**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

André Luiz Ferraz Lourenço

**Roteamento, alocação de espectro e proteção
compartilhada em redes ópticas elásticas utilizando
aprendizagem profunda com sistema de inferência fuzzy**

São Carlos

2021

André Luiz Ferraz Lourenço

**Roteamento, alocação de espectro e proteção
compartilhada em redes ópticas elásticas utilizando
aprendizagem profunda com sistema de inferência fuzzy**

Tese apresentada à Escola de Engenharia de
São Carlos da Universidade de São Paulo para
obtenção do título de Doutor em Ciências -
Programa de Pós-Graduação em Engenharia
Elétrica.

Área de concentração: Telecomunicações

Orientador: Prof. Dr. Amílcar Careli Cesar

**São Carlos
2021**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

L892r	<p>Lourenço, André Luiz Ferraz</p> <p>Roteamento, alocação de espectro e proteção compartilhada em redes ópticas elásticas utilizando aprendizagem profunda com sistema de inferência fuzzy / André Luiz Ferraz Lourenço; orientador Amílcar Careli César. São Carlos, 2021.</p> <p>Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Telecomunicações -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2021.</p> <p>1. sobrevivência da rede. 2. proteção de caminho de backup compartilhado. 3. sistema de inferência fuzzy. 4. rede neural profunda. 5. algoritmo de roteamento e alocação de espectro. 6. rede óptica elástica. I. Título.</p>
-------	---

FOLHA DE JULGAMENTO

Candidato: Engenheiro **ANDRÉ LUIZ FERRAZ LOURENÇO**.

Título da tese: "Roteamento, alocação de espectro e proteção compartilhada em redes ópticas elásticas utilizando aprendizagem profunda com sistema de inferência *fuzzy*".

Data da defesa: 16/04/2021.

Comissão Julgadora

Resultado

Prof. Titular **Amilcar Careli Cesar**
(Orientador)

APROVADO

(Escola de Engenharia de São Carlos – EESC/USP)

Profa. Dra. **Mônica de Lacerda Rocha**

A PROVA DO

(Escola de Engenharia de São Carlos – EESC/USP)

Prof. Dr. **Marcelo Luis Francisco Abbade**

APROVADO

(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP-São João da Boa Vista)

Profa. Dra. **Márcia Helena Moreira Paiva**

APROVADO

(Universidade Federal do Espírito Santo/UFES)

Prof. Dr. **Luiz Henrique Bonani do Nascimento**

APROVADO

(Universidade Federal do ABC/UFABC)

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:

Prof. Dr. **João Bosco Augusto London Junior**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Murilo Araujo Romero**

*Dedico este trabalho aos meus pais
José Rubens e Rosane, à minha irmã Larissa,
e a toda minha família*

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Gostaria de agradecer às pessoas importantes que foram essenciais para a conclusão de mais esta etapa em minha vida.

Agradeço aos meus pais, José Rubens e Rosane e à minha irmã Larissa por ter apoiado em todas as minhas decisões e estarem sempre ao meu lado cuidando de mim com muito amor e carinho.

Ao orientador Dr. Amílcar Careli César, pelo exemplo de professor e pesquisador que me proporcionou esta oportunidade, orientando meu aprendizado.

A professora Dra. Mônica de Lacerda Rocha por me oferecer várias oportunidades de participar de bancas de TCC, engrandecendo o meu aprendizado.

Aos demais docentes do departamento pelo conhecimento transmitido nas disciplinas ministradas.

A CAPES pela bolsa de estudos e ao CNPq pela aprovação do projeto que permitiu a obtenção de novos computadores para o laboratório.

A EESC-USP pela infraestrutura oferecida.

Aos meus amigos que conviveram comigo todos esses anos Pedro Oyama, Yuri Martins, Marcelo Barbosa, Daniel Sírio, Guilherme El Choueiri por proporcionar um momentos de alegria e descontração.

A Chien Ling Tseng, pela amizade atenciosa e pura que sempre me deu todo apoio e me fez crescer espiritualmente.

*“Um mal é um mal. Menor, maior, médio, tanto faz...
As proporções são convencionadas e as fronteiras, imprecisas.
Não sou um santo eremita e não pratiquei apenas o bem ao longo de minha vida.
Mas, se me couber escolher entre dois males,
prefiro abster-me por completo da escolha.”*
The Witcher - Andrzej Sapkowski

RESUMO

LOURENCO, A.L.F. **Roteamento, alocação de espectro e proteção compartilhada em redes ópticas elásticas utilizando aprendizagem profunda com sistema de inferência fuzzy**. 2021. 127p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

Em um cenário de intenso tráfego de dados, as redes ópticas elásticas precisam dispor de mecanismos que gerenciem falhas. Um mecanismo eficiente é o esquema de proteção compartilhada de caminho de *backup* que realiza a pré-reserva de recursos e os compartilham com as conexões que transportam dados nos caminhos de trabalho. O compartilhamento de recursos expressa uma relação de compromisso entre a economia de espectro e a capacidade de restauração de conexões afetadas por falhas. Quanto maior for o número de compartilhamentos, maior será a economia de espectro. No entanto, menor será a capacidade de restauração das conexões interrompidas por falhas por causa da maior competição por recursos de proteção. Várias heurísticas foram propostas para alocar recursos compartilhados eficientemente, mas nenhuma delas tratou o problema de otimização na restauração de falhas. Este trabalho propõe combinar uma nova métrica de fragmentação espectral como função das entradas de um sistema de inferência *fuzzy* (FIS, *Fuzzy Inference System*) para o treinamento de uma rede neural profunda (DNN, *Deep Neural Network*). Os objetivos são resolver o problema RSA compartilhado para obter maior eficiência na proteção dos caminhos de trabalho em um cenário de múltiplas falhas. Os algoritmos foram avaliados em termos de probabilidade de bloqueio, taxa de restauração de falhas, capacidade ociosa da rede e tempo de inatividade das conexões, dentre outras métricas auxiliares. O algoritmo utilizando somente FIS reduziu a probabilidade de bloqueio em 25,85%. O algoritmo DNN treinado com FIS aumentou a taxa de restauração de falha em 2,88% e reduziu a probabilidade de bloqueio em 10,63%.

Palavras-chave: rede óptica elástica, algoritmo de roteamento e alocação de espectro, proteção de caminho de *backup* compartilhado, sobrevivência da rede, sistema de inferência *fuzzy*, rede neural profunda.

ABSTRACT

LOURENCO, A.L.F. **Routing, spectrum allocation and shared protection in elastic optical networks using deep learning with a fuzzy inference system.** 2021. 127p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

In a traffic-intensive data scenario, elastic optical networks must have mechanisms that manage failures. An efficient scheme is the shared backup path protection that reserves resources and shares them with the connections that carry the main data stream. Resource sharing expresses a trade-off between spectrum savings and the ability to restore connections affected by failures. The higher the number of sharing, the greater the spectrum savings. However, the less the capacity to restore connections interrupted by failures, due to the fierce competition for obtaining protection. Heuristics were proposed to allocate shared resources efficiently, but none of them addressed the problem of optimization in restoring failures. This work proposes combining a new fragmentation metric as a function of the inputs of a fuzzy inference system (FIS), then used to train a deep neural network (DNN) to solve the RSA problem to obtain greater efficiency in the protection of working paths in a multi-failure scenario. The algorithms were evaluated in terms of blocking probability, failure restoration rate, idle network capacity, and options downtime, among other auxiliary metrics. The algorithm that only used FIS reduced the blocking probability by 25.85%. When trained with FIS, the DNN algorithm increased the fault restoration rate by 2.88% while lowering the blocking probability by 10.63%.

Keywords: elastic optical network, routing and spectrum allocation algorithm, backup path shared protection, network survivability, fuzzy inference system, deep neural network.

LISTA DE FIGURAS

Figura 1 – Esquemas de sobrevivência (proteção e restauração) de rede. UPSR: unidirectional path switched ring, BLSR: bidirectional line switched ring, SBPP: shared backup path protection, PR: path restoration [17].	46
Figura 2 – Rede <i>perceptron</i> multicamadas contendo camada de entrada (<i>input layer</i>), camada escondida (<i>hidden layer</i>) e camada de saída (<i>output layer</i>).	54
Figura 3 – Rede de Hopfield. Os neurônios da camada de saída realimentam a camada de entrada.	55
Figura 4 – Modelo neuro- <i>fuzzy</i> 1 [109]. As saídas da lógica <i>fuzzy</i> servem de entradas para as redes neurais.	56
Figura 5 – Modelo neuro- <i>fuzzy</i> 2 [109]. As saídas da rede neural servem de entradas para a lógica <i>fuzzy</i>	56
Figura 6 – Exemplos de sub-blocos formados a partir de um bloco contínuo de 5 FSs e uma demanda de 2 FSs.	60
Figura 7 – Estrutura de dados de uma EON e o processamento de duas solicitações de conexões (CRs) que chegam à rede. O algoritmo aloca às CRs os caminhos de trabalho $W_1[(v_1, v_2, v_3, v_4), (f_1, f_2)]$ e $W_2[(v_1, v_5, v_6, v_4), (f_9, f_{10})]$, e os caminhos de <i>backup</i> $B_{1,2}[(v_1, v_7, v_8, v_4), (f_9, f_{10})]$. O conjunto de <i>slots</i> de frequências satisfaz as restrições de continuidade e contiguidade de espectro.	62
Figura 8 – Exemplo do esquema de fragmentação WCAFr de dois caminhos de alocação formado pelos enlaces $\{l_1, l_2, l_3\}$. Os <i>slots</i> de frequência alocados para o Caminho nº 1 e Caminho nº 2 são $\{f_1, f_2\}$ e $\{f_9, f_{10}\}$, respectivamente. Os valores de ex_1 and ex_2 indicam a quantidade de <i>slots</i> livres de frequência do lado esquerdo e direito, respectivamente dos slots de frequência alocados nos enlaces. Portanto, $ex_1 = \{0, 2, 1\}$ e $ex_2 = \{0, 0, 2\}$ para o caminho nº 1 e $ex_1 = \{2, 0, 0\}$, e $ex_2 = \{0, 0, 0\}$ para o caminho nº 2.	64
Figura 9 – Funções de pertinência do conjunto de entradas do FIS 1 [a,b,c] e FIS 2 [a,b,c,d].	67
Figura 10 – Grau de pertinência de saída.	67
Figura 11 – Avaliação de custo de um sub-bloco pelo sistema de inferência <i>fuzzy</i> . . .	70
Figura 12 – Modelo de rede neural profunda para avaliação dos caminhos de alocação com proteção compartilhada.	72
Figura 13 – Fluxograma do algoritmo para obtenção do conjunto de treinamento da DNN.	74

Figura 14 – Progresso do treinamento da rede neural profunda usando Matlab. Cor azul: RMSE (<i>Root-Mean-Square Error</i>); e cor vermelha: <i>Loss Function</i> .	76
Figura 15 – Topologia da rede NSFNet composta por 14 nós e 21 enlaces bidirecionais.	81
Figura 16 – Probabilidade de bloqueio em função da carga de tráfego para métricas de fragmentação de espectro: MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].	87
Figura 17 – Logarítmo da probabilidade de bloqueio em função da carga de tráfego para métricas de fragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].	87
Figura 18 – Relação de restauração de conexão em função da carga de tráfego para métricas de desfragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].	88
Figura 19 – Taxa de utilização de espectro em função da carga de tráfego para métricas de desfragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].	88
Figura 20 – Probabilidade de bloqueio dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN.	90
Figura 21 – Logarítmo da probabilidade de bloqueio dos algoritmos: MFK FASA [64], MFK WCAFr , FIS e DNN.	90
Figura 22 – Taxa de utilização do espectro dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN.	91
Figura 23 – Taxa de restauração de conexão dos algoritmos: MFK FASA [64], MFK WCAFr , FIS e DNN.	92
Figura 24 – Relação de compromisso CRR(1-BP) dos algoritmos : MFK FASA [64], MFK WCAFr , FIS e DNN.	92
Figura 25 – Média de hops dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN	93
Figura 26 – Número médio de blocos livres dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN	94
Figura 27 – Tamanho médio de blocos livres dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN	95
Figura 28 – Capacidade ociosa da rede produzida pelos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN	95
Figura 29 – Tempo de inatividade de conexões produzida pelos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN	97
Figura A.1–Lógica clássica vs lógica <i>fuzzy</i> aplicadas ao conceito de altura [125].	116
Figura A.2–Diagrama de um sistema <i>fuzzy</i> contendo os blocos funcionais: <i>fuzzifier</i> , <i>fuzzy knowledge base</i> , <i>inference engine</i> e <i>defuzzifier</i> .	119

Figura B.1 – Representação do neurônio. As entradas X_n são ponderadas pelos pesos sinápticos W_n subtraído pelo limiar de ativação θ e aplicados a uma função de ativação f 121

LISTA DE TABELAS

Tabela 1 – Valores ex_1 e ex_2 dos caminhos n.º 1 e n.º 2 para computação da métrica WCAFr	64
Tabela 2 – Funções de pertinência de entrada do FIS 1 para computação dos caminhos de trabalho	66
Tabela 3 – Funções de pertinência de entrada do FIS 2 para computação dos caminhos de proteção compartilhada	66
Tabela 4 – Conjunto de regras automáticas para o FIS 1	68
Tabela 5 – Conjunto de regras automáticas para o FIS 2	69
Tabela 6 – Definições das variáveis	69
Tabela 7 – Comprimento máximo de 5-menores caminhos de algumas topologias de redes	75
Tabela 8 – Comparação dos tempos de execução computacional das métricas de fragmentação espectral analisadas nesse trabalho.	79
Tabela 9 – Comparação dos tempos de execução computacional das métricas de fragmentação espectral analisadas neste trabalho (continuação).	79
Tabela 10 – Parâmetros para simulações computacionais.	83
Tabela 11 – BP e CRR médios e variação relativa (VR)	96
Tabela 12 – Inclinação do ajuste linear das curvas de <i>downtime</i>	98
Tabela C.1 – Conjunto de regras automáticas para o FIS 1	123
Tabela C.2 – Conjunto de regras automáticas para o FIS 2	124

LISTA DE ABREVIATURAS E SIGLAS

BP	blocking probability
BPSK	binary phase shift keying
BSUR	backup spectrum utilization ratio
BV-T	bandwidth variable transponder
BV-WXC	bandwidth variable wavelength cross-connector
BV-WSS	bandwidth variable wavelength selective switch
CRR	connection restoration ratio
DPP	dedicated path protection
DNN	deep neural network
DTF	discrete Fourier transform
EON	elastic optical network
GFP	generic framing procedure
IA	inteligência artificial
IoT	internet of things
ILP	integer linear programming
ITU-T	International Telecommunication Union-Telecommunication Standardi- zation Sector
FA	fragmentation and alignment
FASA	fragmentation of spectrum allocation
FIS	fuzzy inference systems
FR	fragmentation ratio
FS	frequency slot
FWM	four wave mixing
H	fragmentation entropy

HOPS	hitless optical path shift
LCoS	liquid crystal on silicon
MBBR	make before break rerouting
MC	multi-carrier
MDT	mean downtime
MNFB	mean number of free blocks
MF	membership function
MFBS	mean free blocksize
MTTF	mean time to failure
MEMS	microelectro-mechanical systems
MFK	minimum fragmentation of k-paths
MILP	mixed-integer linear programming
NFP	neuro-fuzzy protection
NP-hard	non-deterministic polynomial-time hard
OADM	optical add-drop multiplexer
OFDM	orthogonal frequency-division multiplexing
OSNR	optical signal-to-noise ratio
QoS	quality of service
QoT	quality of transmission
ReLU	rectified linear unit
RMSE	root-mean-square error
RSA	routing and spectrum allocation
SBPP	shared backup path protection
SC	single-carrier
SCFDM	single-carrier frequency division multiplexing
SDH	synchronous digital hierarchy

SDM	spatial division multiplexing
SGDM	stochastic gradient descent with momentum
SLA	service level agreement
SLICE	spectrum-sliced elastic optical path network
S-BVT	sliceable bandwidth variable transceiver
SONET	synchronous optical network
TG	traffic grooming
TON	transparent optical networks
XPM	cross phase modulation
WCAFr	weighted cutting avoid fragmentation
WDM	wavelength division multiplexing
WUFR	wasted-unusable-free ratio

SUMÁRIO

1	INTRODUÇÃO	27
1.1	Objetivos	30
1.2	Organização da tese de doutorado	31
2	ESTADO DA ARTE	33
2.1	Conceitos Básicos em Redes Ópticas Elásticas	33
2.2	Alguns Algoritmos RSA para Redes Ópticas Elásticas	36
2.2.1	<i>Distance-adaptive</i> RSA (DA-RSA)	36
2.2.2	<i>Fragmentation-aware</i> RSA	38
2.2.3	<i>Traffic Grooming</i> RSA (TG-RSA)	41
2.2.4	<i>Survivability</i> RSA	45
2.2.4.1	Proteção	46
2.2.4.2	Restauração	49
2.3	Considerações Finais	49
3	SISTEMA DE INFERÊNCIA <i>FUZZY</i> E REDE NEURAL ARTIFICIAL PARA SOLUÇÃO DO PROBLEMA DE ROTEAMENTO E ALOCAÇÃO DE ESPECTRO	51
3.1	Sistemas <i>fuzzy</i> aplicados em Problemas RSA	51
3.2	Redes Neurais Artificiais para Problemas RSA	53
3.2.1	Sistemas Neuro-Fuzzy	55
4	MODELOS DE SISTEMA DE INFERÊNCIA <i>FUZZY</i> E DE REDE NEURAL PROFUNDA PARA O PROBLEMA RSA COM PROTEÇÃO COMPARTILHADA	59
4.1	Algoritmo de Roteamento e Alocação de Espectro com Proteção Compartilhada (RSA-SBPP)	59
4.2	Nova Métrica de Fragmentação do Espectro	63
4.3	Modelo <i>Fuzzy</i>-RSA Multiobjetivo para Proteção Compartilhada	66
4.4	Rede Neural Profunda	71
4.4.1	Camada Convolutacional 2D	71
4.4.2	Construção do Conjunto de Treinamento, Validação e Teste	73
4.4.3	Treinamento e operação da DNN para computação RSA de proteção compartilhada	75
4.5	Complexidade computacional	77
4.5.1	Algoritmo para resolver os caminhos de alocação	77

4.5.2	Métricas	78
4.5.3	Rede neural	79
4.5.4	Sistema de inferência <i>fuzzy</i>	80
5	RESULTADOS COMPUTACIONAIS	81
5.1	Topologia de Rede, Geração de Tráfego e de Falha, e Processamento de Conexão	81
5.2	Análise dos Resultados	83
5.2.1	Comparação das Métricas de Fragmentação Espectral	86
5.2.2	Análise dos algoritmos FIS e DNN	89
6	CONCLUSÕES	99
	REFERÊNCIAS	101
	APÊNDICES	113
	APÊNDICE A – NOÇÕES DE LÓGICA <i>FUZZY</i>	115
	APÊNDICE B – NOÇÕES DE REDES NEURAIS	121
	APÊNDICE C – REGRAS <i>FUZZY</i>	123
	APÊNDICE D – MÉTRICAS DE FRAGMENTAÇÃO DE ESPECTRO	125
D.1	Fragmentation Ratio (FR) [58,60]	125
D.2	Fragmentation Entropy (H) [56,57]	125
D.3	Fragmentation and Alignment (FA) [120]	125
D.4	Fragmentation-aware spectrum allocation (FASA) [64]	126
D.5	Wasted-Unusable-Free Ratio (WUFR) [121]	126

1 INTRODUÇÃO

Durante a última década, a rede óptica elástica (*elastic optical network*, EON) tornou-se uma arquitetura de rede competitiva, habilitando-se para ser a principal candidata para a próxima geração de redes ópticas. A EON processa tráfego de alta capacidade e apresenta elevada eficiência no uso do espectro óptico, além de flexibilidade e escalabilidade [1–7]. Atualmente, a EON passa pelo processo de padronização e estudos de virtualização na camada de *hardware* [8]. O cenário de atuação da EON está configurado em função de demanda de tráfego de serviços como computação em nuvem, transferência de dados entre *data centers*, teleconferências, redes sociais, e compartilhamento de *streaming* de vídeo de alta resolução. A integração com redes sem fio operando na quinta geração (5G) e com a internet das coisas (*internet of things*, IoT), que são tecnologias recentes, impactará ainda mais a rede *backbone* e trará novos desafios. Essa mudança tecnológica vem ocorrendo devido ao aumento crescente por demanda de largura de banda [9].

Se a demanda de largura de faixa de frequências for muito grande ela pode tornar-se intratável por um único canal óptico. Nessa situação, uma solução é construir um conjunto de subportadoras ópticas muito próximas com dados modulados individualmente em taxas mais baixas. Este conjunto de subportadoras é conhecido como supercanal e é tratado pelos dispositivos da rede como entidade única. O supercanal é demultiplexado no receptor. As técnicas de modulação que permitem a composição do espectro do supercanal a partir de subportadoras estão agrupadas em duas categorias: portadora única (*single carrier*, SC) e múltiplas portadoras (*multi-carrier*, MC). Algumas dessas técnicas são a MC *discrete Fourier transform (DFT)-based* utilizando multiplexação por divisão ortogonal de frequência (*orthogonal frequency-division multiplexing*, OFDM), *filter-bank-based* MC, SC *frequency division multiplexing (SCFDM)* ou *DFT-spread OFDM* e *Nyquist pulse-shaped* SC [10–13].

A EON atribui um conjunto de *slots* de frequência (*frequency slot*, FS) estritamente necessário para atender à demanda de largura de banda de uma conexão. Portanto, o espectro é mais bem gerenciado em comparação com a técnica de multiplexação por divisão de comprimento de onda (*wavelength division multiplexing*, WDM). Na arquitetura EON, o espectro óptico é dividido em *slots* de largura fixa e grupos de *slots* contíguos são alocados às conexões de acordo com suas demandas de largura de banda.

O algoritmo de roteamento e atribuição de espectro (*routing and spectrum allocation*, RSA) desempenha papel decisivo no planejamento e operação da EON, pois gerencia os recursos considerando objetivos, restrições, arquitetura de nós e topologia de rede. A partir de uma topologia de rede óptica e um conjunto de demandas de tráfego, o algoritmo RSA encontra uma rota e atribui espectro para cada demanda atendendo a duas restrições:

contiguidade de FS e continuidade de FS ao longo do caminho óptico. O problema RSA se enquadra na classe de problema não determinístico em tempo polinomial (NP-difícil) [14]. Além do mais, é um problema multiobjetivo e lida com várias restrições e topologias de rede extensas [15]. Por isso, soluções heurísticas e meta-heurísticas têm sido utilizadas para resolvê-los [16].

Além da característica principal, que é alocar *slots* de frequência estritamente necessários para atender à demanda, a EON permite a utilização da técnica de distância adaptativa, em que o tipo de modulação digital é escolhido para permitir a transmissão da maior taxa de bits possível, e a agregação de tráfego na camada óptica. Por outro lado, a alocação dinâmica de *slots* de frequência pode produzir fragmentação espectral, em que há *slots* vagos entre *slots* alocados, aumentando a probabilidade de bloqueio de novas demandas e reduzindo a eficiência espectral.

A arquitetura EON pode gerenciar tráfego agregado de centenas de Gbps a dezenas de Tbps e espera-se que as redes de próxima geração operem com acordos de nível de serviço (*service level agreement*, SLA) ainda mais rigorosos, o que inclui o desafio de garantir a sobrevivência da rede [17]. Nesse cenário, as técnicas de proteção e restauração de tráfego desempenharão um papel importante porque uma falha em um dispositivo ou enlace acarreta perda de grande quantidade de dados.

Basicamente, as estratégias de gerenciamento de falhas são agrupadas em esquemas de proteção e esquemas de restauração [18–20]. Essencialmente, os esquemas de proteção incluem enlace e caminho (conjunto de enlaces conectando o nó de origem e o nó de destino), enquanto os esquemas de restauração incluem subcaminho, caminho e enlace. O caminho alocado para o tráfego em condições operacionais sem falha é o caminho de trabalho (primário) e o caminho para o qual o tráfego é redirecionado é o caminho de *backup* (secundário). Um esquema de proteção reserva previamente um caminho de *backup*, enquanto um esquema de restauração busca os recursos disponíveis para restaurar o tráfego após a falha.

Uma das técnicas mais versáteis e eficientes de proteção de rede é a bem conhecida proteção de caminho de *backup* compartilhado (*shared backup path protection*, SBPP). O esquema SBPP atribui um caminho de trabalho a uma conexão e, em seguida, reserva um caminho de *backup* disjunto para essa conexão. O caminho de *backup* é compartilhado com outras conexões, economizando recursos de rede. O esquema SBPP tem sido amplamente analisado em redes WDM [18–22] e tem recebido muita atenção com o advento da EON, principalmente por causa de sua característica de operar com elevada taxa de bits e por causa das consequências prejudiciais da perda de tráfego em caso de falha [23–33]. Embora o uso de compartilhamento de recursos no caminho de *backup* resulte em economia de espectro, a ocorrência de compartilhamento excessivo reduz a capacidade do esquema SBPP de restaurar a conexão de tráfego em um cenário de falha múltipla porque os vários

caminhos competem por recursos de proteção. Esta desvantagem pode ser superada com algoritmos SBPP-RSA projetados para ajustar/controlar o nível de compartilhamento de recursos para melhorar o *trade-off* entre os níveis de compartilhamento e a capacidade de restaurar conexões, conforme abordado em [34].

Desse cenário surge a necessidade de esquemas de proteção e restauração de tráfego não apenas eficientes em economizar recursos, mas também eficientes em agir rapidamente em presença de falhas. Como a solução para o problema de alcançar o *trade-off* ideal entre níveis de compartilhamento e a capacidade de restaurar conexões deve ser obtida a partir da análise de vários objetivos e restrições, a melhor forma de abstrair esse relacionamento, que não é trivial, é utilizar técnicas que fazem parte do escopo da inteligência artificial (IA), como redes neurais profundas (*deep neural network*, DNN) e sistemas de inferência difusa (*fuzzy inference systems*, FIS). DNN e IA foram abordados em redes para prever tráfego [35], resolver o problema RSA [36], otimizar qualidade de serviço (*quality of service*, QoS) [37], e estimar a relação sinal ruído óptico (*optical signal-to-noise ratio*, OSNR) em sistemas de monitoramento de falhas [38].

