Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Provendo Múltiplas Transferências de Dados em Massa em Redes Ópticas Elásticas

Léia Sousa de Sousa

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Orientador
Prof. Dr. André Costa Drummond

Brasília
2016

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SSO725p

Sousa, Léia Sousa de
Provendo Múltiplas Transferências de Dados em Massa em Redes Ópticas Elásticas / Léia Sousa de Sousa; orientador André Costa Drummond. -- Brasília, 2016.
121 p.

Dissertação (Mestrado - Mestrado em Informática) - Universidade de Brasília, 2016.

1. Roteamento Ciente da Aplicação. 2. Transferências de Dados em Massa. 3. Redes Ópticas Elásticas. I. Drummond, André Costa, orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Provendo Múltiplas Transferências de Dados em Massa em Redes Ópticas Elásticas

Léia Sousa de Sousa

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Prof. Dr. André Costa Drummond (Orientador)
CIC/IE/UnB

Prof. Dr. Eduardo Adilio Pelinson Alchieri
CIC/IE/UnB

Prof. Dr. Paulo Roberto de Lira Gondim
ENE/FT/UnB

Prof.a Dr.a Célia Ghedini Ralha
Coordenadora do Programa de Pós-graduação em Informática

Brasília, 14 de Junho de 2016

Dedicatória

Aos meus pais, Ana e Raimundo.

Aos meus avós, Maria Rosa *(in memoriam)* e Juvenal *(in memoriam)*.

Agradecimentos

Agradeço ao meu orientador Prof. Dr. André Costa Drummond pelo incentivo e apoio conferidos a mim em toda a jornada do curso de mestrado e na realização deste trabalho.

Também gostaria de agradecer aos demais professores Programa de Pós-Graduação em Informática, deste Departamento de Ciência da Computação, bem como à Universidade de Brasília, por contribuírem fortemente na minha formação.

Aos amigos que fiz nesta instituição, com os quais aprendi imensamente, agradeço de coração pelas colaborações, incentivos, sugestões e toda atenção que recebi, em especial ao Lucas Rodrigues, Jeremias Gomes, Amanda Cristina, Paula Lima, Nilson Júnior, Felipe Rodopoulos, Henrique Domingues, Kaio Alexandre, Gustavo Sandri, Vera Ramalho e Daniel Souza.

Resumo

A tecnologia das redes ópticas elásticas, EON (do inglês, Elastic Optical Networks), fornece as condições necessárias para o atendimento de múltiplas transferências de dados em massa, MBDT (do inglês, Multiple Bulk Data Transfer), por ser capaz de alocar circuitos ópticos dinâmicos e flexíveis adaptando-se às demandas. Este trabalho propõe soluções de alocação de circuitos em redes ópticas elásticas que são cientes das idiossincrasias das MBDTs em um cenário de ressincronização de bases de dados em redes que interconectam centros de dados. Os resultados obtidos mostram que soluções cientes da aplicação estabelecem três vezes mais atendimentos bem sucedidos de MBDTs em comparação com soluções convencionais. Além disso, o escalonamento dinâmico de requisições de aplicações dessa natureza permite que a aceitação continue ocorrendo quando o tráfego aumenta.

Palavras-chave: Roteamento Ciente da Aplicação, Transferências de Dados em Massa, Redes Ópticas Elásticas

Abstract

The Elastic Optical Networks technology (EON) provides the necessary infrastructure to cope with Multiple Bulk Data Transfer (MBDT), being able to allocate dynamic and flexible optical circuits adapting itself to the demands. This work proposes solutions to circuit allocation in EON that are aware of the idiosyncrasies of MBDTs in a scenario of databases resynchronization in networks which interconnect data centers. The results obtained shows that the application-aware solutions establish up to three times more successful MBDTs calls if compared to conventional solutions.

Keywords: Aplication-Aware Routing, Bulk Data Transfer, Elastic Optical Network

Sumário

Lista de Abreviaturas e Siglas

AA Application-Aware.

AMRA Adaptative Modulation Routing Assignment.

BBR Bandwidth Blocking Ratio.

BDT Bulk Data Transfer.

BER Bit Error Rate.

BP Blocking Probability.

BPSK Binary Phase-Shift Keying.

BV-WSS Bandwidth-Variable Wavelength Selective Switch.

BV-WXC Bandwidth-Variable Wavelength Cross Connect.

BVT Bandwidth Variable Transceiver.

CAPEX Capital Expenditures.

CD Chromatic Dispersion.

CDN Content Distribution Network.

CLD Cross-Layer Design.

D2C Data Center to Client.

D2D Data Center-to-Data Center.

DCN Data Center Network.

DMS Distribution Management System.

DS Distributed Systems.

DVD Digital Versatile Disc.

DWDM Dense Wavelength Division Multiplexing.

E2E End To End.

EDF Earliest Deadline First.

EEM Energy Efficient Manycasting.

EON Elastic Optical Networks.

FEC Forward Error Correction.

FIFO First in, First out.

FS Frequency Spectrum.

GMPLS Generalized Multi-Protocol Label Switching.

GRIPhoN Globally Reconfigurable Intelligent Photonic Network.

IA Impairment-Aware.

IaaS Infrastructure as a Service.

ICD Inter Centro de Dados.

IETF Internet Engineering Task Force.

ILP Integer Linear Programming.

IP Internet Protocol.

IPTV Internet Protocol Television.

ISI Inter-Symbol Interference.

ISP Internet Service Provider.

ITU-T Telecommunication Standardization Sector of the International Telecommunication Union.

KSP K-Shortest Paths.

LASER Light Amplification by Stimulated Emission of Radiation.

LCoS Liquid Crystal on Silicon.

LPF Longest Path First.

LUM Link Utilization Metric.

MBDT Multiple Bulk Data Transfer.

MILP Mixed integer linear programming.

MMF Max-Min Fairness.

MSF Most Subcarriers First.

NP Non-Deterministic Polynomial time.

NSFNET National Science Foundation Network.

O-OFDM Optical-Orthogonal Frequency-Division Multiplexing.

OFDM Orthogonal Frequency-Division Multiplexing.

ONS Optical Network Simulator.

OPEX Operating Expenses.

OSI Open Systems Interconnection.

OTN Optical Transport Network.

OXC Optical Cross-Connects.

PCE Path Computation Element.

PLI Physical Layer Impairments.

PMD Polarization Mode Dispersion.

QAM Quadrature Amplitude Modulation.

QoS Quality of Service.

QoT Quality of Transmission.

QPSK Binary Phase-Shift Keying.

RAM Random Access Memory.

RMLSA Routing, Modulation Level and Spectrum Allocation.

RSA Routing and Spectrum Assignment.

RWA Routing and Wavelength Assignment.

SaaS Software as a Service.

SDN Software-Defined Networking.

SDOT Software-Defined Optical Transmission.

SGDD Sistema de Gerenciamento de Dados Distribuídos.

SJF Shortest Job First.

SLICE Spectrum-Sliced Elastic Optical Path Network.

SMF Single-Mode Fiber.

SMR State Machine Replication.

SnF Store-and-Forward.

SNR Signal-to Noise Ratio.

SRR Successful Resynchronization Rate.

TCP Transmission Control Protocol.

TI Tecnologia da Informação.

UCC User Created Content.

UGC User-Generated Content.

ULAF Ultra-Large Effective Area Fiber.

URL Uniform Resource Locator.

VDC Virtual Data Center.

VM Virtual Machine.

WDM Wavelength Division Multiplexing.

Capítulo 1

Introdução

Múltiplas Transferências de Dados em Massa (do inglês, *Multiple Bulk Data Transfer* - MBDT) são operações tolerantes a atrasos que transportam volumosas massas de dados, na forma de requisições orientadas a dados da ordem de centenas de *gigabytes* de dados ou superior, em redes inter centros de dados (ICD) geo-distribuídas, e que requerem grande capacidade de largura de banda para atender esse serviço [54]. Esse tipo de transporte de dados é empregado tanto em operações de computação intensiva [67], que realizam processamento paralelo e escalável de forma distribuída, quanto em gerenciamento, armazenamento e replicação de dados, que são as principais atividades do emergente paradigma de Computação em Nuvem [66].

Uma das principais características das MBDTs é o grau de tolerância a atrasos que permite executar uma transferência por tempo prolongado ou paralisar uma transferência de dados e reiniciá-la em outro momento, desde que suas restrições de tempo e recurso sejam atendidas [61]. Em qualquer caso, o desafio é encontrar uma fórmula para a alocação de banda suficiente e que atenda às restrições dadas. Geralmente essa alocação é feita por meio de reserva antecipada [68], definindo-se a máxima capacidade disponível ou a mínima taxa de transmissão dentro do prazo previamente especificado [11, 65]. Entretanto, existem outros tipos de tráfego utilizando a mesma infraestrutura e disputando esses mesmos recursos, embora apenas o volume de tráfego trocado entre os CDs já correspondam a 45% do volume de tráfego total nas redes de *backbone* [70].

Como os centros de dados (CDs) hospedam uma variada gama de aplicações, existe atualmente um tráfego diversificado em trânsito na rede ICD. O conjunto desse tráfego são fluxos entre CD e clientes (*Data Center to Client* - D2C) ou apenas entre CDs (*Data Center to Data Center* - D2D). Contudo, as redes não distinguem os tráfegos em curso e aplicações em execução, tampouco existem políticas padronizadas para fazer a diferenciação de todos esses tipos de tráfego, e assim os provedores é que definem os perfis de tráfego e políticas de alocação de banda, de acordo com prioridades pré-estabelecidas

[20, 68]. Tais prioridades são resultantes da classificação do tráfego em interativo, elástico e *background*, sendo que cada um possui diferentes requisitos de desempenho. A principal distinção entre essas classes é o quão sensível ao atraso um tráfego é [98].

As replicações são especialmente relevantes neste cenário porque precisam ocorrer periodicamente para garantir que o sistema seja inteiramente tolerante a falhas de qualquer natureza [21]. Em cada um desses períodos, grandes volumes de dados são transferidos entre os centros de dados espalhados por todo o planeta. Uma consequência imediata desse esforço é o melhor desempenho fim-a-fim percebido na rede. Um pedido de um usuário nunca é rejeitado por causa da sobrecarga de pedidos em um mesmo local [2]. Como existem várias réplicas dispostas com um mesmo conteúdo, o atendimento dos pedidos é distribuído e balanceado entre os centros detentores do dado de interesse. Além disso, um pedido geralmente é direcionado ao centro de dados mais próximo do solicitante, apesar de existirem outros centros de dados igualmente hábeis a respondê-lo. Por esse motivo, as respostas são retornadas cada vez mais com menor atraso [80, 4].

Atualmente, as MBDTs ocorrem com duração de horas a dias [54]. O crescimento gradual do tráfego tem motivado a busca por formas de uso eficiente dos recursos combinados com redução nos custos e diminuição do tempo de execução dessas operações. Estima-se que o tráfego IP (*Internet Protocol*) alcançará o patamar de $7,7$ *zettabytes* ao ano em 2017 [2], quando provavelmente dois terços de todos os dados gerados no mundo serão processados em nuvem. No atual panorama tecnológico, a proliferação de conteúdos gerados pelo usuário (*User-Generated Content* - UGC ou *User Created Content* - UCC) [63] tem contribuído para o aumento do volume de tráfego bem como pressionado a redução do tempo de trânsito, visto que a grande quantidade de dados produzida como fotos, vídeos e conteúdos de redes sociais, é acessada de diferentes partes da rede e com relativa frequência.

Grandes companhias provedoras de serviço de computação em nuvem e armazenamento de dados sob demanda, como *Google*, *Facebook*,*Amazon*, *Yahoo!*, entre outras, estão observando as tendências de crescimento do tráfego com vistas a ampliar suas capacidades de transporte ICD para garantir a longo prazo as transferências de dados em massa [64] visto que os recursos de largura de banda atualmente disponíveis podem estar alcançando o seu limite técnico de crescimento [91]. A capacidade total de $100Tbps$ por fibra já foi alcançada em experimentos em laboratório com a tecnologia atualmente existente. A intenção é ampliar esse número para além de $1Pbps$ a fim de aumentar a capacidade de transporte [69].

A escassez de largura de banda e o custo dos sistemas ópticos de transmissão justificam que aplicações corporativas sejam executadas em um único CD, no entanto, os riscos e ameaças advindos dessa decisão de projeto pode acarretar em severas perdas de dados

de importância vital para a continuidade dos negócios. Problemas como perda de conectividade da rede, catástrofes e invasões maliciosas implicam na necessidade de sistemas altamente redundantes em todos os níveis de sua infraestrutura, especialmente no que diz respeito ao armazenamento dos dados, para que tenham vários desses dados replicados em múltiplos computadores ou unidades de centros de dados [41]. A redundância é executada tendo dois ou mais centros de dados sincronizados, independentes e separados fisicamente ao redor do mundo, entre os quais são implantadas redes de longa distância para interligação, comumente sendo a própria *Internet* pública [38] e em alguns casos, redes particulares de elevado custo de construção e manutenção [40].

É devido aos altos custos que diferentes técnicas de transferência de dados são combinadas, analisando-se uma série de fatores como o custo da largura de banda em diversas partes do mundo, o fuso-horário nas localizações dos centros de dados, considerando-se os momentos de maior e menor utilização da rede, as distâncias físicas de interligação que podem levar a decisões de suspensões momentâneas dos serviços de transferência em diversos pontos do percurso até que seja alcançado o CD final, entre outros fatores ponderantes. Todas essas soluções possuem uma característica em comum: resolvem o problema de gerenciar a alocação de largura de banda para as MBDTs, muito embora o façam de maneira estática, ou seja, os recursos a serem atribuídos a um dado pedido são reservados com base nas previsões de volume de tráfego.

A técnica de armazenamento de dados em nós intermediários da rede, antes do encaminhamento ao destino final (*Store-and-Forward* - SnF), é empregada para transferências de dados por distâncias relativamente grandes, com escalonamento de pedidos [11], ou quando a rede tende a atingir o pico de tráfego de outros serviços, quando as MBDTs podem ser atrasadas devido ao seu menor nível de prioridade e tráfego elástico [71]. Recentemente, as transferências foram feitas mediante a gravação em diversas mídias como discos rígidos e DVDs (*Digital Versatile Discs*) para então serem movimentados por meio de serviços postais [54]. No entanto, o volume de dados, mesmo aproveitando as técnicas de compressão, exigia grande quantidade de equipamentos para gravação, além de altos custos com serviço de entrega e tempo desvantajoso. A transmissão fim-a-fim (*End To End* - E2E) é comumente empregada quando não há grandes variações nos custos e impactos no tráfego tanto para o CD transmissor, quanto para o CD receptor [54]. Em todas as modalidades envolvidas existe a preocupação quanto a ocupação da rede que é compartilhada com diversas outras aplicações.

Nesse contexto, a realização de ressincronização entre CDs é de grande importância para manter a consistência dos dados, a disponibilidade geográfica e a tolerância a faltas, e por este motivo, MBDT para replicação de dados precisa encontrar outras maneiras mais ágeis e de menor custo para serem realizadas.

A tecnologia de Multiplexação Densa por Divisão de Comprimento de Onda (*Dense Wavelength Division Multiplexing* - DWDM) [17] é utilizada hoje nas redes de *backbone* que sustentam essas transações. Ela garante a combinação de múltiplos sinais ópticos, ou canais ópticos, em diferentes comprimentos de onda e transmissão simultânea em uma mesma fibra óptica sob altas velocidades e capacidade de banda de $10Gbps$, $40Gbps$, e mais recentemente $100Gbps$ [39]. Essa tecnologia usa uma grade de espaçamento fixo fracionada em canais uniformes de $50GHz$ ou $100GHz$, desenhando a configuração de como as demandas são acomodadas, preenchendo ou não o canal completamente [17]. Quando uma conexão é estabelecida toda sua capacidade de largura de banda é alocada para o tráfego que a requisitou, acarretando em desperdício caso este fluxo não seja suficiente para ocupar todo o canal alocado. Esse desperdício de largura de banda pode ser evitado com a implementação de espaçamento flexível, de tal maneira que as demandas de transmissão possam adaptar-se a ele, aproveitando-se melhor dos recursos disponíveis.

Assim, as redes ópticas elásticas (*Elastic Optical Network* - EONs) têm emergido como forte aposta para o futuro, já que permitem que a largura de banda possa contrair-se ou expandir-se para acomodar o fluxo, sendo possíveis taxas de transmissão de até $400Gbps$ ou $1Tbps$ sob diversos formatos de modulação [106]. Dessa forma, os recursos disponíveis são utilizados sob demanda podendo atender a muito mais requisições se comparado à atual tecnologia WDM. Outra vantagem dessa nova tecnologia é a possibilidade de enviar mais dados com o mesmo número de portadoras de sinais ópticos, já que os formatos de modulação desses sinais podem ser definidos de acordo com a distância que esses sinais percorrerão, reduzindo as perdas.

De olho nesse novo horizonte, este trabalho traz as aplicações MBDTs para serem executadas no ambiente de EON, onde é possível alocar recursos de maneira elástica e transportar mais dados na mesma fatia de espectro. Tais possibilidades abrem precedentes para a redução no tempo de transporte desses dados e possível alocação dinâmica de recursos para esse tipo de tráfego, sem prejudicar as demais aplicações que compartilham tais recursos. Indo mais adiante, este trabalho também aborda o paradigma de projeto em camada cruzada (*Cross Layer Design* - CLD) através do qual uma camada da arquitetura de rede adquire visão sobre o que acontece em outras camadas. Com particular interesse em fazer com que a camada de rede enxergue e reconheça as aplicações MBDTs, soluções de roteamento e alocação de recursos com base nas características dessas aplicações são propostas como forma de conduzir o campo de pesquisa das soluções de roteamento ciente da aplicação, tradicionais nas áreas de Redes Definidas por *Software* (*Software Defined Networks* - SDN) [113, 20], para o campo das redes ópticas.

1.1 Motivação

A replicação de conteúdos é uma forma de garantir a sobrevivência da rede diante da ocorrência de falhas ou desastres, ou simplesmente para manter a consistência e disponibilidade aos usuários. As ressincronizações ocorrem nos casos em que um CD volta a compor a rede depois de um tempo indisponível, estando assim com o estado desatualizado. Para restabelecer o sincronismo é necessário receber as atualizações de várias réplicas provenientes de outros CDs.

Nesse processo de encaminhamento de atualizações, aplicações MBDTs são acionadas para realizar as transferências dos dados entre as réplicas relacionadas a um mesmo conjunto de replicação. Esse tipo de tráfego é um dos grandes responsáveis por consumir enormes capacidades de largura de banda da rede, e por esse motivo, as transferências ocorrem de maneira estática, com reserva antecipada de banda em específicos períodos de tempo da rede, onde o nível de utilização por outros tipos de aplicações é menor. Nos horários de pico de utilização, os provedores de serviços precisam alugar hospedagem em outros CDs para guardar momentaneamente os dados em transferência e evitar sobrecarga de utilização da largura de banda.

Com o advento da EON será possível transportar mais dados utilizando a mesma porção de espectro que é alocada hoje. Com tal aumento de capacidade, é possível vislumbrar alocações dinâmicas para as aplicações MBDTs, de maneira que as despesas com aluguel de armazenamento sejam eliminadas e o tempo de transferência de todos os dados seja menor.

A dificuldade encontrada hoje em garantir a prestação de serviços de transporte de dados a longo prazo é devida ao crescente aumento do volume de tráfego a ser combinado com as atuais capacidades dos canais de transmissão. No passado, foi possível escalar tal capacidade sem aumentar a largura de banda de cada um desses canais. No entanto, a tecnologia WDM requer que os canais sejam espaçados em uma grade espectral em grossa granularidade. Com a largura de banda quase exaurida, é necessário explorar outras tecnologias que utilizem granularidades mais finas, capazes de empregar a largura de banda de forma mais eficiente.

Nesse contexto de mudança tecnológica também é necessário viabilizar soluções com maior poder de decisão sobre recursos da rede e maior visão sobre as aplicações que deles fazem uso. Com as soluções de roteamento tradicionais, que não possuem conhecimento sobre a entidade que solicitou atendimento, todas as aplicações são tratadas da mesma maneira, aparecendo com características e necessidades padrões. Esse ponto de vista acaba limitando o desempenho dessas aplicações por parte da rede, que mesmo dispondo de recursos suficientes para o devido atendimento, acaba resultando em bloqueio devido à peculiaridades de cada aplicação que não foram levadas em consideração, como por

exemplo, aspectos de flexibilidade quanto ao tempo de espera, que varia de aplicação para aplicação, mesmo entre aquelas que fazem parte da mesma classe de tráfego.

1.2 Objetivos

O objetivo geral deste trabalho é propor uma solução para a ressincronização entre CDs com roteamento ciente da aplicação em EON, capaz de realizar transferências de dados em massa. Para isso, são definidos alguns objetivos específicos:

- Levantamento de requisitos da nova tecnologia EON;
- Implementação de um simulador em ambiente EON para resolver o problema de Roteamento e Alocação de Espectro (*Routing and Spectrum Allocation* - RSA) e lidar com requisições orientadas a dados;
- Implementação de soluções cientes da aplicação para o problema RSA;
- Simulação, análise e comparação de desempenho entre soluções cientes e soluções convencionais, ou seja, que não são cientes da aplicação.

1.3 Contribuições

As principais contribuições desta Dissertação são as seguintes:

1. Levantamento do estado da arte de soluções para MBDTs (Capítulo 3). As estratégias mais comuns identificadas incluem encaminhamento direto de dados entre origem e destino ou encaminhamentos com pausas, onde os dados podem ser armazenados temporariamente até que a rede disponha de banda disponível para a continuação da transferência até o destino final.
2. Desenvolvimento do simulador para redes ópticas elásticas capaz de lidar com requisições orientadas a dados, que oferecem liberdade quanto a largura de banda definida para a transmissão e suportam atrasos, mas com rígidas exigências de cumprimento do tempo de transmissão.
3. Proposição de algoritmos de roteamento e alocação de banda dinâmicos cientes da aplicação. As soluções propostas nos Capítulos 4 e 5 realizam transferências de dados em uma rede que implementa a tecnologia de transporte EON.

Os resultados parciais de tais contribuições foram:

- Autoria em publicação de artigo apresentado no $XXXIV$ Simpósio Brasileiro de Rede de Computadores e Sistemas Distribuídos - SBRC 2016, intitulado "Roteamento e Alocação de Espectro Ciente da Aplicação em Redes Ópticas Elásticas".

- Co-autoria em trabalho apresentado no Salão de Ferramentas do $XXXIV$ Simpósio Brasileiro de Rede de Computadores e Sistemas Distribuídos - SBRC 2016, intitulado "ONS: Simulador de Eventos Discretos para Redes Ópticas WDM/EON".

1.4 Organização do Documento

Este trabalho é constituído de 6 capítulos que fundamentam a proposta de solução para o problema MBDT em EON para ressincronização de CDs em uma rede geo-distribuída.

O Capítulo 2 contextualiza a tecnologia de transporte óptico EON trazendo as principais definições para compreender o seu funcionamento, as mudanças tecnológicas diante do sistema implantado hoje nos *backbones* da *Internet*, o sistema de modulação multinível por trás dessas grandes mudanças e os principais problemas ligados às buscas de caminhos ópticos e fatias suficientes do espectro. O capítulo também destaca o ambiente ICD, que lida com tais demandas de alta capacidade e podem enfrentar problemas de escalabilidade da capacidade com a tecnologia de transporte que utilizam. O paradigma CLD, amplamente abordado nas redes e comunicações *wireless*, conclui o capítulo fornecendo um aparato teórico que permite ser empregado na extração de informações dos aspectos da camada de aplicação e utilizá-las na camada de rede.

O Capítulo 3 apresenta o atual estado da arte dessas tecnologias. São informadas as principais propostas da literatura para tratar de EON em 4 vertentes principais: os dispositivos e equipamentos que implementam a característica elástica da tecnologia, as novas diretrizes e padronizações que podem surgir para implantação de uma abordagem formal e pragmática que, por sua vez, permitirá a incorporação da EON na indústria, as novas exigências para gerenciamento e controle da rede e os problemas fundamentais de roteamento e alocação de espectro óptico (RSA) e sua versão que também escolhe o nível de modulação mais adequada (*Routing, Modulation Level and Spectrum Allocation* - RMLSA), uma generalização do problema RSA. Além disso, são destacadas as propostas elaboradas para tratar do problema MBDT que são identificadas na literatura e tentam reduzir os custos de transporte ao mesmo tempo em que procuram melhorar o atendimento dos seus serviços.

No Capítulo 4 será apresentada uma solução de roteamento e atribuição de espectro ciente da aplicação MBDT em EON (*Application-Aware Routing and Spectrum Allocation* - AA-RSA), que alcança um patamar de ressincronizações bem sucedidas em torno de 32% maiores do que seria possível com um algoritmo de roteamento convencional. Além

disso, os resultados em termos de bloqueio de banda passante são nitidamente menores, o que é alcançado com um esquema de roteamento mais complexo, capaz de buscar e selecionar o máximo de possíveis demandas a serem atendidas.

Já o Capítulo 5 estende o problema do Capítulo 4 para um cenário com mais de uma aplicação de transferência de dados disputando banda para atendimento. Diversas possibilidades de aprovisionamento de recursos são desenhadas em busca de favorecer as ressincronizações sem, ao mesmo tempo, elevar o bloqueio de outros tipos de aplicações. Para aumentar as chances de uma solicitação de ressincronização ser atendida, uma janela de escalonamento de requisições ciente do prazo da chamada é empregada em algumas alternativas de solução. Os resultados de tais mecanismos mostram que, quando essa janela trata com prioridade as chamadas do tipo MBDT, uma maior taxa de aceitação é alcançada. E por fim, o Capítulo 6 apresenta as considerações finais do trabalho e aponta as possíveis propostas futuras de investigação.

Capítulo 2

Conceitos Básicos

Neste capítulo é apresentada a nova tecnologia das redes ópticas de transporte (*Optical Transport Network* - OTN), a rede óptica elástica (*Elastic Optical Network* - EON) que definirá o ambiente de rede a ser empregado na solução do problema das múltiplas transferências de dados em massa (*Multiple Bulk Data Transfers* - MBDTs). São apresentados o problema de roteamento e alocação de espectro (*Routing and Spectrum Allocation* - RSA), bem como sua generalização, o problema de roteamento e alocação de nível de modulação e espectro (*Routing, Modulation Level and Spectrum Allocation* - RMLSA). A organização da rede de CD também é descrita para um melhor entendimento das operações de transferências de dados geo-distribuídas, o que é reforçado por conceitos referentes a sistemas distribuídos e replicação de dados. Ao final, é apresentado o Projeto de Camadas Cruzadas (*Cross-Layer Design* - CLD) que ampliará a visão do roteamento para além da camada de rede.

2.1 Rede Óptica Elástica (EON)

O aumento da capacidade de transmissão dos sistemas ópticos pode ser alcançado de três maneiras possíveis: instalando-se novas fibras ópticas e equipamentos de transmissão, aumentando o número de canais de transmissão disponíveis nos sistemas ópticos ou aumentando a taxa de transmissão desses canais nesses sistemas ópticos. A primeira possibilidade implica no aumento dos custos de capital (*CAPital EXpenditure* - CAPEX) e operacional (*OPerational EXpenditure* - OPEX) embora possam ocorrer instantaneamente, e as outras duas soluções podem ser de menor custo e melhor eficiência, em contrapartida, implicam em intensas pesquisas a longo prazo [45].

A atual tecnologia DWDM utiliza os comprimentos de onda da banda C (hachurado na Figura 2.1) do espectro eletromagnético como sua grade de espectro óptico padronizada pela ITU-T (*Telecommunication Standardization Sector of the International Telecommu-*

nications Union). A banda C é dividida em canais de espaçamentos fixos e no futuro, seria necessário a busca por novas bandas de transmissão para aumentar o número de canais, expandindo a banda C, de $1530nm$ a $1565nm$, para outras faixas do espectro eletromagnético, como a banda L, por exemplo [39].

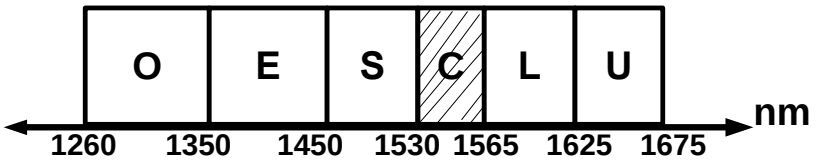

Figura 2.1: Grade de comprimentos de onda padronizada pela ITU-T. Em destaque, a banda C (hachurada) utilizada nos sistemas DWDM.

Obedecendo a esse espaçamento uniforme, a criação e transmissão desses comprimentos de onda é feita por amplificação da luz por emissão estimulada de radiação (*Light Amplification through Stimulated Emission of Radiation* - LASERs) [74], tecnologia amplamente utilizada nas redes DWDM, visto que para cada comprimento de onda específico é exigido um desses equipamentos [7]. A multiplexação de comprimentos de onda ocorre nos *transponders* da rede, elemento que adapta o sinal vindo da rede cliente em um sinal viável para uso na rede óptica, similarmente na direção inversa, o sinal óptico é adaptado para ser enviado à rede cliente, ambos tipicamente feitos através de uma conversão óptica-elétrico-óptica (O/E/O) [31, 7].

A EON [83] propõe modificações no funcionamento desses LASERs e *transponders* para explorar melhor o espectro óptico, dando origem à moderna arquitetura da rede de caminhos ópticos elásticos com espectro fatiado (*Spectrum-Sliced Elastic Optical Path Network* - SLICE) [45], mostrada na Figura 2.2.

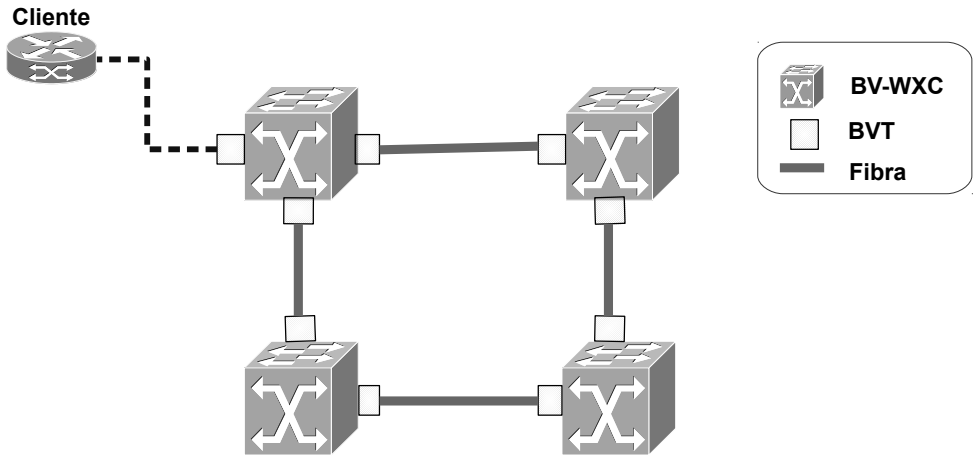

Figura 2.2: Representação da arquitetura SLICE.

Os *transponders* elásticos ou *transponders* com largura de banda variável (*Bandwidth Variable Transponders* - BVTs), Figura 2.2, recebem o sinal emitido pela rede cliente e o converte para ser comutado no domínio óptico com apenas a largura de banda suficiente. Cada BVT pode implementar vários *transceivers* virtuais, daí a justificativa para o conceito elástico da tecnologia. A capacidade de compartilhamento do BVT reduz o desperdício de recursos do espectro óptico alocando apenas a porção suficiente para atender a uma demanda [22]. Os BVTs estão conectados aos comutadores ópticos de largura de banda variável (*Bandwidth-Variable Wavelength Cross-Connects* - BV-WXC) que são equipamentos responsáveis por estabelecer uma conexão fim-a fim com o correspondente espectro para criar um caminho óptico de tamanho apropriado ao que foi solicitado pela requisição de conexão [90]. A essa porção do espectro definida em cada caminho óptico dá-se o nome de janela de encaminhamento [106].

Dentro dos BV-WXC existem numerosos comutadores seletivos em comprimento de onda (*Bandwidth-Variable Wavelength Selective Switch* - BV-WSS), mostrados na Figura 2.3 [69]. Cada BV-WSS recebe a luz emitida através da fibra óptica conectada ao BV-WXC e a divide em componentes espectrais usando um elemento dispersivo de espectro, que na sequência são repassados para um arranjo de espelhos unidimensionais, e em seguida, são redirecionados para a porta de saída, isto é, para outra fibra ligada a outro nó [45]. O BV-WSS pode ser implementado por dispositivos de cristal líquido sobre silício (*Liquid Crystal on Silicon* - LCoS), que é uma tecnologia já existente. Com o LCoS é possível ajustar dinamicamente a capacidade de cada enlace, filtrando apenas o espectro de frequência correspondente ao dado que é transferido para uma outra porta de saída apropriada [47].

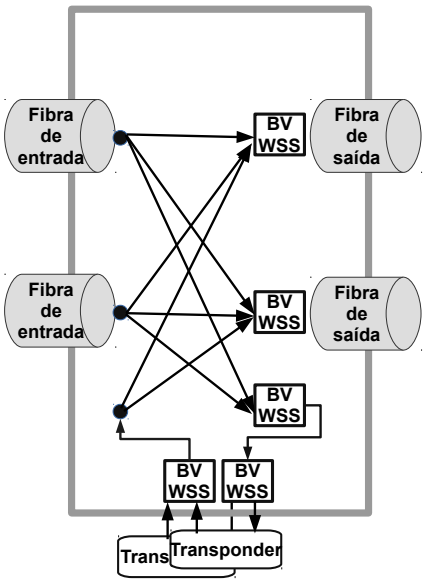

Figura 2.3: Modelo de um BV-WSS[69].

Para a operação dessa arquitetura, uma nova divisão de grade é proposta, utilizando a mesma banda C, mas com espaçamentos bem menores, denominadas *slots* de frequência ou fatias da frequência (*Frequency Slices* - FS). A alocação de recursos do espectro para uma requisição é feita designando-se parte dessas fatias, que assim passam a formar um canal [83]. A Figura 2.4 representa a grade fixa e a míni grade fixa. A grade fixa possui canais uniformes de grossa granularidade, tendo uma largura de espectro de $50GHz$ [27]. Por outro lado, a mini grade fixa possui espaçamentos menores, com canais de menor granularidade, cuja largura espectral é de $12.5GHz$ [6, 83]. Nas duas grades foram configuradas conexões de $40Gbps$, representadas pela parte hachurada nas duas figuras. Percebe-se que essa taxa de dados pode ser alocada em uma porção menor do espectro. A tecnologia EON pode transmitir a mesma quantidade de dados condensando o número de *bits* em um número menor de FS. Cabe ressaltar que, para alocar banda para transmissão em qualquer uma dessas grades, é necessário designar como banda de guarda os canais imediatamente anterior e imediatamente posterior ao canal ocupado. Com isso, o estabelecimento do circuito óptico para atender uma demanda de $40Gbps$ tomaria um canal de $50GHz$ para receber essa demanda e mais $11GHz$ [79] como banda de guarda bilateral nos dois canais adjacentes no exemplo da grade fixa. Em contrapartida, se o mesmo atendimento é feito na mini grade fixa apenas 4 fatias são necessárias, o que totalizaria uma largura espectral de $50GHz$.

