



Débora Andréa de Oliveira Santos

**Decomposição em Programação Matemática
Aplicada a Redes Computacionais Verdes**

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio.

Orientador : Prof. Luiz Alencar Reis da Silva Mello
Co-Orientador: Prof. Marco Antonio Grivet Mattoso Maia

Rio de Janeiro
Abril de 2015



Débora Andréa de Oliveira Santos

**Decomposição em Programação Matemática
Aplicada a Redes Computacionais Verdes**

Tese apresentada como requisito parcial para obtenção do grau de
Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica
do Departamento de Engenharia Elétrica do Centro Técnico
Científico da PUC-Rio.

Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Luiz Alencar Reis da Silva Mello
Orientador

Centro de Estudos em Telecomunicações /PUC-Rio

Prof. Marco Antonio Grivet Mattoso Maia
Co-Orientador

Centro de Estudos em Telecomunicações /PUC-Rio

Prof. Abilio Pereira de Lucena Filho
UFRJ

Prof. Helios Malebranche Olbrisch Freres Filho
UFRJ

Prof. Alexandre Street de Aguiar
Departamento de Engenharia Elétrica /PUC-Rio

Prof. Marcus Vinicius Soledade Poggi de Aragão
Departamento de Informática /PUC-Rio

Prof. Sinesio Pesco
Departamento de Matemática /PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de Janeiro, 10 de Abril de 2015

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Débora Andréa de Oliveira Santos

Graduou-se em Engenharia Elétrica com ênfase em Eletrônica no CEFET-RJ em 1999. Concluiu o Mestrado em Engenharia Elétrica com área de concentração em Sistemas de Comunicações pela PUC-Rio em 2001. Publicou artigos nacionais e internacionais nas áreas de Redes Verdes e de Reconhecimento de Voz em presença de Ruído. Como funcionária da Petrobras participou de vários cursos e congressos na área de Redes de Telecomunicações. Nesta empresa atuou diversos anos como suporte à operação das redes de dados, voz e vídeo presentes no Brasil e no exterior, sendo responsável também pela implantação e manutenção do *backbone* IP/MPLS desta empresa. Atualmente integra a equipe de planejamento, sendo uma das responsáveis pelo plano de implantação do protocolo IPv6 na rede da Petrobras e pela elaboração de um plano diretor que definirá os serviços e redes de telecomunicações que doravante suportarão os Negócios da referida Companhia.

Ficha Catalográfica

Santos, Débora Andréa de Oliveira

Decomposição em Programação Matemática Aplicada a Redes Computacionais Verdes / Débora Andréa de Oliveira Santos; orientador: Luiz Alencar Reis da Silva Mello; co-orientador: Marco Antonio Grivet Mattoso Maia. – 2015.

163 f. : il. (color) ; 29,7 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2015.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Decomposição. 3. Redes Verdes. 4. Benders. 5. Engenharia de Tráfego. I. Mello, Luiz Alencar Reis da Silva. II. Maia, Marco Antonio Grivet Mattoso. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

A minha amada e venerada família,
tão companheira, presente e compreensiva
e a minha saudosa avó Conceição,
que, sem dúvida, continua olhando por mim.

Agradecimentos

A **Deus**, primeiramente e acima de tudo, por carregar-me sempre em seus braços e por permitir que eu chegasse ao fim de mais esta etapa de minha vida.

A minha santinha de devoção **Nossa Senhora de Fátima**, que nunca deixa de zelar por mim e cujo doce olhar me acalmava nas muitas horas de desespero e de angústia.

Aos meus amados pais, **Jeremias** e **Gloria Marta**, pelo simples fato de existirem. Tudo que sou e que consegui devo a eles.

Às minhas queridas irmãs, **Kátia Cristina** e **Shirley**, que, a cada dia, apresentam-me provas incontestáveis de que sem elas a vida não teria a menor graça.

Ao meu supersobrinho, **Eduardo Henrique**, que consegue ser mais poderoso que o Punho de Ferro.

Ao engenheiro **Jorge Machado Lisboa**, que tanto me motivou, antes e durante o Doutorado, sendo muito mais do que um ídolo e mentor para mim.

Aos secretários do CETUC e do DEE, em particular, a **Ana Maria de Sousa e Sérgio** e a **Maria Alcina Portes**, cujo auxílio e incentivo foram decisivos para que eu pudesse transpor as pedras que surgiram no caminho.

Aos professores que participaram da Comissão Examinadora, em especial, ao professor **Alexandre Street de Aguiar**, por ajudar a desfazer algumas nuvens que teimavam em encobrir o sol.

À **PUC-Rio**, por ser uma excelente instituição de ensino e por propiciar o ambiente ideal para o desenvolvimento acadêmico e científico.

À **Petrobras**, por me proporcionar crescer profissionalmente e pelo patrocínio; e ao meu gerente **Antonio Carlos Guimarães de Faria**, por ser um exemplo, fonte de inspiração e por garantir os meios e as condições para que eu cursasse o Doutorado.

E, em especial, ao meu orientador **Luiz Alencar Reis da Silva Mello** e ao meu co-orientador **Marco Antonio Grivet Mattoso Maia**, pela paciência, por mostrarem-se sempre tão solícitos e por conduzirem-me de forma ímpar. Eles personificam o verbo “orientar” e sem ambos, com certeza absoluta, eu não teria realizado este sonho.

Resumo

Santos, Débora Andréa de Oliveira; Mello, Luiz Alencar Reis da Silva; Maia, Marco Antonio Grivet Mattoso. **Decomposição em Programação Matemática Aplicada a Redes Computacionais Verdes**. Rio de Janeiro, 2015. 163p. Tese de Doutorado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O crescente consumo de energia já se tornou uma preocupação mundial e atualmente mais de quarenta países estão envolvidos em pesquisas e programas para criar mecanismos para economizá-la. No presente trabalho é tratado o problema de Engenharia de Tráfego com base na energia (em inglês, *energy-aware Traffic Engineering*) aplicado ao *backbone* de uma rede IP que utiliza como protocolo de roteamento um SPF (*Shortest Path First*), como o OSPF (*Open Shortest Path First*), por exemplo. Na abordagem proposta são considerados os problemas de desligamento de nós (roteadores) e circuitos, para a economia de energia; e da garantia de um nível de máxima utilização dos circuitos, para assegurar os requisitos de QoS. Para a resolução do problema de otimização, em lugar de adotarem-se métodos heurísticos, propõe-se o tratamento direto por meio de decomposição de Benders, segmentando um problema complicado e de elevada carga computacional em vários menores cuja resolução é mais simples e cuja convergência é mais rápida.

Palavras-chave

Decomposição; Benders; Redes Verdes; Engenharia de Tráfego.

Abstract

Santos, Débora Andréa de Oliveira; Mello, Luiz Alencar Reis da Silva (Advisor); Maia, Marco Antonio Grivet Mattoso (Co-Advisor). **Decomposition in Mathematical Programming Applied to Computational Green Networks**. Rio de Janeiro, 2015. 163p. PhD Thesis — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The growing energy consumption has already become a global concern and currently more than forty countries are involved in researches and programs in order to create mechanisms to save it. This work deals with the energy-aware Traffic Engineering problem applied to the backbone of an IP network in which the used routing protocol is a SPF (Shortest Path First) one, such as OSPF (Open Shortest Path First), for example. The proposed approach considers the problem of switching-off nodes (routers) and circuits, for energy saving; and it also considers the problem of ensuring a maximum utilization level by the circuits, towards to assure QoS requirements. In order to solve the optimization problem, rather than adopting heuristic methods, we propose the direct processing by means of Benders decomposition, crumbling a complicated and hard to solve problem into several smaller ones whose resolution is more simple and whose convergence is faster.

Keywords

Decomposition; Benders; Green Networks; Traffic Engineering.

Sumário

1	Introdução	15
1.1	Organização do Trabalho	23
2	Estratégias em Redes Verdes	25
2.1	Consolidação de Recursos	26
2.2	Virtualização	28
2.3	Computação Proporcional	29
2.4	Desconexão Seletiva	29
2.5	Trabalhos Relacionados	30
2.6	Motivação	37
3	Programação Matemática em Redes IP	39
3.1	Engenharia de Tráfego Baseada em Economia de Energia	47
4	Decomposição Aplicada a Engenharia de Tráfego	58
4.1	Parâmetros e Variáveis Envolvidos	63
4.2	Restrições Envolvidas	65
4.2.1	Restrições de Conservação de Fluxo	65
4.2.2	Restrição de Capacidade dos Circuitos	67
4.2.3	Restrições ECMP	68
4.2.4	Restrições de Roteamento pelo Caminho Mais Curto	69
4.2.5	Restrições de Condição do Nó ou do Circuito	71
4.3	Formulação do Método	71
4.4	Decomposição Aplicada ao Problema de Otimização com Foco em Eficiência	75
4.5	Decomposição Aplicada ao Problema de Otimização com Foco em Economia de Energia	87
4.6	Resultados Computacionais	94
5	Conclusão	99
6	Referências Bibliográficas	102
A	Técnicas ECMP	118
B	Teoria da Dualidade	121
C	Condições KKT	130
D	Método de Decomposição de Benders	133
E	Problemas Mestre e Escravo	140
F	Aplicação Encadeada do Método de Decomposição de Benders	144

Lista de figuras

Figura 1.1	Custo de energia elétrica por <i>MWh</i>	17
Figura 1.2	Valores estimados de consumo global de eletricidade por equipamentos de TIC e por usuários no setor residencial distribuídos por região, no período de 1990 a 2030	18
Figura 1.3	Gráfico ilustrando a crise energética enfrentada pelo Brasil	19
Figura 1.4	Análise comparativa entre a evolução da capacidade dos roteadores, o crescimento do volume de tráfego (Lei de Moore) e o aumento da eficiência energética dos semicondutores (Lei de Escalabilidade de Dennard), no período de 1993 a 2010	19
Figura 2.1	<i>Backbone</i> de uma rede IP interligando diversas redes, sem emprego de técnicas de Consolidação de Recursos: todos os nós e circuitos ativos	27
Figura 2.2	<i>Backbone</i> de uma rede IP interligando diversas redes, empregando técnicas de Consolidação de Recursos: ativos somente os nós e circuitos necessários	27
Figura 2.3	Aproveitamento de capacidade: uma máquina fazendo o papel de três	28
Figura 2.4	<i>Switch</i> atuando como <i>proxy</i> para o tráfego de pacotes ARP	30
Figura 2.5	Método da colônia de formigas	34
Figura 2.6	Relação entre economia de energia e resiliência	37
Figura 3.1	Representação esquemática de um problema de Pesquisa Operacional	40
Figura 3.2	Grafo não orientado, com 5 nós e 7 arestas	40
Figura 3.3	Problema das sete pontes de Königsberg	41
Figura 3.4	Grafo com pesos representando a rede Internet2 em abril de 2009	43
Figura 3.5	Algumas estruturas de problemas suscetíveis a decomposição	46
Figura 3.6	Exemplos de topologias de rede extraídas do SNDlib	51
Figura 3.7	Representação da regra ECMP	53
Figura 4.1	Comportamento típico do tráfego em um circuito de <i>backbone</i>	59
Figura 4.2	Definição do SPF numa rede cujo IGP é o OSPF	60
Figura 4.3	Consumo de energia para diferentes configurações do roteador Cisco 12000	62
Figura 4.4	Representação esquemática das restrições de conservação de fluxo	67
Figura 4.5	Representação esquemática da restrição de capacidade dos circuitos	68
Figura 4.6	Representação esquemática das restrições ECMP	69
Figura 4.7	Representação esquemática das restrições de roteamento pelo caminho mais curto	70

Figura 4.8	Representação esquemática das restrições de condição do nó ou do circuito	72
Figura 4.9	Matriz de restrições para o problema com foco em desempenho	74
Figura 4.10	Matriz de restrições para o problema com foco em economia de energia	74
Figura 4.11	Matriz de restrições para o problema com foco em eficiência ou desempenho, após rearranjos	76
Figura 4.12	Relação entre os elementos da matriz de restrições e as variáveis do problema de otimização simplificado, com foco em eficiência	77
Figura 4.13	Representação esquemática da matriz \underline{V}_2 - Foco em eficiência	78
Figura 4.14	Representação esquemática da matriz \underline{T}_1 - Foco em eficiência	78
Figura 4.15	Representação esquemática da matriz \underline{T}_2 - Foco em eficiência	78
Figura 4.16	Relação entre os elementos da matriz de restrições e as variáveis no Passo 1 do Benders-0 , viabilizando o Benders-1 - Foco em eficiência	80
Figura 4.17	Representação esquemática da matriz \underline{V}_{2h} - Foco em eficiência	80
Figura 4.18	Representação esquemática da matriz \underline{T}_{2w} - Foco em eficiência	81
Figura 4.19	Representação esquemática da matriz \underline{T}_{2h} - Foco em eficiência	81
Figura 4.20	Relação entre os elementos da matriz de restrições e as variáveis no Passo 1 do Benders-1 - Foco em eficiência	83
Figura 4.21	Matriz de restrições no Passo 1 do Benders-2 - Foco em eficiência	86
Figura 4.22	Aplicação encadeada do método de decomposição de Benders	87
Figura 4.23	Matriz de restrições para o problema com foco em economia de energia, após rearranjos	88
Figura 4.24	Relação entre os elementos da matriz de restrições e as variáveis do problema de otimização simplificado, com foco em economia de energia	89
Figura 4.25	Representação esquemática da matriz \underline{V}_2 - Foco em economia de energia	89
Figura 4.26	Representação esquemática da matriz \underline{T}_1 - Foco em economia de energia	90
Figura 4.27	Representação esquemática da matriz \underline{T}_2 - Foco em economia de energia	91
Figura 4.28	Relação entre os elementos da matriz de restrições e as variáveis no Passo 1 do Benders-0 , viabilizando o Benders-1 - Foco em economia de energia	92
Figura 4.29	Relação entre os elementos da matriz de restrições e as variáveis no Passo 1 do Benders-1 - Foco em economia de energia	93
Figura 4.30	Rede Nano	95
Figura 4.31	Rede Spider	95
Figura 4.32	Rede Penta	95
Figura 4.33	Topologias resultantes para a solução ótima obtida para a rede Spider nos dois diferentes cenários testados	98
Figura A.1	ECMP como mecanismo de balanceamento de carga	119
Figura A.2	ECMP como mecanismo de proteção	119

Figura B.1	Relação entre as variáveis do Problema Primal e as restrições do Problema Dual	123
Figura B.2	Relação entre as restrições do Problema Primal e as variáveis do Problema Dual	123
Figura B.3	Auto-inversibilidade da dualidade	125
Figura D.1	Representação esquemática do algoritmo de Benders	139
Figura F.1	Representação esquemática da matriz de restrições	147
Figura F.2	Visão geral do método proposto	163

Lista de tabelas e quadros

Quadro 1.1	Considerações para o cálculo do consumo de energia numa análise comparativa de <i>switches</i> de diferentes fabricantes	20
Quadro 1.2	Consumo de potência, em Watts, dos <i>switches</i> Avaya ERS 4548GT-PWR, HP E4500G-48G-PoE, Cisco WS-3750-X-48P-S e Juniper EX4200-48P sujeitos a diferentes cargas de tráfego	21
Quadro 1.3	Comparação de custos totais inerentes ao consumo de energia de quatro diferentes modelos de <i>switches</i> , ao longo de três anos	21
Quadro 3.1	Exemplos de emprego de grafos	42
Quadro 3.2	Resultados computacionais obtidos em seis diferentes topologias de rede descritas no SNDlib (com cerca de 10 nós e menos de 50 circuitos em cada uma)	56
Quadro 4.1	Potência consumida e capacidade suportada referente a alguns modelos de roteadores Juniper	61
Quadro 4.2	Placas do roteador Cisco 12000	62
Quadro 4.3	Parâmetros envolvidos na modelagem matemática	63
Quadro 4.4	Variáveis envolvidas na modelagem matemática	64
Quadro 4.5	Composição das redes empregadas nos testes	96
Quadro 4.6	Soluções para o Cenário 1	97
Quadro 4.7	Soluções para o Cenário 2	97
Quadro B.1	Relações de conversão entre os problemas Primal e Dual	124
Quadro B.2	Resumo dos corolários da Teoria da Dualidade	126
Quadro B.3	Propriedades intrínsecas entre o PP e o PD	127

Lista de abreviaturas e siglas

ADSL2	<i>Asymmetric Digital Subscriber Line version 2</i>
ALR	<i>Adaptive Link Rate</i>
ARP	<i>Address Resolution Protocol</i>
AS	<i>Autonomous System</i>
BTU	<i>British Thermal Unit</i>
CAPEX	<i>CAPital EXpenditure</i>
CMCF	<i>Capacitated Multicommodity minimumCost Flow</i>
DVS	<i>Dynamic Voltage Scaling</i>
E2-MCRA	<i>Energy Efficient Multi-Constrained Routing Algorithm</i>
ECMP	<i>Equal-Cost Multi-Path</i>
EIGRP	<i>Enhanced Interior Gateway Routing Protocol</i>
EPA	<i>Environmental Protection Agency</i>
EPE	<i>Empresa de Pesquisa Energética</i>
IEA	<i>International Energy Agency</i>
IEEE	<i>Instituto de Engenheiros Eletricistas e Eletrônicos</i>
IETF	<i>Internet Engineering Task Force</i>
IGP	<i>Interior Gateway Protocol</i>
IGRP	<i>Interior Gateway Routing Protocol</i>
IP	<i>Internet Protocol</i>
LAN	<i>Local Area Network</i>
LBNL	<i>Lawrence Berkeley National Laboratory</i>
LPM	<i>Low Power Modes</i>
MIP	<i>Mixed Integer Problem</i>
MILP	<i>Mixed-Integer Linear Programming</i>
MPLS	<i>MultiProtocol Label Switching</i>
MTU	<i>Maximum Transmission Unit</i>
OECD	<i>Organization for Economic Co-Operation and Development</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
OPEX	<i>OPerational EXpenditure</i>
OSPF	<i>Open Shortest Path First</i>
PD	<i>Problema Dual</i>
PL	<i>Programação Linear</i>
PO	<i>Pesquisa Operacional</i>
PP	<i>Problema Primal</i>
QoS	<i>Quality of Service</i>
SPF	<i>Shortest Path First</i>
SPR	<i>Shortest Path Routing</i>
TCP	<i>Transmission Control Protocol</i>
TE	<i>Traffic Engineering</i>
TIC	<i>Tecnologia da Informação e Comunicações</i>
WDM	<i>Wavelength Division Multiplexing</i>

Todas las teorías son legítimas y ninguna tiene importancia. Lo que importa es lo que se hace con ellas.

Jorge Luis Borges

1

Introdução

O mundo hoje não tem fronteiras. O processo de Globalização, que diz respeito à forma como os países interagem e aproximam as pessoas, transformou o planeta numa grande aldeia global. Tal processo foi acelerado, em grande parte, pelo crescimento da Internet e pelo desenvolvimento e barateamento de soluções de informática e telecomunicações, que tornaram a tecnologia acessível a uma quantidade maior de pessoas.

Apesar dos aspectos positivos, a intensificação do uso da Internet e dos recursos de rede em geral tem também consequências negativas, como o aumento do consumo de energia, uma vez que, tradicionalmente, os sistemas de Tecnologia da Informação e Comunicações (TIC) têm sido desenvolvidos com foco unicamente no desempenho, sem preocupação com a eficiência energética.

Segundo Cianfrani et al. [61], há oito anos a energia total demandada pela Internet, por exemplo, era de cerca de 900 bilhões de kWh, o que representava, na época, 5,5% do consumo mundial de energia elétrica, sendo a taxa de crescimento estimada entre 20 e 25% ao ano. Além disso, a eficiência energética da Internet, ou seja, a razão entre a energia consumida e a quantidade total de dados trafegados, era muito baixa, cerca de 8 a 10 vezes inferior a de redes sem fio [75], por exemplo, que já é considerada baixa.

Recentes relatórios estatísticos apresentados por operadoras de telecomunicações, provedores de serviço Internet e organizações públicas presentes em todo o mundo registraram um crescimento vertiginoso na demanda de recursos energéticos [76], e, de acordo com Avallone e Ventre [99], os sistemas de TIC, considerando as operadoras de telecomunicações e também os computadores pessoais, já são responsáveis por até 10% do consumo mundial de energia.

Segundo o *Greenpeace*, se nenhuma medida for tomada, a energia elétrica

consumida pelos *Data Centers* globais e pelas redes de telecomunicações deve triplicar até 2020. Além disso, se o tráfego Internet continuar a crescer no ritmo atual (entre 50 e 100% ao ano), as redes ópticas mundiais de longa distância demandarão 3,3 GigaWatts de energia em 2018 (consumo equivalente ao de sete usinas de médio porte). Outro dado fornecido pelo *The Climate Group* [1, 43] é que as emissões de dióxido de carbono dos dispositivos de telecomunicações triplicarão até 2020, passando das 18 milhões de toneladas em 2002 para 51 milhões. Nesse mesmo período, as emissões geradas pelas infraestruturas de TIC dobrarão, passando de 133 para 299 milhões de toneladas [62].

