Title

Traffic-Adaptive Networking Solutions for Next-Generation Wide-Area Optical Networks

Author

Ahmed, Tanjila

Publication Date

2021

Traffic-Adaptive Networking Solutions for Next-Generation Wide-Area Optical Networks

By

Tanjila Ahmed

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Biswanath Mukherjee, Chair

Massimo Tornatore, Co-Chair

Khaled Abdel-Ghaffar

Committee in Charge

2021

To my family,

teachers,

and to all the people who have enriched my life.

Contents

List of Figures

List of Tables

Abstract

Traffic-Adaptive Networking Solutions for Next-Generation Wide-Area Optical Networks

As we continually churn out more and more information that must be transmitted over our networks, the networking challenges of handling this large-scale, highly-diverse, and varied-speed traffic opens exciting new research and development problems. More bandwidth and lower latency is required to accommodate the complex applications over wide-area networks. Although a complete traffic-adaptive network is still a long way to go, our networks are gradually incorporating changes through flexible network architectures such as Precision Time Synchronization Protocol (PTP), Elastic Optical Networks (EON), C+L bands expansion, Artificial Intelligence (AI), etc. This dissertation comprises of four contributions: i) high-precision time synchronization techniques for optical datacenter networks (Chapter 2); ii) dynamic resource allocation in mixed-grid optical networks (Chapter 3); iii) C+L bands upgrade strategies to sustain capacity crunch (Chapter 4); and iv) C to C+L bands upgrade with resource re-provisioning (Chapter 5). The dissertation concludes with a summary and future research directions (Chapter 6).

The main topics of this dissertation are the following:

1. A datacenter, which is a highly-distributed multiprocessing system, needs to keep accurate track of time across a large number of machines. Precise time synchronization is critical due to stringent requirements of time-critical applications such as real-time big-data analytics, high-performance computing, Internet of things (IoT) networks, and financial trading. To achieve this time accuracy, we consider Precision Time Protocol (PTP) to synchronize the server clocks. Zero overhead is maintained by using data traffic to carry the time messages instead of a separate control channel. We showed that microsecond level of time accuracy can be achieved and discussed the dependency of the accuracy on different traffic loads, traffic distributions, and packet lengths.

2. A rapid change in traffic type, volume, and dynamicity is presented by cloud-based services, datacenters, 5G, Internet of Everything (IoE), etc. Although introduced as a

promising solution to this change in 2008, EON technologies are not fully deployed yet. Rather a logical and gradual migration technique is adapted which investigates bottleneck points of the network and takes migration decision while keeping the rest of the fixed-grid network operational. Therefore, a co-existing fixed-grid and flex-grid (which can be called a "mixed-grid") is a cost-effective solution for current circumstances. However, this introduces new challenges by the interoperability issues between fixed and flex-grid technologies. We proposed a solution to a RSMA problem in a mixed-grid considering interoperability constraints. The solution proposes routes, spectrum, and modulation format to provision a dynamic, heterogeneous traffic on two US-wide network topologies ensuring maximum network throughput and minimum blocking.

3. As more traffic is put into the access network, the backbone network also needs to have higher capacity and better network planning. Although EON helped to maximize the available spectrum utilization in C band, the exploitation of bandwidth potential of single-mode fiber (SMF) can be achieved through opening up other spectrum bands. We investigate cost-efficient upgrade strategies for capacity enhancement in wide-area networks enabled by C+L bands. A multi-period strategy for upgrading network links from C band to C+L bands is proposed, ensuring physical-layer awareness, cost effectiveness, and less than 0.1% blocking. Results indicate that performance of an upgrade strategy depends on efficient selection of the sequence of links to be upgraded and on the time instant to upgrade, which are both topology- and traffic-dependent.

4. We study efficient allocation of resources during C to C+L bands network upgrade. After an upgrade, resource allocation may become sub-optimal, leading to lower utilization of spectrum resources causing requests blocking, early upgrade, and higher cost. Thus, we investigate pro-active re-provisioning of lightpaths after each upgrade for cost benefit. Our strategy locates highly-utilized links and upgrades them in batches. After each batch upgrade, existing traffic in C band is re-provisioned to L band. This re-provisioning frees up high-OSNR lightpaths in C band, leading to improved quality of future transmissions, delayed upgrades, and cost benefits. Results show that re-provisioning of a shorter lightpath provides the most cost-effective upgrade strategy.

Acknowledgments

I would like to express my gratitude to Professor Biswanath Mukherjee for providing me with the opportunity to pursue a PhD under his supervision. His guidance and insights have taught me important lessons for finding excellence in career and in life.

I acknowledge Professor Massimo Tornatore for being a patient co-supervisor and guide for all my research works. His detailed feedback have contributed significantly towards my research output. I would also like to thank Professor Khaled Abdel-Ghaffar for his guidance and support as my PhD qualifying exam chair and dissertation committee member.

I would like to thank Prof. Charles U. Martel for his feedback on my projects. I would also thank Prof. Zhi Ding for being part of my PhD qualifying exam committee.

I am immensely grateful to the Electrical and Computer Engineering Department for all their help. This dissertation would not be possible without their support. I would like to acknowledge National Science Foundation (NSF) (Grant No. 1716945) and Institute for Information communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (Grant no. B0117-16-1008) for funding our projects.

I would like to thank all my NetLab colleagues and peers for creating a friendly and creative environment which encourages open discussion on research. I thank all my collaborators for enriching my research, specially Dr. Kwangjoon Kim, Dr. Sifat Ferdousi, Dr. Abhijit Mitra, and Dr. Andrew Lord.

I would like to thank my father Sujauddin Ahmed, mother Rumana Emdad, sister Tasnuva Ahmed Raisa, and my husband Sabidur Rahman, whose unwavering support and love have made my journey easier.

I am lucky to learn from excellent educators throughout schools, college, and universities. I am grateful to all of them and will always cherish their teachings.

Chapter 1

Introduction

A traffic-adaptive networking solution leads to the customization of network operations according to the increasing unpredictability of traffic patterns. With the advent of cloud computing, social networking, wireless Internet access, and multimedia on demand, our current wide-area networks are being exposed to traffic demands with diverse throughput, speed, and QoS requirements. 5G, Wi-Fi 6, digitized spaces, SD-WAN, and artificial intelligence are some of the current technologies which enable flexibility to the network for better adaptation to the fast-changing traffic.

The advent of cloud computing allows content and service providers to 'migrate' content and computing resources from one location to another, based on where they need to be consumed. This creates substantial shifts in traffic patterns, and their delay requirement as sources and sinks of information can change quickly.

The increased mobility of users' contents present additional challenges. In the past, there was a clear relationship between the user and the users' location when accessing the network, as users' were physically tethered to the network. Today's radio access networks are increasingly capable of supporting high-bandwidth, low-latency applications, including streaming video; and a plethora of mobile devices allow people to consume content no matter where they are without any degradation.

Exceptional events such as sports events (e.g., football finals, Olympics, etc.) and other events such as political rallies, can generate large short-term demands between particular network nodes (i.e., telecommunication nodes serving the different venues of that event).

The wide-area network needs to handle this fast-changing traffic without any service degradation. Aggregation networks are static and are built based on knowing where the users are, where the content is stored, and where the applications are running. All of this is now fluid and dynamic, and hence the core transport network needs to provide flexibility in terms of ultra-reliable, ultra-low-latency, and high-bandwidth services. In contrast, the traditional network technologies are more suitable for static traffic and are not well adapted for ever-changing traffic patterns.

This chapter summarizes how the above-mentioned emerging challenges lead to the following research contributions.

1.1 Research Contributions

1.1.1 High-Precision Time-Synchronization Techniques for Optical Datacenter Networks

The first research contribution (Chapter 2) explores how an efficient time-synchronization technique helps to manage traffic flows in a distributed computing system such as a datacenter. As more organizations adopt cloud services, it is becoming more complex to maintain all the services with required QoS. To satisfy the latency constraint of this complex network, novel methods are needed for time synchronization between all computing nodes. As many of these cloud services are highly latency sensitive, the cloud operator has to ensure precise time synchronization between all compute nodes to ensure ultra low latency.

Time synchronization is an essential property for any distributed system. Network time protocol (NTP) [1] has been used for network time synchronization for several decades now. As network computing becomes more complex and the world becomes more interconnected, the need for more precise time synchronization has greatly increased. Network time protocol (NTP) [1] and inter-range instrumentation group (IRIG) [2] time code have been the most common protocols governing time transfer, but the later-developed IEEE 1588 Precise Time Protocol (PTP) [3] promises to revolutionize time synchronization by improving accuracy and reducing cost. PTP networks aim to achieve nanosecond- or even picosecond-level synchronization. While certain other precise synchronization protocols require significant investment in hardware and cabling, PTP makes highly-precise timekeeping possible using the most widely-deployed medium for network connectivity, namely Ethernet.

In Chapter 2, we describe and compare the state-of-the-art solutions for time synchronization. We also propose an implementation of PTP with zero overhead for an optical datacenter network. Most prior studies considered PTP over a wide-area network, and only a few of them considered a datacenter scenario without proposing any improvement to the algorithm. Our algorithm considers a controller as a master clock and the Top-of-the-Rack (ToR) switches as slave clocks. The controller sends protocol messages using the data traffic which results in no additional control messages. Our study demonstrates that, for a dynamic traffic load, average delay of synchronization remains within a few microseconds in each case. Our key contribution to this problem is a detailed survey on time synchronization and a zero-overhead microsecond accuracy solution for a datacenter network.

1.1.2 Dynamic Routing, Spectrum, and Modulation-Format Allocation in Mixed-Grid Optical Networks

The second research contribution (Chapter 3) proposes a dynamic route, spectrum, and modulation format allocation (RSMA) technique, ensuring maximum spectral utilization for co-existing fixed/flex-grid wide-area optical networks. Drastic changes in traffic patterns have created new challenges on the existing static network, which is gearing towards a more traffic-adaptive network. With recent efforts on migration from a fixed-grid wavelength-division multiplexing (WDM) network to a more flexible-grid network comes new planning and operational challenges for network operators.

A traditional fixed-grid network with its fixed resource-allocation policies cannot utilize available spectrum resources for dynamic and heterogeneous traffic demands, whereas a flexible network can modify the spectral efficiency (SE) of an optical channel to better exploit its resources when provisioning the related demand for service. To that end, it can increase SE to save spectral resources to create larger total network capacity, or it can lower SE to extend the transmission reach, if needed. Utilizing benefits of a flex-grid network in presence of existing fixed-grid technologies is crucial until a complete migration. However, until complete migration is done, network operators need to ensure interoperability with maximum resource utilization in a co-existing fixed/flex-grid network. We analyze this interoperability challenge in detail with our proposed solution techniques.

Most prior studies propose RSMA solutions on a fully-flexible network, called "Elastic Optical Network", without addressing any intermediate stages. Although a few of them consider a mixed-grid network, they do not provide in-depth analysis on evolving network operations. We analyze the migration strategies proposed by recent studies and make a realistic assumption on intermediate mixed-grid networks. Various traffic profiles are taken into account to maintain heterogeneity of the load. To maintain a smooth transition into different type of grid, logical assumptions are made regarding spectrum occupation and modulation-format adaptation. Comparisons with baseline RSMA strategies are evaluated in terms of their bandwidth blocking ratio (BBR) and spectrum utilization. We also observe the effect of changing the number of fixed vs. flex-grid nodes in the network, assuming an ongoing migration. Our proposed solution routes heterogeneous traffic with lowest spectrum allocation and negligible cost. Illustrative results show significant improvement compared to baseline solution techniques in terms of BBR.

1.1.3 C+L Bands Upgrade Strategies to Sustain Traffic Growth in Optical Backbone Networks

A long-term capacity scaling in optical backbone networks is necessary to accommodate today's fast-growing traffic. Studies have debated over multiple solutions, among which spatial-division multiplexing (SDM) [4] and low-loss spectrum optical bands (L, O, E, S, U bands) of single-mode fiber (SMF) [5] have emerged as potential solutions. SDM consists in transmitting over multi-fibers (MF), multi-core fiber (MCF), and multi-mode fiber (MMF). MF technology requires rolling out new optical fiber infrastructure either by installing new fibers or lighting up existing dark fibers, both being expensive options. MCF technology require deployment of novel type of fibers with complex multiple-input-multiple-output (MIMO) transceivers, which are not yet commercially available. MMF technology has limited deployment due to inter-modal dispersion for longer distance communication. Therefore, expansion of the operating band of existing SMFs beyond C band [6], [7] is a nearer-term viable solution to handle the capacity crunch. This solution maximizes return on investment of already-deployed optical infrastructure. A gradual progression from C band to L, O, E, S, U bands (multi-band (MB)) is envisioned to be a longer-term solution as technology matures for all bands. Submarine networks already have active C+L bands systems to extend cable life-time [8].

Our work in Chapter 4 provides a detailed analysis on different multi-period upgrade strategies from C to C+L bands, and evaluates their performance on capacity enhancement in optical backbone networks. We propose upgrade strategies for two types of traffic: predictable and unpredictable. Predictable traffic provides information on exact arrival times of future connection requests whereas unpredictable traffic does not have this information. For a given traffic matrix and network topology, traffic predictability results in the most cost-efficient upgrade. Although the case of unpredictable traffic is more challenging to manage, our proposed upgrade strategy for unpredictable traffic achieves comparable results as predictable traffic. We also explored different upgrade batches (i.e., number of links that can be upgraded at the same time) for both traffic types. Different number of batches result in different cost of upgrade. For predictable traffic, as connection requests and their arrival times are known, the links to be exhausted at a future time can be predicted. Upgrading links before their exhaustion times avoids blocking. In contrast, for unpredictable traffic, connection request sequence and their arrival times are unknown. Therefore, we propose an upgrade strategy which analyzes the traffic matrix and topology, to devise the correct link sequence and times to upgrade.

1.1.4 C to C+L Bands Upgrade with Resource Re-provisioning in Optical Backbone Networks

Chapter 5 explores efficient allocation of resources during upgrade of network links from C to C+L bands for a wide-area optical network. After an upgrade, resource allocation may become sub-optimal, leading to lower utilization of spectrum resources. Such inefficient spectrum utilization can block future requests and require early upgrade, which leads to higher cost. Thus, we investigate pro-active re-provisioning of lightpaths to C+L bands after each upgrade for cost benefit. Prior works show benefits of lightpath re-provisioning for restorability of optical mesh networks [9, 10], network capacity maximization [11], service chaining in optical metro networks [12], etc. But, to the best of our knowledge, our work is the first to study the benefits of lightpath re-provisioning during upgrade to C+L bands.

Our strategy locates highly-utilized links and upgrades them in batches. After each batch upgrade, existing traffic in C band is re-provisioned to L band using various methods. This re-provisioning frees up high-OSNR lightpaths in C band, leading to improved quality of future transmissions, delayed upgrades, and cost benefits. Results show that re-provisioning of a shorter lightpath provides the most cost-effective upgrade strategy.

1.2 Organization

This dissertation is organized as follows. Chapter 2 presents a detailed survey on time synchronization and a zero-overhead microsecond-accuracy solution for a packet-switched optical datacenter network architecture. This work has been published in Photonic Network Communications, Aug. 2018 [13].

Chapter 3 investigates the dynamic routing, spectrum, and modulation format assignment problem in a mixed-grid optical network, which provisions routing, spectrum allocation, and modulation format assignment techniques for dynamic, heterogeneous traffic, ensuring maximum spectrum utilization and minimum blocking. A paper documenting this study was published in IEEE/OSA Journal of Optical Communication and Networking 2020 [14], after presentation at IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Dec. 2018 [15].

Chapter 4 investigates cost-efficient upgrade strategies for capacity enhancement in optical backbone networks enabled by C+L bands optical line systems. A multi-period strategy for upgrading network links from C band to C+L bands is proposed, ensuring physical-layer awareness, cost-effectiveness, and less than 0.1% blocking. This contribution is under review by the IEEE/OSA Journal of Optical Communications and Networking [16].

Chapter 5 investigates resource re-provisioning during network upgrade from C to C+L bands by optimizing resource allocation and postponing upgrade cost. Results show that our proposed method exploiting re-provisioning shorter lightpaths to L band leads to a more cost-effective upgrade. This contribution was submitted to Optical Fiber Communications Conference (OFC) [17].

Chapter 6 concludes this dissertation, discussing important future research directions.

Chapter 2

High-Precision Time Synchronization Techniques for Optical Datacenter Networks

2.1 Introduction

Time synchronization is a major requirement for a distributed computing system. Inside a datacenter, the network interconnection faces several engineering challenges to support a large number of highly-interdependent and homogenous nodes in a relatively-small physical space. With emergence of cloud computing and high-speed networks, computing systems need even more precise time synchronization than before. Applications such as big-data analytics, online gaming, smart-grid, IoT networks, and financial trading highly rely on time synchronization accuracy. Ideally, highly-accurate time synchronization guarantees that all the clocks in a system have the same time information at any time instance.

Today, Network Time Protocol (NTP) [1] is a widely-implemented synchronization protocol throughout the Internet and in datacenters. Despite its ease of implementation, low cost, and efficiency in synchronizing a distributed system, NTP falls short in terms of accuracy for many of today's real-time applications. Recently, a more-accurate protocol, called Precision Time Protocol (PTP), is being considered for these applications [3]. PTP was initially designed for financial trading, industrial automation, power systems, and telecommunication applications. Although PTP's basic communication protocol is similar to NTP's, PTP differs mainly for its precise hardware-assisted time recording of any event (e.g., message reception/transmission), called timestamping, and additional clocks deployed to account for any error which might affect timestamps. No additional hardware is required for NTP, while PTP requires PTP-enabled hardware. Although PTP needs hardware support, its nanosecond accuracy is crucial for some applications compared to NTP's millisecond accuracy.

To the best of our knowledge, no previous work has comprehensively surveyed time-synchronization mechanisms in optical datacenters. In this study, we describe and compare the state-of-the-art solutions for time synchronization. We also propose an implementation of PTP with zero-overhead for an optical datacenter network (ODCN), and justify our claim through simulation results. Our simulation considers a PSON datacenter with a centralized controller [18]. The controller is assumed to have an accurate clock and an array of time counters, which keep track of synchronization of Top-of-the-Rack (ToR) switches. Using PTP time messages, the controller sends latest time information to ToR switches. As the ToR switches are directly connected to servers, servers get synchronized as soon as their ToR switch is synchronized. Moreover, our method exchanges protocol messages using data traffic, i.e., incoming/outgoing packets to/from ToR switches. Thus, it can synchronize clocks without involving any control channel. This feature reduces protocol overhead (zero-overhead method). We tested with three traffic distributions—Pareto, lognormal, and uniform—to find accuracy of our proposed method. Results indicate that, as traffic load increases, average delay per successful synchronization decreases for each distribution. Our study demonstrates that average delay of synchronization remains within a few microseconds in each case. Our key contribution to this problem is a detailed survey on time synchronization and a zero-overhead microsecond-accuracy solution for PSON architecture.

The rest of the chapter is organized as follows. We provide a brief introduction on time synchronization in distributed systems in Section 2.2, which includes definition, necessity, types, and comparison among state-of-the-art techniques. In Section 2.3, we focus on PTP, which we implemented in our simulation. Various aspects of PTP including types of clocks it uses, its mechanism, and its accuracy are discussed here. In Section 2.4, we review the most relevant prior works on network time synchronization. In Section 2.5, we introduce a specific datacenter architecture for which we propose a zero-overhead implementation of PTP, and discuss the time accuracy of this approach through simulations. Section 2.6 concludes this study, with suggestions for improvements.

2.2 Time Synchronization

2.2.1 What and Why?

Time (or, equivalently, clock) synchronization is an important problem that coordinates independent clocks running on different system processes or different systems [19]. Computer clocks are usually made of inexpensive oscillator circuits or battery-backed quartz crystals. As crystal properties change due to heat, power, and aging (a vibration range for these oscillators are provided on their data sheets), no two crystals are ensured to be oscillating at the same exact frequency and time. These limitations make one oscillator run marginally faster than the other. Typically, drift rate of a computer clock is 0.6 seconds per week [20]. Replacing the built-in inexpensive clocks of computers with an expensive one is not feasible for large systems [21]. Therefore, a robust and efficient method to synchronize clocks of a distributed system is needed.

2.2.2 Types of Synchronization

Synchronization of two clocks can be referred to time, frequency, or phase; and it can be applied to any one or to all three types. While frequency and phase synchronizations mean correctness of clock-generation rate (zero skew and zero drift respectively), time synchronization means being at the same position (zero offset) on time axis. In frequency-synchronized systems, the significant instants occur at the same rate for all synchronized nodes [22]; they do not necessarily occur at the same time. For phase-synchronized systems, the significant instants not only happen at same rate but also they happen in unison [22]. However, in both cases, time information can be incorrect so they are not necessarily time synchronized.

Two major distributed time-synchronization schemes are categorized according to “continuity” and “discontinuity” of time over the network. Continuous-time (CT) approach interprets transmission start and finish times continuously over the network. A high-precision clock is necessary to provide time information consistently. In contrast, discrete-time (DT) approach, characterized by time discontinuity, divides the time into slots of constant duration. Utilization depends on time-slot dimensions and propagation-delay deviations, whereas CT approach only depends on a highly-accurate clock-synchronization mechanism [23]. For these reasons, CT approach is considered more suitable for today's high-speed networking. We have used CT approach in this study.

2.2.3 Importance of Time Synchronization

Synchronized clocks have become more crucial than before, due to emergence of new services including high-frequency trading (HFT), smart grid, smart home, autonomous vehicle system, IoT, and cloud computing. Below are some examples.

HFT requires to reduce latency of financial market data to allow trading firms to have advance market knowledge as fast as possible. The latency for real-time market data to travel over cables, switches, and routers matters with time scale of one millisecond, and accurate coordinated universal time (UTC) delivery is important [24]. Similarly, in smart-grid networks, time synchronization is a necessity to allow various parts of the grid to connect or disconnect without disruption. In IoT networks, sensor data from distributed networks need to be collected in real time to perform a particular application. In a software-defined network (SDN), high-precision controller clocks help to achieve accurate network updates with low packet loss [25].

Similarly, a datacenter, being a distributed system, requires accurate clocks within all its machines. For today's traffic-intensive, high-speed datacenters running time-critical applications, a time synchronization of higher accuracy than NTP is essential. Cisco Global Cloud Index, 2015-2020, forecasts that, by 2020, about 77% of the total traffic of datacenter will remain within the datacenter [26]. Big data is a significant contributor to this huge traffic. Moreover, synchronized clocks with 100 ns precision will allow packet-level scheduling of minimum-sized packets at a finer granularity, which can minimize congestion in rack-scale systems [27].

2.2.4 Comparison of State-of-the-Art Time-Synchronization Techniques

We discussed about the high-precision reference clock required for CT approach in Section 2.2.2. Atomic clocks are considered to be the most accurate reference clocks. These clocks use electronic transition frequency in microwave, optical, or ultraviolet region of the electromagnetic spectrum of atoms as a frequency standard for their oscillators [28]. National standard agencies tightly synchronize these clocks to produce an accurate global time. This time is then sent to Global Positioning System (GPS) or similar systems. GPS radio signals bring back the time information to GPS clocks which can be used as reference clocks. NTP, GPS time synchronization, and PTP are three well-known CT approaches which use these standard clocks. Below, we discuss these three CT-based synchronization techniques.

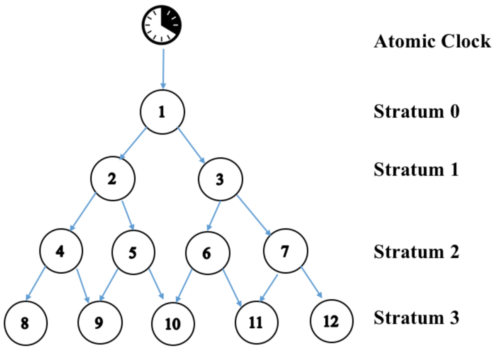


Figure 2.1: NTP mechanism.

Being one of the oldest time-synchronization protocols, NTP was developed to synchronize packet-switched networks with millisecond-level precision. It uses User Datagram Protocol (UDP) to exchange time messages. A hierarchical tree topology is constructed (Fig. 2.1) where the top-level timeserver (stratum-0) is synchronized with a standard clock and passes time information to the next layer (stratum-1). Here, stratum denotes the distance from the reference clock. The more the distance, the higher is the stratum value. Timeserver at stratum-1 distributes the time information to the following layer of timeservers (stratum-2). As time information is distributed in multiple layers, loss of accuracy is inevitable. Moreover, each stratum level adds a delay which aggregates till the furthest stratum. Therefore, NTP is not suitable for today's delay-sensitive traffic demands (which require accuracy between microsecond to nanosecond level).

GPS-based time synchronization ideally is the best solution to achieve nanosecond accuracy [29]. GPS satellites have atomic clock information. Therefore, GPS receivers receive accurate time information. This information is distributed to all the connected devices (Fig. 2.2) from the timeserver. To have GPS-based time synchronization, each server in a datacenter requires a GPS signal receiver. However, GPS receivers and extra cables make it both expensive and non-scalable. Moreover, GPS signals are difficult to receive inside

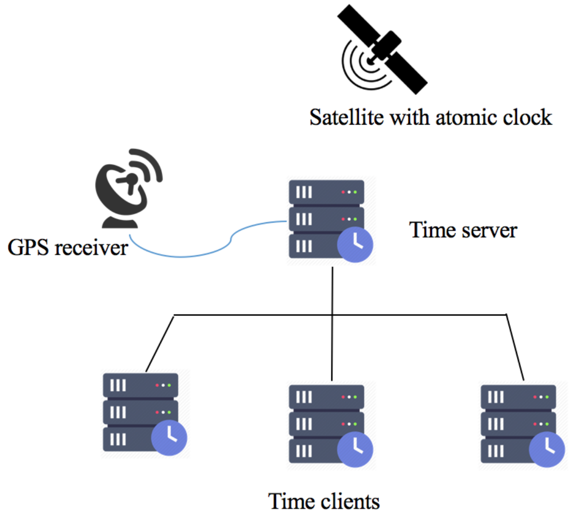


Figure 2.2: GPS time synchronization.

datacenters. So, this is not attractive even though it may give highest accuracy.