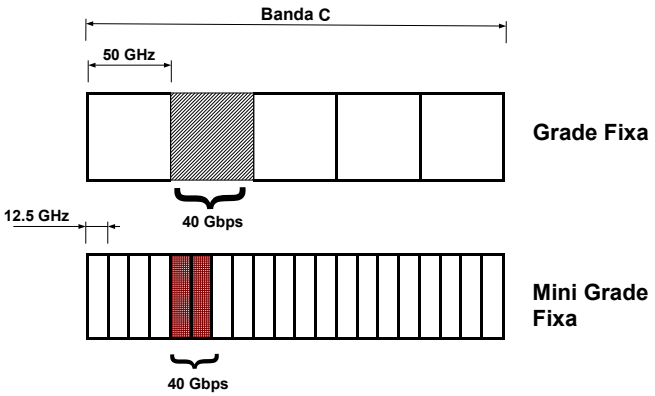

Figura 2.4: Grade fixa e mini grade fixa do espectro óptico.

Mas, só é possível concentrar tantos *bits* por símbolo em uma menor porção espectral devido aos BVTs implementarem uma versão óptica do sistema de modulação de Multiplexação por Divisão de Freqüências. Ortogonais (*Orthogonal Frequency Division Multiplexing* - OFDM) [57, 58], que permite definir a quantidade de *bits* ideal para diferentes distâncias de transmissão, de maneira que as perdas de sinais sejam mínimas.

2.1.1 Multiplexação por Divisão de Frequência Ortogonal Óptica (O-OFDM)

A versão óptica da OFDM tem sido proposta como uma forte candidata para garantir a escalabilidade nos sistemas ópticos de transmissão, através do transporte de dados com mais alta velocidade, dividindo-os em múltiplos canais paralelos (ou múltiplas portadoras) de baixa velocidade [106]. Esse sistema de modulação é efetivo em combater interferências inter-símbolos (*Inter-Symbols Interferences* - ISIs), é mais resistente ao enfraquecimento do sinal durante o transporte dos símbolos, além de possibilitar que diferentes quantidades de *bits* sejam transportados nesses símbolos [17].

A largura de banda é ocupada por uma constelação de pontos (Figura 2.5), usados para plotar várias combinações diferentes de níveis de modulação. Uma constelação é um conjunto de pontos dispostos no plano. Um conjunto de constelações é um sistema de modulação. Os pontos são os *bits* que transportam a informação. Esses *bits* são transmitidos em um único símbolo ou *slot* de tempo. Aumentar o número de pontos significa aumentar o nível de modulação [8]. Como o dado a ser transmitido constitui uma palavra de comprimento b *bits*, o modulador utiliza esses *bits* para formar ondas, onde para cada constelação (ou modulação) são produzidas $m = 2^b$ formas de ondas (ou

portadoras de sinais). As ondas geradas são ortogonalmente posicionadas umas às outras e os dados são simultaneamente transmitidos [8, 82].

Na Figura 2.5 são representados alguns níveis de modulações, ou constelações, que compõem o sistema O-OFDM. Cada subportadora pode ser modulada individualmente usando alguma das seguintes modulações:

- um *bit* por símbolo binário ($b = 1$) é transportado no formato BPSK (*Binary Phase Shift Keying*), produzindo assim $m = 2^1 = 2$, ou 2 formas de onda;
- ou dois *bits* por símbolo binário ($b = 2$) são transportados no formato QPSK (*Quadrature Phase Shift Keying*), sendo produzidas $m = 2^2 = 4$, 4 formas de onda;
- ou três *bits* por símbolo binário ($b = 3$) são transportados no formato 8QAM (*Quadrature Amplitude Modulation*), produzindo $m = 2^3 = 8$, ou seja, 8 formas de onda;
- ou ainda, quatro *bits* por símbolo binário ($b = 4$) são transportados no formato 16QAM, produzindo $m = 2^4 = 16$, de 16 formas de onda;

Existem outras possibilidades ainda mais densas de modulação, e consequentemente mais eficientes, todas viáveis a partir de um único BVT. Essa densidade de símbolos da constelação, devido a sua grande proximidade entre símbolos, pode acarretar em erros de detecção de sinal ao chegar no receptor ou mesmo erros de interpretação, por esse motivo o projeto da EON tem inicialmente elencando apenas esses 4 principais níveis de modulação, enquanto planeja avanços em níveis mais densos [8, 81].

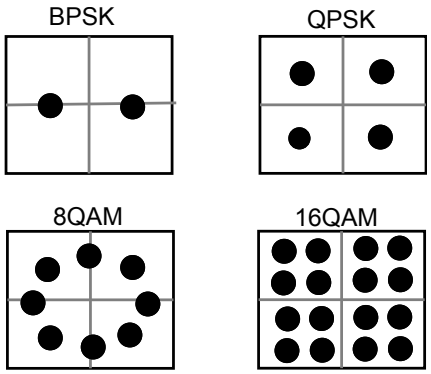

Figura 2.5: Exemplos de modulações (constelações) com 1, 2, 3 e 4 *bits* por símbolo.

O número e tamanho das subportadoras também variam de acordo com as diferentes modulações, de forma que o espectro óptico é fatiado em subportadoras ou *slots* com granularidade mínima de 12.5 GHz de frequência, conforme recomendação da ITU-T, e

futuramente 6.25 GHz, permitindo um grau de flexibilidade tanto no nível de modulação quanto no modo de fatiar esse espectro [13, 106].

As modulações O-OFDM são menos sensíveis à interferência de sinais entre os canais. Essa interferência geralmente ocorre quando muitos sinais ópticos viajam em diferentes frequências e velocidades, podendo levar a atrasos na sua chegada, bem como invasão de outros canais adjacentes ao seu. Além disso, maiores distâncias de transmissão passam a ser alcançadas, o que se traduz em menos custos operacionais com infraestrutura e dispensa o uso de parte dos equipamentos regeneradores de sinal [106]. A escolha de qualquer nível de modulação leva em conta a qualidade da transmissão (*Quality of Transmission* - QoT), fator cujo valor deve ser o mais alto possível. Esse fator determina o formato mais adequado para que o sinal viaje a uma distância segura para a preservação do dado, e assim, caminhos mais curtos tendem a empregar modulaçãos de mais alto nível, bem como caminhos mais longos são atendidos com modulações de mais baixo nível [22]. Um fator QoT muito baixo indica que o sinal necessita de regeneração.

Em [56] são propostos limites de alcances teóricos para os níveis de modulação dos sinais ópticos: sinais modulados em BPSK tem alcance de até 2500km e capacidade de $12,5Gbps$, ocupando 1 FS; com QPSK chega até a 1200km com 25$Gbps$ de capacidade por *slot*, ocupando 2 FS; usando a modulação 8-QAM o alcance é de 625km com capacidade de $37,5Gbps$ sendo que 3 FS são ocupados; e os sinais modulados em 16-QAM têm alcance de no máximo 625km com 50$Gbps$ de capacidade, sendo que 4 FS precisam ser ocupados.

Em [97] é definida a lei da meia distância, na qual o máximo alcance de transmissão do sinal diminui pela metade quando o nível de modulação aumenta em uma unidade. Nota-se que a diferença de vasão é inversamente proporcional ao alcance do caminho óptico, pois a maior vazão acontece em distâncias menores, da mesma forma que quando a distância aumenta, a modulação aplicada é de uma ordem menor. Na prática, esse dinamismo contribui para uma maior robustez dos canais de transmissão, não admitindo prejuízos ao tráfego em virtude de características específicas desses caminhos. Como a pesquisa da tecnologia EON é recente, não existem ainda padrões ou normas a respeito dos alcances relacionados aos formatos de modulação O-OFDM [19].

A programabilidade das taxas de dados pelo ajuste dos esquemas de modulação resultará em uma variedade de modulações e taxas de dados sendo arbitrariamente utilizadas em uma mesma fibra óptica e assim, é imprescindível adicionar bandas de guardas de tamanhos arbitrários entre conexões adjacentes para que tais combinações não gerem interferências de sinais entre esses canais e para que ocorra a correta detecção dos dados transmitidos ao chegar no destino final. Geralmente as bandas de guarda ocupam 1 ou 2 FS [56, 45, 22].

2.1.2 Problema do Roteamento e Atribuição de Espectro em EON

Para o estabelecimento de uma conexão entre uma dada origem e um dado destino na EON é necessário que na rota definida existam os recursos necessários e suficientes para atender a demanda requisitada. A busca por tal caminho óptico e respectiva largura de banda corresponde ao problema fundamental em redes ópticas elásticas, que é o problema do roteamento e atribuição de espectro (RSA) [97].

A Figura 2.6 mostra, à esquerda, uma rede composta pelos nós $1, 2, 3$ e 4, onde nenhuma conexão foi estabelecida ainda. Os seus enlaces 1-2, 1-3, 2-4 e 4-3 estão representados à direita, sendo que cada um deles está dividido em 7 FS, todas disponíveis. As requisições A e B solicitam atendimento. A deseja estabelecer uma conexão do nó 1 para o nó 4 e necessita de 3 FS, e B deseja estabelecer uma conexão do nó 2 para o nó 3 e necessita de 2 FS.

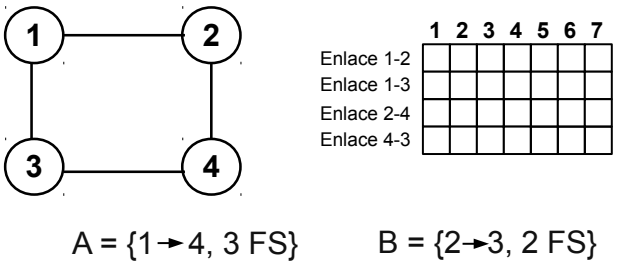

Figura 2.6: Cenário EON e requisições A e B a serem atendidas. .

Para que tais requisições sejam atendidas é preciso resolver o problema RSA. O problema só pode ser satisfatoriamente resolvido se as seguintes restrições forem atendidas:

- Restrição de continuidade do espectro: o circuito óptico deve ocupar o mesmo FS ou conjunto de FS ao longo de toda a rota;
- Restrição de contiguidade do espectro: se uma conexão precisa ser alocada em mais de um FS, esses FS devem ser adjacentes;
- Necessidade de banda de guarda: deve existir um espaçamento mínimo entre duas conexões adjacentes. Esse espaçamento refere-se ao número de FS que deve ser padrão para todas as conexões;
- Restrição de não sobreposição do espectro: duas conexões não podem ocupar simultaneamente um mesmo FS;

O esquema da Figura 2.7 mostra uma solução do problema RSA para atender as demandas A e B. A legenda na parte de baixo da figura identifica a rota da demanda A como um seguimento em tracejado fino e os suas respectivas fatias de frequência, hachuradas na horizontal. A rota da demanda B é representada com um seguimento em tracejado ultra-fino e seus FS hachurados na diagonal. Os FSs marcados com as letras "BG"representam bandas de guarda. A fatia marcada com "x"representa FS já atribuído a uma dada chamada e que ainda não foi desocupado, o que representa um impedimento para uma nova alocação.

O caminho definido para a requisição A é composto pelos enlaces 1-2 e 2-4. Nesses enlaces foi possível designar os três *slots* solicitados, atendidos com as fatias 2,3 e 4. Além disso, essa conexão está devidamente separada das demais fatias com uma banda de guarda nas suas extremidades, em cada um dos enlaces. Como todas as restrições do problema RSA foram atendidas, a conexão pode ser aceita na rede.

Agora veja o caso da requisição B, cuja rota encontrada é composta pelos enlaces 2-4 e 4-3. Note que o enlace 2-4 deve ser compartilhado por ambas as requisições. O enlace 4-3 está com todo o seu espectro desocupado e poderia aceitar B, entretanto, como é necessário designar os mesmos 2 FS no enlace compartilhado e esse enlace não dispõe desse recurso, as restrições do problema não são atendidas. Os FSs 4 e 5 no enlace 2-4 já foram atribuídos à chamada A, e os FSs 6 e 7 são insuficientes para atender B. Como resultado, a requisição B é bloqueada.

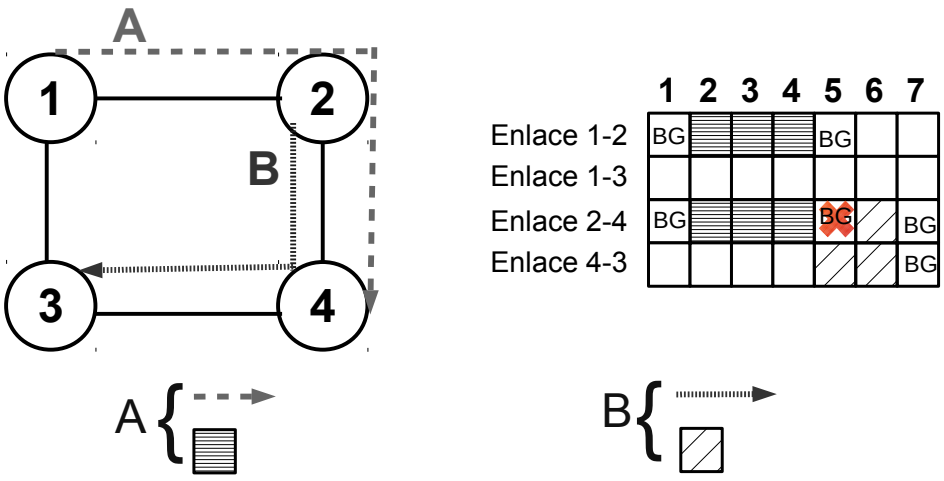

Figura 2.7: Problema do roteamento e alocação de espectro (RSA): Requisições A e B solicitando uma rota na rede (à esquerda) e recursos de espectro (à direita)..

A generalização do problema RSA com maior flexibilidade é o problema de roteamento

e alocação de nível de modulação e espectro (RMLSA), onde além da busca por um caminho óptico e *slots* de frequência suficientes, também é possível escolher o número de *bits* modulados por símbolo para cada subportadora ou para o conjunto de subportadoras que correspondem a uma conexão. Considere na Tabela 2.1 as relações de níveis de modulação e alcances postulados em [56] e destacados anteriormente na Subseção 2.1.1. Ao serem computadas as rotas que podem atender as requisições, suas distâncias são comparadas aos máximos alcances para que um nível de modulação seja determinado. De maneira resumida, os caminhos relativamente maiores são atendidos com os níveis mais baixos de modulação, que transportam menos *bits* por símbolo, de maneira a reduzir as degradações dos sinais, que são maiores quanto maior for a distância. O inverso também é válido para as menores distâncias, onde os níveis mais altos de modulação são mais viáveis.

Modulação	Alcance
BPSK	$2500km$
QPSK	$1200km$
8QAM	$625km$
16QAM	$325km$

Tabela 2.1: Relação de níveis de modulação e seus máximos alcances [56]

2.2 Rede Inter Centros de Dados (ICD)

Um centro de dados (CD) é uma infraestrutura que oferece hospedagem de recursos de computação e aplicações, armazenamento e distribuição de dados para apoiar as operações de tecnologia da informação (TI) das organizações [99]. A coleção de centros de dados em uma rede é a base para os serviços de computação em nuvem. As requisições de usuários desses serviços são distribuídas ao longo de tais centros de dados para prover balanceamento de carga. Os usuários podem ser organizações ou mesmo dispositivos físicos como telescópios astronômicos, por exemplo. Cabe aos provedores de serviço de Internet (*Internet Service Providers* - ISP) a responsabilidade de gerenciar a distribuição desses serviços, a conectividade e as políticas de engenharia de tráfego [64, 101].

A interligação entre vários CD ocorre por meio da *Internet* [38] ou por meio de redes proprietárias como fazem *Google*, *Yahoo!* e *Microsoft* [40], o que exige investimentos substanciais [3]. Essas conexões são ópticas e têm suas capacidades super dimensionadas, ou seja, seus recursos são planejados com excesso de capacidade até um determinado limite para atender picos de demandas que eventualmente podem ocorrer em conexões estáticas.

A literatura faz referências a três tipos principais de redes no que tange à comunicação de CD: rede de CD (*Data Center Network* - DCN), rede intra e inter CD. A CDN é definida como o conjunto de interconexões internas ao ambiente de CD [108, 99]. Rede intra-CD é outro termo encontrado em [77] e [105] para referir-se a CDN. Já rede inter-CD (ICD) é encontrada em [35, 64, 105, 105] e define a interconexão entre mais de um CD, com compartilhamento de aplicações entre os diversos servidores de cada um dos CD. De forma simplificada, [101] trata rede ICD como nuvem (*cloud*).

Um exemplo de rede ICD é a rede do *Google* (Figura 2.8), que atualmente possui pontos de presença em quatro continentes [42]. Por meio dessa infraestrutura completa são processadas buscas diárias por mais de 30 trilhões de URLs[1] únicas na *web*, são hospedados mais de 230 milhões de domínios *web*, são realizadas mais de 3 bilhões de consultas de pesquisas por conteúdos diariamente, sendo que a cada dia pelo menos 15% desse total representa novas buscas, oferecendo serviços em 55 países em 146 diferentes idiomas, para os quais são disponibilizados ainda serviços de *e-mail*, vídeo, armazenamento em nuvem, entre outros [110].

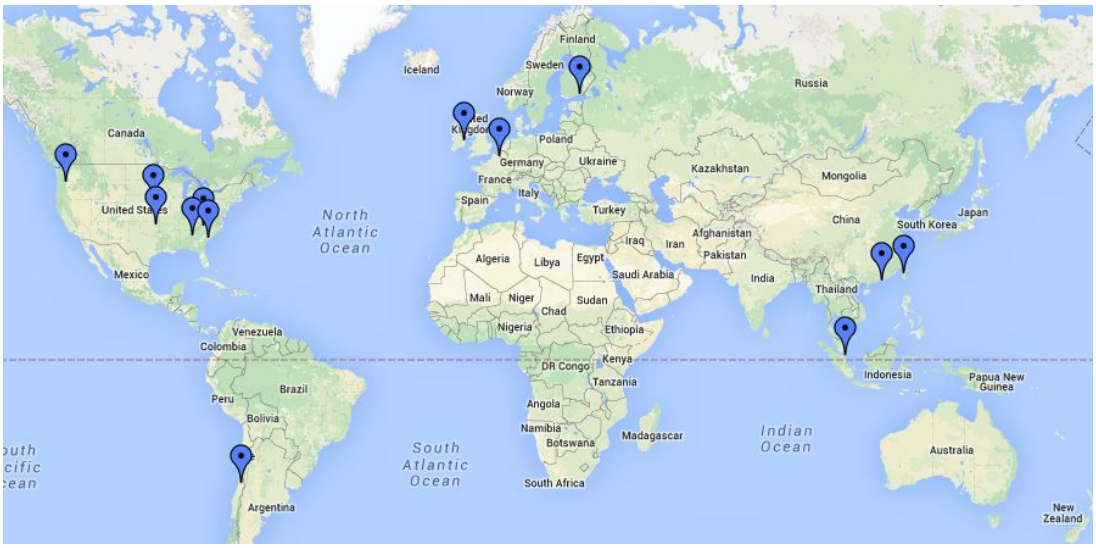

Figura 2.8: Os centros de dados do *Google* constituem uma rede difundida ao longo da América do Norte, América do Sul, Europa e Ásia [42].

[1] *Uniform Resource Locator*

2.3 Aplicações Distribuídas

Um sistema distribuído (*Distributed System* - DS) é definido como uma coleção de computadores independentes que aparece para os seus usuários como um único sistema coerente, escondendo assim, a existência de redes de computadores individuais que suportam esse sistema [88]. As aplicações distribuídas são oferecidas por meio desses sistemas e com tais características de plataforma única. Para os serviços em nuvem, as aplicações são modeladas na forma de *software* como uma serviço (*Software as a Service* - SaaS), hospedados remotamente e prontamente disponíveis para serem acessados de qualquer parte do mundo enquanto detalhes de implementação são ignorados . O aspecto elástico dos recursos distribuídos tem facilitado a migração dessas aplicações para a nuvem [3].

Um serviço pode ser entregue a muitos clientes ao mesmo tempo e o ISP tem o poder de gerenciar o acesso desses clientes aos recursos, garantindo que apenas aquele com permissão apropriada possa acessá-lo. Da mesma forma, um cliente pode fazer solicitações para múltiplos serviços ao mesmo tempo. Entre os serviços mais comuns, as aplicações intensivas em dados se destacam devido aos rígidos requisitos de gerenciamento escalável para manter a consistência dos dados diante desses múltiplos acessos, o que depende do tipo de aplicação e do tipo de dado com o qual lida. Por razões administrativas, diferentes tipos de dados são confinados em diferentes tipos de bases de dados. Por exemplo, bases de dados científicos lidam com dados resultantes de simulações e imagens que demandam enorme poder de processamento e gerenciamento de processos altamente colaborativos [25, 44].

De maneira geral, além da transparência de localização dessas aplicações, também é transparente o acesso a um dado servidor, visto que cada cliente realizando essa ação não pode perceber a presença do outro, assim como é transparente a localização dos dados e processos do servidor, que podem migrar livremente sem que o cliente saiba [44]. Devido a tais restrições, as aplicações distribuídas precisam assegurar que seus serviços sejam entregues de acordo com as especificações do sistema, constituindo a propriedade de dependabilidade. Esse conceito engloba os seguintes atributos [10]:

- Disponibilidade: prontidão para oferecer o serviço correto imediatamente. Isso diz respeito a probabilidade que o sistema opere corretamente em qualquer dado momento;
- Confiabilidade: oferecer os serviços corretos continuamente sem falhas;
- Segurança: ausência de consequências catastróficas sobre o usuário ou o ambiente;
- Integridade: ausência de alterações impróprias no sistema;

- Manutenibilidade: capacidade de sofrer modificações e reparos e ainda assim continuar funcionando conforme suas especificações.

A não identificação de algum desses atributos pode ocasionar problemas de funcionamento ou relacionados à segurança [25].

2.3.1 Segurança e Funcionamento das Aplicações Distribuídas

As aplicações distribuídas funcionam em um arranjo de CD (ou nós) interligados em rede. Em cada um desses CD um ou mais processos são executados e existe uma memória local que armazena informações de estado dessa aplicação. Supondo que exista um processo por nó, tais processos se comunicam trocando mensagens através da rede [89]. Um sistema de gerenciamento distribuído (*Distribution Management System* - DMS) mantém um correto e consistente mapeamento dos CDs, controlando a distribuição de requisições ao longo da rede e serviços de proteção a falhas [78].

O estado da rede é representado pela sequência de mensagens ao longo dos canais de comunicação. O processo é composto de um conjunto de estados, um estado inicial uma sequência de operações, sendo que cada estado do conjunto é representado por um valor atribuído mediante operações. O estado da rede e o estado do processo juntos constituem o estado global do sistema. Como cada processo possui um estado inicial, a reunião de todos os estados iniciais dos processos em geral, representam o estado inicial global. Esse estado global pode ser mudado mediante a ocorrência de um determinado evento que de antemão, modifique o estado de algum processo ou mesmo de algum canal da rede [44].

O modelo de comunicação entre nós dotados de processos é caracterizado principalmente pelo período de tempo aceitável para a espera de uma resposta relacionada a uma solicitação entre nós. Esse modelo de comunicação pode ser síncrono, assíncrono ou parcialmente síncrono [25, 32]. No modelo síncrono, a comunicação ocorre em tempo real sob o gerenciamento de um único relógio e cada mensagem é recebida ao mesmo tempo em que é enviada. No modelo assíncrono, também chamado de modelo de tempo livre, cada processo tem o seu relógio e a comunicação não é simultânea, o que pode gerar falhas e incertezas quanto ao seu estado. Já no modelo parcialmente síncrono, existem processos tanto síncronos quanto assíncronos, e a comunicação entre eles ocorre com sincronia alternada. Muitos serviços, como os oferecidos por computação em nuvem e a própria *Internet* são assíncronos, e por este motivo, podem experimentar atrasos de comunicação e velocidade dos seus processos, e assim vivenciam o problema do consenso [89].

O problema do consenso é o problema de alcançar um acordo entre processos remotos, desde que os processos participantes e a rede sejam confiáveis [36]. Quando a comunicação é síncrona, passos simultâneos e previsíveis de troca de mensagens são realizadas para

se chegar a um acordo. Se ocorrer alguma falha em algum processo, estas podem ser detectadas simplesmente porque não há uma resposta à uma dada requisição em tempo hábil [34].

Se a comunicação é assíncrona, existe uma imprevisibilidade de atrasos nas trocas de mensagem, tornando muito difícil chegar ao acordo, por esse motivo não há uma solução determinística. Se ocorrer alguma falha com um dos processos, essa falha pode não ser detectada e impedir que um acordo seja alcançado, uma vez que é difícil dizer se realmente trata-se de uma falha ou a quantidade de tempo requerida pelo processador é muito grande. Para que tanto os processos quanto a rede sejam confiáveis, é necessário que acessos não autorizados sejam inviabilizados, para assim manter a integridade dos dados e contribuir com a sua disponibilidade [36].

Para burlar tal impossibilidade, os sistemas assíncronos podem assumir a propriedade de serem eventualmente síncronos, supondo que em uma sincronia momentânea exista um tempo de estabilização global; outra possibilidade é supor que o sistema assíncrono possui um oráculo chamado "detector de falhas", para os quais várias classes de falhas (ou faltas) podem ser definidas [16].

2.3.2 Tolerância a Faltas

Um sistema é dito falho quando não pode cumprir as suas promessas, e no caso dos DSs, quando não consegue fornecer um ou mais serviços (completamente) para os quais foi especificado. Nos DSs em larga escala é esperado que seu funcionamento continue aceitável quando ocorre uma falta [75].

Faltas e falhas são intercambiavelmente discutidas [44, 12]. As faltas podem ser atribuídas a um escopo maior, como na dimensão da rede. As falhas são causadas por faltas nos componentes do sistema. Já para [37], faltas são defeitos de mais baixo nível de abstração e falhas, são desvios de especificação. Como não há um consenso definido sobre o uso dos termos falhas e faltas, ambos podem ser empregados entremeadamente.

A rede ICD oferece uma série de recursos e aplicações cujos clientes são altamente dependentes e confiam na recuperação e disponibilidade. No entanto, problemas como queda de enlaces, falhas nos servidores ou perda parcial de conexão são possíveis de ocorrer em uma rede geo-distribuída. A fim de reduzir os impactos causados pelas falhas, [44] propõe classificações de faltas nos sistemas distribuídos (Figura 2.9) baseado em como um componente faltoso se comporta quando ocorre uma falha e que suposições podem ser feitas a respeito:

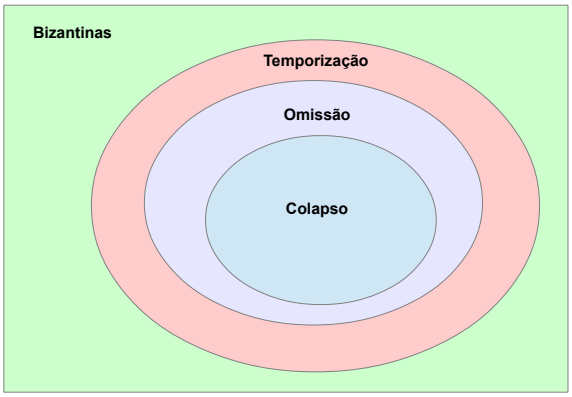

Figura 2.9: Classes de faltas [44].

- Colapso (*Crash*) é a classe de faltas caracterizada pela perda do estado interno de um componente. Durante as transações, seu estado nunca muda após a ocorrência da falha.

- Omissão é a classe de faltas relacionadas com omissões de respostas. Na comunicação entre processos, a sequência de execuções pode ser interrompida devido a falta de uma reposta cujas as ações subsequentes são dependentes. O processo afetado deixa de enviar mensagens ou deixa de receber, cessando em algum nível a comunicação.

- Temporização é a classe de faltas que enquadra falhas decorrentes de respostas inoportunas, que ocorrem fora do momento especificado, podendo ser atrasadas ou adiantadas.

- Bizantinas correspondem à classe cujos processos afetados enviam mensagens que não seguem os princípios e especificações do sistema, se comportando de maneira totalmente arbitrária após a falha. Trata-se do modelo mais severo de faltas.

Enquanto é possível supor que um colapso pode ter ocorrido devido a uma pane elétrica, uma omissão pode ser resultante de uma falha de comunicação e uma falta de temporização é provável de ocorrer por problemas de configurações no sistema. A respeito de faltas bizantinas, nenhuma suposição pode ser feita para justificar o comportamento do processo faltoso, e por este motivo representa o pior tipo de falta [44].

A técnica chave para prover tolerância a faltas é a redundância, que pode ser de informação, de tempo e física [25, 44]. Com redundância de informação, *bits* extras podem ser transmitidos para permitir a recuperação de *bits* ilegíveis. Com a redundância de tempo, uma ação pode ser realizada mais de uma vez, como por exemplo, quando uma requisição enviada a um servidor não obteve resposta, então essa requisição pode ser retransmitida.

A redundância física pode ser feita em *hardware* e *software*, adicionando equipamentos extra à rede e adicionando processos extras, respectivamente. Com a redundância física de *hardware*, o mau funcionamento de um equipamento pode ser mascarado com a utilização de um equipamento similar que já esteja disponível. Com a redundância física de *software*, cada processo pode ter processos similares correspondentes com as mesmas características. Se algum desses processos deixar de responder a uma requisição, o sistema continuará funcionando corretamente porque outro processo similar é capaz de responder. A replicação de processos oferece um alto grau de tolerância a faltas.

Para tolerar faltas bizantinas é necessário haver várias boas réplicas dos processos para compensar os processos ruins (faltosos). Entretanto, uma vez que nenhuma suposição pode ser feita a cerca de quantos e quais são os processos ruins, é impossível chegar a um consenso onde o acordo é feito apenas entre as boas réplicas e o resultado seja verdadeiro e confiável. Tem sido mostrado [44, 15, 34] que, se um grupo de nós busca o consenso mesmo com a existência de uma falta f, um estado final confiável é obtido com a garantia de que pelo menos dois terços desses nós seja não faltoso [70].

2.3.3 Replicação

A replicação é uma técnica efetiva para fornecer tolerância a faltas. Mas, além disso, com mais réplicas disponíveis ao longo da rede, a latência de acesso aos dados é reduzida, minimizando a quantidade de dados trocados na rede e melhorando o desempenho [70].

Tais replicações são geridas pelo Sistema de Gerenciamento de Dados Distribuídos (SGDD). O SGDD executa protocolos para assegurar a confiabilidade dos dados replicados. Esses dados são organizados em partições correlacionadas. Em cada um dos CDs geo-distribuídos, o volume total de dados disponibilizados para replicação é dividido em partições lógicas. Cada partição é replicada n vezes, onde n é dito o fator de replicação, ou seja, n refere-se ao número de cópias que existem do dado [93]. Um grupo de replicação é um grupo de partições cujos CDs participantes possuem a mesma partição sincronizada [70, 3].

Na Figura 2.10 é mostrado um exemplo de replicação entre centros de dados. São mostrados 3 grupos de replicação: G_1, formado pelas partições P_1 localizadas nos centros de dados CD_1, CD_3 e CD_5; G_{k-1}, formado pelas partições P_{k-1} localizadas nos centros de dados CD_2, CD_3, CD_4 e CD_5; e G_k, formado pelas partições P_k localizadas nos centros de dados CD_2, CD_3, CD_4 e CD_5.

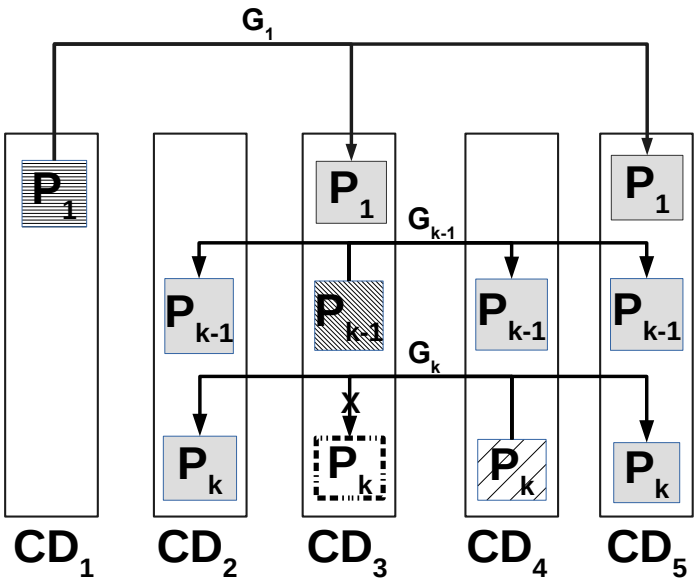

Figura 2.10: Centros de dados e grupos de replicação .

Cada um dos 5 CD participa de algum grupo de replicação de partições e existem no máximo k partições do volume total de dados em cada um desses CD. O grupo de replicação G_1 possui a partição P_1 sincronizada, tendo um fator de replicação $n = 2$. O grupo G_{k-1} mantém a partição P_{k-1} sincronizada com um fator de replicação $n = 3$. Já o grupo G_k representa um grupo com fator de replicação $n = 3$, mas atualmente está com fator de replicação $n = 2$, porque o CD_3 mantém uma cópia de P_k (tracejado na figura) desatualizada. A respeito de desatualizações, nas configurações padrão dos sistemas que gereciam armazenamento de dados distribuídos, em qualquer ponto do tempo é preciso que existam ao menos duas réplicas de dados disponíveis [70]. Tal limite é definido por várias medidas de confiabilidade, como por exemplo, o tempo médio de perda de dados, que é o tempo medido quando acontece de ao menos um bloco ou partição "perdido"pelo sistema não puder ser recuperado [93].

No exemplo, os grupos G_1 e G_k não atendem ao requisito de tolerância à falhas bizantinas, que exige $3f + 1$ cópias para combater f falhas [15], embora, ainda assim, componham grupos de replicação que suportem outros tipos de falhas de menor complexidade. O grupo G_{k-1} é o único tolerante a falhas bizantinas. O grupo G_k pode realizar uma ressincronização da partição P_k no CD_3, aumentando seu fator de replicação para $n = 3$. Embora este número possa ser aumentado para permitir que mais réplicas falhem, na prática utiliza-se poucas réplicas [94]. A réplica de P_k nesse CD_3 pode estar inconsistente com o grupo por diversas razões, como por exemplo, grande atraso na gravação dos dados durante o processo de sincronização, muitas vezes gerado pelos gargalos formados

dentro dos CDs devido aos diversos outros tráfegos e graus de prioridade locais, agravados se a partição for um considerável volume de dado.

Os principais desafios das replicações são manter a consistência dos dados e o bom desempenho das réplicas. A manutenção da consistência é uma tarefa complexa para algumas aplicações específicas que requerem estrita precisão por parte das transações durante o processo de sincronia. Já o desempenho das réplicas pode ser afetado quando novas réplicas são criadas, isto porque a ocupação dos espaços de armazenamento pode incorrer em sobrecargas, ou mesmo durante o processo de escrita, em razão do grande número de atualizações. Além disso, o gerenciamento de réplicas se torna um grande desafio com o crescimento do volume de dados. Quando uma réplica tem que ser criada ou migrada para um novo local, não estará disponível até que todo o seu conteúdo seja copiado de outras réplicas existentes. Se esse processo leva muito tempo, a disponibilidade geral de dados pode ser prejudicada se o número de réplicas disponível não é suficiente para acomodar todas as solicitações do usuário [70, 4].