Estes números são extremamente alarmantes e contribuíram para que a economia de energia se tornasse uma preocupação mundial, por motivos financeiros, ambientais, sociais e comerciais, envolvendo mais de quarenta países na busca de mecanismos para economizá-la.

O aspecto financeiro prescinde de explicação, pois, conhecendo-se a fórmula do cálculo de energia ($E = P \cdot t$, onde E é Energia, P é Potência e t é Tempo), é evidente que a energia consumida é diretamente proporcional à potência consumida pelo dispositivo e ao respectivo tempo durante o qual o mesmo fica ligado. Em outras palavras, quanto maior a potência e o tempo de utilização, maior será a energia consumida e, conseqüentemente, a conta a ser paga por sua utilização.

Para exemplificar, medindo a Potência em Watts (W) e a Energia, em Watts-hora (Wh), considere-se um telefone sem fio, que permaneça ligado 24 horas por dia, ao longo dos 30 dias do mês, e cujo consumo de potência para funcionamento seja de $3W$. A energia consumida pelo telefone no período de um mês seria:

$$E = 3W \cdot 24h/dia \cdot 30 dias$$

$$E = 2,16 kWh$$

Neste caso, considerando o preço de $1 kWh$ como sendo de cerca de R\$ 0,50 (que é o valor atual no Brasil, como ilustrado na Figura 1.1), o custo mensal da energia consumida pelo telefone sem fio seria de R\$ 1,08.

É um valor bem pequeno, chegando mesmo a ser desprezível. No entanto, considerando-se o consumo de empresas, os valores passam a ficar expressivos, representando uma parcela significativa do seu orçamento. Em 2009, por exemplo, os valores anuais de consumo de energia divulgados pelas empresas de telecomunicações Telecom Italia, França Telecom - Orange, Telefonica e Verizon eram, respectivamente, de $2,1 TWh$, $3,7 TWh$, $4,5 TWh$ e $9,9 TWh$ [77], que, convertidos em valores monetários, resultam em quantias vultosas.

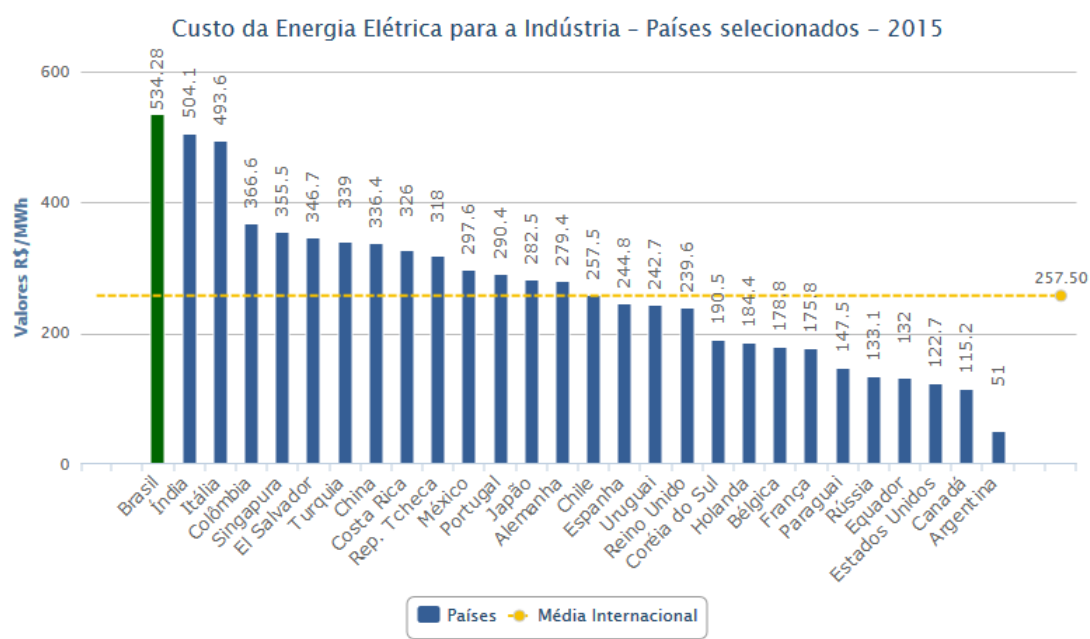


Figura 1.1 – Custo de energia elétrica por MWh

Fonte: FIRJAN [2]

No tocante ao aspecto ambiental, vários estudos têm demonstrado os efeitos devastadores da emissão de gases poluentes e suas consequências nas mudanças climáticas. De acordo com Bianzino et al. [100], é necessária, até o ano de 2020, uma redução de 15 a 30% no volume de gases emitidos para manter o aumento da temperatura global abaixo de $2^{\circ}C$. E segundo a OECD e a IEA, se não houver um planejamento voltado para desenvolver mecanismos com o intuito de limitar o aquecimento global e melhorar a eficiência energética, nos próximos anos, para cada dólar aplicado no setor de energia até 2020, deverão ser gastos outros US\$ 4,3 para compensar o aumento de emissões de gases do efeito estufa e assim manter a mudança climática em margens gerenciáveis. Em um relatório apresentado pela empresa de consultoria *Gartner Group*, ainda

em 2007, o setor de TIC era apontado como o responsável por 2% da emissão global de dióxido de carbono (CO_2), número equivalente ao produzido pelo setor de aviação, com a ressalva de que a perspectiva de crescimento deste setor mostrava-se infinitamente inferior ao de TIC.

No que concerne ao aspecto social, um ponto bastante preocupante é que o crescente consumo de energia pelo setor de TIC (como representado na Figura 1.2), apresenta um ritmo muito maior do que o crescimento da oferta [53], provocando crises energéticas cíclicas [131], como a que o Brasil ora enfrenta, ilustrada na Figura 1.3.

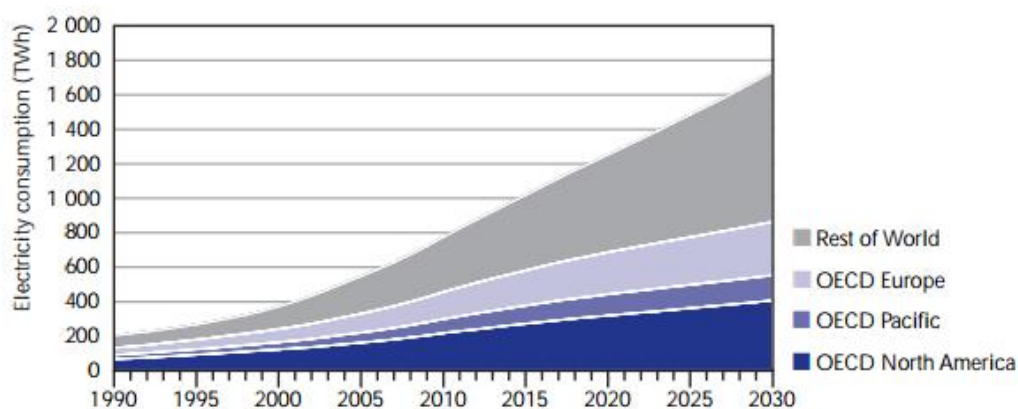


Figura 1.2 – Valores estimados de consumo global de eletricidade por equipamentos de TIC e por usuários no setor residencial distribuídos por região, no período de 1990 a 2030

Fonte: IEA [54]

E, por fim, os aspectos comerciais estão relacionados à própria sobrevivência de um determinado provedor de sistemas de TIC, pois sabe-se que, para suportar as infraestruturas de rede de nova geração e os serviços associados com o intuito de atender a um público que cresce num ritmo acelerado, as áreas de TIC precisam de um número cada vez maior de dispositivos, com arquiteturas sofisticadas capazes de realizar cada vez mais operações complexas de forma escalável. O processamento desses dispositivos, no entanto, tem um fator de aumento de 2,5 a cada 18 meses, que é superior às taxas de aumento do volume de tráfego e de eficiência energética dos semicondutores, que são, respectivamente, de 2 e 1,65, como ilustrado na Figura 1.4 [78]. Frisa-se, ainda, que, além do consumo direto, sistemas de TIC maiores tendem a requerer sis-

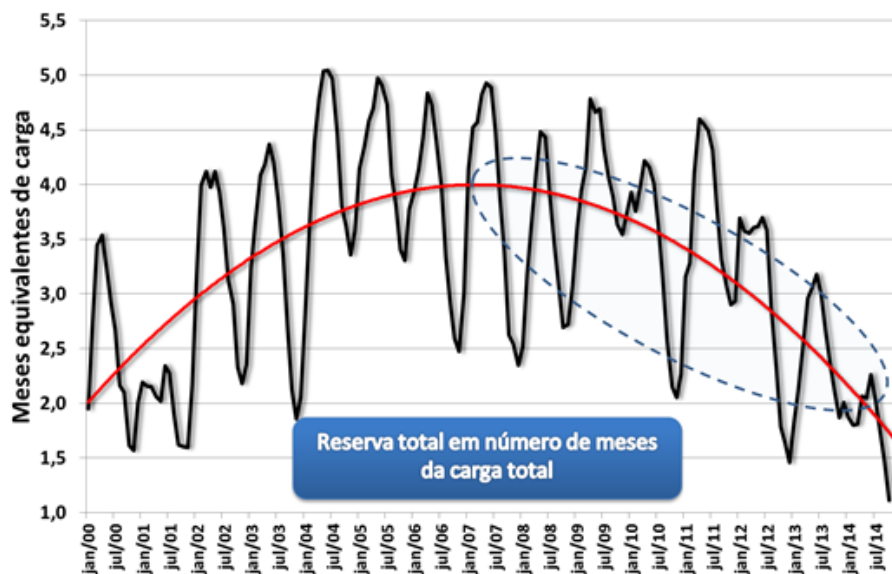


Figura 1.3 – Gráfico ilustrando a crise energética enfrentada pelo Brasil
 Fonte: ILUMINA [132]

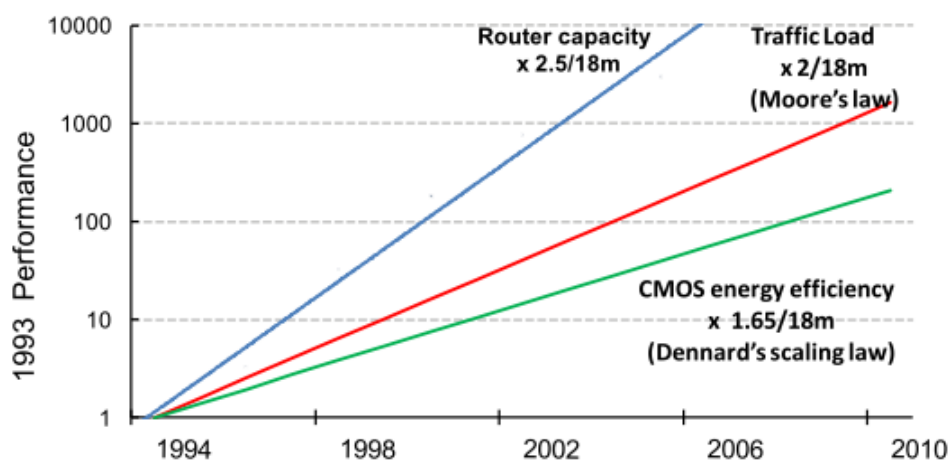


Figura 1.4 – Análise comparativa entre a evolução da capacidade dos roteadores, o crescimento do volume de tráfego (Lei de Moore) e o aumento da eficiência energética dos semicondutores (Lei de Escalabilidade de Dennard), no período de 1993 a 2010
 Fonte: Bolla et al. [78]

temas de ar condicionado proporcionalmente maiores, de tal forma a garantir a refrigeração adequada e o funcionamento correto. Assim, uma vez que tem aumentando a preocupação com a economia de energia, pelos motivos anteriormente citados, o provedor de sistemas de TIC que conseguir ofertar soluções com alto desempenho e reduzido consumo de energia tenderá a ser escolhido em detrimento de outro cujo consumo energético seja maior.

Para exemplificar, considere-se o caso hipotético de uma empresa que, desejando adquirir *switches* para a sua rede, efetue uma análise comparativa entre os produtos dos fabricantes Avaya, HP, Cisco e Juniper, e pré-selecione os seguintes modelos que apresentam características similares: Avaya ERS 4548GT-PWR, HP E4500-48G-PoE, Cisco WS-3750X-48P-S e Juniper EX4200-48P. Para definir a escolha, a referida empresa leva em consideração os impactos no seu CAPEX e no seu OPEX. Como os quatro modelos possuem preços bastante similares, o impacto no CAPEX não é usado como critério de decisão. Na análise do impacto no OPEX, em função dos elevados custos associados ao consumo de energia, é realizado um estudo comparativo da eficiência energética dos quatro equipamentos. Para tanto, são assumidas as premissas apresentadas nos Quadros 1.1 e 1.2 [79].

Quadro 1.1 – Considerações para o cálculo do consumo de energia numa análise comparativa de *switches* de diferentes fabricantes

Considerações para o Cálculo do Consumo de Energia	
Custo médio por kW por hora	US\$ 0,10
Quantidade de dias no ano	365
<i>Uptime</i> do <i>switch</i> , em horas, ao longo do dia	24
Calor dissipado em BTUs para cada Watt	3,41
Watts consumidos para refrigerar 1 BTU	0,105

Fonte: Miercom [79]

Aplicando-se os dados dos Quadros 1.1 e 1.2 a uma rede com 10.000 portas, com uma carga de 70% e pacotes de 512 bytes, são obtidos os valores apresentados no Quadro 1.3 para os custos totais inerentes ao consumo de

energia de cada um dos *switches* pré-selecionados para um período de três anos.

Quadro 1.2 – Consumo de potência, em Watts, dos *switches* Avaya ERS 4548GT-PWR, HP E4500G-48G-PoE, Cisco WS-3750-X-48P-S e Juniper EX4200-48P sujeitos a diferentes cargas de tráfego

	Avaya	HP	Cisco	Juniper
Circuito Ocioso	99,7	126,3	130,7	206,0
Carga de tráfego de 70% e diferentes tamanhos de <i>frames</i>				
64 bytes	101,2	128,1	135,3	210,0
512 bytes	100,7	127,3	133,5	208,0
1518 bytes	100,7	127,1	133,2	207,0
Carga de tráfego de 100% e diferentes tamanhos de <i>frames</i>				
64 bytes	102,1	129,1	136,4	212,0
512 bytes	101,2	128,1	134,6	208,0
1518 bytes	100,9	127,9	133,9	208,0

Fonte: Miercom [79]

Quadro 1.3 – Comparação de custos totais inerentes ao consumo de energia de quatro diferentes modelos de *switches*, ao longo de três anos

	Custo por <i>switch</i>	Com 10.000 portas
Avaya ERS 4548GT-PWR	US\$ 359	US\$ 74.874,00
HP E4500-48G-PoE	US\$ 454	US\$ 94.652,00
Cisco WS-3750-X-48P-S	US\$ 476	US\$ 99.262,00
Juniper EX4200-48P	US\$ 742	US\$ 154.655,00

Fonte: Miercom [79]

Com base nos valores obtidos a empresa decide, então, comprar o *switch* Avaya ERS 4548GT-PWR, por atender a todos os requisitos técnicos e apresentar um custo inerente ao consumo de energia cerca de 36% inferior aos demais.

Diante do exposto é inquestionável a necessidade de desenvolvimento de mecanismos e artifícios visando à economia de energia, o que já motivou a criação de algumas políticas ambientais internacionais aplicadas em programas de suporte à pesquisa nessa área. Além disso, empresas e provedores de serviços, que desejam continuar crescendo, mas cujos gastos com energia têm representado a principal parcela dos seus custos operacionais [101], também têm apoiado a comunidade científica para que realize estudos em diversos campos de atuação voltados para a eficiência energética. Um deles que vem merecendo bastante atenção ao longo dos últimos anos concerne ao desenvolvimento de soluções para a economia de energia em sistemas de TIC, o que deu origem ao conceito de **Redes Verdes**.

Redes Verdes podem ser definidas como aquelas em que se adota um conjunto de práticas visando ao desenvolvimento sustentável, e engloba o estudo e o emprego de tecnologias e produtos (computadores, servidores e subsistemas associados) que sejam energeticamente eficientes, ou seja, que utilizem o mínimo possível de recursos, reduzindo, assim, o impacto para o meio ambiente [44].

O nome, portanto, é simples, mas o objetivo é extremamente desafiador, uma vez que deve ser mantido o crescimento, mas reduzido o consumo de energia [102].

Além disso, considerando-se os critérios tradicionais de planejamento de redes, que primam por superdimensionamento e resiliência, o desafio é ainda maior, pois tais critérios são diametralmente opostos à ideia de economia de energia, uma vez que:

superdimensionamento – É o princípio pelo qual as redes são planejadas e dimensionadas para atender às demandas de tráfego durante os horários de pico. Entretanto, como as demandas variam consideravelmente ao longo do tempo, tais redes caracterizam-se por serem *over-energy consuming*, ou sejam, consomem mais energia do que seria realmente necessário, principalmente nos períodos de baixo tráfego; e

resiliência – Consiste na capacidade de uma rede prover e manter um nível aceitável de funcionamento dos serviços por ela suportados em situações de falhas de elementos que a compõem. Para garantir esta alta disponibilidade são instalados e mantidos ligados dois ou mais equipamentos ou circuitos, de tal forma que os serviços continuem disponíveis ainda que ocorram indisponibilidades em componentes dos sistemas. Em condições normais, portanto, os recursos ficam ociosos, mas consomem o mesmo nível de energia que consumiriam se estivessem em plena carga.

Em suma, portanto, **Redes Verdes** representam uma verdadeira revolução, estabelecendo um novo paradigma para o ambiente de Informática e para as redes de telecomunicações, como bem afirmou Jeong Kim, então presidente do *Bell Labs*, numa das apresentações do projeto Earth (*Energy Aware Radio and neTwork tecHnologies*) que teve início em 2010:

estamos sendo testemunhas de uma mudança substancial na forma de conceber soluções de TIC, uma vez que o objetivo de otimizar as redes para que alcancem sua capacidade máxima está dando lugar ao objetivo de otimizá-las para que alcancem uma maior eficiência energética.

1.1

Organização do Trabalho

Face ao apresentado, fica evidente, portanto, a importância do tema **Redes Verdes**, por sua abrangência mundial e pelos diversos motivos sociais, econômicos, técnicos e ambientais.

E, embora vários trabalhos tenham sido realizados nesta área (dentre os quais pode-se citar Medina et al. [18], Pickavet et al. [45], Lanzisera et al. [63], Fisher et al. [64], Lange et al. [80], Cuomo et al. [81], Bolla et al. [82], Gelenbe e Mahmoodi [83], Lambert et al. [103], Chiaraviglio et al. [104], Rizzelli et al. [105], Krief [106], Lin et al. [122] e Khan e Mauri [128]), muito ainda existe a ser pesquisado, principalmente no tocante à proposição de técnicas que, além de propiciar a economia de energia, assegurem a qualidade dos serviços trafegados nas redes.

Ciente disso, desenvolveu-se o presente trabalho, que consiste na definição de uma política de Engenharia de Tráfego aplicada a **Redes Verdes**. O estudo está direcionado ao *backbone* de uma rede IP e visa a definir, por meio de técnicas de otimização, a menor quantidade de roteadores e circuitos que precisam ser mantidos ativos de tal forma a assegurar o perfeito funcionamento dos serviços suportados pela rede.

Para facilitar o entendimento, o trabalho está organizado da seguinte forma: no presente capítulo é feita uma contextualização do cenário de **Redes Verdes**, apresentando a importância do tema.

No Capítulo 2 é exposto um breve resumo de algumas técnicas propostas na literatura, com o intuito de apresentar o estágio de evolução das pesquisas científicas até então desenvolvidas. Além disso, é detalhada um pouco mais a motivação do presente estudo e apresentados os principais problemas já tratados no campo de **Redes Verdes**.

No Capítulo 3 são fundamentadas a formulação e a modelagem do problema de interesse que visa à economia de energia em redes IP, por meio do desligamento de roteadores e circuitos que compõem seu *backbone*, sem degradar o desempenho das aplicações por ele suportadas.

No Capítulo 4 descreve-se efetivamente a técnica ora proposta, que consiste no emprego do método de decomposição de Benders para a obtenção da solução ótima para o problema de otimização de interesse.

O Capítulo 5 finaliza o trabalho com algumas conclusões globais e sugestões para estudos futuros.

Os Apêndices A, B e C apresentam, respectivamente, os principais conceitos de *Equal-Cost Multi-Path*, Teoria da Dualidade e Condições de Karush-Kuhn-Tucker, necessários ao entendimento de algumas considerações feitas ao longo do trabalho.

O método de decomposição de Benders e premissas importantes a ele associadas que foram consideradas no desenvolvimento do presente trabalho são apresentados nos Apêndices D e E.

No Apêndice F o cerne da presente tese, que consiste na aplicação de forma encadeada do algoritmo de Benders, é fundamentado matematicamente.

O campo de **Redes Verdes** é extremamente vasto e promissor. E embora seu cerne seja único, ou seja, o desenvolvimento sustentável, existem inúmeras áreas de atuação, técnicas e mecanismos que podem ser explorados para que tal objetivo seja atingido.

Visando a facilitar o entendimento e o desenvolvimento das técnicas e de novas propostas, as áreas de pesquisa podem ser agrupadas em categorias que refletem o segmento da rede em que são aplicadas.