2.3 Precision Time Protocol

2.3.1 PTP Clock Types

PTP has a hierarchical master-slave architecture which can contain one or more communication media (network segments) with multiple clocks. An algorithm called Best Master Clock (BMC) is used to select the most-accurate clock in the network. This selected clock is called the grandmaster clock; usually, it has a good oscillator with standard time. The grandmaster clock is usually connected to an external GPS or atomic clock to receive correct time information. A master clock is selected among each network segment which is primarily synchronized by the grandmaster clock. Other clocks in the network synchronize to the grandmaster/master clock, and are called slave clocks. Usually, slave clocks do not redistribute the time to other clocks. Both the master and slave clocks are end devices on a network as opposed to switch or router. Usually, a simplified PTP system only has one network segment with one master clock which is the grandmaster clock.

Fig. 2.3 shows locations of the PTP clocks. PTP has two additional clocks which help to circumvent any additional delays in queues. Worst queuing delay occurs in switches/routers. Therefore, IEEE 1588-2008 introduced two types of switches/routers to mitigate these delays. One of them is a boundary clock, located at an intermediate node between a grandmaster

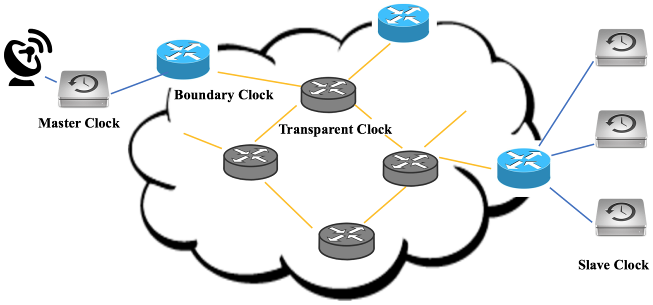


Figure 2.3: PTP clock types.

and its slave clocks of a network segment. It has one port in slave slate, and all the others in master slate. It is synchronized through grandmaster using slave slate, and synchronizes the downstream slave clocks using the master slate. In addition, a transparent clock is located inside a network segment. It modifies the time information as it passes through this switch/router. The modification is done to incorporate the queuing delay which occurred due to passing through the device.

2.3.2 PTP Algorithm

PTP starts with identifying master and slave clocks using the Best Master Clock (BMC) algorithm. Once the hierarchy has been established, considering the grandmaster clock as the root, other clocks get time information by exchanging request-response messages with it. Table 2.1 shows the basic protocol messages exchanged between these clocks along with their size and functionalities.

Table 2.1: Basic messages used in PTP synchronization [30].

Message	Length (Bit)	Function
Sync	352	Timestamped when sent by Master and received by Slave
Delay_Req	352	Timestamped when sent by Slave and received by Master
Delay_Resp	432	Includes Master Delay_Req timestamp
Follow_Up	352	Timestamped with Master's Sync message information

Other management and signaling messages are used for proper functionality. Messages are of two types: event and general message. Event messages (Sync, Delay_Req, etc.) are time critical whereas general messages (Follow_Up, Delay_Resp, etc.) are not. Fig. 2.4 shows the sequence in which all these messages are sent and received.

1. Once the hierarchy is established, master clock sends Sync messages to slave clocks with its timestamp, t_{m1}. This can be done in two ways, i.e., using a one-step or a two-step method. In the one-step method, time information is read from master clock and instantly transmitted to slave clock in the Sync message. For the two-step method, an additional Follow_Up message is sent with same time information as Sync message. Two-step system is used to eliminate the need to precisely read the timestamp and insert it into the protocol message simultaneously. This makes the Follow_Up message not time critical. Not all master clocks can read the timestamp and send it through Sync message instantly; they use Follow_Up message to send the timestamp.

2. A slave clock, upon receiving this Sync message, notes the time of reception as t_{s1}.

3. Delay_Req messages are sent timestamped by the slave to the master as t_{s2} and received by the master at t_{m2}.

4. Finally, master clock sends Delay_Resp message timestamped at t_{m2} to the slave clock. Slave clock calculates its offset and synchronizes its clock with master's time information.

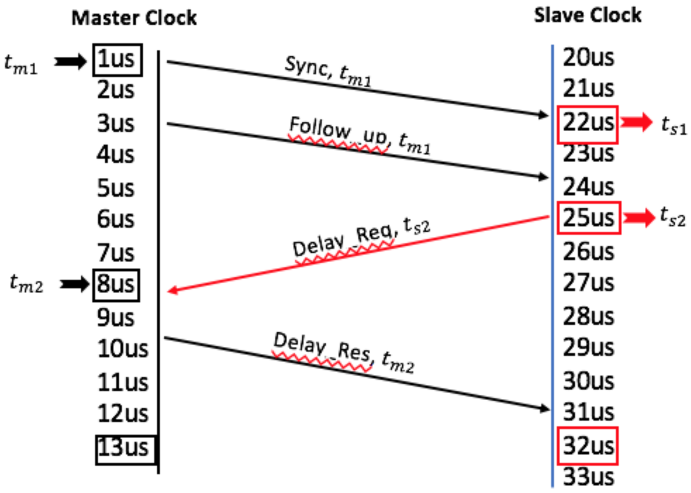


Figure 2.4: PTP synchronization mechanism example.

Clock offset (time difference between master and slave clocks) is calculated as follows:

$$Clock\_offset = t_{s1} - t_{m1} - avgPathDelay \tag{2.1}$$

where IEEE 1588-2008 assumes that the forward and reverse path delays between the master and slave clocks are always symmetrical. The average path delay is calculated as follows (numbers refer to the example in Fig. 2.4):

$$avgPathDelay = ((t_{s1} - t_{m1}) + (t_{m2} - t_{s2}))/2 \tag{2.2}$$

Hence,
$avgPathDelay = ((22 - 1) + (8 - 25))/2 = 2\mu s$ and
$Clock\_offset = t_{s1} - t_{m1} - avgPathDelay = 22 - 1 - 2 = 19\mu s$

This means the slave clock is ahead of the master clock by 19 microseconds. So, the slave clock needs to be adjusted to 32-19 = 13 microseconds. Now, the slave clock is synchronized with the master clock.

2.3.3 What Makes PTP Accurate?

Although NTP and PTP rely on similar concepts, they differ in implementation. First, PTP is more accurate mainly due to its hardware-assisted timestamping which makes the timestamping close to actual receive/transmit ports of the reference clock. Special integrated circuits are available to help the hardware timestamping such as National Semiconductor DP83640 [30]. Second, PTP has more than one clock to aid its synchronization operation such as boundary clock, transparent clock, etc. Boundary clock helps to connect multiple network segments, and transparent clock manages queuing or processing delay. Third, NTP generates a hierarchical tree, where time information is redistributed at each level which decreases the accuracy and adds delay to the system. In contrast, in PTP, a grandmaster clock synchronizes all slave clocks and no additional delays are experienced due to redistribution.

2.4 Survey on Related Work

Time synchronization in datacenter networks (DCN) is crucial for recent low-latency applications. There are studies based on the implementation of PTP in wide-area networks (WANs).

However, only few of these studies focus on a DCN. In this section, we review relevant prior works on time synchronization. Refs. [23][31][32] discuss datacenter time synchronization, which closely relates to our problem, while Refs. [33]-[36] explore PTP implementations in general. Table 2.2 summarizes these studies.

Table 2.2: Summary of state-of-the-art time-synchronization solutions.

Ref.	Objective	Algorithm/Simulation model	Results
[23]	Study and compare time-synchronization aspect of transmission scheduling in OGDCN.	Authors measure network throughput and average delay to compare CT and DT schemes in a simulation environment.	CT approach appears to be more promising than DT approach for OGDCN in terms of propagation delay variance and packet-length.
[31]	Implement a datacenter network with centralized arbiter having control over packet latency.	Time-slot allocation, path assignment, and replication strategies to overcome failures.	Results show that, throughput penalty is small, queueing delays reduced.
[32]	Implement a zero-overhead, decentralized datacenter time-synchronization protocol.	DTP measures one-way delay between two peer nodes. These nodes periodically exchange time information.	Precision from DTP in the network is about 25 nanoseconds for directly-connected nodes and about 150 nanoseconds.
[33]	PTP-based OMNeT++ model, investigates the synchronization accuracy.	PTP++ is proposed which has PTP nodes, hardware and software clocks.	To improve PTP accuracy, new techniques are introduced.
[34]	Analyze the performance of the proposed PTP clock model.	A model of PTP clock based on OMNeT++ framework is proposed.	Performance of the proposed clock model is evaluated.
[35]	Overview on LibPTP, which is an OMNeT++ based PTP simulation framework.	Performance evaluation of LibPLN and LibPTP is done by testing different PTP features.	Shows that LibPLN and LibPTP provide a way to gain insight in the domains of clock noise.
[36]	Implemented PTP on power-system networks to achieve sub-microsecond accuracy.	Authors tested six different network designs and measured synchronization error for each design.	Accuracy of PTP depends on device-specific characteristics and network design.

Refs. [23][31][32] focus on accuracy requirement, zero-overhead transmission, and CT approach of time synchronization in DCN. Ref. [23] studies time synchronization in an optically-groomed datacenter network (OGDCN). The authors compare CT- and DT-based synchronization for scheduling in terms of link utilization of each scheme. Results show that, although CT approach needs an accurate clock, it is better than DT because of its non-dependence on propagation delay and packet-length-distribution variance.

Fastpass [31] introduces a centralized datacenter architecture where a controller determines the time and associated path on which each packet should be transmitted. This centralized arbiter schedules packet flows to reduce queuing delay. To schedule packet transmission with high precision, Fastpass uses PTP to achieve a "Zero Queue" datacenter network. Although PTP is mentioned as a synchronization protocol, not much implementation details on how to achieve scheduling with zero queue are provided.

Ref. [32] introduces a new synchronization protocol, called Datacenter Time Protocol (DTP), which uses modified control block messages to exchange protocol messages. It provides a decentralized clock synchronization with nanosecond accuracy. The authors discuss drawbacks of NTP and PTP in terms of network jitter, packet buffering, scheduling limitations, network stack overheads, etc. However, DTP requires the physical layer to be modified such as new switching chip, or NIC (network interface card), is required, which comes with additional cost. Moreover, DTP cannot be deployed on routers or network devices with multiple line cards without compromising accuracy.

Moving to more generic studies for PTP implementations in a WAN, we note that OMNET++ simulator, for its per-packet discrete-event nature, has been often used for PTP evaluation. Ref. [33] gives a PTP simulation model for OMNeT++ framework to investigate the synchronization accuracy of PTP under different network load conditions. Simulation results show that synchronization accuracy improves when using prioritized quality of service (QoS).

In [34], the authors focus on modeling PTP clocks. They conclude with two observations: (1) hardware timestamps minimize synchronization errors and (2) clock-frequency-skew correction suppresses errors further, to the range of less than a hundred nanoseconds. This work gives a general idea of how to adjust the time error of a clock using a simple scenario of two clocks.

Ref. [35] shows detailed implementation of most of the PTP features in OMNeT++ with a goal to provide a tool for PTP design-space exploration. To capture a realistic scenario, it includes noise properties in the clock model. This model includes noise generation, estimation, and cancellation features, and analyzes the choice of synchronization intervals, impact of path asymmetry, and daisy chaining of clocks. The authors designed and implemented portable library for PTP called LibPTP and LibPLN which includes Power-Law noise.

A practical application scenario is analyzed in [36]. Here, PTP is applied to a power-system network to achieve sub-microsecond accuracy. The authors analyze various configurations of the master-slave devices and switches. Gradually, power-system devices are introduced to inject traffic in the system. This study compares frequency of time error occurring between master and slave clocks for each of these configurations. There are some studies on analog synchronization [37] of C-RoFN (Cloud-based Radio-over-optical Fiber Network) architecture using software-defined networking. However, they do not consider time synchronization for the network.

2.5 Time Synchronization in PSON Datacenter

2.5.1 PSON Architecture

In PSON (Fig. 2.5) datacenter architecture [18], a group of servers is connected to a single top-of-rack (ToR) switch, and n such ToR switches are interconnected by a core switching network.

Each ToR switch connects to an ingress/egress module which contains electrical framer and de-framer, respectively. In these modules, framer/de-framer does the wrapping/unwrapping of Ethernet packets into photonic frames. Framers consist of n Virtual Output Queues (VoQ) corresponding to each ToR switch. Each framer wraps the outgoing Ethernet packets from a connected ToR switch and stores it in the corresponding destination-specific VoQ. In contrast, each deframer unwraps the incoming data packets for its connected ToR switch. An optical transmitter uses a fast tunable-wavelength light source to transmit photonic frames from the ingress module to the connected core network. Each ingress module is connected with a 1xm space switch to connect with m arrayed waveguide grating routers

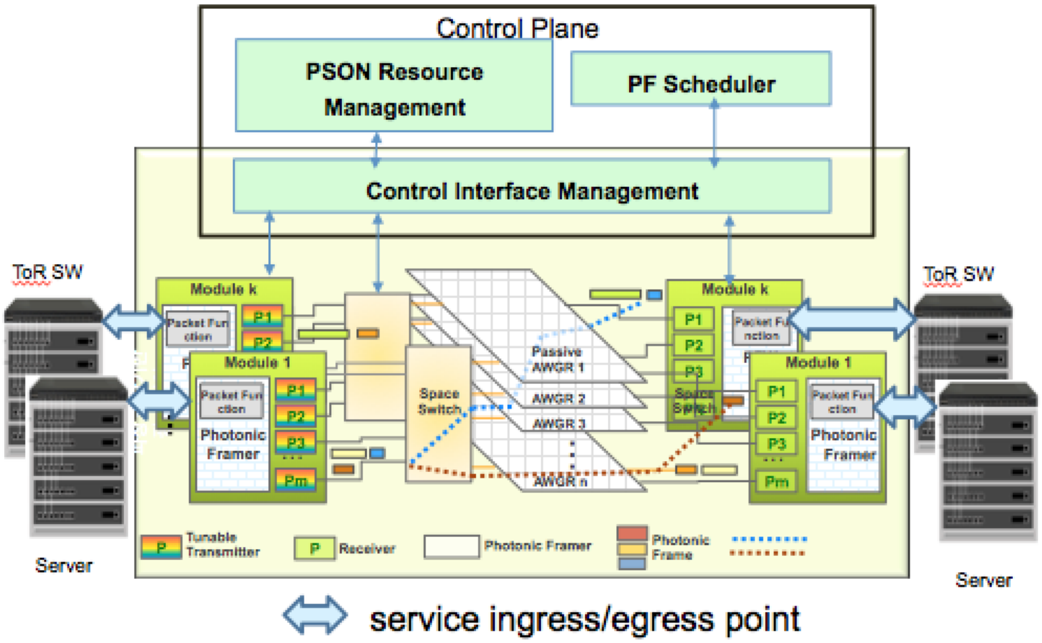
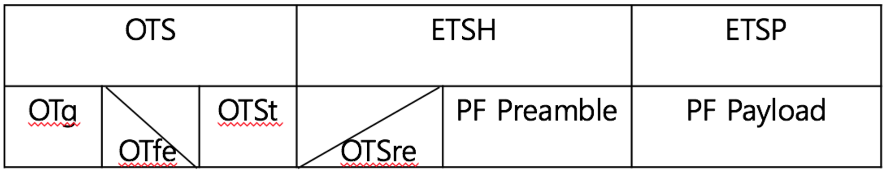


Figure 2.5: PSON architecture.



OTS: Optical Time Slot

ETS: Electrical Time Slot

ETSH: ETS Header

ETSP: ETS Payload

PF: Photonic Frame

OTg: Optical guard time

OTfe: Optical Falling Edge time

OTSt: Optical wavelength Switching Time

OTSre: Optical Rising Edge time

Figure 2.6: PSON photonic frame structure.

(AWGRs), and an AWGR exclusively delivers photonic frames to n/m egress modules.

Ethernet/IP packets generated by the servers are collected by the connected ToR switches and sent to connected ingress modules. In an ingress module, a packet is wrapped (to the payload of photonic frame shown in Fig. 2.6 by a framer according to its destination address and placed into its respective VoQ. A centralized controller decides which photonic packet from a VoQ can be transmitted into the core network. When an ingress is allocated with a transmission grant by the control plane, it sends a photonic frame to the associated space switch in the next time slot, and the photonic frame is switched by a certain AWGR which connects to the destined egress module, where inverse operations are performed so that packets can be delivered to destination servers.

In PSON, if time information of servers, ToR switches, controller and space switches, etc. are not the same, then accurate packet transmission would not be possible. For example, as a controller sends transmission grants to ingress/egress modules, these modules should be in sync with the controller clocks to be able to execute the transmission at the right time; otherwise, a delay between clocks of different modules and the controller can create additional delay to the transmission time.

2.5.2 Proposed Synchronization Mechanism for PSON

In this study, we propose and evaluate a mechanism for time synchronization in an optical datacenter using PTP. Given a PSON datacenter architecture, various traffic models, PTP-enabled ToR switches, and controller with reference time, our aim is to synchronize the ToR switches without using any separate control channel to send time information/messages to the ToR switches. So, we are constrained to use the data plane to send the timing messages and still perform better than NTP level of accuracy. Our proposed method ensures that the PTP-enabled ToRs are synchronized within a few microseconds. Data packets carry the timestamps to the clocks to be synchronized.

2.5.3 Methodology

2.5.3.1 Assumptions

To achieve zero-overhead and microsecond accuracy, we assume that:

- PTP runs the BMC algorithm (see Section 2.3.2) to find the most accurate clock in the system. However, we assume the controller to have the most accurate clock, so BMC is not required in our simulation.

- Controller has time counters for each ToR switch, to keep track of when a ToR switch needs synchronization. A resynchronization interval time (explained in Section 2.5.3.2) is preselected. When resynchronization interval is over for a ToR switch, controller will send Sync message to start synchronizing it.

- In PSON architecture, a group of servers are connected to a single ToR switch (Fig. 2.5). We assume that, as the servers are physically connected to ToR switches, if the ToR switch gets synchronized, so do the servers connected to it.
- All ToR switches and controller clocks are PTP-enabled.

2.5.3.2 Mechanism and Implementation

As shown in Fig. 2.4, PTP clocks are synchronized using a request-response mechanism. The resynchronization interval time can be changed as per level of accuracy needed. For short interval time, network will be frequently synchronized, creating more workload in the system. In contrast, for long intervals, the network will have less precise synchronization and less workload.

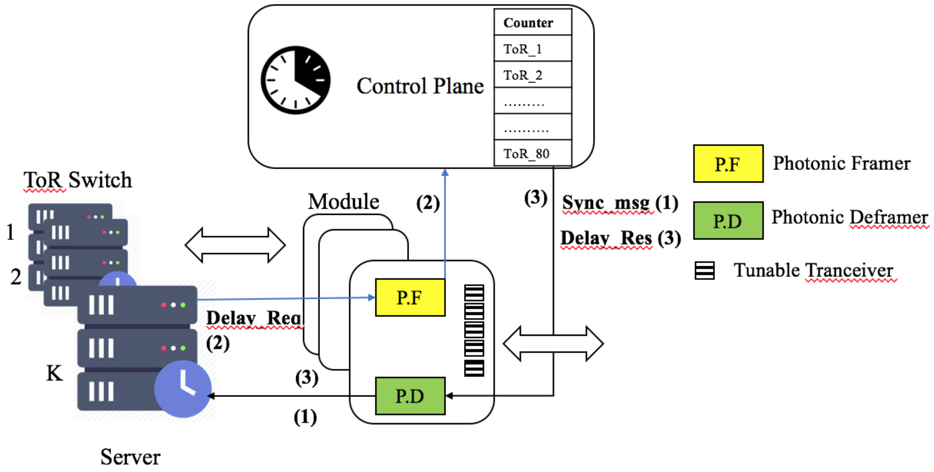


Figure 2.7: PTP mechanism on PSON architecture.

When the resynchronization interval of a counter expires, synchronization process for that particular ToR switch initializes. The goal is to send accurate time information from controller to that ToR switch using PTP standard messages. Fig. 2.7 shows the PTP mechanism on PSON architecture. It only captures the transmission path of PTP messages such as controller to framer/deframer to ToR switches. Data packet transmissions from the rest of the network are not shown. In Fig. 2.7, controller generates a Sync_msg including current reference time and sends it to the photonic deframer, connected to that ToR switch. Deframer unwraps the photonic packet, attaches the Sync_msg to the available data packet to be sent, and transmits it to the destination ToR switch. This process can also be referred as piggy-backing. This operation is indicated as path (1) in Fig. 2.7. If there are no data packets available for that ToR switch in the deframer, the controller has to wait and send only when data becomes available. This waiting time depends on traffic load of the datacenter network.

On receiving part, a PTP-enabled ToR switch receives Sync_msg and records its time of reception. It replies via a Delay_Req message to the egress module with its recorded timestamp of Sync_msg reception. This Delay_Req message is carried by the outgoing data packet from that ToR switch to its connected framer indicated as path (2). Again, ToR switch has to wait until it can piggy-back this message on outgoing data packets. Controller reads this request message (Delay_Req) from the framer and records the time of its reception. Finally, the controller sends this timestamp to deframer via a Delay_Resp message. Accordingly, this Delay_Resp message is sent to ToR switch via piggy-backing on data packet and ends the synchronization process shown in path (3). Delay_Resp message is not time sensitive, so synchronization is done when the controller gets the Delay_Resp message.

ToR switch calculates clock offset using Eqns. (2.1) and (2.2) to adjust its clock. Once the ToR sends Delay_Req message, resynchronization-interval counter is reset. After one successful synchronization process, controller waits for the resynchronization interval to reach the set value. This process continues simultaneously for all ToR switches. The steps needed to conduct this process are summarized in Algorithm.

2.5.4 Simulation Results

In this section, we evaluate the performance of our proposed mechanism. Our simulation results show that all the values are within ±5% with 95% confidence interval. As our simulation is traffic dependent, we considered three different traffic distributions: lognormal (which captures inter-arrival times of datacenter traffic [38]), Pareto (to capture the bursty nature of datacenter traffic [18][39]), and uniform. The considered PSON architecture consists of 80 TOR switches and 80 modules with framers and de-framers. Maximum length of PTP messages is 432 bits [30]. Data packet length is assumed to be between 200-1400 bytes [40]. Links have 10 Gbps bandwidth, and traffic load over such links is varied from 20% to 100% [41]. Resynchronization interval is set to 1 second [42].

Performance metric of our simulation is the "time error" which depends on traffic load,

Algorithm 1 PTP time synchronization through in-band transmission

```
1: Input: Packet size, traffic load, link bandwidth of datacenter, number of ToR switches, resynchronization-interval time;
2: Output: Time error;
3: Initialize counter times for all ToR switches (limit time at resynchronization-interval time);
4: Create a list of candidate ToR switches which have counter time equal to resynchronization-interval time (that means this ToR switch needs to be synchronized);
5: if ToR switch $\epsilon$ candidate list then
6:     wait_sync_time $\leftarrow$ wait time for data packets in deframer to send ‘sync’ message to
7:     candidate ToR switches with packets generated from a specific traffic distribution;
8:     wait_delay_req_time $\leftarrow$ wait time for data packets in framer to send ‘delay request’
9:     message to controller with packets generated from the same traffic distribution;
10:    counter time of that ToR $\leftarrow$ zero;
11:    Time error $\leftarrow$ wait_sync_time + wait_delay_req_time;
12:    Candidate list $\leftarrow$ candidate list – synced TOR;
13: else
14:    counter time of that ToR $\leftarrow$ counter time++;
15: end if
16: Repeat from Step 4;
```

packet length, and traffic distribution. Time error indicates the total time to synchronize server clocks connected to ToR switches with the controller clock. This can be quantified as follows:

Time error = wait time for data packets in deframer to send ‘sync’ message to candidate ToR switches + wait time for data packets in framer to send ‘delay request’ message to controller.

According to our assumption, as the traffic load increases, time to synchronize a ToR switch should decrease, thus time error decreases. Similarly, if the traffic load decreases, time error increases. This parameter can also imply the accuracy of the synchronization method that we are using. For PTP, precision should be from a few microseconds to sub-nanoseconds.

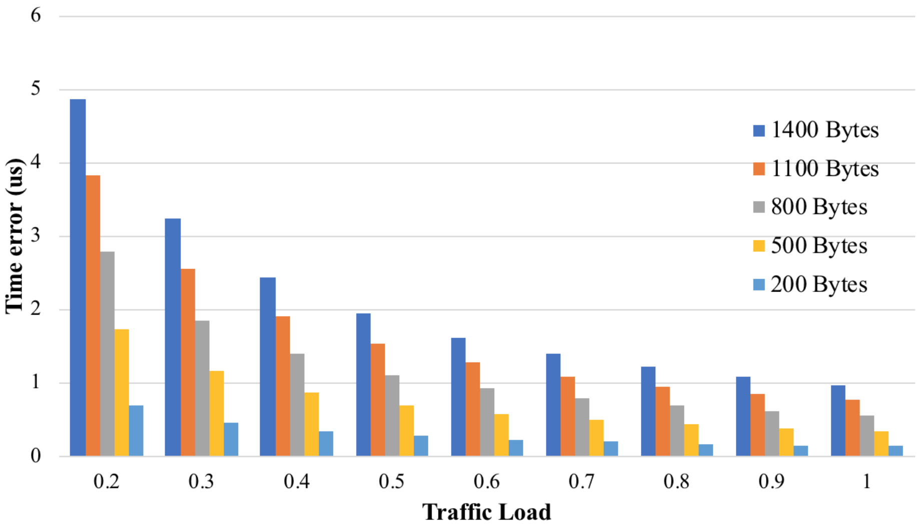


Figure 2.8: Time error depending on traffic load and packet length for lognormal traffic distribution.

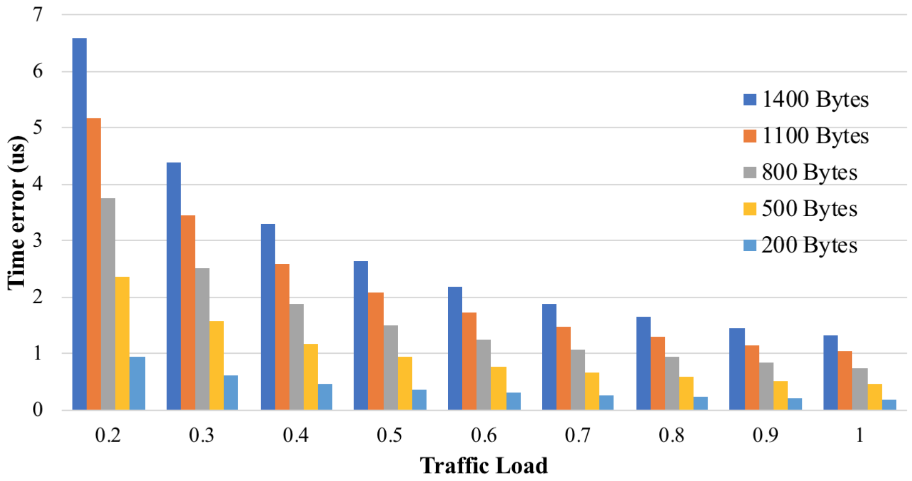


Figure 2.9: Time error depending on traffic load and packet length for Pareto traffic distribution.