Um dos métodos mais comuns para implementar serviços de tolerância a faltas são as replicações de máquina de estados (*State Machine Replication* - SMR) [75], que é baseada em máquinas determinísticas implementadas nas réplicas. Essas máquinas iniciam a execução de requisições cada uma no mesmo estado das demais. Todas as execuções ocorrem na mesma ordem. No final do processo, se as réplicas são corretas seus estados são uniformes [37].

Outra técnica comum é o sistema de *quóruns*. Dado um conjunto de réplicas, organizados em vários subconjuntos denominados *quóruns*, que são capazes de tomar decisões (votar sobre um valor proposto). O voto da maioria dos *quóruns* autoriza a realização de uma operação de leitura ou escrita que eventualmente realiza uma sincronização de dados [37, 50].

2.3.4 Múltiplas Transferências de Dados em Massa

O tráfego entre CD é classificado em três classes baseadas na sensibilidade a atrasos: interativo, elástico e *background*. O tráfego interativo é muito sensível ao tempo [107, 20, 68]. A classe interativa, de mais alta ordem de prioridade, requer o menor retardo possível e é proveniente de aplicações voltados para o cliente, como os serviços de telefonia e vídeo, por exemplo. O tráfego elástico é menos crítico para a experiência do usuário final e, embora seja menos sensível a atrasos, precisa cumprir os prazos estabelecidos, o que o torna o nível dois de prioridade. Um exemplo de aplicação com tráfego elástico são as aplicações *web*. A terceira ordem de prioridade é a classe de tráfego *background*, que possui os prazos mais longos para atendimento e cujos serviços precisam ser completados tão rápido quanto possível, como a ressincronização de CD [98, 43]. Em termos de volume

total de dados, o tráfego interativo representa a menor porção e o tráfego *background*, a maior [43].

As aplicações de Múltiplas Transferências de Dados em Massa (MBDT) são um tipo de tráfego *background*. Essas aplicações realizam a movimentação de volumosas quantidades de dados ao longo da rede geograficamente distribuída. São implementadas nos ambientes de computação em nuvem para replicar e sincronizar as bases de conteúdos de grandes provedores desse tipo de serviço [54].

O problema MBDT é encontrar uma alocação de banda que possa garantir as transferências dentro de um tempo estabelecido. Hoje, a alocação de recursos para MBDT é estática com taxas previamente definidas, onde uma operação completa de transferência pode levar de dias a semanas [100], o que é feito tipicamente nos momentos de menor utilização da rede (padrão noturno) e leva os provedores a utilizarem armazenamento de dados em CD alugados durante o dia, sendo que a rede não é ciente das aplicações que são executadas [100, 53, 54]; além disso, não existem técnicas de engenharia de tráfego que garantam as transferências dentro de um específico período de tempo (prazo) [107].

O tráfego elástico das MBDT pode ser paralisado e reiniciado conforme prioridade definida pelos ISPs, sendo altamente dependente da largura de banda disponível na rede [52]. Além disso, existe total liberdade para a definição de banda que atenda tais requisições. Atribuir banda fixa pode permitir maior vazão de transferências de dados, ou seja, aumentar o volume. Se a banda é mínima, mais requisições podem ser efetuadas, aumentando o quantidade.

O problema MBDT pode ser reduzido ao problema de transferência de dados em massa (*Bulk Data Transfer* - BDT), sua variação mais simplificada. Em um DS, uma BDT ocorre na distribuição de conteúdo, *backups*, migração de máquinas virtuais, entre outras operações. As MBDTs são realizadas através de várias BDTs[18].

Escalonamento de Requisições

Os provedores utilizam escalonamento de requisições para aumentar as chances de uma chamada ser atendida [54]. Geralmente quando trata-se de requisições com prazo implícito, associado a entrega de dados, como é o caso tanto das BDT quanto MBDT, as políticas aplicadas que definen prioridade é a de prazo mais recente primeiro (*Earliest Deadline First*- EDF) e a de menor quantidade de dados primeiro (*Shortest Job First* - SJF), por reduzirem o número de requisições bloqueadas [92], também chamada de método oportunístico [11].

Quando ocorre de várias requisições terem o mesmo prazo, a política mais justa é o método FIFO (*first-in-first-out*) [14], onde as chamadas vão sendo atendidas segundo a ordem de chegada. Num cenário multi tráfego, existem prioridades já estabelecidas que

constituem a política de alocação hierárquica. As diferentes requisições são prontamente executadas, e devido às políticas de QoS, as chamadas de mais alta prioridade são garantidas em detrimento das chamadas de prioridade menor. Isso significa que em um ambiente com escassez de recurso, o tráfego de classes inferiores pode ser bloqueado para garantir os fluxos de maior prioridade. Por este motivo, não há garantias de conclusão dentro do prazo para as MBDTs, resultando em bloqueios prejudiciais. [98, 107].

Uma maneira de garantir atendimento dentro do prazo é fazendo reserva antecipada, garantindo a capacidade necessária. Entretanto, para efetuar uma reserva é preciso fazer uma análise sobre os recursos da rede, utilizando ferramentas de previsão de tráfego e detecção de erros. A reserva imediata muitas vezes é inviável devido ao grande volume de dados a serem transferidos, que pode ocupar toda a capacidade disponível, assim, este tipo reserva é comumente feito por aplicações de classes de maior prioridade [60].

2.4 O Paradigma CLD

O projeto em camadas cruzadas (CLD) pode ser entendido como a violação das configurações tradicionais das arquiteturas em camadas através de um novo modelo de comunicação, de forma direta, possível até mesmo entre camadas que não estão imediatamente próximas [76, 84]. Na maneira lógica como as camadas são estruturadas, onde cada uma delas mantém isolado o seu próprio conjunto de regras individuais, a comunicação só acontece entre camadas adjacentes e por meio de procedimentos de chamadas e respostas. Nenhuma camada é capaz de enxergar a estruturação interna de outra camada.

O modelo de rede de interconexão de sistemas abertos (*Open Systems Interconnection*-OSI), onde a comunicação na rede ocorre através de sete camadas individuais, isoladas de acessos oriundo das camadas subjacentes, e o modelo de rede TCP/IP (*Transmission Control Protocol/Internet Protocol*), onde o TCP é responsável por definir a forma de comunicação da rede através das aplicações e o IP define como os dados são trocados nessa comunicação, constituem arquiteturas padronizadas que são aplicadas para o funcionamento das redes de computação [84]. Esses dois modelos são exemplos de organização padrão convencional de camadas sequenciais que preservam a tradicional comunicação de chamadas e repostas.

Ambas as arquiteturas permitem a rápida implementação de sistemas interoperáveis, mas o desempenho desses sistemas é limitado por não haver coordenação entre tais camadas. Visando ganho de desempenho e maior facilidade de gerenciamento das redes, o paradigma CLD oferece acesso e compartilhamento entre essas camadas [85].

Embora CLD seja bastante empregado em soluções de comunicação sem fio, estudos recentes tem buscado adotá-la em redes ópticas [85]. Um exemplo de arquitetura de rede

CLD é a rede IP sobre WDM, onde as requisições na rede lógica IP são mapeadas e transmitidas através de interligações físicas na camada WDM, e esse mapeamento é uma relação em camadas cruzadas [112].

Uma linha de pesquisa CLD em redes ópticas é o roteamento ciente das limitações na camada física (*Impairment-Aware Routing* - IA-R). Em redes WDM, o problema IA-R é o problema de roteamento e atribuição de comprimento de onda ciente das limitações da camada física (*Impairment-Aware Routing and Wavelength Assignment* - IA-RWA) [51]. Já em EON, o problema que considera tais limitações é o problema de roteamento e alocação de espectro ciente das limitações da camada física (*Impairment-Aware Routing and Spectrum Alocation* - IA-RSA) [102].

Em ambos os problemas, fatores decorrentes da camada física, como taxa de erro de *bits* (*Bit Error Rate* - BER), efeitos de dispersão e atenuação da luz que incide na fibra, modos de propagação, entre outros, são considerados na tarefa de aprimorar o roteamento [102, 51]. A camada de rede adquire visão sobre os impactos na qualidade do sinal óptico que podem interferir diretamente da alocação dos recursos. No modelo convencional de roteamento, que não dispõe de tais informações, os recursos alocados são muitas vezes bloqueados devido a problemas externos à camada de rede [85]. Com a visão CLD, a rede pode tomar decisões mais complexas e viabilizar o estabelecimento de conexões que, na maneira tradicional seriam prejudicadas [113].

A ideia por trás deste trabalho é explorar a visão CLD na perspectiva da camada de aplicação em um ambiente EON. Com o variedade de aplicações oferecidas pelos serviços em nuvem, as medidas de QoS tem servido não apenas para avaliar os requisitos aceitáveis, mas também para fornecer informações separadas sobre o comportamento de cada uma das aplicações e serviços. Com tais informações, aplicações pertencentes à mesma classe de tráfego podem receber tratamento diferenciado na camada de rede. Por exemplo, suponha que, em uma organização, um grupo de pessoas em um momento de lazer, decidam assistir a um vídeo no *Youtube*, enquanto que outro grupo, no mesmo momento, precisa realizar uma vídeo conferência. Ambas as aplicações são da mesma classe de tráfego, mas com diferentes níveis de importância para a organização. Se tal problema é transportado para uma rede de maior dimensão, como uma rede ICD, é possível imaginar que uma rede ciente da aplicação poderia sofrer menos impactos ao enxergar aplicações de maneira diferenciada. Em EON, a rede ciente da aplicação estaria apta a resolver o problema de roteamento e alocação de espectro ciente da aplicação (*Application-Aware Routing and Spectrum Alocation* - AA-RSA), atendendo aplicação de mesma classe de tráfego, sem que a aceitação de uma delas acarrete em aumento de bloqueio para a outra aplicação.

Essas abordagens CLD em redes ópticas, IA-RWA, IA-RSA e AA-RSA são esquematizadas na Figura 2.11. Com a visão sobre as outras camadas, a camada de rede pode tomar

decisões sobre os seus recursos com maior eficiência operacional, realizando roteamento inteligente [29].

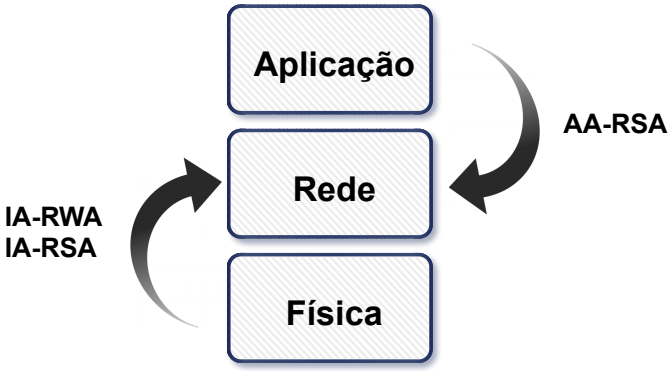

Figura 2.11: Modelo de interação no paradigma CLD para redes ópticas..

Operações fundamentais de ressincronizações entre CDs são acionadas quando um CD anteriormente indisponível, volta a fazer parte da rede. Esse tipo de operação é base para o protocolo de funcionamento dos CDs geo-distribuídos, e apesar de serem rotineiramente realizados e exigirem altas demandas de largura de banda da rede, as aplicações MBDTs que tornam possível essas ressincronizações não estão entre as classes de tráfego para as quais o atendimento é prioritário. Por esse motivo, a camada de rede pode tirar proveito do CLD, reconhecendo os requisitos da aplicação MBDT e absorvendo tais requisitos às suas decisões de roteamento. A resolução de um novo problema AA-RSA com discernimento sobre o cenário de ressincronização ICD pode aumentar a taxa de serviços realizados, em comparação com as soluções de roteamento tradicionais, que no máximo, conseguem tomar decisões de acordo com as classes de tráfego, mas nunca com relação às diferentes aplicações em cada uma dessas classes.

2.5 Resumo Conclusivo

Este capítulo tratou dos conceitos básicos que ajudarão na compreensão de como prover MBDT em EON no que tange a ressincronização entre CD. Para tanto, inicialmente apresentou-se a tecnologia EON a ser implantada futuramente nas redes de núcleo ICD, bem como os problemas fundamentais nessa OTN, que são o problema de roteamento e alocação de espectro (RSA) e o problema de roteamento e alocação de nível de modulação e espectro (RMLSA).

Também foram retratadas brevemente a composição das redes ICD geograficamente distribuídas, destacando-se suas propriedades de DS, onde uma variada gama de apli-

cações distribuídas são hospedadas e executadas para prover os serviços de computação em nuvem. Pontos-chaves de operações fundamentais como as replicações entre centros de dados foram discorridos, onde foram explicitadas as MBDTs que possibilitam a movimentação de dados para que as replicações ocorram. Por final, o paradigma CLD foi apresentado para destacar a potencial inteligência que a camada de rede pode adquirir ao receber informações de outras camadas. Nesse sentido, o interesse foi definir que a camada de aplicação pode encaminhar informação a respeito das aplicações MBDTs diretamente para a camada de rede, tornando o roteamento inteligente.

Capítulo 3

Revisão de Literatura

Este capítulo traz um levantamento bibliográfico a respeito dos conceitos destacados no capítulo anterior. A Seção 3.1 apresenta os principais apontamentos da literatura para a tecnologia EON. A Subseção 3.1.1 destaca as principais soluções de roteamento das quais a literatura trata. Por último, são levantadas as principais soluções para os problemas BDT e MBDT na Seção 3.2.

3.1 Redes Ópticas Elásticas (EON)

A tecnologia EON foi apresentada em [45], onde foi proposta inicialmente como arquitetura SLICE, também mostrada em [87]. Foram destacados o potencial de fornecimento de caminhos ópticos com largura de banda variável a partir de modernos equipamentos BV-WXC e BVT, e da adoção do sistema de modulação especialmente implementado para essas redes ópticas, o O-OFDM. Já é possível alcançar transmissões de 1 $Tbps$ [26]. Os esforços das pesquisas também estão focados em reduzir o número de subportadoras configuradas como bandas de guardas. Dessa maneira, a parte do espectro reservada para transmissão vai ser aumentada, e a separação entre canais será feita cada vez mais com menos espectro [109].

Tal arquitetura promete alta capacidade e chama a atenção para os possíveis novos padrões de consumo de energia elétrica. Segundo [28] essa preocupação surge ao relacionar tanto o aumento da capacidade quanto o aumento do volume de tráfego. Enquanto o consumo de energia das redes de acesso (distribuidora de serviços como TV por assinatura, por exemplo) é proporcional ao número de assinantes, o consumo de energia da rede de núcleo depende do volume de tráfego. As previsões para 2020 sugerem que o consumo energético com esses equipamentos será, em média, 10.5 $W/Mbps$ para as redes de *backbones*, e 8 $W/Mbps$ para as redes de acesso. Além disso, com a mudança de tecnologia os

ganhos com eficiência energética para este mesmo período serão de 64% para *backbones* e 449% para redes de acesso, se comparados ao ano de 2010.

A implantação da nova tecnologia tem como desafios futuros os seguintes aspectos [45]: a grade do espectro óptico mudará sua granularidade, o que já tem sido considerado pela ITU-T; o sistema O-OFDM está em experimentação e tem sido percebido que as altas taxas de picos de potências dos sinais exigirão componentes transmissores e receptores com bom desempenho em lidar com essas variações; os BVTs serão auto-adaptáveis na escolha de modulações. Os formatos de modulação mais considerados atualmente são QPSK e 8/16-QAM, isso porque as modulações de mais alta ordem ocupam menos recursos de espectro [26]; os LASERs empregados na geração dos sinais ópticos serão reprojetados para adquirir características como frequência controlada e temperatura estabilizada [87, 106, 90].

Os BVTs implementarão vários *transceivers* virtuais, uma referência à transmissão óptica definida por *software* (*Software-Defined Optical Transmission*-SDOT) [81]. Também apresentarão boa acurácia em reportar parâmetros de condições do canal como SNR óptico e elétrico, dispersão cromática e PMD[1]. Com isso, existirão boas chances para as redes realizarem IA-RSA já que as medidas de desempenho da camada física serão facilmente obtidas.

Um novo plano de controle adaptativo será projetado para lidar com as propriedades elásticas da EON [90]. É prevista a extensão do protocolo OSPF-TE (*Open Shortest Path First - Traffic Engineering)*, que já é utilizado para obtenção de rotas, mas precisará lidar com verificação das fatias de espectro disponível. A comunidade do IETF (*Internet Engineering Task Force*) tem focado no projeto do conjunto de protocolos do plano de controle, GMPLS (*Generalized Multi-Protocol Label Switching*), associado a extensões SDN, para suportar as configurações dos elementos da EON, onde os parâmetros serão controlados via Elementos de Computação de Caminhos (*Paths Computation Elements*-PCEs). Essas extensões poderão ser usadas tanto em caso de estabelecimento de um novo caminho óptico quanto na adaptação de caminhos já estabelecidos [26].

3.1.1 Problemas RSA e RMLSA

A literatura recente tem apresentado várias propostas para solucionar os problemas de roteamento e alocação de espectro (RSA) e de roteamento e alocação de nível de modulação e espectro (RMLSA). Ambos os problemas utilizam as modulações Algumas soluções são destacadas a seguir:

[1] *Polarization Mode Dispersion*

Roteamento Convencional

Em [48] foi mostrado o conceito de alocação de espectro óptico baseando-se nas distâncias de transmissão e [86] empregou tal conceito em uma proposta RSA. Esses trabalhos indicaram o quão promissora a EON poderia ser em transportar mais dados e acomodar mais caminhos ópticos, se comparado às rede DWDM. Já [46] demonstrou que o problema RSA é NP-difícil, e mais complexo de ser resolvido do que o problema de roteamento e atribuição de comprimento de onda (RWA).

Em [22] foi mostrado que o problema RMLSA estático é NP-Completo. Esse mesmo problema foi abordado de forma decomposta em dois subproblemas, denominados ILP-(RM+SA), *Routing and Modulation+Spectrum Assingment*, e ainda, uma terceira proposta foi apresentada, consistindo da heurística RMLSA para atendimento de demandas de maneira sequencial. Trabalhos futuros incluiriam examinar os mesmos problemas no cenário semi estático ou dinâmicos em redes O-OFDM onde sobreposição de espectro é permitida com base no tempo e/ou modelos probabilísticos de tráfego, como forma de melhorar ainda mais a eficiência de utilização do espectro. As demandas foram ordenadas de acordo com as políticas MSF (*Most Subcarriers First*) e LPF (*Longest Path First*).

Um esquema de reciclagem de espectro fragmentado é proposto em [60]. Aplicando um método de programação dinâmica, o espectro fragmentado após o atendimento de chamadas orientadas a fluxos, com reserva imediata, é identificado e então utilizado para realizar BDT por meio de reserva antecipada, um esquema denominado reserva maleável. Este trabalho é estendido em [62] onde uma solução é proposta para maximizar o número médio de atendimento de requisições orientadas a dados com garantia de atendimento para, pelo menos, um limite mínimo, bem como, um controle de admissão para requisições que podem ser aceitas dentro das restrições estabelecidas, e que tem o poder de rejeitar requisições não completáveis logo no aprovisionamento.

O trabalho de [96] trata o problema RSA dinâmico, que reúne o subproblema de seleção do formato de sinal, roteamento e atribuição do espectro. O problema foi modelado utilizando programação linear inteira mista (*Mixed Integer Linear Programming*-MILP), mas como é não linear, o problema foi decomposto em duas abordagens heurísticas: RSA com modulação fixa e RSA com modulação adaptativa. Nesse primeiro caso, o RSA foi testado com pelo menos quatro modulações fixas diferentes e a atribuição de espectro dependia dessa modulação; no segundo caso, o algoritmo começa utilizando a modulação de nível mais alto, e modificações eram introduzidas caso a distância da transmissão fosse incompatível com o alcance fornecido. Foram considerados no experimentos dois tipos de fibra óptica: SMF (*Sigle Mode Fiber*) e ULAF (Ultra-Large-Area Fiber). SMF é um tipo de fibra monomodo, na qual se usa detecção direta e amplificação, com modulação 4QAM em uma distância de $1200km$; e ULAF é um tipo de fibra para áreas ultra-distantes, que

empregou modulação 14QAM para uma distância de $2000km$. A diferença entre os dois tipos de fibra é a redução das limitações de não linearidade pelas diferentes formas de propagação da luz que incide em cada uma.

Em [57] é proposto um algoritmo RSA *multicast* no cenário estático e dinâmico e que emprega um grafo auxiliar. A cada requisição *multicast* feita, a topologia física é decomposta em vários grafos auxiliares em camadas de acordo com a utilização do espectro da rede. Em seguida, com base na requisição de banda passante do pedido, a camada mais adequada é selecionada e uma árvore óptica é calculada nessa camada. Os algoritmos não realizam *multicast* óptico, uma vez que os comutadores ópticos cruzados (*Optical Cross-Connects* -OXCs) empregados são incapazes de *multicast*. Em [58] também é proposto um algoritmo RSA *multicast* dinâmico, mas este calcula a fragmentação do espectro decorrente da alocação de recursos.

Já [6] propõe uma solução de roteamento RMLSA para caminhos individuais e multicaminho. A granularidade mínima da largura de banda foi utilizada para a definição de outros tamanhos de bandas que são maiores que a mínima em 20%, 40%, 60%, 80% e 100%. Todos os caminhos definidos pelo roteamento são classificados em 5 níveis de fragmentação, do menos fragmentado para o mais fragmentado. Então, quanto mais fragmentado um caminho, menor é a banda alocada nele.

Em [19] foi apresentada uma solução de RSA multicaminhos com modulação adaptável à distância em um cenário de tráfego dinâmico, para o qual foi proposto um algoritmo de múltiplos caminhos computados sequencialmente e outro com caminhos pré-computados. Os autores alertam para o cuidado quanto a fragmentação do espectro nesse cenário dinâmico, cujos *slots* podem resultar em porções não contíguas e demostram que o roteamento multicaminho em EON resulta em um maior consumo de recursos espectrais na forma de banda de guarda.

Roteamento com CLD

As soluções de roteamento ciente destacadas na literatura, estão divididas entre roteamento ciente das limitações da camada física e ciente da aplicação, conforme segue:

Roteamento ciente da limitação da camada física Em [5] é proposto o algoritmo de atribuição de rota e modulação adaptativa (*Adaptative Modulation Routing Assignment* - AMRA), um algoritmo de roteamento ciente do uso de regeneradores e que emprega modulação adaptativa, em duas versões, sendo a primeira para transferências de dados *unicast* e outra, transferência *anycast*. A estratégia das duas soluções identifica as causas de rejeição de uma requisição, que pode ser a falta de recursos ou a falta de regenerador óptico na rota. Se o problema identificado for a inexistência de regeneradores

de sinais, o algoritmo emprega uma modulação de nível mais baixo compatível com a distância. Os caminhos candidatos com saltos são calculados com base na métrica de utilização do enlace (*Link Utilization Metric*-LUM), que permite encontrar os enlaces menos utilizados para alocar as requisições.

As propostas de [33] são baseadas em *manycast* eficiente em energia (*Energy Efficient Manycast*-EEM) nos cenários estáticos e dinâmicos, que são: um modelo de programação linear inteira (ILP) para roteamento *manycast* e alocação de recursos que visa o uso eficiente de energia elétrica por parte dos dispositivos físicos dispostos na rota definida, um RSA heurístico *Pure*-EEM, um RSA heurístico ciente do bloqueio (*Blocking Aware*-EEM) e um RSA heurístico baseado nos menores caminhos. Para avaliar o uso eficiente de energia elétrica pelos dispositivos da rede, uma função custo de consumo de energia foi proposta com restrições ponderadas das localizações de comutadores ópticos, conversores, roteadores e processadores de sinais, considerando valores típicos do mundo real. O modelo ILP busca minimizar o consumo de energia, o P-EEM-RSA escolhe rotas baseadas no menor consumo de energia de acordo com o modelo ILP, o BA-EEM-RSA considera a probabilidade de bloqueio de requisições devido a indisponibilidade de espectro óptico nos enlaces, e o RSA com menores caminhos, que não é ciente dos recursos nos enlaces, é executado em cenários onde equipamentos inativos são desligados para reduzir o consumo de energia.

Em [95] é proposto um RMLSA dinâmico ciente das limitações físicas para roteamento *unicast* e *anycast* com o objetivo de minimizar a taxa de bloqueio das requisições e fragmentação do espectro, usando a política *First-Fit*. O roteamento é baseado nos caminhos que possuem menos regeneradores. A proposta considera regeneradores e BVTs de três capacidades distintas, resultando em diferentes perfis de consumo de energia elétrica. Com os experimentos realizados foi mostrado que o uso de variados níveis de modulação conduz à melhor eficiência espectral, porém, eleva o consumo de energia pelos elementos da rede. Além disso, os níveis mais altos de modulação, com maior eficiência, requerem mais regeneradores de sinais em redes de longa distância.

Roteamento ciente da limitação da camada de aplicação Um esquema de roteamento ciente da aplicação é proposto em [72]. O AA-RSA é empregado na criação de CD virtuais (*Virtual Data Center* -VDC). O provisionamento desses VDCs é feito a partir da coordenação da rede e o CD. Primeiramente, quando chega um pedido de máquina virtual, os CDs físicos com mais recurso disponível são verificados. Em seguida, são verificadas as condições RSA nos caminhos candidatos entre esses CD físicos. Essa solução melhorou a escalabilidade, permitindo a criação de mais VDCs e reduziu a interação com a infraestrutura física, o que resultou em menor tempo de provisionamento de recursos.

3.2 Múltiplas Transferências de Dados em Massa (MBDT)

Para lidar com as grandes demandas de dados e aproveitar de maneira mais eficiente a largura de banda disponível, os operadores de CD e provedores de serviços de computação em nuvem podem explorar a proposta de [53] do sistema *NetStitcher*. Esse sistema localiza os recursos ociosos e os emprega para as MBDTs em momentos de menor utilização da rede. A estratégia *store-and-forward* armazena dados temporariamente em CD com serviços de hospedagem de se alcançar o destino final. Devido a natureza de distintos fusos-horários entre os CDs geo-distribuídos, os horários de pico de utilização da rede em cada lugar podem não coincidir, levando à ocorrência de ociosidades de recursos que é eficientemente aproveitada pelo *NetStitcher*. Entretanto, o sistema lida apenas com tráfego estático.

Visando a utilização mais eficiente dos recursos da rede, em [103] é proposto um *framework* de serviços e infraestrutura de nuvem (CISF) para informar sobre os tipos de tráfegos utilizando a rede, como arquivos de vídeo, BDT e aplicação de televisão por IP (*Internet Protocol Television*-IPTV), e suas respectivas e adequadas alocações de recursos. Essa relação é organizada em um *ranking* de requisitos de QoS, ordenados dos mais restritos para os menos restritos. Também é previsto um mecanismo corretor orientado a serviço, com informações a respeito da rede, tais como custo, prazo máximo e mínima largura de banda. Esse mecanismo recebe as requisições da camada de aplicação, decide sobre a possibilidade de atendimento de acordo com o seu destino final e então mapeia tais requisições aos recursos. Essa solução trata tanto de transferência de dados sensíveis a atrasos, quanto transferências tolerantes a atrasos, e considera distinções de classes, não de aplicações dentro das classes.

Em [92] é proposta uma solução de roteamento e escalonamento de alocação de espectro (RSSA), a partir da interconexão de CD em uma ambiente *Cloud*, com gerenciamento de recursos e controle implementados via SDN, sendo que esse controlador traduz as requisições de transferências em requisições de conexões no plano de controle da rede óptica. Para decidir se uma requisição de transferência pode ser aceita, o controlador executa o RSSA e se for possível, tenta reduzir o recurso alocado para outra transferência que esteja ocorrendo, desde que sua conclusão seja garantida. Mais recentemente em [9], outra solução semelhante trata diretamente de requisições de transferência, liberando o controlador da função de tradutor. Em [11] também é proposta uma solução de escalonamento de múltiplas requisições para garantir a atribuição de recursos variando no tempo fazendo-se reserva antecipada.

Já em [65], partido do preceito da utilização eficiente, é proposto o NetEx, um sistema de transferências que explora oportunisticamente os excessos de capacidade nos enlaces da rede e os utiliza para transferir dois tipos de classes de tráfego, normal e *bulk*, cujo nível

de prioridade é menor. Esse sistema conta com um componente de engenharia de tráfego ciente da largura de banda aplicado ao esquema de roteamento padrão que, a partir de estimativas de demandas futuras por largura de banda, consegue maximizar o recurso atribuído às requisições e atender a uma quantidade adicional de tráfego. Essa solução é estática e utiliza ferramentas extras para fazer previsões do consumo de recursos.

As BDTs realizadas para replicação ICD geo-distribuídas são destacadas em [64], que propõe uma arquitetura de rede de comunicação ICD capaz de oferecer largura de banda sob demanda. Trata-se da rede fotônica inteligente e globalmente reconfigurável (*Globally Reconfigurable Intelligent Photonic Network* - GRIPhoN), que dinamicamente permite ajustar a largura de banda baseada nas distintas classes de tráfego. A GRIPhoN prevê o particionamento da rede de transporte em duas camadas, OTN e DWDM, para oferecer, respectivamente, taxas mais baixas e altas. Também é possível compartilhar transmissores no lado cliente da OTN e utilizar nós de encaminhamento eletrônico para manter o baixo custo. A proposta ainda prevê um controlador que identifica o tipo de tráfego requisitado e encaminha essa requisição para a camada a qual compete atendê-la.

Em [18], trata-se do aprovisionamento de serviços inter domínio, para o qual foi mostrado que é possível reduzir a taxa de bloqueio de conexões para requisições de aplicações BDT. Dois algoritmos são apresentados para roteamento multicaminho inter domínio, sendo um para aplicações de *streaming* em tempo real e outro para BDT. Os domínios, na topologia estática virtual são interligados com túneis de trânsito super provisionados entre os nós de borda. Como o tunelamento acaba reservando recursos que podem não ser utilizados por completo, o roteamento multicaminho ameniza este problema de desperdício realizando agregação de largura de banda, que pode desencadear o problema do atraso diferencial, pois uma vez que o fluxo é dividido para ser transportado simultaneamente por diversos caminhos entre um nó de origem e um nó de destino, não há garantias quanto a ordem de chegada dos pacotes, nem quanto ao tempo gasto para percorrer estes caminhos. Assim, um *buffer* no nó de destino poderia receber estes pacotes e ordená-los.

Custo e capacidade são os problemas destacados em [21]. Os autores realizaram as transferências utilizando armazenamento intermediário com um custo fixo e com custo variável. Foi apontado que as movimentações que ocorrem entre lugares com diferentes fusos horários, impactam nos custos finais, uma vez que, quando parte do caminho está em um horário fora de pico de tráfego, a outra parte do mesmo caminho definido pode estar em situação diferente.

Em [71], os autores sugerem a expansão da capacidade da rede utilizando técnicas de virtualização para lidar com as transferências. A solução *Effingo*, adotada pelo *Google*, é destacada. Esse sistema se apoia em um mecanismo que reúne a técnica do SnF, associada a decisões de escalonamento de recursos tomadas de forma distribuída, e utilizando

uma eficiente topologia multiponto que tira proveito de históricos locais da utilização de recursos.

Em [54], foi abordado o padrão de uso da rede, que evidenciou os períodos das madrugadas como o momento em que os recursos tendem a ficar ociosos. As MBDTs são realizadas nesses períodos, reduzindo os custos da largura de banda. São empregados escalonamentos de chamadas de acordo com o fuso horário do local onde está situado cada CD. São assumidas transferências fim-a-fim (E2E-Sched) com múltiplas requisições paralelas, e também SnF, com armazenamento em trânsito, entre os ISPs. O uso de E2E e SnF resultou em redução dos custos com armazenamento alugado.

Baseado nisto, [100] infere que, alocar banda com base nas previsões feitas pelo sistema, acaba levando à subutilização de largura de banda, quando o dado fluxo requisitante for pequeno. Para resolver esse problema, foi proposto o algoritmo estático de justiça máxima e mínima (*Max-Min Fairness*-MMF), que aloca largura de banda iterativamente de forma crescente. Uma versão dinâmica desse problema foi modelado em PLI.

Sem distinção do tipo de tráfego inter-CD, [59] propõe um escalonamento de tráfegos orientado ao lucro de acordo com o uso da largura de banda e com o custo das transferências de dados, ao qual é empregada a técnica *Lyapunov* de otimização de tempo. O escalonamento considera uma estratégia estocástica que aproveita o esquema SnF para provisionar requisições de transferência de dados ICD. As decisões são relacionadas às transferências cujos dados podem ser enviados entre dois CD, visto de nem todos os nós podem assumir o papel de destinatário.

Em [49] é proposto um projeto para uma infra-estrutura de alocação de banda hierárquica global com suporte para computação distribuída e transferência de dados, com foco em redes privadas dedicadas. As políticas de alocação obedecem as prioridades relacionadas aos níveis de serviços oferecidos pela rede, as definições dos usuários e as políticas de engenharia de tráfego. Por exemplo, um grupo provedor de determinado serviço pode necessitar de garantia mínima de banda ao longo de todos os nós de uma rede onde seus serviços são oferecidos, enquanto que um outro usuário individual pode precisar de garantias de banda apenas em um determinado par de nós. Contudo, tal projeto é construído sob o conceito de SDN.

Em [107] o objetivo é utilizar totalmente a banda disponível e realizar tantas transferências quanto possível, respeitando o limite de tempo total de cada requisição. Assim, é proposta uma abstração da rede ICD baseada no prazo, ou seja, atenta às restrições de limite de tempo total de cada requisição. A ideia é permitir que clientes definam seus prazo apropriados e então oferecer-lhes alocação flexível de banda. Essa abstração é oferecida através do sistema *Amoeba*, cujo propósito é aumentar a taxa de aceitação de chamadas e utilizar a rede eficientemente. O *Amoeba* funciona com múltiplos tipos de trá-

fego. Suas políticas de alocação de banda são baseadas em classes prioritárias e no prazo. Além disso, duas políticas de escalamento são implementadas: escalonamento adaptativo e reescalonamento oportunístico. No primeiro caso, novas requisições são escalonadas sem mudar a banda das requisições existentes, e a seleção de caminhos obedece a ordem de grandeza das larguras de banda residuais nos enlaces, e no segundo caso, para cada nova requisição que deve ser acomodada, as requisições existentes podem sofrer uma redução de tempo e serem reescalonadas fora do prazo das requisições acomodadas. Entretanto, essa solução lida apenas com requisições individuais.

O mesmo problema de restrição baseado no prazo foi abordado em [111], que propõe uma solução de roteamento para transferências de dados em massa cuja restrição de prazo seja garantida e assim, diminua a probabilidade de falhas de transferências devido à falhas na rede, tratando-se de um roteamento ciente das limitações na camada física(probabilidade de falhas). Como as requisições transportam uma banda relativamente grande de dados pela fibra, falhas de rede (por exemplo, fibra ou falha no amplificador) são inevitáveis. E, normalmente,em tais movimentações, o tempo de transferência é comparável com o tempo de reparo falha. Daí as falhas de rede geralmente causarem a violação dos prazos. Todas as requisições são atendidas com taxa mínima.