No artigo publicado por Bianzino et al. [100] foi apresentada uma divisão em quatro categorias, a saber: Consolidação de Recursos, Virtualização, Computação Proporcional e Desconexão Seletiva. O enquadramento das técnicas em cada um dos grupos é feito em função das seguintes características, respectivamente:

1. por sua aplicação nos elementos que compõem o núcleo da rede e dos sistemas;
2. pela criação de elementos virtuais para substituição de dispositivos;
3. pelo emprego de mecanismos que atuam mais fortemente nos componentes eletrônicos dos sistemas; e
4. por sua aplicação nos elementos de borda da rede.

Ressalta-se que algumas categorias tiveram mais contribuições que outras e alguns trabalhos exploraram também o emprego conjunto de técnicas de diferentes conjuntos. Nas seções seguintes são detalhadas as categorias propostas por Bianzino et al. [100] e apresentam-se exemplos de trabalhos anteriores,

com o intuito de fornecer uma visão do grau de desenvolvimento em cada um dos grupos.

2.1

Consolidação de Recursos

Esta categoria engloba todas as estratégias de dimensionamento voltadas para a redução do consumo global, partindo da premissa de que os dispositivos apresentam períodos de subutilização.

As técnicas propostas neste grupo consideram que, em virtude de o nível de tráfego numa rede seguir um padrão de comportamento conhecido, existe a possibilidade de “adequar” o nível de superdimensionamento às condições gerais de funcionamento da rede. Em outras palavras, o nível requerido de desempenho pode continuar sendo assegurado, usando, no entanto, uma quantidade de recursos dimensionada para a demanda usual de tráfego, em lugar da demanda de pico.

Consolidação de Recursos já é uma técnica popular em muitas áreas da Informática, em particular, em sistemas com *cluster* de servidores ou com mais de um processador, caracterizando-se pelo controle centralizado das decisões de concentração ou distribuição do processamento, de acordo com os requisitos das aplicações.

Na área de redes, no entanto, esta metodologia ainda está incipiente e há muito a ser explorado. Em linhas gerais, o objetivo das técnicas aqui enquadradas é definir a melhor estratégia para o desligamento (ou hibernação) de nós e circuitos que compõem o núcleo de uma rede. Assim, considere-se, por exemplo, o *backbone* de uma rede IP formado pelos nós N1, N2, N3, N4 e N5, que interligam as redes A, B, C e D, conforme representado na Figura 2.1. Por meio do emprego de técnicas de Consolidação de Recursos, poder-se-ia definir que somente os nós N2 e N3 necessitam ficar ativos de tal forma a assegurar a conexão adequada entre todas as redes A, B, C e D, consumindo, contudo, menos recursos, conforme representado na Figura 2.2.

Salienta-se que a proposta apresentada no presente trabalho enquadra-se nesta categoria.

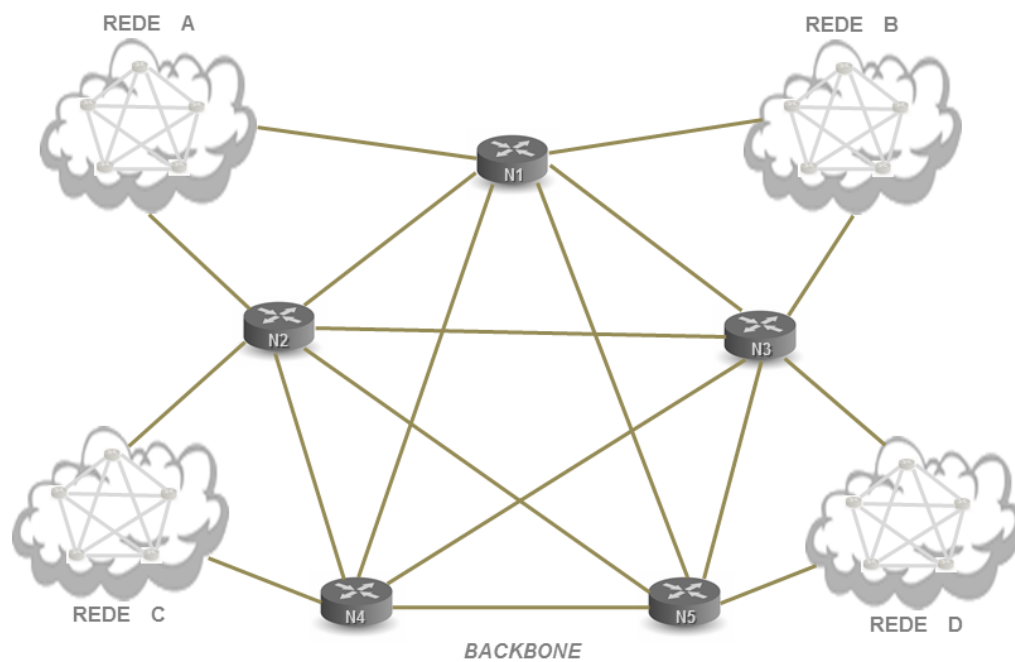


Figura 2.1 – *Backbone* de uma rede IP interligando diversas redes, sem emprego de técnicas de Consolidação de Recursos: todos os nós e circuitos ativos

Fonte: Elaborada pela autora

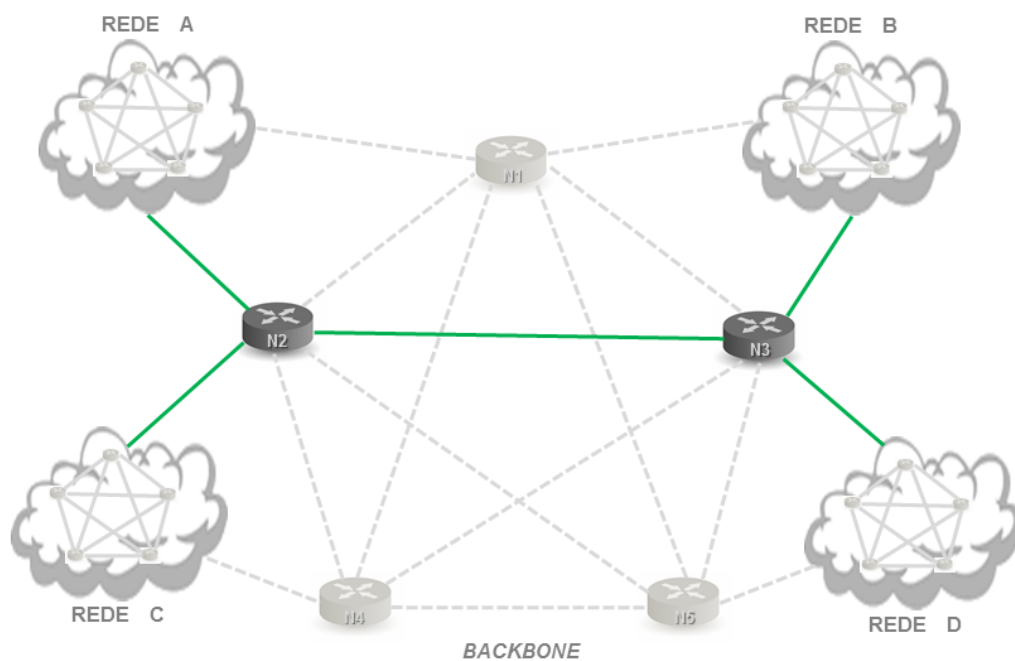


Figura 2.2 – *Backbone* de uma rede IP interligando diversas redes, empregando técnicas de Consolidação de Recursos: ativos somente os nós e circuitos necessários

Fonte: Elaborada pela autora

2.2 Virtualização

Esta categoria contempla o conjunto de mecanismos que permite que mais de um serviço compartilhe o mesmo *hardware*, aumentando, assim, sua utilização, como ilustrado na Figura 2.3.

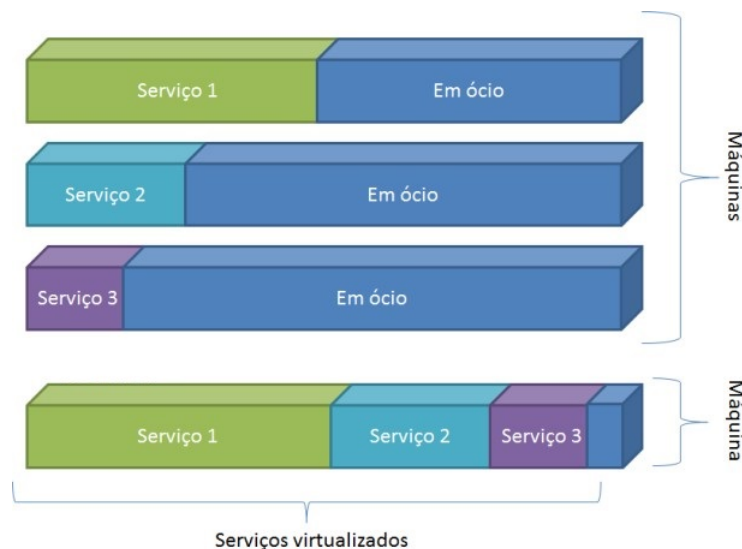


Figura 2.3 – Aproveitamento de capacidade: uma máquina fazendo o papel de três
Fonte: InfoWester [123]

A redução do consumo de energia, neste caso, advém do fato de que uma única máquina com carga elevada consome menos energia do que diversas máquinas com baixa carga.

Cumprе salientar que o conceito de Virtualização pode ser aplicado a diversos tipos de recursos, incluindo dispositivos de armazenamento, recursos de *software* etc. Um exemplo típico de Virtualização consiste em compartilhar servidores em *Data Centers*, reduzindo, assim, os custos de *hardware*, melhorando o gerenciamento de energia, reduzindo os custos energéticos e de refrigeração, e, consequentemente, reduzindo a emissão de CO_2 .

Este campo de pesquisa já está bem maduro. No trabalho publicado por Nanda e Chiueh [31] é apresentada uma pesquisa detalhada de técnicas de Virtualização sob a perspectiva da arquitetura de computadores; enquanto o trabalho de Mosharaf et al. [46] está focado na sua aplicação em redes.

Cumprе salientar que diversas empresas já adotam Virtualização em seus setores de TIC, e em algumas delas o consumo de energia foi reduzido pela metade.

2.3 Computação Proporcional

Este conceito foi introduzido por Barroso e Hölzle [39] e consiste, em linhas gerais, em técnicas de ajuste do consumo de energia de tal forma a torná-lo proporcional ao nível de utilização do recurso. Pode ser aplicado, de forma holística, ao sistema e aos protocolos de rede; e, de forma individual, aos dispositivos e componentes.

Diferentemente das técnicas de Consolidação de Recursos, as aqui enquadradas não visam aos desligamento de equipamentos e interfaces, mas em ajustes na arquitetura eletrônica. *Dynamic Voltage Scaling* (DVS) e *Adaptive Link Rate* (ALR) são exemplos típicos de Computação Proporcional [36, 47, 48]. O primeiro mecanismo reduz o consumo de energia da CPU, em função da carga do sistema, e o último aplica um conceito similar às interfaces de rede, reduzindo sua capacidade, e, assim, seu consumo, em função da carga dos circuitos a elas conectados.

2.4 Desconexão Seletiva

Consiste, como descrito nos trabalhos de Christensen et al. [28], Gunaratne et al. [32] e Allman et al. [40], em mecanismos distribuídos que permitem que algumas partes ou módulos dos equipamentos fiquem inativos por um período, o mais transparentemente possível para o restante dos dispositivos da rede. Esta categoria às vezes é confundida com a de Consolidação de Recursos, mas existe uma sutil diferença entre ambas. Conforme explicado anteriormente, a Consolidação de Recursos aplica-se a recursos que são compartilhados **no núcleo** da infraestrutura da rede, enquanto a Desconexão Seletiva permite o desligamento de recursos ociosos na **borda** da rede.

Alguns exemplos de técnicas enquadradas nesta categoria são aquelas que preveem a desabilitação de nós de borda que estejam enviando apenas sinais

de conectividade com a rede (ex.: envios periódicos de *heartbeats*, resposta desnecessária a pacotes de *broadcasts* etc), transferindo a execução desta tarefa para outros nós, que passam a atuar, então, como *proxies*, simulando momentaneamente o papel dos dispositivos inativos, de tal forma a não ser necessária nenhuma mudança radical nos protocolos de rede. Na Figura 2.4 é apresentado um exemplo de emprego deste tipo de técnica, onde um *switch* atua como *proxy* para o tráfego de pacotes ARP (*Address Resolution Protocol*), permitindo que a máquina de destino fique inativa até que pacotes de dados sejam efetivamente enviados a ela.

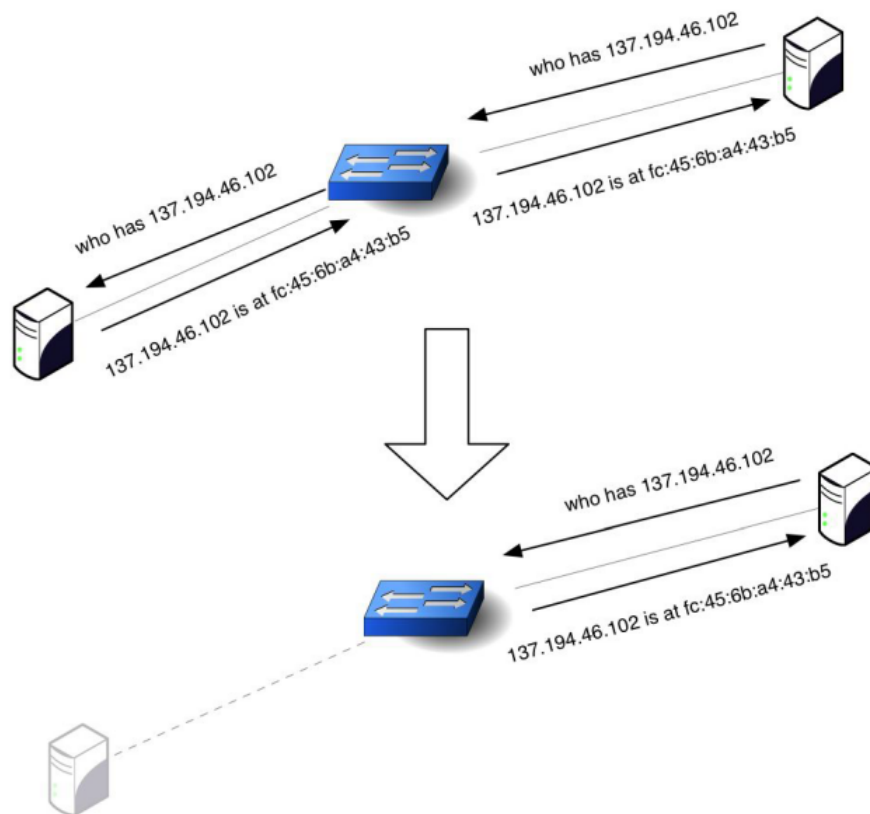


Figura 2.4 – *Switch* atuando como *proxy* para o tráfego de pacotes ARP

Fonte: Bianzino et.al [100]

2.5

Trabalhos Relacionados

O trabalho de Gupta e Singh [26] é um dos marcos da pesquisa em **Redes Verdes**. E desde a sua publicação em 2003 a comunidade científica vem contribuindo com diversos trabalhos. Neste artigo os autores analisam os

impactos da economia de energia nos protocolos de rede a partir da desativação (ou hibernação) de interfaces e outros componentes de roteadores e *switches*. Além de mostrarem que é realmente razoável adotar este tipo de procedimento, discutem as mudanças que precisam ser aplicadas aos protocolos de Internet usados para suportar estratégias de hibernação mais agressivas. Embora não apresentem resultados, por ser um artigo de direcionamento, descrevem de forma clara os benefícios da aplicação de mecanismos para a economia de energia e sugerem importantes campos de exploração para a pesquisa voltada para o núcleo das redes.

No artigo publicado por Chabarek et al. [49] os autores assumem que as demandas de tráfego são conhecidas e formulam o problema denominado de CMCF (do inglês, *Capacitated Multicommodity minimumCost Flow*) para reduzir o consumo de energia. A complexidade de tal abordagem, no entanto, torna-a bastante cara, mesmo para pequenas redes. Os autores medem também o consumo de energia dos dispositivos de rede e executam algumas simulações com base no modelo proposto para representar o consumo de energia.

Nos trabalhos publicados por Chiaraviglio et al. [55, 56] os autores também abordam o problema CMCF, mas apresentam uma série de heurísticas para limitar a complexidade computacional.

Assumindo que a matriz de tráfego é conhecida, Zhang et al. [65] apresentam uma formulação MIP (do inglês, *Mixed Integer Program*) para maximizar a economia total de energia, sujeito às restrições de que a utilização máxima em cada circuito não deve exceder um limiar pré-definido e de que a demanda entre cada par origem-destino só pode ser roteada por um conjunto pré-definido de caminhos.

Similarmente, o trabalho apresentado por Amaldi et al. [84] assume o conhecimento da matriz de tráfego e a utilização do protocolo OSPF e propõe uma heurística que tem como principal objetivo a redução, ao mínimo possível, do consumo de energia; e, como objetivo secundário, a minimização do custo total de utilização do circuito.

Uma análise das vantagens advindas dos algoritmos para eficiência energética é feita por Bianzino et al. [66]. Os autores consideram diferentes matrizes

de tráfego, topologias de rede e perfis de energia dos dispositivos e avaliam a economia de energia em diferentes casos.

No trabalho publicado por Prem et al. [85] o problema de eficiência energética é modelado como um torneio entre ligas, onde cada nó é um jogador, cada liga é uma configuração de rede e a parcela de tráfego entregue corresponde ao rendimento da liga. Os roteadores que devem permanecer ligados são definidos com base no Valor de Shapley [9, 13].

Os autores Cianfrani et al. [61, 86] propõem uma modificação no protocolo OSPF para desligar circuitos e demonstram no primeiro artigo que um número superior a 60% dos circuitos pode ser potencialmente desligado em uma rede real.

Seguindo a mesma filosofia, um algoritmo de roteamento distribuído energeticamente eficiente é proposto por Bianzino et al. [87].

Cientes de que o desligamento de nós e circuitos com o intuito de economizar energia pode causar inconsistências nas tabelas de roteamento durante o período necessário para a convergência do protocolo de roteamento distribuído, Lee et al. [88] propõem um esquema distribuído de atualização de roteamento livre de *loops*, para determinar uma sequência para a atualização das tabelas de roteamento que permite evitar os *loops* transientes.

No artigo publicado por Bolla et al. [107] os autores apresentam uma visão bem organizada de técnicas propostas para melhorar o desempenho energético da Internet. Todas as abordagens possíveis são, então, avaliadas estimando o potencial possível de economia de energia e analisando o equilíbrio entre eficiência energética e desempenho da rede. Para a avaliação são considerados dois cenários de redes reais, um de um grande provedor de Internet e outro de um provedor mediano, segmentando as contribuições em diferentes setores da rede (residência, acesso, transporte e núcleo).

Dentre as possíveis abordagens para a redução do consumo de energia da Internet, uma extremamente promissora baseia-se no emprego de modos de repouso para roteadores e interfaces de acordo com os níveis de tráfego. Quando a carga está baixa, alguns dos elementos de rede podem ser colocados para hibernar para economizar energia e o tráfego pode ser desviado pelos

elementos ativos.

Neste contexto os autores Kim et al. [108] consideram o problema de minimização do número de elementos ativos para um dado tráfego e resolvem-no por meio de um algoritmo heurístico baseado no método da Colônia de Formigas, que consiste em uma técnica probabilística para resolver problemas computacionais por meio da busca de bons caminhos em grafos. O algoritmo recebe este nome pois é baseado nas formigas e em sua movimentação fora do formigueiro, ou seja, como as formigas liberam uma substância chamada feromônio à medida que caminham, tornam o caminho por onde passam mais atrativo para as demais. Assim, quando estão levando comida para o formigueiro, o caminho inicialmente escolhido pode não ser necessariamente o melhor, mas as concentrações de feromônio vão sendo alteradas ao longo do caminho com o passar do tempo e o trajeto vai se modificando à medida que elas transitam entre sua fonte de alimento e o formigueiro. A Figura 2.5 representa justamente esse trânsito das formigas e o ajuste do caminho de acordo com a concentração de feromônio. Nela fica evidenciado que o caminho mais curto é favorecido, porque, com o passar do tempo, o feromônio evapora, reduzindo a sua força atrativa. Quanto menor é o tempo que uma formiga leva para ir e voltar por um determinado caminho, maior é o tempo que o feromônio tem para evaporar. Portanto, como os caminhos menores são percorridos mais rapidamente, a densidade de feromônio ao longo deles fica mais elevada, e consequentemente, torna-os mais atrativos que os caminhos maiores [67].

No trabalho publicado por Cuomo et al. [109] é proposta uma heurística diferente, que, usando algumas propriedades topológicas do grafo representativo da rede de um provedor de serviços de Internet, identifica as interfaces dos roteadores que podem ser desativadas.