Uma rede neural artificial é um modelo de computação de entrada e saída que utiliza mecanismos semelhantes aos neurônios do cérebro humano. Ela é composta por camadas: de entrada, escondidas e de saída. As camadas escondidas são responsáveis por processar os dados de entrada para determinar o resultado na camada de saída. O termo “profunda” (*deep*) refere-se à quantidade de camadas que compõem a rede neural artificial. Embora não haja consenso a respeito da quantidade, uma rede com mais de três camadas, que pode incluir as camadas de entrada e saída, pode ser dita “profunda” [39]. Por outro lado, “profunda” pode referir-se também à quantidade de camadas escondidas. Neste caso, “profunda” significa quantidade maior que duas camadas escondidas [40]. A rede neural utilizada neste trabalho é formada por quatro camadas escondidas, além das camadas de entrada e saída. A rede neural artificial forma a base de técnicas para o aprendizado de máquinas, uma subárea de conhecimentos da inteligência artificial.

A lógica *fuzzy* (nebulosa, difusa) é baseada na maneira de expressar a relação de pertinência multivalorada de elementos a conjuntos. Diferentemente da lógica binária, na qual um elemento pertence ou não a um conjunto, na lógica *fuzzy* um elemento pode exibir graus intermediários de pertinência. A lógica *fuzzy* lida com informação vaga e imprecisa. A utilização da lógica *fuzzy* é vantajosa em modelagens em que o projetista conhece a dinâmica do processo, porém não tem certeza sobre as informações obtidas, enquanto as redes neurais artificiais são ideais para obter o conhecimento (ou comportamento) implícito por meio do aprendizado de um conjunto conhecido de dados. A integração dessas duas técnicas, denominada sistema neuro-*fuzzy* (SNF), forma um modelo híbrido que une as suas vantagens. De um lado, a característica de aprendizagem da rede neural artificial; de outro lado, o desempenho do modelo de inferência *fuzzy* [41].

Os sistemas *fuzzy* e as redes neurais têm sido amplamente utilizados em várias áreas da ciência. Entretanto, a pesquisa bibliográfica realizada para esta tese revelou a existência de poucas abordagens ao problema de roteamento e alocação de espectro em redes ópticas elásticas. Abordagens mais antigas apontaram para soluções para redes WDM que poderiam ser adaptadas para rede EON. No entanto, tais adaptações precisam incluir as características da EON, como fragmentação e restrições de continuidade e contiguidade de *slots* de frequências. Além do mais, até onde foi possível saber, nenhum dos trabalhos pesquisados utilizaram técnicas neuro-*fuzzy* para resolver o problema de sobrevivência de rede envolvendo proteção compartilhada em EON.

O esquema de proteção compartilhada é projetado para atuar em um cenário de uma única falha, pois as restrições impostas consideram caminhos disjuntos entre os caminhos de trabalhos do espectro compartilhado. Quando ocorre múltiplas falhas simultâneas nos enlaces, essa restrição é violada e a restauração da conexão afetada pela falha tem seu desempenho comprometido devido à concorrência pelo espectro compartilhado. Sendo assim, o nível de comprometimento é proporcional ao nível de compartilhamentos efetuados. Nesse sentido, temos uma relação de compromisso entre o número de compartilhamentos e a restauração desejada cujo algoritmo original não provê nenhum controle [42]. Nessa tese utilizamos as técnicas de redes neurais e lógica *fuzzy* para otimizar essa relação de compromisso para adaptar melhor o esquema SBPP a um cenário de múltiplas falhas.

1.1 Objetivos

Nessa tese de doutorado, propomos um modelo baseado em rede neural profunda com sistema de inferência difusa multiobjetivo para resolver problemas de RSA com SBPP em EON visando à otimização do nível de compartilhamento e da taxa de restauração de conexões. A entrada do modelo DNN é o conjunto de todo espectro ao longo do maior comprimento dos k -menores caminhos da rede e a saída é um valor de custo que avalia o caminho de alocação. O método de treinamento é baseado na avaliação de cada caminho de alocação computador por um sistema *fuzzy* Takagi-Sugeno que combina três entradas para cálculo do caminho de trabalho e quatro entradas para cálculo do caminho de proteção compartilhada. As entradas do sistema *fuzzy* são coeficientes numéricos normalizados calculados pelos critérios: fragmentação, primeiro índice de *slot* livre, disponibilidade de *slot* no enlace, e nível de compartilhamento de recursos de proteção. A saída é um valor de custo normalizado entre $[0, 1]$ de forma que quanto menor o custo, melhor é a solução.

Há três contribuições principais nessa tese de doutorado: (1) um modelo de *fuzzy inference system* (FIS) com conjunto reduzido de regras e um modelo de *deep neural network* (DNN) com número reduzido de entradas para resolver o problema RSA com esquema SBPP. Os dois modelos exibem menor complexidade computacional devido à redução da dimensão do problema; (2) uma nova figura de mérito para avaliar o grau

de fragmentação do espectro para melhorar a probabilidade de bloqueio de solicitações de conexão; e (3) uma otimização entre o nível de compartilhamento e a capacidade de restauração de conexões afetadas por falhas para o esquema SBPP em cenário de múltiplas falhas.

O desempenho do algoritmo proposto é comparado com os resultados apresentados na literatura. As principais figuras de mérito utilizadas para avaliar o desempenho do algoritmo e compará-lo com alguns algoritmos semelhantes relacionados na literatura são a probabilidade de bloqueio de solicitações de conexão, a taxa de utilização do espectro de *backup* e a taxa de restauração de conexão. O algoritmo proposto organiza o tráfego de forma menos fragmentada enquanto utiliza os recursos de proteção de forma eficiente.

1.2 Organização da tese de doutorado

A tese de doutorado está organizada da seguinte maneira. O Capítulo 2 apresenta os conceitos básicos relacionados às redes elásticas e os principais algoritmos estudados. O Capítulo 3 apresenta os conceitos de lógica *fuzzy* e redes neurais artificiais e suas aplicações em problemas RSA. No Capítulo 4, propomos uma nova métrica de fragmentação, um modelo de sistema *fuzzy* e de rede neural profunda para RSA com proteção compartilhada. Os resultados obtidos por meio de simulações computacionais são apresentados no Capítulo 5, e as conclusões são reunidas no Capítulo 6.

2 ESTADO DA ARTE

Neste capítulo, apresentamos os conceitos básicos de uma rede óptica elástica (EON), incluindo distância adaptativa, fragmentação/desfragmentação, agregação de tráfego e sobrevivência de rede. Descrevemos sucintamente alguns algoritmos de roteamento e alocação de espectro, de proteção e de restauração de tráfego apresentados na literatura.

2.1 Conceitos Básicos em Redes Ópticas Elásticas

A rede óptica elástica (*elastic optical network*, EON) é uma arquitetura de rede óptica baseada no conceito de espectro elástico, capaz de promover maior escalabilidade e eficiência espectral do que as tradicionais redes de multiplexação por divisão de comprimento de onda (*wavelength division multiplexing*, WDM). Uma rede óptica funciona como uma rede de comutação de circuitos em que os recursos precisam ser reservados e mantidos durante o período de atendimento à conexão do usuário. Na rede WDM, o recurso utilizado é o comprimento de onda e ela opera a partir de grade fixa de frequências. Uma demanda de conexão é atendida com sucesso quando o mesmo comprimento de onda é atribuído a todos os enlaces da rota. Nesse tipo de rede, a conexão ocupa um comprimento de onda inteiro mesmo que o espectro do sinal modulado seja menor do que a largura de banda da grade. Na EON, o recurso é um espectro de largura variável alocado de acordo com a largura de banda requerida por uma demanda de conexão. As técnicas de geração de multiportadoras como o OFDM óptico permitem gerar subportadoras de granularidade fina e cujas larguras espectrais se sobrepõem parcialmente formando uma envoltória. O espectro de largura variável é obtido por uma quantidade de subportadoras adjacentes que são selecionadas pelo plano de controle e, dessa forma, a envoltória pode ter a sua largura espectral expandida ou comprimida [4]. A arquitetura EON apresenta as seguintes vantagens em relação à WDM [2]:

1. **Resiliência em relação aos efeitos físicos não lineares de degradação de sinal.** Como cada subportadora opera em taxa de bits baixa, a interferência entre símbolos é reduzida e os efeitos de dispersão cromática (*chromatic dispersion*, CD) e de dispersão de modo de polarização (*polarization mode dispersion*, PMD) da fibra são reduzidos;
2. **Eficiência espectral.** Há duas características das técnicas de multiplexação que contribuem para o uso altamente eficiente dos recursos espectrais. Em uma delas, as subportadoras adjacentes podem se sobrepor no espectro devido à ortogonalidade. Na outra, a melhor acomodação espectral com o tamanho da demanda elimina o

desperdício de espectro que ocorre no modelo de grade fixa de frequências, resultando em melhor utilização do espectro;

3. **Taxas de dados elásticas.** O número de subportadoras alocadas e o formato de modulação podem ser ajustados dinamicamente via plano de controle, possibilitando a expansão e contração do espectro. Essa característica elástica do espectro se adapta melhor às demandas de bandas heterogêneas e variáveis no tempo, permitindo suportar taxas de dados de Gbps até Tbps;
4. **Escalabilidade.** A capacidade elástica do espectro aumenta a capacidade da rede de atender maior número de demandas e/ou com maiores taxas de bits sem que haja necessidade de mudar a estrutura física da rede.

A arquitetura de nós de uma EON consiste em dois componentes básicos: o transceptor de banda variável (*bandwidth variable transceiver*, BV-T) e o comutador de comprimento de onda (*wavelength cross-connects*, WXC). A tecnologia WSS (*wavelength-selective switch*) de largura de banda variável foi incorporada ao WXC básico, compondo o BV-WSS (*bandwidth variable wavelength selective switch*). Este componente permite que a EON atribua recursos de forma flexível. Quando uma demanda de conexão chega à EON, o BV-T gera um sinal óptico com banda suficiente para que a conexão seja atendida. Ao longo da rota designada para o tráfego, todos os BV-WSSs alocam banda suficiente à conexão. Assim, o BV-WSS permite comutar (chavear) fatias do espectro de tamanhos diferentes, roteando os respectivos tráfegos para diversos destinos. O BV-WSS realiza as funções de multiplexação/demultiplexação e comutação usando dispositivos em tecnologia óptica integrada, como *microelectro-mechanical systems* (MEMS) ou *liquid crystal on silicon* (LCoS) [2]. Outra tecnologia que integra a EON é o transceptor divisível de banda variável (*sliceable bandwidth variable transceiver*, S-BVT). Esse transceptor pode ser dividido em transceptores virtuais para atender rotas/bandas distintas. Assim, evita-se ainda mais o desperdício de banda, pois demandas de taxas de bits baixas não ocupam toda a banda de um transceptor de banda fixa [4]

O espectro gerenciado por uma EON é dividido em *frequency slots* (FS). Uma EON atribui às demandas de tráfego a quantidade necessária de FSs contíguos, permitindo a formação de subcanais ou até de supercanais ópticos. O padrão do *International Telecommunication Union-Telecommunication Standardization Sector* (ITU-T) define uma grade flexível centrada em 193,1 GHz cujas frequências centrais das subportadoras estão espaçadas de 6,25 GHz e com 12,5 GHz de largura espectral. Existem duas abordagens de FS no espectro flexível: (a) *single-slot*, na qual o FS é definido pela largura espectral da subportadoras; (b) *double-sided half slot*, na qual o FS é definido pelo espaçamento entre as frequências centrais.

O problema de encontrar um caminho na rede e reservar recursos espectrais em uma EON é chamado de roteamento e alocação de espectro (*routing and spectrum allocation*, RSA) e é um problema NP-hard, como provado em [43]. Nas duas abordagens, *single-slot* e *double-sided half slot*, o algoritmo que resolve o problema RSA executa a alocação do espectro em função do conjunto de FSs disponíveis na rede, respeitando as restrições impostas e otimizando uma ou mais funções objetivo. Existem duas restrições no RSA relacionadas ao FS: a continuidade e a contiguidade de espectro. De acordo com a continuidade, um FS tem que estar livre em todos os enlaces da rota. De acordo com a restrição de contiguidade, os FSs disponíveis devem ser contíguos em todos os enlaces da rota. A contiguidade do espectro é a diferença mais importante entre as redes EON e WDM e aumenta a complexidade das abordagens RSA existentes. Nas abordagens de algoritmos RSA, geralmente essas restrições são impositivas considerando que nos nós da rede não há dispositivos que realizam a conversão e a divisão da demanda. Os algoritmos RSA geralmente são caracterizados por sua função objetivo, dependendo dos parâmetros. Alguns algoritmos RSA objetivam diminuir a probabilidade de bloqueio, enquanto outros visam a melhorar o uso espectral ou a eficiência energética. No entanto, o objetivo comum é o uso eficiente dos recursos de acordo com as demandas de tráfego. Os algoritmos RSA ainda podem ser classificados em duas categorias: (1) os que suportam tráfego estático e (2) os que suportam tráfego dinâmico.

No cenário de tráfego estático o problema RSA é tratado de modo *offline*. As entradas consistem em um conjunto de demandas de tráfego previamente conhecido e o objetivo é atribuir um caminho óptico contínuo a cada demanda de modo a minimizar a quantidade total de FSs. O RSA *offline* é utilizado quando os padrões de tráfego na rede são conhecidos antecipadamente ou quando as conexões permanecem ativas por um longo período.

Sob um cenário de tráfego dinâmico, os usuários enviam as solicitações para que os caminhos ópticos sejam configurados conforme a necessidade. Assim, os pedidos de conexão são iniciados de forma aleatória e, dependendo do estado da rede no momento de uma solicitação, os recursos espectrais disponíveis podem ou não ser suficientes para estabelecer uma conexão. O estado da rede consiste em caminho físico (rota) e atribuição do espectro para todas as conexões ativas. Conforme novas conexões são admitidas e as conexões atendidas são liberadas, não é possível prever o estado da rede em determinado tempo de funcionamento, já que o seu estado é constantemente alterado pelos eventos de tráfego aleatório. Assim, cada vez que uma solicitação é feita, um algoritmo RSA deve ser executado em tempo real para determinar se é possível estabelecer a conexão. Adota-se que se um pedido de conexão não puder ser aceito por falta de recursos ele será bloqueado. Portanto, a probabilidade de bloqueio dos pedidos de conexão é a principal métrica de desempenho em um cenário dinâmico. Os algoritmos RSA *online* devem ser simples e rápidos e, por isso, algoritmos heurísticos são uma escolha apropriada. Ademais,

o RSA *online* pode ser classificado em duas categorias: solução conjunta e solução em duas etapas. A solução conjunta, ou *joint*-RSA, tenta resolver o roteamento e a alocação de espectro em um único passo. Na solução disjunta, *disjoint*-RSA, o problema é dividido em dois subproblemas: (1) roteamento e (2) alocação de espectro (R-SA). O algoritmo R-SA aborda primeiro o subproblema de roteamento e depois considera a alocação de espectro. A abordagem do RSA em duas etapas (R-SA) é preferível no processo de restauração dinâmica de uma falha de conexão já que as restrições de tempo são ainda mais rígidas.

2.2 Alguns Algoritmos RSA para Redes Ópticas Elásticas

2.2.1 *Distance-adaptive* RSA (DA-RSA)

O problema RSA adaptativo à distância (*distance-adaptive*, DA), primeiramente abordado em [44], é uma extensão do RSA com requisitos adicionais para selecionar um formato de modulação apropriado e largura do espectro de acordo com a distância de transmissão. A distância de transmissão que um sinal óptico com um formato de modulação específico pode alcançar é determinada com base nos parâmetros de transmissão de um caminho óptico, incluindo a distância de propagação do sinal que contabiliza a degradação do *optical signal noise ratio* (OSNR) ao longo da fibra, o número de *hops* (saltos) na rede e o tipo do filtro óptico [45]. Nas redes WDM, o formato de modulação é selecionado na fase de configuração e permanece durante todo o funcionamento da rede. Como a largura espectral máxima é definida pela largura da grade fixa de frequências, o formato de modulação é determinado de modo a garantir que o sinal seja transmitido com qualidade (*quality of transmission*, QoT) ao longo do pior caminho, que é, geralmente, o caminho mais longo ou com maior número de hops na rede. A configuração do formato de modulação para o pior caso leva à baixa utilização do espectro.

Por outro lado, em uma EON, os BV-Ts são capazes de selecionar dinamicamente o formato de modulação e modular subportadoras individualmente [46]. Assim, é possível ajustar o formato de modulação e/ou o número de bits por símbolo para considerar as deficiências do enlace (como a degradação OSNR) para que às demandas com a mesma taxa de dados sejam atribuídas diferentes quantidades de recursos espectrais, dependendo da qualidade de seu caminho. A alocação do espectro adaptativo à distância aproveita a relação de compromisso que existe entre a largura do espectro modulado e o alcance máximo, para a mesma taxa de dados e para um QoT aceitável. Assim, um formato de modulação de alto nível que produz um espectro estreito e com baixa tolerância a ruídos pode ser selecionado para um caminho curto, enquanto que uma modulação de baixo nível com um espectro mais amplo e alta tolerância a ruídos pode ser usada para um caminho mais longo [47]. Por exemplo, para a mesma taxa de dados, o formato 16-QAM é composto por duas vezes o número de bits por símbolo do que o formato QPSK e, conseqüentemente, requer metade da largura de banda do espectro, enquanto a tolerância OSNR é menor do

que a do QPSK, o que significa menor alcance de distância de transmissão.

Os algoritmos DA-RSA foram propostos em versões *offline* e *online*. A versão *offline* formulada em ILP denominado *routing, modulation level, and spectrum allocation* (RMLSA) [48] considera que cada demanda é mapeada para um nível de modulação com base na taxa de dados solicitada e na distância do caminho sobre o qual é encaminhada. Em [49] foi proposta uma versão separada desse problema (RML+SA) que primeiro considera a seleção do formato de modulação em função da rota e depois a alocação de espectro. Finalmente, uma meta-heurística usando o algoritmo *simulated annealing* também foi apresentada em [49].

A maioria das abordagens dos algoritmos DA-RSA apresenta versões *online* do problema e segue uma abordagem de dois passos. A maior diferença entre os algoritmos DA-RSA consiste no modelo de alocação de espectro utilizado para determinar o número de subportadoras em função da taxa de dados e do comprimento do caminho. Geralmente, no segundo passo do algoritmo, o número de *slots* varia dependendo do formato de modulação selecionado em função do comprimento do caminho considerado. Um exemplo é o algoritmo QoT-aware RSA proposto em [50] que consiste em três etapas: cálculo do caminho, seleção do caminho e atribuição do espectro. Os algoritmos Dijkstra e *k-shortest path* foram adaptados para calcular os caminhos enquanto consideram os efeitos lineares e não lineares usando uma expressão fechada que estima o QoT ao longo de um determinado caminho. Para cada solicitação, a rota mais viável é escolhida em relação ao ONSR.

Outro algoritmo DA-RSA, proposto em [51], é utilizado para calcular uma série de caminhos alternativos fixos (roteamento estático) para cada par de origem-destino e com número de *hops* ordenado de forma crescente. No segundo passo, a política de alocação de espectro é utilizada para ajustar a quantidade de *slots* de acordo com o formato de modulação para aquele comprimento (em *hops*) de caminho e para buscar sequencialmente cada caminho até encontrar *slots* contíguos que possam acomodar a taxa de dados solicitada.

Em [52], os autores apresentam um algoritmo DA-RSA para selecionar de forma dinâmica o formato de modulação apropriado utilizando dois modelos: alcance de transmissão e seleção do formato de modulação. O modelo de alcance de transmissão leva em consideração restrições de camada física, como atenuação da fibra, limites OSNR, efeitos não lineares, como *cross phase modulation* (XPM) e *four wave mixing* (FWM), interferência entre canais vizinhos, bem como uma margem de OSNR adicional de 3dB para os BV-WXCs ao longo do caminho. O modelo de seleção do formato de modulação considera o alcance da transmissão para os diferentes formatos e utiliza a lei de meia distância proposta em [53]. A lei define que para cada ganho de 3dB em OSNR a ordem de modulação pode ser aumentada em 1 bit, reduzindo a distância de transmissão em 50%. O algoritmo proposto em [51] utiliza um modelo de alcance de transmissão tabelado em

função do número de *hops* e o algoritmo em [52] estabelece um limite fixo de um caminho de 15 *hops* e calcula a distância de transmissão em função do comprimento máximo do caminho.

Um algoritmo que resolve o problema em um passo foi apresentado em [47] com o nome de *impairment-aware* RSA. O algoritmo apresenta uma função de custo para cada formato de modulação e cada enlace na rede recebe um peso igual à proporção do espectro requerido em relação ao número de *slots* livres neste enlace. Para cada formato de modulação, uma versão modificada do algoritmo de Dijkstra é usada para encontrar um caminho de custo mínimo com um espectro contíguo suficiente para essa solicitação.

2.2.2 Fragmentation-aware RSA

Um dos desafios que têm sido abordados na arquitetura EON é como os algoritmos RSA podem minimizar a fragmentação do espectro. A fragmentação espectral é decorrente do processo de alocação e liberação de conexões de bandas heterogêneas que resultam em lacunas espectrais que não são utilizadas, principalmente em um cenário de tráfego dinâmico. Existem dois tipos de fragmentação espectral: a horizontal e a vertical. A fragmentação horizontal é causada pela restrição de continuidade do espectro na medida em que o mesmo bloco de espectro pode estar indisponível ao longo de enlaces sucessivos de um caminho, apesar de cada enlace ter espectro suficiente para alocar uma solicitação. Esse problema é semelhante à fragmentação de comprimento de onda em redes WDM. Por outro lado, a fragmentação vertical é causada pela restrição da contiguidade do espectro e pelas solicitações de bandas heterogêneas de maneira que o espectro é fragmentado em pequenos blocos não contíguos que não podem ser alocados a uma única grande demanda.

A fragmentação espectral influencia diretamente as métricas probabilidade de bloqueio (*blocking probability*, BP) e imparcialidade (*fairness*). A medida de imparcialidade (equidade no atendimento) é usada para determinar se diferentes demandas de conexões estão recebendo um atendimento imparcial em relação às diferentes requisições de banda. Suponha a existência de dois tipos de requisições de banda: BW_1 e BW_2 . Um das medidas de imparcialidade é calculada pela razão entre as respectivas BPs, ou seja, $F = \frac{BP_1}{BP_2}$ [54]. Devido à fragmentação nos enlaces da rede, as requisições de maiores larguras de banda são bloqueadas com maior probabilidade de ocorrência, tornando os pedidos de conexão heterogêneos fortemente dependentes da taxa de dados e/ou do comprimento do caminho, e causando parcialidade (*unfairness*) no atendimento.

Os algoritmos RSA cientes da fragmentação (*fragmentation-aware* RSA) têm como objetivo melhorar a BP, a imparcialidade e a utilização do espectro, minimizando a extensão da fragmentação espectral. Para isso, é preciso o desenvolvimento de métricas que quantificam o grau de fragmentação na rede. A primeira medida de fragmentação em EONs foi introduzida em [55], descrevendo o conceito de entropia de fragmentação.

Essa métrica avalia a fragmentação do espectro de um enlace contando o número de pares de *slots* vizinhos que têm *status* diferente (ou seja, um *slot* está alocado, mas o outro está livre) e normaliza o valor no intervalo $[0, 1]$, de modo que os valores representam diretamente a fragmentação vertical. A medida de entropia remete à “quantidade de informação” necessária para descrever um conjunto de elementos que, no caso, serve para caracterizar as variações de status dos *slots* adjacentes. A medida de entropia foi estudada também em [56,57]. Outra medida, denominada taxa de fragmentação, proposta em [58,59], mensura a fragmentação em um enlace ou caminho de modo semelhante aos sistemas de armazenamento e memória. Essa medida considera o status de *slots* vizinhos ou o de *slots* em enlaces adjacentes e também o tamanho de blocos livres (e contíguos) de *slots*. Em [60] é apresentado o índice de fragmentação em que a cada bloco contíguo de *slots* livres é atribuído um valor igual à taxa de dados máxima que suporta. O índice de fragmentação de um caminho é calculado pela razão entre a soma do valor dos blocos disponíveis e o valor da soma dos *slots* em todos os blocos; a última soma representa o valor desses *slots* como se eles fossem contíguos.

Os algoritmos FA-RSA são classificados em proativo e reativo. O esquema FA-RSA proativo tenta prevenir ou minimizar a fragmentação do espectro no momento em que uma nova solicitação é admitida na rede. O estudo conduzido em [60] identifica quatro formas de particionar o espectro para atender variadas taxas de bits e testa o seu desempenho em um cenário de tráfego dinâmico. Um conceito semelhante de reservas de espectro é estudado em [61]. Em vez de dividir o espectro de modo que cada partição seja compartilhada apenas entre conexões com a mesma taxa, os autores do estudo propõem que um bloco de subportadoras contíguas seja reservado para cada par origem-destino.

Os algoritmos FA-RSA proativos tomam as decisões de alocação de espectro com base no estado atual da rede. Dois algoritmos FA-RSA, apresentados em [58], realizam as admissões com base na taxa de fragmentação. O primeiro algoritmo atribui a cada nova solicitação um caminho que minimiza uma métrica de fragmentação global. O segundo algoritmo tenta utilizar os *slots* que já são mais usados na rede. Dois outros algoritmos FA-RSA tentam controlar heurísticamente o tamanho dos blocos para cada alocação realizada. O primeiro algoritmo, denominado *maximize common large segment* (MCLS), proposto em [62], gera uma série de caminhos candidatos para uma solicitação e, para cada caminho, calcula a métrica que captura a disponibilidade de *slots* contíguos entre os enlaces candidatos após a alocação do espectro. A alocação de caminho e espectro que tem o menor valor para essa métrica é selecionada, de modo a maximizar a probabilidade de que pedidos futuros encontrem um número suficiente de *slots* contíguos a serem aceitos. Outro algoritmo, proposto em [63], consiste em uma heurística de “*cut avoiding*” que objetiva minimizar a divisão dos blocos contíguos de *slots*, evitando assim a fragmentação do espectro.