Figs. 2.8-2.10 show the effect of increasing traffic load and packet length on time error. Fig. 2.8 shows the results for a lognormally-distributed traffic. With increasing load, the number of generated packets increases which helps to synchronize faster (lower time error). Time error follows a decreasing pattern with increasing traffic load (from 0.2 to 1). On the contrary, time error decreases with decreasing packet length (1400-200 bytes), as PTP messages can be transferred faster with smaller packets. For a datacenter experiencing load between 0.3 and 0.6, average time error is between 1.8 and 0.9 μs, which is better than sub-millisecond accuracy of NTP [21].

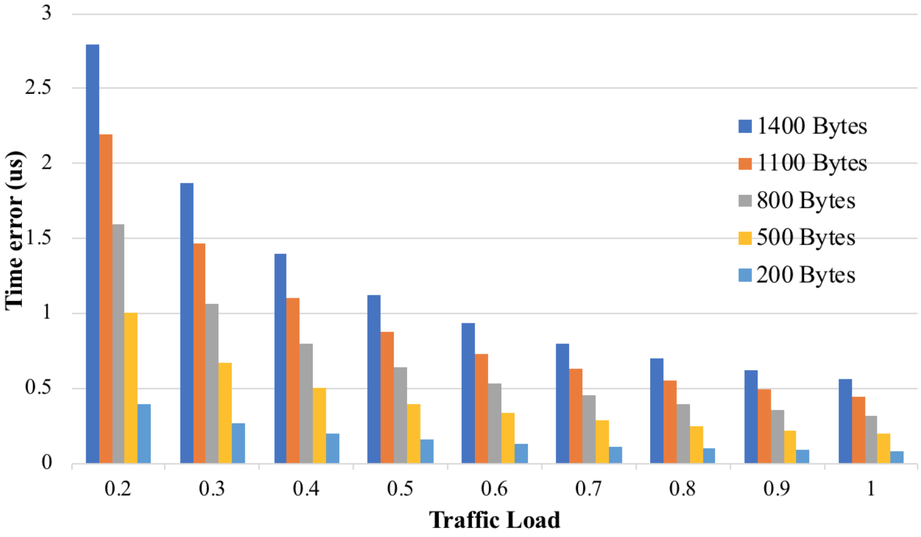


Figure 2.10: Time error depending on traffic load and packet length for uniform traffic distribution.

Fig. 2.9 shows time error for Pareto-distributed traffic with varying packet size and load. As the time error of the proposed mechanism depends on load and number of packets, it is expected that, as number of packets increases, synchronization speed will improve. For example, in case of 200-byte packets and 50% load, synchronization delay is 0.37 μs only. As datacenter traffic is expected to handle a large number of mice traffic (smaller packets), lower delay in this case is an advantage of using the proposed algorithm.

Fig. 2.10 shows the time error for uniformly-distributed traffic. Results follow the same decreasing time error pattern as the traffic load increases. As the packet length decreases, time error decreases as well.

Comparisons of time errors found for three different traffic distributions are shown in Figs. 2.11 and 2.12 for two types of packet lengths. Through this analysis, we can comment on which traffic distribution gives the least time error for our implementation. First observation considers only the mice traffic (200-byte packets). As expected, lognormal traffic faces lesser delay than Pareto traffic, as Pareto traffic has a more bursty nature. Uniformly-distributed traffic performs the best with lowest time error because of its non-bursty nature. As the inter-arrival time of each packet is same in case of uniform distribution, average delay of time synchronization is the lowest. In Fig. 2.12, we consider elephant traffic (1400-byte packets). For this case, we plot again all three examined distributions. Again, lognormal traffic is characterized by lower time error than Pareto traffic, and uniform observes lowest time error

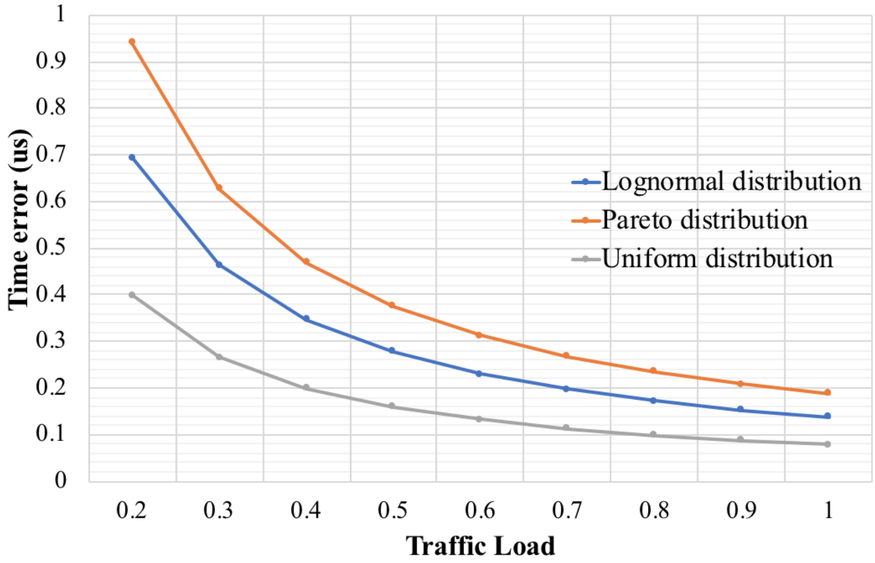


Figure 2.11: Comparison of synchronization time error among lognormal, Pareto, and uniformly-distributed mice traffic (200-byte packets).

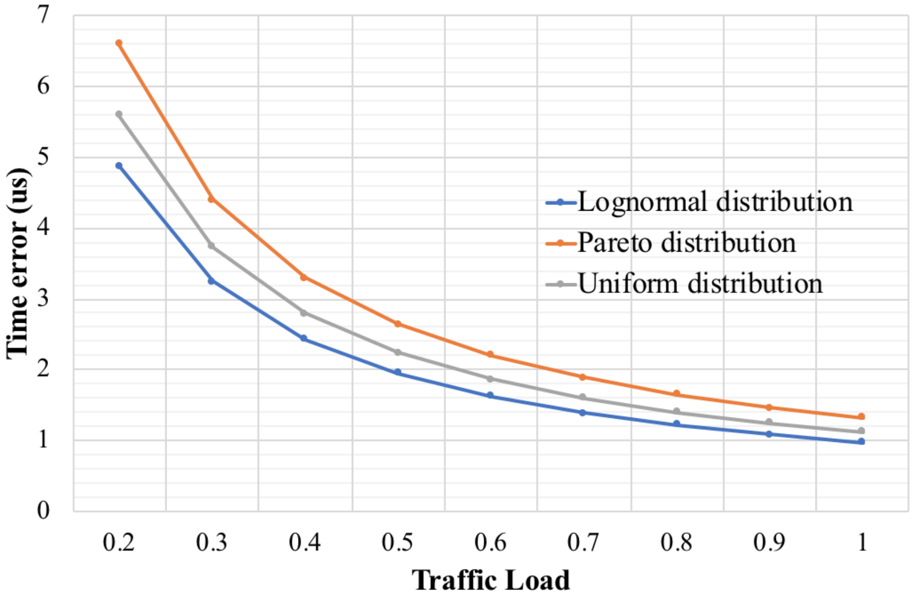


Figure 2.12: Comparison of synchronization time error among lognormal, Pareto, and uniformly-distributed elephant traffic (1400-byte packets).

among all three distributions. For all our simulation results, the confidence intervals with the error bars are too close, so we did not report them. For example, in case of load 1 in Fig. 2.12, the result is 1.1185 with high error bar +0.00154492 and low error bar $-1.8x10^{-07}$.

2.6 Conclusion

Time synchronization is a very important property for datacenters to serve ultra-low-latency applications. We reviewed synchronization techniques applicable for distributed networks, and discussed the ones suited for DCN. We tested the performance of PTP for an in-band transmission system which provides zero-overhead by appending synchronization information to data packets. Data packets were generated from three different traffic distributions. We achieved the desired level of accuracy (microseconds) for all three cases. Our technique maintained zero-overhead by piggy-backing time information on data packets. However, slight reduction in accuracy in low-load conditions were observed. Nonetheless, the advantage achieved by zero-overhead and in-band transmission compensates for this slightly lower accuracy. For PTP-enabled DCNs, PTP-supported hardware and software are already introduced in the market. For example, Cisco's PTP-enabled Nexus 3100 platform and Nexus 9000 switches for datacenters [43] and Arista's 7150S datacenter switch segment offer complete hardware support for PTP boundary and transparent clock functions [44].

Increasing usage of datacenter network is also contributing to increased traffic volume in backbone optical networks. To keep pace, backbone optical networks need to increase capacity. In the next chapters (Chapters 3, 4, and 5), we explore methods and technologies to increase capacity for next-generation optical networks.

Chapter 3

Dynamic Resource Allocation in Mixed-Grid Optical Networks

3.1 Introduction

Massive increase in Global IP traffic volume (with compound annual growth rate (CAGR) of 26%) has been forecast in the Cisco global visual networking index (VNI) for 2017-2022 [45], where the CAGR is dominated primarily by video traffic (82% of IP traffic), while most of this traffic is generated from wireless and mobile devices (71% of IP traffic). This chapter proposes novel methods to support these ever-increasing traffic by utilizing mixed-grid optical networks.

Existing wavelength-division-multiplexing (WDM) backbone networks based on a fixed-grid subdivision of the spectrum need to evolve to carry such heterogeneous, high-volume, and high-bit-rate traffic, while ensuring high resource utilization. Elastic optical networks (EON), thanks to its flexible assignment of spectrum resources and its adaptive transponder technologies, offers an effective solution to serve this evolving traffic. However, given the large amount of currently-operational fixed-grid networks, in some cases (e.g., when the network is single-vendor, or in future scenarios where equipment disaggregation is supported), migration towards a flex-grid EON can happen through a gradual process. In fact, gradual (or brown-field) migration towards a flex-grid infrastructure provides an opportunity to optimize cost of deployment, minimizes wastage of previously-deployed WDM equipment, and prevents disruption in regular network operations.

Various studies show how to migrate from fixed to flex-grid network [46] [47] [48]. They suggest to localize bottleneck nodes/links in the network as a initial point to start upgrading them to operate on flex-grid. During this upgrade, some existing switching nodes operating with fixed spectrum slots of 50 GHz will be substituted with optical architectures capable to manage variable-width optical channels consisting of multiples of basic frequency slots at 12.5 or 25 GHz.

These flex-grid nodes are typically equipped with wavelength-selective switches (WSSs) [49] [50], and symbol-rate adaptable transponders [51] to offer flexibility. This will result in lightpaths operating at different bit rates (e.g., 10, 40, 100, 200, 400 Gb/s) that can be allocated over different channel widths using different modulation formats, e.g., BPSK, QPSK, 8QAM, 16QAM, 32QAM, etc. Also, larger bit rates (e.g., 400 Gb/s, 1 Tb/s) can be achieved by using super-channels.

Although these technologies are available now, they are not widely deployed yet. Gradual migration strategies have been proposed [46] [47] [48] using higher bit rates and advanced modulation formats only for specific connections by upgrading nodes based on some node-merit metric such as: (1) upgrade nodes that generate/carry most traffic first, (2) upgrade nodes that generate/carry high traffic variation first, (3) upgrade nodes that generate/carry most low/high-bandwidth traffic first, or (4) upgrade nodes with highest nodal degree first.

During this migration process, fixed-grid and flex-grid technologies would need to inter-operate, introducing new planning and operational challenges for network operators. Most prior works either have studied migration strategies or proposed resource allocation in a EON. Few works address the operational challenges in a mixed-grid environment, e.g., migration-aware routing [52], static routing and spectrum allocation techniques [53], dynamic routing (shortest path) and spectrum allocation (first-fit) in a pre- and post-migration scenario [47], modulation-format and spectrum allocation in EON [54], etc. Some recent works propose solutions such as split-spectrum or sub-band virtual concatenation [55] [56] where traffic demand is split and transmitted via multiple optical sub-channels for better flexibility. However, these works either use standard routing and spectrum-allocation techniques in a mixed-grid network or propose solutions with higher complexity and cost. In our work, a dynamic routing, spectrum, and modulation format allocation (RSMA) is proposed which exploits diverse modulation formats and provides higher spectral efficiency while maintaining complexity close to standard techniques.

Clearly, a mixed-grid network raises new challenges which require modifications to traditional network operations. Our work focuses on resource allocation while ensuring seamless adaptation to network heterogeneity. We propose an algorithm Mixed-grid-aware Dynamic Resource Allocation (MDRA) which includes Spectrum-Efficient Dynamic Route Allocation (SEDRA) algorithm, Reusable Spectrum Allocation First (RSAF), and a distance-adaptive modulation-format allocation. Performance evaluation is done with respect to bandwidth blocking over two large network topologies with various traffic profiles. Our results depict 50% reduction in bandwidth blocking ratio (BBR) for practical load values while using our solution compared to benchmark techniques.

The rest of this chapter is organized as follows. In Section 3.2, related works are reviewed. In Section 3.3, lightpath provisioning challenges in a mixed-grid network are introduced and represented through examples. Section 3.4 formally states the RSMA problem in a mixed-grid network, and describes a possible strategy to solve the problem based on existing approaches for EON. Section 3.5 provides our proposed algorithm. Section 3.6 introduces the performance evaluation metrics which have been considered as well as numerical results with explanations. Section 3.7 concludes the chapter.

3.2 Related Works

Most prior studies propose RSMA solutions on a fully-flexible EON without addressing any intermediate migration stages. Ref. [52] discusses brown-field migration from fixed grid to flexible grid in optical networks. The authors proposed a migration-aware routing (MAR) algorithm for resource provisioning. Their proposed algorithm first calculates the probability of each node in the network to be upgraded to a flex-grid node. Based on these probabilities, it routes lightpaths to avoid any interruption due to any future migration. Ref. [53] focuses on routing and spectrum allocation (RSA) in a mixed-grid network considering post-migration scenario. The authors proposed integer linear programming (ILP) formulations along with static heuristic algorithms to minimize spectrum utilization.

Ref. [47] presented a comparison on various migration strategies. It adopted traditional

k-shortest path and first-fit technique for route and spectrum allocation, respectively, to compare the performance of these migration strategies. Ref. [57] evaluated the impact on network capacity of deploying a flex-grid solution over a network which is partially loaded with fixed-grid channels. The authors proposed several migration strategies from fixed-grid to flex-grid networks.

Considering modulation-format adaptability in EON, Ref. [54] proposed distance-adaptive spectrum allocation, where minimum spectral resource is adaptively allocated to make better use of the resource. The study considered both modulation formats and optical filter width to determine the necessary spectral resources to be allocated to an optical path. It adopted a traditional fixed-alternate routing and a first-fit spectrum assignment algorithm to provision lightpaths. Most studies on modulation-format adaptability of flex-grid are limited to pure flex-grid networks (not mixed-grid scenario).

In [55] [56], authors introduce the concept of sub-band virtual concatenation (VCAT) in mixed-grid optical network, improving spectrum utilization. They propose sub-band VCAT to enable lightpath connections to be established between different types of nodes, and allow the traffic demand to be split and transmitted via multiple optical sub-channels for better flexibility and greater spectral efficiency. They proposed mixed integer linear program models and heuristic algorithm based on spectrum window planes for RSA optimization. Although VCAT can help with better spectrum utilization, the guard band required between neighboring split sub-channels may waste fiber spectra and also increase the number of transponders and signal regenerators used.

In the preliminary version of this work [15], we proposed a novel routing algorithm, called SEDRA, in a mixed-grid network. It provisions routes for dynamic, heterogeneous traffic, ensuring maximum spectrum utilization and minimum blocking in a mixed-grid network. We evaluated BBR for both Uniform and Poisson distribution of traffic arrivals.

In the extended version of this work reported in this chapter, various additional contributions are included. First, we propose a new resource-allocation algorithm MDRA which includes SEDRA, as well as RSAF and distance-adaptive modulation-format allocation, the latter being the most significant addition. For this algorithm, we evaluated various baseline routing, spectrum allocation, and modulation-format allocation strategies. Second, we evaluated the performance of MDRA on a denser network (24-node USnet topology). Third, we investigated the effect of different numbers of flex-grid nodes in the network. Fourth, we made detailed comparison of MDRA with baseline strategies by investigating metrices such as Average Number of Hops per Path and Percentage of Requests Blocked.

3.3 Challenges due to Migration Strategies

Migration strategy depends on network topology, traffic distribution, locality of traffic, network bottlenecks, traffic profiles, etc. It also depends on the type (fixed/flex) of neighboring nodes. If a fixed-grid node with a flex-grid neighbor node is being upgraded, a high-rate super-channel can be set up between them. Also, a higher modulation format can be adopted on the route. Therefore, studies [46] [47] have recommended migration through creating multiple independently-growing flex-grid islands. A flex-grid island is defined as a subset of network nodes with flexible-grid technology. Multiple such islands are required to grow, based on the traffic distribution in the network. Fig. 3.1 shows an example of flex-grid islands. Here, we consider a US-wide backbone network where flex-grid islands are being formed with nodes located in east and west coast areas, where the traffic is assumed to be higher than in the rest of the network.

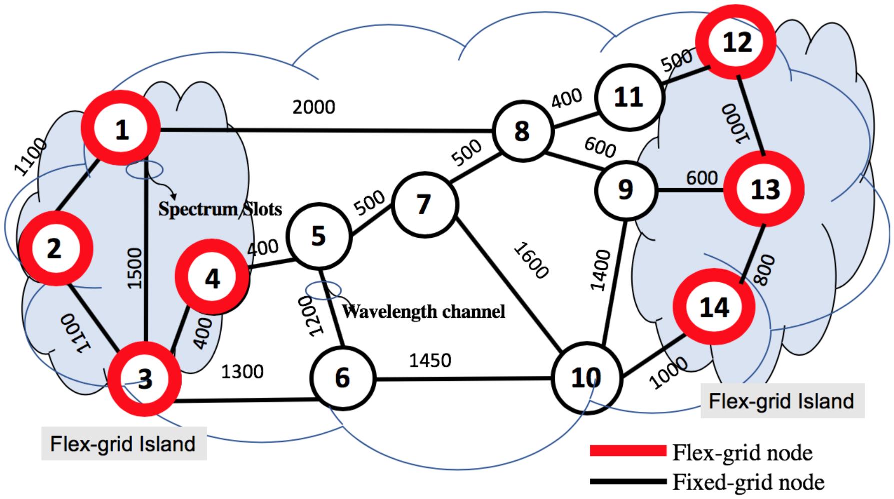


Figure 3.1: Co-existing fixed/flex-grid in 14-node NSFnet topology.

The next two sub-sections explain (through an example) spectrum assignment in a mixed-grid network with and without distance-adaptive modulation formats, respectively. A fixed modulation format of Dual Polarization Quadrature Phase Shift Keying (DP-QPSK) is assumed for non-distance-adaptive case, irrespective of the distance between source and destination. Figs. 3.2 and 3.3 demonstrate different cases of spectrum assignment in mixed-grid scenarios. We assume that fixed-grid and flex-grid have a basic frequency slice of 50 GHz and 12.5 GHz, respectively. Note that wavelength continuity and contiguity constraints must be respected at node B.

3.3.1 Spectrum Assignment in a Mixed-Grid Network without Distance-Adaptive Modulation

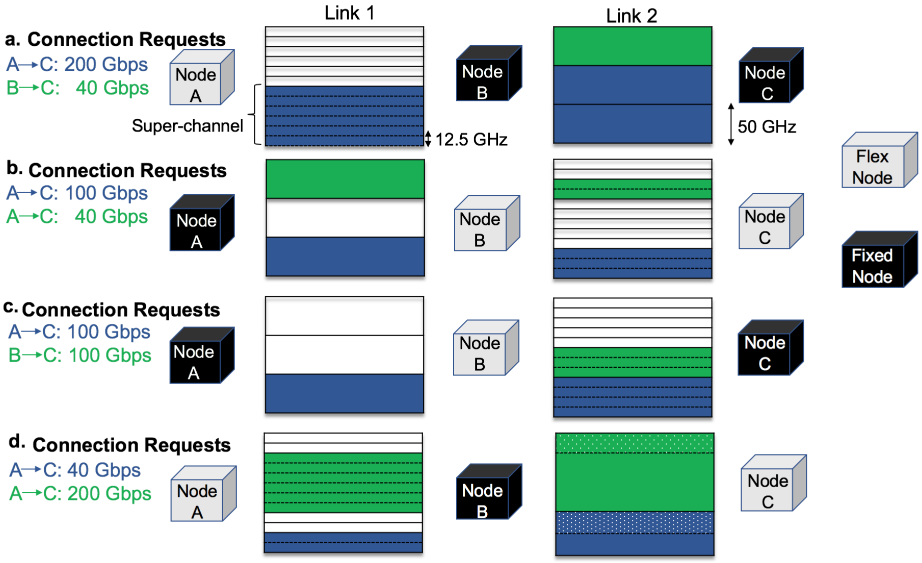


Figure 3.2: Spectrum assignment in different mixed-grid scenarios. (a) connection requests A → C: 200 Gbps, B → C: 40 Gbps; (b) connection requests A → C: 100 Gbps, A → C: 40 Gbps, (c) connection requests A → C: 100 Gbps, B → C: 100 Gbps; and (d) connection requests A → C: 40 Gbps, A → C: 200 Gbps.

Fig. 3.2 shows part of a mixed-grid network where lightpaths traverse both flex-grid and fixed-grid links. Spectrum occupation of signals with various bit rates are reported in Table 1 [47]. There are three nodes and two links in this example. We assume that a link has 150 GHz capacity, where fixed-grid and flex-grid links would have three wavelength channels and 12 frequency slots, respectively. In Fig. 3.2(a), a lightpath request of 200 Gbps, originating at a flex-grid node (node A), terminates into a fixed-grid island of two fixed-grid nodes (nodes B and C). According to Table 3.1, link 1 needs six slots (75 GHz) whereas link 2, being in a fixed-grid island, needs two lightpaths of 50 GHz (total 100 GHz) to allocate the same

Table 3.1: Spectrum occupation for various bit rates.

Traffic Demand (Gb/s)	Fixed-Grid		Flex-Grid	
	Bandwidth (GHz)	#Wave-lengths	Bandwidth Gap (GHz)	# Slots
40	50	1	25	2
100	50	1	37.5	3
200	100	2	75	6
400	200	4	150	12

200 Gbps connection request (hence an O/E/O conversion is required at node B). In link 1, flex-grid uses super-channel; on the contrary, in link 2, limitation of fixed-grid to allocate higher bit rates in a single channel is observed. The second connection request of 40 Gbps, which originates from node B, stays in a fixed-grid island, and is assigned a 50 GHz slot. For the 200 Gbps traffic request at each link, 25 GHz (2x12.5 GHz) of spectrum is saved in the flex-grid link compared to the fixed-grid link.

On the contrary, in Fig. 3.2(b), lightpaths originating from a fixed-grid node are ending in a flex-grid island. 100 Gbps connection requests from node A to C occupied 50 GHz in link 1 and 37.5 GHz in link 2. Now, a 40 Gbps connection request is assigned between nodes A to C, with 50 GHz and 25 GHz occupation in links 1 and 2, respectively. Here, for the same connection requests, flex-grid link occupies 37.5 GHz less spectrum in total than the fixed-grid link.

Figs. 3.2(c) and 3.2(d) represent scenarios where lightpaths originate and terminate at same type of islands but traverse through a different one. Lightpaths should maintain transparency while traversing through different islands. In Fig. 3.2(c), a 100 Gbps connection request is set up between nodes A to C. This request originates and terminates into fixed-grid island, traversing through a flex-grid island. To maintain transparency, a lightpath starts with 50 GHz on link 1 and comes out from the flex-grid island with the same 50 GHz (4 slots) signal. On the contrary, the second connection request of 100 Gbps originating from node B occupies only (3 slots) 37.5 GHz instead. Similarly, in Fig. 3.2(d), a 40 Gbps connection request occupies 2 slots (25 GHz) in link 1 but takes up 50 GHz channel in link

2, where the signal occupies only 25 GHz in the channel, while the rest of the spectrum is not used (blue with dotted white). Same happens with the 200 Gbps connection request.

3.3.2 Spectrum Assignment in a Mixed-Grid Network with Distance-Adaptive Modulation

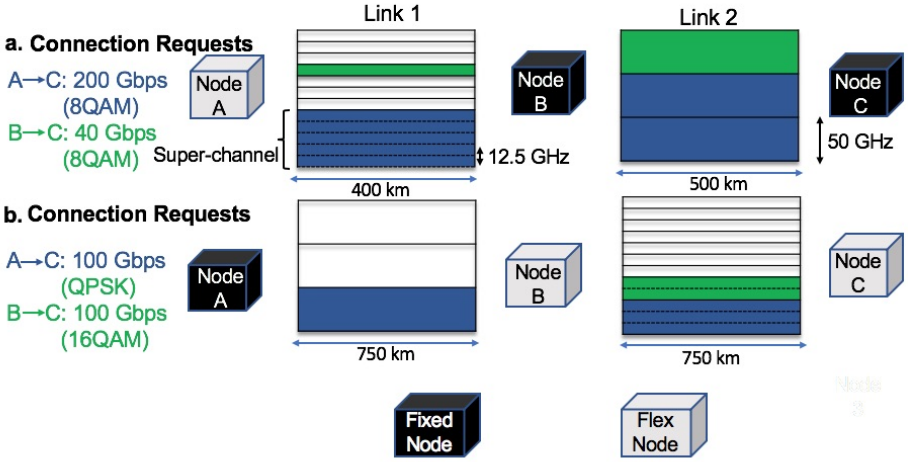


Figure 3.3: Spectrum assignment in different mixed-grid scenarios using different modulation formats. (a) connection requests A → C: 200 Gbps (8QAM), B → C: 40 Gbps (8QAM); and (b) connection requests A → C: 100 Gbps (QPSK), B → C: 100 Gbps (16QAM).