Uma abordagem diferente baseada em dados de roteamento *on-line* e em controle de admissão é proposta por Avallone e Ventre [99]. Seu trabalho trata do problema de roteamento de fluxos em uma rede de tal forma a maximizar o número de nós e circuitos não utilizados, que podem, por conseguinte, ser desligados com o intuito de aumentar a economia de energia. Este trabalho parte do princípio de que a performance requerida por um fluxo deve ser

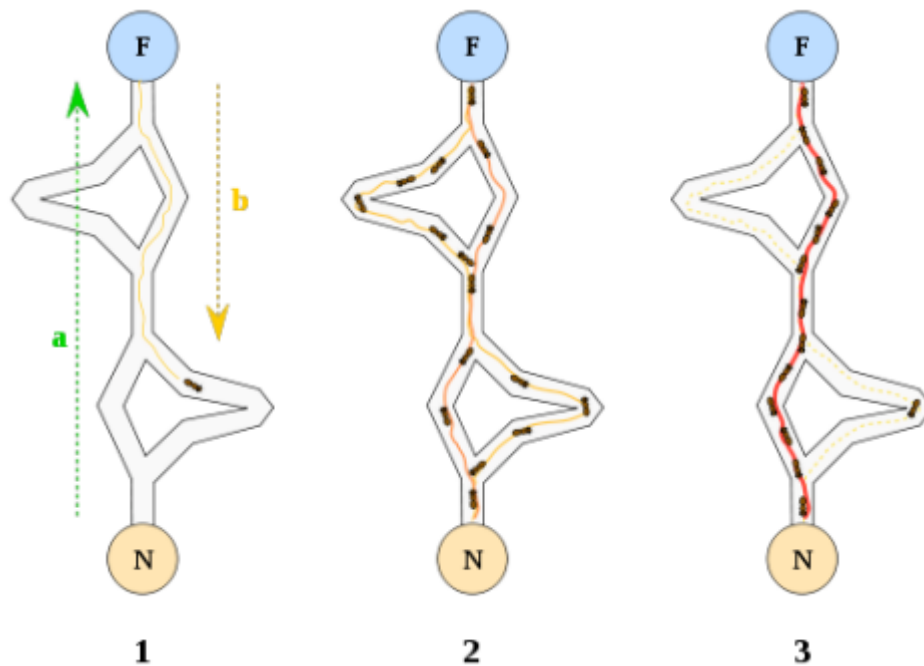


Figura 2.5 – Método da colônia de formigas
 Fonte: De Aquino [67]

garantida e considera requisitos dos fluxos em termos da largura de banda disponível e de outras medidas de QoS, como *jitter* e retardo, que podem ser modeladas como restrições aditivas. O algoritmo proposto, E2-MCRA (*Energy Efficient Multi-Constrained Routing Algorithm*), tenta descobrir um caminho que satisfaça aos requisitos de cada fluxo. Se o algoritmo de roteamento não é capaz de encontrar tal caminho o fluxo não é admitido. O objetivo do algoritmo, portanto, é admitir tantos fluxos quanto possível, enquanto minimiza o número de nós e circuitos que devem ser mantidos ligados para este fim. Os conceitos básicos do E2-MCRA são as técnicas de aproximação *look-ahead* e busca *depth-first* e uma definição de comprimento de caminho como função dos requisitos citados anteriormente e da condição atual (ligado ou desligado) dos nós e circuitos ao longo do caminho.

Sob o ponto de vista energético tecnologias de transporte baseadas em redes ópticas são, em geral, eficientes. No entanto, com o tráfego IP a eficiência do transporte óptico depende muito da possibilidade de agrupar fluxos de tráfego com o intuito de reduzir o número de portas ópticas/IP.

Os autores Hou et al. [110] investigam o problema de agrupamento ro-

busto de tráfego em redes IP sobre WDM com múltiplos níveis de granularidade.

Uma abordagem alternativa para o WDM, que devido a sua grande capacidade pode levar à ineficiência energética quando o tráfego é baixo, baseia-se na utilização de uma técnica de modulação óptica mais flexível, conhecida como *Orthogonal Frequency Division Multiplexing* (OFDM). Os autores Vizcaíno et al. [111] comparam o desempenho de OFDM com o de WDM, em termos de consumo de energia.

Os problemas de roteamento integrado e atribuição de comprimento de onda são investigados no trabalho publicado por Ricciardi et al. [112], no qual a eficiência energética é introduzida no processo de decisão, em adição aos objetivos tradicionais de gerenciamento de rede.

No segmento de acesso de rede existe uma diversidade muito grande de tecnologias que podem ser adotadas para a conexão dos dispositivos à rede, o que dificulta muito a padronização de técnicas para a economia de energia.

O caso da tecnologia Ethernet é considerado no artigo publicado por Herrería-Alonso et al. [89] no qual os algoritmos de hibernação do padrão 802.3az do IEEE [68] são analisados por meio de modelos matemáticos capazes de estimar a economia de energia e a média do retardo dos pacotes de tráfego de saída. Os modelos são, então, usados para derivar parâmetros de configuração ótimos para determinados requisitos de energia e qualidade.

O padrão 802.3az é analisado também nos trabalhos publicados por Kohl [41], Nordman [50], Reviriego et al. [57, 69, 90], Christensen et al. [70], Marsan et al. [91], Herrería-Alonso et al. [92] e Chatzipapas [113].

Estudos relacionados ao gerenciamento de energia em ambiente de LAN (*Local Area Network*) são apresentados por Gupta e Singh [29, 42], Rodríguez-Pérez et al. [58], Herrería-Alonso et al. [93] e Larrabeiti et al. [94].

A tecnologia de acesso ADSL2, que atualmente é amplamente adotada como mecanismo para conexão das residências à Internet, é analisado por Wolkerstorfer et al. [114]. O trabalho considera o caso da já padronizada técnica LPM (*Low Power Modes*) e estuda o problema de se otimizar o nível de taxa do LPM com o intuito de garantir eficiência energética.

Ao longo dos últimos anos as redes sem fio tornaram-se uma tecnologia muito popular para acesso à Internet, devido ao uso comum de potentes dispositivos pessoais, como *Smart Phones*, *Tablets* e *Phablets*. Tal fato aumentou o desafio de melhorar a eficiência energética em redes de comunicação devido ao fato de os níveis de consumo de energia de dispositivos sem fio serem muito maiores que os de elementos cabeados.

Nesta linha de pesquisa Mancuso e Alouf [115], consideram o caso de conectividade contínua com dispositivos inteligentes em redes sem fio e analisam os modos de baixa potência por meio de modelos de filas.

A análise de modos de baixa potência também é considerada no trabalho publicado por De Turck et al. [116], no qual o foco é sobre o tráfego bidirecional.

Os autores Gomez et al. [117] caracterizam o perfil de consumo de energia de estações de acesso sem fio com uma abordagem experimental com foco em tecnologias Wi-Fi e WiMax.

O uso de modo de hibernação para o Ponto de Acesso em Redes Locais Sem Fio é considerado no trabalho publicado por Da Silva et al. [118] no qual o foco é em algoritmos de associação que atribuam tráfego a um subconjunto de Pontos de Acesso.

O tema de serviços em nuvem ou *Cloud Computing* é considerado no artigo publicado por Nguyen et al. [119], onde se considera a integração de *Data Centers* e redes de banda larga que permite um uso mais flexível dos recursos computacionais por meio de uma alocação dinâmica de recursos distribuídos geograficamente, de tal forma a explorar a disponibilidade de fontes de energia verde.

Considerando o conceito de Engenharia de Tráfego baseada em economia de energia, Chu et al. [95] propõem a criação de caminhos alternativos numa rede MPLS para desvio de tráfego durante horários pré-definidos, permitindo o desligamento de elementos e, conseqüentemente, economizando energia. Já Amaldi et al. [124] propõem uma heurística para reduzir a quantidade de nós ativos, considerando também, *a posteriori*, um nível máximo de utilização dos circuitos que compõem a rede.

A relação entre a economia de energia e a robustez da rede é minuciosa-

mente analisada por Addis et al. [129]. Neste artigo é apresentada a Figura 2.6, que ilustra bem o paradoxo entre eficiência energética e resiliência das redes.

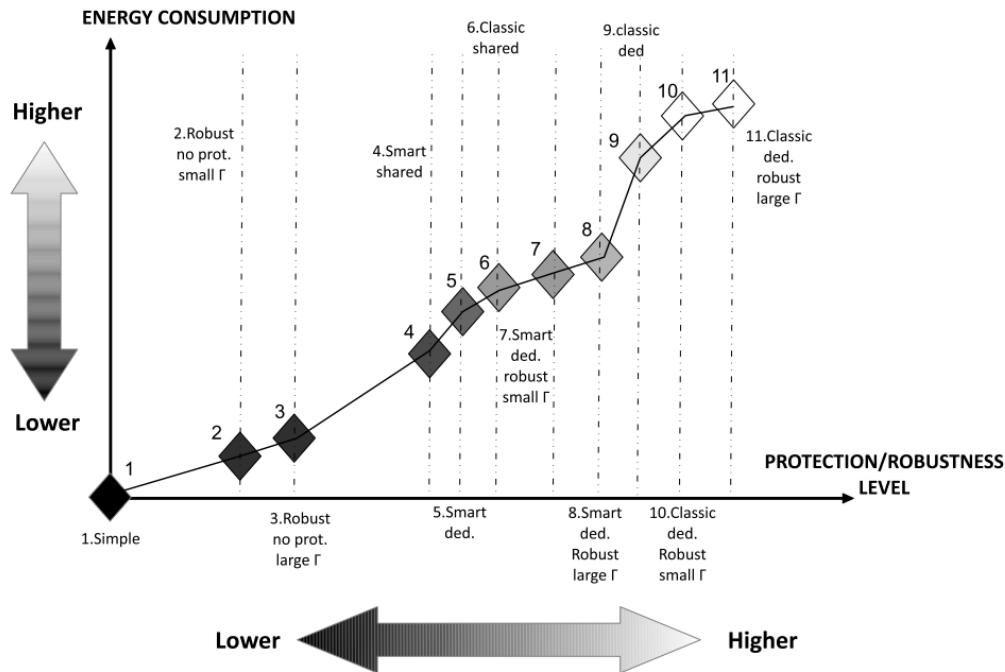


Figura 2.6 – Relação entre economia de energia e resiliência

Fonte: Addis et al. [129]

2.6 Motivação

Embora diversos trabalhos tenham sido publicados no campo de **Redes Verdes**, conforme mostrado nas seções anteriores, existem muitas áreas a serem exploradas e algumas delas como a de Engenharia de Tráfego baseada em economia de energia ainda têm um número relativamente pequeno de contribuições [124].

Assim, apresenta-se o presente trabalho, no qual se considera o problema de otimização do gerenciamento de energia em redes IP, mediante o desligamento planejado de roteadores e circuitos, tendo como objetivo também a garantia do desempenho dos tráfegos cursando por tais redes.

Assume-se que a matriz de tráfego é conhecida e que o IGP (*Interior Gateway Protocol*) usado permite o balanceamento uniforme por mais de um circuito entre origem e destino.

Pelo fato de o OSPF (*Open Shortest Path First*) ser um dos protocolos de roteamento interno mais empregado, podendo ser adotado em redes de pequeno e grande portes, ele é o IGP escolhido.

A definição dos circuitos que devem desligados é feita atribuindo-se a eles custo elevado. E para que um roteador possa ser desativado, todos os circuitos a ele conectados devem possuir custo alto também.

A técnica a ser apresentada, portanto, visa à definição *off-line* dos custos, provocando o redesenho da rede IP, de tal forma a economizar o máximo possível de energia, assegurando, também, que a utilização dos circuitos não ultrapasse um determinado limiar.

A proposta não considera otimização *on-line*, por dois principais motivos:

- recidivas alterações nos custos podem provocar instabilidades na rede em função da quantidade de atualizações na tabela de rotas;
- caso os recursos sofram muitas alterações de estado, o consumo de energia durante as fases de transição pode ser um fator negativo [130].

Sabe-se que outros trabalhos, como o publicado por Amaldi et al. [124], propuseram técnicas de Engenharia de Tráfego baseadas em economia de energia, tendo como preocupação também os aspectos relacionados ao congestionamento dos circuitos, mas, diferentemente das contribuições encontradas na literatura pesquisada, o presente trabalho apresenta um tratamento para a questão sem recorrer a métodos heurísticos.

Ressalta-se, que o fato de outros trabalhos terem recorrido a heurísticas para a resolução dos problemas de otimização em função da excessiva carga computacional destaca ainda mais a importância do presente trabalho, que, propondo a resolução por métodos de decomposição, visa a chegar à solução exata e não aproximada, de forma direta, mas sem comprometer o processamento dos computadores envolvidos.

O termo “Pesquisa Operacional” (PO), em inglês *Operations Research*, foi cunhado pelo matemático russo Leonid Kantorovich, um dos pioneiros no seu emprego como instrumento do planejamento econômico, na tentativa de englobar, sob uma única denominação, todas as técnicas existentes, ou que viessem a ser desenvolvidas, que tivessem a finalidade de prever e comparar estratégias ou decisões alternativas com o intuito de oferecer suporte à definição de políticas e determinação de ações. Atualmente o termo engloba um conjunto de disciplinas isoladas tais como Programação Matemática, Análise Estatística, Teoria das Filas, Simulação, Teoria dos Jogos, Cadeias de Markov, Programação em Redes e Teoria dos Grafos, e suas contribuições estendem-se por praticamente todos os domínios da atividade humana, da Engenharia à Medicina, passando pela Economia e a Gestão Empresarial [125].

Para apoiar a tomada de decisão, a PO busca a solução de problemas que podem ser representados por modelos, como esquematizado na Figura 3.1 [96]. Os modelos em PO podem ser classificados como icônicos ou físicos (por exemplo, maquetes), analógicos (por exemplo, organogramas e fluxogramas) e matemáticos (representações quantitativas de processos e problemas reais, por meio de relações matemáticas, como equações, inequações, funções ou lógica simbólica) [71, 97]. A elaboração de modelos em Pesquisa Operacional possui inúmeras vantagens, dentre as quais pode-se destacar o fato de que a construção de um modelo revela aspectos não evidentes *a priori*, propiciando, deste modo, um melhor entendimento do problema. Para a elaboração dos modelos um dos artifícios usado é o emprego de grafos, que consistem numa representação abstrata do problema real formada por dois conjuntos finitos identificados pelo par $G = (V, E)$, onde V é o conjunto de nós ou vértices e E

é o conjunto de arestas [11, 19, 133], conforme representado na Figura 3.2.

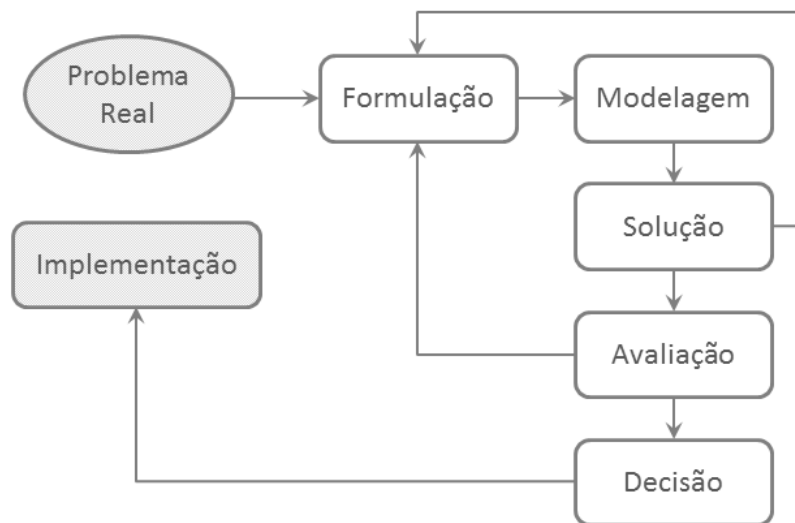


Figura 3.1 – Representação esquemática de um problema de Pesquisa Operacional
Fonte: Cardoso [96]

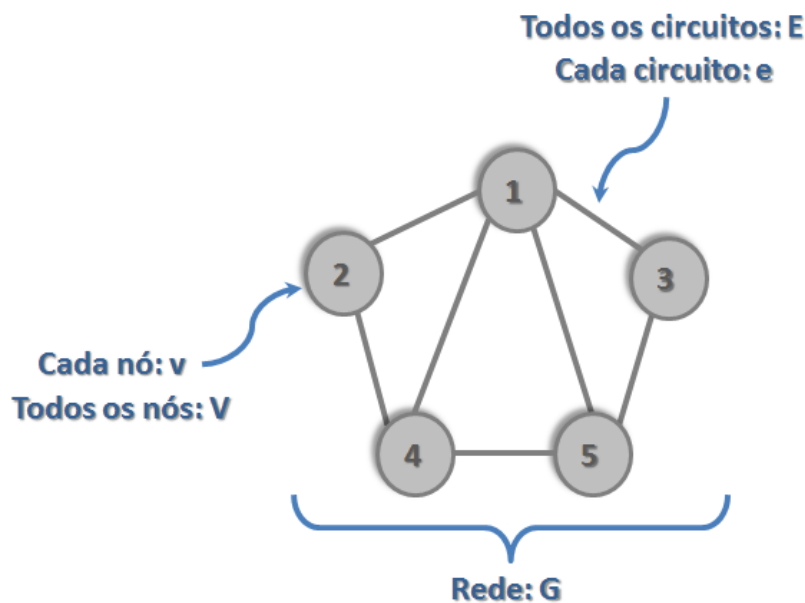
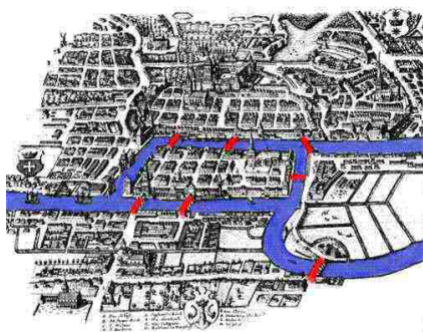


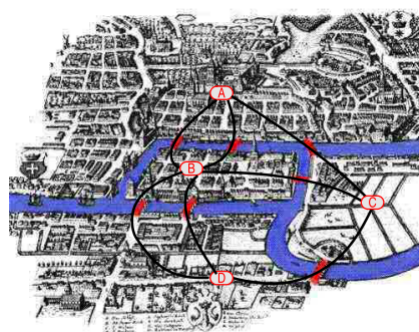
Figura 3.2 – Grafo não orientado, com 5 nós e 7 arestas
Fonte: Elaborada pela autora

A área de Teoria dos Grafos teve início em 1736 quando o matemático Leonhard Euler publicou o artigo *Solutio problematis ad geometriam situs pertinentis* contendo a solução para o problema das sete pontes de Königsberg (atualmente Kaliningrado, na Rússia). A dúvida que não conseguia ser esclarecida pelos habitantes desta cidade é se era possível efetuar um percurso de

tal forma que se iniciasse e voltasse à mesma posição passando por todas as pontes uma única vez. Euler resolveu este problema interpretando cada zona como um nó de um grafo, e cada ponte interligando as duas margens como uma aresta, como mostrado na Figura 3.3 [133].



3.3(a): Mapa de Königsberg.



3.3(b): Grafo que serviu de modelo.

Figura 3.3 – Problema das sete pontes de Königsberg

Fonte: Loureiro e Goussevskaja [133]

Ressalta-se que desde a sua criação a Teoria dos Grafos já evoluiu bastante e hoje é largamente empregada em diversas áreas, como as exemplificadas no Quadro 3.1. Na Figura 3.4 é ilustrado o emprego de um grafo com pesos (também chamado de grafo valorado) para representar a conectividade entre os roteadores do *backbone* IP da rede Internet2, bem como as métricas do IGP nele configurado.

Outro importante campo de estudo da Pesquisa Operacional que vem se destacando como uma poderosa ferramenta matemática para vários segmentos, propiciando melhorias mensuráveis nos processos e suas respectivas automações é o da Programação Matemática, que pode ser definida como a área que estuda Problemas de Otimização [96].

Em um Problema de Otimização empregam-se modelos matemáticos e pretende-se maximizar ou minimizar uma função numérica, chamada *função objetivo*. Esta função é a representação matemática do objetivo a ser atendido pelo modelo, podendo vir a representar critérios econômicos (minimização de custos, maximização de lucros), ou operacionais (minimização de atrasos), por exemplo. O processo de minimização ou maximização deve ser realizado através da determinação dos parâmetros que definem o sistema. Tais parâmetros são

Quadro 3.1 – Exemplos de emprego de grafos

Grafo	Vértice	Aresta
Comunicação	Centrais telefônicas, computadores, satélites	Cabos, fibras ópticas, enlaces de microondas
Circuitos eletrônicos	Portas lógicas, registradores, processadores	Filamentos
Hidráulico	Reservatórios, estações de bombeamento	Tubulações
Financeiro	Ações, moeda	Transações
Transporte	Cidades, aeroportos	Rodovias, vias aéreas
Arquitetura funcional de <i>softwares</i>	Módulos	Interações entre os módulos
Internet	Páginas Web	<i>Links</i>
Jogos de tabuleiro	Posições no tabuleiro	Movimentos permitidos
Redes	Roteadores	Circuitos

Fonte: Loureiro e Goussevskaia [133]

chamados de *variáveis de decisão*. Na maioria dos problemas encontrar-se-ão *restrições* impostas para que o projeto seja *admissível* ou *viável*, devido às leis físicas da natureza, leis políticas, limitações de orçamento, entre outras condições que restrinjam os cenários de soluções possíveis. A solução para o problema é considerada *factível* ou *viável* quando respeita todas as restrições do problema e, portanto, é possível de ser implementada. O conjunto de todas as soluções factíveis gera a *região factível*, que é, portanto, a região formada pela intersecção das restrições do problema. Dentro da região factível, a melhor solução factível encontrada, de acordo com a função objetivo, é denominada de *solução ótima* [24].