O algoritmo proposto em [64] obtém os caminhos utilizando o roteamento *k-shortest path* e, na etapa de alocação de espectro, o algoritmo *best-fit* busca a melhor alocação de acordo com uma função de custo definida. Um bloco grande é dividido em vários sub-blocos de acordo com o tamanho da solicitação. Se uma solicitação precisar de n FSs em um bloco contíguo de m FSs, ela terá $m - n + 1$ sub-blocos. A função de custo é baseada no estado dos *slots* vizinhos para minimizar a fragmentação após a alocação de espectro. Um FS livre recebe valor 1 e o ocupado recebe valor 0 e, assim, o custo de cada sub-bloco é definido pela sua posição no espectro e pelo estado dos FSs adjacentes. Os sub-blocos posicionados na extremidade esquerda avaliam o estado do FS à direita e os posicionados na extremidade direita avaliam o estado do FS à esquerda. Sub-blocos intermediários realiza a média do custo do estado do FS à esquerda e à direita. Essa heurística é semelhante ao demonstrado em [63], pois também avalia o custo do estado dos FS adjacentes. A diferença é que o valor do custo de *cut* é 1 somente quando o bloco é dividido pela alocação.

Em [65] são apresentados dois esquemas de alocação de espectro dinâmica baseada no aprimoramento da imparcialidade e na redução da fragmentação. O primeiro algoritmo, denominado *advanced fairness-aware* (AFA), considera a imparcialidade no momento da alocação e o segundo é um algoritmo híbrido, denominado *fairness-aware and fragmentation-reduction* (FAFR) que, além de considerar a imparcialidade, objetiva otimizar a medida de fragmentação no instante de alocação. O AFA estima um número total de FSs alocados para uma solicitação de tamanho i . Seja p_i a probabilidade de chegada de uma solicitação de banda de tamanho i e NFS o número total de FSs. O total é obtido por $i \times p_i \times NFS$ dividido pela esperança da distribuição de probabilidade das solicitações de banda. A busca por FSs contíguos é feita dentro desse intervalo estimado e a primeira alocação é selecionada. Essa função probabilística ajusta a imparcialidade para diferentes requisições de banda. O FAFR segue os mesmos passos do AFA, porém ao invés de selecionar a primeira alocação, o algoritmo seleciona a alocação com o menor grau de fragmentação (definido em [58, 59]) dentre as soluções candidatas.

Em um ambiente de tráfego dinâmico, os algoritmos proativos não eliminam completamente a fragmentação e, por isso, os algoritmos FA-RSA reativos eventualmente são utilizados, pois realizam a desfragmentação da rede ou de parte dela. O objetivo da desfragmentação é reorganizar a alocação de espectro das conexões existentes de modo a consolidar *slots* disponíveis em grandes blocos contíguos e contínuos que podem ser usados para estabelecer pedidos futuros de conexão. A desfragmentação pode ser acionada periodicamente ou desencadeada pelo estabelecimento do caminho (gatilho). A desfragmentação periódica é executada em escalas longas de tempo e é iniciada em intervalos regulares ou sempre que uma métrica indicativa de fragmentação em toda a rede exceda um determinado limite. Esse processo considera a fragmentação dos recursos espectrais em toda a rede e, portanto, é computacionalmente caro e pode interromper inúmeras conexões. Por outro lado, a desfragmentação acionada por gatilho atua localmente,

uma vez que é acionada quando há uma nova solicitação de conexão ou quando uma conexão é liberada. Essa desfragmentação é executada em escalas de tempo mais curtas, pois apenas interrompe conexões que compartilham enlaces do caminho que a desencadeou.

O problema de desfragmentação em toda a rede foi tratado em [66] por um algoritmo heurístico guloso (*greedy*) com dois objetivos. O primeiro é minimizar o espectro utilizado pelas conexões e o segundo é minimizar o número de conexões interrompidas durante o processo de reconfiguração. Um algoritmo de desfragmentação, estudado em [67], atua para estabelecer um caminho de uma nova solicitação que seria bloqueada. Se os recursos totais de *slots* disponíveis no caminho estabelecido (mas não necessariamente contínuos ou contíguos) não forem suficientes em número de *slots*, a solicitação é definitivamente bloqueada. Em caso contrário, o algoritmo tenta criar um bloco de *slots* contínuos e contíguos ao longo do caminho mais curto. O estudo [67] formula o problema em ILP e para resolvê-lo em caso de redes grandes os autores propõem uma meta-heurística gulosa. Outro algoritmo desencadeado pelo caminho, proposto em [68], usa uma técnica simples para consolidar o espectro liberado por uma conexão terminada com outros blocos de espectro disponíveis ao longo dos enlaces de seu caminho. Após a liberação de uma conexão, as conexões alocadas no espectro adjacente e que compartilham um enlace com a conexão finalizada são deslocadas, sempre que possível, para a extremidade inferior do espectro, criando mais blocos disponíveis na parte superior para futuras conexões.

Uma das preocupações ao realizar a desfragmentação do espectro é o intervalo de interrupção das conexões ativas porque as operadoras de rede precisam cumprir o contrato de qualidade de serviço definida no *service level agreement* (SLA). Um dos parâmetros de qualidade é a disponibilidade das conexões ativas ou *uptime*. Algumas técnicas utilizadas conseguem realizar a desfragmentação sem interromper as conexões ativas, como o *hitless optical path shift* (HOPS) proposto em [68]. A técnica HOPS foi avaliada em [69] na qual é referida como desfragmentação *push-pull*. Uma outra técnica, a *make before break rerouting* (MBBR), apresentada em [70], minimiza a interrupção das conexões existentes já que as conexões ativas são movidas para um novo bloco de espectro antes de interromper a reserva de recursos atual.

2.2.3 Traffic Grooming RSA (TG-RSA)

A agregação de tráfego (*traffic grooming*, TG) é um problema de otimização que surge no projeto e controle de redes que transportam múltiplos fluxos de dados com requisitos de banda inferiores à largura de banda oferecida pela rede. Esse problema foi resolvido tradicionalmente em sistemas digitais de conexões cruzadas (*digital cross-connect system*, DCS) [71] que combinavam fluxos de taxa de bits constante em um contêiner usando protocolo *synchronous optical networking/synchronous digital hierarchy* (SONET/SDH) ou *generic framing procedure* (GFP) (ITU-T G.7041). Esses protocolos realizam multiplexação

por divisão de tempo fornecendo a cada fluxo de bits um número fixo de *slots* de tempo que são incorporados em um contêiner (ou quadro) e transportados em um fluxo de dados de taxa de bits mais alta. Nas redes WDM convencionais, o objetivo principal é agregar as demandas granulares em caminhos ópticos de capacidade fixa para melhorar a utilização do espectro e reduzir o número de *optical add-drop multiplexer* (OADM) [72]. Por exemplo, pode-se combinar dois fluxos de 15 Gbps em um único fluxo de 40 Gbps transportados em um dos comprimentos de onda de uma rede WDM.

Nas redes EONs, os caminhos ópticos flexíveis podem ser configurados com a quantidade necessária de espectro para atender a taxa de dados de uma demanda de tráfego [73]. Essa tarefa é realizada pelo BV-T que pode gerar uma ou mais subportadoras para acomodar requisições de banda heterogênea e granular. Em um cenário com muitas requisições de baixa taxa de dados surgem dois problemas que são resolvidos ou amenizados pela técnica de agregação de tráfego. Primeiramente, ao acomodar um serviço de baixa taxa, um BV-T é reservado para transmitir uma (ou poucas) subportadoras. Devido ao número limitado de *transponders* nos equipamentos de rede, as demandas de baixa taxas podem reduzir drasticamente ou até esgotar a disponibilidade de *transponders* para atender novas solicitações. Nesse caso, a agregação de tráfego pode condensar múltiplas demandas de baixa taxa em um supercanal óptico utilizando todo potencial de transmissão OFDM de um BV-T. Além do mais, o transporte de uma única conexão de alta taxa (por exemplo, 400 Gbps) geralmente é mais eficiente do que o transporte de várias conexões de baixa taxa com a mesma taxa de dados total (por exemplo, 4 demandas de 100 Gbps) [74]. Em segundo lugar, quanto maior o número de subcanais independentes, maior é o número de bandas de guarda em torno desses subcanais, causando um desperdício espectral adicional. Para resolver esta questão, o ideal é agregar os vários serviços em um único canal óptico, reduzindo assim o espectro necessário para atender um determinado conjunto de demandas e, conseqüentemente, aumentando a quantidade de espectro disponível para atender futuras demandas de tráfego. Portanto, a agregação de tráfego em EONs tem o potencial de melhorar a utilização do espectro, além de produzir uma redução significativa de custos em termos de *transponders*.

As soluções totalmente ópticas, ou *transparent optical networks* (TONs), (incluindo os roteadores ópticos) que se tornaram disponíveis nos últimos anos reduziram a complexidade das redes ópticas ao reduzir o número de conversões óptica-elétrica-ópticas (O-E-O) na rede. Dessa forma, a agregação do tráfego nas EONs pode ser realizada diretamente na camada óptica por meio da configuração de um túnel óptico, um conceito introduzido em [73]. A ideia do túnel óptico é combinar serviços de subcomprimento de onda em um BV-T e transmiti-los de forma alternada, separados por uma banda de guarda interna ao túnel. Em [75] foram estudados algoritmos de agregação de tráfego na camada óptica que consideram a agregação de serviços que partem do mesmo nó fonte.

Os algoritmos que consideram a agregação de tráfego em EONs são chamados de TG-RSAs. O problema TG-RSA, seja *online* ou *offline*, é significativamente mais difícil de resolver de forma otimizada. O problema *offline* é tratado em [75–77] e versões *online* do problema são apresentados em [78–81]. Além do mais, algoritmos TG-RSA geralmente fazem parte da classe de algoritmos que focam na eficiência energética da rede, já que é possível minimizar o consumo de energia dos equipamentos por meio da otimização da utilização dos *transponders*, como apresentado em [82, 83].

O estudo em [76] compara o desempenho entre os algoritmos RSA e TG-RSA utilizando formulação MILP com o objetivo de minimizar a taxa média de utilização do espectro ponderada pelo comprimento do enlace. Em [75], as formulações são baseadas em caminho e utiliza uma formulação do RSA *offline* com restrições específicas de agregação de tráfego. Como a resolução do problema via ILP é impraticável para grandes redes devido ao custo computacional, também são apresentados dois algoritmos heurísticos TG-RSA: *least spectrum grooming* (LSG) e *minimum transmitter grooming* (MTG). O algoritmo LSG minimiza o uso total do espectro agregando os caminhos ópticos que compartilham as rotas mais longas com o mesmo nó de origem. Esse algoritmo visa maximizar o ganho de espectro pelas bandas de guarda reduzidas. O algoritmo MTG visa minimizar o número de *transponders*. Os passos do algoritmo são semelhantes ao LSG; a diferença é que na busca de uma alocação de um túnel óptico, ou seja, um grupo de caminhos ópticos agregados, o algoritmo MTG avalia o caminho com a máxima capacidade do túnel óptico ao invés da real banda requerida da conexão. Os dois algoritmos heurísticos utilizam a computação de caminhos *k-shortest path*. Em [77] é considerada a agregação na camada óptica com o roteamento *multicast* e visa reduzir a quantidade de *splitters*. Uma conexão *multicast* possui mesma fonte e destinos diferentes e essa operação é realizada pelo *splitter* óptico. O *splitter* óptico é um componente passivo da rede que serve para dividir o sinal óptico. A desvantagem do *splitter* é que ele causa atenuação do sinal e, portanto, o algoritmo tenta reduzir o número de *splitters* necessários em uma transmissão *multicast* por meio da agregação de tráfego em conexões de possuam destinos próximos.

Os algoritmos TG-RSA online foram tratados usando grafo auxiliar que possibilita a aplicação de uma variedade de políticas de agregação, como minimizar o número de *transponders* e o número de enlaces físicos ou lógicos. O grafo auxiliar é muito utilizado para resolver problemas RSA, pois ele permite aplicar restrições na rede por meio da remoção, adição ou modificações dos pesos das arestas. Em [78], cada nó físico da rede é representado em camadas no grafo auxiliar. A camada do *transponder* abstrai a habilidade de conversão O-E-O e as camadas de espectro representam os recursos disponíveis. A restrição de continuidade de espectro é aplicada detectando arestas entre as mesmas camadas do espectro de nós adjacentes do grafo auxiliar, e a restrição de contiguidade de espectro é aplicada pela exclusão de arestas entre camadas de espectro se a quantidade de espectro contíguo não for suficiente para alocar uma demanda. Um caminho para uma demanda

é encontrado executando o algoritmo *shortest path* no grafo auxiliar que representa o estado da rede no instante da solicitação. O modelo de grafo auxiliar também foi usado em [79] para encontrar um caminho primário e um caminho secundário de proteção compartilhada. O caminho primário é selecionado usando o algoritmo *k-shortest path* no gráfico auxiliar e o caminho secundário de proteção é disjunto do caminho primário. Os autores do estudo propõem um esquema de proteção designado *elastic separate-protection-at-connection* (ESPAC), que permite que os caminhos ópticos de proteção de duas conexões compartilhem espectro se os respectivos caminhos primários forem disjuntos em enlace e, dessa forma, o ESPAC realiza a agregação do espectro compartilhado nos caminhos de proteção reduzindo assim a utilização do espectro. Os algoritmos RSA de proteção e restauração para EONs são discutidos com mais detalhes na seção a seguir.

Várias políticas de agregação na camada óptica são comparadas em [80], em que múltiplos fluxos ópticos de diferentes fontes e destinos podem ser agregados em um *sliceable* BV-T [84]. Nesse estudo, o modelo de grafo auxiliar possui três camadas para abordar o problema de agregação mista entre transmissão *all-optical* e com conversão O-E-O: camada elétrica, camada de *sub-transponder* e camada de *transponder*. Na camada elétrica, os nós representam a abstração do roteador ou *switch* elétrico. Na camada *sub-transponder*, os nós representam a abstração das subportadoras. Por fim, na camada do *transponder*, os nós representam o fluxo do túnel óptico. Esse modelo permite ajustar os pesos das arestas do grafo auxiliar para implementar as seguintes políticas de agregação de tráfego: *maximal optical grooming* (MOG) e *maximal electrical grooming* (MEG), que são propostas para minimizar o número de caminhos ópticos recém-estabelecidos; *minimal number of virtual hops* (MVH), que visa minimizar o número de *hops* virtuais que a nova conexão atravessa para reduzir o número de conversões O-E-O; e *minimal number of physical hops* (MPH), cujo objetivo é minimizar o número de *hops* da rede. Além disso, foram propostos dois esquemas de reserva de espectro que podem utilizar eficientemente a capacidade de um *transponder*. O primeiro é chamado de *spectrum reservation for each lightpath* (SRLP), no qual um canal de um *sub-transponder* é reservado para cada novo caminho. O segundo esquema é chamado de *spectrum reservation for each node-pair* (SRNP) que faz apenas uma reserva de caminho óptico para cada par de nós na topologia física. Os caminhos ópticos com o mesmo par de nós são reservados até que a sua capacidade seja totalmente utilizada. Em seguida, uma nova reserva de um caminho óptico entre o par de nós é feita para as futuras conexões. Em [81] os autores apresentam uma extensão das funcionalidades do estudo apresentado em [80]. Usando a mesma forma de abstração em camadas, o problema tratado em [81] inclui a camada *spatial division multiplexing* (SDM). A tecnologia SDM, detalhada em [85], considera que os enlaces de rede possuem fibras com múltiplos núcleos e que o tráfego possa ser multiplexado e agregado entre os diferentes núcleos da fibras.

2.2.4 *Survivability* RSA

A taxa de bits nas EONs pode variar de 400 a 1000 Gbps por canal, correspondendo à capacidade agregada de dezenas a centenas de Tbps por fibra. Com o aumento significativo da taxa de transmissão, as falhas nos componentes de rede podem interromper inúmeras conexões e acarretar uma grande perda de dados. Por isso, existem mecanismos que garantem a sobrevivência da rede até que os componentes com falha sejam reparados pelos operadores de rede. A sobrevivência da rede (*survivability*) refere-se à capacidade da rede de ser reconfigurada para restaurar as conexões afetadas por uma ou mais falhas. Em uma rede óptica, três tipos de falhas são geralmente considerados: falhas de enlace (por exemplo, causadas por cortes de cabos de fibra óptica), falhas de nó (por exemplo, devido a mau funcionamento do equipamento em um *switch* ou roteador) e falhas de canal (por exemplo, causada por falha no *transponder* que pode interromper a transmissão em um canal específico).

Vários mecanismos de sobrevivência foram explorados em redes ópticas e podem ser classificados em uma das duas grandes categorias: proteção e restauração. Os esquemas de proteção de rede reservam previamente recursos para recuperar o tráfego afetado por uma falha. Os recursos de proteção são pré-configurados na fase de planejamento da rede ou no momento do estabelecimento de uma conexão e podem ser dedicados a uma única conexão ou compartilhados entre várias. Durante a operação normal da rede, os recursos reservados permanecem inativos e, na ocorrência de falha, as conexões afetadas são redirecionadas para os recursos de proteção correspondentes. Quando uma conexão está protegida por esses esquemas, a recuperação da conexão é rápida e geralmente mais garantida (caso o caminho de proteção não falhe também). Um esquema de restauração, por outro lado, não realiza a reserva prévia de recursos; estes são alocados após a detecção da falha. Em um esquema de restauração, a recuperação da conexão não é garantida, pois depende dos recursos disponíveis distribuídos por toda a rede no momento da ocorrência da falha; e o tempo de reestabelecimento da conexão é mais longo que nos esquemas de proteção devido à execução do algoritmo RSA.

Como mostrado na Fig. 1 [17], os esquemas de proteção e restauração de rede podem ser divididos em categorias: baseadas em enlaces (*links*) ou em caminhos (*paths*). As técnicas baseadas em enlace recuperam uma falha de rede entre os nós em torno do enlace com falha, enquanto as técnicas baseadas em caminho recuperam a falha entre os nós origem-destino de um caminho. As técnicas de proteção também podem ser divididas com base nas topologias de rede, como a proteção baseada em anel (*ring*); como *unidirectional path switched ring* (UPSR) e *bidirectional line switched ring* (BLSR); e a proteção baseada em malha (*mesh*). Essas técnicas podem realizar a reserva de recursos de forma dedicada ou compartilhada.

Figura 1: Esquemas de sobrevivência (proteção e restauração) de rede. UPSR: unidirectional path switched ring, BLSR: bidirectional line switched ring, SBPP: shared backup path protection, PR: path restoration [17].

2.2.4.1 Proteção

O problema de garantir proteção dedicada ou compartilhada de conexões estabelecidas em uma EON pode ser visto como uma variante do problema RSA com restrições adicionais para considerar os caminhos de *backup* e o compartilhamento (se houver) do espectro. Além disso, ao estender os esquemas de proteção propostos para redes WDM para a arquitetura EON, podem-se considerar as técnicas de distância adaptativa, agregação de tráfego e tratar o problema da fragmentação.

Os critérios mais importantes para medir as vantagens e desvantagens dos esquemas de proteção são o tempo de restauração e a capacidade ociosa de recursos. O tempo de restauração é o intervalo decorrido a partir da ocorrência de uma falha até a sua recuperação. Um intervalo curto de restauração é frequentemente esperado para uma técnica de proteção. Essa medida está associada à disponibilidade (*availability*) da rede que é uma métrica que faz parte do SLA. A eficiência na reserva de recursos é outro critério importante para avaliar uma técnica de proteção. Uma métrica denominada redundância de capacidade de reserva mede a relação entre a capacidade total de proteção reservada em toda a rede pela capacidade total da rede. Outra métrica é a taxa de utilização de espectro, que é a relação entre o total de recursos utilizados pelos caminhos de proteção e o total de recursos utilizados pelos caminhos de trabalho. Quanto maior a eficiência na utilização do espectro

por um esquema de proteção, maior a capacidade disponível da rede. Alguns esquemas de proteção são listados a seguir:

1. *p-cycle: pre-configured cycle* é um tipo de proteção dedicada que realiza a reserva de recursos na fase de planejamento da rede provendo proteção no enlace semelhante a uma proteção em anel. Os recursos reservados formam um ciclo hamiltoniano na rede. Apesar de o esquema ser pré-configurado, é possível realizar a reconfiguração *online* acionada mediante a ocorrência de falha na rede. Alguns estudos sobre *p-cycle* em EONS foram apresentados em [86–88].
2. *Dedicated path protection* (DPP): O DPP é um esquema de proteção baseado em caminho em que dois caminhos ópticos disjuntos são reservados no momento do estabelecimento da conexão, sendo um o caminho primário (de trabalho) e o outro o de proteção (*backup*). Os caminhos ópticos podem ser disjuntos em enlace ou nó [89–93].
3. *Shared backup path protection* (SBPP): Assim como o esquema DPP, o esquema SBPP também é baseado em caminho. A ideia do SBPP é compartilhar um caminho de proteção entre diferentes caminhos primários, ou seja, um caminho de proteção dá suporte a vários caminhos primários. O esquema SBPP oferece a vantagem de melhorar a utilização de recursos da rede e, no caso de uma única falha, apresenta o mesmo nível de proteção do que o DPP [30, 32, 34, 79, 90, 91, 94–96].

Nos trabalhos realizados em [90, 91] os autores comparam o DPP e o SBPP em diferentes situações. O primeiro trabalho [90] realiza uma abordagem de atribuir diferentes esquemas de proteção para lidar com várias classes de tráfego oferecendo diferentes níveis de proteção necessário. O objetivo principal é mostrar a melhoria da eficiência de transmissão com o uso de três classes distintas de tráfego, resultando em uma diminuição da probabilidade de bloqueio em relação aos métodos tradicionais de proteção de caminho DPP e SBPP.

Dentre os esquemas citados, o SBPP é o mais vantajoso em relação à economia de recursos espectrais por causa da característica de compartilhamento [97]. Além disso, o SBPP permite implementar as várias técnicas apresentadas anteriormente. Por esse motivo, os algoritmos que envolvem esquemas SBPP são o foco deste trabalho e serão apresentados com mais detalhes a seguir.

O esquema *dynamic load balancing shared-path protection* (DLBSPP) [30] é uma heurística de roteamento que considera o balanceamento dinâmico de carga na computação do caminho, ou seja, os enlaces menos carregados são preferidos na computação tanto do caminho primário quanto no de proteção para cada conexão. Uma característica interessante do algoritmo é que, para fins de computação dos caminhos, os custos dos

enlaces são definidos proporcionalmente ao número de slots de frequência disponíveis, em uma tentativa de balancear a carga pelos enlaces da rede. Para a computação do caminho de *backup*, a função de custo também é proporcional à quantidade de compartilhamentos para maximizar o número de compartilhamentos.

O estudo em [32] considerou o problema RSA *online* com proteção compartilhada e introduziu duas políticas para compartilhamento de espectro entre os caminhos de *backup*: a *conservative sharing policy* (CSP) e a *aggressive sharing policy* (ASP). Sob o CSP, o compartilhamento é permitido apenas se os caminhos de *backup* tiverem a mesma largura de banda, enquanto que sob o ASP, o compartilhamento é permitido mesmo se os caminhos de *backup* tiverem larguras de banda diferentes. A política agressiva leva a mais oportunidades de compartilhamento, mas ela pode fragmentar o espectro ao longo dos caminhos de *backup*. Em [94] é considerado o algoritmo DA-RSA com proteção SBPP. O algoritmo heurístico utiliza um grafo auxiliar para os sub-blocos contínuos como apresentado em [64] e define uma função de custo para cada FS compartilhável. O custo do sub-bloco de alocação é calculado pelo somatório dos custos dos FSs que o compõe. A técnica proposta é eficaz na redução de probabilidade de bloqueio e na melhora da eficiência espectral em comparação com esquemas SBPP convencionais.

O problema da fragmentação com proteção compartilhada sob uma demanda dinâmica de tráfego é tratado em [95]. Dois algoritmos heurísticos de desfragmentação *hitless* foram propostos: *sequentially releasing defragmentation* (SD-R) e *jointly releasing defragmentation* (JR-D). Nos dois algoritmos, para uma desfragmentação eficiente, a rota e o espectro de cada caminho de proteção é alterada. O SR-D é acionado quando um determinado número de caminhos é liberado. Primeiro, os caminhos compartilhados estabelecidos na rede são classificados em ordem decrescente com base no índice máximo dos FSs usados por cada caminho de proteção e, em seguida, liberado e realocado um a um. Em contraste, quando uma desfragmentação é acionada no esquema JR-D, todos os recursos espectrais de todos os caminhos de proteção são liberados e depois rearranjados conjuntamente. A maior vantagem da desfragmentação *hitless* é que os caminhos ópticos que transportam tráfego de usuário não é afetado quando os caminhos de proteção são desfragmentados. Em [34] é apresentada uma heurística semelhante de alocação e realocação *hitless* para caminhos compartilhados. As políticas *first-fit* e *last-fit* são utilizadas para estabelecer os caminhos de trabalho e de proteção compartilhada, respectivamente. Quando uma conexão é liberada, a realocação do espectro é acionada e os caminhos de proteção das conexões dos caminhos primários cujos recursos de *backup* coincidem com os enlaces do caminho liberado são rearranjados. Como a realocação ocorre apenas em um intervalo limitado, o algoritmo atua localmente e, portanto, apresenta menor complexidade.

O algoritmo ESPAC [79], citado na Seção 2.2.3, implementa um esquema de proteção compartilhada independente de falha em um contexto de agregação de tráfego. A alocação

de espectro é realizada usando *first-fit* para atribuir tráfego de trabalho e *last-fit* para atribuir tráfego de *backup* e permitir a sobreposição de espectro entre comprimentos de onda de *backup* adjacentes. A elasticidade do *transponder* permite a expansão e contração dos caminhos ópticos e, assim, quando ocorre uma única falha na rede, os caminhos ópticos que transportam fluxos de *backup* podem ser sintonizados de tal forma que o espectro de sobreposição é usado por apenas um dos caminhos ópticos adjacentes.

2.2.4.2 Restauração

A granularidade espectral nas EONs possibilita um novo esquema de sobrevivência chamado *bandwidth squeezed restoration* (BSR), que foi primeiramente proposto em [98] e refinado em [99]. Nesse esquema, há três possíveis modelos de comportamento de restauração de conexão: *full bandwidth guaranteed recovery* (FBGR), *partial bandwidth guaranteed recovery* (PBGR), e *best effort recovery* (BEFR). Os modelos FBGR e PBGR funcionam, respectivamente, com proteção dedicada e proteção dedicada com banda reduzida. A recuperação de uma conexão que utiliza esses modelos é feita pela comutação dos caminhos de trabalho para os caminhos de *backup*. O modelo BEFR, exclusivo das EONs, não reserva tais recursos e a recuperação é feita pela restauração dinâmica por melhor esforço. A restauração por melhor esforço permite alocar qualquer quantidade de recursos espectrais disponíveis no momento da falha, de modo que o espectro alocado possa ter a sua banda reduzida (*squeezed*) e que uma conexão continue ativa, mesmo com uma taxa de bit baixa. Em [30] é estudado o caso de recuperação de conexões afetadas por múltiplas falhas utilizando SBPP. No caso de uma falha de vários enlaces, o algoritmo tenta restaurar dinamicamente todas as conexões para as quais os caminhos de trabalho e *backup* ficam indisponíveis. Para este fim, as conexões afetadas são priorizadas em ordem decrescente de suas demandas de espectro, e o algoritmo calcula em tempo real um novo caminho para cada conexão, começando com a conexão com a maior prioridade (ou seja, a maior taxa de dados). A motivação para essa abordagem, conhecida como *traffic aware restoration* (TAR), baseia-se na observação de que é mais difícil encontrar caminhos para conexões de alta taxa de bit e, portanto, essas devem ser acomodadas antes de alocar recursos para conexões de baixa taxa de bit. Porém, o estudo realizado em [100] provou que o TAR não obteve melhora significativa comparado à uma restauração dinâmica não ordenada, ou seja, no caso em que as conexões são recuperadas de forma aleatória.