Uma solução de roteamento EON que também utiliza tipos diferentes de requisições, incluinto BDT, é encontrada em [62]. O propósito dessa solução é maximizar o percentual de transferências completadas ou minimizar o número de transferências não finalizadas e ainda, reciclar fragmentos do espectro, possível a partir de algoritmos de otimização *offline* e de roteamento convencional de aprovisionamento dinâmico, para requisições orientadas a fluxos e a dados. Fluxos são atendidos com maior prioridade em relação às transferências de dados. Somente após ao atendimento dos fluxos e mediante a disponibilidade suficiente de fragmentos do espectro, um esquema de reserva de banda é acionado para atribuir recurso às transferências. Ademais, os *bulks* podem ser pausados quando necessário em certos intervalos de tempo.

Em [55] as transferências de dados são feitas utilizando a política SnF, com armazenamento temporário de dados ao longo do caminho, antes da entrega final ao destino. O escalonamento de requisições obedece a disponibilidade de largura de banda e de recurso de armazenamento nos pontos da rota definida. Um *framework* com arquitetura de grafo multicamada é construído para a alocação coordenada de recursos espaço-temporal. Porém, tal solução acarreta em maior complexidade devido ao número de camadas que reflete o número de requisições ativas, sendo viável apenas em pequenas redes.

3.3 Literatura de Referência

A proposta deste trabalho é desenvolver uma solução de RSA que seja ciente da aplicação, capaz de realizar ressincronizações ICD. A tecnologia pensada para o subtrato dessa rede ICD é a EON. A Seção 3.1 destacou as pesquisas que apontam o potencial dessa tecnologia, tanto em termos de capacidade quanto de redução de custos, uma vez que sua viabilidade também tem sido garantida com relação ao consumo de energia de maneira sustentável. Do ponto de vista da rede, soluções para os problemas de roteamento em EON são apresentadas em duas principais frente de pesquisa: roteamento convencional e roteamento ciente.

Com relação às soluções convencionais, focadas apenas na camada de rede, [46] e [57] apresentam um RSA estático, sendo este último voltado para *multicast*. Já [86, 96] apresentam RSAs dinâmico, sendo que este último resolve os problemas, roteamento e alocação, de forma reunida (única etapa) ou de forma separa (em duas etapas), para então propor adaptações quanto à modulações escolhidas. Por sua vez, [58] propõe um RSA *multicast* dinâmico utilizando medidas de fragmentação do espectro para escolher os caminhos com mais recursos disponíveis, enquanto que [6] e [19] propõe um RSA dinâmico multicaminho. Ademais, [60] e [62] propõem RSAs considerando chamadas orientadas a fluxos e a dados, sendo que esse primeiro faz alocações dinâmicas e estáticas, respectivamente, para ambos os tipos de chamadas, enquanto que o segunda, trata de alocações dinâmicas, também modelando como problema de otimização (PLI). Já [22] propõe um RMLSA estático.

Considerando-se as soluções de roteamento CLD, uma grande parte foca em roteamento ciente das limitações da camada física. Contudo, foi identificado em [72] uma solução AA-RSA empregada na criação de VDC de maneira coordenada com os recursos dos CD físicos e da rede, melhorando a agilidade e limitando a tendência dos operadores das aplicações de solicitarem mais servidores virtuais do que eles realmente precisam. A proposta desta dissertação também é aplicada ao cenário ICD, mas o problema tratado se refere a ressincronização de CD.

Desta maneira, do ponto de vista da rede, no panorama de soluções RSA não foram encontradas propostas que resolvam o problema das MBDTs. Assim, para atender esse tipo de requisição seria necessário tratar o grupo de requisições como uma coleção de requisições individuais sem nenhuma relação entre elas, para então tentar atendê-las.

Agora, do ponto de vista da aplicação MBDT, a literatura tem demonstrado as alternativas empregadas atualmente no atendimento da maioria dos serviços que lidam com grandes volumes de dados e larga variedade de tráfego. Entre os trabalhos identificados, [18] realiza BDT multidomínio dinâmico com encaminhamento E2E em um cenário com duas aplicações de classes diferentes; [103] também faz BDT via E2E através de um *fra-*

mework projetado na camada elétrica em um ambiente com uma única classe de tráfego; [73] realiza BDT dinâmica via E2E na rede ICD e também considera apenas tráfego *background*; [65] trata de BDT com menor prioridade poque faz diferenciação de classes de tráfego em uma rede de sobreposição (camada eletrônica). Nenhuma dessas soluções são apropriadas para EON e também não abordam CLD.

O encaminhamento E2E também é realizado com escalonamento de requisições. [62] faz BDT com encaminhamento E2E em EON, aproveitando a sobra de espectro remanescente dos fluxos atendidos para escalonar tais chamadas, mas não garante o prazo de atendimento das transferências; [107] considera as classes de tráfego e, devido a menor prioridade atribuída às MBDTs, essas chamadas são escalonadas e o encaminhamento é E2E na camada eletrônica; [11] realiza escalonamento de BDT com roteamento dinâmico e encaminhamento E2E em uma rede de sobreposição; [92] e [9] fazem escalonamento de BDTs dinâmicas com encaminhamento E2E, mas os dados transportados são relacionados a uma aplicação de criação de VMs. Ambos utilizam EON, mas enquanto o primeiro implementa uma camada extra para traduzir as requisições que chegam como sendo de fluxo e se tornam de dados, o segundo emprega elementos de SDN; Nesse grupo existem soluções para EON, mas não tratam de CLD.

A única solução identificada que trata de CLD é apresentada em [111]. Na proposta, as BDT são escalonadas e atendidas com encaminhamento E2E, considerando a probabilidade de falhas devido a problemas na camada física da rede. Entretanto, a solução é implementada na camada elétrica e os prazos das transferências não são garantidos.

As soluções apresentadas em [53, 21, 54, 100, 59, 55] implementam a técnica SnF para encaminhamento de dados, precisando alugar serviços de armazenamento em diversos pontos da rede geo-distribuída. Nenhuma das propostas são voltadas para EON. A solução proposta nesta dissertação dispensa esses serviços de armazenamento. Além do mais, nesse grupo de soluções apresentadas, apenas [100, 53, 54] tratam de MBDTs, mas sem especificação do tipo de aplicação que esteja executando tais solicitações. Como não se trata de CLD, suas múltiplas transferências são um processo no qual são desencadeadas várias BDTs ou simultaneamente, ou em um espaço de tempo muito pequeno.

Dessa maneira, as soluções apresentadas neste trabalho foram baseadas em [96], especificamente na proposta de resolução do problema RSA dinâmico completo, definido em uma única etapa, ou seja, a busca de rota e verificação da disponibilidade do espectro é feita no mesmo passo. Essa solução RSA é um roteamento convencional (RSAC), que não é baseado em CLD. Os parâmetros da EON definidos nesse trabalho serviram para a configuração do ambiente de simulação empregado. O objetivo principal de comparar essa solução convencional com as soluções cientes da aplicação, propostas nos próximos capítulos, é verificar o ganho de desempenho que é possível alcançar com relação a serviços

específicos executados, como as ressincronizações ICD.

3.4 Resumo Conclusivo

Este capítulo faz um levantamento da literatura de EON e seus problemas fundamentais, bem como das operações BDT e MBDT. Destacou-se as perspectivas futuras do paradigma de redes elásticas que será implementado nas redes de núcleo, bem como as soluções desenvolvidas para tratar dos problemas fundamentais RSA e RMLSA, no provimento de recursos às demandas. O paradigma CLD, já amplamente disseminado no campo de pesquisa das redes ópticas no que tange às limitações da camada física, ainda é timidamente abordado quando se refere à rede ciente da camada de aplicação. A solução identificada [72], que realiza um RSA ciente da aplicação, distingue-se da proposta deste trabalho por focar em uma aplicação diferente, que impede comparações justas com as aplicação de ressincronização.

As soluções de transferências de dados que foram ressaltadas, tanto de operações BDT quanto MBDT, também não versam sobre o paradigma CLD. Por esse motivo, a solução RSA convencional proposta por [96], que realiza roteamento e designação de espectro óptico de maneira dinâmica e em um único passo, será considerada como referencial no decorrer deste trabalho.

Capítulo 4

Roteamento e Alocação de Espectro Ciente da Aplicação em Redes Ópticas Elásticas

Este capítulo propõe uma solução de RSA com modulação fixa no âmbito do roteamento ciente das aplicações, que servirá para efetuar a ressincronização de uma partição de dados em um CD que estava indisponível na rede, e ao retornar, precisa atualizar o seu estado enquanto participante de um grupo de replicação. Na solução mostrada será considerado o número de falhas $f = 1$. Os resultados comparativos, feitos com uma outra solução RSA convencional (RSAC) [96] são apresentados para demonstrar o ganho de eficiência que pode ser alcançado quando a camada de rede possui mais informações para auxiliar a tomada de decisões. Verificou-se que o algoritmo ciente da aplicação estabelece até três vezes mais conexões para ressincronizações em comparação com o roteamento convencional, mantendo seu desempenho superior, em até 32%, mesmo em condições de tráfego pesado na rede.

Para chegar a tal resultado, os seguintes passos foram tomados: *(i)* preparação do simulador de redes ópticas elásticas e configuração do ambiente de rede ICD assim como dos mecanismos de controle para lidar com requisições do tipo orientadas a dados; *(ii)* proposta de um algoritmo de roteamento em EON ciente da aplicação de MBDT, chamado AA-RSA; e *(iii)* realização de uma série de simulações para avaliar o desempenho do algoritmo proposto face a um algoritmo RSA convencional, as quais demonstram claramente os ganhos advindos com a troca de informações entre as camadas de aplicação e de rede.

4.1 Problema das Múltiplas Transferências de Dados em Massa na Ressincronização ICD

Um sistema distribuído capaz de tolerar falhas de *hardware* e até intrusões precisa realizar replicações geo-distribuídas dos seus dados. Cada nó de armazenamento desse sistema é responsável por um conjunto de informações, chamado de partição, que são replicados separadamente. Um grupo de replicação é um grupo de nós responsável pela mesma partição. Assim, cada CD pode participar simultaneamente de vários grupos diferentes. Os sistemas de gerenciamento de dados distribuídos (SGDD) que lidam com esses CD possuem um correto e consistente mapeamento dos grupos de replicação em seus membros e suas respectivas localizações [78].

Várias rodadas de trocas de mensagens dentro de determinado grupo, que são definidas de acordo com cada protocolo específico, são realizadas quando CD submetem requisições entre si. O pressuposto fundamental para que este mecanismo possa ocorrer é o estado comum das réplicas, alcançado mediante sincronizações e ressincronizações [15]. Um nó membro inativo que deseja voltar à rede após um período de indisponibilidade e cujo seu estado encontra-se desatualizado, pode submeter uma solicitação de integração a esse grupo e assim, acionar os serviços de ressincronização. O grupo de réplicas designado pelo SGDD possui o mesmo estado e um mapeamento das partições a serem replicadas [4].

Para atender a solicitação de ressincronização do nó que está retornando, o sistemas de gerenciamento de dados distribuídos define os respectivos nós candidatos capazes de atender a essa ressincronização e dispara as MBDTs dentro de um período de tempo suficiente para a completa atualização do nó solicitante [3].

Do ponto de vista da camada de rede, normalmente a aplicação MBDT é atendida juntamente com outras solicitações de roteamento, embora suas taxas de dados sejam muito variadas, sem distinção do tipo de tráfego ou do seu nível de prioridade, para as quais o roteamento estabelece um caminho com largura de banda disponível e move o tráfego fim-a-fim (E2E). Logo, embora as MBDTs sejam tolerantes ao atraso, algumas requisições podem não ser atendidas devido a indisponibilidade de recurso suficiente por um período estendido de tempo e, consequentemente, estrangulamento do seu prazo [107].

A solução de roteamento em EON ciente da aplicação MBDT para ressincronização de CDs, AA-RSA, impõe maior complexidade por incorporar tomadas de decisões relacionadas a busca de combinações de um subconjunto dos nós aptos à transferência. Assim, o problema RSA tem seu espaço de busca ampliado de tal maneira que, além de tentar estabelecer caminhos ópticos com fatias espectrais suficientes, essas tentativas são experimentadas para todas os subconjuntos de 3 CDs transmissores que integram o grupo de replicação. Neste trabalho, foram considerados grupos de transferências de 3 CDs porque

o número de nós que compõem um grupo de ressincronização são geralmente 4 ou 5, sendo que as combinações de requisições consideram apenas os nós que farão transferências, que são 3 ou 4 [94].

Esta solução é possível a partir da comunicação vertical com o DMS, que é encarregado de informar o tipo de tráfego, de aplicação e respectivo QoS, a ser movido, e de fornecer uma abstração da localização dos membros participantes do grupo de replicação, normalmente inacessível pelo roteamento. Essa troca de informações é garantida via CLD. A decisão por algum subconjunto de replicação atende aos requisitos de tolerância a falhas da aplicação, garante melhor desempenho na ressincronização (mais requisições são aceitas) e reduz a utilização de recursos atendendo somente ao mínimo de transferências necessárias.

Para resolver o problema AA-RSA, a topologia física da EON foi modelada como um grafo direcionado $G(V, E)$, onde V e E denotam o conjunto de nós e enlaces da fibra, respectivamente. Assume-se que V é constituído por BV-WXCs e CDs geograficamente distribuídos e interconectados, representados como V^h. O conjunto V^h representa as localizações de CDs e comutadores ópticos simultaneamente, e portanto, esse tipo de nó é capaz de solicitar e receber uma ressincronização. Cada enlace pode acomodar, no máximo, BW *slots* de frequência do espectro.

Sobre essa rede é realizado o roteamento e alocação desses *slots* para determinar um caminho k com largura de banda B disponível a fim de transportar o tráfego ICD do tipo BDT, que são chamadas orientadas a dados [61]. Uma requisição BDT é um *bulk*, modelado como $r_u = \{s_u, d_u, C_u, Dl_u\}$, onde $s_u \in V^h$ e $d_u \in V^h - \{s_u\}$ são a origem e o destino da chamada, C_u é a quantidade de dados a ser transferida e Dl_u é o prazo da transferência.

A requisição de ressincronização MBDT encaminhada do DMS e recebida no roteamento, é um conjunto (lote) de *bulks* r_u de tamanho n representado como $R = \{r_{u_1}, r_{u_2}, ..., r_{u_n}\}$, onde existe um conjunto S contendo todas as origens das transmissões, sendo d e Dl iguais para todas as requisições.

Como principal característica deste tipo de aplicação, sua redundância inerente requer que a rede ICD seja altamente tolerante a falhas multi-nós. Por isso, um conjunto R_u deve ter uma cardinalidade mínima de 3 *bulks*. Com este conhecimento, as demandas de conexão terão semânticas que consideram a camada de aplicação, especificando o grupo de redundância participante, o que restringe o espaço de buscas por nós candidatos por parte da camada de rede. Devido a esta restrição, isto é, alocar recursos para as três transferências, o roteamento não tenta realizar alocações se a requisição viola esta condição.

Para exemplificar, a Figura 4.1 mostra um processo de ressincronização para o CD_5. Um lote MBDT com $S = \{CD_1, CD_2, CD_3, CD_4\}$ é enviado para o $d = \{CD_5\}$, movimentando um volume $C = \{1TB\}$, que deve ser recebido no destino dentro de um limite de tempo $Dl = \{30min\}$.

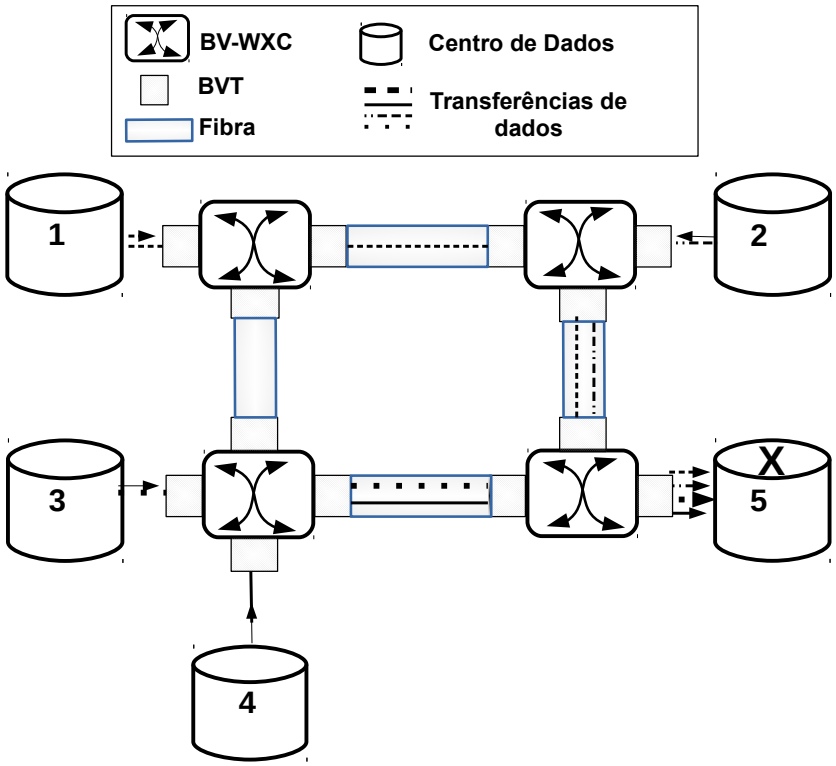

Figura 4.1: MBDT dos CD_1, CD_2, CD_3, CD_4 para ressincronizar o CD_5.

Um algoritmo convencional, como é o caso do RSAC [96], adaptado para tratar de requisições orientadas a dados, considera esse lote como um conjunto de quatro chamadas independentes para as quais a busca por recursos é feita de maneira sequencial conforme a chegada dessas requisições. Nesse cenário, pode ocorrer de menos do que três chamadas serem atendidas e as demais serem bloqueadas devido ao esgotamento do tempo enquanto se aguardava por recursos de largura de banda ou mesmo devido a sua indisponibilidade nos enlaces compartilhados. Quando isso acontece, mesmo que algumas chamadas individuais sejam atendidas, não ocorre uma ressincronização com garantias de tolerância a falhas.

Ciente da importância dessa restrição para a aplicação, o roteamento pode esforçar-se em atender ao menos 3 *bulks* desse lote, limite inferior denominado fator de replicação, evitando bloqueá-las antes de ter certeza sobre a inexistência de recursos. Para isso, um algoritmo de combinações de lote pode acessar um lote e buscar por um subconjunto mínimo com essas características. Na Figura 4.1, as possíveis combinações seriam:

$$Comb(S) = \begin{cases} \{CD_1, CD_2, CD_3\} \\ \{CD_1, CD_2, CD_4\} \\ \{CD_1, CD_3, CD_4\} \\ \{CD_2, CD_3, CD_4\} \end{cases}$$

O AA-RSA propõe a busca por uma combinação mínima dos elementos de S. O atendimento de menos do que três demandas quebra a restrição de tolerância a falhas da aplicação de ressincronização de CDs. Por outro lado, atender mais do que 3 chamadas levaria ao desperdício de recursos [94].

4.2 Algoritmo AA-RSA

O algoritmo AA-RSA (Algoritmo 1) baseia-se no algoritmo RSA com modulação fixa [96]. Nas linhas 1 e 2 são inicializados o fator de replicação b, que representa a quantidade mínima de réplicas de CDs para a ressincronização, e K, a quantidade de menores caminhos entre quaisquer dois nós. A função $Combinação\binom{R_u}{b}$, na linha 3, calcula todas as combinações possíveis de b requisições dentro do universo do lote R_u, que contém n chamadas para efetuar uma ressincronização. O seu processamento retorna uma matriz m de $\frac{n!}{b!(n-b)!}$ linhas e b colunas. Cada linha de m é um subconjunto de combinações *comb* de tamanho b. Em cada *comb* executa-se para cada um de seus *bulks* i a função $KSP(i, K)$ [1], linha 5, que retorna os K menores caminhos entre $s(i)$ e $d(i)$, origem e destino da chamada i, respectivamente, em ordem crescente de tamanho.

Na busca do k-ésimo caminho viável, nas linhas 6-13, para cada candidato $k \in K$ calcula-se a distância de k, linha 7, que é comparada com a fórmula $\mathcal{D} \times \tau$, onde $\mathcal{D}$ é o diâmetro da rede e τ é um parâmetro que restringe esse diâmetro a um limite de tamanho que permita utilizar a largura de banda máxima $MaxRate$ do caminho, equivalente à capacidade máxima de um BVT. Essa fórmula impede que requisições de conexões para os caminhos menores do que um determinado limite do maior caminho sejam penalizadas, com a alocação da mínima largura de banda disponível. A ideia é que recursos de caminhos relativamente pequenos sejam liberados mais rapidamente, desocupando recursos para as chamadas seguintes. Às requisições dos caminhos que não atendem a essa condição, na linha 10, é atribuída uma largura de banda B mínima, equivalente ao quociente da quantidade de dados transferida dentro do prazo total da sua chamada, $C_i \div Dl_i$.

Definidos o caminho e a largura de banda, na linha 12 é verificada a restrição de continuidade, de contiguidade e de não-sobreposição do espectro óptico. As linhas $14-17$ analisam se um caminho e sua respectiva taxa foi obtida para cada requisição do

Algoritmo 1 AA-RSA(G, R)

1: $b \leftarrow 3$
2: $K \leftarrow 3$
3: Combinação $\binom{R_u}{b}$
4: **para** $i \leftarrow 1$ **até** b **faça**
5: $KSP(i, K)$
6: **para** $k \leftarrow 1$ **até** $|K|$ **faça**
7: $dist\,(k) = \sum_{l=0}^{|k|-1} v_l, v_l \in k$
8: **se** $dist(k) \leqslant \mathcal{D} \times \tau$ **então**
9: $B \leftarrow MaxRate$
10: **senão** $B \leftarrow C_i \div Dl_i$
11: **fim se**
12: Testa restrições RSA
13: **fim para**
14: **se** $\exists k \in K \mid k$ é viável **então**
15: $\mathcal{P} \leftarrow k$
16: $\mathcal{B} \leftarrow B$
17: **fim se**
18: **fim para**
19: **se** *comb* pode ser atendida **então**
20: Aceita $(R, \mathcal{P}, \mathcal{B})$
21: **senão**
22: Bloqueia (R)
23: **fim se**

subconjunto de combinações, que em caso positivo, são previamente alocadas e guardadas em $\mathcal{P}$ e $\mathcal{B}$, conjunto dos caminhos da combinação e conjunto das larguras de banda das requisições dessa mesma combinação. A linha 19 verifica se um dado subconjunto de requisições, obtido na combinação, pode ser completamente atendido. São feitas tentativas com todos os subconjuntos de combinações até exaurir a matriz de combinações m ou até que encontre uma combinação viável. A aceitação das MBDTs para a ressincronização é feita na linha 20, caso um subconjunto possa ser aceito. Se não for o caso, a requisição é bloqueada (linha 22).

O algoritmo recebe um lote R como entrada, mas a aceitação do serviço se dá com o atendimento de apenas um subconjunto de R, o que significa que chamadas remanescentes são marcadas como bloqueadas pelo plano de controle da rede. Isso é possível porque na comunicação cruzada em camadas, o roteamento, em contato com a aplicação, sabe que se R é proveniente de um grupo de replicação e portanto, os dados das transferências são similares, então algumas das requisições podem ser bloqueadas sem prejuízo do serviço.

Complexidade do Algoritmo

O RSAC dinâmico com caminho fixo proposto por [96], que é dividido em dois passos, computa os k menores caminhos sem *loops* usando o algoritmo KSP [1] no primeiro desses passos com tempo $O(K|V|^3)$, e em seguida, utiliza operações de detecção e intersecção de espectro no segundo passo, levando $O(K^2)$ para a verificação em todos os enlaces que compõem o caminho. Assim, seu tempo total é $O(K^3|V|^3)$.

O AA-RSA dinâmico, baseado no RSAC acima, para obter as combinações de *bulks* do lote na função $Combinação\binom{R_u}{b}$, com um lote R de tamanho n e um fator de replicação b, tem complexidade de tempo $O\left(\frac{n!}{b!(n-b)!}\right)$ que é exponencial. Para cada elemento de uma combinação é invocada a função $KSP\,(i, K)$, com complexidade de $O(V^3)$. A atribuição de espectro é feita pela política *First-Fit* de complexidade linear. O diâmetro da rede $\mathcal{D}$ é calculado uma única vez. A partir do algoritmo de *Dijkstra* [23], que leva tempo $O(V^2)$, é encontrada a distância de cada nó s do grafo para todos os demais nós desse mesmo grafo. A maior distância dentre todas as menores distâncias representa $\mathcal{D}$. Assim, a complexidade de tempo é $O\left(\left(\frac{n!}{b!(n-b)!}\right) * (V^3)\right)$. No entanto, a literatura mostra que o tamanho de R geralmente é de 3 requisições de conexões de ressincronizações, visto que no mundo real é desvantajoso, do ponto de vista do custo capital e operacional, possuir um conglomerado muito grande de recursos que são poucos solicitados [94].

4.3 Avaliação de Desempenho

Através do simulador ONS (*Optical Network Simulator*) [24][1] desenvolvido com base no simulador WDMSim [30], simulações foram realizadas para verificar o ganho de desempenho do algoritmo proposto em relação ao algoritmo RSAC [96]. Eventos dinâmicos de chegadas e partidas de requisições foram simuladas na topologia NSFNET (Figura 4.2) com 14 nós, dos quais cinco deles $(0, 7, 11, 12, 13)$ foram definidos como CDs, e nas topologias USA e Pan-Euro reunidas por cabo de fibra óptica submarino de capacidade suficiente, com 24 e 28 nós, respectivamente (Figura 4.3), onde os nós $\{1, 10, 12, 20, 21, 22, 28, 30, 46\}$ são as representações de CD. As distâncias físicas são destacadas nos enlaces. As definições dos CDs foram feitas baseadas nas localizações de CDs do *Google* [42], nas quais foram consideradas ressincronizações de uma única partição.

Na implementação foram considerados 15 *transponders* por nó, cada um com capacidade de 8 *slots*. Cada *slot* possui largura de banda de $12.5GHz$ [96] e cada enlace possui 120 *slots* de frequência. Assumiu-se ainda que são empregados dois *slots* como banda de guarda. Nos dois cenários, os dois algoritmos foram testados com as modulações BPSK,

[1] O simulador ONS é um simulador híbrido para redes WDM e EON, disponibilizado em http://comnet.unb.br/br/grupos/get/ons/download

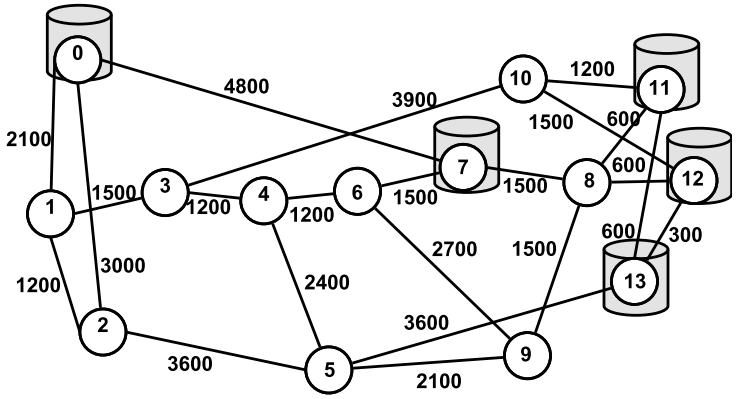

Figura 4.2: Topologia da rede NSFNET .

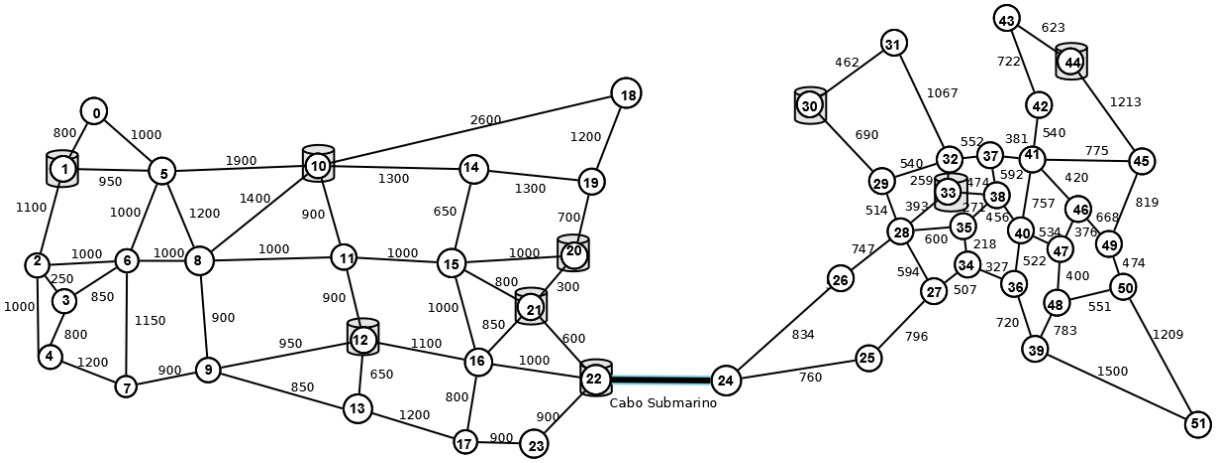

Figura 4.3: Topologia das redes USA e Pan-Euro, reunidas.

que transmite um *bit* por símbolo e por isso fornece baixa taxa de dados, mostrando-se ideal para as distâncias mais longas, e QPSK, com dois *bits* por símbolo e uma taxa de dados mais elevada, que permite transmitir duas vezes mais quantidade de dados no mesmo canal com a mesma taxa oferecida à BPSK, e que portanto, é mais adequada para distâncias menores. Além disso, por se tratar do problema RSA, a solução não considera distâncias para definição de qualquer modulação, no entanto, o ajuste da modulação implica na definição da quantidade de *bits* por símbolo empregados no transporte dos dados. Nos testes, essas modulações tiveram resultados próximos quanto a taxa de bloqueio, embora QPSK tenha se sobressaído.

Cada simulação foi realizada 5 vezes utilizando o método de replicações independentes. Para os resultados apresentados foram calculados intervalos de confiança com 95% de confiabilidade. Foram realizadas 100.000 chamadas com origens e destinos distribuídos uniformemente dentro do subconjunto de localizações dos CDs. Os prazos das chamadas

foram de 10, 15 e 20 unidades de tempo. Foram empregados dois cenários de carga, um deles considerado leve, com taxas de chegadas de 2 a 10 chamadas por unidade de tempo com incrementos de 2 chegadas, e um outro cenário com tráfego pesado, cujos números de chegadas de chamadas variam de 20 a 100 por unidade de tempo com incrementos de 20 chegadas. Quanto aos volumes dos *bulks*, foram definidas chamadas de $100GB$, $500GB$ e $1TB$. Para essas chamadas foram configurados lotes com 3 e 4 requisições, que são parâmetros largamente encontrados na literatura [4].

A simulação foi executada em uma máquina *Intel Core* $2Quad$ de 2.66 GHz com $4GB$ de RAM, onde verificou-se que o tempo médio de execução do algoritmo AA-RSA, para atendimento de um lote, é da ordem de $1,27ms$ com a menor topologia e $90,31ms$ com a topologia reunida.

A atribuição de espectro é realizada utilizando a política *First-Fit*, onde para cada enlace pertencente a rota estabelecida, são considerados os seus respectivos BW *slots* de frequência em ordem crescente dos índices, e a demanda é acomodada iniciado-se pelos *slots* de índices menores que estão com a capacidade livre. O parâmetro τ, de restrição do diâmetro da rede utilizado no algoritmo, foi assumido como $0,5$ de modo a comparar os tamanhos dos caminhos candidatos à metade do diâmetro da rede.

Taxa de Bloqueio de Banda Passante (BBR)

A taxa de bloqueio de largura de banda (BBR) de *bulks* equivale a taxa do uso de largura de banda correspondente aos *bulks* bloqueados dividida pelo total de largura de banda em uso por todas as chamadas. A Figura 4.4 mostra que os algoritmos RSACs bloqueia mais *bulks* do que os algoritmos AA-RSAs no cenário das topologias reunidas, onde os caminhos ópticos estabelecidos são maiores. No cenário de tráfego leve, os algoritmos mantêm o seu comportamento até o limite de 10 chegadas. Para as duas modulações testadas, o AA-RSA apresenta melhor desempenho do que o RSAC. Para ambos os algoritmos, a modulação QPSK resultou em menores taxas de bloqueio de largura de banda do que a modulação BPSK. O resultado de BBR do AA-RSA_QPSK inicia com uma taxa de cerca de 16% e chega a atingir 55% de bloqueio, ao passo que o RSAC com essa mesma modulação inicializa com 28% de bloqueio, chegando a 80%.

Cabe ressaltar que parte dos bloqueios do AA-RSA são na verdade chamadas descartadas dos lotes com 4 *bulks* e, portanto, não significa prejuízo no atendimento à aplicações. Quando ocorrem duas chegadas por unidade de tempo, cerca de 90% dos bloqueios dizem respeito a descartes, mas esta taxa cai para 17% quando o número de chegadas aumenta para 10. O AA-RSA_BPSK descarta menos chamadas do que o AA-RSA_QPSK. A explicação para o bom desempenho da modulação QPSK é a alta capacidade de transporte em relação a BPSK.

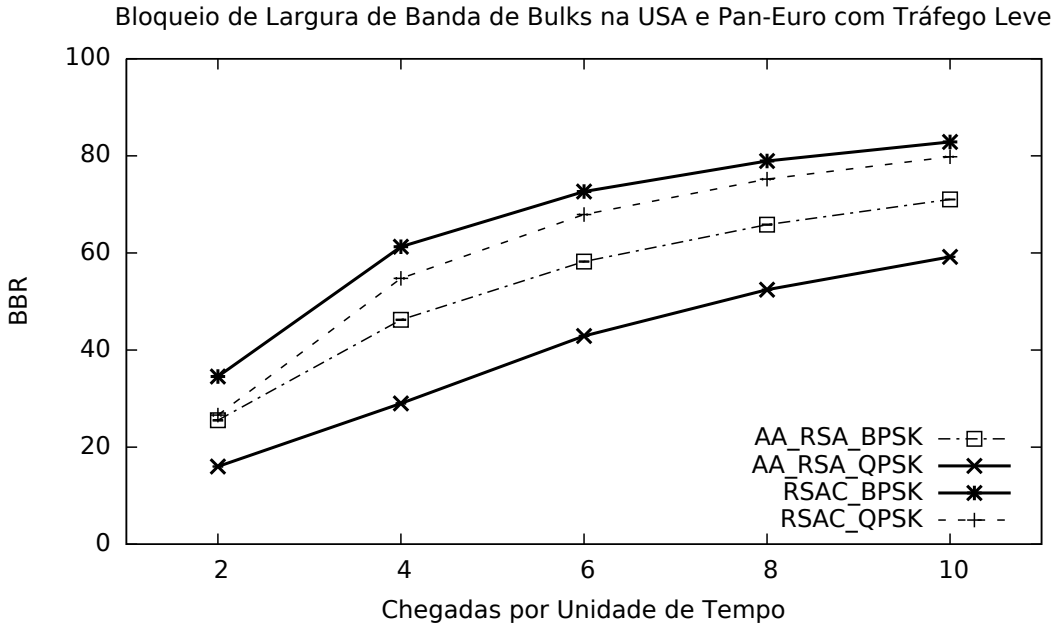

Figura 4.4: BBR de *bulks* nas topologias USA e Pan-Euro reunidas com tráfego leve.