Formalmente, um modelo para um problema de Programação Matemática é escrito da seguinte maneira [20]:

$$\begin{aligned}
 &\text{minimize} && \underline{c}^T \cdot \underline{x} \\
 &\text{sujeito a:} && \underline{A} \cdot \underline{x} = \underline{b} \\
 &&& \underline{x} \geq \underline{0}
 \end{aligned}$$

onde $\underline{c} \in \mathbb{R}^n$, $\underline{x} \in \mathbb{R}^n$, $\underline{b} \in \mathbb{R}^m$ e $\underline{A} \in \mathbb{R}^{m \times n}$.

A desigualdade $\underline{x} \geq \underline{0}$ expressa que cada componente de \underline{x} assume apenas valores não-negativos. Existem diversas variações para o problema acima, por exemplo, em vez de minimização, poderia ser um problema de maximização, ou as restrições poderiam estar na forma de desigualdades, tais como $\underline{A} \cdot \underline{x} > \underline{b}$ ou $\underline{A} \cdot \underline{x} < \underline{b}$.

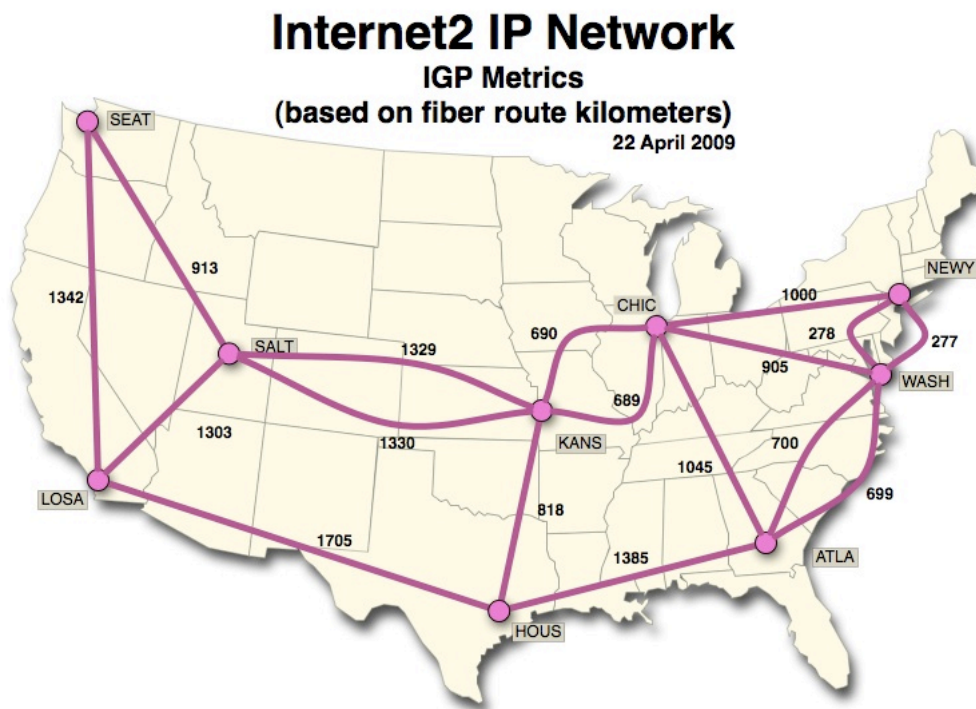


Figura 3.4 – Grafo com pesos representando a rede Internet2 em abril de 2009
 Fonte: Internet2 Network NOC [3]

Dependendo do comportamento de \underline{x} e das restrições existem diferentes classes de problemas de otimização, como descrito a seguir:

- **Programação Linear** ou **Otimização Linear** – Aplicável a problemas caracterizados por envolverem função objetivo e restrições (equações e/ou inequações) lineares. Esta classe subdivide-se em duas:

- **Programação Linear Contínua** – Aplicável a problemas nos quais as restrições e a função objetivo são funções lineares, mas que não contenham variáveis inteiras em sua formulação.

- **Programação Linear Inteira** – Aplicável a problemas nos quais as restrições e a função objetivo são funções lineares, mas algumas ou todas as variáveis pertencem ao conjunto dos números inteiros. No primeiro caso, é também conhecida simplesmente como Programação Inteira e no segundo, como Programação Linear Inteira Mista. Um exemplo típico de emprego de Programação Inteira é a modelagem de situações em que se deve decidir se uma ação deve ser realizada ou não, pois, nesses casos, a variável de decisão só pode assumir os valores 1 e 0, respectivamente.

- **Programação Não-Linear** ou **Otimização Não-Linear** – Aplicável a problemas em que pelo menos uma das funções envolvida (função objetivo ou restrições) não é linear.

A dificuldade de obtenção de solução de um modelo está intimamente ligada à classe de programação em que se enquadra. Por exemplo, um problema de Programação Linear Contínua pode ser resolvido de maneira exata e eficientemente mesmo que contenha milhares (e, às vezes, milhões) de variáveis e restrições, pois há métodos eficientes para isso. Já se o problema for de Programação Linear Inteira Mista, sua resolução pode demandar uma carga computacional muito grande ou levar tempo demais para ser resolvido, comprometendo a utilidade do modelo. No tocante aos problemas de Programação Não-Linear existem métodos de solução eficazes apenas para alguns casos particulares, como funções quadráticas ou funções convexas, e problemas envolvendo funções mais complicadas com algumas dezenas de variáveis e restrições já podem ser impossíveis de serem tratados.

Uma alternativa à resolução de maneira exata, quando esta se torna inviável ou extremamente onerosa em termos computacionais (tempo e memória) é o emprego de heurísticas, que se constituem basicamente em algoritmos [21] que não buscam diretamente a otimização pura, ou seja, não garantem que seja

obtida a solução ótima em todos os cenários resolvidos por eles, mas geram soluções aceitáveis (“boas soluções” ou de “boa qualidade”) e com baixo custo computacional. A desvantagem dos métodos heurísticos é que, em alguns casos, eles podem não ser muito precisos ou previsíveis. Além disso, ocasionalmente incorrem em falhas, devido à escalabilidade do problema (explosão combinatória à medida que o problema cresce), a hipóteses errôneas que estejam sendo consideradas ou à própria complexidade da modelagem do problema (que pode ser devido, por exemplo, a variáveis inteiras ou discretas, função objetivo não linear e não diferenciável, restrições não lineares ou região factível não convexa) [120].

Outra alternativa é adaptar os modelos de tal modo que possam ser resolvidos de forma exata por meio do emprego de métodos de decomposição [4]. Assim, em Programação Matemática, decomposição refere-se ao processo pelo qual um problema ou sistema complexo pode ter seu conjunto de variáveis ou restrições reorganizado de tal maneira a gerar subproblemas ou subsistemas menores e mais fáceis de serem resolvidos. Esta reorganização gera dois grupos de variáveis e restrições:

- **independentes** – são aquelas que têm uma estrutura especial e permitem uma fácil divisão em blocos disjuntos;
- **complicadoras** – são aquelas que dificultam a decomposição do problema e impedem que um algoritmo potencialmente mais simples seja aplicado na resolução do problema. Restrições complicadoras envolvem variáveis diferentes; e variáveis complicadoras, por sua vez, são aquelas presentes em diversas restrições.

Com base na estrutura dos problemas foram propostos na literatura diversos métodos de decomposição, dentre os quais podem ser citados os clássicos de Dantzig-Wolfe e de Benders. O primeiro adapta-se muito bem a problemas com estrutura bloco-angular primal ou sistemas com restrições complicadoras (Figura 3.5(a)), enquanto o segundo é ideal para estruturas bloco-angular dual ou sistemas com variáveis complicadoras (Figura 3.5(b)) [12]. E como os sistemas aos quais tais métodos podem ser aplicados estão intimamente ligados

pela propriedade da dualidade, detalhada no Apêndice B, o dual de um sistema com restrições complicadoras é um sistema com variáveis complicadoras, e vice-versa. Desta forma, pode-se aplicar o método de Dantzig-Wolfe diretamente para decompor problemas com restrições complicadoras, ou aplicar Benders ao seu dual. De maneira análoga, caso o problema tenha variáveis complicadoras, pode-se decompô-lo diretamente por Benders, ou aplicar Dantzig-Wolfe ao seu dual.

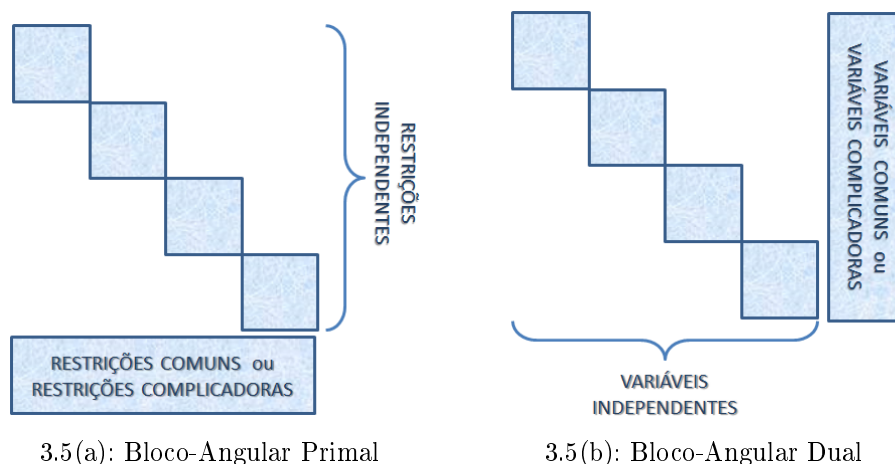


Figura 3.5 – Algumas estruturas de problemas suscetíveis a decomposição
Fonte: Elaborada pela autora

Os métodos de decomposição são, portanto, uma alternativa eficiente para a resolução de certos problemas de otimização de larga escala. Em vez de considerar simultaneamente todas as variáveis de decisão e restrições do problema, as técnicas de decomposição particionam o problema original em múltiplos problemas menores. E uma vez que a dificuldade computacional dos problemas de otimização aumenta significativamente à medida que aumenta o número de variáveis e restrições, resolver problemas menores iterativamente tende a ser mais eficiente do que resolver um único problema grande.

Ao contrário dos métodos tradicionais, os métodos de decomposição dividem o processo de tomada de decisões em várias etapas. No método de decomposição de Benders, por exemplo, conforme descrito no Apêndice D, em um primeiro estágio um problema denominado Mestre é resolvido para um subconjunto de variáveis, e os valores das variáveis restantes são determinados

em um segundo estágio por um problema denominado de Escravo (ou Subproblema), que leva em consideração os valores das variáveis obtidos no primeiro estágio. Se o Subproblema determina que as decisões propostas no primeiro estágio são inviáveis, uma ou mais restrições são geradas e adicionadas ao Problema Mestre, o qual é, então, novamente resolvido. Assim, uma solução que poderia ser demorada ou até mesmo impossível de ser obtida a partir da resolução direta de um problema grande e complexo pode ser encontrada de forma mais simples por meio da resolução iterativa de diversos problemas pequenos.

3.1

Engenharia de Tráfego Baseada em Economia de Energia

O emprego da Teoria dos Grafos na modelagem de problemas de otimização de redes IP ou MPLS (onde os roteadores podem ser representados pelos nós e os circuitos, pelas arestas) para o tratamento de problemas de fluxo máximo ou de escolha de caminho mais curto já está bastante sedimentado, e sua extensão para os casos nos quais se deseja minimizar o consumo de energia pode ser feita de forma trivial, como demonstrado, por exemplo em [124] e detalhado a seguir.

Primeiramente, cabe lembrar que o roteamento do tráfego é o ponto principal a ser considerado na concepção e no gerenciamento de qualquer rede de comunicação, que inclui as redes IP, por exemplo. O termo “roteamento” indica que se deseja forçar os fluxos a usarem rotas apropriadas, as quais são frequentemente pré-definidas [72]. Na prática, o tráfego de fluxos aparece sempre que dois usuários finais precisam se comunicar ou quando um usuário final requer algum conteúdo de um servidor distante. Uma vez que os fluxos aparecem dinamicamente, as decisões de roteamento devem, essencialmente, ser realizadas pela própria rede: se uma nova conexão é demandada para o tráfego de um fluxo entre um usuário final no nó **A** e um usuário final no nó **B**, alguns equipamentos de rede devem decidir ao longo de qual rota de **A** para **B** esta conexão deve ser estabelecida. Esta decisão de roteamento deve ser feita em uma fração de segundo e de tal forma a atender o usuário com uma qualidade de serviço satisfatória e a minimizar o consumo de recursos

de rede. Um protocolo de roteamento é, portanto, um conjunto de regras e mecanismos implementados em uma rede com o intuito de possibilitar decisões de roteamento adequadas.

Os protocolos de roteamento que são implementados para decidir como rotear o tráfego de um nó a outro dentro do mesmo Sistema Autônomo (cuja sigla é AS, derivada do nome em inglês *Autonomous System*) são chamados *Interior Gateway Protocols* (IGPs). Os IGPs “clássicos” baseiam-se em um paradigma de roteamento simples chamado Roteamento de Caminho Mais Curto (em inglês *Shortest Path Routing*, SPR), no qual cada pacote é transmitido por circuitos IP ao longo do caminho mais curto entre sua fonte e os nós de destino do AS. Os protocolos IGP mais utilizados frequentemente são o OSPF e o IS-IS. De acordo com eles, cada roteador no AS deve adquirir e manter individualmente uma visão completa e precisa da topologia do AS (isso é feito por meio de trocas frequentes de mensagens do protocolo de roteamento entre os roteadores), além de informação adequada sobre todos os circuitos desta topologia. A informação relativa ao circuito limita-se ao “custo” ou “peso administrativo” do enlace, que é um valor inteiro atribuído pelo administrador da rede (dentro dos limites definidos pela versão do protocolo de roteamento). Usando a topologia e os pesos administrativos dos circuitos, cada roteador é capaz de calcular sua árvore de caminho mais curto cobrindo a topologia do grafo, ou, em outras palavras, um caminho mais curto em direção a cada um dos roteadores existentes na rede. Os custos administrativos são o único artifício de que o administrador da rede dispõe para influenciar e controlar o roteamento de tráfego na rede. Este controle é feito de forma indireta, pois embora o administrador da rede esteja interessado em otimizar os caminhos usados pelas demandas de tráfego individuais, por meio dos protocolos SPF ele não pode definir diretamente um caminho e atribuí-lo a um fluxo particular. É preciso que ele defina os valores dos custos administrativos de tal forma que o caminho definido torne-se um dos caminhos mais curtos, ou mesmo o único mais curto, entre os nós de origem e destino do fluxo. Esta tarefa não é tão óbvia quanto pode parecer, uma vez que a alteração dos valores dos pesos dos circuitos envolvidos na definição de um caminho para um fluxo em particular

pode mudar os caminhos que tenham sido definidos para outros fluxos.

Tradicionalmente as técnicas de Engenharia de Tráfego (em inglês, *Traffic Engineering*, TE) existentes definem os valores dos pesos de forma otimizada, tendo como restrição apenas o grau de utilização dos circuitos [22, 25, 72]. Nos casos de TE baseada na economia de energia, nos quais se enquadra o presente trabalho e outros como Amaldi et al. [124], o que se busca é a definição ótima dos custos de tal forma a definir caminhos mais curtos energeticamente eficientes. Para tanto, parte-se da premissa de que um elemento de rede pode ser desligado quando nenhum tráfego está fluindo por ele, assim, um circuito pode ser desligado atribuindo-lhe um valor de custo muito alto, de tal forma a excluí-lo das árvores de caminhos mais curtos para todas as fontes. De forma análoga, os roteadores podem ser desligados a partir de um aumento suficiente dos custos de todos os circuitos que nele estejam conectados. Assim, assume-se que dispositivos de rede que não estejam em uso (placas de rede e dispositivos inteiros) podem ser colocados em hibernação, seja através de procedimentos centralizados de gerenciamento de rede ou mecanismos automatizados baseados na monitoração do tráfego [124].

De forma elucidativa, considerando-se $G = (V, E)$ o grafo orientado que representa uma rede IP composta por roteadores (chassis) e circuitos (ou interfaces às quais os enlaces estão conectados), a decisão de quais elementos (roteadores e circuitos) devem ser desligados, ou seja, a determinação dos custos dos enlaces de tal forma a minimizar o consumo total de energia da rede, garantindo também que todas as demandas de tráfego sejam roteadas e que um limiar de utilização máxima dos circuitos seja satisfeito, pode ser claramente tratada como um Problema de Otimização, conforme apresentado por exemplo em Amaldi et al. [124], sendo a abordagem escolhida na presente tese e cuja descrição detalhada é transcrita a seguir.

A quantidade de circuitos, enlaces ou *links*, que equivale à cardinalidade de E , é dada por nE . A quantidade de roteadores, ou cardinalidade de V , é dada por nV . Numa rede hierárquica os nós são divididos em dois conjuntos: nós de borda e nós de núcleo. Os nós de borda podem ser tanto a origem quanto o destino dos fluxos, enquanto os nós de núcleo exercem apenas o

papel de roteadores de trânsito. Quando a rede não é hierárquica estes papéis se confundem e um nó de núcleo pode também funcionar como nó de destino ou origem do tráfego. Na Figura 3.6 são apresentados exemplos de topologias de rede extraídas do banco de dados SNDlib [73]. A Figura 3.6(a) é um exemplo de rede não hierárquica, onde não há distinção entre nós de núcleo e nós de entrada. A Figura 3.6(b), por sua vez, é um exemplo de rede hierárquica, possuindo nítida distinção entre nós de núcleo e nós de borda ou entrada.

Para cada nó $v \in V$ definem-se o conjunto de circuitos $\delta^+(v)$ saíntes do nó v e o conjunto de circuitos $\delta^-(v)$ entrantes no nó v , ou seja, $\delta^+(v) = \{e \in E : a(e) = v\}$ e $\delta^-(v) = \{e \in E : b(e) = v\}$, onde $a(e) \in V$ e $b(e) \in V$ representam, respectivamente, os nós de origem e de destino do circuito $e \in E$.

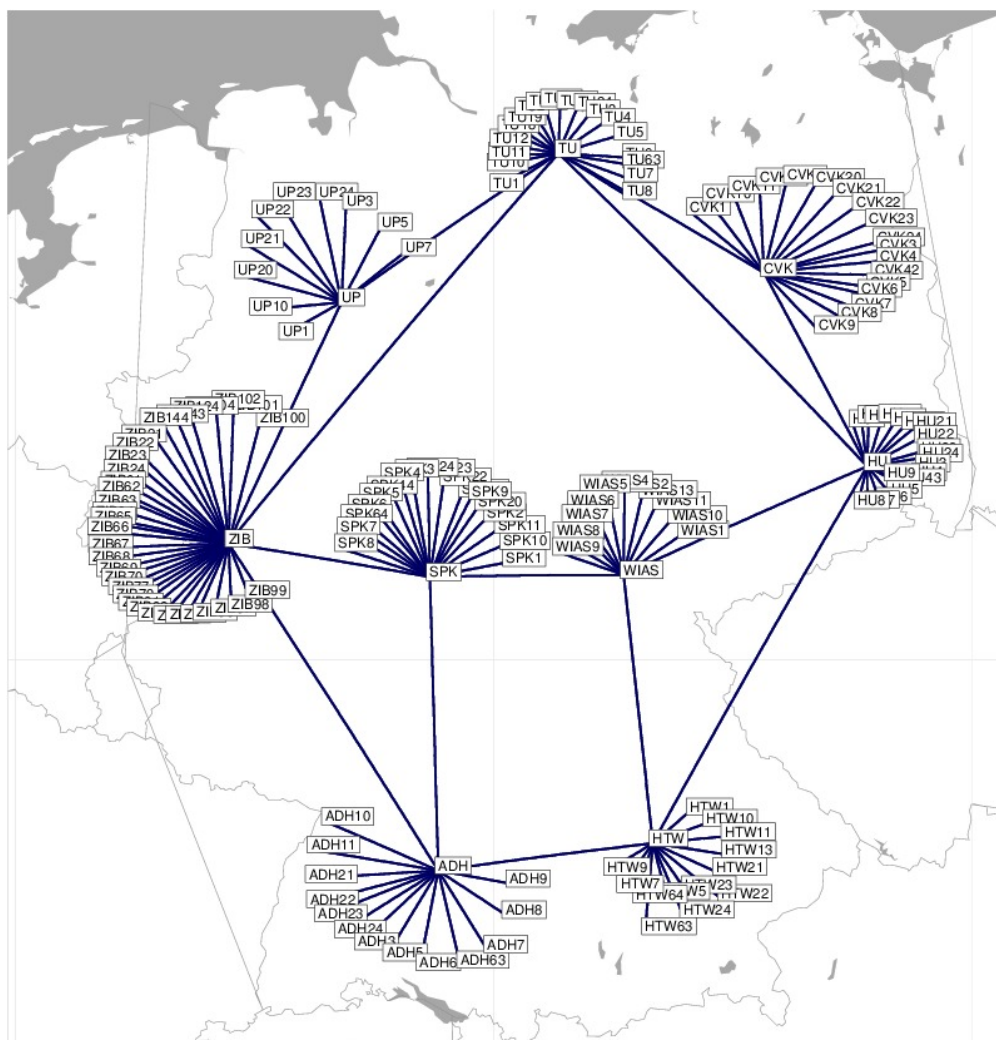
Cada circuito $e \in E$ tem uma capacidade denotada por C_e . Além disso, a cada circuito $e \in E$ é atribuído um custo representado por w_e . Dependendo do caso, o custo pode ser constante ou variável. O vetor $w = w_e : e \in E$ é denominado sistema de pesos (ou vetor de pesos, ou ainda sequencia de pesos). Em geral os pesos dos circuitos devem assumir valores inteiros positivos limitados superiormente (o OSPF, por exemplo, assume que $w_e \in \{1, 2, \dots, \kappa\}$, onde $\kappa = 2^{16} - 1$).