Another key technical advancement towards EON is the introduction of dynamically-adjustable modulation formats. Advanced modulation formats offer higher bit rate and spectral efficiency (bits/sec/Hz), at cost of a lower optical reach. Distance adaptivity [58] [59] is achieved using modulation-adaptive transmitters [51] [60]. Combination of distance-adaptive coherent transceivers with flex-grid links enables even higher spectrum utilization. Fig. 3.3 shows spectrum assignment in mixed-grid scenarios assuming spectrum occupancies for various bit rates as reported in Tables 3.1 and 3.2 [56] [61] [62] [63]. Transmission performance (e.g., reach, operating bandwidth, optical signal-to-noise ratio (OSNR)) of any optical lightpath depends on various factors (fiber, load, and system characteristics) which requires accurate physical-layer models. In our study, we employ a simplification (commonly used in network-layer studies) consisting of setting a maximum optical reach value for a given set of possible bit rates. These values are reported in Table 3.2 and are taken from studies which considered the Gaussian Noise Model [64]. Fig. 3.3 has same settings as Fig. 3.2 with additional capability of assigning spectrum based on different modulation formats which satisfy distance between the source and destination.

In Fig. 3.3(a), a lightpath request of 200 Gbps originates at a flex-grid node (node A), and terminates into a fixed-grid island of two fixed-grid nodes (nodes B and C), having a source-destination distance of 900 km. We observed in Fig. 3.2 that link 1 needs six slots (75 GHz) whereas link 2, being in a fixed-grid island, would need two lightpaths of 50 GHz (100 GHz) to allocate this request using DP-QPSK. However, with inclusion of distance-adaptive properties in node A (flex-grid node), it can use higher modulation format such as 8QAM which still satisfies 900 km reach requirement (see Table 3.2). The spectrum occupation in link 1 is 62.5 GHz whereas link 2, being in a fixed-grid island, would need two lightpaths of 50 GHz (100 GHz) as before. However, the overall spectrum occupation (62.5 + 100 = 162.5 GHz, compared to 175 GHz) is reduced in this distance-adaptive approach using higher modulation. Similarly, a second connection request of 40 Gbps requires only one slot (12.5 GHz) in link 1 but one wavelength channel (50 GHz) in link 2 using modulation format of 8QAM. With a non-distance-adaptive route and spectrum allocation technique, link 1 would need 2 slots (25 GHz) to allocate this 40 Gbps request using DP-QPSK.

Now, in Fig. 3.3(b), a lightpath originating from a fixed-grid node is ending in a flex-grid island. A 100 Gbps connection request from node A to C occupies 50 GHz in link 1 and 37.5 GHz in link 2 using DP-QPSK from fixed-grid node 1. If node 1 were also a flex-grid node, for a distance of 1500 km, it could use 8QAM which needs only 25 GHz in each link. A second lightpath request of 100 Gbps from nodes B to C occupies only 2 slots (25 GHz) from link 2 (flex-grid) using 16QAM.

Standard strategies for resource assignment in EON are not effective for mixed-grid networks. Therefore, we propose a “mixed-grid-aware” algorithm for a novel solution to the dynamic RSMA problem in a mixed-grid network.

3.4 Problem Statement and Solution Strategies

3.4.1 Problem Statement

In this study, we address the RSMA problem in a mixed-grid network, where dynamic traffic requests with heterogeneous bit rate and various traffic profiles are being provisioned. We propose a spectrally-efficient route-selection technique ensuring maximum re-usability of

Table 3.2: Distance and spectrum occupation for various bit rates in flex-grid.

Traffic Demand (Gb/s)	Modulation Format	Operating Bandwidth (GHz)	Distance (km)	#Slots
40	BPSK	50	6000	4
	QPSK	25	3000	2
	8QAM	25	1000	1
100	BPSK	75	4500	6
	QPSK	50	3500	4
	QPSK	37.5	3000	3
	8QAM	25	2500	2
	16QAM	25	1500	2
200	BPSK	100	2500	8
	QPSK	75	1500	6
	8QAM	62.5	1000	5
	16QAM	43.75	700	4
	32QAM	37.5	500	3
400	BPSK	200	2000	16
	QPSK	150	1000	12
	8QAM	100	800	8
	16QAM	75	600	6
	32QAM	56.25	200	5

resources, and distance-adaptive modulation formats are assigned to achieve higher spectrum utilization and lower BBR compared to benchmark techniques.

The dynamic 'on-demand' traffic provisioning problem can be defined as follows: given a network topology (with a set of fixed/flex-grid nodes and links with limited spectrum resources), and incoming traffic requests, find optimal route, spectrum, and modulation format to satisfy the requests while minimizing the BBR.

3.4.2 Solution Strategies

Several strategies can be devised, resulting from the combination of different routing and spectrum allocation policies, as shown below.

1. Routing Policies

Shortest-Path First (SPF): SPF [65] [66] pre-computes a single fixed route for each source-destination pair using a shortest-path algorithm, such as Dijkstra‘s algorithm [67]. When a connection request arrives in the network, it tries to establish a lightpath along the pre-computed fixed route. It checks whether the desired slot is free on each link of the pre-computed route or not. The request is blocked if one link of the pre-computed route does not have the desired slot.

Most Slots First (MSF): This policy [68] keeps track of available slots at each link of a path. It pre-calculates k-shortest paths using k-SPF and arranges them in descending order based on their total available slots obtained from slot availability on their links. Finally, it selects the path with the most available slots. This policy avoids congestion and uniformly distributes the traffic load. In the process, it may take longer routes compared to the route along the shortest path.

Largest Slot-over-Hops First (LSoHF): This policy [68] keeps track of available slots on each link and path length, which is measured in terms of hop count (links). It pre-calculates the k-shortest paths and arranges them in descending order based on the ratio between total available slots and corresponding path length in hops of each path. Finally, it selects the path with the highest value of this ratio (available slots/hops). This policy takes care of the problem of MSF taking longer routes by taking into account path lengths. If a path is taking too many hops, it would automatically be eliminated from being the first to be selected.

Spectrum-Efficient Dynamic Route Allocation (SEDRA): SEDRA is the applied routing policy in this work. It finds the route which requires least spectral allocation among k-shortest routes. For example, in Fig. 3.4, let us consider a 100 Gb/s traffic demand from node 5 to node 1. The first three shortest paths in terms of hops are calculated and marked in three different traits shown with dotted, solid, and dashed lines, respectively in Fig. 3.4. According to Fig. 3.2 and Table 3.1, spectrum requirement for these three paths (assuming

DP-QPSK) can be calculated as follows:

- Path 1, 5-7-8-1 (Three fixed-grid and one flex-grid nodes): (50*3) GHz = 150 GHz.
- Path 2, 5-4-3-1 (One fixed-grid and three flex-grid nodes): (50 + 37.5*2) GHz = 125 GHz.
- Path 3, 5-6-3-1 (Two fixed-grid and two flex-grid nodes): (50*2 + 37.5) GHz = 137.5 GHz.

Path 2 is the most spectrally-efficient route for the request. Although all three paths have same number of links, paths 1 and 3 will waste more spectrum. SPF routing may choose any of these paths as their hop count is same. However, SEDRA chooses path 2 which has higher spectral efficiency.

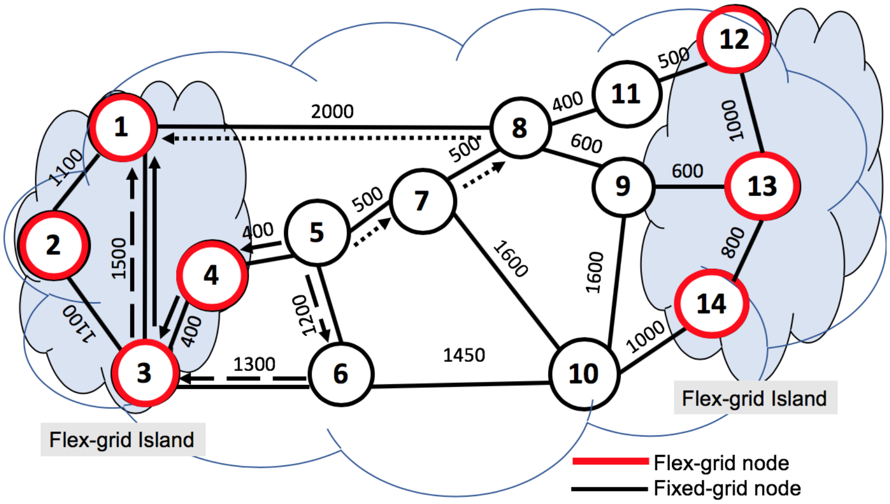


Figure 3.4: Route, spectrum, and modulation-format allocation using SEDRA.

2. Spectrum-Allocation Strategies

First Fit (FF): This policy [69] tries to find the lower-most indexed slot in available spectrum slots. By choosing spectrum in this way, lightpaths are gathered into fewer spectrum slots, which helps to increase the contiguous-aligned available slots in the network [70].

Random Fit (RF): This policy [66] maintains a list of available spectrum slots. When a lightpath request arrives, it arbitrarily selects slots from available slots for lightpath provisioning. It continuously updates available spectrum slots in the process of lightpath allocation and de-allocation. By selecting spectrum slots in a random manner, a network operator tries to reduce the possibility of some specific slots to be used too often [70]. Allocated spectrum slots are expected to be uniformly distributed over the entire spectrum.

Reusable Spectrum Allocation First (RSAF): This policy maintains two separate lists of available slots: at-least-once-used slots and never-used slots. When a lightpath request arrives, this policy selects slots from the used slots first using FF policy. If no slots are available in this list, it selects from the list of never-used slots using FF policy. Choosing spectrum slots this way is an effort to enhance the reuse of spectrum in the network.

3. Modulation-Format Assignment Strategies

We consider two assumptions regarding modulation-format assignment: i) fixed or non-distance-adaptive modulation-format assignment (which has been assumed to be always DP-QPSK). In this case, k-shortest paths are calculated for minimizing number of hops, as this is the most spectrally-efficient choice; ii) distance-adaptive modulation-format assignment in which we incorporate the distances in km to calculate the shortest path. Modulation formats are selected depending on the distance needed to reach the destination.

In Fig. 3.4, let us consider the same 100 Gb/s traffic demand from node 5 to node 1 as we did while explaining SEDRA. The shortest path is selected as 5-4-3-1 with distance of 2300 km. According to Fig. 3.3 and Table 3.2, spectrum requirement and modulation formats can be selected as follows:

- Non-distance-adaptive approach, 5-4-3-1 (One fixed-grid and three flex-grid nodes, QPSK, 3000 kms): (50 + 37.5*2) GHz = 125 GHz.
- Distance-adaptive approach, 5-4-3-1 (One fixed-grid and three flex-grid nodes, 8QAM, 2500 kms): (50 + 25*2) GHz = 100 GHz.

By using distance-adaptive modulation format, we can use even less spectrum to provision the same lightpath request.

3.5 Proposed Algorithm: Mixed-grid-aware Dynamic Resource Allocation (MDRA)

In this section, we describe our proposed algorithm, MDRA, which is a combination of SEDRA and RSAF with modulation format allocation. We show that this combination performs the best in the next section. Given parameters:

- $N(V, E)$: Network topology, with V set of nodes and E set of edges.
- V_{FI}: Set of fixed-grid nodes.
- V_{FL}: Set of flex-grid nodes, where $V = V_{FI} \cup V_{FL}$.
- $n_s(l)$: Start node of a link l, where $n_s(l) \in V$.
- $n_e(l)$: End node of a link l, where $n_e(l) \in V$.
- C_l: Capacity of link in GHz.
- W_{FI}: Frequency slice in fixed-grid links in GHz.
- W_{FL}: Frequency slice in flex-grid links in GHz.
- $\alpha_{s,d}$: Traffic request between nodes S and node D in Gbps.
- ϕ_v: Boolean value which defines if a node v is fixed(0)/flex-grid(1), where v ∈ V.
- ϕ_s: Boolean value which defines if a source node is fixed(0)/ flex-grid(1), where s ∈ V.
- $\phi_{n_s(l)}$: Boolean value which defines if start node of link l is fixed(0)/flex-grid(1).
- $\phi_{n_e(l)}$: Boolean value which defines if end node of link l is fixed(0)/flex-grid(1).
- $P_{s,d}$: Set of k shortest paths $p_{s,d}$ between source s and destination d, where $p_{s,d} \in P_{s,d}$.

-ψ_l^n: Boolean value which defines if link l has n contiguous available slots for a request.

- $\kappa_{s,d}$: Set of candidate paths with requested contiguous slot availability, where $\kappa_{s,d} \subseteq P_{s,d}$.
- γ_l^p: Required spectrum on link l of $p_{s,d}$ for $\alpha_{s,d}$.
- γ_T^p: Total required spectrum slice over all links of $p_{s,d}$ for $\alpha_{s,d}$.
- γ_{min}^p: Lowest required spectrum over all links of $p_{s,d}$ for $\alpha_{s,d}$.
- n: Number of slots required in fixed/flex-grid link for $\alpha_{s,d}$.
- n_o^l: List of available slots on link l which were used at least once.
- n_n^l: List of available slots on link l which were never used.
- $p_{s,d}^{best}$: Path requiring lowest spectrum to allocate $\alpha_{s,d}$.
- m^{best}: Best modulation for given request and distance.

- p^l: Path length.

- p^{fixed}: Boolean value which denotes whether the path consists of all fixed-grid nodes.

Algorithm 2 Mixed-grid aware Dynamic Resource Allocation(MDRA)

```
**Input:** $N(V,E), V_{FI}, V_{FL}, C_l, W_{FI}, W_{FL}, p_{s,d}, \alpha_{s,d}$;
**Output:** Route, Spectrum, and Modulation Format;
**for each** connection request ($\alpha_{s,d}$) **do**
    $P_{s,d} \leftarrow$ find set of k-shortest paths $\alpha_{s,d}$;
                                        ▷ list of candidate paths with available spectrum
    **for each** $p_{s,d}$ **in** $P_{s,d}$ **do**
        **if** ($spectrum\_avail(p_{s,d}, \alpha_{s,d}) == True$) **then**
            $\kappa_{s,d} \leftarrow \kappa_{s,d} \cup p_{s,d}$;
        **end if**
    **end for**
    **for each** $p_{s,d}$ **in** $\kappa_{s,d}$ **do**
        $m \leftarrow modulation\_format(p_{s,d}, \alpha_{s,d})$;
        $\gamma_T^p \leftarrow calculate\_spectrum(p_{s,d}, \alpha_{s,d}, m)$;
                                        ▷ find path requiring least spectrum for $\alpha_{s,d}$
        **if** $\gamma_T^p$ is lowest **then**
            $\gamma_{min}^p \leftarrow \gamma_T^p$;
            $p_{s,d}^{best} \leftarrow p_{s,d}$;
            $m^{best} \leftarrow m$;
        **end if**
    **end for**
    Allocate lightpath on $p_{s,d}^{best}$ using modulation format $m^{best}$ to achieve minimum spectrum allocation of $\gamma_{min}^p$;
**end for**
```

In MDRA, a mixed-grid-aware spectrally-efficient RSMA is applied for varying traffic requests (see Algorithm 2 for detailed pseudo-code). This algorithm is a combination of

SEDRA, RSAF, and distance-adaptive modulation-format allocation. The algorithm finds k-shortest path $P_{s,d}$ for a given traffic request $\alpha_{s,d}$ (lines 1-4 in Algorithm 2). Next, it checks which of these paths has enough spectrum availability (lines 6-10) for requested $\alpha_{s,d}$. Function *spectrum_avail()* calculates contiguous slot availability for each path using function *mixed_grid_spectrum()*, and returns 'true' if slots are available on a path. Function *mixed-grid_spectrum()* identifies location of fixed and flex-grid nodes along the path and returns γ_l^p as required spectrum for $\alpha_{s,d}$. The paths which have the required contiguous slots are listed in a candidate path list (line 8). Now, modulation format and corresponding spectrum allocation for each of the candidate paths are calculated (lines 11-13). Modulation format for corresponding path length p^l is calculated using Table 3.2. For SEDRA without distance-adaptive modulation-format property, modulation format is fixed to DP-QPSK.

Algorithm 3 spectrum_avail()

```
Input: $p_{s,d}, \alpha_{s,d}$;
Output: Boolean, spectrum available or not;
$m \leftarrow modulation\_format(p_{s,d}, \alpha_{s,d})$;
for each link $l$ in $p_{s,d}$ do
    $\gamma_l^p \leftarrow mixed\_grid\_spectrum(s, n_s(l), n_e(l), \alpha_{s,d}, m)$;
    Requested number of slots, $n \leftarrow \gamma_l^p / W_{FL}$;
                        ▷ first find $n$ contiguous slots on $n_o^l$ slots else find in $n_n^l$ of link $l$
    if $\psi_l^n == false$ then
        return $false$;
    end if
end for
return $true$;
```

Algorithm 4 mixed_grid_spectrum()

```
Input: $s, n_s(l), n_e(l), \alpha_{s,d}, m$;
Output: $\gamma_l^p$;
if $\phi_s == 0$ then
    if $\phi_{n_s(l)} == 0$ then
        $calculate\_spectrum(0, \alpha_{s,d}, m)$
                                        ▷ check node type: fixed/flex-grid;
    else if $(\phi_{n_s(l)} == 1 \ \& \ \phi_{n_e(l)} == 0)$ then
        $calculate\_spectrum(0, \alpha_{s,d}, m)$;
    else if $(\phi_{n_s(l)} == 1 \ \& \ \phi_{n_e(l)} == 1)$ then
        $calculate\_spectrum(1, \alpha_{s,d}, m)$;
    end if
else
    if $\phi_{n_s(l)} == 1$ then
        $calculate\_spectrum(1, \alpha_{s,d}, m)$;
    else if $(\phi_{n_s(l)} == 0 \ \& \ \phi_{n_e(l)} == 1)$ then
        $calculate\_spectrum(0, \alpha_{s,d}, m)$;
    else if $(\phi_{n_s(l)} == 0 \ \& \ \phi_{n_e(l)} == 0)$ then
        $calculate\_spectrum(0, \alpha_{s,d}, m)$;
    end if
end if
return $\gamma_l^p$;
```

Algorithm 5 calculate_spectrum()

```
Input: $\phi_v, \alpha_{s,d}, m$;
Output: $\gamma_T^p$;
$\gamma_T^p \leftarrow 0$;
for each link $l$ in $p_{s,d}$ do
    $\gamma_l^p \leftarrow$ find minimum required spectrum for $\alpha_{s,d}$ and modulation format $m$ from Tables 3.1 and 3.2;
    $\gamma_T^p \leftarrow \gamma_T^p + \gamma_l^p$;
end for
return $\gamma_T^p$;
```

Algorithm 6 modulation_format()

```
Input: $p_{s,d}, \alpha_{s,d}$;
Output: $m$;
$p^l \leftarrow$ find path length of path $p_{s,d}$;
$p^{fixed} \leftarrow$ find if $p_{s,d}$ has all fixed-grid nodes;
if $p_{fixed} == True$ then
    return DP-QPSK;
else
    return highest modulation format with reach $p^l$ for $\alpha_{s,d}$ using Table 3.2;
end if
```

Function *calculate_spectrum* then calculates the minimum spectrum required, γ_{min}^p, on path p for $\alpha_{s,d}$. Path which requires minimum spectrum γ_{min}^p is called the best path, $p_{s,d}^{best}$, and modulation format used to achieve minimum spectrum allocation is denoted by m^{best} (lines 15-22).

3.6 Illustrative Numerical Results

Results were obtained over two US network topologies with variation in number of nodes and links. We first consider the 14-node NSFnet topology for analysis. We also considered 24-node USnet backbone network topology, to verify our findings for a larger network. Selection of fixed-grid and flex-grid nodes is pre-determined in both. Half of the nodes are considered to be fixed-grid and another half to be flex-grid. Flex-grid nodes are located at east and west coastal areas. For 14-node NSFnet network (Figs. 3.1 and 3.4), number of fixed-grid links = 14, and number of flex-grid links = 6. For 24-node USnet network (Figs. 3.5, number of fixed-grid links = 29, and number of flex-grid links = 14. Capacity of each optical fiber link is assumed to be 5 THz. This leads to 100 wavelengths for a fixed-grid link with spectrum width of 50 GHz; and 400 frequency slots, each of 12.5 GHz for a flex-grid link. For traffic demand, random pair-connection requests with Poisson inter-arrival and exponential holding time of mean 15 seconds are generated. Today, optical network traffic is of mostly semi-static or static nature. However, traffic is evolving towards a more heterogeneous and application-oriented nature for which the dynamicity is expected to rise. Even today some use cases for dynamic traffic can be found, as science data exchanges over a network such as ESnet [71], or as dynamic ligthpath provisioning in response to important social events (Olympics, concerts, etc.). When we consider a dynamic scenario, we evaluate the absolute performance of our algorithm for this kind of future traffic. Moreover, dynamic traffic studies give an indication also of how to effectively allocate resources in presence of new-arriving traffic requests (incremental traffic demands). To represent heterogeneous traffic, three traffic profiles (Table 3.3) are considered. Profile 1 mimics predominantly low-bandwidth traffic. In profile 2, 100 Gb/s traffic is predominant, representing moderate load. In profile 3, all traffic is 100 Gb/s or higher with significant increase in 400 Gb/s, representing heavy load.

Table 3.3: Traffic profiles.

Traffic Demand (Gb/s)	Profile 1	Profile 2	Profile 3
40	50%	20%	0%
100	30%	50%	40%
200	15%	20%	40%
400	5%	10%	20%

3.6.1 Performance Evaluation Metrics

Performance of the proposed algorithm is evaluated based on the blocked bandwidth and spectrum utilization with gradual increment of normalized offered traffic load. Following are

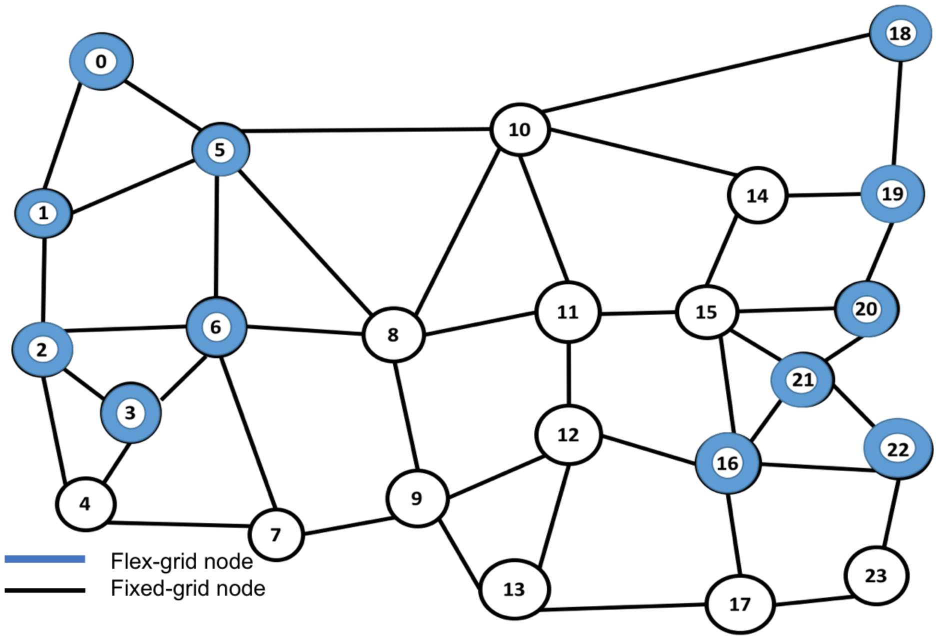


Figure 3.5: Co-existing fixed/flex-grid in 24-node USnet topology.

the performance evaluation metrics considered to evaluate MDRA against other strategies:

BBR = rejected bandwidth divided by total requested bandwidth.

Normalized offered load is calculated based on the amount of traffic arrival compared to the total spectrum capacity of the network. However, total network capacity varies with selection of different modulation formats. It is difficult to calculate network capacity based on each modulation format and make the comparison. Therefore, we assumed 100 Gbps and DP-QPSK to be our baseline standards for network capacity calculation.

Offered load = (connection arrival rate × average request size × average holding time × average path length) ÷ network capacity.

Network capacity = number of fixed-grid links × channel capacity (in GHz) × spectral efficiency of fixed-grid + number of flex-grid nodes × channel capacity (in GHz) × spectral efficiency of flex-grid.

where

- Spectral efficiency of fixed-grid links = 100 ÷ 50 = 2 bits/sec/Hz
- Spectral efficiency of flex-grid links = 100 ÷ 37.5 = 2.6 bits/sec/Hz

Average hops traversed for each path and percentage request blocking in terms of individual traffic demands were also computed for in-depth analysis.

3.6.2 Simulation Results

For a route-selection problem, performance of the algorithm depends on the number of shortest paths, k. In the following graphs, k is chosen to be 10, as we have simulatively verified that no significant gain is achieved above k = *10.* The graphs that are shown in this section correspond to the results on the NSFnet, with the exception of Fig. 3.11, which shows results from USnet.

Fig. 3.6 plots BBR of four routing techniques with spectrum allocation policy RSAF for increasing traffic load, using traffic profile 1. MDRA is a combination of SEDRA and RSAF. As MDRA allocates least spectrum, it achieves the lowest blocking among all, confirming the intuition that, in a mixed-grid, MDRA can outperform existing strategies. For example, for 50% offered load, SPF (0.12) blocks 50% more bandwidth requests compared to MDRA (0.06). MSF has the worst BBR performance (0.28 for 50% offered load) of all four as it does not constrain spectrum usage.

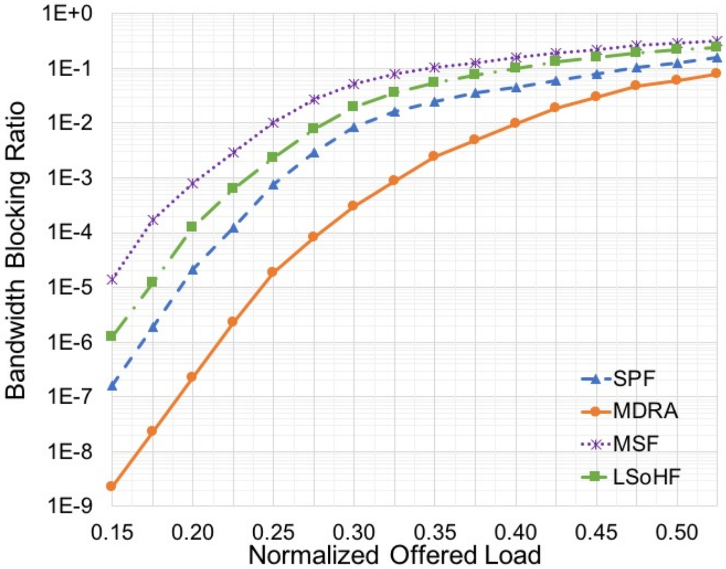


Figure 3.6: Comparison of bandwidth blocking ratio (for NSFnet).