2.3 Considerações Finais

A revisão bibliográfica realizada neste estudo abordou alguns dos principais algoritmos RSA que tratam de: distância adaptativa, agregação de tráfego, fragmentação e esquemas de proteção. Vemos que alguns desses algoritmos, como [36, 37], utilizaram lógica *fuzzy* para resolver o problema e otimizar o QoT na alocação da conexão. Em um escopo diferente dos principais algoritmos abordados, vemos também o uso de redes

neurais para RSA, como em [35], que utiliza um sistema probabilístico de aprendizagem de máquina para predição de tráfego. Contudo, até o momento não foram encontrados estudos envolvendo proteção compartilhada que utilizasse técnicas de lógica *fuzzy* e redes neurais e otimizando a fragmentação do espectro, inclusive dos caminhos de proteção. Além do mais, as métricas de fragmentação estudadas avaliam o estado de alocação dos *slots* de frequência antes da realização da conexão. Para se adaptar melhor aos nossos objetivos, apresentamos uma métrica de fragmentação que considera a minimização do nível de fragmentação posterior ao estabelecimento da conexão candidata.

3 SISTEMA DE INFERÊNCIA FUZZY E REDE NEURAL ARTIFICIAL PARA SOLUÇÃO DO PROBLEMA DE ROTEAMENTO E ALOCAÇÃO DE ESPECTRO

Nesse capítulo abordaremos algumas técnicas de lógica fuzzy e redes neurais para resolver problemas de roteamento e alocação de espectro em redes ópticas elásticas.

3.1 Sistemas fuzzy aplicados em Problemas RSA

Nesta seção apresentamos alguns problemas RSA que utilizam a lógica *fuzzy* como ferramenta matemática principal. As noções sobre o funcionamento da lógica *fuzzy* está no Apêndice A.

As dificuldades encontradas na construção de um modelo matemático preciso para a caracterização das diferentes métricas utilizadas no processo de roteamento em EON tornam a lógica *fuzzy* uma ferramenta muito atraente. A lógica *fuzzy* tem sido intensamente utilizada na engenharia em sistemas de controle e em outras áreas. No entanto, até onde se sabe, poucos estudos foram conduzidos em problemas de roteamento em redes ópticas. Alguns desses estudos foram apresentados em [37, 101, 102].

O estudo apresentado em [37] utiliza um controlador *fuzzy* para resolver o problema de roteamento de comprimento de onda em redes WDM. Embora o foco desse estudo seja para as redes WDM, o modelo proposto pode ser estendido para as redes EON acrescentando a restrição de contiguidade de alocação do espectro. A técnica de roteamento *fuzzy* proposta incorpora atributos de qualidade de serviço (QoS), como latência, comprimento físico do enlace, taxa de perda de pacote de dados e atributos de roteamento, como número de *hops* e disponibilidade de comprimento de onda no caminho.

O algoritmo computa os k -menores caminhos e o sistema de inferência baseado em regras *fuzzy* seleciona o caminho com melhor qualidade e, em seguida, realiza a alocação do espectro usando o algoritmo *first-fit*. O modelo de inferência utiliza dois controladores com funções de pertinência triangular para variáveis linguísticas de entrada (*excelente, very good, good*) e variáveis linguísticas de saída (*excelent, very good, good, medium, poor, very poor*), além de compor um conjunto de 27 regras. O operador de composição mínimo [103] determina o relacionamento entre os conjuntos de entrada. O operador de implicação relaciona os conjuntos difusos de entrada (antecedente) com a saída do conjunto *fuzzy* (consequente) para cada regra. O processo de implicação de Mamdani [104] é implementado no algoritmo proposto. A unidade de agregação usa o operador *max* para agregar cada conjunto *fuzzy* de saída derivado do processo de implicação para cada regra. Após a realização do processo de agregação, o conjunto *fuzzy* de saída é obtido pelo processo de *defuzzification* utilizando o método centroide. O primeiro controlador utiliza os atributos

relacionados ao QoS como entradas no processo de *fuzzyfication* e obtém uma saída de custo. O custo do primeiro controlador *fuzzy* é utilizado como entrada do segundo controlador, juntamente com os atributos relacionados ao roteamento (número de *hops* e disponibilidade de comprimento de onda). A saída do segundo controlador define o custo final e a rota de menor custo é selecionada.

No segundo estudo, conduzido em [101], os autores propõem um algoritmo baseado em lógica *fuzzy* para resolver o problema RSA em EONs. O algoritmo seleciona a melhor rota para atender à demanda de tráfego com base nos seguintes parâmetros de entrada:

1. Taxa de ocupação dos *slots* de frequência contíguos: após a aplicação da restrição de continuidade do espectro, a medida é calculada pela relação entre o número de *slots* ocupados e o número total de *slots* do enlace;
2. Taxa de ocupação dos *slots* de frequência do percurso óptico: para cada enlace calcula-se a relação entre o número de *slots* ocupados e o número total de *slots* do enlace;
3. Comprimento físico da rota: é uma medida normalizada que corresponde ao comprimento da rota dividido pelo comprimento da maior rota.

Para cada solicitação de conexão na rede, são computadas k rotas entre os nós de origem e de destino utilizando o algoritmo de Yen (algoritmo clássico *k-shortest path*) [105]. Em seguida, o algoritmo avalia o custo total de cada rota por meio de um conjunto de regras de decisão e seleciona a rota com o menor custo total para atender à solicitação. O processo de decisão é realizado pela lógica *fuzzy* que, independentemente das imprecisões apresentadas nos parâmetros de entrada, oferece soluções robustas. O modelo define funções de pertinência do tipo trapezoidal, isto é, três variáveis linguísticas para cada entrada e um conjunto de regras de decisão. O processo de *defuzzification* é baseado na abordagem Tagaki-Sugeno [104] para determinar os valores numéricos da função de custo.

Em outro estudo, realizado em [102], os autores apresentam um modelo de lógica *fuzzy* para tratar o problema de fragmentação em algoritmos RSA. O objeto do algoritmo é otimizar a seleção dos caminhos utilizando um sistema de inferência *fuzzy* tendo como entradas as métricas de fragmentação externa e de utilização da rede. O primeiro passo do algoritmo é a realização de um roteamento *offline* para obtenção do conjunto de menores caminhos para cada par de nós origem-destino. Na etapa de operação da rede, para cada solicitação de conexão, as rotas com pares origem-destino correspondentes são avaliadas pelo sistema de inferência *fuzzy*. O sistema de inferência *fuzzy* utiliza o modelo Mamdani e as entradas seguem os seguintes critérios:

1. Grau de fragmentação relativa da rota: essa medida permite verificar quanto uma rota é fragmentada para atender às solicitações que exigem diferentes números de *slots*. No algoritmo proposto, as rotas de maior qualidade são aquelas que podem ser alocadas e exibem maior fragmentação relativa;
2. Índice de primeiro espaço alocável: este critério permite a escolha de uma rota com menor índice espectral resultando em melhor organização do espectro. A escolha de rotas com o primeiro intervalo alocável perto do início do espectro permite uma melhor implementação da política *first-fit* no processo de alocação do espectro;
3. Número de *slots* ocupados na rota: corresponde ao total de *slots* continuamente ocupados na rota. Rotas com muitos *slots* ocupados devem ser evitadas para não criar gargalos na rede.

Dessa forma, as informações sobre o nível de fragmentação permitem escolher uma rota com o menor grau de fragmentação e possibilita uma redução na probabilidade de bloqueio. As informações sobre o índice do *slot* formam uma política de *first-fit* e o número de *slots* ocupados permite um melhor balanceamento de carga na escolha das rotas. A função de pertinência é uma função gaussiana que mapeia as entradas em cinco variáveis linguísticas. O sistema permite inferir o grau de pertinência de um elemento em um conjunto pelo uso de variáveis linguísticas e regras de controle. As variáveis classificam o elemento, assim como as regras reúnem as diferentes classificações e definem uma qualidade que representa o elemento avaliado. O conjunto de variáveis linguísticas selecionadas para qualificar a rota é transmitido para o modelo de inferência de Mamdani e, juntamente com o conjunto de regras, seleciona a variável linguística que melhor define a rota. Finalmente, o processo de *defuzzification* obtém um valor numérico que caracteriza o estado da rota considerada.

3.2 Redes Neurais Artificiais para Problemas RSA

Nesta seção apresentamos alguns problemas RSA que utilizam redes neurais como ferramenta matemática principal. As noções sobre o funcionamento de uma rede neural está no Apêndice B.

Nas aplicações de roteamento em redes ópticas, dois tipos de redes neurais se destacam: *perceptron* multicamadas (PCM) e redes de Hopfield. As redes *perceptron* multicamadas possuem uma camada de entrada, uma de saída e pelo menos uma camada intermediária (escondida) de neurônios, como ilustra a Fig 2. Um das aplicações desse tipo de rede neural é o mapeamento de problemas envolvendo sistemas variantes no tempo como a predição de eventos futuros. Em [35], uma rede PCM foi modelada para prever o tempo de manutenção das conexões estabelecidas e, dessa forma, otimizar o processo

de roteamento para conexões futuras. Nesse estudo, a rede neural foi utilizada como uma função heurística para alterar os custos das rotas obtidas pelo algoritmo *k-shortest path*.

Figura 2: Rede *perceptron* multicamadas contendo camada de entrada (*input layer*), camada escondida (*hidden layer*) e camada de saída (*output layer*).

Os autores dos estudos conduzidos em [106, 107] apresentam redes de Hopfield que realizam o roteamento de forma direta. A rede de Hopfield é constituída de uma única camada em que todos os neurônios são completamente interligados, isto é, todas as saídas da rede realimentam todas as suas entradas, conforme mostra a Fig 3. Uma das aplicações mais difundidas das redes de Hopfield diz respeito às memórias associativas binárias que podem ser usadas para resolver problemas de otimização combinatória [108]. No caso do roteamento como apresentado em [106] a rede de Hopfield possui uma função de energia que provê uma medida de desempenho para o problema de otimização. Cada elemento C_{ij} na matriz de custo de tamanho $N \times N$ é representado por um neurônio, portanto, o número de neurônios da rede de Hopfield para este problema é $N(N-1)$. A cada iteração, a rede converge para o estado de menor energia e, dessa forma, a saída binária Y_{ij} representa se o enlace de i para j faz parte ou não do menor caminho. Em [107], o roteamento utiliza uma função de energia para considerar as deficiências físicas do enlace e selecionar a rota com melhor QoT.

Figura 3: Rede de Hopfield. Os neurônios da camada de saída realimentam a camada de entrada.

3.2.1 Sistemas Neuro-Fuzzy

Um sistema neuro-*fuzzy* (SNF) [109] é um sistema híbrido constituído pela combinação de duas técnicas de modelagem: a *fuzzy* e rede neural artificial (RNA). Os SNFs têm chamado a atenção, pois trazem os benefícios tanto da RNA quanto de sistemas *fuzzy*, removendo assim as desvantagens individuais ao combinar as características comuns. Além disso, esses sistemas podem ainda incorporar algoritmos meta-heurísticos e evolucionários como *particle swarm optimization* (PSO) e *genetic algorithm* (GA).

Toda técnica inteligente tem propriedades computacionais como capacidade de aprender e tomar de decisões que as tornam adequadas para a solução de problemas. Por exemplo, enquanto as RNAs são adequadas para reconhecer padrões, elas não são adequadas para explicar como elas atingem suas decisões. Por outro lado, os sistemas de lógica *fuzzy*, que podem trabalhar com informações imprecisas, são bons para explicar suas decisões, mas não capazes de estabelecer automaticamente as regras que usam para tomar tais decisões. Essas limitações têm sido a motivação para a criação de sistemas híbridos inteligentes. Basicamente, um SNF pode ser modelado de duas formas:

1. A resposta das inferências linguísticas do bloco de interface *fuzzy* fornece um vetor de entrada para uma rede neural, que pode ser treinada para produzir saídas ou decisões de comando (Fig. 4);
2. A saída de uma rede neural conduz o mecanismo de inferência *fuzzy*. As redes neurais podem ser utilizadas para ajustar as funções de associação que definem quantitativamente as variáveis linguísticas (Fig. 5).

Figura 4: Modelo neuro-*fuzzy* 1 [109]. As saídas da lógica *fuzzy* servem de entradas para as redes neurais.

Figura 5: Modelo neuro-*fuzzy* 2 [109]. As saídas da rede neural servem de entradas para a lógica *fuzzy*.

Os SNFs possuem características que motivam o estudo de modelos para tratar o problema RSA e suas variantes, principalmente em um cenário dinâmico. Os sistemas

fuzzy são apropriados para modelar soluções em que é preciso de um especialista que conheça bem o comportamento do problema. E as RNAs são adequadas para reproduzir o comportamento de um problema com base no conhecimento embutido no conjunto de dados. O projeto de um SNF para resolver o problema RSA deve considerar:

1. **As relações de compromisso entre os parâmetros de rede.** Algumas relações de compromisso são explícitas como a que existe entre o alcance de transmissão e a largura do espectro no algoritmo DA-RSA. Porém, pode haver relações de compromisso entre outros parâmetros (ou características) de redes que não são intuitivas de se abstrair. Neste caso, o SNF pode aprender com a dinâmica da rede e correlacionar os parâmetros analisados [110, 111];
2. **A aleatoriedade das demandas em um cenário de tráfego dinâmico.** O critério para a escolha da melhor rota e/ou alocação de espectro é obtido unicamente pelas informações explicitadas no estado de alocação atual da rede. A melhor solução RSA no instante atual pode não ser a solução ideal que garanta a disponibilidade de espectro para as demandas futuras. Mesmo que os algoritmos RSA utilizem heurísticas para organizar o espectro da melhor forma possível, a variável temporal ainda tem grande influência nas alocações. Dessa forma, os SNFs podem aprender com o tempo e predizer a melhor alocação atual tendo em vista perspectivas de demandas futuras [35];
3. **A otimização do problema RSA na abordagem multiobjetivo.** A dificuldade de tratar um problema multiobjetivo está na escolha de uma solução dentre muitas soluções sub-ótimas. Neste caso, um sistema neuro-*fuzzy* desempenharia bem esse papel de analisar as imprecisões das entradas do problema. Além do mais, uma combinação específica dos parâmetros de entrada pode ser um fator determinante na qualidade da solução multiobjetivo. Tal problema combinatório pode ser otimizado pelo bloco neural do SNF [37, 102].

4 MODELOS DE SISTEMA DE INFERÊNCIA FUZZY E DE REDE NEURAL PROFUNDA PARA O PROBLEMA RSA COM PROTEÇÃO COMPARTILHADA

Nesse capítulo apresentamos os algoritmos de roteamento e alocação de espectro para gerar a base de soluções necessárias para incorporar os sistemas de avaliação de custo dos caminhos de alocação. Em seguida, é proposto um modelo de sistema de inferência fuzzy para avaliar os caminhos de trabalho e proteção utilizando a fragmentação como métrica principal. Por último, é modelada uma rede neural profunda que utiliza o sistema fuzzy de avaliação dos caminhos para simplificar e generalizar o modo de resolver o problema RSA com proteção compartilhada.

4.1 Algoritmo de Roteamento e Alocação de Espectro com Proteção Compartilhada (RSA-SBPP)

Nesse trabalho, o problema RSA foi dividido em dois subproblemas: o problema de roteamento e o problema de alocação de espectro. Primeiramente, resolvemos o problema de roteamento com o algoritmo de Yen (*k-shortest path*) obtendo, assim, K menores caminhos. A primeira etapa é chamada de roteamento estático, em que calculamos previamente quais os caminhos disponíveis para cada par de nós (i, j) . O conjunto de *slots* de frequências ao longo da rota candidata é chamado de caminho de alocação, ao qual são aplicadas restrições de continuidade e contiguidade do espectro para se obter um conjunto de blocos de tamanho suficiente para acomodar a largura de banda requerida pela demanda de conexão. Cada bloco contínuo e contíguo é subdividido em sub-blocos. Um sub-bloco é também um bloco contínuo de mesmo tamanho da banda requerida e que compreende o bloco contínuo livre. Esse conceito de sub-blocos é utilizado em [64].

Na Figura 6 é mostrado como sub-blocos são obtidos a partir de um bloco contínuo de 5 FSS e uma demanda BW_{req} de 2 FSS. O número de sub-blocos NSB é dado por todas as combinações possíveis em que o bloco de largura da banda requerida BW_{req} ocupa dentro da largura do bloco disponível W_k . A quantidade é $NSB = W_k - BW_{req} + 1$. Assim, como no exemplo da Figura 6, o número de sub-blocos (alocações possíveis) de uma demanda de tamanho 2 para um bloco de tamanho 5 é 4. A obtenção dos caminhos e dos blocos de alocação de espectro constituem o caminho de alocação.

No caso de uma rede que adota um esquema de proteção compartilhada, uma conexão solicitada só será aceita se tanto o caminho de trabalho quanto o caminho de proteção forem disjuntos e respeitarem as restrições de continuidade e contiguidade do espectro. No entanto, a proteção compartilhada de recursos em redes ópticas está ainda sujeita a mais restrições. Por esse motivo, o esquema de proteção compartilhada escolhido utiliza a política agressiva de compartilhamento, como apresentado em [32]. A política

Figura 6: Exemplos de sub-blocos formados a partir de um bloco contínuo de 5 FSs e uma demanda de 2 FSs.

agressiva de compartilhamento permite que dois caminhos ópticos de proteção com larguras de banda diferentes compartilhem recursos, desde que os respectivos caminhos de trabalho sejam disjuntos por enlace. O benefício desta política de compartilhamento é que haverá mais oportunidades de compartilhamento, implicando o uso eficiente dos recursos de proteção.

Suponha que $W1$ e $B1$ sejam, respectivamente, o caminho de trabalho e o caminho de proteção (*backup*) da solicitação de conexão $k1$, e que $W2$ e $B2$ sejam, respectivamente, o caminho de trabalho e o caminho de proteção da solicitação de conexão $k2$. Assim, o compartilhamento de proteção das solicitações de conexão $k1$ e $k2$ está sujeito às seguintes restrições:

Restrição 1: $W1$ e $B1$ são disjuntos por enlace; $W2$ e $B2$ são disjuntos por enlace;

Restrição 2: O espectro não pode ser compartilhado entre: $W1$ e $W2$, $W1$ e $B2$, $W2$ e $B1$;

Restrição 3: $B1$ e $B2$ podem compartilhar espectro se e somente se $W1$ e $W2$ forem disjuntos por enlace.

Cabe aqui observar que a restrição de disjunção do algoritmo SBPP é apenas uma das restrições impostas. Por esse motivo, a utilização de um algoritmo dedicado para encontrar dois caminhos disjuntos pode encontrar soluções que violam as demais restrições, tornando-as inválidas. Além dos caminhos disjuntos, as demais restrições

devem ser consideradas. No caso de proteção dedicada seria possível utilizar, por exemplo, o algoritmo de Suurballe [112] que permite encontrar dois caminhos disjuntos com o mesmo custo computacional. No entanto, as restrições que dependem da disjunção dos respectivos caminhos de trabalho e de proteção a ser compartilhado adiciona um nível de complexidade. O algoritmo de Suurballe teria que ser modificado para incorporar as outras restrições e resolver o problema de forma conjunta. Por isso, adotamos nessa tese computar primeiramente o caminho de trabalho e sequencialmente aplicar as restrições no grafo auxiliar que computará o caminho de proteção compartilhada. Dessa forma, a cada solicitação de conexão o algoritmo RSA com proteção compartilhada executa os seguintes passos:

Passo 1: Computa o caminho de trabalho

- a) Calcula os *k-shortest paths* (de hops) para o par de nós origem-destino da solicitação;
- b) Obtém os sub-blocos de alocação ([64]) aplicando as restrições de continuidade e contiguidade de espectro;
- c) Atribui um custo ao caminho de alocação/sub-bloco;
- d) Obtém a alocação de menor custo ou, se indisponível, bloquear a conexão;

Passo 2: Inserir as restrições do esquema *Shared Backup Path Protection* (SBPP) alterando o grafo da seguinte forma:

- a) Elimina todos os *slots* de frequência usados por todos os caminhos de trabalho existentes, incluindo o caminho de trabalho atual;
- b) Se um caminho de trabalho atravessa os mesmos enlaces que o caminho de trabalho atual, elimina todas as subportadoras usadas pelo seu respectivo caminho de proteção;

Passo 3: Computa o caminho de proteção compartilhada da mesma forma que o Passo 1 e, se indisponível, bloquear a conexão.

O princípio de operação do esquema de proteção compartilhada é mostrado na Figura 7. O caminho de trabalho W_1 compreende a rota $v_1 - v_2 - v_3 - v_4$ e o caminho de trabalho W_2 compreende a rota $v_1 - v_5 - v_6 - v_4$. As duas rotas compartilham o caminho de proteção $B_{1,2}$ formado pelos nós $v_1 - v_7 - v_8 - v_4$. Se houver uma falha em qualquer dos caminhos W_1 ou W_2 , o tráfego é restaurado pelo caminho $B_{1,2}$. Os caminhos estão estabelecidos de forma que respeitam as restrições impostas, ou seja, os caminhos W_1 , W_2 e $B_{1,2}$ são mutualmente disjuntos.

Figura 7: Estrutura de dados de uma EON e o processamento de duas solicitações de conexões (CRs) que chegam à rede. O algoritmo aloca às CRs os caminhos de trabalho $W_1[(v_1, v_2, v_3, v_4), (f_1, f_2)]$ e $W_2[(v_1, v_5, v_6, v_4), (f_9, f_{10})]$, e os caminhos de backup $B_{1,2}[(v_1, v_7, v_8, v_4), (f_9, f_{10})]$. O conjunto de slots de frequências satisfaz as restrições de continuidade e contiguidade de espectro.

4.2 Nova Métrica de Fragmentação do Espectro

Nesta subseção é apresentada uma nova métrica de fragmentação espectral chamada *Weighted Cutting Avoid Fragmentation* (WCAFr). O WCAFr proposto nesta tese é baseado no algoritmo FASA [64] porque considera a borda do bloco de alocação. Incluímos dois recursos no algoritmo FASA para melhorar seu desempenho: balanceamento do índice (BI) e fator de proporcionalidade (FP). O BI representa a razão entre a distância (em número de *slots*) de cada extremidade do bloco de alocação até a borda do bloco disponível. O FP considera o tamanho de bloco disponível selecionado pelo tamanho máximo de todos os blocos disponíveis. Para explicar os dois conceitos, vamos considerar a ocupação do link mostrada na Figura 8. O comprimento dos *slots* livres nas adjacências à esquerda (acima na Figura 8) e à direita (abaixo na referida figura) do grupo de alocação candidato (solução viável que ainda não foi alocada) é calculado como ex_1 e ex_2 . O custo básico da fragmentação é calculado por

$$Cost_{base} = \begin{cases} 0 & \text{para } ex_1 = 0 \text{ and } ex_2 = 0 \\ \varepsilon & \text{para } ex_1 \text{ XOR } ex_2 \\ \frac{\min(ex_1, ex_2)}{\max(ex_1, ex_2)} & \text{caso contrário} \end{cases} \quad (4.1)$$

Se $ex_1 = 0$ e $ex_2 = 0$, a alocação do grupo atinge o melhor ajuste e, portanto, o custo é zero. Se $ex_1 = 0$ ou $ex_2 = 0$, o grupo alocado é alinhado à borda e não está "cortando" o grupo em dois. Nesse caso, definimos um custo próximo ao valor mínimo representado pela constante ε . Caso contrário, o grupo alocado custaria uma proporção entre ex_1 e ex_2 . Após o cálculo básico do custo de fragmentação, (4.2) é usado para calcular o fator de custo proporcional do tamanho do bloco contínuo selecionado em comparação com o tamanho máximo de bloco disponível (representado pela variável "MaxSlotsLivresDoGrupo"). Isso significa que quanto menor o tamanho do bloco contínuo, menor o custo. O custo final é calculado por meio de (4.3).

$$Cost_{factor} = \frac{ex_1 + ex_2 + BW_{req}}{MaxSlotsLivresDoGrupo} \quad (4.2)$$

$$Cost_{final} = Cost_{base} \times Cost_{factor} \quad (4.3)$$

A seguir, apresentamos o cálculo do custo dos caminhos da Figura 8 conforme explicado acima. Os valores de cada variável "extent", ex_1 e ex_2 , e do número máximo de slots livres contíguos ("MaxSlotsLivresDoGrupo") em cada enlace para cada caminho são apresentados na Tabela 1. O custo final de cada caminho é a média dos custos de fragmentação dos enlaces $\{l_1, l_2, l_3\}$ dado pela Eq. 4.3.

Figura 8: Exemplo do esquema de fragmentação WCAFr de dois caminhos de alocação formado pelos enlaces $\{\ell_1, \ell_2, \ell_3\}$. Os slots de frequência alocados para o Caminho nº 1 e Caminho nº 2 são $\{f_1, f_2\}$ e $\{f_9, f_{10}\}$, respectivamente. Os valores de ex_1 and ex_2 indicam a quantidade de slots livres de frequência do lado esquerdo e direito, respectivamente, dos slots de frequência alocados nos enlaces. Portanto, $ex_1 = \{0, 2, 1\}$ e $ex_2 = \{0, 0, 2\}$ para o caminho nº 1 e $ex_1 = \{2, 0, 0\}$, e $ex_2 = \{0, 0, 0\}$ para o caminho nº 2.