Para fins de otimização é ciente de que esta relaxação não invalida o resultado, consideram-se os sistemas de pesos w como sendo contínuos, com $1 \leq w_e \leq \kappa$, $e \in E$, assumindo que eles são regulares. Um sistema de pesos w é dito regular se, para quaisquer dois caminhos que tenham comprimentos diferentes, estes comprimentos possuem uma diferença de no mínimo 1. Certamente, esta condição de regularidade é satisfeita quando os pesos são números inteiros positivos. Se os pesos podem eventualmente assumir valores racionais positivos, então multiplicando-se todos os pesos w_e por um valor positivo grande o suficiente assegurar-se-á a regularidade.

A demanda de tráfego gerada no nó $s \in V$ e direcionada ao nó de destino $t \in V$ é definida como d_{st} . Se não há demanda de tráfego entre um par de nós $(s, t) \in V^{[2]}$, então, representa-se simplesmente $d_{st} = 0$. O volume total de demanda de tráfego destinado ao nó $t \in V$ é definido como D_t , de tal forma



3.6(a): Rede Abilene (faz parte do projeto Internet2, interligando as principais universidades dos Estados Unidos)



3.6(b): Rede Brain (rede de dados de alta velocidade que interliga instituições científicas e culturais em Berlim)

Figura 3.6 – Exemplos de topologias de rede extraídas do SNDlib

Fonte: SNDlib [73]

que $D_t = \sum_{s \in V \setminus \{t\}} d_{st}$.

O conjunto de todos caminhos elementares (isto é, sem *loop*) no grafo de rede é denotado por P . Cada caminho $p \in P$ é representado por seu conjunto de circuitos de tal forma que $p \subseteq E$. O comprimento do caminho $p \in P$ com relação ao sistema de pesos w será representado por $w(p)$, e qualquer caminho mais curto com relação ao sistema pesos w será referenciado como um caminho mais curto, com custos w . Claramente $w(p) = \sum_{e \in p} w_e$.

Assume-se ainda que as capacidades dos circuitos $C = (C_e : e \in E)$ e as demandas de tráfego $d = (d_{st} : (s, t) \in V^{[2]})$ são fixas e conhecidas.

Por ser um dos IGPs mais amplamente adotado, assume-se como protocolo de roteamento o OSPF, com suporte ao protocolo ECMP, no qual as demandas de tráfego são uniformemente distribuídas, a cada nó, por dois ou mais circuitos de saída que pertençam aos caminhos mais curtos até o destino. Em outras palavras, a cada nó $v \in V$ o fluxo total X_{vt} de v destinado a qualquer nó $t \in V$, $v \neq t$ (X_{vt} é composto pelo tráfego que passa pelo nó v e também pelo tráfego que é originado nele) é dividido igualmente entre todos os circuitos de saída do nó v que pertencem aos caminhos mais curtos, com custos w , de v para t . Esta regra é ilustrada na Figura 3.7 para uma rede onde os custos de todos os circuitos são iguais e unitários ($w_e = 1$) e só há uma demanda de fluxo, também unitária, entre os nós s e t ($d_{st} = 1$). Há três caminhos mais curtos entre s e t : $s - a - c - t$, $s - a - d - t$ e $s - b - e - t$. Com base nas regras do protocolo ECMP, o fluxo no nó s é dividido em duas partes iguais, o mesmo ocorre com o fluxo no nó a , mas nos nós restantes o fluxo não é dividido. Como resultado tem-se que os fluxos nos circuitos são os seguintes: $x_{sa} = x_{sb} = x_{be} = x_{et} = 1/2$ e $x_{ac} = x_{ad} = x_{ct} = x_{dt} = 1/4$. Ressalta-se que, uma vez que o grafo é orientado, se t passasse a ser o nó de origem do fluxo e s o nó de destino, os fluxos ECMP resultantes nos circuitos passariam a ser: $x_{tc} = x_{td} = x_{te} = x_{ca} = x_{da} = x_{eb} = x_{bs} = 1/3$ e $x_{as} = 2/3$.

Denotando-se como $x_{et}(w)$ o fluxo ECMP induzido pelo sistema de custos w no enlace $e \in E$, para uma dada rede com demandas de tráfego definidas e conhecidas, os fluxos $x_e(w)$, $e \in E$, dependem exclusivamente do sistema de

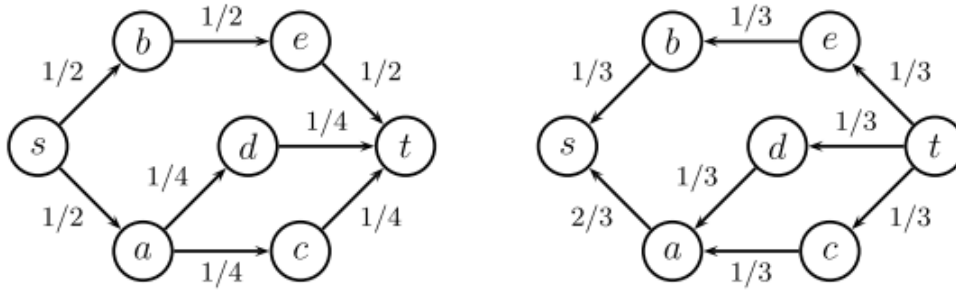


Figura 3.7 – Representação da regra ECMP

Fonte: Koster e Muñoz [72]

custos w .

São definidas ainda as variáveis $u_{et} \in \{0, 1\}$ para todo $t \in V$ e $e \in E \setminus \{\delta^+(t)\}$. Estas variáveis descrevem os chamados grafos de caminho mínimo (em inglês, *shortest path graphs*) para todos os destinos $t \in V$. A variável u_{et} assume o valor igual a 1 se, e somente se, há um caminho mais curto do nó $a(e)$ para o nó t que contenha o circuito e . Ou seja, u_{et} indica se o circuito e pertence ao caminho mais curto até o nó t .

A variável x_{et} representa o fluxo agregado que flui pelo circuito e a partir de todas as possíveis origens com destino ao nó t . Se num dado nó v o fluxo agregado em direção a t é uniformemente dividido e enviado por diversos caminhos mais curtos, então, este valor (ou parcela) comum de fluxo é representado pela variável z_{vt} . De fato, a mesma quantidade de fluxo z_{vt} é enviada por todos os circuitos $e \in \delta^+(v)$ que pertencem a pelo menos um destes caminhos mais curtos.

Com o intuito de evitar o congestionamento é definida a variável Z que representa o limiar de máxima utilização permitida nos circuitos, ou seja, a máxima fração de utilização do enlace em relação a sua capacidade.

Define-se ainda a variável r_{vt} como sendo o potencial de cada nó $v \in V$ com respeito a cada destino $t \in V$ e os pesos w . Assim, se $r_{tt} = 0$, o menor potencial possível r_{vt} do nó v é exatamente a distância do nó v ao nó t com respeito aos pesos w . Em outras palavras, r_{vt} pode ser definida como o comprimento do caminho mais curto, com custos w , entre os nós v e t .

Finalmente, representando-se a potência consumida pelo circuito e como

P_e^E e a potência consumida pelo nó t como P_t^V , e definindo-se as variáveis binárias α_e e β_t , que indicam respectivamente se o **circuito** e ou o **roteador** t estão ou não ativos, a formulação do problema MILP visando à minimização do consumo total de energia da rede é dada por:

$$\min z = \sum_{e \in E} P_e^E \cdot \alpha_e + \sum_{t \in V} P_t^V \cdot \beta_t \quad (3-1)$$

sujeito a:

$$\sum_{e \in \delta^-(t)} x_{et} = D_t \quad t \in V \quad (3-2)$$

$$\sum_{e \in \delta^+(v)} x_{et} - \sum_{e \in \delta^-(v)} x_{et} = d_{vt} \quad t \in V, v \in V \setminus \{t\} \quad (3-3)$$

$$\sum_{t \in V \setminus \{a(e)\}} x_{et} \leq Z \cdot C_e \quad e \in E \quad (3-4)$$

$$0 \leq x_{et} \leq D_t \cdot u_{et} \quad e \in E, t \in V \setminus \{a(e)\} \quad (3-5)$$

$$0 \leq z_{a(e),t} - x_{et} \leq D_t \cdot (1 - u_{et}) \quad e \in E, t \in V \setminus \{a(e)\} \quad (3-6)$$

$$1 - u_{et} \leq w_e + r_{b(e),t} - r_{a(e),t} \quad e \in E, t \in V \setminus \{a(e)\} \quad (3-7)$$

$$w_e + r_{b(e),t} - r_{a(e),t} \leq M \cdot (1 - u_{et}) \quad e \in E, t \in V \setminus \{a(e)\} \quad (3-8)$$

$$r_{st} \geq 1 \quad (s, t) \in V^{|2|} \quad (3-9)$$

$$r_{vv} = 0 \quad v \in V \quad (3-10)$$

$$\alpha_e \leq \beta_{a(e)} \quad e \in E \quad (3-11)$$

$$\alpha_e \leq \beta_{b(e)} \quad e \in E \quad (3-12)$$

$$u_{et} \leq \alpha_e \quad \forall t \in V \quad (3-13)$$

$$u_{et} \in \{0, 1\} \quad e \in E, t \in V \setminus \{a(e)\} \quad (3-14)$$

$$w_e \in \{1, 2, \dots, \kappa\} \quad e \in E \quad (3-15)$$

$$x, z, r, Z \geq 0 \quad (3-16)$$

$$x, z, r, Z \in \mathbb{R} \quad (3-17)$$

Na eq. (3-8), M é uma constante (“*big-M*”) cujo valor não deve ser inferior à diferença em comprimento entre quaisquer dois caminhos no grafo da rede. Assim, por exemplo, $M = \kappa \cdot nE$ seria um valor suficiente.

A eq. (3-2) e a eq. (3-3) descrevem as restrições clássicas de conservação ou preservação de fluxo. A eq. (3-4) é a restrição de máxima utilização do circuito, impondo que o fluxo total que trafega em cada circuito não pode exceder o limiar de utilização Z , sendo $0 \leq Z \leq 1$. As duas equações seguintes expressam a regra ECMP de roteamento do tráfego por meio da variável binária u : a eq. (3-5) força o tráfego destinado ao nó t a usar apenas os circuitos permitidos pela configuração de roteamento definida por u (isto é, circuitos $e \in E$ com $u_{et} = 1$), enquanto a eq. (3-6) garante que em cada nó o tráfego destinado ao nó t é uniformemente distribuído pelos circuitos atribuídos a este destino. Para o nó $v \in V$ e o destino $t \in V$ este valor em comum de distribuição uniforme é expresso pela variável z_{vt} . As restrições de roteamento pelo caminho mais curto expressas a partir da eq. (3-7) à eq. (3-10) garantem que o vetor de roteamento u define os caminhos mais curtos de forma consistente com o sistema de custos w e evita que circuitos desligados pertençam a um caminho mais curto. Cada variável r_{vt} expressa a distância (comprimento do caminho mais curto com relação a w) entre o nó v e o nó t . A quantidade $q_e = w_e + r_{b(e)t} - r_{a(e)t}$ mede a diferença entre o comprimento do caminho mais curto que começa no nó $a(e)$, passa pelo circuito e e termina no nó t e a distância entre o nó de início de e e o nó t . Assim, o circuito e está em um caminho mais curto até t se, e somente se, $w_e + r_{b(e)t} - r_{a(e)t} = 0$. Finalmente, as restrições expressas pela eq. (3-11), pela eq. (3-12) e pela eq. (3-13) asseguram que se um nó for desligado todos os circuitos a ele conectados também o serão. Obviamente um nó só pode ser desligado se não houver nenhuma demanda de tráfego oriunda ou direcionada a ele.

Infelizmente, a formulação MILP anteriormente descrita, apesar de caracterizar bem o problema de interesse, é de elevada complexidade computacional, mesmo para redes de pequeno porte. Amaldi et al. [124], por exemplo, empregaram formulação semelhante e a testaram com seis diferentes redes extraídas do banco de dados SNDlib [73] (com cerca de 10 nós e menos de 50 circuitos em cada uma). Os resultados computacionais são apresentados no Quadro 3.2 e fica evidente que o modelo MILP, neste caso, foi competitivo apenas quando a demanda de tráfego era muito baixa (porcentagens de 1 e 10% da matriz

de tráfego) e a solução final era uma árvore simples. Para os casos com demanda de tráfego maior (30% da matriz de tráfego), o programa de otimização CPLEX 12.2, rodando em um computador com processador Intel i7, com 4 núcleos, *multi-thread* 8x, e equipado com 8GB de memória RAM, não conseguiu encontrar uma solução não-trivial mesmo após um dia de processamento. Cabe esclarecer que é considerada trivial uma rede onde todos os elementos estão ativos e todos os custos são unitários.

Quadro 3.2 – Resultados computacionais obtidos em seis diferentes topologias de rede descritas no SNDlib (com cerca de 10 nós e menos de 50 circuitos em cada uma)

Rede e Tráfego	V-E-D	Não-trivial	Ótima	t(s)
abilene-1%	12-15-132	sim	sim	0,6
abilene-10%	12-15-132	sim	sim	0,6
abilene-30%	12-15-132	sim	sim	339,5
dfn-bwin-1%	10-45-90	sim	sim	182,1
dfn-bwin-10%	10-45-90	sim	sim	257.838,0
dfn-bwin-30%	10-45-90	não	não	648.927,0
dfn-gwin-1%	11-47-110	sim	sim	348,4
dfn-gwin-10%	11-47-110	sim	sim	591.561,0
dfn-gwin-30%	11-47-110	não	não	848.884,0
di-yuan-1%	11-42-22	sim	sim	2.551,5
di-yuan-10%	11-42-22	sim	sim	1.867,1
di-yuan-30%	11-42-22	não	não	73.835,9*
pdh-1%	11-34-24	sim	sim	11,1
pdh-10%	11-34-24	sim	sim	1.630,6
pdh-30%	11-34-24	sim	não	25.137,1*
polska-1%	12-18-66	sim	sim	6,8
polska-10%	12-18-66	sim	sim	16,9
polska-30%	12-18-66	sim	sim	29.348,5

*Memória excedida

Fonte: Amaldi et al. [124]

Face a esta complexidade computacional, diversos trabalhos têm proposto

métodos heurísticos para a sua resolução [124].

Ratifica-se, portanto, que o grande diferencial do presente trabalho é o emprego de técnicas de decomposição [37] na busca da solução ótima para o problema de otimização aqui formulado no qual se deseja determinar os elementos e circuitos que devem ser desligados com o intuito de economizar o máximo de energia possível, assegurando, contudo, que a utilização dos circuitos não fique acima de um limiar definido como ideal para que as aplicações funcionem adequadamente, conforme detalhado no próximo capítulo.

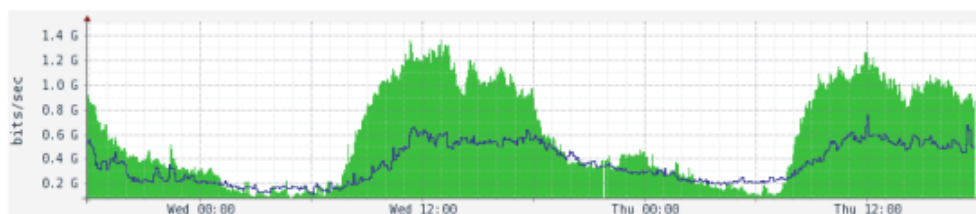
O emprego de Programação Matemática para a resolução do problema de Engenharia de Tráfego baseada na economia de energia caracteriza-se como uma técnica extremamente adequada para a definição da arquitetura de rede que melhor atenda às necessidades dos serviços por ela suportados. No entanto, como bem descrito no capítulo anterior, a obtenção da solução ótima (ou mesmo das soluções factíveis) para esse tipo de problema de otimização não é simples e muitas vezes inviável computacionalmente.

No presente capítulo é detalhado, então, o método ora proposto com o intuito de transpor esta inviabilidade, o qual assume como premissas a modelagem matemática apresentada no capítulo anterior e os seguintes aspectos inerentes às redes IP, que já foram brevemente citados anteriormente, mas sobre os quais se discorre mais profundamente a seguir:

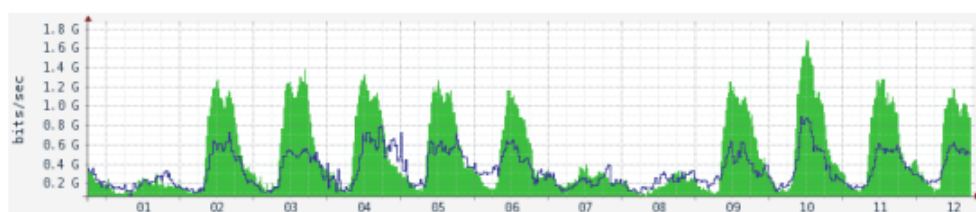
1. Comportamento do tráfego na rede típico e conhecido

Sabe-se que o comportamento do tráfego no *backbone* de redes IP varia bastante ao longo do dia, mas este mesmo padrão de comportamento se repete ao longo da semana e, conseqüentemente, ao longo dos meses, ou seja, ele possui um perfil típico e conhecido [33, 59]. Para exemplificar considere-se a Figura 4.1, que apresenta as estatísticas de tráfego observadas em uma interface de 2,5 Gbps de um dos roteadores que compunha em 2012 o *backbone* da Rede Italiana Acadêmica e de Pesquisa (GARR, sigla derivada do nome em italiano *Gestione Ampliamento Rete Ricerca* [5]). A Figura 4.1(a) mostra o tráfego diário, evidenciando os baixos níveis de utilização do circuito tipicamente durante o período noturno (das 20h às 7h). Na Figura 4.1(b) é ilustrado o tráfego semanal, no qual se pode

observar que as flutuações diárias de tráfego apresentam um padrão de comportamento típico ao longo de todos os dias da semana, com exceção dos sábados e domingos, quando os níveis de utilização do circuito assemelham-se aos observados durante a noite dos dias úteis.



4.1(a): Visão diária



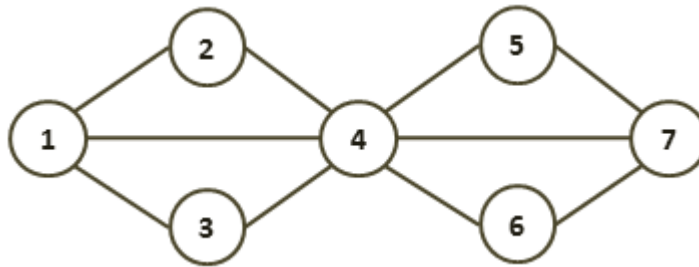
4.1(b): Visão semanal

Figura 4.1 – Comportamento típico do tráfego em um circuito de *backbone*
Fonte: Cianfrani et al. [121]

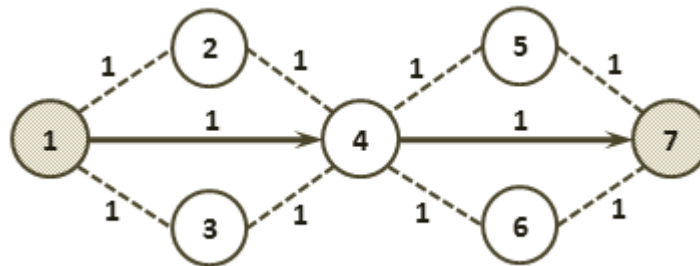
2. Matriz de tráfego conhecida e IGP flexível

Um dos protocolos de roteamento interno mais usado é o OSPF. Numa rede empregando este protocolo cada roteador calcula sua própria tabela de roteamento com base nas métricas (custos administrativos) associados a cada enlace. É definida pelo algoritmo de Dijkstra como melhor rota (ou caminho mais curto) aquela que apresenta menor custo total. Tais custos podem ser definidos manualmente, possibilitando ao administrador da rede influenciar a definição dos caminhos mais curtos e, conseqüentemente, a escolha dos circuitos e roteadores a serem usados para o encaminhamento dos fluxos de tráfego. Para exemplificar, na rede representada na Figura 4.2(a), considerando unitários todos os custos OSPF, o caminho mais curto definido pelo algoritmo de Dijkstra para um tráfego tendo como origem o nó 1 e destino o nó 7 seria o trecho 1–4–7,

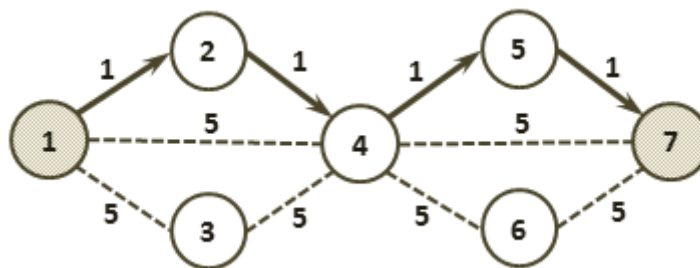
cujo custo total é 2, conforme destacado na Figura 4.2(b). Caso o administrador da rede desejasse definir um caminho mais curto diferente entre os mesmos nós 1 e 7, seria necessário apenas alterar os custos envolvidos, sem alterar o protocolo, como representado na Figura 4.2(c), onde o novo trecho escolhido é o 1-2-4-5-7, com custo total 4.