Fig. 3.7 compares BBR of the three spectrum-allocation strategies, when applied with SEDRA. SEDRA with RSAF (MDRA) performs the best in terms of BBR as it promotes spectrum re-usability which helps to accommodate more requests. RF has the highest BBR, as it results in sparse spectrum allocation, causing fragmentation, and lack of contiguous slots for new connections. FF has intermediate performance but still worse than RSAF.

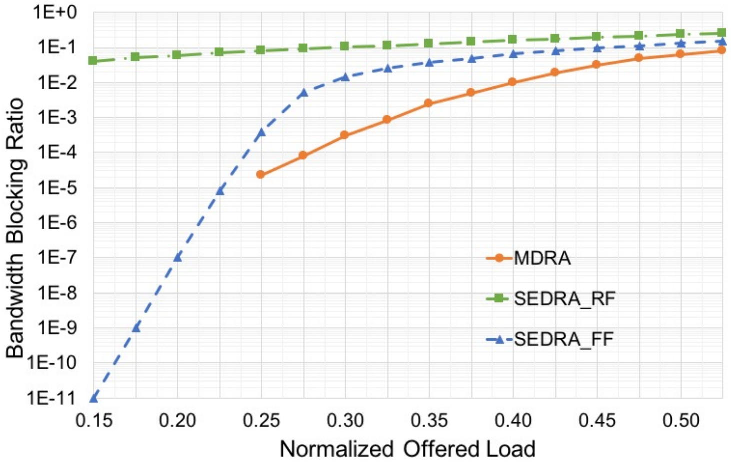


Figure 3.7: Comparison of spectrum allocation strategies (for NSFnet).

Fig. 3.8 represents the average hop count taken by all four routing strategies. MDRA and SPF both allocate average number of hops around 2.4 (for low loads, below 30%). The difference starts at 35% load (see Fig. 3.8). SPF experiences resource shortage and blocks the connection request from this point onwards. Most of the network spectrum becomes exhausted now, therefore average number of hops per connection request gradually decreases (to 2.3 in SPF). MDRA with comparatively lower blocking ratio takes longer routes (up to 2.56 average hop count) to allocate requests when shorter paths are congested. As MSF does not minimize spectrum allocation, and only focuses on paths with highest available spectrum, it takes longer paths (almost up to 7 hops) compared to all four strategies. LSoHF does balance between provisioning shorter path with highest availability. Therefore, the average hop count (up to 4.3 hops) is lower than MSF but not lower than SPF and MDRA. In summary, for 50% offered load, MDRA achieves 50% lower BBR than SPF, with the cost of 11% increased hop count.

Fig. 3.9 shows a breakdown of lightpath blocking for traffic demands with different bit rates. All routing strategies block lowest number of 40 Gbps connections. As expected, blocking increases with increasing bit rate, but MDRA blocks fewer requests due to its mixed-grid-aware properties.

Fig. 3.10 compares BBR of MDRA and SPF with and without distance-adaptive

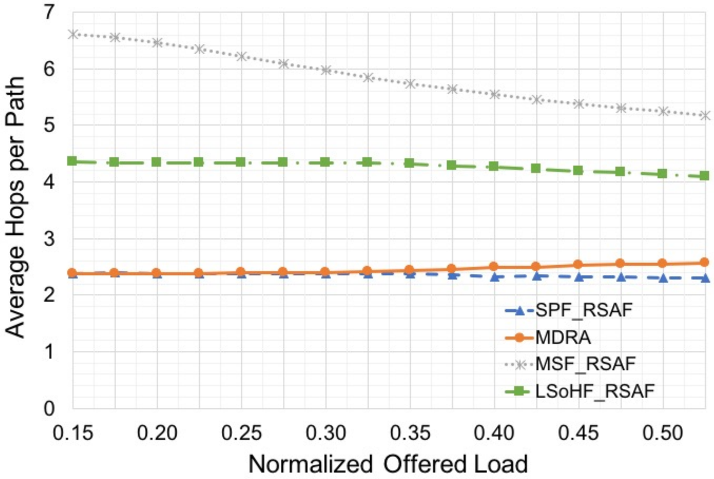


Figure 3.8: Comparison of average hop count among different routing strategies (for NSFnet).

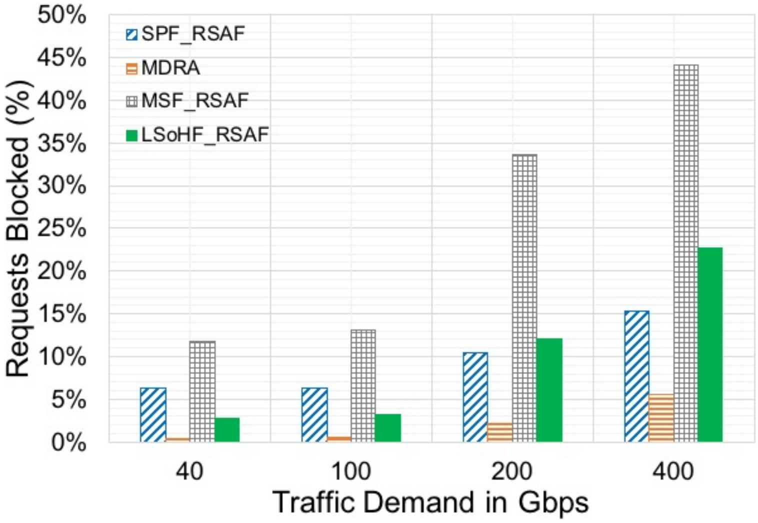


Figure 3.9: Requests blocked from individual bit rates (for NSFnet).

modulation-format allocation (denoted with DA in the figure). Inclusion of DA modulation format increases spectral efficiency, resulting in less blocking. It is worth noting that this decrement in BBR is more significant in MDRA than in SPF (e.g., for 50% load, improvement in BBR is 25% for MDRA and only 15% for SPF).

Fig. 3.11 considers a migration scenario where number of flex-grid nodes is increased gradually. Here, a comparison of BBR is done by setting an increasing number of flex-grid nodes for MDRA and MDRA_DA. It is already shown that MDRA_DA has lower BBR than MDRA without DA modulation format. Number of flex-grid nodes also plays a role in BBR performance of MDRA. As the number of flex-grid nodes grows, capacity of a network to

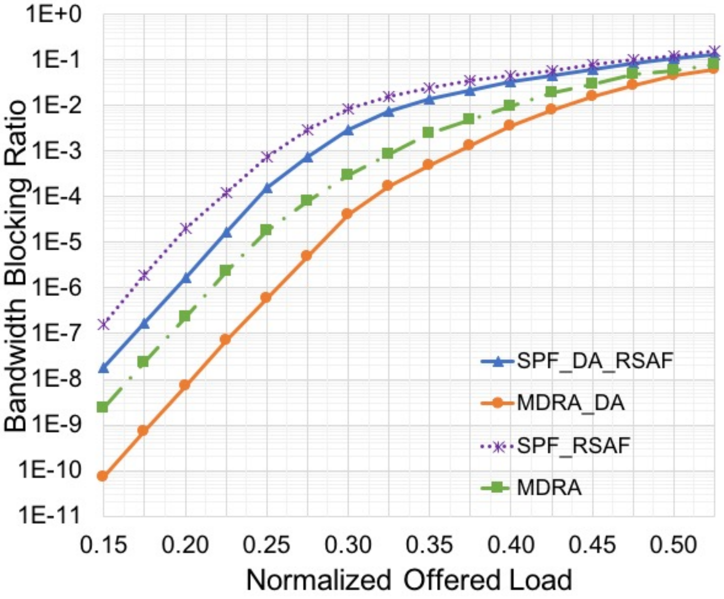


Figure 3.10: Comparison of BBR with and without distance-adaptive modulation (for NSFnet).

accommodate more connection requests also grows.

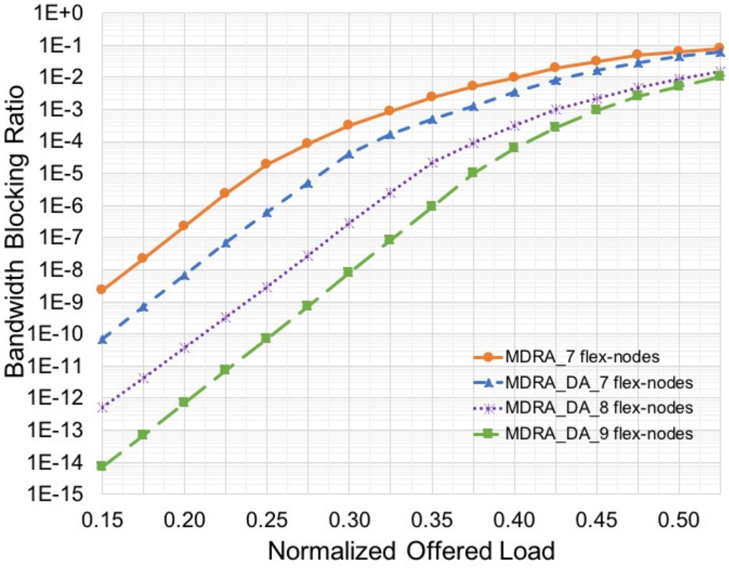


Figure 3.11: Comparison of BBR with varying flex-grid nodes (for NSFnet).

Fig. 3.12 depicts BBR comparison between with and without distance-adaptive modulation formats in a 24-node USnet topology. Observations seen for 14-node NSFnet are confirmed. SPF and MDRA both improve their BBR using DA modulation formats. However, MDRA_DA achieves 40% BBR reduction whereas SPF_DA achieves 16% BBR reduction compared to without DA modulation formats (at 50% offered load). USnet topology, having

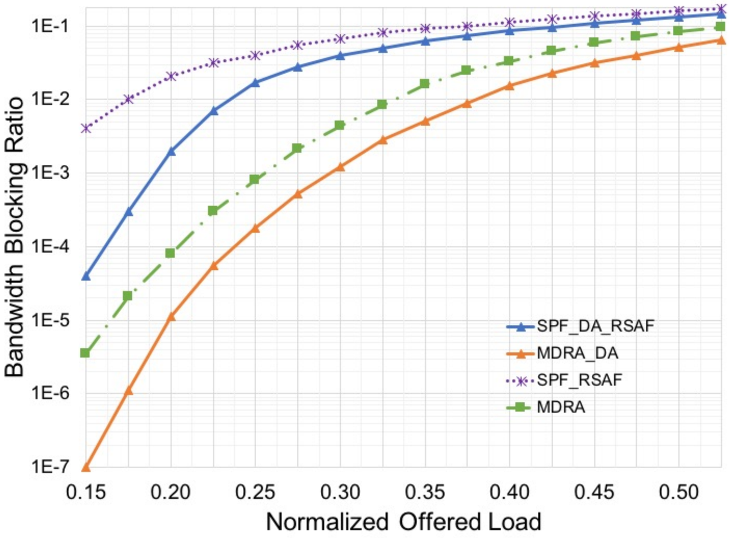


Figure 3.12: Comparison of BBR with and without distance-adaptive modulation-format (for US-net).

higher nodal degree, gives more route options to MDRA to achieve lower blocking.

3.7 Conclusion

Migration towards a flex-grid network is eminent to meet the ever-growing traffic demands. Network operations need to be adaptive to any changes during the process of this migration. RSMA in a mixed-grid network introduces new challenges for network orchestration. In this study, a mixed-grid-aware spectrum-efficient solution, called MDRA, is proposed for dynamic traffic. MDRA routes heterogeneous traffic with lower spectrum allocation. Distance adaptivity is obtained by dynamically adjusting modulation formats, achieving even higher spectrum efficiency. Illustrative results show up to 50% BBR reduction compared to baseline solutions. Also, 25% BBR reduction is achieved with DA modulation-format allocation compared to non-DA approach. We also performed detailed analysis of impact from different traffic profiles, number of flex-grid nodes, modulation formats, and network topology, to gain more insights on RSMA for mixed-grid networks.

In addition to migration to flexible grid, another migration trend is on the rise in optical networks. In order to increase the capacity in optical networks, researchers have proposed migration from C to C+L bands. In Chapters 4 and 5, we discuss novel methods to analyze upgrade strategies for such high-capacity networks.

Chapter 4

C+L Bands Upgrade Strategies to Sustain Capacity Crunch

4.1 Introduction

Global IP traffic volume with a compound annual growth rate (CAGR) of 26% has been forecast by the Cisco global visual networking index (VNI) for 2017-2022 [45]. A long-term capacity scaling in optical backbone networks is necessary to accommodate this fast-growing traffic. In this chapter, we explore C to C+L bands upgrade strategies to utilize the spectrum of single-mode fibers (SMF) in the quest for increasing network capacity.

Studies have debated over multiple solutions, among which spatial-division multiplexing (SDM) [4] and low-loss spectrum optical bands (L, O, E, S, U bands) of single-mode fiber (SMF) [5] have emerged as potential solutions. SDM consists in transmitting over multi-fibers (MF), multi-core (MCF), and multi-mode fiber (MMF). MF technology requires rolling out new optical fiber infrastructure either by installing new fibers or lighting up existing dark fibers, both being expensive options. MCF technologies require deployment of novel type of fibers with complex multiple-input-multiple-output (MIMO) transceivers, which are not yet commercially available. MMF technology deployed only for shorter distance due to inter-modal dispersion. Therefore, expansion of the operating band of existing SMFs beyond C band [6], [7] is a nearer-term viable solution to handle the capacity crunch. This solution maximizes return on investment of already-deployed optical infrastructure. A gradual progression from C band to L, O, E, S, U bands (multi-band (MB)) is envisioned to be a longer-term solution as technology matures for all bands. Submarine networks already have active C+L bands system to extend cable life-time [8].

As a first step towards multi-band transmission, C+L bands expansion allows us to use the existing technologies available for C band, maximizing return on capital expenditure (CAPEX) [72]. Moreover, the attenuation co-efficient variation between C and L bands is negligible, and the in-line erbium-doped fiber amplifier (EDFA) used in C band can be tuned to amplify L band as well. Capacity gain due to this additional band is from 5 THz to 10 THz. However, there are transmission penalties caused by nonlinear interference (NLI) due to inter-channel stimulated Raman scattering (ISRS) [73]. Our work shows how a well-devised upgrade strategy is a key to successful capacity enhancement of the network despite ISRS. The upgrade strategy should account for quality of transmission (QoT) degradation of lightpaths, interoperability issues between C and L bands, unwanted connection blocking, etc.

Network upgrade can be done either all at once or periodically. Operators might choose to upgrade all links of the network as soon as possible in order to absorb the performance benefits of C+L bands early or delay the upgrade process for cost benefits. Continuous traffic growth, technology developments, and cost decrement of equipment require to properly time the upgrades. Therefore, an optimal upgrade strategy is a complex problem. Studies [6], [72] on C+L or multi-band transmission usually compare with SDM techniques by quantifying the amount of added capacity and cost to the network. No study, to the best of our knowledge, has explored cost-effective, multi-period upgrade from C to C+L bands. We assume all links initially to be in C band. As traffic grows, link capacities exhaust, and an upgrade decision to L band needs to be taken. A multi-period batch upgrade strategy [74] is considered, which upgrades batches of links at a time over multiple years. Multi-period network planning with batch upgrade can provide cost-efficient long-term solutions for costly backbone networks [75]. Our work provides a detailed analysis on different multi-period upgrade strategies for C to C+L bands and evaluates their performance on capacity enhancement in optical backbone networks.

We propose upgrade strategies for two types of traffic: predictable and unpredictable. Predictable traffic provides information on exact arrival times of future connection requests whereas unpredictable traffic does not have this information. For a given traffic matrix and network topology, traffic predictability results in the most cost-efficient upgrade. Although the case of unpredictable traffic is more challenging to manage, our proposed upgrade strategy for unpredictable traffic achieves comparable results as predictable traffic. We also explored different upgrade batches (i.e., number of links that can be upgraded at the same time) for both traffic types. Different number of batches results in different cost of upgrade. For predictable traffic, as connection requests and their arrival times are known, the links to be exhausted at future time can be predicted. Upgrading links before their exhaustion times avoids blocking. In contrast, for unpredictable traffic, connection request sequence and their arrival times are unknown. Therefore, we propose an upgrade strategy which analyzes the traffic matrix and topology, to devise the correct link sequence and times to upgrade.

The rest of this chapter is organized as follows. In Section 4.2, related works are reviewed. In Section 4.3, the physical-layer model is described. Section 4.4 provides detailed description of multi-period batch upgrade strategies from C to C+L bands, and introduces some baseline strategies for comparison. Section 4.5 contains the cost model. Section 4.6 shows the proposed cost-efficient upgrade algorithms for both predictable and unpredictable traffic. Section 4.7 introduces the simulation settings and traffic matrix considered. Section 4.8 shows numerical results with explanations. Section 4.9 concludes the chapter.

4.2 Related Works

Researchers have debated over adapting SDM or low-loss spectrum optical band expansion to solve the capacity crunch problem in backbone network. A gradual expansion from C band to other bands is considered to be both deployable and cost-effective.

Ref. [76] investigated several multi-band and multi-fiber upgrade strategies. It found that total capacity served by a multi-band system is slightly better than multi-fiber systems. However, this solution (multi-band) shows a significantly higher network upgrade cost. Ref. [77] shows optical degradation in terms of generalized signal-to-noise ratio, on different bands (C, L, S, U, E, O), resulting from successive channel upgrades until the complete low-loss window is occupied. Given these limitations, Refs. [6] and [72] build a strong motivation on C+L expansion over lighting dark fiber or 2C fiber (two fibers). Simulation results show that C+L bands system does not exhibit capacity penalties compared to parallel C-band systems, but it can unlock more capacity. In light of the above studies, we incorporated noise penalties observed by C+L bands and analyzed capacity gain accordingly.

A critical aspect of C+L bands expansion is a practical physical-layer implementation. Authors in [78] and [79] investigated C+L bands systems, accounting for ISRS and amplified spontaneous emission (ASE) noise generated by in-line amplifiers. They use frequency-dependant dynamic spectral occupancy to account for NLI which imposes complex scenarios with chances of inaccuracy, specially for an on-going migration process. Similarly, Ref. [80] proposes frequency-dependant power control strategies for C+L bands. For every lightpath, there exists an optimal power, maximizing OSNR and transmission capacity. According to [79], ideal range of launch power is -1.5 dBm to -3 dBm. We choose a frequency-independant, fully-filled, worst-case NLI at -1.5 dBm launch power as it gives maximum capacity in the BT-UK network (which is an example network in our study shown in Fig. 4.1) while considering maximum NLI.

Ref. [5] presents a lightpath-provisioning scheme for multi-bands with different upgrade scenarios (C only, C+L, C+L+S, C+L+S+E, C+L+S+E+O) showing significant capacity increase. However, they do not provide any multi-period plan for upgrade. Some studies have discussed the importance of careful link selection for C+L upgrade. Authors in [81] presented a network design framework to exploit C+L bands, focusing on geographically-dependent fiber upgrade expenditures. Simulation results highlight the savings that can be realized by optimizing the usage of L band and carefully selecting the links to be upgraded. With respect to [81], our link-selection technique considers not only geographical locations of links (network topology), but also traffic matrix, and yearly traffic growth. Moreover, we propose a multi-period batch upgrade strategy with associated cost estimation which was not considered in [81].

In the past few years, clear advancements in industrial implementation of C+L bands have been observed [82], [83], [84]. Given this advancement in technology, a cost-efficient, flexible upgrade strategy is crucial for backbone networks. We provide a complete upgrade plan ensuring the most cost efficiency to the network operator.

4.3 Physical-Layer Model

C+L bands transmission can cause significant physical-layer signal impairments. To evaluate the performance of C+L bands upgrade strategy, a thorough physical-layer lightpath OSNR estimation model should be employed to account for the noise contribution due to in-line equipments and effect of ISRS on non-linear interference (NLI) among C+L bands channels. ISRS involves power transfer from C to L band due to Raman effect, and NLI happens due to phase-mismatch nonlinear interactions between non-degenerate frequency triplets [78]. These signal impairments limit OSNR of lightpaths, which determine the set of modulation formats and hence capacity of lightpaths.

We assume a worst-case scenario for both spectrum bands (C and L) while calculating OSNR. This is based on a fully-filled spectrum, where more NLI interactions occur due to higher number of active channels, which degrades OSNR of active C band channels. Therefore, as a link is upgraded to include operations over L band, OSNR of existing C band lightpaths degrades due to higher NLI and ISRS process interactions. Following equation [79] shows OSNR with its noise components:

$$\frac{1}{OSNR(f)} = \sum_{i=0}^{N_L-1} \frac{P^i_{ASE}(f) + P^i_{NLI}(f)}{P_{ch}} + \frac{P^R_{ASE}}{P_{ch}} N_R \tag{4.1}$$

where $P^i_{ASE}(f)$ is total ASE noise from in-line EDFAs and $P^i_{NLI}(f)$ is cumulative NLI due to ISRS in i^{th} optical link. $P^R_{ASE}(f)$ is ASE noise generated in ROADM post amplification. N_L is number of links traversed by the lightpath, and N_R is total number of traversed intermediate ROADMs.

In this chapter, we assume that NLI is accumulated incoherently across multiple spans. Following equation is used to calculate noise power (P_{NLI}) for all intermediate links based on their current state of spectral occupancy [79]:

$$P^i_{NLI}(f_z) = P^3_{ch} N^i_s \eta_1(f_z) \tag{4.2}$$

where P_{ch} is channel launch power, N^i_s is number of spans in i^{th} link, η_1 is NLI co-efficient for a single span, and $P^i_{NLI}(f_z)$ is NLI power of i^{th} link for the channel of interest (COI) f_z. η_1 denotes the total noise contribution across all the active interfering channels [79] which is given by:

$$\eta_1(f_z) = \eta_{XPM}(f_z) + \eta_{SPM}(f_z) \tag{4.3}$$

where η_{XPM} is due to cross-channel and η_{SPM} is due to self-channel interference. For C+L bands, cross-channel interference dominates. Closed-form expression of above total NLI contribution due to η_{XPM} [85] is given by:

$$\eta_{XPM}(f_z) \approx \frac{32}{27} \sum_{k=1,k\neq z}^{N_{ch}} (\frac{P_k}{P_{ch}})^2 \frac{\gamma^2}{B_k \phi_{z,k} \bar{\alpha}(2\alpha + \bar{\alpha})} \\ [\frac{T_k - \alpha^2}{\alpha}] a\tan(\frac{\phi_{z,k} B_z}{\alpha}) + \frac{A^2 - T_K}{A} a\tan(\frac{\phi_{z,k} B_z}{A})] \tag{4.4}$$

where N_{ch} is total number of active channels. Higher the value of N_{ch}, more significant will be $\eta_{XPM}(f_z)$. P_k is power of k^{th} interfering channel, γ is fiber non-linear co-efficient, $\phi_{z,k}$ is phase mismatch term between k^{th} channel and COI, and T_k is frequency-dependent constant of k^{th} channel for ISRS power transfer [85].

4.4 Multi-Period Batch Upgrade for C+L Bands

We explore multi-period batch upgrade strategies under two different cases. If accurate traffic predictions are available, then upgrade decisions are relatively simple. But, both magnitude and exact time of occurrence of future traffic are hard to predict [86], so we also study unpredictable-traffic upgrade strategies.

1. **Upgrade strategies for unpredictable traffic:** Unpredictable traffic does not provide any information on connection request arrivals ahead of time. So, we use insights from initial traffic matrix, network topology, and a large set of simulations representing stochastic evolution of network spectrum occupancy. Goal is to upgrade links that are likely to exhaust in capacity at the right time instant to avoid future blocking. Wrong selection of links and un-timely upgrade can cause cost-inefficiency and blocking. So, to upgrade links to C+L bands, a stochastic link-selection and upgrade-time-selection technique is proposed to identify the link sequence and their probable blocking times. Below, we propose several upgrade strategies for unpredictable traffic based on different link-selection techniques.

Link-selection techniques: Each link is given a weight according to following parameters:

(a) Spectrum utilization of the link.

(b) Number of highly-utilized links in the shortest paths containing the link.

(c) Number of highly-utilized nodes in the shortest paths containing the link.

(d) Number of high-joint-probability node-pairs in shortest paths containing the link.

(e) Betweenness centrality of the link.

Each technique leads to a different strategy, i.e., a different sequence of links to upgrade. Then, a cost analysis of each strategy lets us identify the best strategy among all. Below, we describe each link-selection technique:

(a) **Link spectrum utilization:** Links with higher spectrum utilization are given priority for upgrade. Higher utilization indicates higher probability to exhaust capacity, eventually requiring upgrade to L band. Link spectrum utilization is calculated as follows:

$$W_u(L) = \frac{\text{Spectrum occupation of link L at t}}{\text{Total spectrum of link L}} \tag{4.5}$$

where t is connection request number at which the utilization is calculated and $W_u(L)$ is spectrum utilization of link L.

(b) **Highly-utilized links:** Links which are in a shortest path containing any highly-utilized links are prioritized for upgrade. Highly-utilized links are specified by a threshold obtained from average link utilization. Any link with higher utilization than this threshold is considered as a highly-utilized link. Number of times a link appears in the shortest paths containing any of these highly-utilized links is counted as the weight for this link. Links with higher weight are given higher upgrade priority. Weight of link L is calculated as:

$$W_{hl}(L) = \sum_{U_L} \frac{C(U_L|L)}{C(U_L)} \tag{4.6}$$

where U_L is set of highly-utilized links, $C(U_L)$ is set of shortest paths containing highly-utilized links, $C(U_L|L)$ is set of shortest paths containing both link L and one or multiple highly-utilized links, and $W_{hl}(L)$ is weight of link L from this strategy.

(c) **Highly-utilized nodes:** Links which are in a shortest path containing high-utilized nodes are prioritized. Utilization of a node is calculated from the traffic-generation probability of that node, given by the traffic matrix (Fig. 4.1). Number of times a link lies in the shortest paths containing any of these highly-utilized nodes is considered as weight of that link. Link with higher weight is given upgrade priority. Weight of link L is calculated as follows:

$$W_{hn}(L) = \sum_{U_N} \frac{C(U_N|L)}{C(U_N)} \tag{4.7}$$

where U_N is set of highly-utilized nodes, $C(U_N)$ is set of shortest paths containing highly-utilized nodes, $C(U_N|L)$ is set of shortest paths containing both link L and one or multiple highly-utilized nodes, and $W_{hn}(L)$ is weight of link L from this strategy.