Para o tráfego pesado (Figura 4.5), os dois RSACs tiveram taxa de desempenho próximas e seus percentuais de utilização de BVTs e de espectro óptico foram praticamente os mesmos, embora o RSAC_BPSK tenha bloqueado mais recursos. As diferenças em torno de 1% entre eles, mostram que em condições pesadas de tráfego, BPSK e QPSK exibem o mesmo comportamento. Já no caso do AA-RSA, o uso da modulação BPSK levou a mais bloqueio com uma desvantagem de cerca de 4% em comparação com QPSK. O mecanismo de busca por recurso disponível desse algoritmo eleva a taxa de utilização de BVTs e de espectro óptico em pelo menos 3%, levando a uma menor BBR. Os AA-RSAs, comparados com os RSACs, aumentam a taxa de utilização desses recursos em quase 50% a medida que o número de chegadas também cresce.

Já no cenário da rede NSFNET, percebeu-se que a utilização de BVTs manteve-se em torno de 22 e 24% pelo RSAC_BPSK e RSAC_QPSK, respectivamente, sendo que os resultados das taxas de bloqueio de largura de banda estiveram muito próximos em condições leves de tráfego, conforme mostra a Figura 4.6. A justificativa está na quantidade e nas localizações dos nós CDs. Em um subconjunto de CDs relativamente pequeno, a possibilidade de sorteio de um nó cujos recursos nos caminhos adjacentes a ele ainda não tenha sido desocupados é muito alta. Além disso, 80% deles são concentrados no lado leste da topologia enquanto que existe um único CD situado no lado oposto, resultando em mais transferências de dados entre os membros a leste do que entre qualquer um desses membros e o CD na região oposta da rede. Os algoritmos AA-RSAs bloquearam menos

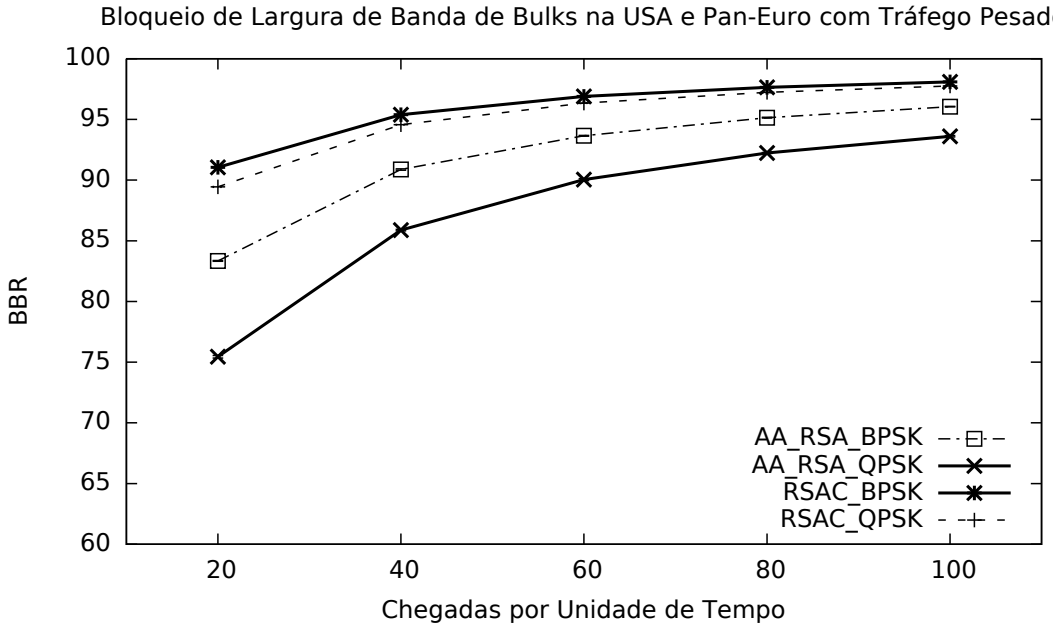

Figura 4.5: BBR de *bulks* nas topologias USA e Pan-Euro reunidas com tráfego pesado.

recursos e novamente a modulação QPSK resultou em menor taxa de bloqueio e menor taxa de utilização de BVTs e espectro.

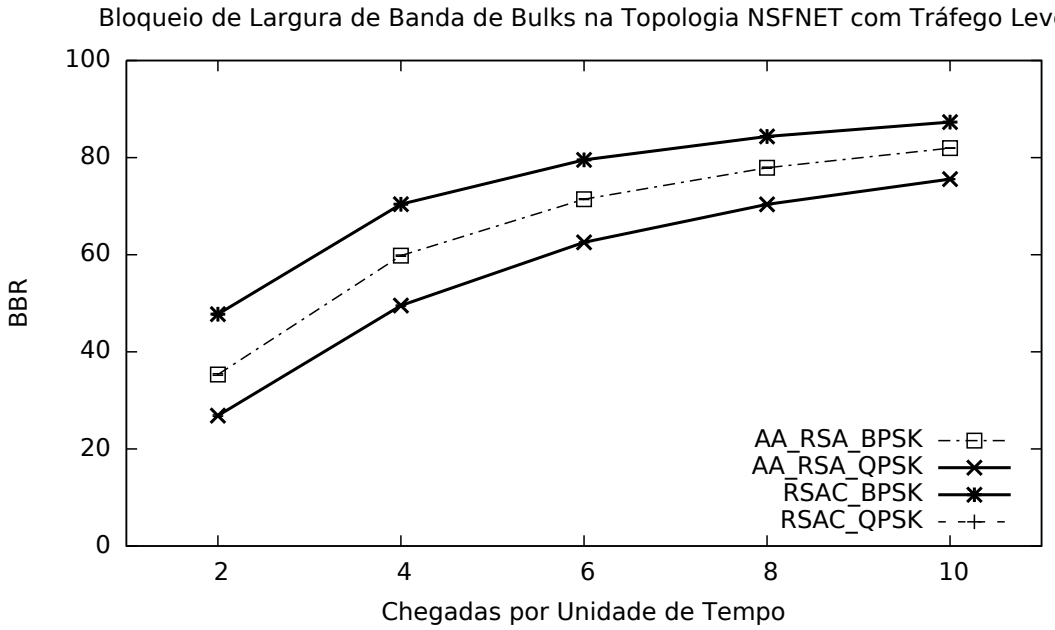

Figura 4.6: BBR de *bulks* na topologia NSFNET com tráfego leve .

No gráfico da Figura 4.7, percebe-se que o RSAC_QPSK tem uma taxa de bloqueio

por volta de 94% com 20 chegadas registradas, que aumenta para 96% com 40 chegadas e em seguida, atinge 97% na ocorrência de 60 chegadas por unidade de tempo, e de maneira análoga, o RSAC_BPSK repete esse comportamento. Novamente, as localizações dos nós sorteados e a ocupação de recursos dos canais adjacentes com tráfego pesado justificam esses resultados, visto que as condições de tráfego são as mesmas, independente do algoritmo utilizado. O AA-RSA_QPSK segue com menor bloqueio e experimenta uma rápida elevação logo com o aumento de tráfego, onde ocorre uma ligeira diminuição de utilização de BVTs, o que leva a entender que a taxa de ocupação da rede não permitiu o estabelecimento de novos caminhos ópticos para as requisições que solicitavam conexões. Sua diferença percentual para o AA-RSA_BPSK foi de $0,6\%$ com a máxima taxa de chegadas registradas.

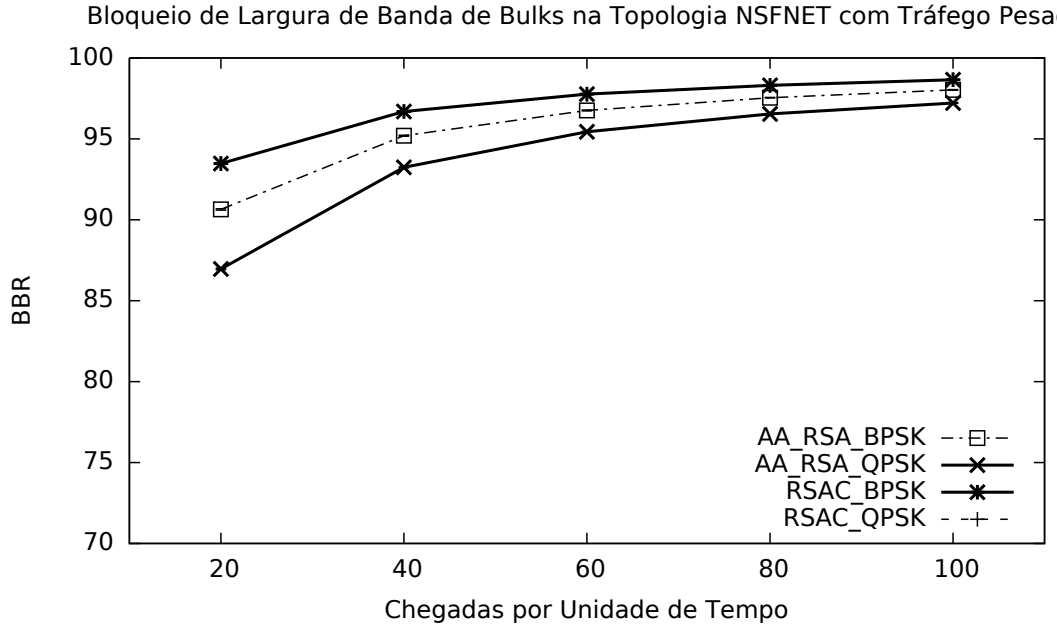

Figura 4.7: BBR de *bulks* na topologia NSFNET com tráfego pesado .

Note que nas simulações, os RSAs cientes da aplicação resultaram em menores taxas de bloqueio de largura de banda e maior utilização de BVTs devido a sua política de busca de uma combinação de chamadas para a qual existem recursos. Os RSACs tiveram taxas mais elevadas de bloqueio embora os número de recursos disponíveis como BVTs e espectro de frequência estivesse entre 6% e 12% a mais que os outros algoritmos. Cabe destacar ainda que, entre os RSACs, todos os resultados mostrados são muito próximos ou sobrepostos, o que leva a entender que não há diferença relevante quanto à escolha de qualquer modulação. Isso demonstra que o investimento na infraestrutura da rede,

como a aquisição de *transporders* com maior eficiência espectral, pode não resultar nas melhorias esperadas.

Avaliação da Qualidade do Serviço

A taxa de sucesso do serviço diz respeito às ressincronizações que se efetuaram, ou seja, o percentual de lotes atendidos. Nas topologias reunidas onde se registra o tráfego leve no gráfico da Figura 4.8, verifica-se que, como esperado, o AA-RSA_QPSK leva a um maior estabelecimento de conexões de lotes e consegue manter sua taxa de aceitação em nível elevado. O AA-RSA_BPSK tem desempenho inferior mas ainda assim, apresenta vantagens quando comparado com os RSACs. No gráfico da Figura 4.9, os algoritmos cientes da aplicação tendem a manter o mesmo comportamento, mesmo com uma queda um pouco acentuada para o QPSK até a ocorrência de 40 chegadas por unidade de tempo. Como a taxa de aceitação vai diminuindo, a taxa de utilização de recursos vai se reduzindo no mesmo passo. Por fim, com número máximo de chegadas, a versão com QPSK aceita $9,94\%$ das conexões enquanto que a versão com BPSK tem taxa de $6,44\%$.

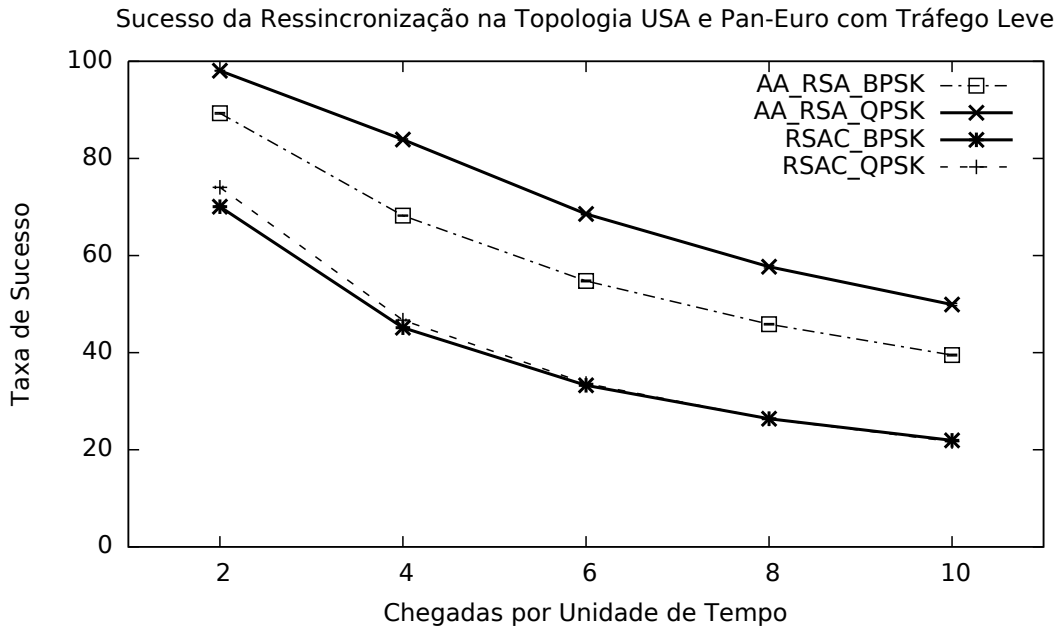

Figura 4.8: Taxa de sucesso de conexões atendidas nas topologias USA e Pan-Euro reunidas mediante tráfego leve.

Quanto aos dois RSACs na Figura 4.8, como as taxas de bloqueio de largura de banda foram altas nos resultados anteriores, esse fator se reflete nas taxas de aceitação de conexões. O QPSK começa com aceitação de pouco mais de 73% enquanto que o BPSK tem 69%. Como esses dois algoritmos tratam as requisições na sequência em que

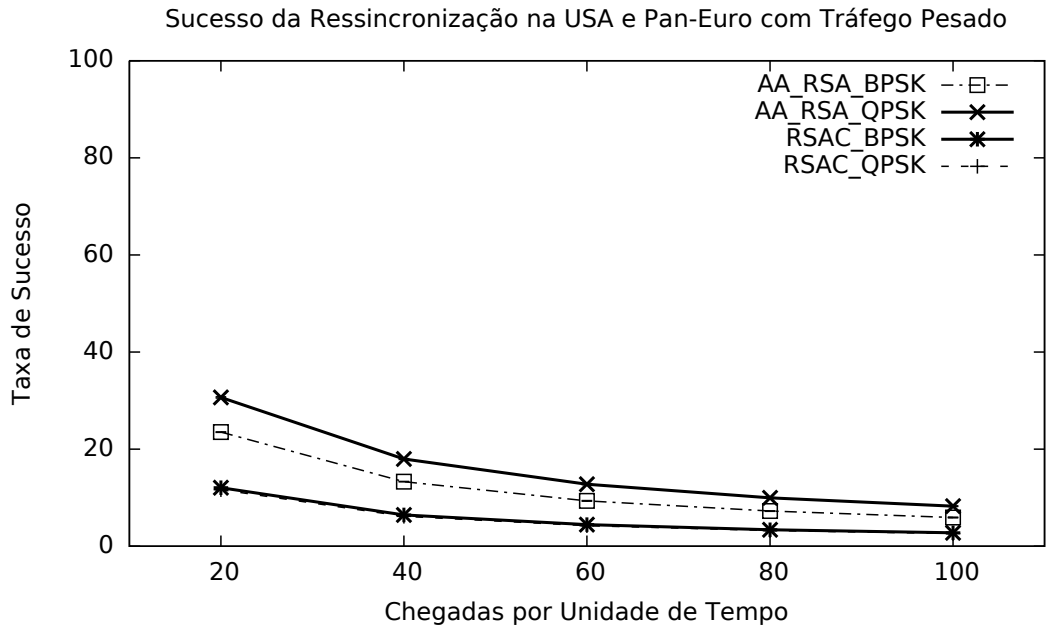

Figura 4.9: Taxa de sucesso de conexões atendidas nas topologias USA e Pan-Euro reunidas mediante tráfego pesado.

são dadas, a probabilidade de atender a pelo menos 3 chamadas de um lote e resultar em uma ressincronização é pequena. O mesmo desempenho é percebido no gráfico da Figura 4.9, quando seus resultados tendem a ser sobrepostos, uma característica própria desses algoritmos quando os recursos estão saturados.

Já com relação a topologia NSFNET (Figura 4.10 e Figura 4.11) foi registrado que a taxa de aceitação nessa rede é inferior se comparada com as topologias reunidas, mostradas anteriormente, visto que nessa rede os recursos se esgotam muito mais rápido. O gráfico da Figura 4.11 mostra queda acentuada da taxa de aceitação do AA-RSA_QPSK em condições de tráfego pesado, em uma rede com concentração de nós. Os RSACs conservam a taxa de utilização de BVTs na faixa dos 43 a 44% e 11 a 13% de utilização do espectro, enquanto que os AA-RSAs tem variações de 28 a 32% de utilização de BVTs e até 25% de utilização do espectro. Com a diminuição de aceitações, o número de chamadas descartadas tende a diminuir enquanto que o número de chamadas efetivamente bloqueadas aumenta, e lotes acabam sendo inteiramente bloqueados devido ao alto congestionamento na rede por conexões ativas.

Em todos os cenários de rede analisados, os algoritmos AA-RSAs obtiveram bons resultados em termos de ressincronizações efetuadas e reduzida taxa de bloqueio. Também é possível atingir diferentes resultados de acordo com os níveis de eficiência espectral aplicado na transmissão, com margem de diferença entre o AA-RSA_BPSK e AA-

RSA_QPSK em torno de 16%. Já entre os RSACs praticamente não há diferença com relação às duas eficiências espectrais utilizada (BPSK e QPSK).

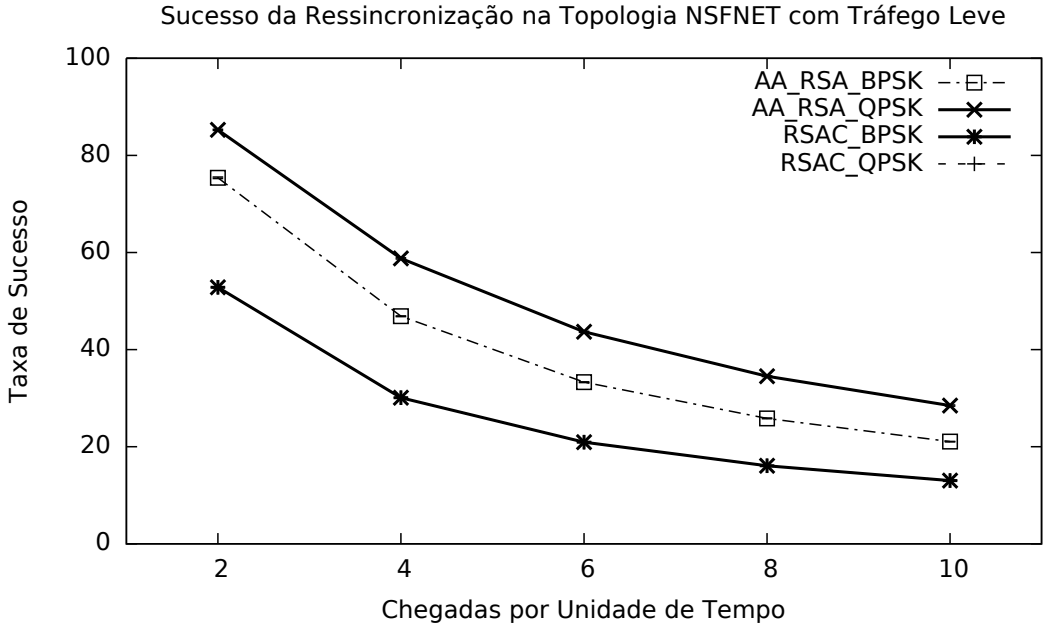

Figura 4.10: Taxa de sucesso de conexões atendidas na topologia NSFNET mediante tráfego leve.

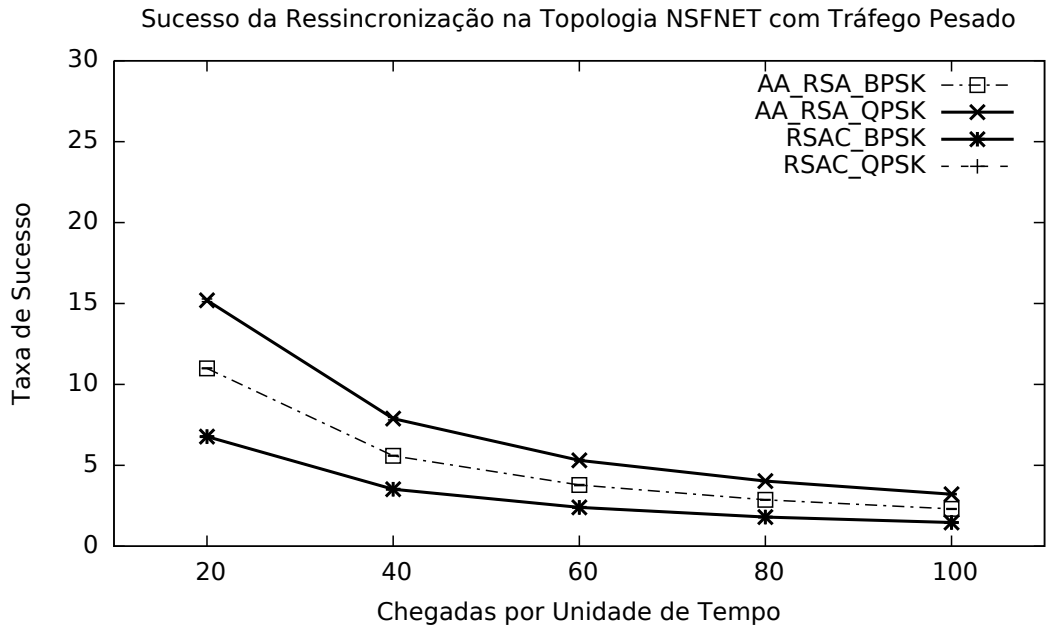

Figura 4.11: Taxa de sucesso de conexões atendidas na topologia NSFNET mediante tráfego pesado.

De maneira geral, os resultados mostram que algoritmos de roteamento que conhecem a aplicação para a qual precisam fornecer recursos tendem a ter melhor desempenho do ponto de vista dos serviços se comparado com algoritmos convencionais. Além disso, fazer investimentos na infraestrutura da rede para acrescentar novas interfaces de modulações não representa melhoras significativas de desempenho se o algoritmo de roteamento for convencional.

4.4 Resumo Conclusivo

Este trabalho propõe uma solução de roteamento ciente da aplicação que eficientemente atende ao serviço de ressincronização ICD e reduz as taxas de bloqueio de banda de chamadas orientadas a dados. Os resultados comparativos mostram que a aplicação MBDT alcança maior sucesso quando a abordagem CLD é aplicada, visto que as características da aplicação podem servir de entrada para as tomadas de decisões na camada de rede. As soluções cientes demonstraram atender até 30% a mais de requisições, em comparação com a soluções convencional. Interfaces de modulação com maior eficiência espectral podem representar um aumento de até 16% no atendimento entre soluções cientes, enquanto que entre as soluções convencionais, não há vantagens significativas. Ademais,

percebeu-se que, para os ISP, é mais vantajoso investir em infraestrutura se os protocolos de roteamento são construídos baseados em CLD.

Capítulo 5

Escalonamento de Múltiplas Transferências de Dados em Massa

No Capítulo 4 foi feita a comparação entre algoritmos de roteamento cientes da aplicação e convencional, aplicada à ressincronização de centros de dados, onde se obteve um ganho de até 30% no estabelecimento de conexões bem sucedidas em um cenário onde todas as requisições eram da mesma natureza. Os CDs que participavam das transferências estavam definidos em localizações realistas e notou-se que, nas duas topologias adotadas, NSFNET e Pan-Euro reunida à USA, a maior concentração de nós em parte da rede levou ao rápido esgotamento de recursos, e nessa segunda rede, o gargalo identificado no enlace que as conectavam, resultou em grande impacto no bloqueio de requisições. A partir deste ponto, uma nova variação desse problema é obtida, agora considerando-se a realização de ressincronizações em um ambiente onde existe competição por recursos devido a execução de um segundo tipo de aplicação. Um serviço de rede viável para essas aplicações deve suportá-las com QoS admissível e sem degradações. Além disso, essas análises serão feitas em topologias de redes isoladas para minimizar os impactos associados à características de um cenário específico.

Este capítulo propõe soluções cientes da aplicação que realizam ressincronizações de dados e ao mesmo tempo, mantêm reduzida a probabilidade de bloqueio de outras conexões. Devido a disputa por recursos disponíveis, novas oportunidades de atendimento serão oferecidas para cada chamada no decorrer do seu prazo de atendimento. Para tanto, será empregado escalonamento de requisições. Uma requisição que deseje transferir dados para um determinado centro de dados e não encontre largura de banda suficiente em qualquer um dos seus caminhos candidatos, poderá aguardar para que uma nova busca seja feita em um tempo posterior, atendendo-se ao limite de restrição do seu prazo.

Os problemas de escalonamento de requisições onde existe conflito de tempo e recurso (bi-conflituoso) são da classe NP-Díficil [11], o que significa dizer que se cada problema

pertencente à classe NP é redutível ao problema de escalonamento bi-conflituoso, então esse problema de escalonamento é tão difícil quanto todos os problemas da classe NP. Cabe ressaltar que a classe NP reúne todos os problemas de decisão para os quais é possível verificar em tempo polinomial todos os casos em que a resposta é sim [23].

Entretanto, as heurísticas empregadas para o escalonamento baseiam-se na política FIFO (*First in, First out*) e fila de prioridades. As redes de melhor esforço de hoje priorizam igualmente todos os fluxos de dados e não distinguem operações de alta prioridade de operações de movimentação de dados sensíveis ao tempo com o resto dos fluxos de dados concorrentes na rede, embora o ideal não seja manter prioridade fixa para fluxos individuais [49].

5.1 Problema do Escalonamento de Requisições de Aplicações Diferentes

Quando duas aplicações distintas precisam ser executadas na rede ICD, os provedores analisam as classes de tráfego referentes e utilizam tal informação para a definição da largura de banda. Esse dado é utilizado pelo algoritmo de roteamento, que segue as prioridades identificadas pelas classes, o que constitui a política de alocação hierárquica. As diferentes requisições são prontamente executadas, e devido às políticas de QoS, as chamadas de mais alta prioridade são garantidas em detrimento das chamadas de prioridade menor [18]. Isso significa que em um ambiente com escassez de recurso, o tráfego de classes inferiores pode ser bloqueado para garantir os fluxos de maior prioridade. É por esse motivo que grandes provedores de CDs preveem grande parte das suas despesas dedicadas ao aluguel de serviços de hospedagem de dados em *hosts* ao longo do caminho da transmissão, enquanto aguardam que a banda suficiente seja disponibilizada para completar o serviço de transferência [65, 107].

Se as requisições solicitando recurso forem de mesma classe, às conexões a serem estabelecidas poderão ser atribuídas a mesma banda para o atendimento de forma justa. Entretanto, com o aumento no número de solicitações simultâneas, os canais da rede tornam-se saturados limitando o atendimento, levando a uma elevada taxa de bloqueio [100, 53].

No Capítulo 4 foi proposta uma solução de roteamento ciente da aplicação (AA-RSA), capaz de lidar com requisições em lote, isto é, MBDT, que teve seu desempenho comparado a uma solução de roteamento convencional, RSAC [96], em um cenário de uma única aplicação sendo executada para ressincronizar CDs. O algoritmo AA-RSA faz combinações de chamadas dentro de um lote para aumentar as chances de encontrar um subconjunto de três requisições para o qual seja possível encontrar recursos disponíveis. Além disso, a quantidade de banda atribuída a uma chamada é diretamente ligada à referência do

diâmetro da rede. Se o caminho definido no roteamento é inferior à metade do diâmetro, a chamada é alocada com taxa máxima de transmissão. Caso contrário, a taxa mínima é atribuída. Já o algoritmo RSAC [96], que não é capaz de identificar o tipo de requisição com a qual lida, trata as chamadas de um lote como requisições individuais.

Neste capítulo propõem-se que uma segunda aplicação seja executada, cujo tráfego seja da mesma classe que o tráfego da primeira aplicação, com o objetivo de garantir que o serviço de ressincronização seja atendido satisfatoriamente, afinal, CDs subutilizados ou inativos simbolizam utilização ineficiente de recursos, risco de perda de dados, ameaça à proteção contra falhas e prejuízo para os negócios, e além disso, os enlaces ligados aos CDs ativos são sobrecarregados com o aumento de solicitações de serviços [94].

As aplicações definidas são da classe de tráfego *backgroud* e incorrem em elevado volume de dados sendo transportado. Uma das aplicações aciona requisições de conexões individuais, como um simples serviço de *backup* de dados por exemplo, com restrição de tempo e volume de dados definidos. Os serviços de *backups* são parte essenciais do planejamento de proteção contra desastres, oferecendo capacidade satisfatória de recuperação [104]. A outra é uma aplicação de MBDT voltada a ressincronização de dados em um data-center com estado desatualizado, conforme proposto no Capítulo 4. Pelas garantias de tolerância a faltas, ao menos três nós participarão da operação e enviarão os dados solicitados pelo nó inativo, sendo que a ressincronização é definida sobre uma partição de dados [4, 3, 15].

Então, para aprovisionar recursos para requisições de conexão provenientes de aplicações de *backups* e ressincronizações, é preciso destacar as possíveis configurações de recursos que podem ser feitas, a implementação de um mecanismo de fila de espera e os tipos diferentes de taxas máximas e taxas mínimas que podem ser oferecidos. Dessa forma, alguns casos podem ser elaborados, conforme descrição a seguir:

1. Em uma situação hipotética simples, se tanto as requisições de *backup* quanto as requisições de ressincronização forem atendidas com a banda mínima suficiente compatível com o prazo da transmissão, é provável que por longos períodos os enlaces da rede que compõem o caminho, permaneçam ocupados. Por outro lado, ao atribuir a taxa mínima, essa granularidade permite que mais conexões possam ser estabelecidas ao mesmo tempo. No entanto, com o aumento da ocupação da banda do canal, o limite de atendimento pode ser alcançado rapidamente. Essa situação é designada como solução sem janela (SSJ).

2. Agora, considere utilizar escalonamento para reconfiguração de recursos na situação anterior. Ao atingir a escassez de banda, naturalmente algumas requisições seriam bloqueadas. Com o escalonamento, ao invés de bloquear imediatamente por falta

de recursos, as requisições podem ser encaminhadas para uma janela de espera permanente ciente do prazo. Essa janela pode avisar quando uma dada chamada se aproxima do risco de bloqueio, uma vez que é capaz de perceber o esgotamento do prazo de transmissão até o ponto em que não é possível designar a quantidade de banda passante necessária. É importante que alguma política seja implementada para que não se repita dentro da janela, a mesma situação percebida antes do encaminhamento das chamadas a essa janela, quando os recursos se esgotam e as chamadas tendem a ser bloqueadas. Desta maneira, para garantir a satisfação das ressincronizações e melhorar o percentual de conexões bem sucedidas, ao chegar um novo tipo de requisição dessa natureza, a máxima taxa disponível pode ser atribuída, e se não é possível atendê-la nesse momento, a requisição pode ser encaminhada para a janela, onde a partir das reconfigurações, novamente a banda máxima seja atribuída. Já as requisições de *backups* continuam recebendo a taxa mínima e nunca são encaminhadas para a janela. Assim, as ressincronizações possuem mais chances de serem atendidas do que os *backups*. Essa alternativa de solução recebe o nome de solução para o máximo de ressincronizações sem requisições de backups na janela (MRSBJ).

3. Em uma variação do caso anterior, pode-se igualar o número de oportunidades para ambos os tipos de requisições. Assim, as requisições de *backups* continuam recebendo a taxa mínima na primeira tentativa de atendimento, e agora podem ser encaminhadas para a janela, onde na segunda tentativa de atendimento, a mínima banda continue sendo garantida. Como a janela contém agora dois tipos distintos de requisição de conexão, pode-se estabelecer prioridade de atendimento para todas as chamadas do tipo MBDT. Esse proposta é designada como solução para atender o máximo de ressincronizações com janela (MBCJ - Máximo de Batchs com Janela). Note que, MRSBJ e MBCJ pretendem melhorar a taxa de sucesso das ressincronizações, mas as longas execução de *backups* podem refletir em longos períodos de ocupação dos enlaces da rede. Com a elevação no número de chegadas de pedidos, a escassez de banda pode impactar a alocação de alta taxas para satisfazer as ressincronizações.

4. Em uma nova alternativa de solução, propõem-se que ambas as requisições sejam atendidas com a taxa máxima. A ideia é que, como os *backups* necessitam de menos banda, satisfazê-los com a taxa máxima pode contribuir com a rápida desocupação dos recursos. Com recursos disponíveis, as ressincronizações podem ser inicialmente atendidas com a taxa mínima. Assim, a rapidez de atendimento de BDT é muito maior do que de MBDT. Com isso, pode-se dispensar o encaminhamento de re-

quisições de *backups* para a janela e dedicá-la às reconfigurações para atender as ressincronizações. Dentro dessa janela composta apenas de MBDT, as tentativas de atendimento podem ser feitas com a taxa máxima, para que o número de requisições aguardando seja o menor possível. Essa alternativa pode ser chamada de solução da rápida execução de *backups* e diversificadas chances para ressincronizações (RBDCR).

5. Por outro lado, se na proposta RBDCR, onde as ressincronizações são primeiro realizadas com taxa mínima e em seguida, após entrarem na janela, serem executadas com taxa máxima, for sugerido sempre admitir MBDT com o máximo de banda, é possível que, com a rápida satisfação de todas as requisições, um número bem maior de ressincronizações bem sucedidas possa ser obtido do que nas alternativas de solução anteriores. Repare que os recursos são rapidamente desocupados, e embora as MBDTs sejam proporcionalmente maiores e demandem prazos superiores do que as BDTs, as chances de se conseguir alocar a banda máxima são muito maiores. Essa solução é chamada de rápidos *backups* e ressincronizações (RBR).

Com as propostas SSJ, MRSBJ, MBCJ, RBDCR e RBR direcionadas para a máxima satisfação das ressincronizações, é preciso verificar a política mais vantajosa no cenário destacado. Por serem de mesma classe, requisições de aplicações de *backup* e ressincronizações podem ser representadas com semânticas semelhantes.

Sejam:

- r uma requisição única de BDT denotada por (s_r, d_r, C_r, Dl_r), onde s_r e d_r são respectivamente a origem e destino da requisição r, C_r é a quantidade de dados transferidos e Dl_r o intervalo de tempo dentro do qual a transferência da requisição r deve ser concluída.

- R uma requisição de ressincronização em lote de MBDT, encaminhada da aplicação, representada como $(r_1, ..., r_n)$, com $n \geqslant 3$. Faz sentido que cada $C_i, 1 \leqslant i \leqslant n$ representem o mesmo volume de dados, uma vez que as partições de um grupo em comum foram designadas. Existe um conjunto S contendo todas as origens das transmissões e um único destino d. Quanto ao prazo da transmissão, como o lote possui múltiplas chamadas, os prazos podem sofrer alguma variação devido à distância entre os nós, no entanto, por se tratar de uma operação ciente da aplicação, os prazos distintos são bastante próximos na linha do tempo.

- W uma janela de espera permanente ciente do prazo para o escalonamento de requisições que não puderam ser atendidas. A janela pode receber tanto r quanto R.