Tabela 1: Valores ex_1 e ex_2 dos caminhos n.º 1 e n.º 2 para computação da métrica WCAFr

Caminho	Variável	ℓ_1	ℓ_2	ℓ_3
1	ex_1	0	2	1
	ex_2	0	0	2
2	ex_1	2	0	0
	ex_2	0	0	0
1 & 2	Max. cont. slots livres	4	4	5

Primeiramente, calculemos a fragmentação do caminho n.º 1. A seguir, é listado o cálculo do custo base (Eq. 4.1) de cada enlace pertencente ao primeiro caminho:

$Cost_{base_1}^{l_1}$: Como ex_1 e ex_2 são iguais a zero, então $Cost_{base}^{l_1} = 0$;

$Cost_{base_1}^{l_2}$: Como $ex_1 = 2$ e $ex_2 = 0$, então $Cost_{base}^{l_2} = \varepsilon$

$Cost_{base_1}^{l_3}$: Como $ex_1 = 1$ e $ex_2 = 2$, então $Cost_{base}^{l_3} = \min(1, 2)/\max(1, 2) = 1/2 = 0.5$

Em seguida, calcula-se o fator de proporcionalidade $Cost_{factor}$ (Eq. 4.2) de cada enlace.

$$Cost_{factor_1}^{l_1} = \frac{0+0+2}{4} = 0,5;$$

$$Cost_{factor_1}^{l_2} = \frac{2+0+2}{4} = 1;$$

$$Cost_{factor_1}^{l_3} = \frac{1+2+2}{5} = 1.$$

Assumindo $\varepsilon = 0,01$, o custo final é a média dos valores computados pela Eq. 4.3 em cada enlace. Sendo assim, o custo final é $Cost_{final}^1 = (0 \times 0,5 + 0,01 \times 1 + 0,5 \times 1)/3 = 0,51$.

Da mesma forma que o cálculo realizado para o caminho n.º 1, o cálculo do custo de fragmentação do caminho n.º 2 segue os mesmos passos. O custo base para cada enlace é apresentado a seguir:

$Cost_{base_2}^{l_1}$: Como $ex_1 = 2$ e $ex_2 = 0$, então $Cost_{base}^{l_1} = \varepsilon$

$Cost_{base_2}^{l_2}$: Como ex_1 e ex_2 iguais a zero, então $Cost_{base}^{l_2} = 0$

$Cost_{base_2}^{l_3}$: Como ex_1 e ex_2 iguais a zero, então $Cost_{base}^{l_3} = 0$

O fator de proporcionalidade do caminho n.º 2 é calculado a seguir:

$$Cost_{factor_2}^{l_1} = Cost_{factor}^{l_1} = \frac{2+0+2}{4} = 1$$

$$Cost_{factor_2}^{l_2} = Cost_{factor}^{l_2} = \frac{0+0+2}{4} = 0,5$$

$$Cost_{factor_2}^{l_3} = Cost_{factor}^{l_3} = \frac{0+0+2}{5} = 0,4$$

O custo final é $Cost_{final}^2 = (0,01 \times 1 + 0 \times 0,5 + 0 \times 0,4)/3 = 0,01$.

4.3 Modelo *Fuzzy*-RSA Multiobjetivo para Proteção Compartilhada

O algoritmo *fuzzy*-RSA com proteção compartilhada é um algoritmo dinâmico para resolver dois subproblemas: roteamento e alocação que utiliza a abordagem apresentada na Seção 4.1. O item c) do Passo 1 do algoritmo corresponde à etapa de avaliação de custo do caminho de alocação e, nesse modelo, é realizado por meio de sistema de inferência *fuzzy* do tipo Takagi-Sugeno [113]. O caminho de alocação obtido pelos itens a) e b) são avaliados por métricas heurísticas que são usadas como parâmetros de entrada para um sistema de inferência *fuzzy* (*fuzzy inference system*, FIS) e, dessa forma, um custo é calculado. Em seguida, o caminho de alocação de menor custo é selecionado como solução. Nesse trabalho, definimos dois FIS para otimização RSA baseados em critérios multiobjetivos: FIS 1, para computação de caminhos de trabalho e FIS 2 para computação de caminhos de proteção compartilhada. O FIS 2 é uma extensão do FIS 1, conforme explicação a seguir.

O primeiro sistema, FIS 1, descrito na Tabela 2, é responsável pelo cálculo do custo das alocações dos caminhos primários e o segundo sistema, FIS 2, descrito na Tabela 3, pelo cálculo dos custos para alocação dos recursos de proteção compartilhada. A função de pertinência das entradas é mostrada na Fig. 9 e a função de pertinência de saída, na Fig. 10. O conjunto de regras do FIS 1 é apresentado na Tabela 4 e o conjunto de regras adicionais para FIS 2 está na Tabela 5. A implicação da lógica *fuzzy* é por produto algébrico e o processo de *defuzzification* é por média ponderada.

Tabela 2: Funções de pertinência de entrada do FIS 1 para computação dos caminhos de trabalho

Nome	Tipo	Intervalo	Func. de pertinência	Var. linguísticas
in1	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
in2	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
in3	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
cost	Saída	[0,1]	constante	COST0,COST1,...,COST8

Tabela 3: Funções de pertinência de entrada do FIS 2 para computação dos caminhos de proteção compartilhada

Nome	Tipo	Intervalo	Func. de pertinência	Var. linguísticas
in1	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
in2	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
in3	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
in4	Entrada	[0,1]	triangular	BAD,MEDIUM,GOOD
cost	Saída	[0,1]	constante	COST0,COST1,...,COST8

(a) Funções de pertinência da entrada 1

(b) Funções de pertinência da entrada 2

(c) Funções de pertinência da entrada 3

(d) Funções de pertinência da entrada 4

Figura 9: Funções de pertinência do conjunto de entradas do FIS 1 [a,b,c] e FIS 2 [a,b,c,d].

Figura 10: Grau de pertinência de saída.

As regras do modelo FIS 1 são construídas no formato "Se In1 é BAD e In2 é BAD e In3 é BAD então o custo é COST0". O modelo FIS 1 possui três entradas e, para cada entrada, adotam-se três variáveis linguísticas $\{Bad, Medium, Good\}$ definidas como funções de pertinência do tipo triângulo no intervalo $[0, 1]$, como está mostrado na Tabela 2. A saída é mapeada em nove variáveis linguísticas $\{cost0, cost1, \dots, cost8\}$.

Sejam n_{terms} o total de funções de pertinência para cada entrada, i , j e k os índices de cada variável linguística no intervalo $[0, n_{terms} - 1]$, e n_{out} o número de variáveis linguísticas de saída. O indexador de saída out_idx , dadas as posições i, j e k , é calculado pela Eq. 4.4. O conjunto de regras do FIS 1 é construído calculando-se todas as combinações entre i, j e k e suas respectivas saídas out_idx . Na Tabela 4 é mostrado um exemplo das regras e o respectivo conjunto completo é mostrado no Apêndice C.1.

$$out_idx = floor \left\{ (1.3i + j + k) \times \left[\frac{n_{out} - 1.0}{n_{terms} \times (n_{terms} - 1.0)} \right] \right\} \quad (4.4)$$

Tabela 4: Conjunto de regras automáticas para o FIS 1

Regra n ^o	In1	In2	In3	cost
1	B	B	B	0
2	B	B	M	1
3	B	B	G	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
26	G	M	G	4
27	G	G	G	6

Para o modelo FIS 2 adotamos as mesmas regras do modelo FIS 1 e acrescentamos novas regras que são geradas. As regras adicionais do FIS 2 são construídas no formato "**Se** custo é COST0 e In4 é GOOD então custo é COST1". Sejam w o índice da variável linguística da entrada $in4$ e $out2_idx$ o índice da variável linguística de saída modificada, então as novas regras são calculadas pela Eq. 4.5. Na Tabela 5 é mostrado um exemplo das regras e o respectivo conjunto completo é apresentado com detalhes no Apêndice C.2.

$$out2_idx = min \left\{ max \left[out_idx + (w + 1) - round \left(\frac{n_{terms}}{2} \right), 0 \right], (n_{out} - 1.0) \right\} \quad (4.5)$$

Há duas abordagens para a seleção de um caminho de alocação usando a métrica de fragmentação: (1) escolher um caminho óptico que é configurado sobre enlaces que têm custo mínimo baseado na fragmentação atual (antes da alocação); ou (2) escolher um caminho óptico que minimize a fragmentação após a alocação de recursos atual. Dada a característica de nossa nova métrica de fragmentação, consideramos a segunda abordagem.

O modelo proposto combina três entradas para o cálculo do caminho de trabalho (FIS 1) e quatro entradas para o cálculo do caminho da proteção compartilhada (FIS 2). As entradas são definidas como coeficientes numéricos no intervalo $[0, 1]$ e calculados pelas seguintes métricas: (a) fragmentação (*Fragmentation*, Eq. 4.6); (b) primeiro índice de

Tabela 5: Conjunto de regras automáticas para o FIS 2

Regra n ^o	cost	In4	cost2
1	0	B	0
2	0	M	0
3	0	G	1
⋮	⋮	⋮	⋮
26	8	M	8
27	8	G	8

slot livre (*FirstIndex*, Eq. 4.7); (c) disponibilidade de *slot* de enlace (ou balanceamento de carga) (*SlotAvailability*, Eq. 4.8); e (d) nível de compartilhamento (*SharingLevel*, Eq. 4.9). A Tabela 6 define as variáveis utilizadas para o cálculo das métricas. A otimização *fuzzy* RSA proposta nesse trabalho é realizada dinamicamente em três etapas. O algoritmo é uma abordagem aprimorada do algoritmo de duas etapas que separa o roteamento e a alocação em subproblemas com a adição da etapa de avaliação e seleção da lógica difusa. A cada solicitação de conexão, o algoritmo RSA executa o fluxo da Figura 11.

Tabela 6: Definições das variáveis

Parâmetro	Descrição
FR	Medida de fragmentação definida em [64]
index	Índice da primeira posição do bloco contíguo de <i>slots</i> [102]
k	Número de caminhos computados pelo k -shortest path
D	Diâmetro da rede
NFS	Número de FS de um enlace
BW_{req}	Banda requerida pela conexão. Pode ser em número de FS
NFS_{free}^l	Número de <i>slots</i> livres no enlace l
NHOPS	Número de <i>hops</i> de um caminho
MaxSharing	Número máximo de compartilhamentos permitido pelo SBPP
$SharingCounter_l^w$	Número de compartilhamentos no <i>slot</i> w do enlace l

$$Fragmentation = 1 - FR \quad (4.6)$$

$$FirstIndex = 1 - \frac{index}{NFS} \quad (4.7)$$

$$SlotAvailability = \sum_{l \in R} \frac{NFS_{free}^l}{NFS_{total} \times NHOPS} \quad (4.8)$$

$$SharingLevel = 1 - \frac{MaxSharing - (1 + \sum \sum Sharing_i^w)}{MaxSharing} \quad (4.9)$$

Definidos os modelos FIS 1 e FIS 2, o fluxograma do algoritmo para computação e avaliação dos caminhos de trabalho e de proteção compartilhada é apresentado na Fig. 11.

Figura 11: Avaliação de custo de um sub-bloco pelo sistema de inferência *fuzzy*.

4.4 Rede Neural Profunda

Uma rede neural profunda (*deep neural network*, DNN) é uma rede neural com maior profundidade de camadas, que em tese é acima de três camadas, incluídas as camadas de entrada e a saída [114]. A maioria das redes neurais profundas são *feed-forward*, o que significa que fluem em uma direção apenas, da entrada para a saída, e podem ser treinadas por meio do algoritmo *backpropagation* (retropropagação). O *backpropagation* nos permite calcular e atribuir o erro associado a cada neurônio, permitindo ajustar o algoritmo de forma adequada.

A DNN é útil para extrair características complicadas que não podem ser projetadas por expressão analítica ou fórmula fechada. Sua estrutura é composta por várias camadas não lineares e pode fornecer relacionamentos não lineares entre entradas e saídas. O modelo de DNN apresentado a seguir é constituído por uma camada de entrada em duas dimensões, uma camada convolucional 2D (ver seção 4.4.1), 3 camadas totalmente conectadas e uma camada de saída com um único neurônio que representa a função de custo do bloco de alocação. A função de ativação é a unidade linear retificada (ReLU) escolhida pela sua simplicidade e rapidez que proporciona no processo de treinamento. A estratégia de treinamento é a descida gradiente estocástica com momentum (*stochastic gradient descent with momentum*, SGDM). O modelo DNN projetado é mostrado na Fig. 12. A primeira camada oculta é uma camada convolucional com *kernel* de tamanho $[3; 3]$ e 10 filtros para extração de características (seção 4.4.1). Esta camada difere da camada totalmente conectada (*fully connected network*) e exibe algumas características importantes na solução do problema RSA e que veremos a seguir.

4.4.1 Camada Convolucional 2D

Camadas convolucionais são os principais blocos de construção usados em redes neurais para processamento de imagens. Algumas redes podem ser constituídas de apenas redes convolucionais ou mescladas com camadas totalmente conectadas (*fully connected layer*). As camadas convolucionais são preferidas na utilização de segmentação e classificação de imagens, pois elas possuem um número reduzido de conexões sinápticas, o que melhora o desempenho dos algoritmos de treinamento.

Cada camada convolucional contém uma série de filtros conhecidos como núcleos convolucionais (*kernel*). O filtro é uma matriz de inteiros que é aplicada em um subconjunto dos valores de entrada, de mesmo tamanho do *kernel*. Para exemplificar, se tomarmos uma entrada 2D como uma imagem em escala de cinza que tem apenas um canal (uma matriz bidimensional) e um kernel convolucional 3×3 , o kernel se desloca sobre a matriz de entrada movendo-se horizontalmente coluna por coluna, varrendo as primeiras linhas na matriz contendo os valores de pixel das imagens. Em seguida, o kernel desce verticalmente para as fileiras subsequentes. A convolução funciona como a aplicação repetida do mesmo filtro a

Figura 12: Modelo de rede neural profunda para avaliação dos caminhos de alocação com proteção compartilhada.

uma entrada que resulta em um mapa de ativações denominado *feature map*, realçando em uma região a característica a ser detectada. Uma camada convolucional 2D corresponde ao cálculo de filtros convolucionais deslizando sobre a entrada. A camada é convoluída sobre a entrada movendo-se os filtros ao longo da entrada vertical e horizontalmente e calculando-se o produto escalar dos pesos e da entrada e, em seguida, um termo constante de *bias* é adicionado. O passo em que essa janela deslizante do filtro se desloca sobre a imagem é chamada de *stride*.

Uma das consequências da convolução é a redução do tamanho da entrada por causa da largura do filtro. Uma técnica utilizada para obtenção de um *feature map* de mesmo tamanho que a entrada é o preenchimento das bordas com valores nulos, chamada de *padding*. A aplicação do *padding* ajusta o tamanho do *feature map*, porém causa efeito de borda, pois valores nulos não associados ao fenômeno representado pelos valores de entrada são incluídos no cálculo. Outra característica das redes neurais convolucionais é a capacidade de aprender automaticamente uma série de filtros em paralelo para um conjunto de dados de treinamento. A utilização de uma camada convolucional é adequada

para resolver o problema RSA, já que o estado de alocação pode ser representado por uma imagem em que cada pixel representa um valor de estado de alocação do *slot*. Dessa forma, podemos detectar padrões referentes à forma que o espectro está alocado e, assim, detectar fragmentações e quebra de continuidade do espectro.

4.4.2 Construção do Conjunto de Treinamento, Validação e Teste

O objetivo de treinar um modelo DNN é obter uma função de custo geral dado o status de alocação da rede. Em uma abordagem direta, alimentaríamos a DNN com o estado da rede tendo o número total de entradas, N_{in} , que é o número de enlaces multiplicado pelo número de *slots* de frequência por enlace. Para grandes redes, o tamanho da DNN e o processo de treinamento tornam-se impraticáveis porque produzirá um grande número de entradas. Para contornar esse problema, introduzimos um conjunto simplificado de entradas, substituindo o estado da rede inteira pelo estado de alocação de caminho mais longo obtido pelo algoritmo k -caminhos mais curtos para todo par de nós (i, j) . O caminho máximo em qualquer par de nós (i, j) aproxima-se do diâmetro da rede que, por definição, é sempre menor ou igual que o número de nós, sendo que para a maioria das topologias de redes é menor. Em redes reais, o aumento do diâmetro D com relação ao aumento do número de nós N é proporcional a $D \propto \ln(N)$ no melhor caso e $D = N$ no pior caso (topologia em linha). O diâmetro de um grafo (topologia de rede) é a excentricidade máxima de qualquer vértice do grafo. Ou seja, ele é a maior distância entre qualquer par de vértices. Para achar o diâmetro de um grafo, primeiro encontramos o caminho mínimo entre cada par de vértices. O maior comprimento de qualquer um desses caminhos é o diâmetro do grafo. Analogamente, obtemos o diâmetro da rede porém ao invés de usar o menor caminho, utilizamos com base no k -caminhos mais curtos para garantir que o conjunto de entrada projetada comporte outras soluções que tenham comprimento de rota maiores que o diâmetro.

Para efeito de comparação, apresentamos o comprimento máximo dos k -caminhos mais curtos de diferentes topologias de rede na Tabela 7 para um valor arbitrário de $k = 5$. A principal vantagem desta abordagem é que várias topologias de rede podem apresentar o mesmo comprimento máximo dos k -menores caminhos, mesmo que seu número de nós e enlaces sejam diferentes, conforme mostrado na Tabela 7. Por exemplo, as topologias NSFNET, USNET e Geant têm tamanhos diferentes em termos de nós e enlaces. No entanto, elas têm o mesmo comprimento máximo de caminho mais curto, o que significa que a mesma DNN treinada pode ser usada para avaliar seus caminhos de alocação. Assim, dado este *design* de entrada, definimos o conjunto de dados de treinamento com tamanho $[NFSU; MaxHops]$. Cada valor desta matriz de entrada corresponde a um status de *slots*

e é calculado por

$$Input_i = \begin{cases} 0 & \text{se o slot está livre} \\ -1 & \text{se o slot está ocupado} \\ \frac{Sharing_{Count}}{Sharing_{max}} & \text{se o slot está compartilhado} \end{cases} \quad (4.10)$$

A obtenção dos dados de treinamento durante a simulação acontece no momento em que o algoritmo RSA computa as soluções para a demanda de conexão. Na Figura 13 apresentada o fluxograma da obtenção dos caminhos e a caixa em vermelho representa a extração das soluções para o conjunto de treinamento. No fluxograma está implícito que a determinação do caminho de alocação inclui o caminho de trabalho e o caminho de proteção compartilhada.

Figura 13: Fluxograma do algoritmo para obtenção do conjunto de treinamento da DNN.

Tabela 7: Comprimento máximo de 5-menores caminhos de algumas topologias de redes

Topologia	Nós	Enlaces	Comp. máximo de 5-menores caminhos
NFSNET	14	21	8
USNET	24	43	8
Geant 2012 [115]	40	61	8
Canerie [115]	32	41	9
Oxford [115]	20	26	10

4.4.3 Treinamento e operação da DNN para computação RSA de proteção compartilhada

Para treinar a DNN usamos o otimizador gradiente descendente estocástico com *momentum* (*stochastic gradient descent with momentum*, SGDM), que possui dois parâmetros essenciais: a taxa de aprendizado L e *momentum* M e o tamanho de minilote Bs . A taxa de aprendizado é um parâmetro de ajuste em um algoritmo de treinamento de otimização que determina o tamanho do passo que altera os pesos sinápticos em cada iteração para fazê-lo convergir em um mínimo de uma função de perda. Este parâmetro influencia o quão rapidamente a DNN é treinada. Uma taxa de aprendizado muito pequena resulta em um processo de treinamento mais lento, implicando em tempo longo para convergir. Para uma taxa de aprendizado muito grande, o algoritmo pode ultrapassar a solução mínima. Portanto, há uma compensação entre a taxa de convergência e o *overshooting* na definição de uma taxa de aprendizado.

O parâmetro momentum adiciona uma fração da atualização do peso anterior ao atual. Quando o gradiente continua apontando na mesma direção, aumentará o tamanho dos passos dados em direção ao mínimo. O momentum é a aceleração inercial se considerarmos a taxa de aprendizagem que representa a velocidade do processo de treinamento.

O algoritmo do gradiente descendente divide o conjunto de dados de treinamento em pequenos lotes usados para calcular o erro e atualizar os coeficientes do modelo. O minilote é usado para atualizar parâmetros com base em um pequeno lote de gradientes em vez de cada item. Essa estratégia ajuda a reduzir a variância e torna o processo de atualização mais suave. Em [116,117] há um estudo do *trade-off* entre tamanho do minilote e taxa de aprendizado. Quando o tamanho do minilote é maior há falta de generalização e um processo de aprendizado mais lento. Além disso, existe uma alta correlação entre a taxa de aprendizagem e o tamanho do lote. Ao configurar um tamanho de lote maior, o treinamento da rede exibe desempenho melhor ao definir uma alta taxa de aprendizado. No entanto, é recomendado escolher um tamanho de lote pequeno com baixa taxa de aprendizado [116].

O conjunto de dados é subdividido aleatoriamente em três subconjuntos: conjunto de treinamento, de validação e de teste. O conjunto de treinamento é usado para treinar o

modelo e realizar os ajustes de pesos sinápticos e *biases* na DNN. O conjunto de validação é a amostra de dados usada durante o treinamento e fornece uma avaliação imparcial de um ajuste do modelo no conjunto de treinamento durante o ajuste dos hiperparâmetros do modelo. O conjunto de validação é usado para avaliar um determinado modelo e ajustar os hiperparâmetros, como alteração da taxa de aprendizagem ou condição de parada. Finalmente, o conjunto de teste é a amostra de dados usada para fornecer uma avaliação imparcial de um ajuste do modelo final sobre conjunto de dados de treinamento. O conjunto de teste fornece o padrão final usado para avaliar o modelo. Ele só é usado quando um modelo é totalmente treinado (usando os conjuntos de treinamento e validação). Nosso conjunto de dados foi dividido em 70% para treinamento, 15% para validação e 15% para testes. Na Fig. 14 apresenta o progresso de treinamento de rede neural profunda usando o Deep Learning Toolbox do Matlab [118]. O resultado do treinamento atingiu um *root mean square error* (RMSE) de 0,0239 e uma *loss function* [119] de 0,0186.

Figura 14: Progresso do treinamento da rede neural profunda usando Matlab. Cor azul: RMSE (*Root-Mean-Square Error*); e cor vermelha: *Loss Function*.

Cada conexão de entrada gerará muitos caminhos de alocação com seus respectivos custos. Por causa da grande quantidade de dados, o conjunto de dados de treinamento é classificado aleatoriamente e apenas uma porcentagem é escolhida. Foi constatado empiricamente nesta simulação, que para cada cálculo de caminho são geradas mais de 500 soluções temporárias. Considerando 10^6 solicitações de conexões, obtemos cerca de 500×10^6 de conjunto de dados. O treinamento do conjunto de dados completo pode obter

overfitting indesejado em vez de obter uma função generalizada.

Conforme explicado na seção 4.1, em situação de ocorrência de falhas, o tráfego dos caminhos afetados serão comutados para seus respectivos caminhos de *backup*. Ainda assim, eventualmente, nem todas conexões serão restauradas devido ao compartilhamento de recursos em um cenário de múltiplas falhas. Dada a alocação de cada caminho de *backup* compartilhado e sua respectiva taxa de competição de caminho, podemos adicionar uma penalidade à esta entrada de dados para que a DNN que associe cada perfil de compartilhamento a uma taxa de restauração de falha estimada. Dessa forma, o custo gerado pela DNN treinada nos permite escolher uma solução cujo perfil de alocação produz um menor número de competições entre os caminhos de trabalho das conexões afetadas.

4.5 Complexidade computacional

Nesta seção, analisamos a complexidade computacional dos algoritmos usados e propostos neste trabalho usando a conhecida notação $\mathcal{O}$.

4.5.1 Algoritmo para resolver os caminhos de alocação

Os caminhos de alocação são obtidos por meio do cálculo dos k caminhos mais curtos e o cálculo dos sub-blocos para cada caminho. O algoritmo de k caminhos mais curtos é o algoritmo de Yen, no qual a complexidade computacional depende do algoritmo de caminho mais curto usado para calcular os caminhos de spur. Nesse trabalho, optamos pelo algoritmo de Dijkstra com a implementação de heap de Fibonacci porque é assintoticamente mais rápido para lidar com gráficos esparsos. Se N é o número de nós e L é o número de enlaces, a complexidade do algoritmo é $\mathcal{O}(L + N \log N)$. Uma vez que o algoritmo de Yen faz Kl chamadas do algoritmo de Dijkstra para calcular os subcaminhos (em inglês, “*spur path*” ou “*spur route*”), em que $l = 1, 2, \dots, N$ é o comprimento do subcaminho, a complexidade computacional torna-se $\mathcal{O}(KN(L + N \log N))$ [48]. Após a execução do algoritmo de seleção dos caminhos disponíveis, inicia-se a busca dos blocos disponíveis para alocação de recursos nos caminhos escolhidos com base nas restrições de continuidade e contiguidade. O comprimento máximo do caminho é proporcional a N e o número total de sub-blocos é $W - BW_{req} + 1$, em que W é o número de slots e BW_{req} é a largura de banda necessária. Como $BW_{req} \ll W$, o espaço de busca ao longo do espectro é proporcional a W e, portanto, a complexidade da computação de sub-blocos é $\mathcal{O}(NW)$. Esses algoritmos descritos determinam um conjunto de soluções. Para calcular os k caminhos e todos os sub-blocos, a complexidade computacional torna-se $\mathcal{O}(KN(L + N \log N) \times NW) \approx \mathcal{O}(KN^3W)$.

4.5.2 Métricas

A seguir, é descrita a complexidade computacional das métricas utilizadas. As métricas são responsáveis por atribuir um custo para cada solução encontrada. O menor custo determina a melhor solução do conjunto. As métricas são *fragmentation ratio* (FR) [58, 60], *fragmentation entropy* (H) [56, 57], *fragmentation and alignment* (FA) [120], *fragmentation-aware spectrum allocation* (FASA) [64], *wasted-unusable-free ratio* (WUFR) [121] e o algoritmo de fragmentação WCAFr proposto nesse trabalho.

Suponhamos que o espectro esteja dividido em slots e seja organizado em uma lista de blocos subsequentes livres. Dada uma localização de bloco livre, o acesso ao bloco livre anterior e ao próximo é $\mathcal{O}(1)$. Para as métricas FR e H, todos os blocos livres são visitados uma vez. Como o número de blocos livres é proporcional a W e o caminho é proporcional a N , a complexidade desses cálculos é $\mathcal{O}(NW)$. No entanto, o cálculo da entropia utiliza a função logaritmo que é extremamente custosa.

