researching future IP + optical network technologies, and application over photonic network technologies. He has led several GMPLS-related interoperability trials in Japan, such as the Photonic Internet Lab (PIL), the Optical Internetworking Forum (OIF) Worldwide Interoperability Demo, and the Kei-han-na Interoperability Working Group. He is a vice co-chair of the Interoperability Working Group of the Kei-han-na Info-communication Open Laboratory. He is now promoting several research projects related to photonic networks. He has published over 70 peer-reviewed journal and transaction articles, written over 140 international conference papers, and been awarded 50 patents including 5 international patents. He received the Young Researchers’ Award and the Achievement Award in 1995 and 2000 respectively from the IEICE of Japan. He also received the IEICE/IEEE HPSR2002 Outstanding Paper Award, Certification of Appreciation ISO-CORE and PIL in 2008, IEICE Communications Society Best Paper Award and IEEE ISAS2011 Best Paper Award in 2011, and Certification of Appreciation for iPOP Conferences 2005–2014 in 2014. He was an associate editor of the IEICE Transactions on Communications (2006–2011), and was an associate editor of the Optical Express of the Optical Society of America (OSA) (2006–2012). He is an IEEE Fellow and an IEICE Senior Member.

Naoaki Yamanaka graduated from Keio University, Japan where he received B.E., M.E. and Ph.D. degrees in engineering in 1981, 1983 and 1991, respectively. In 1983 he joined Nippon Telegraph and Telephone Corporation’s (NTT’s) Communication Switching Laboratories, Tokyo Japan. He has been active in the development of ATM base backbone network and system including Tb/s electrical/optical backbone switching as NTT’s Distinguished Technical Member. He moved to Keio University in 2004. He is currently a Professor in Dept. of Information and Computer Science, Keio University, Japan, Vice Chair of Keio Leading-edge Laboratory of Science and Technology and chair of Photonic Internet Labs. He is an IEEE Fellow and an IEICE Fellow.