- $G(V, E)$ grafo direcionado que modela a topologia física, onde V é o conjunto de nós e E, o conjunto de enlaces.

- V compreende dois tipos de nós: i) Nós que são apenas BV-WXCs, denotados por V^W e ii) Nós BV-WXCs que possuem CDs locais, denotados por V^{CD}; sendo que $V^W \cup V^{CD} = V$ e $V^W \cap V^{CD} = \varnothing$.

- E compreende os enlaces, sendo que cada um deles é representado por (u, v) e possui largura de banda equivalente a $B_{(u,v)}$ *slots* de frequência do espectro óptico.

A uma requisição r pode ser atribuída uma taxa mínima $\beta^r_{min} = \dfrac{C_r}{Dl_r}$ ou uma máxima β^r_{max} equivalente à capacidade disponível. Analogamente, para uma requisição R constituída de múltiplas requisições r e cujos *prazos* são ligeiramente distintos, a mínima taxa atribuída é representada por $\beta^R_{min} = \sum_1^n \beta^r_{min}$ e a taxa máxima β^R_{max} é o total de todas as máximas taxas alocadas para atender as requisições $r \in R$.

Se tais requisições se encontram dentro da janela, suas taxas máximas e mínimas precisam ser atualizadas. Tanto para r quanto para R, suas respectivas taxas máximas, β^r_{max} e β^R_{max}, são redefinidas a partir da máxima disponibilidade no momento do seu encaminhamento, passando a ser representadas por $\beta^r_{max_w}$ e $\beta^R_{max_w}$.

A taxa mínima obtida dentro da janela para ambas as requisições, consideramo prazo atualizado como sendo o período entre o tempo atual e o tempo final desse prazo, ou seja, $tempoRestante = prazo - tempoAtual$. Com o prazo de transferência atualizado, nova taxa mínima é obtida $\beta^r_{mim_w} = \frac{Q}{tempoRestante}$. Como $prazo > tempoRestante$, então $\beta^r_{mim} < \beta^r_{mim_w}$. Analogamente, $\beta^R_{mim} < \beta^R_{mim_w}$.

Algumas condições serão assumidas:

- Não será utilizado armazenamento intermediário;

- Transferências ocorrem sem preempção;

- A janela, onde as requisições aguardam pela reconfiguração, não possui restrição de tempo, mas pode comportar até 100 requisições simultâneas.

Os algoritmos propostos a seguir abordam essas margens distintas de atribuição de taxas para os diferentes tipos de requisições.

5.2 Algoritmos Propostos

As soluções SSJ, MRSBJ, MBCJ, RBDCR e RBR destacadas anteriormente são cientes da aplicação, ou seja, lidam com MBDT buscando a melhor combinação possível para

favorecer o atendimento de ressincronizações. A principal diferença entre essas propostas é o uso ou não de janela (W) e a maneira como aprovisionam largura de banda para atender as requisições. A janela W é supervisionada pelo plano de controle da rede, que verifica se o prazo de determinada chamada em espera está se esgotando. As oportunidades de reconfiguração das chamadas na janela são dadas nos eventos de chegadas, partidas e quando o prazo de uma chamada em espera na janela chega ao limite.

As soluções que seguirão são baseadas em algumas rotinas a serem previamente definidas. Quando uma chamada recém chegada solicita atendimento pela primeira vez, as soluções genéricas primárias, Seção 5.2.1 abaixo, são aplicadas, onde o atendimento das requisições r é feito pelo Algoritmo 2 e, o atendimento dos lotes R, pelo Algoritmo 3. Quando uma chamada não foi atendida com solução primária, ela é encaminhada para a janela W de espera, onde são processadas por outros algoritmos específicos, mostrados na Seção 5.2.2, onde o atendimento de r é feito pelo Algoritmo 4, o atendimento de R é feito pelo Algoritmo 5, e quando ambas são tomadas simultaneamente, é utilizado o Algoritmo 6.

Os parâmetros $taxa(r)$ (Equação 5.1) e $taxa(R)$ (Equação 5.2), que aparecem nos algoritmos, representam taxas solicitadas pelos referidos algoritmos, que podem ser as taxas máximas ou mínimas para cada tipo de requisição:

$$taxa(r) = \begin{cases} \beta^r_{mim} \\ \beta^r_{max} \end{cases} \tag{5.1}$$

$$taxa(R) = \begin{cases} \beta^R_{mim} \\ \beta^R_{max} \end{cases} \tag{5.2}$$

5.2.1 Soluções Primárias

Para atender r, o Algoritmo 2 busca os K menores caminhos entre a origem e destino da requisição (linha 2), usando o algoritmo KSP [1]. Em cada caminho candidato k é verificado se a taxa solicitada $taxa(r)$ pode ser alocada (linha 4). A expressão $taxa(r)$ pode ser a taxa máxima β^r_{max} ou a taxa mínima β^r_{mim}, dependendo da solução que utiliza essa rotina. Em seguida, verifica-se se a solução atende às restrições RSA (continuidade e contiguidade do espectro). Em caso positivo a requisição r é aceita (linha 6). Em caso negativo, a requisição é bloqueada.

Algoritmo 2 ServeBulks($r, taxa(r)$)

```
1: para r ← 1 até Σ(r) faça
2:     KSP(r, K)
3:     para k ← 1 até |K| faça
4:         se taxa(r) está disponível então
5:             se Atende restrições RSA então
6:                 Aceita (r, k ∈ K, taxa(r))
7:                 Retorna verdadeiro
8:             senão
9:                 Bloqueia r
10:                Retorna falso
11:            fim se
12:        fim se
13:    fim para
14: fim para
```

O Algoritmo 3 atende a requisição R com a taxa de transmissão desejada $taxa(R)$, que pode ser a taxa máxima β^R_{max} ou a taxa mínima β^R_{mim}. As chamadas de um lote são combinadas b a b (linha 2), onde b é o fator de replicação (igual a 3). Em cada combinação obtida, o Algoritmo pega os elementos dessa combinação, que são requisições r, para as quais encontra os K menores caminhos na linha 5 (KSP [1]), verifica se há $taxa(r)$ disponível (linha 7) e se as restrições RSA são atendidas (linhas 8). Esses procedimentos são realizados em todas as chamadas de uma combinação, e para todas as combinações possíveis, até que se encontre um subconjunto viável, quando finalmente ocorre a aceitação (linha 16).

Algoritmo 3 ServeLotes$(R, taxa(R))$

```
1: para R ← 1 até Σ(R) faça
2:     Combinação(R b)
3:     para subconjunto ← 1 até |Combinação| faça
4:         para  r ← 1 até subconjunto faça
5:             KSP(r, K)
6:             para k ← 1 até |K| faça
7:                 se taxa(r) está disponível então
8:                     se Atende restrições RSA então
9:                         taxa(R) ← taxa(r)
10:                        P ← k ∈ K
11:                    fim se
12:                fim se
13:            fim para
14:        fim para
15:        se |taxa(R)| = 3 ∧ |P| = 3  então
16:            Aceita (R, P, ratio(R))
17:            Retorna verdadeiro
18:        senão
19:            Bloqueia r
20:            Retorna falso
21:        fim se
22:    fim para
23: fim para
```

5.2.2 Soluções na Janela

Se a requisição não pode ser servida na primeira tentativa, ela é encaminhada para a janela. Dentro da janela, o procedimento para atender as chamadas são: o processamento de r é feito pelo Algoritmo 4, o de R é feito pelo Algoritmo 5, e quando ambos são considerados, r e R simultaneamente, executa-se o Algoritmo 6.

O Algoritmo 4 recebe o pedido de uma $taxa(r)$ para uma requisição r. Quando um pedido chega, ele vai para a fila de espera $\mathcal{F}^r$ (linha 1), que é ordenada pelos prazos dessas requisições. A prioridade do atendimento é dada para a requisição com menor prazo de espera. Em seguida, para cada pedido na fila, verifica-se se a taxa solicitada $taxa(r)$ é a taxa mínima β^r_{mim} ou a máxima β^r_{max}. Caso seja a taxa mínima, esse valor é atualizado para $\beta^r_{mim_w}$ correspondendo ao tempo restante da chamada (linhas 4 e 5). Se a taxa

solicitada é a máxima, a capacidade de banda disponível naquele momento é atualizada para $\beta^r_{max_w}$ (linha 7). Na linha 9 o Algoritmo 2 é chamado para resolver os problema da requisição r com sua taxa atualizada. A janela de r também é capaz de avisar se o tempo restante de uma chamada está se aproximando do limite máximo de tempo no qual a chamada pode ser atendida (linhas $11 - 13$).

Algoritmo 4 ServeBulksNaJanela($r, taxa(r)$)

1: $\mathcal{F}^r \leftarrow r$ ▷ Fila ordenada em ordem crescente de *prazo*
2: **para** $r_w \leftarrow 1$ **até** $|\mathcal{F}^r|$ **faça**
3: **se** Taxa solicitada é mínima **então**
4: $tempoRestante = prazo - tempoAtual$
5: $taxa(r)_w = (Q/tempoRestante)$ ▷ Atualiza $taxa(r)$ para $taxa(r)_w$
6: **senão**
7: $taxa(r)_w \leftarrow MaxCapacidade$ disponível
8: **fim se**
9: $ServeBulk(r_w, taxa(r)_w)$
10: **fim para**
11: **se** $\exists r_w$ com *prazo* no limite máximo **então**
12: $ServeBulk(r_w, taxa(r)_w)$
13: **fim se**

O Algoritmo 5 atende R. Quando a janela de lotes recebe R e a respectiva taxa $taxa(R)$, que pode ser a taxa máxima β^R_{max} ou a taxa mínima β^R_{mim}, essa requisição é enfileirada em $\mathcal{F}^R$ (linha 1). Para a ordenação dessa fila, primeiramente verifica-se o menor prazo referente a uma requisição dentro dos lotes. Os menores prazos em cada lote determinam a ordenação desses lotes. A partir daí, para cada lote na fila, verifica-se a taxa solicitada para atualização de acordo com o tempo restante para o atendimento (linhas 3-11). Depois que as taxas são atualizadas, o algoritmo $ServeLote(R_w, taxa(R)_w)$ (Algoritmo 3) é chamado (linha 12). A janela pode avisar o último momento que uma chamada pode esperar para ser atendida (linhas 14-16).

Algoritmo 5 ServeLotesNaJanela($R, taxa(R)$)

1: $\mathcal{F}^R \leftarrow R$ ▷ $\forall r \in R$, verifica-se o mínimo $prazo(r)$
2: **para** $R_w \leftarrow 1$ **até** $|\mathcal{F}^R|$ **faça**
3: **para** $r_w \leftarrow 1$ **até** $|R_w|$ **faça**
4: **se** Taxa solicitada é mínima **então**
5: $tempoRestante = prazo - tempoAtual$
6: $taxa(r)_w = (Q/tempoRestante)$ ▷ Atualiza cada $taxa(r)$ para $taxa(r)_w, \forall r \in R$
7: **senão**
8: $taxa(R)_w \leftarrow MaxCapacidade$ disponível
9: **fim se**
10: $taxa(R)_w \leftarrow Máxima(taxa(r)_w)$
11: **fim para**
12: $ServeLote(R_w, taxa(R)_w)$
13: **fim para**
14: **se** $\exists R_w$ com $prazo$ no limite máximo **então**
15: $ServeLote(R_w, taxa(R)_w)$
16: **fim se**

Há casos em que, ao invés de lidar com um único tipo de chamada, a janela precisará acumular tanto r quanto R, conforme mostra o Algoritmo 6. Se os dois tipos de chamadas r e R são encaminhados para a janela, a fila $\mathcal{F}$ conterá ambos os tipos de chamadas, ordenados pelo prazo, como nos casos anteriores de janela. Se o pedido retirado da fila é um lote R, o procedimento de atualização das suas taxas é seguido nas linhas 6-15. Se o pedido retirado da fila é r, a fila é verificada novamente para detectar se existem lotes esperando (linha 18), visto que a prioridade de atendimento na janela mista são as requisições de ressincronização. Caso não exista, segue o procedimento de atualização da taxa requisitada (linhas 19-27), e na sequência, o Algoritmo $ServeBulk(r_w, taxa(r)_w)$ é chamado para executar o roteamento para os parâmetros atualizados (linha 35).

Algoritmo 6 ServePedidoNaJanela($(r, taxa(r))$,$(R, taxa(R))$)

1: $\mathcal{F} \leftarrow R$
2: $\mathcal{F} \leftarrow r$
3: Ordena $\mathcal{F}$ dando preferência para R
4: **para** $f \leftarrow 1$ **até** $|\mathcal{F}|$ **faça**
5: **se** $f = R$ **então**
6: **para** $r_w \leftarrow 1$ **até** $|R_w|$ **faça**
7: **se** Taxa solicitada é mínima **então**
8: $tempoRestante = prazo - tempoAtual$
9: $taxa(r)_w = (Q/tempoRestante)$ ▷ Atualiza $taxa(r)$ para $taxa(r)_w, \forall r \in R$
10: **senão**
11: $taxa(R)_w \leftarrow MaxCapacidade$ disponível
12: **fim se**
13: $taxa(R)_w \leftarrow taxa(r)_w$
14: $ServeLote(R_w, taxa(R)_w)$
15: **fim para**
16: **senão**
17: **se** $f = r$ **então**
18: **se** $\neg\exists R_w(R_w \in \mathcal{F})$ **então**
19: **para** $r_w \leftarrow 1$ **até** $|\mathcal{F}|$ **faça**
20: **se** Taxa solicitada é mínima **então**
21: $tempoRestante = prazo - tempoAtual$
22: $taxa(r)_w = (Q/tempoRestante)$ ▷ Atualiza $taxa(r)$ para $taxa(r)_w$
23: **senão**
24: $taxa(r)_w \leftarrow MaxCapacidade$ disponível
25: **fim se**
26: $ServeBulk(r_w, taxa(r)_w)$
27: **fim para**
28: **fim se**
29: **fim se**
30: **fim se**
31: **se** $\exists R_w \in \mathcal{F}$ com $prazo$ no limite máximo **então**
32: $ServeLote(R_w, taxa(R)_w)$
33: **senão**
34: **se** $\exists r_w \in \mathcal{F}$ com $prazo$ no limite máximo $\wedge$ $\neg\exists R_w(R_w \in \mathcal{F})$ com $prazo$ no limite máximo **então**
35: $ServeBulk(r_w, taxa(r)_w)$
36: **fim se**
37: **fim se**
38: **fim para**

Os algoritmos da Seção 5.2.1 e desta Seção 5.2.2 serão rotinas utilizadas pelas propostas SSJ, MRSBJ, MBCJ, RBDCR e RBR, mostrados à seguir.

5.2.3 Implementação das Soluções

O cenário proposto é composto por requisições r e R distribuídas uniformemente. São propostos 5 novos algoritmos cientes da aplicação: SSJ, MRSBJ, MBCJ, RBDCR e RBR. Isso motiva a realização de testes com o algoritmo ciente da aplicação proposto no Capítulo 4, naturalmente sem encaminhamento de chamadas para a janela (W). Para esses algoritmos cientes são válidas as seguintes premissas:

- Lidam com requisições em lote (R), realizando combinações de requisições individuais desses lotes, podendo descartar chamadas redundantes.
- A ocorrência de uma ressincronização em tais algoritmos está condicionada ao atendimento de exatamente três chamadas $r \in R$;

O algoritmo RSAC [96] utilizado para comparação é o algoritmo de roteamento convencional também empregado no Capítulo 4. Assim, ao todo, 7 soluções, relacionadas à seguir, serão comparadas. Os algoritmos genéricos mostrados anteriormente serão empregados nas rotinas das seguintes soluções:

1. AA-RSA Sem Janela (AARSASJ): O algortimo ciente da aplicação proposto no Capítulo 4 foi adaptado para atender chamadas individuais de transferências de dados. Para lidar com os dois tipos de chamadas r e R, sua estratégia é:
   - Atende r com a taxa mínima β^r_{min};
   - Atende os lotes R com a taxa mínima β^R_{min} se o caminho selecionado é menor que a metade do diâmetro da rede, caso contrário, atribui a taxa máxima β^R_{max};
   - Não utiliza janela de escalonamento;

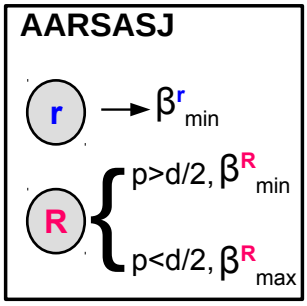

Figura 5.1: Esquema da solução AARSASJ.

2. RSA Convencional (RSAC) [96]: É o único algoritmo convencional a ser testado. Por não ser ciente da aplicação, todas as chamadas são tratadas da mesma maneira, e assim, quando um lote é recebido, todas as chamadas que o compõe são extraídas, e de maneira individual e sequencial são submetidas ao roteamento.

   - Atende requisições r com a taxa mínima β^r_{min} (Figura 5.2);
   - Em cada lote R, cada uma das chamadas r são atendidas com a taxa mínima β^r_{min};
   - Não tem ciência de R, e portanto, atende as requisições $r \in R$ de maneira sequencial. A probabilidade de que uma ressincronização ocorra é reduzida, uma vez que o atendimento de requisições redundantes acaba consumindo recurso desnecessariamente, tornando-o escasso para requisições seguintes;
   - Não utiliza escalonamento.

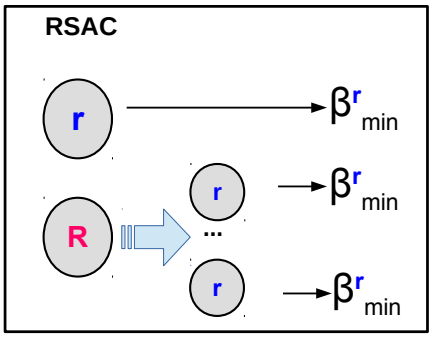

Figura 5.2: Esquema da solução RSAC.

3. Solução Sem Janela (SSJ): procedimentos realizados pelo Algoritmo 7 e mostrados no esquema da Figura 5.3 .

   - Atende os *bulks* r com a taxa mínima β^r_{min}, conforme mostrado na linha 2 do Algoritmo 7;
   - Atende lotes R com a taxa mínima β^R_{min}, mostrado na linha 7;
   - Não utiliza janela de escalonamento (Figura 5.3);

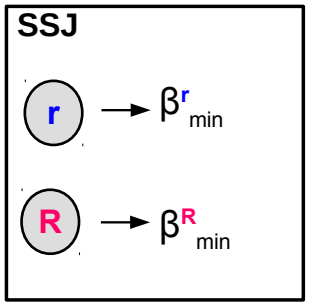

Figura 5.3: Esquema da solução SSJ.

Algoritmo 7 SSJ(G, r, R)

1: **para** r $\leftarrow$ 1 **até** $\Sigma(r)$ **faça**
2: **se** $ServeBulk(r, \beta^r_{min}) = verdadeiro$ **então**
3: **Retorne** *verdadeiro*
4: **fim se**
5: **fim para**
6: **para** R $\leftarrow$ 1 **até** $\Sigma(R)$ **faça**
7: **se** $ServeLote(r, \beta^R_{min}) = verdadeiro$ **então**
8: **Retorne** *verdadeiro*
9: **fim se**
10: **fim para**

4. Máximo de Ressincronizações sem requisições de *backups* na Janela (MRSBJ): Conforme mostrado na Algoritmo 8 e no esquema da Figura 5.4 .

   - Utiliza janela de escalonamento;
   - Se houver recurso, requisições r são atendidas com a taxa mínima β^r_{min}. Caso contrário, são bloqueados. A linha 2 do Algoritmo 8 chama o procedimento que verifica essa alocação;
   - Se houver recurso, atende lotes R com a taxa máxima β^R_{max} (linha 7-8). Caso contrário, escalona R, e em uma nova tentativa de reconfiguração dos recursos, se houver disponibilidade, atende a chamada com taxa máxima β^R_{max} (linhas 7-8);

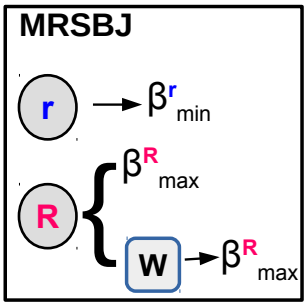

Figura 5.4: Esquema da solução MRSBJ.

Algoritmo 8 MRSBJ(G, r, R)

1: **para** r $\leftarrow$ 1 **até** $\Sigma(r)$ **faça**
2: **se** $ServeBulk(r, \beta^r_{min}) = verdadeiro$ **então**
3: **Retorne** $verdadeiro$
4: **fim se**
5: **fim para**
6: **para** R $\leftarrow$ 1 **até** $\Sigma(R)$ **faça**
7: **se** $ServeLote(r, \beta^R_{max}) = verdadeiro$ **então**
8: **Retorne** $verdadeiro$
9: **senão**
10: **se** $ServeLotesNaJanela(r, \beta^R_{max}) = verdadeiro$ **então**
11: **Retorne** $verdadeiro$
12: **fim se**
13: **fim se**
14: **fim para**

5. Máximo de *Batchs* Com Janela (MBCJ): Algoritmo 9 e respectivo esquema da Figura 5.5.

   - Utiliza escalonamento de ambas as chamadas, mas ao escolher uma chamada da janela para nova tentativa de roteamento, os lotes sempre têm prioridade. Esse procedimento é feito com o Algoritmo 6;
   - Se houver recurso, requisições r são atendidas com a taxa mínima β^r_{min} (linha 2). Caso contrário, cada r é escalonada, e em uma nova tentativa de roteamento, se houver disponibilidade, r recebe β^r_{min} (linha 5);
   - Se houver recurso, atende lotes R com a taxa máxima β^R_{max} (linha 9). Caso contrário, escalona R, e em uma nova tentativa de reconfiguração dos recursos,

se houver disponibilidade, atende a chamada com taxa máxima β^R_{max} (linha 12);

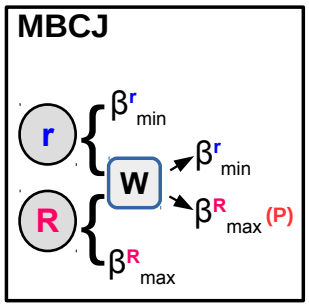

Figura 5.5: Esquema da solução MBCJ.

Algoritmo 9 MBCJ(G, r, R)

1: **para** r $\leftarrow$ 1 **até** $\Sigma(r)$ **faça**
2: **se** $ServeBulk(r, \beta^r_{mim}) = verdadeiro$ **então**
3: **Retorne** $verdadeiro$
4: **senão**
5: $ServePedidoNaJanela \leftarrow (r, \beta^r_{mim})$
6: **fim se**
7: **fim para**
8: **para** R $\leftarrow$ 1 **até** $\Sigma(R)$ **faça**
9: **se** $ServeLote(R, \beta^R_{max}) = verdadeiro$ **então**
10: **Retorne** $verdadeiro$
11: **senão**
12: $ServePedidoNaJanela \leftarrow (R, \beta^R_{max})$
13: **fim se**
14: **fim para**

6. Rápida execução de *Backups* e Dupla Chance para Ressincronizações (RBDCR): Algoritmo 10 e esquema da Figura 5.6.

   - Utiliza janela de escalonamento;
   - Se houver recurso, requisições r são atendidas com a taxa máxima β^r_{max} (linha 2 do Algoritmo 10). Caso contrário, são bloqueadas;

- Se houver recurso, atende lotes R com a taxa mínima β^R_{min} (linha 7 do Algoritmo 10). Caso contrário, escalona R, e em uma nova tentativa de reconfiguração dos recursos, se houver disponibilidade, atende a chamada com taxa máxima β^R_{max} (linha 10 do Algoritmo 10);

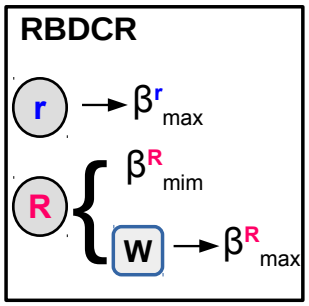

Figura 5.6: Esquema da solução RBDCR.

Algoritmo 10 RBDCR(G, r, R)

```
1: para r ← 1 até Σ(r) faça
2:     se ServeBulk(r, β^r_max) = verdadeiro então
3:         Retorne verdadeiro
4:     fim se
5: fim para
6: para R ← 1 até Σ(R) faça
7:     se ServeLote(r, β^R_mim) = verdadeiro então
8:         Retorne verdadeiro
9:     senão
10:        se ServeLotesNaJanela(r, β^R_max) = verdadeiro então
11:            Retorne verdadeiro
12:        fim se
13:    fim se
14: fim para
```

7. Rápidos *Backups* e Ressincronizações (RBR): Algoritmo 11 e esquema da Figura 5.7.

   - Utiliza escalonamento de lotes (Algoritmo 5);
   - Se houver recurso, requisições r são atendidas com a taxa máxima β^r_{max} (linha 2 do Algoritmo 11). Caso contrário, são bloqueadas;

- Em todas as tentativas de estabelecimento de conexão para lotes R, seja antes ou depois do escalonamento, sempre a taxa máxima β_{max}^{R} é alocada.

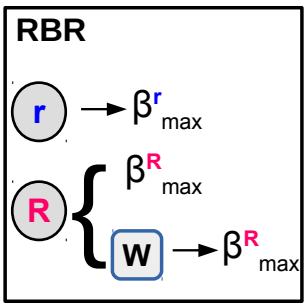

Figura 5.7: Esquema da solução RBR.

Algoritmo 11 RBR(G, r, R)

```
1: para r ← 1 até Σ(r) faça
2:     se ServeBulk(r, β^r_max) = verdadeiro então
3:         Retorne verdadeiro
4:     fim se
5: fim para
6: para R ← 1 até Σ(R) faça
7:     se ServeLote(r, β^R_max) = verdadeiro então
8:         Retorne verdadeiro
9:     senão
10:        se ServeLotesNaJanela(r, β^R_max) = verdadeiro então
11:            Retorne verdadeiro
12:        fim se
13:    fim se
14: fim para
```

Complexidade dos Algoritmos

Todos os algoritmos apresentados utilizam a função $KSP\,(i, K)$ [1], com complexidade de $O(KV^3)$, para a busca de caminhos na rede. A política de atribuição de espectro para todos os algoritmos é a *First-Fit*, de complexidade linear [23], entretanto a verificação de espectro disponível em todos os enlaces nos caminhos disponíveis leva $O(K^2)$.

Alguns algoritmos fazem combinação de chamadas de um lote. Para obter essas combinações, a função $Combinação\binom{R}{b}$ é chamada, com um lote R de tamanho n e um fator de replicação b de tamanho 3. A complexidade de tempo dessa função é $O\left(\frac{n!}{b!(n-b)!}\right)$, que é exponencial [23].

As duas rotinas definidas como Soluções Primárias (Subseção 5.2.1) são o Algoritmo 2 para atender r e o Algoritmo 3 para atender R. Nesse primeiro, toda requisição r chama a função $KSP\,(i, K)$ e verifica a disponibilidade de espectro nos enlaces. Assim, sua complexidade é de:

$$O(K^3V^3) \tag{5.3}$$

Já o Algoritmo 3 faz esse mesmo processo, mas antes chama a função *Combinação*$\binom{R}{b}$. Logo sua complexidade é de:

$$O\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) \tag{5.4}$$

As rotinas definidas como Soluções na Janela (Subseção 5.2.2) são o Algoritmo 4, que atende requisições r dentro da janela, o Algoritmo 5, que atende requisições R, e o Algoritmo 6, que atende r e R também dentro da janela. Todos esses algoritmos implementam uma fila de chamadas e realizam operações para atualizar o prazo e a taxa de transmissão, que são operações constantes. Após esse processo, esses algoritmos das janelas podem chamar os algoritmos primários como rotina, passando-lhes os dados atualizados.

Desta forma, o Algoritmo 4 faz atualizações de dados para todas as chamadas na fila $\mathcal{F}^r$, que no pior caso, pode ser formada por todas as chamadas r, e chama o Algoritmo 2, alcançando uma complexidade de tempo de:

$$O(K^3V^3) \tag{5.5}$$

O Algoritmo 5 também possui um fila $\mathcal{F}^R$, e para essas chamadas são atualizados os dados de prazo e taxa, que são repassados ao Algoritmo 3. Por tanto, sua complexidade é de:

$$O\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) \tag{5.6}$$

Por sua vez, o Algoritmo 6, que atende r e R, possui uma fila única $\mathcal{F}^r$ para ambas as chamadas. Essa fila é ordenada dando-se prioridade para R, por meio de uma busca binária que toma $O(\log n)$ [23]. Assim, a complexidade desse algoritmo é:

$$O\left(\log n\left[\left(K^3V^3\right) + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)\right]\right) \tag{5.7}$$

O algoritmo AARSASJ foi adaptado do algoritmo ciente no Capítulo 4, que atendia somente lotes. A diferença é que agora ele atende requisições r e R. Sua complexidade

total é de:

$$O\left(\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) + K^3V^3\right) \tag{5.8}$$

O algoritmo RSAC [96] é o algoritmo convencional do Capítulo 4, e tem complexidade de:

$$O(K^3V^3) \tag{5.9}$$

O algoritmo SSJ é ciente da aplicação. Ele atende r e R e não faz escalonamento de chamadas. Sua complexidade é de:

$$O\left(\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) + K^3V^3\right) \tag{5.10}$$

O algoritmo MRSBJ, RBDCR e RBR atendem r e R e podem escalonar R. Esses algoritmos são diferentes na maneira em que aprovisionam taxas para atender as chamadas, assim a complexidade de cada um é:

$$O\left(K^3V^3 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right) \tag{5.11}$$

O algoritmo MBCJ atende r e R e escalona ambas, dando prioridade para R, por isso, sua complexidade é:

$$O\left(\log n\left[(K^3V^3)^2 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right]\right) \tag{5.12}$$

Embora as complexidades de tempo dos algoritmos cientes da aplicação sejam exponenciais em n, o tamanho das entradas é pequeno e constante em cenários reais. A Tabela 5.1 relaciona todos os algoritmos tratados e suas complexidades de tempo.

5.3 Avaliação de Desempenho

5.3.1 Parâmetros da simulação

Os parâmetros da simulação, mostrados na Tabela 5.2, foram empregados nas três topologias de redes adotadas, NSFNET, Pan-Euro e USA (Relacionadas na Tabela 5.3), onde foram localizados conjuntos específicos de CDs. Os dois tipos de chamadas orientadas à dados, r e R, respectivamente pedidos de *backup* e ressincronização, são quantificados de maneira específica. Existem dois tipos de requisições r, sendo uma para transferência de $100GB$ e outra para $300GB$, cujos períodos de tempo definido para a entrega desses dados é de 20 unidades de tempo. Além disso, dois tipos de lotes foram definidos, ambos

Algoritmos	Complexidade de Tempo
AARSASJ	$O\left(\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) + K^3V^3\right)$
RSAC	$O(K^3V^3)$
SSJ	$O\left(\left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right) + K^3V^3\right)$
MRSBJ	$O\left(K^3V^3 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right)$
MBCJ	$O\left(\log n \left[(K^3V^3)^2 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right]\right)$
RBDCR	$O\left(K^3V^3 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right)$
RBR	$O\left(K^3V^3 + \left(\left(\frac{n!}{b!(n-b)!}\right) * V^3\right)^2\right)$

Tabela 5.1: Tabela relacionando os algoritmos e suas complexidades de tempo

formados por 4 requisições e um tempo total de transferência de dados de 100 unidades de tempo para um total de $100GB$ e $500GB$ respectivamente, o que significa dizer que em um lote, cada chamada individual transporta essa quantidade previamente definida, em média.

Cada simulação foi realizada 5 vezes. Foram realizadas 100.000 chamadas com origens e destinos distribuídos uniformemente dentro do conjunto de CD dispostos pelas redes. Para os resultados apresentados foram calculados intervalos de confiança com 95% de confiabilidade. O número de chegadas de chamadas varia de 2 a 30 por unidade de tempo com incrementos de 4 chegadas [107]. Como todas as chamadas seguem uma distribuição uniforme, o número de chegadas diz respeito a chegadas simultâneas de r e R.

Nos enlaces da rede, com espectro total de $1.5THz$, foram configurados com 120 *slots*, cada um deles com largura de banda de $12.5GHz$ [96]. As bandas de guardas de separação das conexões são formadas por 2 *slots*, cuja largura de banda total é de $25GHz$. Em cada nó da rede são utilizados 15 transmissores. Cada um dos equipamentos transmissores é capaz de estabelecer um caminho óptico de até 8 *slots*. Esses transmissores são ajustados para utilizar a modulação QPSK, na qual cada dois *bits* de dados são agrupados em um símbolo de transmissão, e assim, alcançam maior vasão e melhor eficiência espectral, com as configurações estabelecidas.

5.3.2 Característica das Topologias de Rede

As topologias de rede utilizadas nas simulações foram a NSFNET(Figura 5.8), Pan-European(Figura 5.9) e USA (Figura 5.10). Em cada uma delas foram localizados e demarcados nós específicos para representar o centro de dados, V^{CD}. Assim, a ressincronização e o *backup* só ocorrem entre esses nós específicos. Os demais nós representam

Parâmetro	**Valor**
Topologia	NSFNET, Pan-Euro e USA
Nós origem/destino	(0,7,11,12,13), (0,1, 8,20,23) e (1,10,12,20,21,22)
Tipos de chamadas	2 bulks, 2 lotes
Transponders	15
Slots por transponders	8
Largura de banda do slot	12.5GHz
Slots por enlace	120
Modulação	QPSK
Número de simulações	5
Número de chamadas	100.000
Taxa de chegadas	[2,6,10,14,18, 22, 26 e 30]
Deadline das chamadas	20u.t. para *bulks* e 100u.t. para lotes
Quantidade de dados/*bulks*	[100,300]GB
Quantidade de dados/lotes	[100,500] GB
Tamanho do lote	4

Tabela 5.2: Parâmetros da simulação

BV-WXCs, V^W. Os enlaces estão numerados de acordo com suas distâncias. O modelo de comunicação adotado considera enlaces bidirecionais, assim esses enlaces são representadas por grafos não direcionados.

O grau de um nó v, denotado por $g(v)$, é o número de vizinhos do nó v, ou seja, o número de enlaces ligados a v. Um grafo é dito conectado se para cada par de nós existe um caminho interligando-os. Caso contrário é desconectado [23].

Em todas as topologias são analisadas o seu grau mínimo de nó e a sua conectividade. Ambos os atributos de topologia dependem da distribuição espacial dos nós e sua faixa de transmissão, conforme a Tabela 5.3.

Rede	Nós	Conectividade		Enlaces	Tamanho
	#	Média entre os nós	Média entre CDs	#	Média
NSFNET	14	3	3	20	1936,3
Pan-Euro	28	2	2,4	41	626,9
USA	24	3	3,5	43	996,5

Tabela 5.3: Principais características das redes utilizadas.

O alcance médio da modulação QPSK, definida para os algoritmos RSAs, é de cerca de 4000 KM, embora esse alcance dependa de muitos fatores como o tipo da fibra e a correção de erros de encaminhamento (*Forward Error Correction*-FEC), entre outros. O tamanho médio dos enlaces nas redes, mostrados na Tabela 5.3, demonstra a compatibilidade com a modulação estabelecida. Nas redes bem conectadas a taxa de aceitação das conexões

é maior do que em uma rede menos conectada, devido a disponibilidade de recursos, qualidade também percebida nas redes mais densas, ou seja, com mais enlaces. Outra característica que chama atenção é o grau de conectividade dos CDs, que geralmente é maior que a média dos demais nós, e permite que enquanto um determinado CD participa de um grupo de replicação recebendo as atualizações para uma partição específica, ao mesmo tempo possa fazer parte de um segundo grupo, operando sobre outra partição.