(d) **High-joint-probability node-pairs:** Links which are in a shortest path containing any high-joint-probability node-pairs are prioritized for upgrade. Joint probability of a node-pair is obtained from multiplying individual traffic-generation probabilities listed in the traffic matrix (Fig. 4.1). Node-pairs are considered to be high-joint-probability node-pairs if their joint probabilities exceed the threshold obtained by the average multiplied probability among all node-pairs. Number of times a link lies in the shortest path containing any of these high-joint-probability node-pairs is counted as weight of this link. Links with higher weight are given higher upgrade priority. Weight of link L is calculated as follows:

$$W_{hj}(L) = \sum_{U_J} \frac{C(U_J|L)}{C(U_J)} \tag{4.8}$$

where U_J is set of high-joint-probability nodes, $C(U_J)$ is set of shortest paths containing high-joint-probability node-pairs, $C(U_J|L)$ is set of shortest paths containing both link L and one or multiple high-joint-probability node-pairs, and $W_{hj}(L)$ is weight of link L from this strategy.

(e) **High-betweenness-centrality:** Frequently-traversed links by most lightpaths

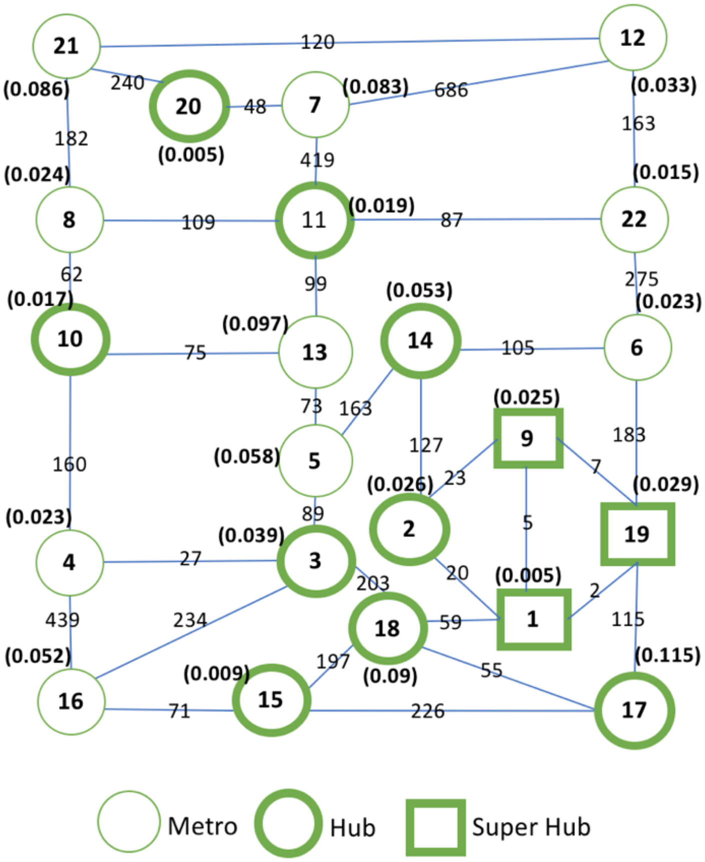


Figure 4.1: BT-UK network with link lengths in km and node metric in parens = (traffic-generation probability).

between nodes are prioritized for upgrade [87]. Links with high-betweenness-centrality have considerable influence within a network by virtue of their control over information passing between nodes. Betweenness centrality of link L is represented as follows:

$$W_b(L) = \sum_{A,B \in V, A!=B,} \frac{C(AB|L)}{C(AB)} \tag{4.9}$$

where V is set of nodes, $C(AB)$ is number of shortest path between A and B, $C(AB|L)$ is number of those paths passing through link L, and $W_b(L)$ is weight of link L using betweenness centrality.

Upgrade-time-selection techniques: The above link-selection techniques help an operator to determine which links should be prioritized for upgrade. However, operators also need to know probable time of blocking of these links to determine when to upgrade them (upgrade time). Upgrading links early or later have significant effect on cost. Upgrading later in time provides cost-efficiency due to reduced equipment cost and CAPEX deferral benefits. However, upgrading later in time might incur connection blocking. So, a cost-efficient upgrade-time-selection technique is critical.

Our proposed method performs a statistical analysis to anticipate time of occurrence of future blocking. To explain how this time is calculated, we refer to an example of finding earliest blocking time. Earliest blocking refers to the very first connection request blocking either before or after any upgrade has been performed. Ideally, no blocking should occur before this earliest blocking instance.

Fig. 4.2 plots the statistical distribution of earliest blocking instances for a large number of independent simulations before any upgrade is done, simulating possible future traffic evolution for BT-UK network. These values follow a normal distribution with finite mean and variance. Depending on targeted blocking probability, an operator can choose an upgrade time.

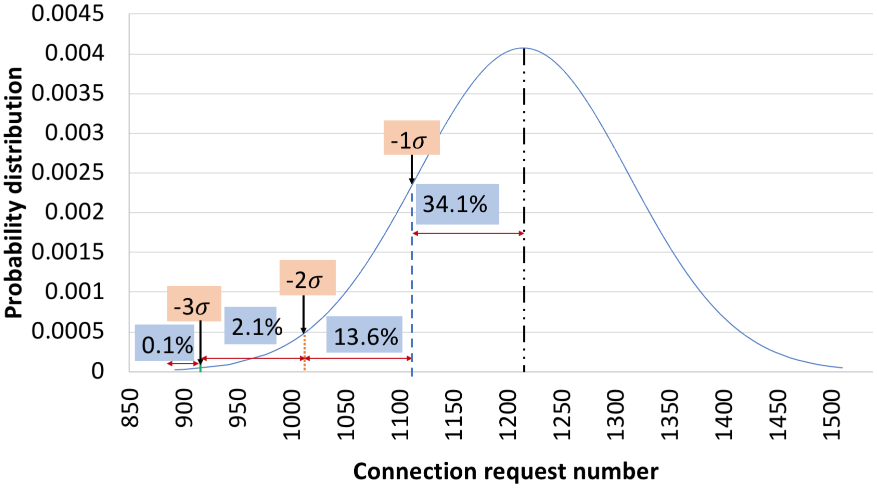


Figure 4.2: Normal distribution of blocking for BT-UK network.

In Fig. 4.2, the normal distribution has mean at connection request number 1241 and standard deviation (σ) of 98. According to the 3 σ rule of 68-95-99.7, 68% of data drawn from this distribution falls within one σ, 95% of data falls within two σ, and 99.7% data falls within three σ away (both negative and positive side) from the mean. For 0.1% blocking probability target, upgrade time will reside within three negative sigma values (connection request number 920) from mean which guarantees least probability of occurrence of any connection request. Any connection request less than 920 has probability of occurrence of 0.1% or less. Therefore, if upgrade to C+L bands is done before connection request number 920, there will be 0.1% or less probability of occurrence of any blocking.

We also assume realistic upgrade completion time (τ). This is the duration to upgrade a batch of links. So, upgrade time can be calculated as follows:

$$M_T = B_T - \tau \tag{4.10}$$

where B_T is earliest blocking time obtained from 3-σ rule, τ is upgrade completion time of one batch of links, and M_T is the obtained upgrade time to upgrade batches of links to C+L bands. Here, time refers to connection request number.

Next, we introduce two base-line strategies that may follow the link-upgrade sequence of any of the above-mentioned link-selection techniques. However, from simulation results, highly-utilized links technique seems to have the most accurate link sequence. Therefore, we use link sequence from highly-utilized links technique for both of these base-line strategies. Although these strategies follow the accurate link sequence, they do not follow the above-mentioned statistical process of time of upgrade.

Early upgrade: This strategy upgrades all links at the beginning without optimizing time of upgrade which causes additional costs due to higher early equipment cost and no CAPEX deferral benefits to the operator (explained in Section 4.5). However, no blocking is observed before the upgrade process is done due to early upgrade.

Blocking un-aware upgrade: This strategy also upgrades all links from highly-utilized links sequence without knowledge of stochastic upgrade times. To keep cost low, the upgrade occurs later causing some blocking.

2. **Upgrade strategy for predictable traffic:** This strategy upgrades links in advance to avoid any future blocking, relying on a perfect prediction of future traffic request arrival times. This strategy can also be used as a baseline for comparison under perfect knowledge. It is assumed that traffic prediction provides information on future requests and their arrival times (in terms of connection request numbers). Hence, both the link sequence to be upgraded and their upgrade times can be deduced from this information. Upgrade completion time is considered using Eq. (4.10) to calculate time of upgrade.

Finally, strategies are formulated to avoid any blocking before all links of the network are upgraded to C+L bands.

Following example explains how upgrade strategy for predictable traffic works for BT-UK network shown in Fig. 4.1.

Table 4.1: Connection requests blocked in C band.

Connection requests number	Blocked links in C band
1147	**8-11, 11-22, 22-6**
1149	**12-22**, 22-6, **6-19, 19-17, 17-18**
...	....
2308	**16-3**, 3-5, 5-14, 14-6

Table 4.1 shows a snapshot of connection requests blocked in C band along with their request number. For a predetermined/predicted set of connection requests, connection blocking and their blocking occurrence times are fixed. Therefore, link sequence and their upgrade times can be deduced from this table. For example, first blocking in C band appears at connection request number 1147 which consists of links 8-11, 11-22, and 22-6, which are all in C band. Similarly, second blocking appears at connection request 1149 which consists of links 12-22, 22-6, 6-19, 19-17, and 17-18 in C band. To avoid connection blocking at 1147 and 1149, associated links need to be upgraded before they get exhausted. So, the sequence of links to be upgraded till connection request 1149 will be (shown in bold in Table 4.1): 8-11, 11-22, 22-6, 12-22, 6-19, 19-17, and 17-18. The time of upgrade will be just before (including upgrade completion time) each link is about to get blocked. The last connection request blocked is at request number 2308, before which all links of the network need to be upgraded to C+L bands to avoid any blocking.

4.5 Upgrade Cost Model

Our multi-period upgrade strategy has periods in years and four quarters per year. An upgrade cost is formulated by summing up the cost of upgrade at each year until all the links are upgraded to C+L bands. Three cost elements are considered in the model:

1. **Equipment cost:** This cost is associated with the equipment required for C+L bands upgrade. Modification in transceiver, digital signal processing units, amplifiers, and filters are required to accommodate C+L bands which results in additional cost. However, as technology matures over the years, equipment cost decreases. Therefore, a depreciation is included in equipment cost calculation. Following is the equation used to calculate equipment cost:

   $$C_E(y) = E * (1 - d\%)^y * l \tag{4.11}$$

   where E is cost to upgrade one link from C to C+L bands, d is yearly equipment cost depreciation value, y is year at which equipment cost is calculated, l is number of links upgraded in year y, and $C_E()$ is resulting depreciated equipment cost at a given year.

2. **Workforce cost:** This cost is associated with the workforce needed to upgrade links from C to C+L bands. It depends on number of links need to be upgraded at certain time. Following is the equation for workforce cost calculation:

   $$C_W(y) = W * l \tag{4.12}$$

   where W is workforce cost to upgrade one link from C to C+L bands, y is year at which workforce cost is calculated, l is number of links upgraded in year y, and $C_W()$ is the resulting workforce cost for a given year.

3. **CAPEX deferral benefit:** Our study applies multi-period upgrade planning by taking time horizon into account. Multi-period planning approaches have objective to minimize network cost. But, a realistic consideration is that a network operator allocates a budget per period (typically, a year), which can be used to build and upgrade the network or, if not used up in the specific period, it can be used in alternate investments. Hence, network operators prefer to invest money in upgrade later than now, i.e., "CAPEX deferral"; and it is a strategy to effectively use a given budget. Following is the equation which accounts for this CAPEX deferral benefit:

   $$B_\delta(y) = \sum_{y=1,n}^{y} C * \delta\% \tag{4.13}$$

   where C is CAPEX budget for each year, δ is yearly CAPEX deferral discount rate earned in alternative investments, y is year at which cost is calculated, y_n is 1 or the year when the last upgrade was performed, and $B_\delta()$ is the resulting CAPEX deferral for a given year.

Total cost is calculated by adding equipment and workforce costs and then subtracting the CAPEX deferral. Following is the total cost of upgrade after Y years:

$$C_T = \sum_{y=1}^{Y} C_E(y) + \sum_{y=1}^{Y} C_W(y) - \sum_{y=1}^{Y} B_\delta(y) \tag{4.14}$$

4.6 Algorithms for Upgrade to C+L bands

Now, we describe the algorithms for our proposed multi-period batch-upgrade strategy to C+L bands for both unpredictable traffic and predictable traffic.

Given parameters:

- $G(V, E)$: Network topology; V set of nodes, E set of links.
- L: Set of links in sequence of upgrade priority obtained from any of the proposed strategies.
- L_C: Set of links in C band.
- L_{C+L}: Set of links in C+L bands, where $E = L_C \cup L_{C+L}$.
- R: Set of connection requests, where $r \in R$.
- R_l: Set of path links of connection requests R.
- R_t: Set of connection requests and their arrival times.
- N: Number of upgrade batches.
- B: Number of links to be upgraded for each batch; also called batch size., where $B = (length(E))/N$.
- T: Traffic matrix.
- α: Yearly increment in traffic in percentage.
- γ: Blocking probability target.
- S: Set of link-selection techniques.
- K: Set of upgrade times in terms of connection request number, where $k_N \in K$.
- $Cost_s$: Cost of an upgrade strategy.
- $Cost_S$: Set of costs for all upgrade strategies.
- S_{best}: Best upgrade strategy with minimum cost.
- L_{best}: Set of links in sequence of upgrade priority obtained from upgrade strategy S_{best}.

- K_{best}: Set of upgrade times in terms of connection request number obtained from strategy S_{best}.

Algorithm 7 takes network topology, number of upgrade batches, traffic matrix, yearly increment in traffic, and blocking probability target from an operator as input. It finds the minimum-cost upgrade strategy in terms of link sequence to upgrade and set of upgrade times. For each upgrade strategy in S, the algorithm finds L-set of links in sequence of upgrade priority using *calc_linkSeq()* (line 4). While the number of upgrade batches, N, is greater than 0, the algorithm finds the current upgrade time, k_1, using normal distribution with given γ. If N is 1, which indicates two cases: operator asked for only one upgrade batch or this is the last remaining batch to upgrade, then all the un-upgraded links will be upgraded to C+L bands at k_1. Next step is to calculate cost at line 16.

If N is not 1, a batch of B un-upgraded links from E is upgraded to L_{C+L} from this set of link sequence L (line 11) at k_N. B is calculated by dividing total number of links by number of upgrade batches, N. This algorithm keeps upgrading B number of links for each upgrade batch until all links are upgraded. It checks whether the number of links in C+L is less than the total number links E; if so, it goes to step 5 and continues to upgrade batches of un-upgraded links until all links are upgraded. Else, if E number of links are being upgraded to C+L, cost calculation is performed using Eq.(4.14). N is set to zero (if not zero) as there is no more link to be upgraded. This algorithm finds cost of individual strategies in S and stores them in $Cost_S$ (lines 22). Finally, it finds the minimum-cost strategy S_{best}, associated link sequence L_{best}, and set of upgrade times K_{best}.

Algorithm 7 Upgrade strategy for unpredictable traffic.

```
1: **Input:** $G(V, E)$, $N$, $T$, $\alpha$, $\gamma$;
2: **Output:** Best upgrade strategy, sequence of links to be upgraded, set of upgrade times, cost of upgrade;
3: **for each** upgrade strategy $s$ **in** $S$ **do**
4:     $L \leftarrow calc\_linkSeq(G(V, E), T, s)$;    ▷ Find set of links ($L$) in sequence of upgrade priority obtained from ($s$);
5:     **while** $N > 0$ **do**
6:         $k_N \leftarrow$ find upgrade times from normal distribution with given $\gamma$;
7:         **if** ($N == 1$) **then**
8:             $L_{C+L} = E \cap L_{C+L}$;    ▷ Upgrade all un-upgraded links from $L$ to $L_{C+L}$ at $k_N$;
9:             Go to step 16;
10:        **else**
11:            $L_{C+L} = L_{C+L} \cup B$; ▷ Upgrade first un-upgraded $B$ links from list sequence $L$ to $L_{C+L}$ at $k_N$;
12:            **if** ($E >$ size of($L_{C+L}$)) **then**
13:                $N = N$ - 1;
14:                Go to step 5;
15:            **else**
16:                Find $Cost_s$ using Eq.(4.14);    ▷ Calculate total cost of upgrade strategy $s$;
17:                $K = K \cup k_N$; ▷ Store the upgrade times for $N$ batches of strategy $s$ in $K$
18:                $N = 0$;
19:            **end if**
20:        **end if**
21:    **end while**
22:    $Cost_S = Cost_S \cup Cost_s$;
23: **end for**
24: $min\_Cost_S$ = find minimum value in $Cost_S$ and associated strategy $S_{min}$
25: $S_{best} \leftarrow S_{min}$
26: $L_{best} \leftarrow L$ associated with $S_{best}$
27: $K_{best} \leftarrow K$ associated with $S_{best}$
```

Algorithm 8 Upgrade strategy for predictable traffic.

```
1: **Input:** $G(V,E)$, $N$, $T$, $\alpha$, $\gamma$, $R_t$;
2: **Output:** Sequence of links to be upgraded, set of upgrade times, cost of upgrade;
3: $(R, R_l) \leftarrow calc\_connLink(R_t)$;   ▷ Find set of connection requests $R$ and their corresponding path links $R_l$ which will be blocked;
4: $L \leftarrow calc\_linkSeq(R, R_l)$;   ▷ Find set of links ($L$) in sequence of upgrade priority;
5: **while** $N > 0$ **do**
6:     $k_N \leftarrow calc\_upgradeTime(R, R_l)$;   ▷ Find set of upgrade times ($k_N$) in sequence of upgrade priority;
7:     **if** ($N == 1$) **then**
8:         $L_{C+L} = E \cap L_{C+L}$;   ▷ Upgrade all un-upgraded links from link sequence $L$ to $L_{C+L}$ at $k_N$;
9:         Break;
10:    **else**
11:        $L_{C+L} = L_{C+L} \cup B$;   ▷ Upgrade first un-upgraded $B$ links from list sequence $L$ to $L_{C+L}$ at $k_N$;
12:        **if** ($E >$ size of($L_{C+L}$)) **then**
13:            $N = N$ - 1;
14:            Go to step 5;
15:        **else**
16:            Break;
17:        **end if**
18:    **end if**
19: **end while**
20: Find $Cost$ using Eq.(4.14);   ▷ Calculate total cost of upgrade;
```

Algorithm 8 takes connection requests and their arrival times as additional inputs from an operator compared to Algorithm 7. Using these inputs, it finds link sequence to upgrade with corresponding set of upgrade times and cost of upgrade. Algorithm 8 works similar to Algorithm 7, except it obtains set of links in sequence of upgrade priority and set of upgrade times from previously-known set of connection requests and their arrival times (R_t). Therefore, it does not need to compare with different link-selection techniques to find the minimum cost strategy.

For both algorithms, physical-layer modeling comes into action during connection request provisioning. At arrival of each connection request, both algorithms check whether the request is to be allocated in C band only, C in C+L bands, or L in C+L bands. OSNR and modulation formats vary according to this connection request allocation in different bands.

4.7 Simulation Setting

4.7.1 Simulation Setup

A custom-built event-driven Java simulator is used to emulate an accurate upgrade environment from C to C+L bands with corresponding physical-layer modeling. BT-UK network (Fig. 4.1) is considered for all our simulations that consists of 35 bi-directional links and 22 nodes, with average link distance of 147 km. As mentioned earlier, we assume all links to be in C band at the beginning. As traffic grows, link capacities exhaust, and an upgrade decision is taken based on different strategies to open L band. C band spectrum ranges from 1530 nm to 1565 nm whereas L band spectrum ranges from 1565 nm to 1625 nm.

A flexible grid is considered with slot width of 37.5 GHz. An uniform channel launch power of -1.5 dbm and ROADM loss of 18 dB are assumed to obtain high-capacity values in BT-UK network [79]. Routing, spectrum, and modulation format selection is done based on availability on k-shortest path, first-fit spectrum allocation, and OSNR threshold parameters [78].

The network simulator takes a network topology and a traffic matrix as inputs, and allocates connection requests with highest modulation format achievable based on physical-layer modeling of different spectrum bands (C and C+L). Connection requests are provisioned between C and L bands to avoid blocking. Initially, connections are provisioned in C band, then to L band if C band does not have needed spectrum. The upgrade time horizon has years and corresponding four quarters as periods of time. An upgrade completion time (τ) of 40 connection requests, or one quarter, is assumed for each batch upgrade. Therefore, an upgrade needs to be performed at least 40 connection requests ahead of the connection request that is about to be blocked. Spectrum continuity and contiguity are maintained for each lightpath.

4.7.2 Traffic Matrix

An incrementally-growing traffic is assumed, with a growth factor of 30% per year. Core networks typically perceive an incremental demand model, i.e., once a lightpath is routed, it stays in the network over all considered periods of time. 3000 connection requests of 100 Gbps are generated by selecting the source and destination from a biased traffic matrix of BT-UK. This traffic matrix is based on connected users in BT-UK network (Fig. 4.1) and corresponding node traffic-generation probability.

4.8 Results and Discussion

We assume yearly equipment depreciation d = 10%, equipment cost to upgrade one link from C to C+L as E = 1 unit, workforce cost for one link upgrade $C_w = 1$ unit, yearly CAPEX budget for upgrade C = 20 units, and yearly CAPEX deferral benefit $B_\delta(y) = 15\%$. This section is divided into three subsections: analysis, results for unpredictable traffic, and results for predictable traffic.

4.8.1 Analysis

Following results show benefit and limitations of C+L bands compared to C band only. We calculate difference in blocking while all links of BT-UK network are in C band and in C+L bands. As expected, number of connection requests blocked is higher when links support only C band. The first blocking occurs much later (at connection request 2714) in C+L bands, compared to C band (at connection request 1281). Fig. 4.3 depicts number of different modulation formats used in BT-UK network. When all links are operated in C band (left part), as BT-UK network has short average link length (about 147 km), most connection requests are provisioned with 16QAM, with few requests provisioned with lower modulation formats (QPSK). But, when all links are in C+L bands (right part), a rise in the number of 8QAM is observed, as OSNR drops when L band gets added at each link due to ISRS.

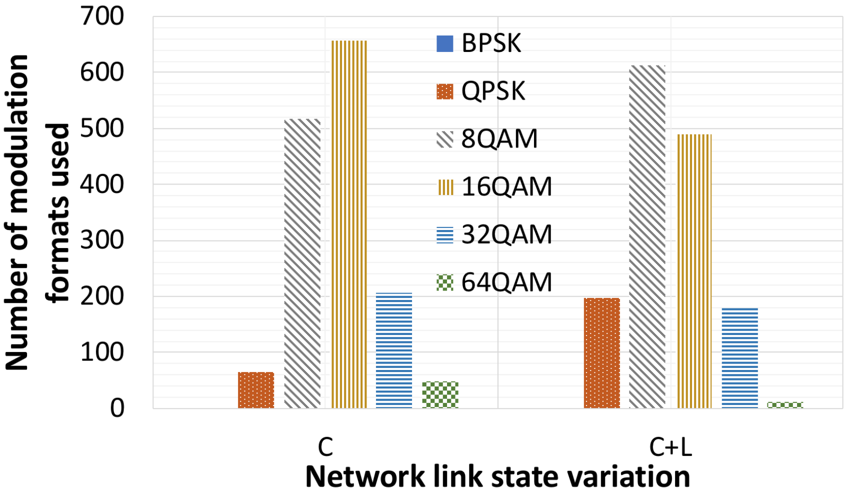


Figure 4.3: Modulation format variation in C and C+L bands.

4.8.2 Unpredictable Traffic

Table 4.2 lists total cost of upgrade strategies for unpredictable traffic where number of upgrade batches is two, which is the least number of batches to select to obtain yearly equipment cost depreciation and CAPEX deferral benefits. Batch size (number of links per batch) is calculated by dividing the number of links in the network by the number of upgrade batches. It is observed that early-upgrade strategy costs the highest (65.2 units) and blocking-unaware strategy costs the least (37.7 units) among all strategies. Early upgrade upgrades all links long before the upgrade time obtained from statistical process mentioned in Fig. 4.2, which causes high cost, as no yearly equipment cost depreciation and capacity deferral benefits are leveraged. In contrary, blocking-unaware strategy upgrades all links after the obtained statistical upgrade time, which gains cost benefits due to yearly equipment cost depreciation and CAPEX deferral but incurs connection blocking before all links are upgraded. Instead, none of the other six strategies experience any blocking before all links are upgraded. Among our proposed strategies, minimum upgrade cost is achieved by the highly-utilized link based strategy, with cost of 48.6 units.

Table 4.3 provides details of different upgrade strategies mentioned in Table 4.2. It shows upgrade time-line of the strategies for unpredictable traffic using two upgrade batches. Two baseline strategies (early upgrade and blocking-unaware upgrade) show two extreme cases where all 35 links are upgraded either early (years: 1, quarter: 1; year: 2, quarter: 1) or late (years: 5, quarter: 3; and year: 7, quarter: 4) compared to the stochastic upgrade times.

Table 4.2: Upgrade cost for unpredictable traffic (two batches).

Upgrade strategy	Upgrade cost (units)
Early upgrade	65.2
Highly-utilized links	48.6
Highly-utilized nodes	52.9
High-joint-probability node-pairs	52.9
High-betweenness-centrality links	57.4
High-spectrum-utilized links	57.4
Blocking-unaware upgrade	37.7

Our proposed approaches schedule the time of first batch (17 links) upgrade at year 3 based on statistical upgrade time, maintaining less than 0.1% blocking. However, the second batch's upgrade (18 links) times varies for the five strategies due to differences in link sequences which eventually differentiates the total upgrade cost. We notice that high-betweenness-centrality and high-spectrum-utilized links strategies require both batches to be upgraded at same time (year: 3, quarter: 2 and year: 3, quarter: 4), which is a cost-inefficient upgrade plan for an operator. Upgrading all links in the same year causes high expenditure in a single year and does not benefit from yearly equipment cost depreciation and CAPEX deferral. It shows the derived link upgrade priority from both high-betweenness-centrality and high-spectrum-utilized links strategy to be inaccurate, resulting in high upgrade cost (Table 4.2). High-betweenness-centrality depends on the topology and ignores the influence of traffic load. High-spectrum-utilized link strategy ranks links based on their spectrum usage only, not considering any associated links.