A métrica FA executa dois cálculos. O primeiro para o número de cortes que verifica a adjacência de cada bloco de alocação com operação $\mathcal{O}(1)$ ao longo do caminho, totalizando $\mathcal{O}(N)$. O segundo é computacionalmente mais custoso porque busca todas as combinações de enlaces adjacentes para cada nó pertencente ao caminho. Se $\langle \kappa_{in} \rangle$ e $\langle \kappa_{out} \rangle$ são, respectivamente, os graus médios dos nós de entrada e saída da rede, a complexidade computacional do cálculo do alinhamento é $\mathcal{O}(N \langle \kappa_{in} \rangle \langle \kappa_{out} \rangle)$.

A métrica WUFR calcula as variáveis ‘wasted’, ‘unused’ e ‘free’. A computação ‘wasted’ e ‘free’ varre o espaço de busca do caminho de alocação, e sua complexidade computacional é $\mathcal{O}(NW)$. A variável ‘unused’ é calculada por meio de um vetor de probabilidade para uma única conexão de entrada que pode caber em um determinado intervalo de *slot* livre disponível. O cálculo do vetor de probabilidade requer conhecimento dos tamanhos possíveis das solicitações de conexão de entrada. Se $|BW_{reqs}|$ é o tamanho das conexões de entrada possíveis, então a complexidade computacional do WUFR torna-se $\mathcal{O}(NW + NW|BW_{reqs}|)$.

As métricas WCAFr e FASA buscam blocos de adjacência (limites superior e inferior) do caminho de alocação selecionado. A complexidade de acessar os blocos anterior e posterior dos blocos fixos é $\mathcal{O}(1)$. O FASA calcula o valor da fronteira e o WCAFr calcula o tamanho da lacuna adjacente. Conforme a métrica pesquisa ao longo do caminho, sua complexidade computacional é $\mathcal{O}(N)$.

A Tabela 8 e a Tabela 9 mostram a comparação entre os tempos de execução computacional das métricas de fragmentação espectral analisadas nessa tese. Os comprimento máximo dos K-menores caminhos estão abreviados na tabela com o nome "Comprimento Max. KSP".

Tabela 8: Comparação dos tempos de execução computacional das métricas de fragmentação espectral analisadas nesse trabalho.

Rede Topologia	Nós/ Enlaces	Comprimento Max. KSP	FASA [38]	FR [40]	H [41], [42]
NSFNET	14/21	8	638.651	79224.6	13039.2
USNET	24/43	8	1149.34	199146	36480.6
Geant 2012	40/61	8	1001.77	175687	35323.9
Canerie	32/41	9	594.432	100550	17401.4
Oxford	20/26	10	521.841	97447.1	12316.1

Tabela 9: Comparação dos tempos de execução computacional das métricas de fragmentação espectral analisadas neste trabalho (continuação).

Rede Topologia	Nós/ Enlaces	Comprimento Max. KSP	FA [43]	WUFR [44]	WCAFr (esta tese)
NSFNET	14/21	8	5459.22	16444.9	863.457
USNET	24/43	8	21605.8	32959.4	1385.07
Geant 2012	40/61	8	37265.5	28945.2	1276.93
Canerie	32/41	9	12835.5	15547.1	775.93
Oxford	20/26	10	5387.62	12621.2	674.025

O cálculo das métricas foi realizado em redes de diferentes tamanhos, fixando apenas o número de *slots* de cada uma. As métricas com menor tempo computacional são o FASA e o WCAFr. Como a métrica WCAFr realiza um número maior de operações constantes de ponderação dos tamanhos de alocação, o tempo é ligeiramente maior. No entanto, o número de operações de complexidade constante dentre dos *loops* não altera a complexidade do algoritmo. Assim, as duas métricas têm a mesma complexidade computacional.

4.5.3 Rede neural

O funcionamento da rede neural é uma multiplicação de matrizes o que implica complexidade computacional assintoticamente $\mathcal{O}(N^3)$. Em nossa abordagem, não há camada intermediária com o número de neurônios maior do que a camada de entrada; portanto, o tamanho do problema é baseado no número de neurônios de entrada N . Em uma abordagem de treinamento de rede óptica direta, alimentaríamos a camada de entrada com o estado da rede, que é todo o status do *slot* de frequência da rede. Nesse caso, o número total de entradas N seria o número de *slots* de frequência multiplicado pelo número de enlaces. Em nossa abordagem, o tamanho da camada de entrada é $D \approx \ln N$. Essa redução é incrivelmente significativa porque em redes do mundo real o número de enlaces

é maior do que o número de nós. Portanto, a complexidade computacional da rede neural é $\mathcal{O}(D^3) \rightarrow \mathcal{O}(\ln^3 N)$.

4.5.4 Sistema de inferência *fuzzy*

A complexidade computacional do FIS é dividida em três partes: fuzzificação, mecanismo de inferência e defuzzificação [122]. A parte de fuzzificação é proporcional ao número de entradas N_i e ao número de conjuntos *fuzzy* de entrada N_{if} . A parte do mecanismo de inferência é o núcleo da lógica *fuzzy* e sua complexidade computacional é proporcional a N_i , ao número de regras L e ao número de discretização do universo de saída MOD . A parte de defuzzificação é proporcional a MOD e L . A parte mais significativa em termos de complexidade é o mecanismo de inferência com complexidade $\mathcal{O}(MOD \times N_i \times L)$. Em nossa abordagem FIS, o conjunto de mapas de regras é a combinação entre conjuntos de entradas e saídas; assim, o número de regras cresce exponencialmente com o aumento do número de entradas e saídas.

5 RESULTADOS COMPUTACIONAIS

Este capítulo é dividido em duas seções. Na primeira seção é apresentada a topologia de rede e o processo de geração de tráfego e falhas e os respectivos parâmetros utilizados na simulação de eventos discretos. Na seção seguinte, os resultados são apresentados e analisados por etapa. Primeiro é feita uma comparação entre as métricas de fragmentação e, em seguida, são avaliados os algoritmos RSA-Fuzzy e RSA-DNN, comparando-os com as duas melhores métricas de fragmentação obtidas na etapa anterior.

5.1 Topologia de Rede, Geração de Tráfego e de Falha, e Processamento de Conexão

A Figura 15 mostra a topologia da rede NSFNet, composta por 14 nós e 21 enlaces bidirecionais, que é utilizada nas simulações computacionais. O recurso de espectro adotado de cada link é 200 *slots* de frequência.

Figura 15: Topologia da rede NSFNet composta por 14 nós e 21 enlaces bidirecionais.

As simulações computacionais são baseadas em eventos discretos e são realizadas para avaliar o desempenho dos algoritmos. Os eventos são apresentados aos pares, a saber, chegada/liberação de solicitação de conexão e ocorrência de falha/restauração de enlace. A chegada de solicitação de conexão e a ocorrência de falha são eventos independentes. Os eventos são descritos a seguir.

1. *Chegada de solicitação de conexão.* As chegadas são eventos independentes que obedecem à distribuição de Poisson com taxa média λ_c . O tempo de ocupação

- (holding time) segue distribuição exponencial negativa com média $1/\mu_c$, na qual $\mu_c = 1$ é a taxa de serviço. A solicitação de conexão é definida por $CR(s, d, BW_{req})$, na qual s , d , e BW_{req} são, respectivamente, o nó fonte, o nó destino e a banda requerida. A banda requerida é uniformemente distribuída no conjunto de slots de frequência $\{3, 6, 10\}$, que corresponde ao conjunto de taxas de bits 10 G, 40 G e 100 Gbps, respectivamente, para formato de modulação *binary phase shift keying* (BPSK). Adota-se uma sobrecarga correspondente à banda de guarda de 2 FSs.
2. *Liberação de conexão.* Os recursos utilizados por uma conexão são liberados quando o tempo de ocupação da conexão expira. A taxa média de serviço é μ_c e o encerramento é computado conforme distribuição exponencial negativa com tempo médio de ocupação $1/\mu_c$.
 3. *Ocorrência de falha.* As ocorrências de falhas são eventos independentes que obedecem à distribuição de Poisson com taxa média $\lambda_f = N_f/\Delta t$, na qual N_f é o número de falhas e Δt é o tempo operacional. Este tempo operacional corresponde ao descarte do transiente da probabilidade de bloqueio, que é o período em que a rede dispõe de muitos recursos e poucas conexões, não sendo afetada por falha de enlace. O número total de falhas é $N_f = 1000$ e cada falha desativa simultaneamente 2 enlaces (cenário de falha múltipla).
 4. *Restauração de enlaces.* O enlace afetado permanece com falha durante o tempo médio de inatividade $MDT = 1/\mu_f$, na qual μ_f é a taxa de restauração. Após esse período, o enlace é restaurado ao seu estado original. Para muitos eventos, a disponibilidade da rede com reparo perfeito é $A \equiv MTTF/(MTTF + MDT)$, na qual $MTTF = 1/\lambda_f$ é o tempo médio para falha [123]. Dados A e λ_f , então μ_f é calculado.

Para uma dada carga de tráfego, 10^6 CRs e 10^3 falhas com $A = 0,99$ são geradas de uma vez e são classificadas sequencialmente. Um par de nós (origem e destino) é selecionado aleatoriamente de acordo com distribuição uniforme. A carga de tráfego oferecida à rede é $\sigma = \lambda_c/\mu_c$ e varia de 80 a 500 erlangs com incremento de 20 erlangs. O número de k -caminhos computados pelo algoritmo RSA é $k = 3$ e o número máximo de compartilhamentos permitidos pelo esquema SBPP é $Sharing_{max} = 15$. A Tabela 10 resume os parâmetros usados nas simulações de computador.

Os detalhes de implementação deste modelo de rede e de geração de tráfego está na seção 5.1 - "Modelagem da simulação" em [42].

Tabela 10: Parâmetros para simulações computacionais.

Parâmetro	Símbolo	Valor
Nº total de conexões	CR_{max}	10^5
Nº total de falhas	N_f	10^3
Banda requerida	BW_{req}	$\{3, 6, 10\}$
Carga de tráfego	σ	80 a 500
Taxa média entre chegadas	λ_c	σ
Taxa de serviço	μ_c	1
Tempo operacional	Δt	$CR_{max} \times \lambda_c$
Taxa de ocorrência de falhas	λ_f	$N_f / \Delta t$
Taxa de restauração	μ_f	ver texto
Disponibilidade	A	0,99
Tempo médio para falha	MTTF	$1/\lambda_f$
Tempo médio de inatividade	MDT	$1/\mu_f$
k – menores caminhos computados	k	3
Compartilhamento máximo permitido	$Sharing_{max}$	15

5.2 Análise dos Resultados

Os resultados obtidos na simulação são apresentados nesta seção. Primeiramente é realizada uma comparação entre os algoritmos que utilizam as principais métricas de fragmentação, comparando-os em termos de BP, BSUR e CRR. Os dois melhores resultados desta comparação são utilizados na etapa seguinte em que os algoritmos FIS e DNN são avaliados. As figuras de mérito utilizadas para avaliação são:

- **Probabilidade de bloqueio (*blocking probability*, BP):** é a relação entre o número de conexões bloqueadas e o total de demandas de conexão. Quanto menor a probabilidade de bloqueio, melhor o atendimento das conexões

$$BP = \frac{BlockedConnections}{TotalConnections}$$

- **Taxa de restauração de conexões (*connection restoration ratio*, CRR):** é a relação entre o número de conexões restauradas e o número de conexões afetadas pela falha. Quanto maior CRR, maior a capacidade de sobrevivência da rede

$$CRR = \frac{RestoredConnections}{FailedConnections}$$

- **Relação de compromisso:** é expressa pela relação $CRR \times (1 - BP)$. Quanto maior, mais otimizada é a relação entre BP e CRR.
- **Taxa de utilização do espectro de backup (*backup spectrum utilization ratio*, BSUR):** é a relação entre o total de recursos utilizados para os caminhos de

proteção e o total de recursos utilizados pelos caminhos de trabalho. Quanto menor a BSUR, melhor, pois melhor é a eficiência em compartilhar de recursos

$$BSUR = \frac{BackupSpectrum}{WorkingSpectrum}$$

- **Média de hops dos caminhos (*average hop*, Ahop):** é a média do tamanho dos caminhos de trabalho e de proteção compartilhada, dada em hops.
- **Número médio de blocos livres (*mean number of free blocks*, MNFB):** é a quantidade média dos blocos livres durante toda a simulação para determinada carga de tráfego.
- **Tamanho médio dos blocos livres (*mean free blocksize*, MFBS):** é a média do tamanho dos blocos livres durante toda simulação para determinada carga de tráfego.
- **Tempo de inatividade (*downtime*):** é o tempo estimado em que as conexões são desligadas antes de serem finalizadas normalmente (pela expiração do tempo de duração). Quanto menor o *downtime*, melhor a resposta da rede perante a ocorrência de falhas.

Ressaltamos que exceto a métrica CRR, as demais métricas mencionadas não foram propostas inicialmente para analisar a situação de proteção e restauração, mas sim roteamento e alocação de espectro. Nesta tese, estamos estendendo as métricas para aplicação em um esquema de proteção e restauração. Por serem soluções heurísticas para desfragmentação do espectro, elas têm como objetivo principal reduzir a probabilidade de bloqueio. A análise dos resultados utilizando tais métricas para o caso de proteção compartilhada segue algumas premissas intuitivas:

1. Quanto mais recursos disponíveis na rede, menor a probabilidade de bloqueio;
2. Um número maior de compartilhamentos economiza recursos;
3. Quanto maior o nível de compartilhamento de um *slot* de frequência de proteção, maior a chance de algum caminho de trabalho não conseguir ser restaurado em cenário de múltiplas falhas;
4. As restrições de contiguidade e continuidade podem impedir que uma conexão seja estabelecida, mesmo que os enlaces da rota tenham mais recursos totais disponíveis que a banda requerida pela conexão;
5. Quanto menos fragmentado está o espectro, melhor a disponibilidade de faixas de espectros contínuos e contíguos para atender as demandas de conexão.

Utilizamos os seguintes algoritmos para avaliar o desempenho do algoritmo RSA SBPP proposto: (1) RSA de k -caminhos mais curtos em número de saltos (hops) (KSP) com First-Fit; (2) RSA de k -caminhos mais curtos e fragmentação mínima (MFK) com heurística ciente de fragmentação; (3) RSA-Fuzzy com heurística ciente de fragmentação WCAFr; (4) rede neural profunda RSA-DNN treinada com RSA-Fuzzy. Primeiramente, analisamos (1) e (2) de acordo com as figuras de mérito BP, CRR e BSUR em função da carga de tráfego. Seleccionamos dois dos algoritmos MFK com melhor desempenho e comparamos com os algoritmos propostos em (3) e (4) realizando uma análise completa com todas as métricas supracitadas.

5.2.1 Comparação das Métricas de Fragmentação Espectral

Primeiramente, analisamos a figura de mérito BP (ver Fig. 16), CRR (ver Fig. 18) e BSUR (ver Fig. 19) em função da carga de tráfego para a referência KSP+FF e as seguintes métricas de fragmentação: *fragmentation ratio* (FR) [58,60], *fragmentation entropy* (H) [56,57], *fragmentation and alignment* (FA) [120], *fragmentation-aware spectrum allocation* (FASA) [64], *wasted-unusable-free ratio* (WUFR) [121] e a métrica de fragmentação WCAFr proposta neste trabalho. Os detalhes de implementação de cada métrica estudada são explicados no Apêndice D.

Para analisar as métricas de fragmentação é preciso entender que a probabilidade de bloqueio está relacionada à disponibilidade de espectro para acomodar as demandas de conexão e existem duas formas de prover essa característica. A primeira é economizando espectro ao selecionar caminhos mais curtos utilizando poucos recursos ou aumentar o número de compartilhamentos para proteção. A segunda forma refere-se à disposição e ao alinhamento dos *slots* de frequência ocupados. Essa última pode ser aprimorada selecionando caminhos com balanceamento de carga ou melhorando o estado de fragmentação do espectro alocado. Conciliar o balanceamento, o nível de fragmentação e o posicionamento no espectro dos caminhos de alocações de proteção e de trabalho é um desafio e, por isso, utilizamos um sistema de inferência *fuzzy* para resolver esse problema multiobjetivo.

O aumento do número de compartilhamento reduz a métrica BSUR, porém prejudica a métrica CRR, uma vez que muitos caminhos dependerão de poucos *slots* de frequência para recuperação. Aqui, temos uma relação de compromisso entre BP e BSUR e entre BSUR e CRR. Aumentar o nível de compartilhamento diminui a BP, porém também diminui a CRR, e vice-versa. A forma encontrada de reduzir a BP e aumentar a CRR foi aumentar a BSUR de forma que não comprometesse a disponibilidade de espectro utilizando outros meios como o balanceamento de carga e de minimização da fragmentação espectral.

A referência KSP+FF não possui controle heurístico sobre os compartilhamentos para estabelecer os caminhos de proteção. Dessa forma, o algoritmo aloca descontroladamente recursos para proteção, não otimizando os níveis de compartilhamento do espectro. O resultado é a obtenção de uma BSUR e BP alta. Devido à grande disponibilidade de proteção, a CRR também é elevada.

A Figura 16 mostra a probabilidade de bloqueio (e Figura 17 na escala logarítmica) em função da carga de tráfego para métricas de desfragmentação de espectro. Com relação à BP, a métrica de fragmentação WCAFr e FASA são as duas que apresentam melhor desempenho comparadas à referência KSP+FF. A métrica WCAFr é a que oferece melhor desempenho da taxa de restauração de conexão, como mostrado na Figura 18.

Figura 16: Probabilidade de bloqueio em função da carga de tráfego para métricas de fragmentação de espectro: MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].

Figura 17: Logaritmo da probabilidade de bloqueio em função da carga de tráfego para métricas de fragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].

Figura 18: Relação de restauração de conexão em função da carga de tráfego para métricas de desfragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].

Figura 19: Taxa de utilização de espectro em função da carga de tráfego para métricas de desfragmentação de espectro. MFK FR [58,60], MFK H [56,57], MFK WUFR [121], MFK FA [120], MFK FASA [64].

Segundo a Figura 19, a maioria das métricas aumentam o nível de compartilhamento

proporcionando uma baixa BSUR. Porém, a métrica WCAFr não produz tantos níveis de compartilhamentos, embora proporcione menor probabilidade de bloqueio comparada às demais. Uma BSUR maior obtida pela métrica WCAFr justifica uma maior CRR, pois com mais recursos para proteção alocados, ela melhora a restauração. Se mesmo utilizando mais recursos a WCAFr é capaz de prover uma baixa BP, isso significa uma maior capacidade de desfragmentação de espectro, permitindo assim que mais conexões possam ser atendidas.

Dado que as métricas FASA e WCAFr apresentam melhor desempenho em termos de BP com relação à referência KSP+FF, essas são selecionadas como comparativos para a próxima etapa da análise. Além do mais, como o nosso objetivo é otimizar a taxa de restauração, a métrica WCAFr é selecionada como Entrada 1 do modelo FIS por apresentar melhor CRR.

5.2.2 Análise dos algoritmos FIS e DNN

Os modelos apresentados nessa tese para solução do problema de roteamento e alocação de espectro com proteção compartilhada são denominados nos gráficos de *Fuzzy Inference System* (FIS) e *Deep Neural Network* (DNN). Eles são comparados com a referência KSP+FF e com os dois melhores resultados avaliados na Seção 5.2.1, algoritmos MFK FASA e MFK WCAFr.

A Figura 20 (e Figura 21 em escala logarítmica), a Figura 22 e a Figura 23 estão relacionadas entre si da mesma forma utilizada para análise das métricas de fragmentação. Adicionalmente, vamos analisar a relação de compromisso entre BP e CRR dada pela expressão da relação de compromisso $CRR \times (1 - BP)$ na Figura 24.

Na Figura 20 o modelo FIS obteve a menor probabilidade de bloqueio por conseguir conciliar uma baixa fragmentação com o balanceamento de carga nos enlaces e o alinhamento do espectro alocado. A entrada 1 do FIS representa o nível de fragmentação, a entrada 2 pondera o balanceamento de carga e a entrada 3 indica o posicionamento preferencial. No caso de um caminho de trabalho, esse modelo seleciona preferencialmente as primeiras posições do espectro e, no caso de um caminho de proteção compartilhada, as últimas posições. Por outro lado, a taxa de restauração do modelo FIS (Figura 23) está comprometida porque o FIS 2, que computa os caminhos de proteção compartilhada, utiliza uma quarta entrada, dada pela Equação 4.9, em que o custo do caminho é ajustado preferindo caminhos em que o nível de compartilhamento seja maximizado. Assim, com o aumento do nível de compartilhamento, a BSUR cai quase pela metade comparada à referência KSP+FF, porém compromete a restauração da conexão via proteção compartilhada, devido à pouca quantidade de recursos alocados para proteção.

Figura 20: Probabilidade de bloqueio dos algoritmos analisados: MFK FASA [64], MFK WCAFr, FIS e DNN.

Figura 21: Logarítmo da probabilidade de bloqueio dos algoritmos: MFK FASA [64], MFK WCAFr, FIS e DNN.

Figura 22: Taxa de utilização do espectro dos algoritmos analisados: MFK FASA [64], MFK WCAFr, FIS e DNN.

Por outro lado, o modelo DNN não reduz tanto a probabilidade de bloqueio, porém oferece uma alta taxa de restauração de conexão. Isso se deve pela forma como ele distribui os caminhos de proteção compartilhada para prover maior sobrevivência de rede. O DNN utiliza o custo gerado pelo FIS, o qual tende a maximizar o número de compartilhamentos, e adiciona uma penalidade ao custo do caminho proporcional à quantidade de recursos alocados nos respectivos caminhos de trabalho do espectro a ser compartilhado. Sendo assim, esse algoritmo passa a utilizar mais recursos para proteção, aumentando consideravelmente a BSUR em relação ao FIS.

Figura 23: Taxa de restauração de conexão dos algoritmos: MFK FASA [64], MFK WCAFr, FIS e DNN.

Figura 24: Relação de compromisso CRR(1-BP) dos algoritmos : MFK FASA [64], MFK WCAFr, FIS e DNN.

Como forma de analisar melhor as ponderações entre BP e CRR das métricas de fragmentação e dos algoritmos propostos, apresentamos na Figura 24 uma medida de relação de compromisso (*trade-off*). Essa relação é um coeficiente no intervalo $[0, 1]$ que multiplica a CRR e a taxa de aceitação de conexões, que é o complementar da BP, ou seja é dado por $1 - BP$. Em todo o intervalo de variação de carga de tráfego o algoritmo DNN consegue otimizar esta relação, oferecendo maior taxa de restauração por menor probabilidade de bloqueio. Na região de baixa carga de tráfego, todos os algoritmos, exceto o DNN, apresentam uma relação de compromisso menor do que a referência. Porém, na região de média e alta carga estes mesmos algoritmos convergem para uma relação de compromisso muito próxima da referência.

A Figura 25 apresenta a quantidade média de hops dos caminhos alocados por cada algoritmo. O KSP+FF sempre escolhe a primeira alocação disponível do menor caminho, por isso, ele é o que apresenta menor média de hops. Já o modelo FIS é o que apresenta maior média de hops por causa da heurística que pondera o balanceamento de carga, pois nem sempre o menor caminho é o caminho mais balanceado. A heurística de balanceamento permite que o algoritmo busque por caminhos alternativos menos carregados. O modelo DNN fica em uma posição intermediária pois a penalidade atribuída ao custo dá preferência para caminhos de trabalho mais curtos.

Figura 25: Média de hops dos algoritmos analisados: MFK FASA [64], MFK WCAFr, FIS e DNN

As métricas apresentadas na Figura 26 e na Figura 27 representam o caminho médio dos blocos livres e número médio de blocos livres, respectivamente. Essas duas métricas podem ser interpretadas como métricas indiretas de fragmentação, pois revelam a quantidade de blocos que não puderam ser alocados e o tamanho deles. A partir dessas duas métricas podemos calcular uma outra medida: a quantidade média de recursos remanescentes na rede, ou seja, a capacidade ociosa, mostrada na Figura 28. Analisando a Figura 27 e a Figura 26, observamos que o algoritmo DNN se destaca por apresentar um número muito grande de pequenos blocos livres. Isso significa que o algoritmo otimiza ao máximo o uso dos recursos de rede de forma que sobram apenas fragmentos de *slots* de frequência que não poderiam ser utilizados de qualquer forma devido às restrições.

Figura 26: Número médio de blocos livres dos algoritmos analisados: MFK FASA [64], MFK WCAFr, FIS e DNN

Figura 27: Tamanho médio de blocos livres dos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN

Figura 28: Capacidade ociosa da rede produzida pelos algoritmos analisados: MFK FASA [64], MFK WCAFr , FIS e DNN

A capacidade ociosa são os *slots* de frequência que não foram utilizados. Existem dois motivos da não alocação desses recursos: 1) havia disponibilidade de espectro, mas não houve demanda para alocá-lo; ou 2) houve demanda, mas a conexão foi recusada por violar as restrições de continuidade e/ou contiguidade de espectro. Analisando a Figura 28, a curva de referência KSP+FF é a que produz maior capacidade ociosa, justamente por não otimizar a fragmentação espectral e o uso dos recursos compartilhados como nos demais algoritmos. Em especial, o resultado produzido pelo DNN obtém uma curva de capacidade ociosa ainda mais baixa, porque ele utiliza mais recursos da rede para proteção compartilhada.

A Tabela 11 mostra os valores médios das figuras de mérito BP e CRR e a respectiva variação relativa dos algoritmos avaliados comparados à referência KSP+FF. A variação relativa é dada pela expressão $VR = 100 \times (x - x_{ref})/(x_{ref})$, em que x é o valor a ser avaliado e x_{ref} é o valor da referência.

Tabela 11: BP e CRR médios e variação relativa (VR)

Algoritmo/ Parâmetro	BP Média	CRR Média	VR BP (%)	VR CRR (%)
KSP+FF (Referência)	0,241	0,848	-	-
MFK FASA [64] (essa tese ↓)	0,195	0,794	-19,088	-6,459
MFK WCAFr	0,204	0,838	-15,260	-1,231
FIS	0,179	0,776	-25,845	-8,550
DNN	0,215	0,873	-10,625	+2,884

A partir dos resultados de variação relativa na Tabela 11, podemos afirmar que, na rede testada, para aplicações em que é preferível uma BP menor em detrimento da sobrevivência da rede, o modelo FIS utilizado é o que apresenta o melhor resultado, conseguindo reduzir a BP em 25,85% em relação à referência. Se a preferência for maior sobrevivência de rede, então a melhor escolha é o algoritmo DNN, que reduziu a BP em 10,63% e aumentou a CRR em 2,88% com relação à referência. Analisando as principais métricas, BP, BSUR e CRR, o algoritmo que superou a referência em todos os casos foi o DNN, obtendo uma redução de BP, aumento de CRR e, ainda, reduziu a BSUR em 5,82%, além de promover menor capacidade ociosa da rede por utilizar melhor o espectro para a proteção compartilhada.