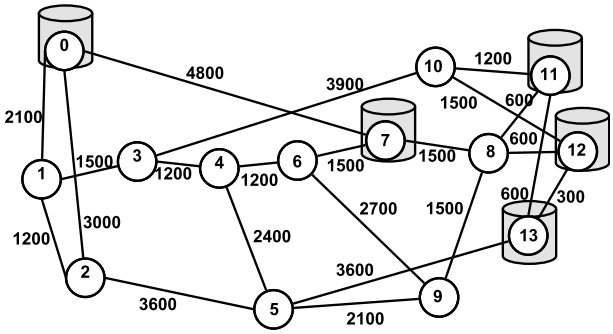

Figura 5.8: Rede NSFNET.

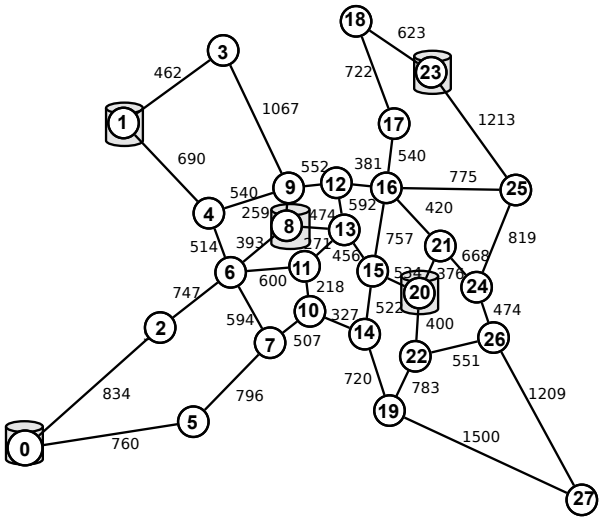

Figura 5.9: Rede PanEuro.

5.3.3 Taxa de Sucesso da Ressincronização

A taxa de sucesso da ressincronização (SRR) é calculada como o número de chamadas em lotes que foram aceitas, dividido pelo número total de chamadas em lote, como mostra a Equação 5.13:

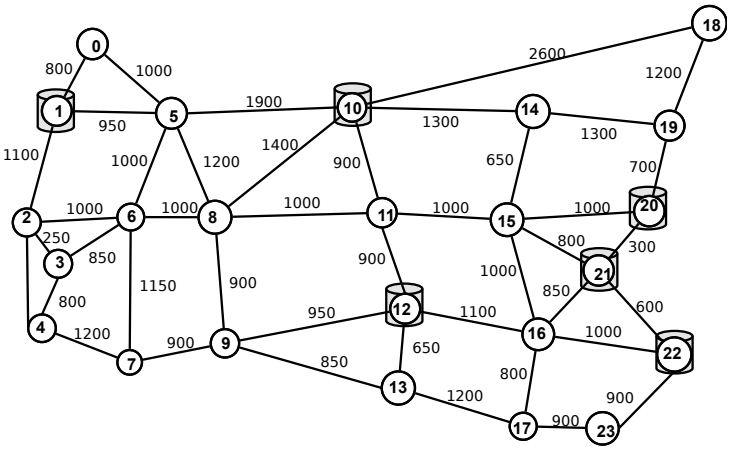

Figura 5.10: Rede USA.

$$SRR = \frac{\sum AcceptedBatchs}{\sum AllTheBatchCalls} \tag{5.13}$$

Um algoritmo de roteamento convencional não trata de lotes, e assim, precisa atender 3 chamadas para que ocorra uma ressincronização. Entretanto, a probabilidade de que cheguem, simultaneamente ou sequencialmente, as três requisições de ressincronização de um mesmo grupo de réplicas, é pequena.

Os algoritmos cientes da aplicação lidam com requisições em lote fazendo combinações em grupos de três entre todas as suas chamadas e selecionando um desses grupos. Isso significa que, se um lote possui 4 solicitações de transferência para um determinado CD, algum desses pedidos será descartado. Os serviços de replicação atrelados à aplicação distribuída observam os estados das réplicas e podem garantir a exclusão segura de uma das solicitações sem prejuízo para a ressincronização, assegurando proteção contra faltas bizantinas. Notoriamente, algoritmos convencionais consomem muito mais banda do que algoritmos cientes da aplicação.

A Figura 5.11 mostra o impacto dos resultados de algoritmos cientes e não cientes na rede NSFNET. O algoritmo RSAC [96] possui um desempenho inferior aos demais porque atende chamadas que não são necessariamente destinadas a atualizar a mesma partição de dados, diferentemente dos algoritmos cientes. Sua estratégia para lidar com requisições em lote é executá-las individualmente. Para atender qualquer tipo de chamada, a taxa mínima é atribuída, tornando os canais ocupados por mais tempo. Com o aumento no número de chegadas, as tentativas de atender todas as solicitações de um lote falham pela falta de banda disponível, e o esgotamento do prazo também pesa sobre esse resultado. Sua taxa de sucesso começa em 13.2% e cai para 0.02%.

O algoritmo RBDCR (Algoritmo 10) primeiramente tenta atender os lotes com a taxa

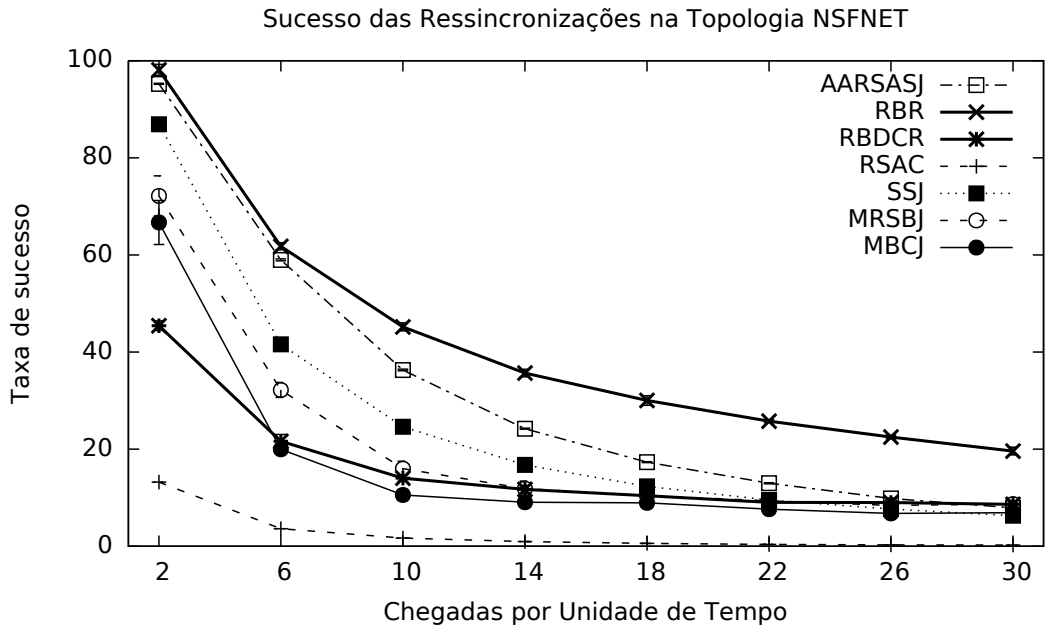

Figura 5.11: Taxa de Sucesso das Ressincronizações na rede NSFNET com modulação QPSK e prazo de $100u.t.$ para as requisições em lote.

mínima, e vai alocando a máxima para as chamadas individuais. Durante o decorrer do prazo, novos lotes vão sendo impedidos de serem atendidos, quando são então escalonados e podem receber a taxa máxima. Como as requisições r rapidamente desocupam a rede, seus recursos foram disponibilizados para as transferências sem recorrer à janela, entretanto o ponto de saturação da rede é atingido. Com os canais permanecendo ocupados por mais tempo, não é possível alocar a banda máxima para as chamadas em espera na janela, e sua taxa de sucesso cai de 46% para 8%.

Ainda no gráfico da Figura 5.11, o MBCJ (Algoritmo 9) começa atendendo lotes já com a taxa máxima, o que resultou em uma taxa de sucesso acima de 66% com carga baixa. A janela de escalonamento desse algoritmo recebe r e R, mas a prioridade de atendimento de chamadas é de R. À medida que os lotes rapidamente desocupam a banda utilizada, muitas conexões individuais são estabelecidas e permanecem ativas por longos períodos, levando os recursos de rede à exaustão. Esse mesmo impacto é sofrido pelo algoritmo MRSBJ (Algoritmo 8), que também consome mais os seus recursos atendendo chamadas individuais, e cai de 75% para 8.5%.

A respeito do RBDCR (Algoritmo 10), MBCJ (Algoritmo 9) e MRSBJ (Algoritmo 8), no gráfico da Figura 5.11, é possível verificar que, atender R com taxa máxima enquanto r é atendida com a taxa mínima resulta em muito recurso liberado de uma única vez, que são rapidamente empregados para uma elevada quantidade de r por longos períodos.

A janela de escalonamento não consegue combater esse efeito porque quando chamadas começam a ser encaminhadas para ela, a banda já está saturada.

O algoritmo AARSASJ (Capítulo 4) também atende r com a taxa mínima, porém, a variabilidade de taxas aprovisionadas para R (máximas e mínimas) alterna os períodos de disponibilidade de banda, garantindo assim o seu resultado entre 95% e 8% com o aumento no número de chamadas. O melhor desempenho destacado é obtido com o algoritmo RBR (Algoritmo 11), que atende ambas as requisições com taxa máxima e mantém a rotatividade de banda em alta, obtendo 98% a 19% de sucesso nas ressincronizações. Por ter o maior percentual de atendimento, o RBR também descarta mais pedidos dos lotes entre todos os algoritmos cientes. Quando o número de chegadas de requisições é igual a 2, quase 25% de pedidos são eliminados e, ao final, com carga alta, o descarte é de cerca de 5% na rede NSFNET (Figura 5.11).

Na rede Pan-Euro (Figura 5.12), menos conectada, os algoritmos que alocam a taxa máxima foram favorecidos. Logo nos primeiros número de chegadas, o AARSASJ (Capítulo 4) alcança taxa de sucesso de mais de 94% enquanto que o RBR (Algoritmo 11) obtém mais de 95%. O AARSASJ (Capítulo 4) atende 68% de lotes com a taxa máxima, inclusive os pedidos que são encaminhados para a janela, porque os caminhos definidos no roteamento são maiores que a metade do diâmetro da rede, o que é justificável pela distribuição dos CDs ao longo da rede. No entanto, a proporção de lotes atendidos com a taxa mínima vai crescendo de 32 a 41%, resultando em queda acentuada de ressincronizações bem sucedidas. Como o RBR (Algoritmo 11) atende todas as chamadas com a taxa máxima, mesmo com carga alta ainda atende 19% do total. Na rede Pan-Euro, o algoritmo RBR descarta o máximo de 24% de pedidos dos lotes.

Logo em seguida, entres os algoritmos SSJ (Algoritmo 7), MRSBJ (Algoritmo 8) e MBCJ (Algoritmo 9), inicialmente com 87%, 67% e 60% de ressincronizações bem sucedidas, respectivamente, a principal diferença é que enquanto o primeiro atende os lotes com taxa mínima, os outros dois faz esse atendimento com taxa máxima e ainda pode encaminhá-los para a janela. O problema é que tanto recurso liberado mais cedo, acaba sendo revertido para as chamadas individuais, que ao chegarem na rede ocupam-na por longos períodos. O MBCJ ainda tem o desempenho agravado por escalonar também as requisições r, no gráfico da Figura 5.12. Os três algoritmos acabam convergindo quando a rede alcança um alto nível de saturação. Já o RBDCR (Algoritmo 10) é menos ambicioso e primeiramente tenta atender os lotes com a taxa mínima. Se não consegue, só após o escalonamento a taxa máxima é atribuída. Devido a isso, quando a carga é baixa, sua taxa de sucesso é de 51%, mas passa a 27% e 19% para 6 e 10 chegadas, conseguindo se manter acima dos índices dos três algoritmos anteriores. Ainda na rede Pan-Euro, o RSAC [96] começa com taxa de 14% e esse índice só piora ficando abaixo de 1%.

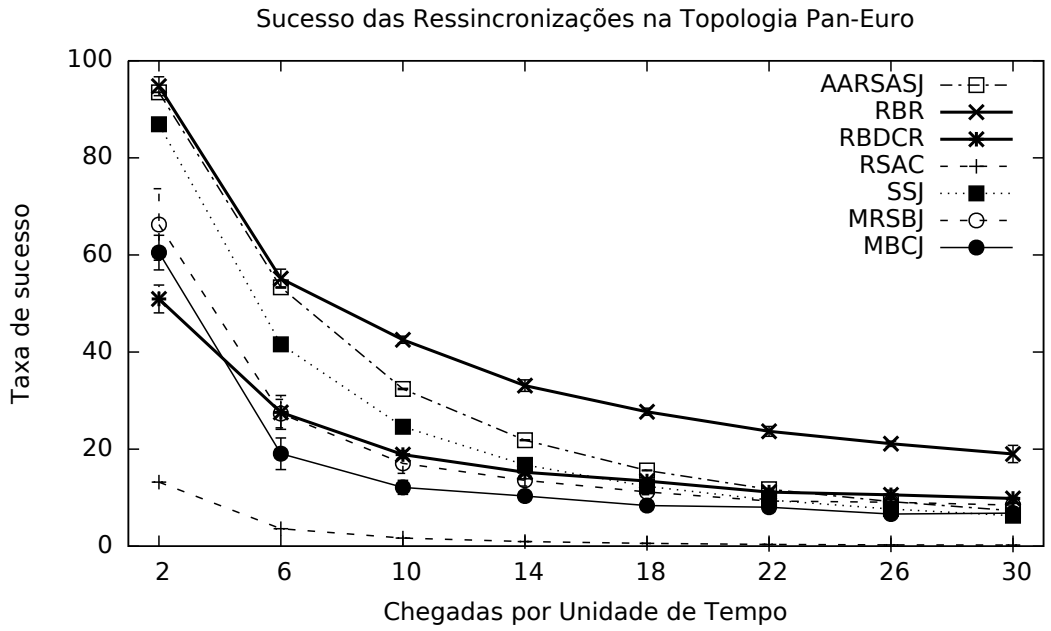

Figura 5.12: Taxa de Sucesso das Ressincronizações na rede Pan-Euro com modulação QPSK e prazo de $100u.t.$ para as requisições em lote.

Os resultados na rede USA (Figura 5.13) sugerem que em uma rede bem conectada, priorizar o atendimento de requisições de ressincronização, e também sempre atendê-las com a taxa máxima, resulta em um bom desempenho de quase 98% com poucas chegadas de requisições. Por esse motivo, os algoritmos AARSASJ (Capítulo 4) (98.4%), RBR (97.4%) e MBCJ (97.2%) alcançam um bom resultado no início da ocupação da rede. Ambos RBR (Algoritmo 11) e MBCJ (Algoritmo 9) escalonam lotes e os atendem com a taxa máxima. O AARSASJ (Capítulo 4) atribui taxa máxima para 52% dos lotes, mas com a crescente indisponibilidade nos menores caminhos, a partir de 10 chegadas de requisições, mais da metade de todas as chamadas são atendidas com a taxa mínima. Nos lotes com mais de três requisições, ao menos uma delas é descartada e exatamente três requisições são atendidas. O descarte de chamadas pelo algoritmo RBR (Algoritmo 11) é de no máximo 24.6%, enquanto que o MBCJ (Algoritmo 9) elimina 24.5% e o AARSASJ (Capítulo 4) começa eliminando 24.7%. No decorrer do aumento no número de chegadas, o MBCJ (Algoritmo 9) alcança os maiores índices de descartes, o que significa que sua eficiência impede o desperdício de quase $\frac{1}{4}$ da banda solicitada.

O SSJ (Algoritmo 7) e o MRSBJ (Algoritmo 8) são penalizados com quedas acentuadas no atendimento. O mais prejudicial entre eles é a liberação de muita banda com rapidez, que são então aprovisionadas para as chamadas com taxa mínima. O MRSBJ (Algoritmo 8) escalona lotes e despacha todos eles com a taxa máxima, daí, a demora em reaver esse

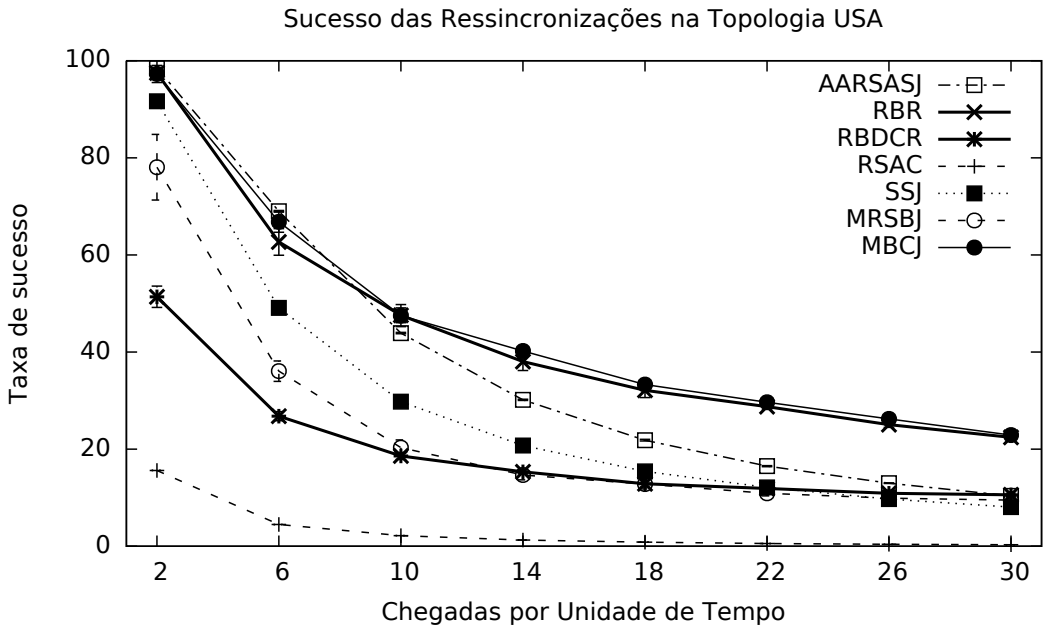

Figura 5.13: Taxa de Sucesso das Ressincronizações na rede USA com QPSK e prazo de $100u.t.$ para todas as requisições em lote.

recurso para as ressincronizações o leva ao declínio. O RBDCR (Algoritmo 10) não se mostrou efetivo no escalonamento de lotes (Figura 5.13). Quando uma chamada de dentro da janela requisita a taxa máxima, as chamadas que foram antedidas antes do escalonamento ainda estão ativas na rede. Sem banda para atender a janela, as requisições são logo bloqueadas. Já o RSAC [96] continua com o desempenho ruim, independentemente da rede onde ele é executado.

Na rede USA, o algoritmo MBCJ (Algoritmo 9) obteve uma elevada taxa de aceitação dos pedidos de ressincronização, mas a taxa de bloqueio de *bulks* esteve entre as mais altas registradas, como será visto na Seção 5.3.4.

Esses resultados refletem que as redes mais conectadas não saturam tão rápido quanto redes menos conectadas. Além disso, os algoritmos cientes da aplicação obtém mais sucesso com as ressincronizações do que algoritmos convencionais. Quanto ao descarte de pedidos de um lote de requisições, é fácil ver que seu percentual é diretamente proporcional à taxa de aceitação. O percentual de banda que deixou de ser alocado para essas requisições que foram descartadas é equivalente a 25% do total de banda solicitada, para lotes formados por 4 pedidos.

O escalonamento de requisições favorece as ressincronizações desde que a segunda aplicação que também solicita recurso também seja rapidamente transmitida. Nos casos em que uma grande quantidade de banda foi liberada rapidamente após o atendimento dos

lotes, e esta foi revertida para as requisições de *backups*, uma reserva antecipada de recursos garantiria o sucesso das ressincronizações. As soluções que promoveram rotatividade de banda empregando a taxa máxima conseguiram resultados superiores.

5.3.4 Taxa de Bloqueio (BR) das Requisições de *Backup*

A taxa de bloqueio (BR) ou probabilidade de bloqueio (BP) de r resulta do número de chamadas r bloqueadas dividido pelo total de chamadas ($\sum r$), conforme mostra a Equação 5.14. Essa medida é tomada considerando-se apenas as requisições de *backup*.

$$BR = \frac{NumberOfBlockedCalls}{TotalNumberofCalls} \tag{5.14}$$

O objetivo das soluções propostas é manter o bom desempenho da aplicação de MBDT em um cenário com mais de um tipo de tráfego. No entanto, espera-se que com o aumento na taxa de aceitação de ressincronizações, a segunda aplicação não seja penalizada com elevado bloqueio.

O gráfico apresentado na Figura 5.14 compara a probabilidade de bloqueio das chamadas do tipo *bulk* (*requisições de backup*), quando o número de chegadas aumenta de 2 para 30 chegadas por unidade de tempo na rede NSFNET. As requisições de *backups* (r) são chamadas que requisitam largura de banda suficiente para transmissão dentro do prazo, e por se tratarem de chamadas unitárias, sua demanda por largura de banda é menor. Observa-se que o maior percentual de bloqueio é obtido com a solução de roteamento convencional, o algoritmo RSAC [96], iniciando em 7.5% e alcançando 16% com máximo número de chegadas, que por atender todas as requisições de um lote, ocupa mais banda e ocasiona o bloqueio de chamadas r. Além do mais, a banda ocupada permanece indisponível por longos períodos devido a atribuição da taxa mínima para todas as chamadas. Em seguida, o segundo pior resultado foi obtido com o algoritmo MRSBJ (Algoritmo 8),com taxas entre 3.5% e 14%, que sacrifica as requisições de *backups*, uma vez que a janela empregada recebe apenas as requisições de ressincronizações. Isso é agravado com o fato de tais requisições em lote receberem a taxa máxima em toda oportunidade de reconfiguração.

O algoritmo RBDCR (Algoritmo 10) também não encaminha requisições r para a janela, mas tenta atendê-las com a taxa máxima, enquanto que os lotes recebem a máxima taxa, caso estejam na janela, e a mínima, caso contrário. Essa dupla possibilidade para as requisições de ressincronização contribuem para a oscilação da banda disponível com o aumento no número de chegadas. Por isso, percebe-se que o bloqueio de requisições r experimenta ligeiras subidas e decidas com as mudanças contantes no estado da rede. Entre 6 e 10 e entre 22 e 26 chegadas de requisições, as taxas de bloqueio são levemente

próximas, 7.1% e 7.8% no primeiro caso, e 12.2% e 12.7% no segundo caso. A taxa máxima de bloqueio desse algoritmo é 13.6%.

Ainda no gráfico da Figura 5.14, o algoritmo SSJ (Algoritmo 7), um dos mais simples, não faz escalonamento e sempre atende as chamadas com taxa mínima. No início das execuções, quando os recursos estão desocupados, o SSJ bloqueia pouquíssimas requisições de *backup* (0.7%), e sua mínima granularidade contribui para esse resultado. Entretanto, nos longos prazos de espera, outras chegadas vão sendo registradas e são diretamente bloqueadas por não haver o mecanismo da janela, alcançando quase 14% de bloqueio total. O algoritmo MBCJ (Algoritmo 9), por outro lado utiliza janela para ambos os tipos de chamadas, e a segunda oportunidade de atendimento de requisições r, contribui para uma taxa de bloqueio não tão elevada (variando entre 1.4% e 11.5%). Claramente percebe-se que o escalonamento de requisições de *backup* contribui para menores índices de chamadas bloqueadas.

Os algoritmos AARSASJ (Capítulo 4) e RBR (Algoritmo 11) não utilizam janela para encaminhamento de r e aprovisionam a mínima e a máxima banda disponível, respectivamente, demonstrando que quanto mais se pode alocar, menor é o bloqueio resultante. Além disso, o AARSASJ se baseia na medida do diâmetro da rede para atribuir a máxima ou mínima taxa aos lotes, enquanto que o RBR é capaz de escalonar lotes e sempre lhes atribui a taxa máxima, sendo que tais mecanismos ajudam na liberação de largura de banda.

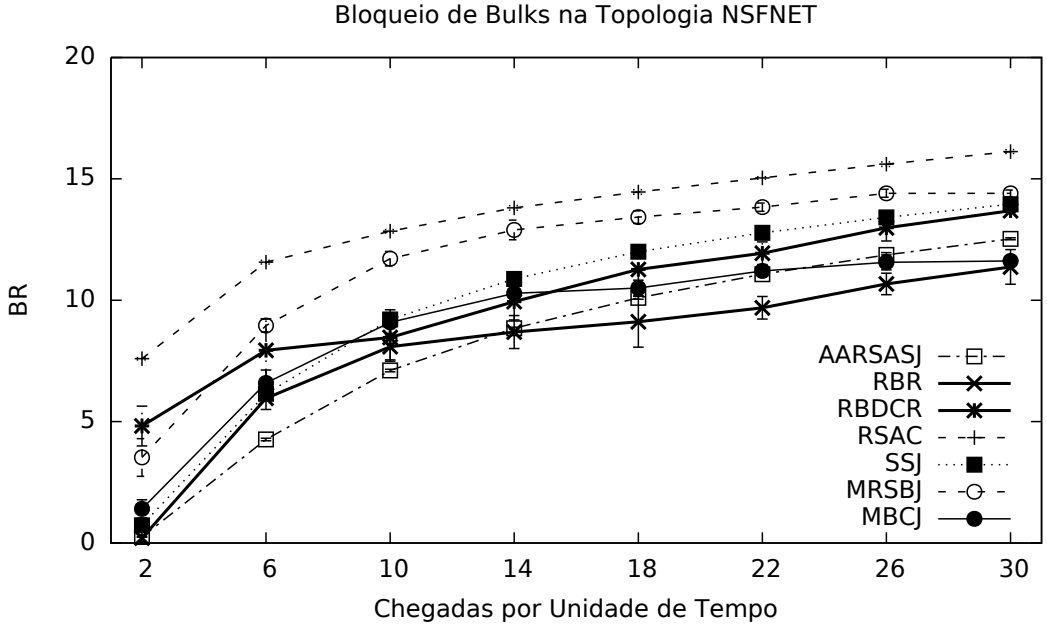

Figura 5.14: Taxa de Bloqueio na rede NSFNET com modulação QPSK e prazo de $20u.t.$ para as requisições de *backup*.

Já os resultados das taxas de bloqueio de r na rede Pan-Euro, mostrados na Figura 5.15, são ligeiramente diferentes dos obtidos na rede NSFNET. É importante destacar que os centros de dados na rede PanEuro estão menos concentrados em pontos específicos da rede. Esse melhor espalhamento contribuiu para a redução dos bloqueios em grande parte dos algoritmos, em comparação com a rede NSFNET, que possui uma leve concentração na parte direita da topologia. O algoritmo convencional RSAC [96] continua com os maiores bloqueios de r (entre 8% e 16%), perceptíveis também na terceira rede analisada, a USA (entre 6.6% e 15.5%), cujos resultados aparecem na Figura 5.16. Também na USA, o algoritmo RBDCR (Algoritmo 10), que atribui a taxa máxima aos *bulks*, possui um bloqueio de 7 a 14% entre o mínimo e máximo de chegadas de chamadas, valores muito próximos do obtido na rede NSFNET(entre 7 a 15%), e seu comportamento ainda apresenta algumas leves oscilações devido a estratégia adotada para atendimento de lotes, com taxas máximas e mínimas, podendo escaloná-las. Como o grau de conectividade da Pan-Euro é menor, dentre todas as topologias, os recursos ficam escassos mais rapidamente.

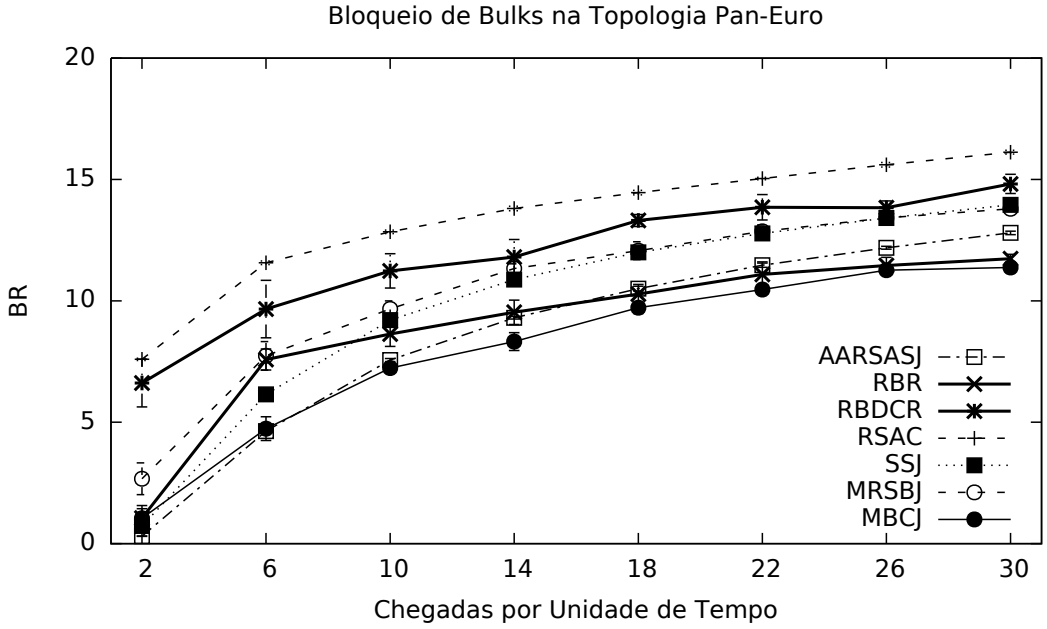

Figura 5.15: Taxa de Bloqueio na rede Pan-Euro com modulação QPSK e prazo de $20u.t.$ para as requisições de *backup*.

Ainda na rede Pan-Euro, os algoritmos SSJ (Algoritmo 7) e MRSBJ (Algoritmo 8) chegam a convergir as taxas de bloqueio a partir de 18 chegadas de requisições, quando atingem respectivamente 11.99% e 12.06%, sendo que a grande diferença entre eles é que o SSJ (Algoritmo 7) não faz escalonamento e atribui a taxa mínima para ambas

as chamadas e o MRSBJ (Algoritmo 8), mais sofisticado, também atende r com a taxa mínima, mas usa a estratégia de atendimento dos lotes fazendo escalonamento, os quais sempre são atendidos com taxa máxima. Devido às condições da rede, tais medidas só fazem diferença enquanto e rede não está saturada.

Os melhores resultados verificados na rede Pan-Euro são dos algoritmos que sempre atendem lotes com a taxa máxima, seja antes do encaminhamento para a janela ou depois dela. Como os lotes demandam largura de banda significativa, e a rede é menos conectada, o atendimento rápido das grandes requisições ajudou na liberação de banda para atendimento das chamadas individuais.

No caso da rede USA (Figura 5.16), com concentração de nós CDs análoga à rede NSFNET, no entanto mais conectada, chama a atenção que o algoritmo MBCJ (Algoritmo 9) mantenha praticamente a mesma taxa de bloqueio para todos os números de chegadas de chamadas, entre 10% e 12.3%. Esse algoritmo faz escalonamento dos dois tipos de requisições sendo que os lotes tem maior prioridade no atendimento, e sempre define a taxa mínima para r e a máxima para os lotes. O encaminhamento de r para a janela não é eficiente porque qualquer lote será atendido primeiro e essas chamadas individuais são ignoradas na maior parte das vezes. Como as ressincronizações sempre são feitas com a máxima banda, o que sobra para alocação de requisições de *backup* geralmente é utilizado imediatamente na oportunidade da chegada de uma nova requisição de *backup*.

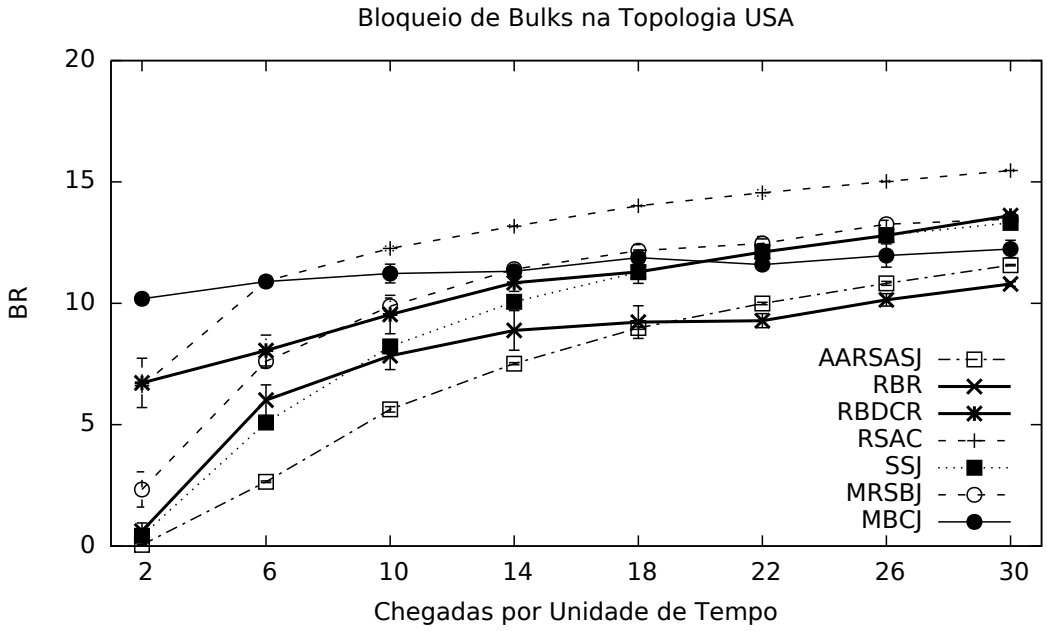

Figura 5.16: Taxa de Bloqueio na rede USA com QPSK e prazo de $20u.t.$ para todas as requisições de *backup*.

O RSAC [96], inicialmente, mostra uma taxa de bloqueio em torno de 6%, que é menor que o algoritmo MBCJ (10%) devido ao estado inicial da rede com ampla disponibilidade, no entanto, a partir de 6 chegadas de chamadas por unidade de tempo, seu índice ultrapassa os 10% de bloqueio e não é mais equiparado com as demais soluções cientes da aplicação. O desempenho do RBDCR (entre 6% e 13%) na rede USA é favorecido inicialmente pela rapidez do atendimento de requisições r com a taxa máxima, mas logo é ultrapassado pelo MRSBJ(entre 2.3% e 13%) que atribui a taxa mínima a eles. O impacto positivo no bloqueio de chamadas neste caso é o fato de o MRSBJ (Algoritmo 8) sempre despachar os lotes com a taxa máxima, gerando rotatividade de banda.