Highly-utilized nodes and high-joint-probability node-pairs upgrade the second batch of links one year later than the first batch, hence lower cost of upgrade is observed. Both strategies consider influence of only high-traffic generating nodes/node-pairs while prioritizing links to upgrade. Both strategies are based on only node traffic information which miss out to correctly identify the bottleneck links.

Finally, highly-utilized links strategy is able to postpone the second batch of upgrade (year: 5, quarter: 3) more than other strategies without causing any blocking. This indicates the accuracy of the link upgrade priority obtained from this strategy. By definition, highly-utilized links strategy prioritizes links in a shortest path containing any highly-utilized links. In practice, upgrading individual link or node-pair does not ensure the entire lightpath to operate in L band. Therefore, upgrading links associated with highly-utilized links makes the upgrade decision closer to accuracy.

Table 4.3: Upgrade time-line for unpredictable traffic (two batches).

Upgrade Strategy	First batch		Second batch	
	Year	Quarter	Year	Quarter
Early upgrade	1	1	2	1
Highly-utilized links	3	2	5	3
Highly-utilized nodes	3	2	4	3
High-joint-probability node-pairs	3	2	4	2
High-betweenness-centrality links	3	2	3	4
High-spectrum-utilized links	3	2	3	4
Blocking-unaware upgrade	5	3	7	4

Table 4.4 shows the effect of different number of batches when using highly-utilized links strategy. Here, we list cost of two types of batch upgrades: two and three. Three batches of upgrades achieve the least cost (44.6 units). Lower cost is obtained in three-batch upgrades due to lower number of average link upgrades in each year and postponing the last batch upgrade to year 6 compared to two-batch upgrades. These factors leveraged equipment cost depreciation, yearly lower workforce cost, and CAPEX deferral. No more batches (four, five, etc.) could be separated, maintaining the less than 0.1% blocking target. More than three batches separation causes more than 0.1% blocking. Therefore, for unpredictable traffic, highly-utilized links upgrade strategy with three batches is the most cost-efficient upgrade strategy.

Fig. 4.4 shows the annual cost break-down of highly-utilized links strategy for two and three-batch upgrades. Pie chart of three-batch upgrade is slightly smaller because of overall

Table 4.4: Highly-utilized links upgrade strategy with different batches.

Batches	Cost	First batch		Second batch		Third batch	
		Year	#links	Year	#links	Year	#links
Two	48.6	3	17	5	18	-	-
Three	44.6	3	11	4	11	6	13

less upgrade cost. This information helps an operator to distribute the upgrade budget over years. In two-batch upgrade, costs of first batch (24.8 units/ 51%) and second batch (23.8 units/ 49%) are substantial. In contrast, in three-batch upgrade, costs of first batch (13.9 units/ 31%), second batch (16 units/ 36%), and third batch (14.7 units/ 33%) are less, on average. Given upgrade cost and annual budget requirement of both batch upgrades, operators have the flexibility to decide which upgrade plan is the best. To obtain lower cost, an upgrade plan which takes advantage of CAPEX deferral, equipment-cost depreciation, and lower yearly workforce cost would be the best. Otherwise, for shorter overall upgrade period and longer time between upgrades, a plan which upgrades links in bulk (larger batch size) is the best.

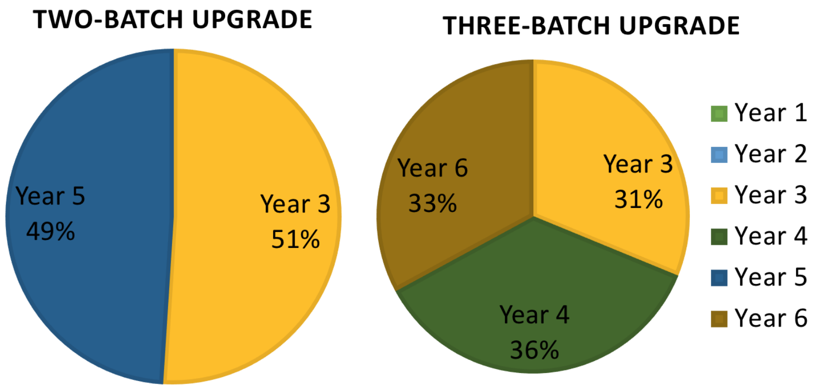


Figure 4.4: Annual cost of different upgrade batches for highly-utilized links strategy.

4.8.3 Predictable Traffic

Table 4.5 compares total cost of upgrade and number of batches achievable for predictable and unpredictable traffic. As highly-utilized link strategy performs the best among all link-selection techniques (shown in Table 4.2), here we call it as ‘unpredictable traffic upgrade

Table 4.5: Cost comparison between predictable and unpredictable traffic upgrade strategies.

Upgrade strategy	Cost	First batch		Second batch		Third batch		Fourth batch	
		Year	#links	Year	#links	Year	#links	Year	#links
Predictable traffic for 2 batches	39.0	4	17	7	18	-	-	-	-
Predictable traffic for 3 batches	38.3	4	11	6	11	7	13	-	-
Predictable traffic for 4 batches	34.6	4	8	5	8	7	8	8	11
Unpredictable traffic for 2 batches	48.6	3	17	5	18	-	-	-	-
Unpredictable traffic for 3 batches	44.6	3	11	4	11	6	13	-	-

strategy'. It is observed that predictable traffic strategies result in lower cost (39, 38.3, 34.6 units) and more number of batches (four batches) compared to unpredictable traffic strategies cost (48.6, 44.6 units) and less number of batches (three batches) due to the known connection requests and corresponding arrivals times. However, more than four batches could not be formed without overlapping two batch upgrades in the same year for predictable traffic upgrade.

Fig. 4.5 shows the annual cost break-down of unpredictable and predictable traffic upgrade strategies with three and four batches. Pie chart of the predictable traffic upgrade is slightly smaller because of lower upgrade cost. Unpredictable traffic upgrade strategy costs more because of early upgrade (years 3, 4, and 6) and higher average batch size (11 links) compared to later (years 4, 5, 7, and 8) and lower average batch size (8 links) of predictable traffic upgrade strategy. Keeping initial batch sizes lower and later batch sizes higher benefits from equipment-cost depreciation. In predictable traffic cost distribution, benefit of CAPEX deferral is observed significantly in years 4 (14%) and 7 (18%) as these upgrades were delayed (3 and 1 years, respectively) from their immediate previous upgrades. Equipment-cost depreciation, workforce cost, and CAPEX-deferral benefits could be exploited more in predictable traffic scenario due to the exact knowledge of traffic arrivals.

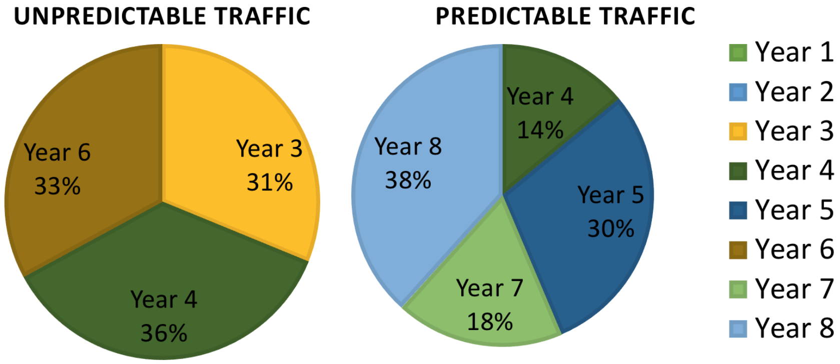


Figure 4.5: Annual cost comparison of predictable and unpredictable traffic upgrade strategies.

4.9 Conclusion

Network upgrade strategies for optical backbone networks from current C band towards C+L bands optical line system has been investigated. A physical-layer-aware, cost-efficient, multi-period, batch upgrade strategy with less than 0.1% blocking was proposed for incrementally-growing traffic (e.g. 30%, per year). Two types of traffic were analyzed to evaluate the performance of the proposed strategy. Predictable traffic upgrade strategy costs less compared to unpredictable ones as the arrival times of connection requests are known ahead of time. However, we showed performance of unpredictable traffic upgrade strategy could be improved by efficient link-selection and upgrade-time-selection techniques. These parameters are network topology and traffic dependent. We proposed several link-selection techniques and analyzed their performance. Finally, for BT-UK network and a given traffic matrix, highly-utilized links strategy showed comparable cost as predictable traffic strategy. This strategy offers an accurate link sequence by considering both the highly-utilized links and associated links which frequently appear in the shortest path. Having an accurate link sequence enables us to postpone batch upgrades in later years, absorbing benefit of equipment cost depreciation, and CAPEX deferral. Our proposed strategy offers flexibility to an operator to chose different number of batches. Generally, more batches lead to less cost due to lower average batch size (lower workforce cost) and CAPEX deferral. However, number of batches cannot be extended arbitrarily due to targeted blocking probability.

In the next chapter, we explore the impact of re-provisioning techniques in C to C+L bands upgrade scenario.

Chapter 5

C to C+L Bands Upgrade with Resource Re-provisioning

5.1 Introduction

To accommodate traffic growth, operators need to upgrade network capacity at minimum cost. Recent studies recommend expansion of low-loss spectrum optical bands (e.g., L, O, E, S, U bands) beyond the current C band [5]. As we discussed in the previous chapter, initial step is expansion to L band [6], which is the second-lowest-loss wavelength band after C band. But L band has higher attenuation, chromatic dispersion, and noise figure compared to C band. Hence, deploying lightpaths in L band leads to lower optical signal-to-noise ratio (OSNR), higher optical power budget, and higher cost. This makes C band signal quality better in terms of modulation format and spectral efficiency compared to L band. In this chapter, we study moving lightpaths from C band to L band judiciously to free up valuable resources of C band.

We study efficient allocation of resources during upgrade of network links from C to C+L bands. After an upgrade, resource allocation may become sub-optimal, leading to lower utilization of spectrum resources. Such inefficient spectrum utilization can block future requests and require early upgrade, which leads to higher cost. Thus, we investigate proactive re-provisioning of lightpaths to C+L bands after each upgrade for cost benefit.

Prior works show benefits of lightpath re-provisioning for restorability of optical mesh networks [9, 10], network capacity maximization [11], service chaining in optical metro networks [12], etc. But, to the best of our knowledge, our work is the first to study the benefits of lightpath re-provisioning during upgrade to C+L bands.

From our previous study we deduced that upgrading highly utilized links leads to cost-effective upgrade strategy. Therefore. in this study, our strategy locates highly-utilized links and upgrades them in batches. After each batch upgrade, existing traffic in C band is re-provisioned to L band using one of these three proposed techniques: "all paths", "shorter paths", and "longer paths" re-provisioning. This re-provisioning frees up high-OSNR lightpaths in C band, leading to improved quality of future transmissions, delayed upgrades, and cost benefits. Results show that re-provisioning of a shorter lightpath provides the most cost-effective upgrade strategy.

5.2 Background on Upgrade Strategy and Cost Model

We use an upgrade strategy based on two steps [16], forming the basis of our re-provisioning strategy in Section 5.3.

- **Link-Selection Technique:** We employ a multi-period strategy where, periodically, a batch of 'highly-utilized links' are selected for upgrade, as this leads to most cost efficiency [16] and as stated in Chapter 4. Batch means a number of links that are upgraded at the same time. Links which are in a shortest path containing any highly-utilized links are prioritized for upgrade. Highly-utilized links are specified by a threshold obtained from average link utilization. Any link with higher spectrum utilization than this threshold is considered as a highly-utilized link. This strategy with three upgrade batches led to the most cost-efficiency, which will be used in this study.

- **Upgrade-Time-Selection Technique:** To handle traffic growth, when to upgrade (Year/Quarter?) is an important question. Upgrading links early or later has significant effect on cost. Our approach performs a statistical analysis to anticipate time of occurrence of future blocking and determines the upgrade time depending on the targeted blocking probability, e.g., 0.1%.

Cost Model: We model the upgrade cost in each year until all links are upgraded to C+L bands. Two cost elements are considered in the model: (i) equipment cost with yearly depreciation; and (ii) CAPEX deferral benefit. Total cost is calculated by adding equipment cost and subtracting CAPEX deferral. Details of the upgrade strategy and cost model can be found in Chapter 4. In this chapter, we focus on the impact of re-provisioning on an upgrade strategy.

5.3 Cost-Efficient Re-Provisioning Strategy

We propose and compare different lightpath re-provisioning strategies applicable during a multi-period batch upgrade. During each upgrade, a batch of C band links are upgraded to L band, causing network links to be in a mix of C and L bands. Re-provisioning moves lightpaths from C band to L band when triggered by a batch upgrade. We assume routes to remain unchanged when moved to L band, so all links of a lightpath need to be upgraded to L band to be re-provisioned. After upgrade, OSNR of re-provisioned lightpaths and lightpaths in C band affected by inter-channel stimulated Raman scattering (ISRS) from extension to L band are re-calculated.

Three lightpath re-provisioning strategies based on lightpath length are investigated:

(1) Re-provision All Lightpaths (R): This strategy re-provisions (from C band to L band) all lightpaths which are routed on links already upgraded to L band;

(2) Re-provision Longer Lightpaths (R^{long}): This strategy re-provisions (from C band to L band) lightpaths whose path length is longer than median path length of already-allocated lightpaths; and

(3) Re-provision Shorter Lightpaths (R^{short}): On the contrary, this strategy re-provisions (from C band to L band) lightpaths whose path length is shorter than median path length of already-allocated lightpaths.

Note that, after upgrade, new requests are allocated (if possible) in L band first in all cases.

No Lightpath Re-Provisioning (NoR, NoR_C): We compare the above strategies with two strategies with no re-provisioning: (i) NoR, i.e., no re-provisioning, assuming new requests after upgrade are routed in L band first; and (ii) NoR_C, i.e., no re-provisioning, but assuming new requests after upgrade are routed in C band first.

OSNR re-calculation is required for all existing C band lightpaths whose links are in both

C and L bands, as OSNR value drops due to L band extension. If new OSNR requires to scale down the modulation format of lightpaths in C band, capacities of these scaled-down lightpaths will decrease. Two situations are now possible:

(i) Re-route C band Lightpaths: It might not be possible to serve some of the requests in the scaled-down lightpaths (overflow requests). Therefore, overflow requests need to be re-routed to a new lightpath; and

(ii) Re-adjust C band Lightpaths: If the scaled-down modulation format of a lightpath can still support all the previously-existing requests, then the OSNR is re-adjusted without needing any request to be re-routed.

Below, we provide a formal description of the proposed re-provisioning algorithms.

Given parameters:

- $G(V, E)$: Network topology; V set of nodes, E set of links.
- T: Traffic matrix.
- α: Yearly increment in traffic in percentage.
- γ: Blocking probability target.
- L_{seq}: Set of links in sequence of upgrade priority.
- L_C: Set of links in C band.
- L_{C+L}: Set of links in C+L bands, where $E = L_C \cup L_{C+L}$.
- R_p: Set of connection requests in a lightpath p.
- N: Number of upgrade batches.
- B: Set of links to be upgraded (batch size) at each batch, where $b \in B$.
- S: Set of re-provisioning strategies, where $s \in S$.
- M_p: Modulation format of a lightpath p.
- O_p: OSNR of a lightpath p.
- P_C: Number of lightpaths in C band.
- P_L: Number of lightpaths in L band.
- $Cost_s$: Cost of an re-provisioning strategy.
- $Cost_S$: Set of costs for all re-provisioning strategies.
- S_{best}: Best re-provisioning strategy with minimum cost.
- K_{best}: Set of upgrade times obtained from strategy S_{best}.

Algorithm 9 finds the minimum-cost re-provisioning strategy. Batch size (number of links per batch) is found by dividing the number of links in the network by the number of upgrade batches, N. While N is greater than 0, the algorithm finds current upgrade time (lines 3-4). A batch of B un-upgraded links from E is upgraded from set of link sequence L_{seq} (line 5) provided by the upgrade strategy. For each re-provisioning strategy in S, the algorithm re-provisions lightpaths (from C to L band) whose links are upgraded to L band, if possible (lines 6-12). After re-provisioning, OSNR of remaining lightpaths in C band whose links are upgraded to L band are re-calculated (lines 13-15). If a request in a path overflows the path capacity due to new OSNR, it is re-routed (lines 16-18). Otherwise, the request remains in the same path with re-adjusted OSNR. After first upgrade, requests are allocated in L band first. Now, the algorithm checks if there are more batches to upgrade (line 19); if so, it goes to step 3 and continues to upgrade batches of un-upgraded links until all links are upgraded. Else, cost calculation is performed (line 20). This algorithm finds cost of individual strategies in S and finds the minimum-cost strategy S_{best} (line 24) and associated set of upgrade times K_{best}.

Algorithm 9 Cost-efficient re-provisioning for C to C+L bands upgrade.

```
Input: $G(V, E)$, $N$, $B$  $T$, $\alpha$, $\gamma$, $L_{seq}$;
Output: Best re-provisioning strategy, set of upgrade times, cost of upgrade;
while ($N > 0$) do
    $k_N \leftarrow$ Find upgrade times for given $\gamma$;
    $L_{C+L} = L_{C+L} \cup b(N)$;   ▷ Upgrade first un-upgraded $b(N)$ links from list sequence $L_{seq}$ to $L_{C+L}$ at $k_N$;
    for each re-provisioning strategy $s$ in $S$ do
        for each lightpath in $p$ in $P_C$ do
            if (All links of $p$ are upgraded to L band) then
                if (Bandwidth available in L band) then
                    Update $O_p$ and $M_p$;
                    Allocate requests $R_p$ in L band, $p \in P_L$;
                    Remove $p$ from C band, update $P_C$;
                end if
            end if
        end for
        for each lightpath in $p$ in $P_c$ do
            if (Links of $p$ is upgraded to L band) then
                Update $O_p$ and $M_p$;
                if Capacity of $R_p$ overflows $p$ then
                    Re-route requests;
                    Update $P_C$ with new lightpaths;
                end if
            end if
        end for
    end for
    $N = N$ - 1;
    Find $Cost_s$;
    $K = K \cup k_N$;
end while
$Cost_S = Cost_S \cup Cost_s$;
$min\_Cost_S \leftarrow$ Find minimum value in $Cost_S$
$S_{best} \leftarrow$ Re-provisioning strategy $S_{min}$ associated to $Cost_S$
$K_{best} \leftarrow K$ associated with $S_{best}$
```

5.4 Simulation Setup

A custom-built event-driven Java simulator is used to emulate a realistic upgrade environment from C to C+L bands. The physical-layer model in [78] is used to find OSNR of a lightpath for different bands (C band only, C and L in C+L bands). BT-UK network with 35 links and 22 nodes is considered (see Fig. 4.1). An incrementally-growing traffic is assumed, with a growth factor of 30% per year. 3000 connection requests of 100 Gbps are generated by selecting the source and destination from a bias traffic matrix of BT-UK. Elastic transponders with five modulation formats (QPSK, QAM, 16QAM, 32QAM, and 64QAM) with corresponding five bit rates (100, 150, 200, 250, and 300 Gbps) are considered. An uniform channel launch power of 0 dbm and ROADM loss of 18 dB are assumed. Routing, spectrum, and modulation format selection is done based on availability on k-shortest path, first-fit spectrum allocation, and OSNR threshold in [78] for modulation format assignment. The upgrade time horizon has years as periods of time.

5.5 Results and Discussion

In this section, we compare total cost of upgrading all links among different re-provisioning strategies. We assume 10% yearly equipment depreciation, equipment cost to upgrade one link from C to C+L as 5 units, yearly CAPEX budget for upgrade of 20 units, and yearly CAPEX deferral benefit of 15%.

Table 5.1: Upgrade cost, years of upgrade batches, and requests in C and L band for all five strategies.

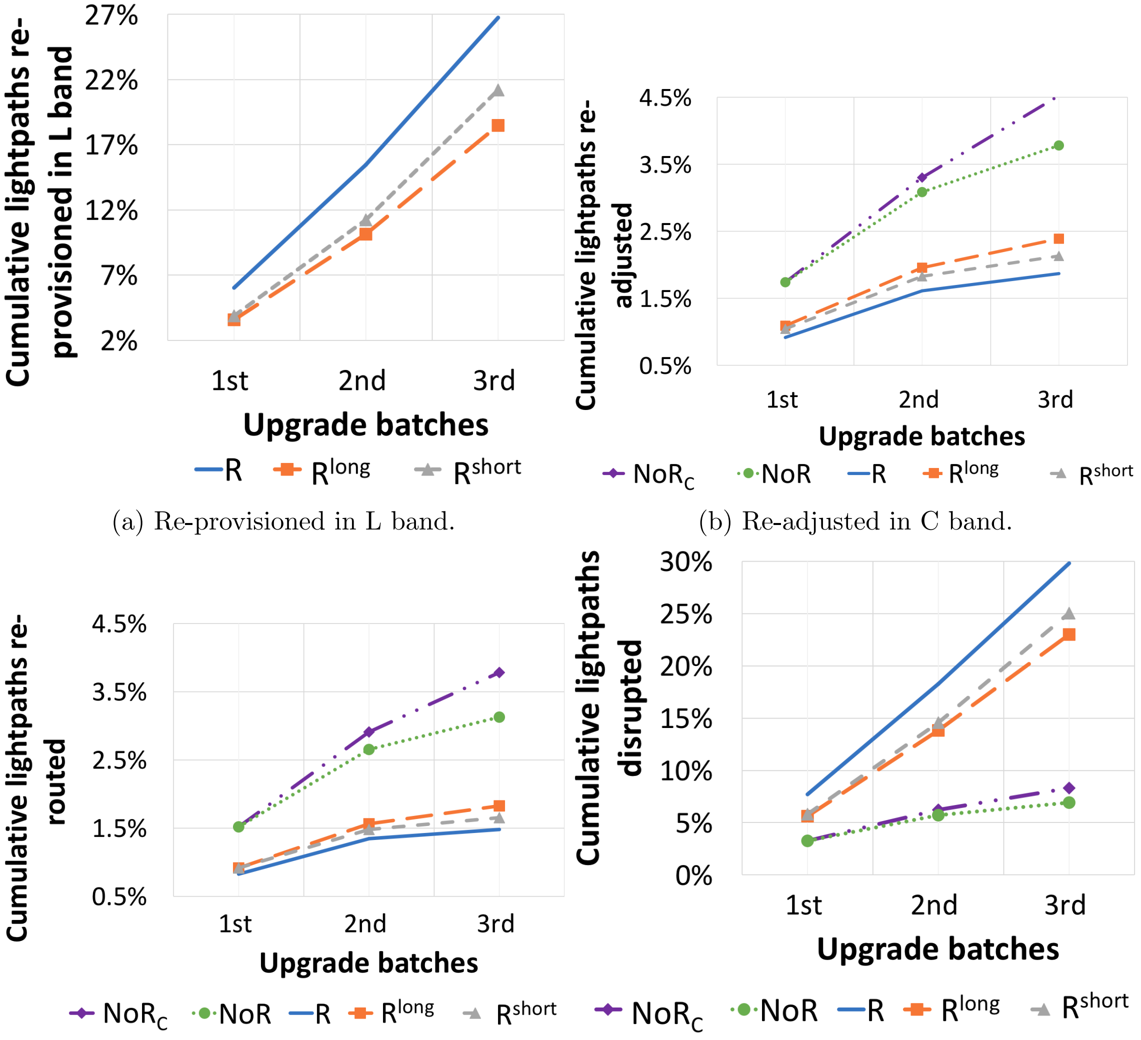
Upgrade strategy	Cost (units)	Year: 1^{st} batch	Year: 2^{nd} batch	Year: 3^{rd} batch	Requests in C	Requests in L
NoR_C	108.0	3	4	6	75.7%	24.3%
NoR	94.7	3	4	8	44.4%	55.6%
R	84.6	3	5	9	32.5%	67.5%
R^{long}	90.7	3	5	8	35.2%	64.8%
R^{short}	78.8	3	5	10	34.1%	65.9%

Table 5.1 lists total cost of upgrade, upgrade years of three batches, and percentage of requests in C and L band for all strategies. It is observed that NoR_C costs the highest (108

units) while R^{short} costs the least (78.8 units). Note that NoR_C upgrades links without re-provisioning while allocating requests in C band first leading to earlier capacity exhaustion in C band (75.7% requests in C band), needing earlier upgrades (years: 3, 4, 6). Instead, R^{short} upgrades links with re-provisioning while allocating traffic in L band first leading to later C band exhaustion (34.1% requests in C band), and needing later upgrades (years: 3, 5, 10) which gains cost benefits due to yearly equipment cost depreciation and CAPEX deferral. R^{short} costs less compared to R^{long} (90.7 units) due to overall less disruption in C band lightpaths created by shorter lightpath re-provisioning. Shorter lightpath re-provisioning causes less disruption by affecting fewer links and lower spectral fragmentation which eventually delays capacity exhaustion and upgrade in C band.

To understand the reasons that lead R^{short} to outperform R and R^{long}, in Fig. 5.1 we provide a detailed analysis of lightpath disruption at different upgrade batches. This helps to understand the trade-off of re-provisioning between creating disruption in the network and reducing the overall cost. Fig. 5.1a shows percentage of lightpaths re-provisioned in L band at the end of upgrade period for the three re-provisioning strategies. For all strategies, a higher number of re-provisioning occurs for later upgrade batches, as more links are already upgraded to L band and hence more re-provisioning options are available. R re-provisions more lightpaths (27%) compared to the other two strategies as all lightpaths in L band are candidates to be re-provisioned. R^{long} re-provisions only longer paths in L, which require bandwidth availability in higher number of links in L band compared to shorter paths. This results in higher number of lightpath re-provisioning in R^{short} (21%) compared to R^{long} (18%).

Figs. 5.1b and 5.1c show cumulative number of lightpaths re-adjusted and re-routed in C band, respectively. For all strategies lower number of lightpaths are re-adjusted and re-routed in C band during later upgrade batches, due to more lightpaths assigned in L band. Overall number of re-adjusted lightpaths (Fig. 5.1b) is higher than re-routed lightpaths (Fig. 5.1c). Higher number of lightpaths are re-adjusted and re-routed by NoR_C (4.5%, 3.7%) and NoR (3.8%, 3.1%), due to larger number of lightpaths located in C band. Among the re-provisioning strategies, R re-provisions more lightpaths in L, thus fewer lightpaths are left in C band to re-adjust (1.9%) and re-route (1.5%). R^{short} compared to R^{long} re-adjusts



(a) Re-provisioned in L band.

(b) Re-adjusted in C band.

(c) Re-routed in C band.

(d) Lightpaths disrupted.