Um dos parâmetros utilizados na simulação de eventos discretos é o tempo de manutenção da conexão que define o tempo que a conexão deve permanecer ativa antes de ser finalizada normalmente. Quando ocorre uma falha em um enlace e as conexões afetadas não conseguem ser restauradas via proteção compartilhada devido à competição entre outras conexões afetadas, a conexão é então desconectada antes do tempo estimado

e permanece inativa. Essa diferença de tempo entre a finalização normal esperada e a interrupção do serviço é o tempo de inatividade (*downtime*) das conexões. A Figura 29 mostra o tempo de inatividade das conexões provocadas por falhas na rede. A primeira característica que devemos notar é como a curva do tempo de inatividade da conexão sem proteção cresce bruscamente comparada com as demais curvas com proteção, que possuem uma elevação mais suave ao longo do tempo. Esse fato revela a importância do uso de um esquema de proteção e restauração para a sobrevivência da rede e manutenção da qualidade do serviço.

Figura 29: Tempo de inatividade de conexões produzida pelos algoritmos analisados: MFK FASA [64], MFK WCAFr, FIS e DNN

Dentre os algoritmos que utilizam proteção compartilhada, o DNN apresentou menor tempo de inatividade das conexões, ou seja, restaurou um número maior de conexões afetadas por falhas. Para comparar a inclinação da curva (*slope*) de inatividade de cada método foi feito um ajuste linear dos pontos e a inclinação da curva apresentada na Tabela 12.

A Tabela 12 foi obtida a partir do gráfico da Figura 29 calculando a regressão linear das curvas de tempo de inatividade. A inclinação de cada curva é um indicativo do quanto de tempo de inatividade cada algoritmo proporciona durante o intervalo de simulação. A qualidade dos ajustes são medidos pelos parâmetros de erro padrão e coeficiente de regressão linear R^2 . Um ajuste com erro padrão menor que 10^{-3} é considerado um excelente

Tabela 12: Inclinação do ajuste linear das curvas de *downtime*

Algoritmo	Inclinação	Erro Padrão	Ajuste R^2
Sem Proteção	5,21243	0,0015	0,99257
KSP+FF SPP	0,29775	2,83957E-4	0,92434
MFk FASA [64]	0,28757	1,73447E-4	0,9683
FIS	0,28436	9,26452E-5	0,99054
MFk WCAFr	0,24629	1,2689E-4	0,97667
DNN	0,18532	1,28891E-4	0,95828

ajuste, dado que a amplitude da curva é na escala de 10^3 . Quanto ao coeficiente de regressão R^2 , quanto mais próximo de 1 melhor é o ajuste.

6 CONCLUSÕES

A flexibilidade e escalabilidade da arquitetura EON permitem aumentar as taxas de transmissão e prover maior eficiência espectral. Com isso, alguns paradigmas tiveram de ser atualizados e novos desafios vieram à tona. Por exemplo, a grade flexível e granular de frequências permitiu que os *transponders* pudessem modular individualmente os subcomprimentos de onda, aumentando a eficiência espectral e, conseqüentemente, as taxas de transmissão. Além disso, permitiu também que mais demandas de conexão com largura de banda granular pudessem ser acomodadas, criando um adensamento de conexões passando por um único enlace. Essas características tornam as EONs mais vantajosas do que a arquitetura WDM. Entretanto, isso trouxe uma preocupação maior quanto à sobrevivência da rede, já que uma falha pode causar enorme perda de informação e inúmeras conexões interrompidas.

Uma das técnicas utilizadas para promover a sobrevivência da rede é o esquema de proteção de caminho de *backup* compartilhado (*shared backup path protection*, SBPP). No caso da EON, essa técnica foi estendida com o que foi chamado de política agressiva de compartilhamento, em que o compartilhamento é realizado individualmente em cada *slot* de frequência com o objetivo de aumentar a possibilidade de estabelecer um caminho de proteção compartilhada. A principal vantagem do esquema SBPP é a economia de espectro, que é proporcional ao número de compartilhamentos realizados, e a principal desvantagem é que a restauração da conexão é comprometida em um cenário de múltiplas falhas. Esse problema surge porque poucos recursos de proteção são compartilhados entre grande número de conexões. Diminuir o nível de compartilhamentos melhora a restauração, porém consome mais recursos que poderiam ser utilizados para atender novas demandas de conexão, impactando a probabilidade de bloqueio. Por isso, é necessário um algoritmo específico que controle o nível de compartilhamento de recursos para melhorar a relação de compromisso entre uma baixa probabilidade de bloqueio e uma alta taxa de restauração de falha.

Nessa tese foi proposto um modelo de *deep neural network* (DNN) com um *fuzzy inference system* (FIS) multiobjetivo para resolver o problema RSA-SBPP em EON visando à otimização da proteção por meio do ajuste da relação de compromisso entre a probabilidade de bloqueio e a taxa de restauração de falha. Para isso, propusemos uma métrica para determinar o grau de fragmentação do espectro que provesse uma menor probabilidade de bloqueio e que não impactasse de forma significativa a taxa de restauração. Seu desempenho foi comparado com métricas publicadas na literatura. Usamos essa métrica como uma das entradas FIS para gerar uma função de custo e depois uma DNN foi treinada com esse custo, adicionando uma penalidade para os caminhos de

alocação que representassem maior nível de competição.

Na primeira parte dos resultados computacionais numéricos, comparamos métricas de fragmentação de outros autores com a métrica de fragmentação proposta nesse trabalho. As duas métricas que obtiveram os melhores desempenhos em termos de BP foram a FASA [64] e a WCAFr (essa tese), obtendo, respectivamente, uma redução relativa em relação à referência de 19,09% e 15,26%. Por outro lado, todas as métricas avaliadas tiveram algum impacto na CRR, que no caso da FASA correspondeu a uma redução de 6,46% e o WCAFr de apenas 1,23%.

Na segunda parte, as duas métricas, FASA e WCAFr, foram comparadas com os modelos FIS e DNN e avaliadas principalmente em termos de probabilidade de bloqueio (BP), taxa de restauração de falha (CRR), capacidade ociosa e tempo de inatividade das conexões. As soluções propostas apresentaram bom desempenho. O melhor desempenho em termos de BP foi obtido pelo FIS, que conseguiu conciliar fragmentação, balanceamento de carga nos enlaces e alinhamento dos blocos de alocação no espectro, obtendo assim uma redução de 25,85% de BP em comparação à referência. Já o modelo DNN treinado com base no modelo FIS e otimizado para restauração das conexões, obteve uma redução relativa de 10,63% de BP. Com relação à CRR, o modelo FIS apresentou uma redução de 8,55% e o modelo DNN obteve um aumento relativo de 2,88%. Com base nas curvas da relação de compromisso dada pela expressão $CRR \times (1 - BP)$, o algoritmo DNN obteve elevação da relação de compromisso em relação a todos os outros algoritmos, demonstrando ser uma versão mais otimizada entre BP e CRR. O algoritmo DNN conseguiu produzir melhor taxa de restauração de falhas pois utilizou melhor os recursos de rede para proteção. Tanto é que produziu menor capacidade ociosa de rede, ou seja, o algoritmo DNN conseguiu reservar mais recursos de rede para proteção do que os outros algoritmos.

Todo o processo de otimização da proteção teve reflexo na medida do tempo de inatividade das conexões, que é útil para visualizar o efeito de um esquema de proteção. Considerando como referência o esquema KSP+FF com SBPP, o esquema sem proteção produziu um tempo de inatividade de +1650,6%. Já com a medida mais otimizada de proteção de rede, o modelo DNN conseguiu reduzir o tempo de inatividade em 37,75%.

Traduzindo os resultados em possíveis aplicações de rede é possível afirmar que em aplicações que requerem baixo BP o modelo FIS é o mais adequado. Em aplicações que é estritamente necessário uma alta sobrevivência de rede para cumprir um acordo de nível de serviço (SLA) com o cliente, a solução fornecida pelo modelo DNN foi a que promoveu o melhor resultado de taxa de restauração de rede e também uma melhor relação de compromisso entre probabilidade de bloqueio e taxa de restauração.

REFERÊNCIAS

- [1] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, T. Yoshimatsu, T. Kobayashi, Y. Miyamoto, K. Yonenaga, A. Takada, O. Ishida, and S. Matsuoka, “Demonstration of novel spectrum-efficient elastic optical path network with per-channel variable capacity of 40 gb/s to over 400 gb/s,” in *2008 34th European Conference on Optical Communication*, 2008, pp. 1–2.
- [2] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, “Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies,” *IEEE Communications Magazine*, vol. 47, no. 11, pp. 66–73, November 2009.
- [3] A. A. M. Saleh and J. M. Simmons, “Technology and architecture to enable the explosive growth of the internet,” *IEEE Communications Magazine*, vol. 49, no. 1, pp. 126–132, 2011.
- [4] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, “Elastic optical networking: a new dawn for the optical layer?” *IEEE Communications Magazine*, vol. 50, no. 2, pp. s12–s20, February 2012.
- [5] I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio, and E. Palkopoulou, “A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges,” *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1317–1337, 2014.
- [6] S. Talebi, F. Alam, I. Katib, M. Khamis, R. Salama, and G. N. Rouskas, “Spectrum management techniques for elastic optical networks: A survey,” *Optical Switching and Networking*, vol. 13, pp. 34 – 48, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1573427714000253>
- [7] A. Lord, P. Wright, and A. Mitra, “Core networks in the flexgrid era,” *J. Lightwave Technol.*, vol. 33, no. 5, pp. 1126–1135, Mar 2015. [Online]. Available: <http://jlt.osa.org/abstract.cfm?URI=jlt-33-5-1126>
- [8] M. Jinno, “Elastic optical networking: Roles and benefits in beyond 100-gb/s era,” *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1116–1124, 2017.
- [9] U. Cisco, “Cisco annual internet report (2018–2023) white paper,” 2020. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>

- [10] G. Zhang, M. De Leenheer, A. Morea, and B. Mukherjee, "A survey on ofdm-based elastic core optical networking," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 65–87, 2013.
- [11] H. N. Tan, T. Inoue, K. Tanizawa, T. Kurosu, and S. Namiki, "Optical nyquist filtering for elastic otdm signals: Fundamentals and demonstrations," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1014–1026, 2015.
- [12] M. Imran, P. M. Anandarajah, A. Kaszubowska-Anandarajah, N. Sambo, and L. Potí, "A survey of optical carrier generation techniques for terabit capacity elastic optical networks," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 211–263, 2018.
- [13] P. Zhu, J. Li, Y. Chen, Y. Xu, N. Zhang, B. Guo, Z. Chen, and Y. He, "Comparison of coherent optical single-carrier and multi-carrier superchannel with fdm and fde," *Optical Fiber Technology*, vol. 20, no. 5, pp. 453–460, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1068520014000674>
- [14] S. Shirazipourazad, C. Zhou, Z. Derakhshandeh, and A. Sen, "On routing and spectrum allocation in spectrum-sliced optical networks," in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 385–389.
- [15] Y. Wang, X. Cao, and Y. Pan, "A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks," in *2011 Proceedings IEEE INFOCOM*, 2011, pp. 1503–1511.
- [16] F. Shirin Abkenar and A. Ghaffarpour Rahbar, "Study and analysis of routing and spectrum allocation (rsa) and routing, modulation and spectrum allocation (rmsa) algorithms in elastic optical networks (eons)," *Optical Switching and Networking*, vol. 23, pp. 5 – 39, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1573427716300741>
- [17] G. Shen, H. Guo, and S. K. Bose, "Survivable elastic optical networks: survey and perspective (invited)," *Photonic Network Communications*, vol. 31, no. 1, pp. 71–87, Feb 2016. [Online]. Available: <https://doi.org/10.1007/s11107-015-0532-0>
- [18] M. Sivakumar, R. K. Shenai, and K. M. Sivalingam, *A Survey of Survivability Techniques for Optical WDM Networks*. Boston, MA: Springer US, 2005, pp. 297–331. [Online]. Available: https://doi.org/10.1007/0-387-22584-6_13
- [19] R. Bhandari, *Survivable networks: algorithms for diverse routing*. Springer Science & Business Media, 1999.
- [20] C. S. Ou and B. Mukherjee, *Survivable optical WDM networks*. Springer Science & Business Media, 2005.

-
- [21] H. Saini and A. K. Garg, "Protection and restoration schemes in optical networks: a comprehensive survey," *International Journal of Microwaves Applications*, vol. 2, no. 1, 2013.
 - [22] M. Bakri, M. Koubàa, A. Bouallègue, and M. Gagnaire, "On the improvement of network resource utilization efficiency for the establishment of static dependable connections in srlg-constrained wdm translucent networks," *Photonic Network Communications*, vol. 31, no. 1, pp. 88–111, 2016.
 - [23] R. Goscién, K. Walkowiak, M. Klinkowski, and J. Rak, "Protection in elastic optical networks," *IEEE Network*, vol. 29, no. 6, pp. 88–96, 2015.
 - [24] H. Liu, R. Li, Y. Chen, and X. Wang, "Resource efficiency improved approach for shared path protection in eons," *Photonic Network Communications*, vol. 33, no. 1, pp. 19–25, 2017.
 - [25] S. Yin, S. Huang, Y. Zhou, H. Huang, J. Zhang, and W. Gu, "Survivable routing, spectrum and waveband assignment strategy in cloud optical and data center network," *Photonic Network Communications*, vol. 31, no. 3, pp. 532–542, 2016.
 - [26] K. D. R. Assis, S. Peng, R. C. Almeida, H. Waldman, A. Hammad, A. F. Santos, and D. Simeonidou, "Network virtualization over elastic optical networks with different protection schemes," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 4, pp. 272–281, 2016.
 - [27] R. Goścień, K. Walkowiak, and M. Tornatore, "Survivable multipath routing of anycast and unicast traffic in elastic optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 6, pp. 343–355, 2016.
 - [28] X. Li, S. Huang, S. Yin, B. Guo, Y. Zhao, J. Zhang, M. Zhang, and W. Gu, "Shared end-to-content backup path protection in k-node (edge) content connected elastic optical datacenter networks," *Opt. Express*, vol. 24, no. 9, pp. 9446–9464, May 2016. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-24-9-9446>
 - [29] C. Wang, G. Shen, and L. Peng, "Protection lightpath-based hitless spectrum defragmentation for distance adaptive elastic optical networks," *Optics Express*, vol. 24, no. 5, pp. 4497–4511, 2016.
 - [30] B. Chen, J. Zhang, Y. Zhao, C. Lv, W. Zhang, S. Huang, X. Zhang, and W. Gu, "Multi-link failure restoration with dynamic load balancing in spectrum-elastic optical path networks," *Optical Fiber Technology*, vol. 18, no. 1, pp. 21 – 28, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1068520011001398>

- [31] A. Tarhan and C. Cavdar, “Shared path protection for distance adaptive elastic optical networks under dynamic traffic,” in *2013 5th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2013, pp. 62–67.
- [32] X. Shao, Y. K. Yeo, Z. Xu, X. Cheng, and L. Zhou, “Shared-path protection in ofdm-based optical networks with elastic bandwidth allocation,” in *OFC/NFOEC*, March 2012, pp. 1–3.
- [33] G. Shen, Y. Wei, and S. K. Bose, “Optimal design for shared backup path protected elastic optical networks under single-link failure,” *Journal of Optical Communications and Networking*, vol. 6, no. 7, pp. 649–659, 2014.
- [34] A. L. F. Lourenço and A. C. César, “Algorithm for shared path protection in elastic optical network based on spectrum partition,” in *2017 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, Aug 2017, pp. 1–5.
- [35] W. B. Jia, Z. Q. Xu, Z. Ding, and K. Wang, “An efficient routing and spectrum assignment algorithm using prediction for elastic optical networks,” in *2016 International Conference on Information System and Artificial Intelligence (ISAI)*, June 2016, pp. 89–93.
- [36] E. Barbosa, Í. B. Brasileiro, R. Rabêlo, A. Soares, and J. Junior, “Nova abordagem para o problema de roteamento nas redes opticas elasticas utilizando sistema fuzzy,” *Simpósio Brasileiro de Automação Inteligente-SBAI*, 2017.
- [37] U. Bhanja and D. Mishra, “Quality of service aware fuzzy dynamic routing and wavelength assignment technique in all optical networks,” *Photonic Network Communications*, vol. 34, no. 2, pp. 155–169, Oct 2017. [Online]. Available: <https://doi.org/10.1007/s11107-017-0689-9>
- [38] M. Zhang, B. Xu, X. Li, M. Yin, B. Wu, and K. Qiu, “Deep neural network-based soft-failure detection and failure aware routing and spectrum allocation for elastic optic networks,” *Optical Engineering*, vol. 58, no. 6, p. 066107, 2019.
- [39] E. Kavlakoglu, “Ai vs. machine learning vs. deep learning vs. neural networks: What’s the difference?” May 2020. [Online]. Available: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>
- [40] S. Sugiyama, *Human Behavior and Another Kind in Consciousness: Emerging Research and Opportunities: Emerging Research and Opportunities*, ser. Advances in Human and Social Aspects of Technology (2328-1316). IGI Global, 2019. [Online]. Available: <https://books.google.com.br/books?id=9CqQDwAAQBAJ>

-
- [41] Y. Kurtgoz and E. Deniz, “Comparison of ann, regression analysis, and anfis models in estimation of global solar radiation for different climatological locations,” in *Exergetic, Energetic and Environmental Dimensions*. Elsevier, 2018, pp. 133–148.
 - [42] A. L. F. Lourenço, “Técnicas de proteção e restauração em redes ópticas elásticas,” Ph.D. dissertation, Universidade de São Paulo, 2017.
 - [43] Y. Wang, X. Cao, and Y. Pan, “A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks,” in *2011 Proceedings IEEE INFOCOM*, April 2011, pp. 1503–1511.
 - [44] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, and A. Hirano, “Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network [topics in optical communications],” *IEEE Communications Magazine*, vol. 48, no. 8, pp. 138–145, August 2010.
 - [45] G. Zhang, M. De Leenheer, A. Morea, and B. Mukherjee, “A survey on ofdm-based elastic core optical networking,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 65–87, 2013.
 - [46] N. Sambo, P. Castoldi, A. D’Errico, E. Riccardi, A. Pagano, M. S. Moreolo, J. M. Fàbrega, D. Rafique, A. Napoli, S. Frigerio, E. H. Salas, G. Zervas, M. Nolle, J. K. Fischer, A. Lord, and J. P. F. P. Giménez, “Next generation sliceable bandwidth variable transponders,” *IEEE Communications Magazine*, vol. 53, no. 2, pp. 163–171, Feb 2015.
 - [47] S. Yang and F. Kuipers, “Impairment-aware routing in translucent spectrum-sliced elastic optical path networks,” in *2012 17th European Conference on Networks and Optical Communications*, June 2012, pp. 1–6.
 - [48] K. Christodoulopoulos, I. Tomkos, and E. A. Varvarigos, “Routing and spectrum allocation in ofdm-based optical networks with elastic bandwidth allocation,” in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Dec 2010, pp. 1–6.
 - [49] —, “Elastic bandwidth allocation in flexible ofdm-based optical networks,” *Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354–1366, May 2011.
 - [50] H. Beyranvand and J. A. Salehi, “A quality-of-transmission aware dynamic routing and spectrum assignment scheme for future elastic optical networks,” *Journal of Lightwave Technology*, vol. 31, no. 18, pp. 3043–3054, Sept 2013.
 - [51] T. Takagi, H. Hasegawa, K. Sato, Y. Sone, B. Kozicki, A. Hirano, and M. Jinno, “Dynamic routing and frequency slot assignment for elastic optical path networks

- that adopt distance adaptive modulation,” in *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, March 2011, pp. 1–3.
- [52] J. Zhao, Q. Yao, X. Liu, W. Li, and M. Maier, “Distance-adaptive routing and spectrum assignment in ofdm-based flexible transparent optical networks,” *Photonic Network Communications*, vol. 27, no. 3, pp. 119–127, Jun 2014. [Online]. Available: <https://doi.org/10.1007/s11107-014-0432-8>
- [53] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C. A. Bunge, and B. Spinnler, “Reach-dependent capacity in optical networks enabled by ofdm,” in *2009 Conference on Optical Fiber Communication - includes post deadline papers*, March 2009, pp. 1–3.
- [54] X. Wang, J. Kim, S. Yan, M. Razo, M. Tacca, and A. Fumagalli, “Blocking probability and fairness in two-rate elastic optical networks,” in *2014 16th International Conference on Transparent Optical Networks (ICTON)*, July 2014, pp. 1–4.
- [55] X. Wang, Q. Zhang, I. Kim, P. Palacharla, and M. Sekiya, “Utilization entropy for assessing resource fragmentation in optical networks,” in *OFC/NFOEC*, March 2012, pp. 1–3.
- [56] P. Wright, M. C. Parker, and A. Lord, “Simulation results of shannon entropy based flexgrid routing and spectrum assignment on a real network topology,” in *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*. IET, 2013, pp. 1–3.
- [57] —, “Minimum-and maximum-entropy routing and spectrum assignment for flexgrid elastic optical networking,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 1, pp. A66–A72, 2015.
- [58] M. Zhang, W. Lu, Z. Zhu, Y. Yin, and S. J. B. Yoo, “Planning and provisioning of elastic o-ofdm networks with fragmentation-aware routing and spectrum assignment (rsa) algorithms,” in *2012 Asia Communications and Photonics Conference (ACP)*, Nov 2012, pp. 1–3.
- [59] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, “Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing,” *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 15–22, Jan 2013.
- [60] R. Wang and B. Mukherjee, “Spectrum management in heterogeneous bandwidth optical networks,” *Optical Switching and Networking*, vol. 11, pp. 83 – 91, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1573427713000799>

-
- [61] K. Christodoulopoulos, I. Tomkos, and E. Varvarigos, "Dynamic bandwidth allocation in flexible ofdm-based networks," in *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, March 2011, pp. 1–3.
 - [62] Y. Sone, A. Hirano, A. Kadohata, M. Jinno, and O. Ishida, "Routing and spectrum assignment algorithm maximizes spectrum utilization in optical networks," in *2011 37th European Conference and Exhibition on Optical Communication*, Sept 2011, pp. 1–3.
 - [63] T. Fukuda, L. Liu, K. i. Baba, S. Shimojo, and S. J. B. Yoo, "Fragmentation-aware spectrum assignment for elastic optical networks with fully-distributed gmpls," in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2015, pp. 1–3.
 - [64] Y. Xu and Y. C. Kim, "Dynamic routing and spectrum allocation to minimize fragmentation in elastic optical networks," in *2017 20th Conference of Open Innovations Association (FRUCT)*, April 2017, pp. 512–518.
 - [65] H. Chen, H. Yang, B. Luo, and L. Yan, "A joint fairness-aware and fragmentation-reduction spectrum allocation scheme in elastic optical networks," in *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, July 2017, pp. 1–2.
 - [66] A. N. Patel, P. N. Ji, J. P. Jue, and T. Wang, "Defragmentation of transparent flexible optical wdm (fwdm) networks," in *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, March 2011, pp. 1–3.
 - [67] A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J. P. Fernández-Palacios, and D. Careglio, "Dynamic routing and spectrum (re)allocation in future flexgrid optical networks," *Computer Networks*, vol. 56, no. 12, pp. 2869 – 2883, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128612001764>
 - [68] R. Wang and B. Mukherjee, "Provisioning in elastic optical networks with non-disruptive defragmentation," *Journal of Lightwave Technology*, vol. 31, no. 15, pp. 2491–2500, Aug 2013.
 - [69] F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Poti, and P. Castoldi, "Push-pull defragmentation without traffic disruption in flexible grid optical networks," *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 125–133, Jan 2013.

- [70] T. Takagi, H. Hasegawa, K. i. Sato, Y. Sone, A. Hirano, and M. Jinno, "Disruption minimized spectrum defragmentation in elastic optical path networks that adopt distance adaptive modulation," in *2011 37th European Conference and Exhibition on Optical Communication*, Sept 2011, pp. 1–3.
- [71] N. A. Jackman, S. H. Patel, B. P. Mikkelsen, and S. K. Korotky, "Optical cross connects for optical networking," *Bell Labs Technical Journal*, vol. 4, no. 1, pp. 262–281, Jan 1999.
- [72] R. Dutta and G. N. Rouskas, "Traffic grooming in wdm networks: past and future," *IEEE Network*, vol. 16, no. 6, pp. 46–56, Nov 2002.
- [73] O. A. Gerstel, "Flexible use of spectrum and photonic grooming," in *Integrated Photonics Research, Silicon and Nanophotonics and Photonics in Switching*. Optical Society of America, 2010, p. PMD3. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=PS-2010-PMD3>
- [74] B. Kozicki, H. Takara, Y. Tsukishima, T. Yoshimatsu, T. Kobayashi, K. Yonenaga, and M. Jinno, "Optical path aggregation for 1-tb/s transmission in spectrum-sliced elastic optical path network," *IEEE Photonics Technology Letters*, vol. 22, no. 17, pp. 1315–1317, Sept 2010.
- [75] G. Zhang, M. D. Leenheer, and B. Mukherjee, "Optical traffic grooming in ofdm-based elastic optical networks [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. B17–B25, Nov 2012.
- [76] Y. Zhang, X. Zheng, Q. Li, N. Hua, Y. Li, and H. Zhang, "Traffic grooming in spectrum-elastic optical path networks," in *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, March 2011, pp. 1–3.
- [77] P. D. Choudhury, N. Agarwal, and T. De, "Spectrum and splitter utilization efficient traffic grooming routing and spectrum assignment in elastic optical networks," in *2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, Aug 2017, pp. 1–6.
- [78] S. Zhang, C. Martel, and B. Mukherjee, "Dynamic traffic grooming in elastic optical networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 1, pp. 4–12, January 2013.
- [79] M. Liu, M. Tornatore, and B. Mukherjee, "Survivable traffic grooming in elastic optical networks - shared protection," *Journal of Lightwave Technology*, vol. 31, no. 6, pp. 903–909, March 2013.