Um simples atendimento de chamadas com a taxa mínima e sem escalonamento, como é o caso do algoritmo SSJ(entre 0.5% e 13%), não é vantajoso com o aumento no número de chegadas. O atendimento de todas as chamadas com taxa máxima em uma rede bem conectada, como faz o RBR(0.6% e 10.7%) - (Algoritmo 11) - não apresenta um bom resultado no início porque requisições r não atendidas são imediatamente bloqueadas, enquanto que os lotes logo no inicio já são prontamente atendidos com a máxima taxa. Depois que os lotes começam a ser encaminhados para a janela, melhores oportunidades surgem para as chamadas individuais. O AARSASJ (Capítulo 4) favoreceu r nesta topologia porque a maior parte das lotes são transferidos em caminhos menores que a metade do diâmetro, e assim, com a taxa máxima, rapidamente desocupam a banda e atendem os *backups*, resultando em taxa de bloqueios entre 0.05% e 11%.

Pode-se inferir a respeito do bloqueio de chamadas individuais r relacionadas com *backups*, que quando a rede é menos conectada, o escalonamento na janela permite esperar que recursos sejam desocupados. Em redes mais conectadas, o ideal foi despachar mais rápido as requisições que demandam muita banda, o que apresentou leve diminuição da carga efetiva na rede e ofereceu mais chances para as solicitações menores.

5.4 Resumo Conclusivo

Este capítulo destaca o uso de escalonamento de requisições para aumentar as chances de uma chamada ser atendida. Também são propostas algumas alternativas de aprovisionamento de largura de banda para todas as chamadas que requerem conexão dentro de um dado limite de tempo (prazo). As simulações mostraram que a execução de duas aplicações de mesma classe de tráfego leva a disputa por recurso de banda na rede.

Se a solução de roteamento é ciente da aplicação, a taxa de sucesso das ressincronizações pode chegar a 80% de diferença, em comparação com uma solução convencional, em uma ambiente com duas aplicações sendo executadas. Esse resultado é alcançado sem

elevar a taxa de bloqueio de requisições de *backup*, visto a solução de roteamento clássica ainda bloqueia 6% a mais.

Apesar de o tráfego *background* suportar grande latência, verificou-se que longos períodos de ocupação da rede pelo mesmo tipo de requisição resultam no bloqueio de chamadas recém-chegadas, principalmente se o recurso solicitado for de granularidade consideravelmente grande. Se a rede é bem conectada, o uso de janela de escalonamento de requisições, além de aumentar as chances de atendimento de um pedido, contribui para uma suave redução de ressincronizações bem sucedidas quando o tráfego aumenta.

Capítulo 6

Considerações Finais

O campo de pesquisa da engenharia de tráfego nas redes de núcleo é bastante promissor, devido ao expressivo aumento e rotatividade de serviços e aplicações hospedados remotamente em centros de dados de grandes provedores de serviços de *Internet* (ISP), que são acessados cada vez mais frequentemente por usuários de muitas partes do mundo.

A tendência natural é que tráfego de diversos tipos de dados se torne cada vez mais intenso, acompanhando a expressiva popularidade da geração de conteúdo pelos usuários, que mais organizações estejam predispostas a deixarem os seus silos isolados de produção e adotem soluções de *software* como serviço (SaaS) e infraestrutura como serviço (IaaS) para simplificar os negócios, e consequentemente, que todo o ecossistema da computação em nuvem sofra vulnerabilidades, tanto com relação a segurança, quanto questões de custos capital e operacional. A boa notícia é a tecnologia das redes EON que poderá ser implantada futuramente, e seu grande potencial de capacidade e eficiência poderá suportar todas essas tendências.

Este trabalho propôs soluções cientes da aplicação para resolver o problema de ressincronização entre centros de dados tolerante a falhas bizantinas, usando uma abordagem de comunicação entre camadas cruzadas (CLD).

Como a OTN adotada nas redes de transporte entre os centros de dados é a de EON, o Capítulo 2 introduziu alguns pontos sobre a arquitetura a ser implementada, o funcionamento dos equipamentos ópticos fundamentais, e a inovação trazida com a OFDM óptica, que aumentará a capacidade de transporte dessas redes. Esse mesmo capítulo definiu o ambiente da rede de centro de dados, apresentou conceitos de aplicações distribuídas que exploram essa infraestrutura, focando em ressincronizações de partições de dados nesses centros de dados.

O Capítulo 3 destacou os principais trabalhos desenvolvidos recentemente para prover o RSA e o RMLSA, bem como roteamento convencional e o roteamento ciente. Também destacou soluções utilizadas para realizar as MBDT.

No Capítulo 4 foram mostrados os primeiros testes com os dois tipos de roteamento, convencional e ciente, executados em uma rede de centro de dados para sincronizar CD com estado desatualizado com relação a um partição de dados. Já o Capítulo 5 apresentou o problema do atendimento de pedidos de ressincronização em um cenário onde uma segunda aplicação também solicitava os mesmos recursos, para o qual foram propostos algoritmos cientes que fazem escalonamento de requisições. Em todos os experimentos, os algoritmos cientes tiveram melhor desempenho que o algoritmo convencional, com taxa de sucesso superior em 30%, quando uma única aplicação é executada. Se duas aplicações disputam os recursos, essa diferença sobe para 80%.

Como a EON é uma solução de longo prazo e oferece inúmeras vantagens, os ISPs podem sentir-se atraídos a fazer grandes investimentos. Entretanto, investimentos substanciais em equipamentos não surtirá muito efeito se a solução de aprovisionamento de recurso continuar sendo convencional.

Durante esse processo, dois pontos chaves ficaram evidentes. Primeiramente, implementar CLD traz melhorias de desempenho recíprocas para as entidades envolvidas, como as camadas de aplicação e de rede. A aplicação executada obteve melhores resultados, com aumento na taxa de atendimento dos seus serviços. Por sua vez, a camada de rede fez um ágil trabalho e o recurso aprovisionado foi habilmente designado para requisições com fortes probabilidades de sucesso, dadas as condições das chamadas em lotes.

Em segundo lugar, percebe-se que o paradigma CLD tem se firmado no campo das redes ópticas em uma abordagem ciente das limitações da camada física, no entanto, a abordagem ciente da aplicação não tem recebido muita atenção da comunidade científica. Foi apontado que, quando a EON estiver em operação, o grau de programabilidade da sua arquitetura permitirá medições dinâmicas de parâmetros da camada física, viabilizando o roteamento ciente de tais limitações [81]. Mas também, existe um grande espaço a ser explorado pelo roteamento ciente da aplicação, uma vez que, o plano de controle da EON incluirá elementos de SDN [26], o que pode reduzir ou condensar camadas que abreviam a comunicação e colaborar para a coordenação entre camadas. Finalmente, soluções cientes das limitações da camada física e das aplicações são o futuro da pesquisa em engenharia de tráfego em EON.

6.1 Trabalhos Futuros

Existem várias possíveis melhorias a serem feitas nos trabalhos propostos:

- Proposição de uma solução RSA que combine IA-RSA com AA-RSA. Dado que a nova arquitetura e configuração EON vai permitir que informações específicas das camadas sejam facilmente obtidas, é possível que soluções CLD explore tais

informações na elaboração de algoritmos de roteamento mais inteligentes e multi criteriosos.

- Proposição de um RMLSA ciente da aplicação. As soluções propostas são baseadas em RSA e são de modulação única para todos os caminhos configurados na rede. Assim, resolver esse problema incorre em determinar um caminho e fazer busca de banda disponível. Já o problema de roteamento e alocação de nível de modulação e espectro (RMLSA) leva em consideração as limitações físicas de alcance do sinal óptico antes de escolher o canal de transmissão. Essa seleção mais criteriosa poderia melhorar a acurácia dos aprovisionamentos, elevando a taxa de atendimento.

- Aprovisionamento para diferentes classes de tráfego. As soluções propostas trataram de ressincronizações e *backups* entre centros de dados. Ambas as aplicações pertencem à mesma classe de tráfego, e portanto, possuem muitas características em comum, como um prazo estabelecido e uma prioridade mais baixa, ainda que tenham características peculiares, como o número de origens e destinos da transmissão, que serve para medir, por exemplo, o grau de impacto que pode ser atingido, se o serviço falhar. Uma ressincronização perdida é mais grave que um *backup* não atendido devido ao número de entidades que poderão sofrer as consequências das perdas. Tratar de múltiplas classes de tráfego exige que uma série de políticas sejam adotadas, não apenas para o atendimento justo de tráfego com diferentes prioridades, mas também, para o atendimento de aplicações suficientemente distintas dentro dessas classes, de maneira justa e eficiente.

- Atribuição de taxas variando no tempo. As requisições orientadas a dados, devido a sua característica elástica, naturalmente possuem um amplo espaço para exploração de taxas dinâmicas. As soluções apresentadas utilizaram taxas de transmissão máximas e mínimas fixas, que eram sempre calculadas após a atualização de recursos disponível na rede. O grau de programabilidade de EON e o amplo espaço de exploração do paradigma CLD podem permitir que transmissões de dados sejam ajustados dinamicamente enquanto estão em curso, abrindo margem para otimizar a utilização de largura de banda.

Referências

[1] Finding the k shortest loopless paths in a network. *Management Science*, 17(11):712–716, 1971. 48, 50, 67, 68, 79

[2] Cisco global cloud index: Forecast and methodology 2013-2018. White Paper, 2014. 2

[3] Divyakant Agrawal, Amr El Abbadi, Shyam Antony, e Sudipto Das. Data management challenges in cloud computing infrastructures. In *Databases in Networked Information Systems*, pages 1–10. Springer, 2010. 18, 20, 24, 45, 63

[4] Divyakant Agrawal, Amr El Abbadi, Hatem A Mahmoud, Faisal Nawab, e Kenneth Salem. Managing geo-replicated data in multi-datacenters. In *Databases in Networked Information Systems*, pages 23–43. Springer, 2013. 2, 26, 45, 52, 63

[5] Michal Aibin e Krzysztof Walkowiak. Adaptive modulation and regenerator-aware dynamic routing algorithm in elastic optical networks. In *Communications (ICC), 2015 IEEE International Conference on*, pages 5138–5143. IEEE, 2015. 35

[6] L. Al-Tarawneh e S. Taebi. Linear dynamic adaptation of the bw granularity allocation for elastic optical ofdm networks. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2015 International Symposium on*, pages 1–7, July 2015. 12, 35, 41

[7] Anwar Al-Yatama. Computing blocking probabilities in survivable wdm optical networks. *Photonic Network Communications*, 27(1):34–46, 2014. 10

[8] Leon-Garcia Alberto e Widjaja Indra. Communication networks: fundamental concepts and key architectures. *Mc GrawHill*, pages 845–857, 2000. 13, 14

[9] A. Asensio, M. Ruiz, e L. Velasco. Routing and scheduled spectrum allocation for transfer-based datacenter connections. In *Transparent Optical Networks (ICTON), 2015 17th International Conference on*, pages 1–4, July 2015. 37, 42

[10] A. Avizienis, J. C. Laprie, B. Randell, e C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1):11–33, Jan 2004. 20

[11] M. Balman. Advance resource provisioning in bulk data scheduling. In *Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on*, pages 984–992, March 2013. 1, 3, 27, 37, 42, 61

[12] Sanjay Bansal, Sanjeev Sharma, e Ishita Trivedi. A detailed review of fault-tolerance techniques in distributed system. *International Journal on Internet and Distributed Computing Systems*, 1(1), 2011. 22

[13] Nabil Bitar, Steven Gringeri, Tiejun J Xia, et al. Technologies and protocols for data center and cloud networking. *IEEE Communications Magazine*, 51(9):24–31, 2013. 15

[14] Giorgio Buttazzo. *Hard real-time computing systems: predictable scheduling algorithms and applications*, volume 24. Springer Science & Business Media, 2011. 27

[15] Miguel Castro e Barbara Liskov. Practical Byzantine fault-tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4):398–461, November 2002. 24, 25, 45, 63

[16] Tushar Deepak Chandra e Sam Toueg. Unreliable failure detectors for reliable distributed systems. *J. ACM*, 43(2):225–267, March 1996. 22

[17] M Channegowda, R Nejabati, M Rashidi Fard, S Peng, N Amaya, G Zervas, D Simeonidou, R Vilalta, R Casellas, R Martínez, et al. Experimental demonstration of an openflow based software-defined optical network employing packet, fixed and flexible dwdm grid technologies on an international multi-domain testbed. *Optics express*, 21(5):5487–5498, 2013. 4, 13

[18] Xiaomin Chen, André C Drummond, Admela Jukan, e Nelson LS da Fonseca. Multipath routing with topology aggregation for scalable inter-domain service provisioning in optical networks. *Optical Switching and Networking*, 9(4):314–322, 2012. 27, 38, 41, 62

[19] Xiaomin Chen, Yuesheng Zhong, e Admela Jukan. Multipath routing in elastic optical networks with distance-adaptive modulation formats. In *Communications (ICC), 2013 IEEE International Conference on*, pages 3915–3920. IEEE, 2013. 15, 35, 41

[20] Li-Chia Cheng, Kuochen Wang, e Yi-Huai Hsu. Application-aware routing scheme for sdn-based cloud datacenters. In *Ubiquitous and Future Networks (ICUFN), 2015 Seventh International Conference on*, pages 820–825. IEEE, 2015. 2, 4, 26

[21] Parminder Chhabra, Vijay Erramilli, Nikolaos Laoutaris, Ravi Sundaram, e Pablo Rodriguez. Algorithms for constrained bulk-transfer of delay-tolerant data. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010. 2, 38, 42

[22] Konstantinos Christodoulopoulos, I Tomkos, e EA Varvarigos. Elastic bandwidth allocation in flexible ofdm-based optical networks. *Journal of Lightwave Technology*, 29(9):1354–1366, 2011. 11, 15, 34, 41

[23] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009. 50, 62, 79, 80, 83

[24] Lucas R. Costa, Léia S. de Sousa, Felipe R. de Oliveira, Kaio A. da Silva, Paulo J. S. Júnior, e André C. Drummond. Ons: Optical network simulator - wdm/eon. http://comnet.unb.br/br/grupos/get/ons. 50

[25] George Coulouris, Jean Dollimore, Tim Kindberg, e Gordon Blair. *Distributed Systems: Concepts and Design.* Addison-Wesley Publishing Company, USA, 5th edition, 2011. 20, 21, 23

[26] Filippo Cugini, Francesco Paolucci, Francesco Fresi, Gianluca Meloni, Nicola Sambo, Luca Potí, Antonio D'Errico, e Piero Castoldi. Toward plug-and-play software-defined elastic optical networks. *J. Lightwave Technol.*, 34(6):1494–1500, Mar 2016. 32, 33, 97

[27] Chris Develder, Marc De Leenheer, Bart Dhoedt, Mario Pickavet, Didier Colle, Filip De Turck, e Piet Demeester. Optical networks for grid and cloud computing applications. *Proceedings of the IEEE*, 100(5):1149–1167, 2012. 12

[28] M.N. Dharmaweera, R. Parthiban, e Y.A. Sekercioglu. Toward a power-efficient backbone network: The state of research. *Communications Surveys Tutorials, IEEE*, 17(1):198–227, Firstquarter 2015. 32

[29] Jerzy Domzal. Intelligent routing in congested approximate flow-aware networks. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 1751–1756. IEEE, 2012. 30

[30] A. C. Drummond. WDMSim: WDM Optical Network Simulator. `http://www.lrc.ic.unicamp.br/wdmsim/`. 50

[31] Andre Costa Drummond. *Agregação de tráfego em redes ópticas com multiplexação por comprimetos de onda.* Tese (Doutorado), Universidade Estadual de Campinas, São Paulo, 2010. 10

[32] Cynthia Dwork, Nancy Lynch, e Larry Stockmeyer. Consensus in the presence of partial synchrony. *J. ACM*, 35(2):288–323, April 1988. 21

[33] Ahmad Fallahpour, Hamzeh Beyranvand, e Jawad A Salehi. Energy-efficient manycast routing and spectrum assignment in elastic optical networks for cloud computing environment. *Journal of Lightwave Technology*, 33(19):4008–4018, 2015. 36

[34] Pesech Feldman e Silvio Micali. An optimal probabilistic protocol for synchronous byzantine agreement. *SIAM J. Comput.*, 26(4):873–933, August 1997. 22, 24

[35] M. Fiorani, S. Aleksic, P. Monti, J. Chen, M. Casoni, e L. Wosinska. Energy efficiency of an integrated intra-data-center and core network with edge caching. *Optical Communications and Networking, IEEE/OSA Journal of*, 6(4):421–432, April 2014. 19

[36] Michael J. Fischer, Nancy A. Lynch, e Michael S. Paterson. Impossibility of distributed consensus with one faulty process. *J. ACM*, 32(2):374–382, April 1985. 21, 22

[37] Felix C Gärtner. Fundamentals of fault-tolerant distributed computing in asynchronous environments. *ACM Computing Surveys (CSUR)*, 31(1):1–26, 1999. 22, 26

[38] A. Gerber e R. Doverspike. Traffic types and growth in backbone networks. In *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, pages 1–3, March 2011. 3, 18

[39] Ori Gerstel, Masahiko Jinno, Andrew Lord, e SJ Ben Yoo. Elastic optical networking: A new dawn for the optical layer? *Communications Magazine, IEEE*, 50(2):s12–s20, 2012. 4, 10

[40] Amitabha Ghosh, Sangtae Ha, Edward Crabbe, e Jennifer Rexford. Scalable multi-class traffic management in data center backbone networks. *Selected Areas in Communications, IEEE Journal on*, 31(12):2673–2684, 2013. 3, 18

[41] Phillipa Gill, Navendu Jain, e Nachiappan Nagappan. Understanding network failures in data centers: measurement, analysis, and implications. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 350–361. ACM, 2011. 3

[42] Google. Google data centers. `https://www.google.com/maps/d/viewer?mid=zXkGqQo6GvyA.kRMGK2Zpmpbg&hl=en`, 2015. Acessed em 16/03/2015. 19, 50

[43] Chi-Yao Hong, Srikanth Kandula, Ratul Mahajan, Ming Zhang, Vijay Gill, Mohan Nanduri, e Roger Wattenhofer. Achieving high utilization with software-driven wan. In *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, SIGCOMM '13, pages 15–26, New York, NY, USA, 2013. ACM. 26, 27

[44] Pankaj Jalote. *Fault Tolerance in Distributed Systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1994. 20, 21, 22, 23, 24

[45] Masahiko Jinno, Hidehiko Takara, Bartlomiej Kozicki, Yukio Tsukishima, Yoshiaki Sone, e Shinji Matsuoka. Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. *Communications Magazine, IEEE*, 47(11):66–73, 2009. 9, 10, 11, 15, 32, 33

[46] M. Klinkowski e K. Walkowiak. Routing and spectrum assignment in spectrum sliced elastic optical path network. *Communications Letters, IEEE*, 15(8):884–886, August 2011. 34, 41

[47] Miroslaw Klinkowski, Davide Careglio, et al. A routing and spectrum assignment problem in optical ofdm networks. In *First European Teletraffic Seminar*, 2011. 11

[48] B. Kozicki, H. Takara, Y. Sone, A. Watanabe, e M. Jinno. Distance-adaptive spectrum allocation in elastic optical path network (slice) with bit per symbol adjustment. In *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pages 1–3, March 2010. 34

[49] Alok Kumar, Sushant Jain, Uday Naik, Nikhil Kasinadhuni, Enrique Cauich Zermeno, C. Stephen Gunn, Jing Ai, Björn Carlin, Mihai Amarandei-Stavila, Mathieu Robin, Aspi Siganporia, Stephen Stuart, e Amin Vahdat. Bwe: Flexible, hierarchical bandwidth allocation for wan distributed computing. In *Sigcomm '15*, 2015. 39, 62

[50] SathiyaPrabhu Kumar, Sylvain Lefebvre, Raja Chiky, e Eric Gressier-Soudan. Calibre: A better consistency-latency tradeoff for quorum based replication systems. In Qiming Chen, Abdelkader Hameurlain, Farouk Toumani, Roland Wagner, e Hendrik Decker, editors, *Database and Expert Systems Applications*, volume 9262 of *Lecture Notes in Computer Science*, pages 491–503. Springer International Publishing, 2015. 26

[51] Caroline P Lai e Keren Bergman. Cross-layer communications for high-bandwidth optical networks. In *Transparent Optical Networks (ICTON), 2010 12th International Conference on*, pages 1–4. IEEE, 2010. 29

[52] Cedric F Lam, Hong Liu, Bikash Koley, Xiaoxue Zhao, Valey Kamalov, e Vijay Gill. Fiber optic communication technologies: What's needed for datacenter network operations. *IEEE Communications Magazine*, 48(7):32–39, 2010. 27

[53] Nikolaos Laoutaris, Michael Sirivianos, Xiaoyuan Yang, e Pablo Rodriguez. Inter-datacenter bulk transfers with netstitcher. *ACM SIGCOMM Computer Communication Review*, 41(4):74–85, 2011. 27, 37, 42, 62

[54] Nikolaos Laoutaris, Georgios Smaragdakis, Pablo Rodriguez, e Ravi Sundaram. Delay tolerant bulk data transfers on the internet. In *ACM SIGMETRICS Performance Evaluation Review*, volume 37, pages 229–238. ACM, 2009. 1, 2, 3, 27, 39, 42

[55] X. Lin, W. Sun, M. Veeraraghavan, e W. Hu. Time-shifted multilayer graph: A routing framework for bulk data transfer in optical circuit-switched networks with assistive storage. *IEEE/OSA Journal of Optical Communications and Networking*, 8(3):162–174, March 2016. 40, 42

[56] X. Liu, L. Gong, e Z. Zhu. On the spectrum-efficient overlay multicast in elastic optical networks built with multicast-incapable switches. *IEEE Communications Letters*, 17(9):1860–1863, September 2013. 15, 18

[57] Xiahe Liu, Long Gong, e Zuqing Zhu. Design integrated rsa for multicast in elastic optical networks with a layered approach. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pages 2346–2351. IEEE, 2013. 13, 35, 41

[58] Xiaoxu Liu, Rentao Gu, Yuefeng Ji, Lin Bai, e Zhitong Huang. Fragmentation-based dynamic multicast routing and spectrum assignment algorithm in spectrum-sliced elastic optical path networks. In *Asia Communications and Photonics Conference*, pages AF2F–37. Optical Society of America, 2013. 13, 35, 41

[59] Ping Lu, Kaiyue Wu, Quanying Sun, e Zuqing Zhu. Toward online profit-driven scheduling of inter-dc data-transfers for cloud applications. In *Communications (ICC), 2015 IEEE International Conference on*, pages 5583–5588. IEEE, 2015. 39, 42

[60] Wei Lu e Zuqing Zhu. Malleable reservation based bulk-data transfer to recycle spectrum fragments in elastic optical networks. *Lightwave Technology, Journal of*, 33(10):2078–2086, May 2015. 28, 34, 41

[61] Wei Lu, Zuqing Zhu, e B. Mukherjee. Data-oriented malleable reservation to revitalize spectrum fragments in elastic optical networks. In *Optical Fiber Communications Conference and Exhibition (OFC), 2015*, pages 1–3, March 2015. 1, 46

[62] Wei Lu, Zuqing Zhu, e Biswanath Mukherjee. Optimizing deadline-driven bulk-data transfer to revitalize spectrum fragments in eons [invited]. *Journal of Optical Communications and Networking*, 7(12):B173–B183, 2015. 34, 40, 41, 42

[63] T. Lynn, P. Healy, S. Kilroy, G. Hunt, L. van der Werff, S. Venkatagiri, e J. Morrison. Towards a general research framework for social media research using big data. In *2015 IEEE International Professional Communication Conference (IPCC)*, pages 1–8, July 2015. 2

[64] Ajay Mahimkar, Angela Chiu, Robert Doverspike, Mark D Feuer, Peter Magill, Emmanuil Mavrogiorgis, Jorge Pastor, Sheryl L Woodward, e Jennifer Yates. Bandwidth on demand for inter-data center communication. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, page 24. ACM, 2011. 2, 18, 19, 38

[65] M. Marcon, N. Santos, K.P. Gummadi, N. Laoutaris, P. Rodriguez, e A. Vahdat. Netex: Efficient and cost-effective internet bulk content delivery. In *Architectures for Networking and Communications Systems (ANCS), 2010 ACM/IEEE Symposium on*, pages 1–2, Oct 2010. 1, 37, 42, 62

[66] Peter Mell e Tim Grance. The nist definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50, 2009. 1

[67] A Middleton. Hpcc systems: data intensive supercomputing solutions. *White paper, LexisNexis Risk Solutions*, 2011. 1

[68] Atanas Mirchev. Survey of concepts for qos improvements via sdn. *Future Internet (FI) and Innovative Internet Technologies and Mobile Communications (IITM)*, 33, 2015. 1, 2, 26

[69] T Morioka, M Jinno, H Takara, e H Kubota. Innovative future optical transport network technologies. *NTT Technical Review*, 9(8):2011, 2011. 2, 11, 12

[70] Amina Mseddi, Mohammad Ali Salahuddin, Mohamed Faten Zhani, Halima Elbiaze, e Roch H Glitho. On optimizing replica migration in distributed cloud storage systems. In *Cloud Networking (CloudNet), 2015 IEEE 4th International Conference on*, pages 191–197. IEEE, 2015. 1, 24, 25, 26

[71] Kim-Khoa Nguyen, Mohamed Cheriet, e Yves Lemieux. Virtual slice assignment in large-scale cloud interconnects. *Internet Computing, IEEE*, 18(4):37–46, 2014. 3, 38

[72] S. Peng, R. Nejabati, M. Channegowda, e D. Simeonidou. Application-aware and adaptive virtual data centre infrastructure provisioning over elastic optical ofdm networks. In *Optical Communication (ECOC 2013), 39th European Conference and Exhibition on*, pages 1–3, Sept 2013. 36, 41, 43

[73] Yongmao Ren, Haina Tang, Jun Li, e Hualin Qian. A novel congestion control algorithm for high performance bulk data transfer. In *Network Computing and Applications, 2009. NCA 2009. Eighth IEEE International Symposium on*, pages 288–291, July 2009. 42

[74] J. Santos-Aguilar, M. Santiago-Bernal, e C. Gutierrez-Martinez. Filtering the spectrum of multi-longitudinal lasers by using optical retarders. In *Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE*, pages 1–4, May 2011. 10

[75] Fred B. Schneider. Implementing fault-tolerant services using the state machine approach: A tutorial. *ACM Comput. Surv.*, 22(4):299–319, December 1990. 22, 26

[76] Nasser Sedaghati-Mokhtari, Mahdi Nazm Bojnordi, e Nasser Yazdani. Cross-layer design: A new paradigm. In *Communications and Information Technologies, 2006. ISCIT'06. International Symposium on*, pages 183–188. IEEE, 2006. 28

[77] Ziyu Shao, Xin Jin, Wenjie Jiang, Minghua Chen, e Mung Chiang. Intra-data-center traffic engineering with ensemble routing. In *INFOCOM, 2013 Proceedings IEEE*, pages 2148–2156. IEEE, 2013. 19

[78] Artyom Sharov, Alexander Shraer, Arif Merchant, e Murray Stokely. Automatic reconfiguration of distributed storage. In *Autonomic Computing (ICAC), 2015 IEEE International Conference on*, pages 133–134. IEEE, 2015. 21, 45

[79] Gangxiang Shen e Qi Yang. From coarse grid to mini-grid to gridless: How much can gridless help contentionless? In *Optical Fiber Communication Conference*, page OTuI3. Optical Society of America, 2011. 12

[80] Min Shen, Ajay D. Kshemkalyani, e Ta Yuan Hsu. Causal consistency for geo-replicated cloud storage under partial replication. In *IPDPS Workshops*, pages 509–518. IEEE, 2015. 2

[81] William Shieh. OFDM for Flexible High-Speed Optical Networks. *IEEE, Journal of Lightwave Technology*, 29(10):1560–1577, May 2011. 14, 33, 97

[82] Jane M Simmons. *Optical network design and planning*. Springer International Publishing, 2 edition, 2014. 14

[83] Joana Sócrates-Dantas, Davide Careglio, Jordi Perelló, Regina Melo Silveira, Wilson Vicente Ruggiero, e Josep Solè-Pareta. Challenges and requirements of a control plane for elastic optical networks. *Computer Networks*, 72:156–171, 2014. 10, 12

[84] Vineet Srivastava e Mehul Motani. Cross-layer design: a survey and the road ahead. *Communications Magazine, IEEE*, 43(12):112–119, 2005. 28

[85] Suresh Subramaniam, Maïté Brandt-Pearce, Piet Demeester, e Chava Vijaya Saradhi. *Cross-layer design in optical networks*, volume 15. Springer, 2013. 28, 29

[86] T. Takagi, H. Hasegawa, K. Sato, Y. Sone, B. Kozicki, A. Hirano, e M. Jinno. Dynamic routing and frequency slot assignment for elastic optical path networks that adopt distance adaptive modulation. In *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, pages 1–3, March 2011. 34, 41

[87] H. Takara, B. Kozicki, Y. Sone, T. Tanaka, A. Watanabe, A. Hirano, K. Yonenaga, e M. Jinno. Distance-adaptive super-wavelength routing in elastic optical path network (slice) with optical ofdm. In *36th European Conference and Exhibition on Optical Communication*, pages 1–3, Sept 2010. 32, 33

[88] Andrew TANENBAUM. *S. Redes de Computadores. São Paulo: Ed.* Campus, 2003. 20

[89] Andrew S Tanenbaum e Maarten Van Steen. *Distributed systems.* Prentice-Hall, 2007. 21

[90] Ioannis Tomkos, Siamak Azodolmolky, Josep Sole-Pareta, Davide Careglio, e Eleni Palkopoulou. A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges. *Proceedings of the IEEE*, 102(9):1317–1337, 2014. 11, 33

[91] Christos Tsekrekos, Wladek Forysiak, Robert Killey, Francesco Matera, Michela Svaluto Moreolo, e Jeroen Nijhof. State of the art on transmission techniques. In *Optical Transmission*, pages 1–52. Springer, 2011. 2

[92] L Velasco, A Asensio, JLl Berral, A Castro, e Víctor López. Towards a carrier sdn: an example for elastic inter-datacenter connectivity. *Optics express*, 22(1):55–61, 2014. 27, 37, 42

[93] V. Venkatesan, I. Iliadis, X. Y. Hu, R. Haas, e C. Fragouli. Effect of replica placement on the reliability of large-scale data storage systems. In *2010 IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 79–88, Aug 2010. 24, 25

[94] Marko Vukolić. The byzantine empire in the intercloud. *ACM SIGACT News*, 41(3):105–111, 2010. 25, 46, 48, 50, 63

[95] Krzysztof Walkowiak, Andrzej Kasprzak, e Mirosław Klinkowski. Dynamic routing of anycast and unicast traffic in elastic optical networks. In *Communications (ICC), 2014 IEEE International Conference on*, pages 3313–3318. IEEE, 2014. 36

[96] Xin Wan, Nan Hua, e Xiaoping Zheng. Dynamic routing and spectrum assignment in spectrum-flexible transparent optical networks. *Journal of Optical Communications and Networking*, 4(8):603–613, 2012. 34, 41, 42, 43, 44, 47, 48, 50, 62, 63, 73, 74, 81, 82, 85, 87, 89, 90, 92, 94

[97] Xin Wan, Lei Wang, Nan Hua, Hanyi Zhang, e Xiaoping Zheng. Dynamic routing and spectrum assignment in flexible optical path networks. In *Optical Fiber Communication Conference*, page JWA055. Optical Society of America, 2011. 15, 16

[98] J. M. Wang, Y. Wang, X. Dai, e B. Bensaou. Sdn-based multi-class qos guarantee in inter-data center communications. *IEEE Transactions on Cloud Computing*, PP(99):1–1, 2016. 2, 26, 28

[99] Ting Wang, Zhiyang Su, Yu Xia, e M. Hamdi. Rethinking the data center networking: Architecture, network protocols, and resource sharing. *Access, IEEE*, 2:1481–1496, 2014. 18, 19

[100] Yiwen Wang, Sen Su, Sujuan Jiang, Zhongbao Zhang, e Kai Shuang. Optimal routing and bandwidth allocation for multiple inter-datacenter bulk data transfers. In *Communications (ICC), 2012 IEEE International Conference on*, pages 5538–5542. IEEE, 2012. 27, 39, 42, 62

[101] Y. Wu, Z. Zhang, C. Wu, C. Guo, Z. Li, e F. Lau. Orchestrating bulk data transfers across geo-distributed datacenters. *Cloud Computing, IEEE Transactions on*, PP(99):1–1, 2015. 18, 19

[102] Song Yang e Fernando Kuipers. Impairment-aware routing in translucent spectrum-sliced elastic optical path networks. In *Networks and Optical Communications (NOC), 2012 17th European Conference on*, pages 1–6. IEEE, 2012. 29

[103] Yichao Yang, Yanbo Zhou, Lei Liang, Dan He, e Zhili Sun. A sevice-oriented broker for bulk data transfer in cloud computing. In *Grid and Cooperative Computing (GCC), 2010 9th International Conference on*, pages 264–269, Nov 2010. 37, 41

[104] Jingjing Yao, Ping Lu, Long Gong, e Zuqing Zhu. On fast and coordinated data backup in geo-distributed optical inter-datacenter networks. *Journal of Lightwave Technology*, 33(14):3005–3015, 2015. 63

[105] SJ Ben Yoo, Yawei Yin, e Ke Wen. Intra and inter datacenter networking: The role of optical packet switching and flexible bandwidth optical networking. In *Optical Network Design and Modeling (ONDM), 2012 16th International Conference on*, pages 1–6. IEEE, 2012. 19

[106] Guoying Zhang, Marc De Leenheer, Annalisa Morea, e Biswanath Mukherjee. A survey on ofdm-based elastic core optical networking. *Communications Surveys & Tutorials, IEEE*, 15(1):65–87, 2013. 4, 11, 13, 15, 33

[107] Hong Zhang, Kai Chen, Wei Bai, Dongsu Han, Chen Tian, Hao Wang, Haibing Guan, e Ming Zhang. Guaranteeing deadlines for inter-datacenter transfers. In *Proceedings of the Tenth European Conference on Computer Systems*, EuroSys '15, pages 20:1–20:14, New York, NY, USA, 2015. ACM. 26, 27, 28, 39, 42, 45, 62, 82

[108] Yan Zhang e N. Ansari. On architecture design, congestion notification, tcp incast and power consumption in data centers. *Communications Surveys Tutorials, IEEE*, 15(1):39–64, First 2013. 19

[109] Juzi Zhao, Henk Wymeersch, e Erik Agrell. Nonlinear impairment aware resource allocation in elastic optical networks. In *Optical Fiber Communication Conference*, pages M2I–1. Optical Society of America, 2015. 32

[110] Xiaoxue Zhao, Vijay Vusirikala, Bikash Koley, Valey Kamalov, e Tad Hofmeister. The prospect of inter-data-center optical networks. *Communications Magazine, IEEE*, 51(9):32–38, 2013. 19

[111] Yaoquan Zhong, Wei Guo, Yaohui Jin, Weiqiang Sun, e Weisheng Hu. Routing for deadline-constrained bulk data transfers based on transfer failure probability. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5, June 2011. 40, 42

[112] Zhili Zhou, Tachun Lin, Krishnaiyan Thulasiraman, Guoliang Xue, e Sartaj Sahni. Cross-layer network survivability under multiple cross-layer metrics. *Journal of Optical Communications and Networking*, 7(6):540–553, 2015. 29

[113] Thomas Zinner, Michael Jarschel, Andreas Blenk, Florian Wamser, e Wolfgang Kellerer. Dynamic application-aware resource management using software-defined networking: Implementation prospects and challenges. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–6. IEEE, 2014. 4, 29