4.2(a): Grafo representando a rede



4.2(b): SPF entre 1 e 7: 1-4-7 (Custo total: 2)



4.2(c): SPF entre 1 e 7: 1-2-4-5-7 (Custo total: 4)

Figura 4.2 – Definição do SPF numa rede cujo IGP é o OSPF

Fonte: Elaborada pela autora

3. Consumo elevado no *backbone* da rede

Conforme proposto por Tucker [51] a relação entre a potência consumida e a capacidade de comutação dos roteadores pode ser representada pela fórmula $P = C^{2/3}$, onde P corresponde à potência

consumida em Watts e C corresponde à capacidade (ou *throughput*) suportada pelo roteador em Mbps. Assim, como os dispositivos presentes no *backbone* da rede concentram elevados volumes de tráfego, consequentemente, consomem excessivas cargas de energia. No Quadro 4.1 são apresentados valores de capacidade suportada e energia consumida por alguns modelos de roteadores de grande porte do fabricante Juniper, evidenciando que, à medida que aumenta a capacidade do equipamento, aumenta também sua potência consumida. Portanto, é evidente que a economia de energia obtida com o desligamento de um único roteador de *backbone* é mais expressiva do que a decorrente do desligamento de um roteador de acesso.

Quadro 4.1 – Potência consumida e capacidade suportada referente a alguns modelos de roteadores Juniper

Modelo	Capacidade	Potência Consumida
M10i	16 Gbps	494,4 W
M120	120 Gbps	2,068 kW
T640	640 Gbps	7,296 kW

Fonte: JUNIPER [6]

4. Potência consumida por um roteador decorre do consumo de energia do chassis e dos módulos que o compõem

A maior parcela da energia consumida pelos elementos de rede está associada à potência demandada pelo chassis isoladamente, mas as placas que o compõem também contribuem de forma significativa para este consumo. Na Figura 4.3 são apresentadas configurações típicas de um roteador Cisco 12000 [7] e o respectivo consumo de energia (cada configuração está designada de acordo com as placas instaladas, cujas abreviaturas estão apresentadas no Quadro 4.2). Fica evidente que a potência consumida aumenta à medida que aumenta a quantidade de módulos instalados no equipamento. Assim, o desligamento apenas de interfaces às quais os circuitos

estão conectados, apesar de inferior ao desligamento do roteador inteiro, também oferece uma economia de energia.

Quadro 4.2 – Placas do roteador Cisco 12000

PLACA	ABREVIACÃO
Route Processor	RP
10 Gb/s Switching fabric	CSC
4 port GE line card	4GE
4 port OC-3/POS line card	OC-3
4 port OC-12/POS line card	OC-12
1 port OC-48/POS line card	OC-48

Fonte: Adaptado de Chabarek et al. [49]

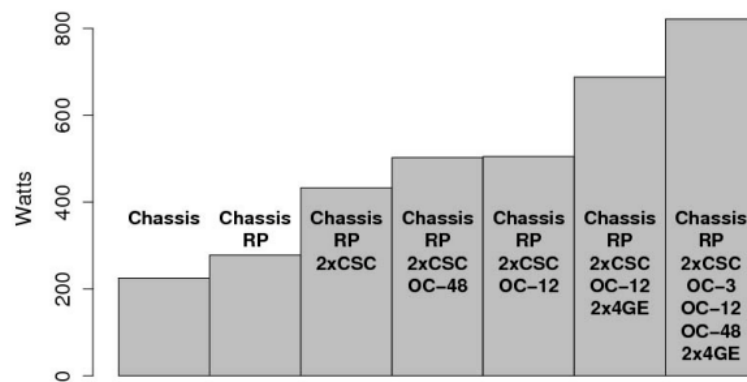


Figura 4.3 – Consumo de energia para diferentes configurações do roteador Cisco 12000

Fonte: Chabarek et al. [49]

5. Superdimensionamento das redes

Os circuitos que compõem a rede são projetados normalmente com capacidades superiores às taxas demandadas durante os horários de pico de tráfego (horários durante os quais o tráfego na rede é maior) para assegurar o correto funcionamento dos serviços em todas as situações, baixa e plena cargas [27], inclusive considerando o reroteamento de tráfego em caso de falhas dos circuitos ou roteadores principais. Como a capacidade dos circuitos é fixa, a taxa de transferência agregada fluindo em cada *link* costuma ser

inferior a 40%, ou seja, durante a maior parte do tempo os recursos ficam ociosos, mas consumindo o mesmo nível de energia de quando estão em plena carga.

6. Matrizes de restrições esparsas

Em função da natureza esparsa e da estrutura em blocos das matrizes de restrição envolvidas, é viável a aplicação do método de decomposição de Benders como meio extremamente eficaz para reduzir a complexidade computacional e o tempo de processamento na resolução do problema.

4.1

Parâmetros e Variáveis Envolvidos

Para facilitar o entendimento da metodologia proposta a ser detalhada no presente capítulo, os parâmetros e variáveis envolvidos são destacados nos Quadros 4.3 e 4.4.

Quadro 4.3 – Parâmetros envolvidos na modelagem matemática

Notação	Definição
V	Conjunto de nós que compõem a rede
E	Conjunto de circuitos que compõem a rede
nV	Quantidade de roteadores que compõem a rede
nE	Quantidade de circuitos que compõem a rede
$\delta^-(v)$	Conjunto de circuitos entrantes no nó v
$\delta^+(v)$	Conjunto de circuitos saíntes do nó v
$\#\delta^+(t)$	Número de <i>links</i> que saem do nó t
$a(e)$	Nó de origem do enlace e
$b(e)$	Nó de destino do enlace e
D_t	Tráfego total destinado ao nó t
d_{vt}	Demanda de tráfego gerada no nó v e destinada ao nó t
C_e	Capacidade do circuito e

Fonte: Adaptado de Koster e Muñoz [72]

Quadro 4.4 – Variáveis envolvidas na modelagem matemática

Var.	Definição	Natureza	Dimensão
x_{et}	Fluxo agregado que trafega pelo circuito e a partir de todas as origens com destino ao nó t	Real	$nE \cdot nV$
z_{vt}	Valor comum em que um fluxo que chega ao nó v em direção ao nó t é dividido numa rede com suporte a ECMP, caso haja diversos caminhos mais curtos entre v e t	Real	$nV \cdot nV$
r_{vt}	Comprimento do caminho mais curto, com custos w , entre os nós v e t	Real	$nV \cdot nV$
w_e	Custo atribuído ao circuito e	“Real”	nE
Z	Máxima fração de utilização do circuito em relação a sua capacidade	Real	1
u_{et}	Variável que indica se o circuito e pertence ao caminho mais curto até o nó t	Binária	$nE \cdot nV$
α_e	Variável que indica se o circuito e está ou não ativo	Binária	nE
β_t	Variável que indica se o roteador t está ou não ativo	Binária	nV

Fonte: Adaptado de Koster e Muñoz [72]

Organizando-se as variáveis de forma matricial, agrupando os conjuntos dos elementos das matrizes de tal modo a aproximá-los de vetores e separando as variáveis contínuas das binárias, têm-se os conjuntos apresentados a seguir:

(i) **Variáveis Binárias**

$$\begin{aligned}
 \underline{u} &= \left[\underbrace{u_{11} \quad u_{21} \quad \dots \quad u_{nE,1}}_{\underline{u}_{,1}} \quad \dots \quad \underbrace{u_{1,nV} \quad u_{2,nV} \quad \dots \quad u_{nE,nV}}_{\underline{u}_{,nV}} \right] \\
 \underline{\alpha} &= [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_{nE}] \\
 \underline{\beta} &= [\beta_1 \quad \beta_2 \quad \dots \quad \beta_{nV}]
 \end{aligned}$$

(ii) Variáveis Contínuas

$$\begin{aligned}\underline{x} &= \begin{bmatrix} \underbrace{x_{11} \ x_{21} \ \dots \ x_{nE,1}}_{\underline{x}_{.,1}} & \dots & \underbrace{x_{1,nV} \ x_{2,nV} \ \dots \ x_{nE,nV}}_{\underline{x}_{.,nV}} \end{bmatrix} \\ \underline{z} &= \begin{bmatrix} \underbrace{z_{11} \ z_{21} \ \dots \ z_{nV,1}}_{\underline{z}_{.,1}} & \dots & \underbrace{z_{1,nV} \ z_{2,nV} \ \dots \ z_{nV,nV}}_{\underline{z}_{.,nV}} \end{bmatrix} \\ \underline{r} &= \begin{bmatrix} \underbrace{r_{11} \ r_{21} \ \dots \ r_{nV,1}}_{\underline{r}_{.,1}} & \dots & \underbrace{r_{1,nV} \ r_{2,nV} \ \dots \ r_{nV,nV}}_{\underline{r}_{.,nV}} \end{bmatrix} \\ \underline{w} &= [w_1 \ w_2 \ \dots \ w_{nE}] \\ Z &\end{aligned}$$

A quantidade total de variáveis pode, portanto, ser expressa por:

$$\underbrace{nV \cdot (2 \cdot nV + nE) + nE + 1}_{\text{Variáveis Contínuas}} + \underbrace{nE \cdot nV + nE + nV}_{\text{Variáveis Binárias}}$$

O que resulta em:

$$\boxed{2 \cdot nV \cdot (nV + nE) + 2 \cdot nE + nV + 1}$$

4.2

Restrições Envolvidas

No tocante às restrições expressas pelas eqs. de (3-2) a (3-13) cabem ainda algumas observações importantes e as mesmas serão detalhadas nas subseções de 4.2.1 a 4.2.5.

4.2.1

Restrições de Conservação de Fluxo

As restrições clássicas de conservação ou preservação de fluxo expressas pela eq. (4-1) e pela eq. (4-2) visam a representar que a quantidade de fluxo saindo do sistema não pode exceder a quantidade entrante.

$$\sum_{e \in \delta^-(t)} x_{et} = D_t \quad t \in V \quad (4-1)$$

$$\sum_{e \in \delta^+(v)} x_{et} - \sum_{e \in \delta^-(v)} x_{et} = d_{vt} \quad t \in V, v \in V \setminus \{t\} \quad (4-2)$$

Escrevendo as restrições de conservação de fluxo na forma algorítmica resulta em:

RESTRIÇÕES DE CONSERVAÇÃO DE FLUXO

```

para cada  $t \in V$  faça
  |
  |    $\sum_{e \in \delta^-(t)} x_{et} = D_t$ 
  |
  |   para cada  $v \in V \setminus \{t\}$  faça
  |   |
  |   |    $\sum_{e \in \delta^+(v)} x_{et} - \sum_{e \in \delta^-(v)} x_{et} = d_{vt}$ 
  |   |
  |   fim
fim

```

Vale elucidar que a notação “ $V \setminus \{t\}$ ” usada advém da Teoria de Conjuntos e visa a representar o conjunto V excluindo-se o conjunto de elementos $\{t\}$.

Com base na expressão algorítmica é possível afirmar facilmente que o número de restrições de preservação de fluxo é dado por $(nV)^2$.

Destaca-se, ainda, que tais restrições podem ser escritas na forma $\underline{A} \cdot \underline{x} = \underline{a}$, onde \underline{A} é uma matriz bloco-diagonal de forma $diag\{A_1, A_2, \dots, A_{nV}\}$ e \underline{a} é um vetor de forma $[a_1^T \ a_2^T \ \dots \ a_{nV}^T]^T$, cujos valores possíveis são representados pela eq. (4-3) e cuja representação esquemática é apresentada na Figura 4.4.

$$A_t(i, j) = \begin{cases} -1 & i \neq t, j \in \delta^+(v) \\ +1 & j \in \delta^-(v) \\ 0 & \text{caso contrário} \end{cases} \quad a_t(k) = \begin{cases} D_t & k = t \\ d_{kt} & k \neq t \end{cases} \quad (4-3)$$

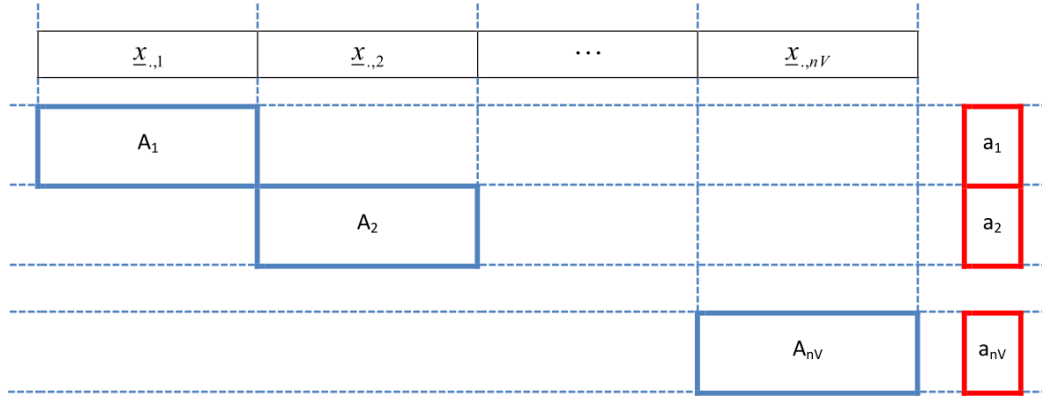


Figura 4.4 – Representação esquemática das restrições de conservação de fluxo

Fonte: Elaborada pela autora

4.2.2

Restrição de Capacidade dos Circuitos

Considerando o requisito de garantia da Qualidade de Serviço, é imposta a restrição de capacidade ou de limitação do nível de utilização máxima dos circuitos, que define que a parcela de fluxo total em cada circuito não pode exceder a fração de máxima utilização Z , e pode ser expressa pela eq. (4-4).

$$\sum_{t \in V \setminus \{a(e)\}} x_{et} \leq Z \cdot C_e \Leftrightarrow \sum_{t \in V \setminus \{a(e)\}} x_{et} - Z \cdot C_e \leq 0 \quad e \in E \quad (4-4)$$

Escrevendo esta restrição na forma algorítmica resulta em:

RESTRIÇÃO DE CAPACIDADE OU DE LIMITAÇÃO DO NÍVEL DE UTILIZAÇÃO MÁXIMA DOS CIRCUITOS

para cada $e \in E$ **faça**

$$\left| \sum_{t \in V \setminus \{a(e)\}} x_{et} - Z \cdot C_e \leq 0 \right.$$

fim

Com base na expressão algorítmica é possível afirmar facilmente que o número de restrições de capacidade dos circuitos é dado por nE . Além disso, observa-se que tais restrições podem ser escritas na forma $\underline{G} \cdot \underline{x} - Z \cdot \underline{K} \leq 0$, onde \underline{G} é uma matriz composta pelas matrizes $\{G_1, G_2, \dots, G_{nV}\}$, como indicado na Figura 4.5, e \underline{K} é um vetor formado pelas capacidades C_j . Os valores que podem ser assumidos por \underline{G} e \underline{K} são representados pela eq. (4-5).

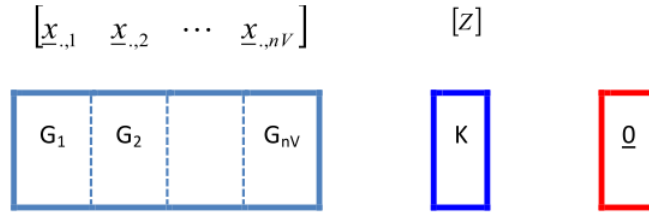


Figura 4.5 – Representação esquemática da restrição de capacidade dos circuitos

Fonte: Elaborada pela autora

$$G_t(i, j) = \begin{cases} 1 & i = j, t \in V \setminus \{a(i)\} \\ 0 & \text{caso contrário} \end{cases} \quad K_t(j) = -C_j \quad (4-5)$$

4.2.3

Restrições ECMP

Uma vez que se adotou a premissa de que os fluxos podem ser distribuídos uniformemente por caminhos distintos desde que estes sejam os caminhos mais curtos, são consideradas também as restrições ECMP, representadas pela eq. (4-6) e pela eq. (4-7).

$$0 \leq x_{et} \leq D_t \cdot u_{et} \quad e \in E, t \in V \setminus \{a(e)\} \quad (4-6)$$

$$0 \leq z_{a(e),t} - x_{et} \leq D_t \cdot (1 - u_{et}) \quad e \in E, t \in V \setminus \{a(e)\} \quad (4-7)$$

Escrevendo de forma algorítmica resulta em:

RESTRITÕES ECMP

para cada $e \in E$ **faça**

para cada $t \in V \setminus \{a(e)\}$ **faça**

$$\left. \begin{array}{l} 0 \leq x_{et} \leq D_t \cdot u_{et} \\ 0 \leq z_{a(e),t} - x_{et} \\ z_{a(e),t} - x_{et} \leq D_t \cdot (1 - u_{et}) \end{array} \right\} \Leftrightarrow \begin{cases} x_{et} - D_t \cdot u_{et} \leq 0 \\ x_{et} - z_{a(e),t} \leq 0 \\ z_{a(e),t} - x_{et} + D_t \cdot u_{et} \leq D_t \end{cases}$$

fim

fim

Com base na expressão algorítmica, afirma-se facilmente que o número total de restrições ECMP é dado por $3 \cdot nE \cdot (nV - 1)$. Tais restrições podem ser escritas na forma $\underline{B} \cdot \underline{x} + \underline{C} \cdot \underline{z} + \underline{D} \cdot \underline{u} = \underline{b}$, onde \underline{B} , \underline{C} e \underline{D} são matrizes bloco-diagonal das formas $\text{diag}\{B_1, B_2, \dots, B_{nV}\}$, $\text{diag}\{C_1, C_2, \dots, C_{nV}\}$ e $\text{diag}\{D_1, D_2, \dots, D_{nV}\}$, respectivamente, conforme representado esquematicamente na Figura 4.6.

O número de linhas de cada uma das matrizes B_t , C_t e D_t é dado por $3 \cdot [nE - \#\delta^+(t)]$. E o vetor \underline{b} é da forma $[\underline{b}_1^T \quad \underline{b}_2^T \quad \dots \quad \underline{b}_{nV}^T]^T$, onde \underline{b}_t possui também dimensão $3 \cdot [nE - \#\delta^+(t)]$.

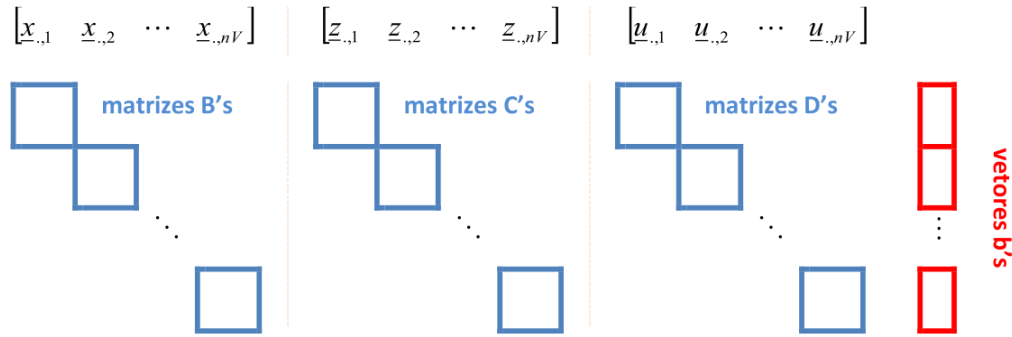


Figura 4.6 – Representação esquemática das restrições ECMP

Fonte: Elaborada pela autora

4.2.4

Restrições de Roteamento pelo Caminho Mais Curto

Uma vez que o protocolo de roteamento escolhido define as rotas com base em um sistema de custos w , deve-se garantir que as rotas ativas sejam, de fato, as que apresentam menor custo, seguindo o SPF, o que é representado pela eq. (4-8) e pela eq. (4-9).

$$1 - u_{et} \leq w_e + r_{b(e),t} - r_{a(e),t} \quad e \in E, t \in V \setminus \{a(e)\} \quad (4-8)$$

$$w_e + r_{b(e),t} - r_{a(e),t} \leq M \cdot (1 - u_{et}) \quad e \in E, t \in V \setminus \{a(e)\} \quad (4-9)$$

Representando de forma algorítmica, tem-se:

RESTRIÇÕES DE ROTEAMENTO PELO CAMINHO MAIS CURTO

para cada $e \in E$ **faça**

para cada $t \in V \setminus \{a(e)\}$ **faça**

$$1 - u_{et} \leq w_e + r_{b(e)t} - r_{a(e)t} \leq M \cdot (1 - u_{et})$$

$$\Updownarrow$$

$$\begin{cases} -w_e - r_{b(e),t} + r_{a(e),t} - u_{et} \leq -1 \\ w_e + r_{b(e),t} - r_{a(e),t} + M \cdot u_{et} \leq M \end{cases}$$

fim

fim

Com base na expressão algorítmica, afirma-se que o número total de restrições de roteamento pelo caminho mais curto é dado por $2 \cdot nE \cdot (nV - 1)$. Tais restrições podem ser escritas na forma $\underline{E} \cdot \underline{u} + \underline{F} \cdot \underline{r} + \underline{H} \cdot \underline{w} = \underline{c}$, onde \underline{E} e \underline{F} são matrizes bloco-diagonal das formas $\text{diag}\{E_1, E_2, \dots, E_{nV}\}$ e $\text{diag}\{F_1, F_2, \dots, F_{nV}\}$ e \underline{H} é da forma $\text{col}\{H_1, H_2, \dots, H_{nV}\}$, conforme representado esquematicamente na Figura 4.7.