-
- [80] J. Zhang, Y. Ji, M. Song, Y. Zhao, X. Yu, J. Zhang, and B. Mukherjee, "Dynamic traffic grooming in sliceable bandwidth-variable transponder-enabled elastic optical networks," *Journal of Lightwave Technology*, vol. 33, no. 1, pp. 183–191, Jan 2015.
 - [81] R. Tian, Y. Zhao, J. Zhang, X. Yu, Y. Li, C. Yu, J. Zhang, C. Liu, and G. Zhang, "Dynamic traffic grooming based on auxiliary graph in spatial division multiplexing enabled elastic optical networks," in *2016 15th International Conference on Optical Communications and Networks (ICOON)*, Sept 2016, pp. 1–3.
 - [82] Y. Wu, W. Hou, L. Guo, Y. Liu, and Z. Sun, "Green grooming in elastic optical networks," in *OFC 2014*, March 2014, pp. 1–3.
 - [83] M. Hadi and M. R. Pakravan, "Energy-efficient fast configuration of flexible transponders and grooming switches in ofdm-based elastic optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 2, pp. 90–103, Feb 2018.
 - [84] N. Sambo, A. D'Errico, C. Porzi, V. Vercesi, M. Imran, F. Cugini, A. Bogoni, L. Potì, and P. Castoldi, "Sliceable transponder architecture including multiwavelength source," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 6, no. 7, pp. 590–600, July 2014.
 - [85] D. Richardson, J. Fini, and L. Nelson, "Space-division multiplexing in optical fibres," *Nature Photonics*, vol. 7, no. 5, p. 354, 2013.
 - [86] R. Asthana, Y. N. Singh, and W. D. Grover, "p-cycles: An overview," *IEEE Communications Surveys Tutorials*, vol. 12, no. 1, pp. 97–111, First 2010.
 - [87] X. Guo, J. Huang, H. Liu, and Y. Chen, "Efficient p-cycle combination protection strategy based on improved genetic algorithm in elastic optical networks," *IET Optoelectronics*, vol. 12, no. 2, pp. 73–79, 2018.
 - [88] X. Chen, M. Zhou, S. Zhu, S. Kang, L. Sun, and Z. Zhu, "Optimizing fipp- p -cycle protection design to realize availability-aware elastic optical networks," *IEEE Communications Letters*, vol. 22, no. 1, pp. 65–68, Jan 2018.
 - [89] K. Walkowiak, M. Klinkowski, B. Rabeiga, and R. Goścień, "Routing and spectrum allocation algorithms for elastic optical networks with dedicated path protection," *Optical Switching and Networking*, vol. 13, pp. 63 – 75, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1573427714000150>
 - [90] M. Aibin and K. Walkowiak, "Adaptive survivability algorithm for path protection with various traffic classes in elastic optical networks," in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, 2015, pp. 56–62.

- [91] —, “Different strategies for dynamic multicast traffic protection in elastic optical networks,” in *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2016, pp. 174–180.
- [92] M. Aibin and K. Walkowiak, “Adaptive survivability algorithm for path protection with various traffic classes in elastic optical networks,” in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, Oct 2015, pp. 56–62.
- [93] F. Cugini, N. Sambo, T. Foggi, M. Ruiz, L. Velasco, and P. Castoldi, “Signal overlap for efficient 1+1 protection in elastic optical networks (eons),” in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2017, pp. 1–3.
- [94] C. Wang, G. Shen, and S. K. Bose, “Distance adaptive dynamic routing and spectrum allocation in elastic optical networks with shared backup path protection,” *Journal of Lightwave Technology*, vol. 33, no. 14, pp. 2955–2964, July 2015.
- [95] C. Wang, G. Shen, B. Chen, and L. Peng, “Protection path-based hitless spectrum defragmentation in elastic optical networks: Shared backup path protection,” in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2015, pp. 1–3.
- [96] M. Liu, M. Tornatore, and B. Mukherjee, “Survivable traffic grooming in elastic optical networks - shared path protection,” in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 6230–6234.
- [97] C. Rocha and B. Jaumard, “Revisiting p-cycles / fipp p-cycles vs. shared link / path protection,” in *2008 Proceedings of 17th International Conference on Computer Communications and Networks*, Aug 2008, pp. 1–6.
- [98] Y. Sone, A. Watanabe, W. Imajuku, Y. Tsukishima, B. Kozicki, H. Takara, and M. Jinno, “Highly survivable restoration scheme employing optical bandwidth squeezing in spectrum-sliced elastic optical path (slice) network,” in *2009 Conference on Optical Fiber Communication - includes post deadline papers*, March 2009, pp. 1–3.
- [99] —, “Bandwidth squeezed restoration in spectrum-sliced elastic optical path networks (slice),” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, no. 3, pp. 223–233, March 2011.
- [100] A. L. F. Lourenço, “Técnicas de proteção e restauração em redes ópticas elásticas,” dissertação de mestrado, Universidade de São Paulo, 2015.
- [101] E. O. Figueirôa, E. A. O. Lima, C. J. A. Bastos-Filho, J. C. da Silva, A. V. S. Xavier, and D. R. B. Araújo, “A routing algorithm based on fuzzy logics for elastic optical networks,” in *2017 IEEE 18th International Conference on High Performance Switching and Routing (HPSR)*, June 2017, pp. 1–6.

-
- [102] I. B. Brasileiro, A. C. B. Soares, J. V. dos Reis, and A. L. Sanches, “A fuzzy solution to routing problem in elastic optical networks,” in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, July 2016, pp. 1–4.
 - [103] M. G. Simões and I. S. Shaw, *Controle e modelagem fuzzy*. Editora Blucher, 2007.
 - [104] M. Andrade and M. A. P. Jaques, “Estudo comparativo de controladores de mamdani e sugeno para controle de tráfego em interseções isoladas,” *TRANSPORTES*, vol. 16, no. 2, 2008.
 - [105] J. Y. Yen, “An algorithm for finding shortest routes from all source nodes to a given destination in general networks,” *Quarterly of Applied Mathematics*, vol. 27, no. 4, pp. 526–530, 1970.
 - [106] N. Kojic, I. Reljin, and B. Reljin, “Routing in optical networks by using neural network,” in *2006 8th Seminar on Neural Network Applications in Electrical Engineering*, Sept 2006, pp. 65–68.
 - [107] C. J. A. Bastos-Filho, R. A. Santana, D. R. C. Silva, J. F. Martins-Filho, and D. A. R. Chaves, “Hopfield neural networks for routing in all-optical networks,” in *2010 12th International Conference on Transparent Optical Networks*, June 2010, pp. 1–4.
 - [108] I. DA SILVA, D. SPATTI, and R. FLAUZINO, *REDES NEURAIAS ARTIFICIAIS PARA ENGENHARIA E: CIENCIAS APLICADAS - CURSO PRATICO*. ARTLIBER, 2010. [Online]. Available: <https://books.google.com.br/books?id=w2VHbwAACAAJ>
 - [109] R. Fullér, *Neural fuzzy systems*. Lectures, 1995.
 - [110] A. Fontana, “An artificial neural network to find correlation patterns among an arbitrary number of variables,” *arXiv preprint arXiv:1606.06564*, 2016.
 - [111] Z. Ahmed and S. Kumar, “Pearson’s correlation coefficient in the theory of networks: A comment,” *arXiv preprint arXiv:1803.06937*, 2018.
 - [112] J. W. Suurballe and R. E. Tarjan, “A quick method for finding shortest pairs of disjoint paths,” *Networks*, vol. 14, no. 2, pp. 325–336, 1984.
 - [113] Matlab®, “Mamdani and sugeno fuzzy inference systems - matlab & simulink,” <https://www.mathworks.com/help/fuzzy/types-of-fuzzy-inference-systems.html>, (Accessed on 05/14/2021).
 - [114] “Ai vs. machine learning vs. deep learning vs. neural networks: What’s the difference? | ibm,” <https://www.ibm.com/cloud/blog/>

- [ai-vs-machine-learning-vs-deep-learning-vs-neural-networks](#), (Accessed on 01/12/2021).
- [115] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 9, pp. 1765–1775, october 2011.
 - [116] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, 2020.
 - [117] K. Shen, "Effect of batch size on training dynamics | by kevin shen | mini distill | medium," June 2018. [Online]. Available: <https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e>
 - [118] Matlab®, "Deep learning toolbox documentation," 2020. [Online]. Available: <https://www.mathworks.com/help/deeplearning/>
 - [119] —, "Create a regression output layer - matlab regressionlayer," 2020. [Online]. Available: <https://www.mathworks.com/help/deeplearning/ref/regressionlayer.html>
 - [120] Y. Yin, M. Zhang, Z. Zhu, and S. Yoo, "Fragmentation-aware routing, modulation and spectrum assignment algorithms in elastic optical networks," in *Optical Fiber Communication Conference*. Optical Society of America, 2013, pp. OW3A–5.
 - [121] F. Pederzoli, D. Siracusa, A. Zanardi, G. Galimberti, D. La Fauci, and G. Martinelli, "Path-based fragmentation metric and rsa algorithms for elastic optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 3, pp. 15–25, 2019.
 - [122] Y. H. Kim, S. C. Ahn, and W. H. Kwon, "Computational complexity of general fuzzy logic control and its simplification for a loop controller," *Fuzzy Sets and Systems*, vol. 111, no. 2, pp. 215–224, 2000.
 - [123] M. Rausand, H. Arnljot *et al.*, *System reliability theory: models, statistical methods, and applications*. John Wiley & Sons, 2004, vol. 396.
 - [124] L. A. Zadeh, "Fuzzy sets," in *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.
 - [125] Matlab®, "Foundations of fuzzy logic - matlab & simulink," <https://www.mathworks.com/help/fuzzy/foundations-of-fuzzy-logic.html>, (Accessed on 05/13/2021).

Apêndices

APÊNDICE A – NOÇÕES DE LÓGICA FUZZY

Os conjuntos *fuzzy*, introduzidos por Zadeh [124], são um meio de representar e manipular dados que são imprecisos, ou seja, difusos ou nebulosos. A lógica *fuzzy* fornece uma ferramenta matemática para capturar as incertezas associadas aos processos cognitivos humanos, como o pensamento e o raciocínio. As abordagens convencionais da representação do conhecimento carecem de meios para representar o significado dos conceitos difusos. Como consequência, as abordagens baseadas na lógica de primeira ordem ou em probabilidade clássica não fornecem uma estrutura conceitual apropriada para lidar com esse tipo de representação do conhecimento, uma vez que tal conhecimento é, por natureza, impreciso e não-lexical.

A lógica *fuzzy* (difusa) começa com o conceito de um conjunto difuso. Um conjunto difuso é um conjunto sem um limite nítido e claramente definido. Ele pode conter elementos com apenas um grau parcial de associação, ou grau de pertinência. Um dos exemplos mais comumente usados de um conjunto difuso é o conjunto de pessoas altas. Nesse caso, existem todas as alturas potenciais, digamos de um metro a três metros, e a variável altura corresponderia a uma curva que define o grau em que qualquer pessoa é alta.

Se analisarmos o conceito de altura segundo a lógica booleana podemos classificar as pessoas em “alta” e “não alta”, o que é um absurdo, pois o conceito de “alto” é subjetivo, uma altura que é considerada alta para um grupo pode não ser para outro. No primeiro gráfico da Figura A.1 é mostrada a altura sendo separada em duas classes, “alta” e “não alta”, por uma função de pertinência μ que é uma função degrau. Já no segundo gráfico, é mostrada uma curva que varia suavemente de “não alta” a “alta”. O eixo de saída é um número conhecido como valor de pertinência entre 0 e 1. A curva é conhecida como função de pertinência (*membership function*, MF) e costuma receber a designação μ . A transição suave permite determinar diferentes graus de altura, ou graus de pertinência. O grau de pertinência indica quão um determinado valor de altura pertence à classe “alta” [125].

Figura A.1: Lógica clássica vs lógica *fuzzy* aplicadas ao conceito de altura [125].

A única condição que uma função de pertinência deve satisfazer é que deve variar entre 0 e 1. Sendo assim, um conjunto *fuzzy* é uma extensão de um conjunto clássico. Se X é o domínio de quantificação e seus elementos são denotados por x , então um conjunto *fuzzy* A em X é definido como um conjunto de pares ordenados do tipo

$$A = \{x, \mu_A(x) | x \in X\} \quad (\text{A.1})$$

na qual $\mu_A(x)$ é a função de pertinência (MF) de x em A . A função de associação mapeia cada elemento de X para um valor de associação entre 0 e 1. Resumidamente, x é a variável de entrada de uma lógica *fuzzy* e "Alto" é a variável linguística em que cada valor de altura é mapeado em um grau de pertinência à variável "Alto".

A lógica *fuzzy* é um superconjunto da lógica booleana padrão, em outras palavras, se forem mantidos os valores difusos em seus extremos de 1 (totalmente verdadeiro) e 0 (totalmente falso), as operações lógicas padrão serão mantidas. Então, os valores de

pertinência intermediários precisam ser computados com operações que não contradizem os seus extremos. Para isso, há operações equivalentes para as operações lógicas binárias "AND", "OR" e "NOT". Sejam A e B limitados ao intervalo $(0, 1)$, então as operações equivalentes são:

- Para A AND B é a função $\min(A, B)$;
- Para A OR B é a função $\max(A, B)$;
- Para NOT A é a operação $(1 - A)$;

Além dos conjuntos e operadores *fuzzy* terem as suas próprias definições, que diferem da lógica clássica, as regras condicionais do tipo “se-então” também têm particularidades. Uma única regra *fuzzy* “se-então” assume a seguinte forma:

“Se x é A então y é B ”,

na qual A e B são valores linguísticos definidos por conjuntos *fuzzy* nos domínios de quantificação X e Y , respectivamente. A parte se da regra " x é A " é chamada de antecedente ou premissa, enquanto a parte da regra " y é B " é chamada de consequente ou conclusão. Um exemplo de tal regra pode ser

“Se o serviço for bom, a gorjeta é mediana.”

O conceito bom é representado como um número entre 0 e 1 e, portanto, o antecedente é uma interpretação que retorna um único número entre 0 e 1. Por outro lado, a média é representada como um conjunto *fuzzy* e, portanto, o consequente é uma atribuição que atribui o conjunto *fuzzy* inteiro B para a variável de saída y . Na regra se-então, a palavra é usada de duas maneiras inteiramente diferentes, dependendo se aparece no antecedente ou no consequente. A interpretação de uma regra “se-então” envolve duas etapas:

1. Avaliação do antecedente - Fuzzificação das entradas e aplicação de quaisquer operadores *fuzzy* necessários;
2. Aplicação do resultado ao consequente.

A segunda etapa é conhecida como implicação. Para uma regra “se-então”, o antecedente, p , implica o consequente, q . Na lógica binária, se p for verdadeiro, então q também é verdadeiro ($p \implies q$). Na lógica *fuzzy*, se p é verdadeiro em algum grau de pertinência, então q também é verdadeiro no mesmo grau ($0,5p \implies 0,5q$). Nos dois casos, se p for falso, o valor de q é indeterminado.

O antecedente de uma regra pode ter várias partes. Exemplo: “Se o serviço for bom E a comida for boa, a gorjeta é alta.”

Um sistema *fuzzy* pode conter várias regras e estas precisam ser agregadas, ou seja, combinadas de alguma maneira para produzir uma saída. A agregação ocorre apenas uma vez para cada variável de saída, antes da etapa final de defuzzificação. A entrada do processo de agregação é a lista de funções de saída truncadas retornadas pelo processo de implicação para cada regra.

A entrada para o processo de defuzzificação é o conjunto fuzzy de saída agregada e a saída é um único número. Por mais que a imprecisão ajude a avaliação da regra durante as etapas intermediárias, a saída final desejada para cada variável é geralmente um único número. No entanto, o agregado de um conjunto difuso abrange uma gama de valores de saída e, portanto, deve ser defuzzificado para obter um único valor de saída do conjunto. Os métodos tradicionais de defuzzificação são: centróide, bissetriz, meio do máximo (a média do valor máximo do conjunto de saída), maior do máximo e menor do máximo. O método de defuzzificação mais popular é o cálculo do centróide, que retorna o centro da área sob o conjunto difuso agregado.

Recapitulando, um sistema *fuzzy* é composto pelas seguintes partes (vide Fig. A.2):

1. *Fuzzy knowledge base*: consiste em um conjunto de regras do tipo “IF-THEN” e funções de pertinência do conjunto *fuzzy* utilizados pelas regras;
2. *Fuzzification*: esse processo converte um valor de entrada em uma variável linguística usando uma função de pertinência;
3. *Inference engine*: utiliza as regras *fuzzy* para converter as variáveis linguísticas de entrada em variáveis linguísticas de saída;
4. *Defuzzification*: converte o conjunto *fuzzy* de saída em um valor de saída.

Figura A.2: Diagrama de um sistema *fuzzy* contendo os blocos funcionais: *fuzzifier*, *fuzzy knowledge base*, *inference engine* e *defuzzifier*.

APÊNDICE B – NOÇÕES DE REDES NEURAIS

As redes neurais artificiais (RNAs) são modelos matemáticos que, de forma simplificada, simulam o comportamento do cérebro. As RNAs funcionam como redes paralelas de computação distribuída e, ao contrário dos programas computacionais convencionais, a maioria das redes neurais deve ser treinada para aprender novas associações, novas dependências funcionais e novos padrões.

A unidade base da RNA é o neurônio artificial que é representado por um nó como mostrado na Fig B.1. Cada conexão representa, de forma simplificada, uma sinapse neural e pode transmitir um sinal de um neurônio para o outro. Às conexões sinápticas são atribuídos pesos w_n que podem ser positivos (excitatórios) ou negativos (inibitórios), e se ajustam à medida que a aprendizagem prossegue. O valor de saída do neurônio corresponde à soma ponderada das entradas. Além do mais, o valor de saída tem um limiar de ativação tal que somente se o sinal agregado ultrapassar esse limiar é que o sinal é enviado. Normalmente, os neurônios artificiais são organizados em camadas. Diferentes camadas podem executar diferentes tipos de transformações em suas entradas.

Figura B.1: Representação do neurônio. As entradas X_n são ponderadas pelos pesos sinápticos W_n subtraído pelo limiar de ativação θ e aplicados a uma função de ativação f .

Dada uma entrada qualquer, a saída de uma camada da rede neural é a multiplicação matricial da entrada com a matriz de pesos sinápticos seguida de uma função de ativação não-linear. Uma rede neural contendo várias camadas efetua repetidamente esse mesmo tipo de operação, sendo que a saída da camada anterior se torna entrada da camada seguinte, até a obtenção do resultado na última camada de saída. Esta operação é chamada de *feedforward*, em que as entradas são computadas no sentido direto até a obtenção da

saída. No processo de treinamento o algoritmo realiza as operações de ajuste nos pesos sinápticos seguindo o sentido reverso (*backpropagation*). Nesta etapa os pesos são ajustados utilizando uma função de erro que compara o valor obtido com o esperado e o erro é propagado da última camada para a primeira.

APÊNDICE C – REGRAS *FUZZY*

Tabela C.1: Conjunto de regras automáticas para o FIS 1

Regra nº	In1	In2	In3	cost
1	B	B	B	0
2	B	B	M	1
3	B	B	G	2
4	B	M	B	1
5	B	M	M	2
6	B	M	G	4
7	B	G	B	2
8	B	G	M	4
9	M	G	G	5
10	M	B	B	1
11	M	B	M	3
12	M	B	G	4
13	M	M	B	3
14	M	M	M	4
15	M	M	G	5
16	M	G	B	4
17	M	G	M	5
18	G	G	G	7
19	G	B	B	3
20	G	B	M	4
21	G	B	G	6
22	G	M	B	4
23	G	M	M	6
24	G	M	G	7
25	G	G	B	6
26	G	G	M	7
27	G	G	G	8

Tabela C.2: Conjunto de regras automáticas para o FIS 2

Regra nº	cost	In4	cost2
1	0	B	0
2	0	M	0
3	0	G	1
4	1	B	0
5	1	M	1
6	1	G	2
7	2	B	1
8	2	M	2
9	2	G	3
10	3	B	2
11	3	M	3
12	3	G	4
13	4	B	3
14	4	M	4
15	4	G	5
16	5	B	4
17	5	M	5
18	5	G	6
19	6	B	5
20	6	M	6
21	6	G	7
22	7	B	6
23	7	M	7
24	7	G	8
25	8	B	7
26	8	M	8
27	8	G	8

APÊNDICE D – MÉTRICAS DE FRAGMENTAÇÃO DE ESPECTRO

Nesse apêndice, apresentamos resumidamente algumas propostas reportadas na literatura [39], [41]-[45] para calcular a fragmentação espectral em EONs e que são utilizadas nessa tese para comparação de desempenho dos algoritmos propostos. O objetivo dessa descrição dos fundamentos dos algoritmos é facilitar a leitura do texto.

D.1 Fragmentation Ratio (FR) [58,60]

O algoritmo atribui uma quantidade de *slots* igual à taxa máxima de dados suportada por cada bloco de *slots* livres e contíguos. A relação de fragmentação de um caminho é

$$FR = 1 - \frac{\sum_i v(G_i)}{v(\sum_i G_i)} \quad (D.1)$$

na qual G_i é o i – éximo bloco de *slots* livres e contíguos; $v(G_i)$ é a taxa de dados máxima que o bloco G_i suporta; $\sum_i v(G_i)$ é o somatório das taxa valor dos blocos disponíveis e $v(\sum_i G_i)$ é a taxa disponível do somatório de todos os blocos livres contínuos.

D.2 Fragmentation Entropy (H) [56,57]

A entropia de fragmentação (H) é baseada na entropia de Shannon. A entropia mede o número de informações necessárias para descrever um conjunto de blocos de alocação. Existem dois tipos de entropia. O primeiro é baseado apenas no número de *slots* não usados e o segundo considera os FSs usados e não usados. A figura de mérito da entropia de fragmentação é

$$H = - \sum_{i=1}^N \frac{D_i}{D} \ln \left(\frac{D_i}{D} \right) \quad (D.2)$$

na qual D é o número total de *slots* no espectro de um enlace ou caminho, D_i é número de *slots* do i – éximo bloco contíguo do espectro, e N é o número de blocos.

D.3 Fragmentation and Alignment (FA) [120]

A métrica é uma combinação de medição de fragmentação com utilização de espectro e um alinhamento com enlaces adjacentes. A primeira etapa é calcular o número de cortes causados pelo caminho óptico candidato. Um corte é gerado quando um caminho divide o espectro disponível em dois blocos. Para cada caminho óptico, o algoritmo calcula o número de saltos (H), o número de *slots* atribuídos (S) e o número de *slots* disponíveis (C). O alinhamento é calculado pareando cada nó conectado a um nó correspondente pertencente ao caminho e contando o número de *slots* não utilizados aplicando a restrição

de continuidade. A figura de mérito é

$$A = \frac{S \times N_{hop} + F_{cut} + M_{align}}{C} \quad (D.3)$$

na qual F_{cut} é o número de cortes e M_{align} é o aumento do alinhamento, que é o custo adicional causado pelo número de cortes do desalinhamento relacionado à fragmentação espectral.

D.4 Fragmentation-aware spectrum allocation (FASA) [64]

Essa métrica faz parte do algoritmo RSA *fragmentation-aware spectrum allocation* (FASA). O custo de um *slot* $C^{u,v}(i)$ do nó u ao nó v é 1 se ele estiver ocioso ou 0 se ele estiver ocupado. O bloco de alocação do nó u ao nó v é o índice inicial i e o comprimento do bloco de alocação contínua n . Adotando N o número total de *slots* de frequência, o custo de fragmentação de cada bloco contíguo, $B^{u,v}(i, n)$, é

$$C^{u,v}(i) = \begin{cases} 0 & \text{se ocupado} \\ 1 & \text{se ocioso} \end{cases} \quad (D.4)$$

$$B^{u,v}(i, n) = \begin{cases} C^{u,v}(i + n) = 1 & \text{se } i = 1 \\ C^{u,v}(i - 1) + C^{u,v}(i + n) & \text{se } 1 < i < N - n + 1, N \\ C^{u,v}(i - 1) & \text{se } i < N - n + 1 \end{cases} \quad (D.5)$$

D.5 Wasted-Unusable-Free Ratio (WUFR) [121]

A WUFR é a relação entre a soma do número de *slots* desperdiçados, W , e a soma do número de *slots* não utilizados, U , e o número de *slots* livres (F) no caminho,

$$WUFR = \begin{cases} \frac{W+U}{F} & \text{if } F > 0 \\ 0 & \text{if } F = 0 \end{cases} \quad (D.6)$$

A disponibilidade de espectro em um enlace é um vetor binário, no qual 0 significa que um *slot* está livre e 1 que está sendo usado por uma conexão óptica. A soma de todos os vetores binários dos enlaces de um determinado caminho fornece um vetor inteiro S que representa a disponibilidade de *slots* nesse caminho. Os valores de cada elemento $s \in S$ são, por construção, confinados no intervalo $[0, NL]$, no qual NL é o número de enlaces que compõem o caminho em estudo. Os *slots* desperdiçados, W , conforme especificados em (D.7), são a soma dos elementos do vetor de disponibilidade que não são 0 nem NL , ou

seja, em termos mais intuitivos, são a contagem dos *slots* que estão livres apenas em alguns dos enlaces desse caminho e, portanto, não podem ser usados para servir uma conexão,

$$WU = \begin{cases} \sum_{s \in S} & \text{if } 0 < s < N_L \\ 0 & \text{caso contrário} \end{cases} \quad (\text{D.7})$$

Os *slots* não utilizados, U , são dados por (D.8), na qual H é o conjunto de todos os blocos de *slots* de frequência disponíveis no caminho; a cardinalidade de um de seus elementos, $|h|$, é o tamanho em *slots* de frequência. A função p codifica a probabilidade de que uma única conexão de entrada possa caber naquele bloco disponível h . Observe que o cálculo de p requer conhecimento dos tamanhos possíveis das solicitações de conexão de entrada,

$$U = \sum_{h \in H} |h| \times N_L \times [1 - p(|h|)] \quad (\text{D.8})$$

Finalmente, o termo de *slots* livres F , dado por (D.9), é a contagem de *slots* livres nos enlaces do caminho e pode ser obtido a partir de S somando, sobre os elementos $s \in S$, o número de enlaces no caminho, N_L menos s , que, por si só, conta os enlaces em que um determinado *slot* está ocupado,

$$U = \sum_{s \in S} (N_L - s) \quad (\text{D.9})$$