Figure 5.1: Cumulative number of lightpaths (a) re-provisioned, (b) re-adjusted, (c) re-routed, and (d) disrupted in all upgrade batches of all strategies.

(2.1%) and re-routes (1.6%) slightly less lightpaths as there are fewer requests in C.

Fig. 5.1d shows the cumulative number of lightpaths disrupted at each of the upgrade batches. Here, disruption is the sum of number of lightpaths re-provisioned, re-adjusted, and re-routed. It is observed that all re-provisioning strategies disrupt more number of lightpaths compared to strategies without re-provisioning. NoR causes the least (total 7%) whereas R causes the highest (total 30%) disruption.

5.6 Conclusion

Benefit of re-provisioning lightpaths in L band in terms of cost and disruption is shown in a C to C+L bands upgrade scenario. Although re-provisioning disrupts a large amount of traffic (25% in R^{short}), it decreases the number of re-adjusted (2.1% in R^{short}) and re-routed (1.6% in R^{short}) lightpaths and improves the resource allocation leading to lower cost of upgrade.

In the next chapter, we discuss future research directions and conclude the dissertation.

Chapter 6

Conclusion

This dissertation contains four important contributions on the study of next-generation optical networks. Here we summarize the contributions with a discussion on future research directions for each.

6.1 Time Synchronization in Next-Generation Optical Datacenters

In Chapter 2, we surveyed and discussed time-synchronization methods for next-generation optical datacenters. We proposed a zero-overhead time-synchronization method by using data traffic to carry the time messages instead of a separate control channel. Through simulation, we showed that microsecond level of time accuracy can be achieved. We also discussed the dependency of the accuracy on different traffic loads, traffic distributions, and packet length.

6.1.1 Future Research Directions

Our findings in Chapter 2 have encouraged to explore further studies in time synchronization for next-generation optical datacenters, such as:

- Our study assumed all the network elements to be PTP-aware. However, in some cases, operators need to transport PTP over existing or third party PTP-unaware networks. In this case, PTP messages can significantly degrade synchronization performance as some of the network elements may not actively support PTP messages. Strategies need to be implemented to eliminate this situation which is called partial timing support (PTS) [88]. ITU-T has published recommendation G.8273.4, which specifies performance requirements for clocks in a PTS network.

- There is scope to improve PTP accuracy which we did not cover in our study. PTP approximates the bi-directional path delay from master to slave to be symmetric. However, this delay comprises of cable delay, as well as transmit and receive latency which varies amongst chip vendors for different demployment methods. This should be considered in future studies.

- Use of optical network connectivity inside datacenter network is increasing. Datacenters are experiencing next-generation optical network technology deployments such as mixed-grid optical networks, optical time slice switching (OTSS), multi-modal optical networks, etc. Appropriate methods to exploit these new deployments to gain more precise time synchronization should be explored.

- Our study focuses on datacenter networks. However, sensor networks such as IoT are an important study for current technology needs [89]. In a IoT network, distributed sensor data are gathered in real-time which requires stringent time accuracy.

6.2 Dynamic Resource Allocation in Mixed-Grid Optical Networks

Chapter 3 proposed a solution to a RSMA problem in a mixed-grid technology environment considering interoperability constraints. The solution proposes routes, spectrum, and modulation format to provision a dynamic, heterogeneous traffic on two US-wide network topologies ensuring maximum network throughput and minimum blocking.

6.2.1 Future Research Directions

While working on Chapter 3, we have come up with some ideas on future research directions, as follows:

- Resource optimization is an important problem for mixed-grid networks. Dynamic RSMA poses sub-optimal resource allocation in a complex scenario such as the mixed-grid. One of the ways can be to study spectrum defragmentation techniques which can improve resource allocation.

- Insights from this study can help make upgrade decisions to flex-grid. Highly-used links and nodes can be detected and prioritized for upgrade.
- Our study assumes a network with a fixed number of already-upgraded flex-grid nodes. However, RSMA techniques should be proposed for a gradually upgrading flex-grid nodes which resembles a realistic migration scenario.
- Ensuring protection during migration can be an important problem for a failure scenario. Strategies should be studied to maintain specific constraints of interoperability between fixed and flex-grid network elements while applying shared or dedicated protection.
- Flex-grid allows to assign as much optical bandwidth as needed by each transponder. In this case, traffic grooming can be a useful way of varying volume of resource management. Resource allocation for grooming in a mixed-grid network is another important problem to be explored.

6.3 C+L Bands Upgrade Strategies to Sustain Capacity Crunch

Chapters 4 investigated cost-efficient upgrade strategies for capacity enhancement in optical backbone networks enabled by C+L bands optical line systems.

6.3.1 Future Research Directions

While investigating upgrade strategies for C+L bands system, we realized some important research problems in this area, as follows:

- Research is moving towards multi-band upgrade to expand the network capacity even more. Our study suggested link-selection and upgrade-time-selection techniques. An efficient band-selection technique can be the next realistic parameter needed for multi-band upgrade strategies.

- Besides considering CAPEX-deferral benefits and equipment-cost depreciation, including cost for upgrading to different bands is also an important parameter to consider. For example, [90] assumed 20% higher equipment cost in L band compared to C band. Based on the technologies needed, different bands may impose different costs.

- The link-selection techniques we suggested requires mathematical modeling which depends on network topology and traffic matrix. Studies should explore other link-selection techniques based on learning-based algorithms backed up by network data.

- Adaptive upgrade strategy based on current state of the network is useful for a dynamic and short-term upgrade scenario. This strategy is flexible in terms of planning period, batch size, and upgrade times.

6.4 C to C+L Bands Upgrade with Resource Re-provisioning

Chapters 5 investigated cost-efficient resource re-provisioning strategies for capacity enhancement in optical backbone networks enabled by C+L bands optical line systems.

6.4.1 Future Research Directions

Resource optimization for a C+L bands system is crucial for better spectrum utilization. Followings are some future work ideas in this direction:

- We have followed a first-fit approach while re-provisioning lightpaths to L band. However, to aid our resource optimization goal, another level of optimization can be done during spectrum allocation such as: fragmentation-aware spectrum allocation.

- Rather than re-provisioning at each upgrade instance, one can propose a smart re-provisioning approach which minimizes the number of re-provisioning instances while still optimizing resource allocation.

- The cost model can consider relevant penalties for lightpath re-allocation in L bands such as traffic disruption, cost of lighting up new transponders, etc.

- Our study re-provisions lightpaths if all the links on the path are in L band. A reasonable update of this approach can be to prioritize lightpaths based on their probability of getting used in future.

References

[1] D. L. Mills, "Internet time synchronization: the network time protocol," *IEEE Transaction on Communications,* vol. 39, no. 10, pp. 1482-1493, Oct. 1991.

[2] G. Wu, L. Hu, H. Zhang, and J. Chen, "High-precision two-way optic-fiber time transfer using an improved time code," *Review of Scientific Instruments*, vol. 85, no. 11, pp. 114701-114706, Nov. 2014.

[3] K. Correll, N. Barendt, and M. Branicky, "Design considerations for software only implementations of the IEEE 1588 precision time protocol," *Proc., Conference on IEEE 1588,* pp. 11-15, Oct. 2005.

[4] C. Antonelli, M. Shtaif, and A. Mecozzi, "Modeling of nonlinear propagation in space-division multiplexed fiber-optic transmission," *IEEE/OSA Journal of Lightwave Technology,* vol. 34, no. 1, pp. 36–54, Dec. 2015.

[5] N. Sambo, A. Ferrari, A. Napoli, N. Costa, J. Pedro, B. S. Krombholz, P. Castoldi, and V. Curri, "Provisioning in multi-band optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2598-2605, 2020.

[6] E. Virgillito, R. Sadeghi, A. Ferrari, G. Borraccini, A. Napoli, and V. Curri, "Network performance assessment of C+L upgrades vs. fiber doubling SDM solutions," *Proc., Optical Fiber Communications Conference (OFC)*, Mar. 2020.

[7] E. Virgillito, R. Sadeghi, A. Ferrari, A. Napoli, B. Correia, and V. Curri, "Network performance assessment with uniform and non-uniform nodes distribution in C+L upgrades vs. fiber doubling SDM solutions," *Proc., Optical Network Design and Modeling (ONDM)*, May 2020.

[8] D. Aguiar, G. Grasso, A. Righetti, and F. Meli, "EDFA with continuous amplification of C and L bands for submarine applications," *Proc., International Microwave and Opto-electronics Conference (IMOC)*, Nov. 2015.

[9] L. Song, J. Zhang, and B. Mukherjee, "A Comprehensive study on backup-bandwidth reprovisioning after network-state updates in survivable telecom mesh networks," *IEEE/ACM Trans. on Networking*, vol. 16, no. 6, pp. 1366-1377, Dec. 2008.

[10] C. Assi and W. Huo, "On the benefits of lightpath re-provisioning in optical mesh networks," *Proc., IEEE International Conference on Communications (ICC)*, May 2005.

[11] I. Kim, X. Wang, O. Vassilieva, P. Palacharla, and T. Ikeuchi, "Maximizing optical network capacity through SNR-availability based provisioning," *Proc., Optical Fiber Communications Conference and Exhibition (OFC),* 2019.

[12] L. Askari, F. Musumeci, and M. Tornatore, "Reprovisioning for latency-aware dynamic service chaining in metro networks," *IEEE/OSA Journal of Optical Communications and Networking*, Nov. 2020.

[13] T. Ahmed, S. Rahman, M. Tornatore, K. Kim, and B Mukherjee, "A survey on high-precision time-synchronization techniques for optical datacenter networks and a zero-overhead microsecond-accuracy solution," *Photonic Network Communications,* vol. 36, no. 1, pp. 56-67, Aug. 2018.

[14] T. Ahmed, S. Rahman, S. Ferdousi, M. Tornatore, A. Mitra, B. C. Chatterjee, and B. Mukherjee, "Dynamic routing, spectrum, and modulation-format allocation in mixed-grid optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 12, no. 5, pp. 79-88, May 2020.

[15] T. Ahmed, S. Rahman, M. Tornatore, X. Yu, K. Kim, and B. Mukherjee, "Dynamic routing and spectrum assignment in co-existing fixed/flex-grid optical networks," *Proc., IEEE Advanced Networks and Telecom Systems*, Indore, India, Dec. 2018.

[16] T. Ahmed, A. Mitra, S. Rahman, M. Tornatore, A. Lord, and B. Mukherjee, "C+L bands upgrade strategies to sustain traffic growth in optical backbone networks," under review by *IEEE/OSA Journal of Optical Communications and Networking*, 2021.

[17] T. Ahmed, S. Rahman, A. Pradhan, A. Mitra, M. Tornatore, A. Lord, and B. Mukherjee, "C to C+L bands upgrade with resource re-provisioning in optical backbone networks," submitted to *Optical Fiber Communications Conference (OFC)*, 2021.

[18] L. Wang, X. Wang, K. J. Kim, S. M. Kim, D. U. Kim, K. E. Han, and B. Mukherjee, "Priority-aware scheduling for packet-switched optical networks in datacenter," *Proc., Optical Network Design and Modeling,* May 2017.

[19] L. Lamport, "Time, clocks, and the ordering of events in a distributed system," *Communications of the ACM,* vol. 21, no. 7, pp. 558-565, July 1978.

[20] H. Marouani and M. R. Dagenais, "Internal clock drift estimation in computer clusters," *Journal of Computer Systems, Networks, and Communications,* vol. 2008, no. 583162, May 2008.

[21] "Time Is an Illusion Lunchtime Doubly So," ACM, [Online]. Available: `https://cacm.acm.org/magazines/2016/1/195723-time-is-an-illusion-lunchtime-doubly-so/fulltext` [Accessed: Feb. 19, 2021].

[22] M. Levesque and D. Tipper, "A survey of clock synchronization over packet-switched networks," *IEEE Communications Surveys and Tutorials,* vol. 18, no. 4, pp. 2926-2947, July 2016.

[23] G. C. Sankaran and K. M. Sivalingam, "Time synchronization mechanisms for an optically groomed data center network," *Proc., Int. Performance Computing and Communications Conference,* Dec. 2016.

[24] M. D. Korreng, "UTC time transfer for high frequency trading using IS-95 CDMA base station transmissions and IEEE-1588 precision time protocol," *Proc., 42nd Annual Precise Time and Time Interval Systems and Apps. Meeting,* pp. 359-368, Nov. 2010.

[25] T. Mizrahi and Y. Moses, "Software defined networks: It's about time," *Proc., IEEE International Conference on Computer Communications,* July 2016.

[26] Cisco global cloud index: forecast and methodology, 2012-2017, Cisco Systems, Inc., San Jose, CA, USA, 2014.

[27] P. Costa, H. Ballani, K. Razavi, and I. Kash, "R2C2: a network stack for rack-scale computers," *Proc., ACM SIGCOMM,* Aug. 2015.

[28] D. D. McCarthy and K. P. Seidelmann, "Optical Atomic Standards," *Time from Earth Rotation to Atomic Physics, Weinheim: Wiley-VCH,* Ch. 11, pp. 181-186, Germany, 2009.

[29] W. Lewandowski, J. Azoubib, and J. W. Klepczynski, "GPS: primary tool for time transfer," *Proceedings of the IEEE,* vol. 87, no. 1, pp. 163-172, Jan. 1999.

[30] "NTP and PTP time transfer," Precise time and frequency LLC, [Online]. Available: `http://www.ptfinc.com/wp-content/uploads/2015/10/NTP-and-PTP-Time-Transfer-Part-2.pdf` [Accessed: Feb. 19, 2020].

[31] J. Perry, A. Ousterhout, H. Balakrishnan, D. Shah, and H. Fugal, "Fastpass: A centralized 'zero-queue' datacenter network," *Proc., ACM SIGCOMM,* pp. 307-318, Aug. 2014.

[32] K. S. Lee, H. Wang, V. Shrivastav, and H. Weatherspoon, "Globally synchronized time via datacenter networks," *Proc., ACM SIGCOMM,* pp. 454-467, Aug. 2016.

[33] M. Levesque and D.Tipper, "PTP++: a precision time protocol simulation model for OMNeT++ / INET," *Proc., OMNeT++ Community Summit,* Sept. 2015.

[34] Y. Liu and C.Yang, "OMNeT++ based modeling and simulation of the IEEE 1588 PTP clock," *Proc., International Conference on Electrical and Control Engineering,* Sept. 2011.

[35] W. Wallner, A. Wasicek, and R. Grosu, "A simulation framework For IEEE 1588," *Proc., International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS),* Sept. 2016.

[36] S. T. Watt, S. Achanta, H. Abubakari, E. Sagen, Z. Korkmaz, and H. Ahmed, "Understanding and applying precision time protocol," *Proc., Power and Energy Automation Conference,* Mar. 2014.

[37] H. Yang, J. Zhang, Y. Ji, and Y. Lee, "C-RoFN: multi-stratum resources optimization for cloud-based radio over optical fiber networks," *IEEE Communications Magazine,* vol. 54, no. 8, pp. 118-125, Aug. 2016.

[38] A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren, "Inside the social network's (datacenter) network," *Proc., ACM SIGCOMM,* pp. 123-137, Aug. 2015.

[39] D. Ersoz, M. S. Yousif, and C. R. Das, "Characterizing network traffic in a cluster-based, multi-tier data center," *Proc., International Conference on Distributed Computing Systems,* June 2007.

[40] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," *Proc., ACM SIGCOMM,* pp. 92-99, Jan. 2010.

[41] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of datacenter traffic: measurements and analysis," *Proc., 9th ACM SIGCOMM Internet Measurement Conference,* pp. 202-208, Nov. 2009.

[42] "Cisco IE3000," Cisco, [Online]. Available: `http://www.cisco.com/c/en/us/td/docs/switches/lan/cisco_ie3000/software/release/12-2_52_se/configuration/guide/ie3000scg/swptp.pdf` [Accessed: Feb. 19, 2020.]

[43] "IEEE 1588 PTP on Cisco Nexus 3100 platform and 9000 series switches," White Paper, Cisco, March 6, 2015.

[44] "Technical solution guide for PTP," Arista, [Online]. Available: `https://www.arista.com/assets/data/pdf/Whitepapers/Technical_Solution_Guide_Precision_Time_Protocol.pdf` [Accessed: Feb. 19, 2020.]

[45] "Cisco Visual Networking Index: Forecast and Trends, 2017-2022 White Paper," Cisco, [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html. [Accessed: Nov. 2019]

[46] M. Ruiz, L. Velasco, A. Lord, D. Fonseca, M. Pioro, R. Wessaly, and J. P. F. Palacios, "Planning fixed to flexgrid gradual migration: drivers and open issues," *IEEE Communications Magazine,* vol. 52, no. 1, pp. 70-76, Jan. 2014.

[47] X. Yu, M. Tornatore, M. Xia, J. Wang, J. Zhang, Y. Zhao, J. Zhang, and B. Mukherjee, "Migration from fixed grid to flexible grid in optical networks," *IEEE Communication Magazine,* vol. 53, no. 2, pp. 34-43, Feb. 2015.

[48] A. Mayoral, V. Lopez, O. Gonzalez de Dios, and J. P. F. Palacios, "Migration steps toward flexi-grid networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 6, no. 11, pp. 988-996, Nov. 2014.

[49] S. Yan, E. H. Salas, A. Hammad, Y. Yan, G. Saridis, S. Bidkar, R. Nejabati, D. Simeonidou, A. Dupas, and P. Layec, "Demonstration of bandwidth maximization between flexi/fixed grid optical networks with real-time BVTs," *Proc., European Conference on Optical Communication*, Sept. 2016.

[50] M. Jinno, "Elastic optical networking: roles and benefits in beyond 100-Gb/s era," *IEEE/OSA Journal of Lightwave Technology,* vol. 35, no. 5, pp. 1116-1124, Mar. 2017.

[51] X. Zhou, L. E. Nelson, and P. Magill, “Rate-adaptable optics for next generation long-haul transport networks,” *IEEE Communications Magazine,* vol. 51, no. 3, pp. 41-49, Mar. 2013.

[52] X. Yu, Y. Zhao, B. Chen, J. Zhang, Y. Li, G. Zhang, X. Chen, and J. Zhang, “Migration-aware dynamic connection provisioning in optical networks evolving from fixed grid to flexible grid,” *Proc., Asia Communications and Photonics Conference (ACP),* Nov. 2016.

[53] X. Yu, Y. Zhao, J. Zhang, B. Mukherjee, J. Zhang, and X Wang, “Static routing and spectrum assignment in co-existing fixed/flex grid optical networks,” *Proc., Optical Fiber Communications Conference and Exhibition,* Mar. 2014.

[54] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, and A. Hirano, “Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network,” *IEEE Communications Magazine*, vol. 48, no. 8, pp. 138-145, Aug. 2010.

[55] X. Wang, G. Shen, Z. Zhu, and X. Fu, “Benefits of sub-band virtual concatenation for enhancing availability of elastic optical networks,” *IEEE/OSA Journal of Lightwave Technology,* vol. 34, no. 4, pp. 1098-1110, Dec. 2015.

[56] Y. Zhang, Y. Zhang, S. K. Bose, and G. Shen, “Migration from fixed to flexible grid optical networks with sub-band virtual concatenation,” *IEEE/OSA Journal of Lightwave Technology,* vol. 35, no. 10, pp. 1752-1765, May 2017.

[57] A. Eira, J. Pedro, J. Pires, D. Fonseca, J. P. F. Palacios, V. Lopez, and S. Spaelter, “Defragmentation-based capacity enhancement for fixed to flexible-grid migration scenarios in DWDM networks,” *Proc., European Conference and Exhibition on Optical Communication (ECOC),* Sept. 2013.

[58] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, “Elastic optical networking: a new dawn for the optical layer?” *IEEE Communications Magazine,* vol. 50, no. 2, pp. s12-s20, Feb. 2012.

[59] M. Ghobadi, J. Gaudette, R. Mahajan, A. Phanishayee, B. Klinkers, and D. Kilper, “Evaluation of elastic modulation gains in Microsoft’s optical backbone in North America,” *Proc., Optical Fiber Communications Conference and Exhibition (OFC),* Mar. 2016.

[60] S. Yan, A. F. Beldachi, F. Qian, K. Kondepu, Y. Yan, C. Jackson, R. Nejabati, and D. Simeonidou, “Demonstration of real-time modulation-adaptable transmitter,” *Proc., European Conference on Optical Communication,* Sept. 2017.

[61] C. Rottondi, M. Tornatore, and G. Gavioli, “Optical ring metro networks with flexible grid and distance-adaptive optical coherent transceivers,” *Bell Labs Technical Journal,* vol. 18, no. 3, pp. 95-110, Dec. 2013.

[62] W. Chen, X. Yu, Y. Zhao, and J. Zhang, “Distance-adaptive routing, modulation, and spectrum assignment (DA-RMSA) algorithm based on signal overlap in elastic optical networks (EONs),” *Proc., International Conference on Optical Communications and Networks,* Feb. 2019.

[63] P. Wright, A. Lord, and S. Nicholas, “Comparison of optical spectrum utilization between flexgrid and fixed grid on a real network topology,” *Proc., Optical Fiber Communications Conference and Exhibition (OFC),* Mar. 2012.

[64] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, “On the Performance of Nyquist-WDM Terabit Superchannels Based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM Subcarriers,” *IEEE/OSA Journal of Lightwave Technology,* vol. 29, no. 1, pp. 53-61, Jan. 2011.

[65] R. Ramamurthy and B. Mukherjee, “Fixed-alternate routing and wavelength conversion in wavelength-routed optical networks,” *IEEE/ACM Transactions on Networking,* vol. 10, no. 3, pp. 351-367, June 2002.

[66] B. Mukherjee, “Optical WDM Networks,” New York, NY, USA, Springer, 2006.

[67] T. H. Cormen, “Introductions to Algorithms,” Cambridge, MA, USA, McGraw-Hill, 2003.

[68] L. Zhang, W. Lu, X. Zhou, and Z. Zhu, “Dynamic RMSA in spectrum-sliced elastic optical networks for high-throughput service provisioning,” *Proc., International Conference on Computing, Networking and Communications*, Jan. 2013.

[69] R. Wang and B. Mukherjee, “Spectrum management in heterogeneous bandwidth optical networks,” *Optical Switching and Networking,* vol. 11, pp. 83-91, Jan. 2014.

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

[71] ESnet, [online]. Available: http://es.net/engineering-services/the-network/. [Accessed: Feb. 19 2021]

[72] M. Cantono, R. Schmogrow, M. Newland, V. Vusirikala, and T. Hofmeister, “Opportunities and challenges of C+L transmission systems,” *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 5, pp. 1050-1060, Dec. 2020.

[73] A. Mitra, D. Ives, A. Lord, S. Savory, S. Kar, and P. Wright, “Network equipment and their procurement strategy for high capacity elastic optical networks,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 7, pp. A201-A211, July 2016.

[74] C. Meusburger, D. A. Schupke, and A. Lord, “Optimizing the migration of channels with higher bitrates,” *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 608-615, Feb. 2010.

[75] C. Meusburger, D. A. Schupke, and J. Eberspacher, "Multiperiod planning for optical networks—approaches based on cost optimization and limited budget," *Proc., IEEE International Conference on Communications (ICC)*, May 2008.

[76] B. Shariati, P. S. Khodashenas, J. M. R. Moscoso, S. B.-Ezra, D. Klonidis, F. Jiménez, L. Velasco, and I. Tomkos, "Investigation of mid-term network migration scenarios comparing multi-band and multi-fiber deployments," *Proc., Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2016.

[77] A. Ferrari, A. Napoli, J. K. Fischer, N. Costa, J. Pedro, N. Sambo, E. Pincemin, B. S.-Krombholz, and V. Curri, "Upgrade capacity scenarios enabled by multi-band optical systems," *Proc., International Conference on Transparent Optical Networks (ICTON)*, July 2019.

[78] A. Mitra, D. Semrau, N. Gahlawat, A. Srivastava, P. Bayvel, and A. Lord, "Effect of reduced link margins on C+L band elastic optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 10, pp. C86-C93, Oct. 2019.

[79] A. Mitra, A. Srivastava, P. Bayvel, and A. Lord, "Effect of channel launch power on fill margin in C+L band elastic optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 5, pp. 1032-1040, Mar. 2020.

[80] A. Ferrari, D. Pilori, E. Virgillito, and V. Curri, "Power control strategies in C+L optical line systems," *Proc., Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2019.

[81] D. Moniz, V. Lopez, and J. Pedro, "Design strategies exploiting C+L-band in networks with geographically-dependent fiber upgrade expenditures," *Proc., Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2020.

[82] Infinera, [Online]. Available: https://www.infinera.com/press-release/windstream-deploys-infinera-c-l-solution-sets-foundation-double-fiber-capacity. [Accessed: Feb. 19 2021]

[83] Corning, [Online]. Available: https://www.corning.com/media/worldwide/coc/ documents/Fiber/R_OFC_Infinera-2018.pdf. [Accessed: Feb. 19 2020]

[84] Ciena, [Online]. Available: https://www.ciena.com/insights/white-papers/a-model-for-business-case-analysis-of-c-l-band-betwork-architectures.html. [Accessed: Feb. 19 2020]

[85] D. Semrau, R. I. Killey, and P. Bayvel, "A closed-form approximation of the gaussian noise model in the presence of inter-channel stimulated raman scattering." *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 5, pp. 1924-1936, May 2019.

[86] D. Applegate and E. Cohen, "Making Routing Robust to Changing Traffic Demands: Algorithms and Evaluation," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1193-1206, Dec. 2006.

[87] N. Kourtellis, G. D. F. Morales, and F. Bonchi, “Scalable online betweenness centrality in evolving graphs,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2494-2506, Sept. 2015.

[88] I. Freire, C. Novaes, I. Almeida, E. Medeiros, M. Berg, and A. Klautau, “Clock synchronization algorithms over PTP-unaware networks: reproducible comparison using an FPGA testbed,” *IEEE Access*, vol. 9, pp. 20575-20601, Jan. 2021.

[89] Z. Idrees, J. Granados, Y. Sun, S. Latif, L. Gong, Z. Zou, and L. Zheng, “IEEE 1588 for clock synchronization in industrial IoT and related applications: a review on contributing technologies, protocols and enhancement methodologies,” *IEEE Access*, vol. 8, pp. 155660-155678, Aug. 2020.

[90] R. K. Jana, A. Mitra, A. Pradhan, K. Grattan, A. Srivastava, B. Mukherjee, and A. Lord, “When is operation over C + L bands more economical than multifiber for capacity upgrade of an optical backbone network?” *Proc., European Conference on Optical Communications (ECOC)*, Dec. 2020.