Cabe ressaltar que o número de linhas de cada uma das matrizes E_t , F_t e H_t é dado por $2 \cdot [nE - \#\delta^+(t)]$. Salienta-se ainda que $\text{col}\{H_1, H_2, \dots, H_N\}$ é uma matriz gerada pelo empilhamento vertical ordenado das matrizes $\{H_1, H_2, \dots, H_N\}$.

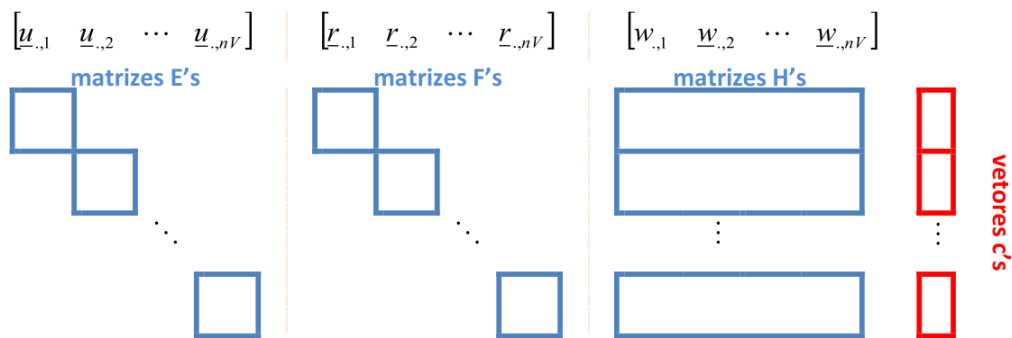


Figura 4.7 – Representação esquemática das restrições de roteamento pelo caminho mais curto

Fonte: Elaborada pela autora

4.2.5

Restrições de Condição do Nó ou do Circuito

Considerando que a economia de energia é obtida a partir do desligamento de nós ou circuitos, são impostas as restrições expressas pelas eqs. de (4-10) a (4-12), de tal forma a assegurar que circuitos que não pertencem aos caminhos mais curtos podem ser desligados, e assegurando também que caso seja possível desligar todos os circuitos (entrantes ou saíntes) de um nó este também pode ser desligado.

$$\alpha_e \leq \beta_{a(e)} \quad e \in E \quad (4-10)$$

$$\alpha_e \leq \beta_{b(e)} \quad e \in E \quad (4-11)$$

$$u_{et} \leq \alpha_e \quad \forall t \in V \quad (4-12)$$

Representando de forma algorítmica tem-se:

RESTRIÇÕES DE CONDIÇÃO DO NÓ OU DO CIRCUITO

para cada $e \in E$ **faça**

$$\alpha_e \leq \beta_{a(e)} \Leftrightarrow \alpha_e - \beta_{a(e)} \leq 0$$

$$\alpha_e \leq \beta_{b(e)} \Leftrightarrow \alpha_e - \beta_{b(e)} \leq 0$$

$$u_{et} \leq \alpha_e \quad \forall t \in V$$

fim

Com base na expressão algorítmica afirma-se que o número total de restrições de condição do nó ou do circuito é dado por $2 \cdot nE + nE \cdot nV$. Estas restrições podem ser convenientemente escritas na forma $\underline{M} \cdot \underline{\alpha} + \underline{N} \cdot \underline{\beta} \leq 0$ e $\underline{R} \cdot \underline{u} - \underline{S} \cdot \underline{\alpha} \leq 0$, conforme representado esquematicamente na Figura 4.8. O número de linhas de \underline{M} e \underline{N} é dado por $2 \cdot nE$ e o número de linhas de \underline{R} e \underline{S} corresponde a $nE \cdot nV$.

4.3

Formulação do Método

Na concepção da formulação matemática apresentada no Capítulo 3, no mesmo problema de otimização, além da preocupação com a economia

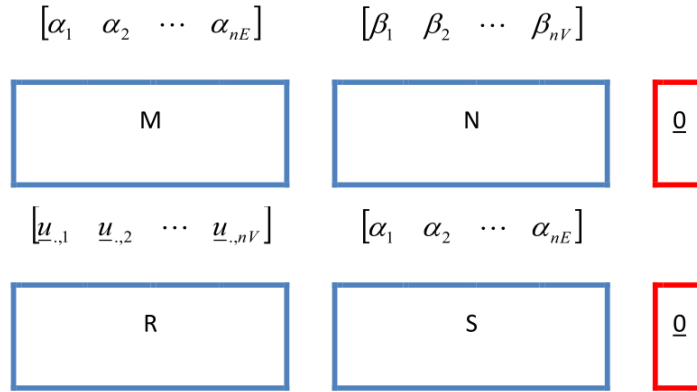


Figura 4.8 – Representação esquemática das restrições de condição do nó ou do circuito

Fonte: Elaborada pela autora

de energia, é considerado outro importante aspecto que atrai a atenção dos administradores de rede: o desempenho. É importante destacar que, fixando isoladamente algumas variáveis, tal problema pode ser facilmente dividido em dois, com funções objetivo e restrições distintas, atendendo a cada um dos aspectos citados separadamente, como mostrado a seguir:

1. Problema de Otimização com Foco em Desempenho ou Eficiência

Neste contexto as variáveis $\alpha's$ e $\beta's$ são fixadas e a função objetivo é dada por:

$$\min z = Z$$

As restrições envolvidas são as de:

- Conservação de Fluxo (Subseção 4.2.1)
- Capacidade dos Circuitos (Subseção 4.2.2)
- ECMP (Subseção 4.2.3)
- Roteamento pelo Caminho Mais Curto (Subseção 4.2.4)

Neste caso, o número total de variáveis é dado por $2 \cdot nV \cdot (nE + nV) + nE + 1$ e o número total de restrições é dado por $(nV)^2 + nE + 5 \cdot nE \cdot (nV - 1)$.

2. Problema de Otimização com Foco em Economia de Energia

Neste contexto a variável Z é eliminada e a função objetivo converte-se,

portanto, em:

$$\min z = \sum_{e \in E} P_e^E \cdot \alpha_e + \sum_{t \in V} P_t^V \cdot \beta_t$$

As restrições envolvidas são as de:

- Conservação de Fluxo (Subseção 4.2.1)
- Capacidade dos Circuitos (Subseção 4.2.2)
- ECMP (Subseção 4.2.3)
- Roteamento pelo Caminho Mais Curto (Subseção 4.2.4)
- Condição do Nó ou do Circuito (Subseção 4.2.5)

Neste caso, o número total de variáveis é dado por $2 \cdot nV \cdot (nE + nV) + 2 \cdot nE + nV$ e o número total de restrições é dado por $nV \cdot (nV + nE) + 4 \cdot nE \cdot (nV - 1) + nE$.

Embora a modelagem matemática dos problemas de otimização supracitados seja de fácil compreensão, eles podem ser reescritos de tal forma a tornarem-se formalmente muito mais simples, como apresentado na eq. (4-13).

$$\begin{aligned} \min_{\underline{y}} \quad & z = \underline{g}^T \cdot \underline{y} \\ \text{sujeito a:} \quad & \\ & \underline{V} \cdot \underline{y} = \underline{r}_0 \\ & \underline{T} \cdot \underline{y} \leq \underline{s}_0 \\ & \underline{y}_L \leq \underline{y} \leq \underline{y}_H \end{aligned} \tag{4-13}$$

onde a variável de decisão \underline{y} é uma agregação adequada das variáveis \underline{x} , \underline{z} , \underline{r} , \underline{w} , \underline{u} , $\underline{\alpha}$ e $\underline{\beta}$, para o problema de otimização com foco em economia de energia; e uma agregação das variáveis \underline{Z} , \underline{x} , \underline{z} , \underline{r} , \underline{w} , e \underline{u} , para o problema de otimização com foco em desempenho. Já \underline{V} e \underline{T} são as matrizes que descrevem as restrições e sua representação gráfica pode ser observada na Figura 4.9 (para o problema de otimização com foco em desempenho) e na Figura 4.10 (para o problema de otimização com foco em economia de energia).

As Figuras 4.9 e 4.10 evidenciam a natureza particular das matrizes \underline{V} e \underline{T} , que não são apenas muito esparsas, mas totalmente compostas por

Figura 4.9 – Matriz de restrições para o problema com foco em desempenho
Fonte: Elaborada pela autora

Figura 4.10 – Matriz de restrições para o problema com foco em economia de energia
Fonte: Elaborada pela autora

blocos de matrizes, o que confirma, como citado anteriormente, que técnicas de decomposição podem ser aplicadas com o intuito de fracionar o problema original em vários outros com dimensão inferior.

Assim, considerando-se que o problema de otimização descrito pela eq. (4-13) pode ser apropriadamente reescrito como apresentado na eq. (4-14), o método de decomposição de Benders, descrito no Apêndice D, pode ser convenientemente aplicado, conforme detalhado nas seções a seguir.

$$\begin{aligned}
 \min_{\underline{y}_1, \underline{y}_2} \quad & z = \underline{g}_1^T \cdot \underline{y}_1 + \underline{g}_2^T \cdot \underline{y}_2 \\
 \text{sujeito a:} \quad & \\
 & \underline{V}_1 \cdot \underline{y}_1 + \underline{V}_2 \cdot \underline{y}_2 = \underline{r}_0 \\
 & \underline{T}_1 \cdot \underline{y}_1 + \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 \\
 & \underline{y}_{1L} \leq \underline{y}_1 \leq \underline{y}_{1H} \\
 & \underline{y}_{2L} \leq \underline{y}_2 \leq \underline{y}_{2H}
 \end{aligned} \tag{4-14}$$

4.4

Decomposição Aplicada ao Problema de Otimização com Foco em Eficiência

Neste contexto ressalta-se que, conforme exposto na seção 4.3, as variáveis $\alpha's$ e $\beta's$ são fixadas, e, conseqüentemente, a matriz \underline{K} , que depende das variáveis $\alpha's$, passa a ser dada por:

$$K_i = -C_i \quad i = 1, 2, \dots, nE$$

Rearrmando linhas e colunas na matriz de restrições (Figura 4.9), obtém-se a estrutura apresentada na Figura 4.11, a partir da qual é possível identificar de maneira mais clara a separação de variáveis necessária ao processo de decomposição.

Usando como referência a eq. (4-14) observa-se que a maneira mais conveniente de dividir as variáveis de decisão é definir as variáveis reais como \underline{y}_2 e as variáveis binárias como a variável complicadora \underline{y}_1 . Além disso, consideram-se \underline{V}_1 e \underline{V}_2 como sendo os coeficientes associados a \underline{y}_1 e

\underline{y}_2 , respectivamente, nas restrições de igualdade, e \underline{T}_1 e \underline{T}_2 , os coeficientes associados a \underline{y}_1 e \underline{y}_2 , respectivamente, nas restrições de desigualdade conforme representado esquematicamente na Figura 4.12.

Variáveis Reais												Variáveis Binárias							
Z	$\underline{x}_{,1}$	$\underline{z}_{,1}$	$\underline{r}_{,1}$	$\underline{x}_{,2}$	$\underline{z}_{,2}$	$\underline{r}_{,2}$...	$\underline{x}_{,nV}$	$\underline{z}_{,nV}$	$\underline{r}_{,nV}$	W	$\underline{u}_{,1}$	$\underline{u}_{,2}$...	$\underline{u}_{,nV}$				
igualdades	A ₁																=	a ₁	
				A ₂													=	a ₂	
			
								A _{nV}									=	a _{nV}	
desigualdades	B ₁	C ₁										D ₁					≤	b ₁	
			F ₁									H ₁	E ₁				≤	c ₁	
			...	B ₂	C ₂								D ₂				...	b ₂	
						F ₂						H ₂	E ₂				≤	c ₂	
								≤	...	
								B _{nV}	C _{nV}							D _{nV}	≤	b _{nV}	
										F _{nV}		H _{nV}			E _{nV}	...	c _{nV}		
	K	G ₁			G ₂			G _{nV}									≤	0	
																≤	0		
																	
																≤	0		

Variáveis Complicadoras

Restrições Complicadoras

Figura 4.11 – Matriz de restrições para o problema com foco em eficiência ou desempenho, após rearranjos

Fonte: Elaborada pela autora

Neste caso, como a componente \underline{V}_1 é nula e a função objetivo depende somente das variáveis reais \underline{y}_2 (ou seja, a parcela $\underline{g}_1^T \cdot \underline{y}_1$ da função objetivo é nula), o problema original pode ser reescrito de forma ainda mais simplificada, conforme representado nas eqs. de (4-15) a (4-19), onde $\underline{y}_1 \in \mathbb{B}^{nY_1}$ e $\underline{y}_2 \in \mathbb{R}^{nY_2}$.

$$\min_{\underline{y}_1, \underline{y}_2} z = \underline{g}_2^T \cdot \underline{y}_2 \quad (4-15)$$

sujeito a:

$$\underline{V}_2 \cdot \underline{y}_2 = r_0 \quad (4-16)$$

$$\underline{T}_1 \cdot \underline{y}_1 + \underline{T}_2 \cdot \underline{y}_2 \leq s_0 \quad (4-17)$$

$$\underline{y}_{1L} \leq \underline{y}_1 \leq \underline{y}_{1H} \quad (4-18)$$

$$\underline{y}_{2L} \leq \underline{y}_2 \leq \underline{y}_{2H} \quad (4-19)$$

Variáveis Reais														Variáveis Binárias					
\mathcal{V}_2	Z	$\underline{x}_{,1}$	$\underline{z}_{,1}$	$\underline{r}_{,1}$	$\underline{x}_{,2}$	$\underline{z}_{,2}$	$\underline{r}_{,2}$...	$\underline{x}_{,nV}$	$\underline{z}_{,nV}$	$\underline{r}_{,nV}$	\underline{w}	$\underline{u}_{,1}$	$\underline{u}_{,2}$...	$\underline{u}_{,nV}$	\mathcal{V}_1		
igualdades		A_1				A_2												=	a_1
																		=	a_2
			
										A_{nV}								=	a_{nV}
desigualdades		B_1	C_1											D_1				≤	b_1
				F_1									H_1	E_1				≤	c_1
				...	B_2	C_2								D_2				...	b_2
							F_2						H_2	E_2				≤	c_2
																		≤	...
										B_{nV}	C_{nV}						D_{nV}	≤	b_{nV}
											F_{nV}	H_{nV}		E_{nV}			...	c_{nV}	
																		≤	0
	K	G_1			G_2					G_{nV}								≤	0
																	
																		≤	0

Figura 4.12 – Relação entre os elementos da matriz de restrições e as variáveis do problema de otimização simplificado, com foco em eficiência

Fonte: Elaborada pela autora

Cabe salientar que $\underline{g}_2 = [1 \ 0 \ 0 \ \dots \ 0]^T$ e as dimensões das matrizes \underline{V}_2 , \underline{T}_1 e \underline{T}_2 , destacadas na Figura 4.12 e representadas isoladamente nas Figuras 4.13, 4.14 e 4.15, são dadas, respectivamente por:

$$\dim(\underline{V}_2) = [nV \cdot nV] \times [1 + nV \cdot (nE + 2 \cdot nV) + nE],$$

$$\dim(\underline{T}_1) = [5 \cdot nE \cdot (nV - 1) + nE] \times [nE \cdot nV] \quad \text{e}$$

$$\dim(\underline{T}_2) = [5 \cdot nE \cdot (nV - 1) + nE] \times [1 + nV \cdot (nE + 2 \cdot nV) + nE].$$

Como também pode ser observado na Figura 4.12, com relação aos termos \underline{r}_0 e \underline{s}_0 , presentes nas eqs. (4-16) e (4-17), tem-se:

$$\underline{r}_0 = \begin{bmatrix} \underline{a}_1^T & \underline{a}_2^T & \dots & \underline{a}_{nV}^T \end{bmatrix}^T$$

$$\underline{s}_0 = \begin{bmatrix} \underline{b}_1^T & \dots & \underline{b}_{nV}^T & \underline{c}_1^T & \dots & \underline{c}_{nV}^T & \underline{0}^T & \dots & \underline{0}^T \end{bmatrix}^T$$

Aplicando-se, então, o método de decomposição de Benders, conforme algoritmo descrito no Apêndice D, ao problema de otimização de interesse, representado pelas eqs. de (4-15) a (4-19), resulta que o **Passo 3** está relacionado a um problema de otimização contendo apenas variáveis binárias enquanto o

	A ₁												
				A ₂									
			...										
									A _{nV}				

Figura 4.13 – Representação esquemática da matriz \underline{V}_2 - Foco em eficiência
Fonte: Elaborada pela autora

D ₁			
E ₁			
	D ₂		
	E ₂		
		...	
			D _{nV}
			E _{nV}

Figura 4.14 – Representação esquemática da matriz \underline{T}_1 - Foco em eficiência
Fonte: Elaborada pela autora

	B ₁	C ₁											
			F ₁										H ₁
			...	B ₂	C ₂								
						F ₂							H ₂
							...						
								B _{nV}	C _{nV}				
										F _{nV}	H _{nV}		
K	G ₁			G ₂				G _{nV}					

Figura 4.15 – Representação esquemática da matriz \underline{T}_2 - Foco em eficiência
Fonte: Elaborada pela autora

Passo 1 envolve apenas variáveis reais. Com o intuito de tornar clara a descrição do método, esta primeira aplicação do algoritmo é denominada como **BENDERS-0**.

Cabe ressaltar que durante a execução do **Passo 1** do **BENDERS-0**, fixando a variável \underline{y}_1 como $\underline{y}_1^{(v)}$, o problema de otimização correspondente para esta etapa pode ser reescrito da seguinte forma:

$$\begin{aligned}
 \min_{\underline{y}_2} \quad & z = \underline{g}_2^T \cdot \underline{y}_2 \\
 \text{sujeito a:} \quad & \\
 & \underline{V}_2 \cdot \underline{y}_2 = r_0 \\
 & \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 - \underline{T}_1 \cdot \underline{y}_1^{(v)} \\
 & \underline{y}_{2L} \leq \underline{y}_2 \leq \underline{y}_{2H}
 \end{aligned} \tag{4-20}$$

Substituindo o termo $\left(\underline{s}_0 - \underline{T}_1 \cdot \underline{y}_1^{(v)}\right)$ por $\underline{s}_0^{(v)}$, tem-se:

$$\begin{aligned}
 \min_{\underline{y}_2} \quad & z = \underline{g}_2^T \cdot \underline{y}_2 \\
 \text{sujeito a:} \quad & \\
 & \underline{V}_2 \cdot \underline{y}_2 = r_0 \\
 & \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0^{(v)} \\
 & \underline{y}_{2L} \leq \underline{y}_2 \leq \underline{y}_{2H}
 \end{aligned} \tag{4-21}$$

A resolução deste problema não é trivial em virtude das dimensões das variáveis \underline{y}_2 envolvidas. No entanto, analisando-se minuciosamente a matriz de restrições associada, observa-se que o problema pode ser reescrito de forma decomponível, conforme representado graficamente na Figura 4.16, considerando-se:

$$\underline{y}_2 = \begin{bmatrix} \underline{h} \\ \underline{w} \end{bmatrix}$$

Cabe salientar que as dimensões das matrizes \underline{V}_{2h} , \underline{T}_{2w} e \underline{T}_{2h} , destacadas na Figura 4.16 e representadas isoladamente nas Figuras 4.17, 4.18 e 4.19, são

dadas, respectivamente por:

$$\dim(\underline{V}_{2h}) = [nV \cdot nV] \times [1 + nV \cdot (nE + 2 \cdot nV)],$$

$$\dim(T_{2w}) = [5 \cdot nE \cdot (nV - 1) + nE] \times nE \quad \text{e}$$

$$\dim(\underline{T}_{2h}) = [5 \cdot nE \cdot (nV - 1) + nE] \times [1 + nV \cdot (nE + 2 \cdot nV)].$$

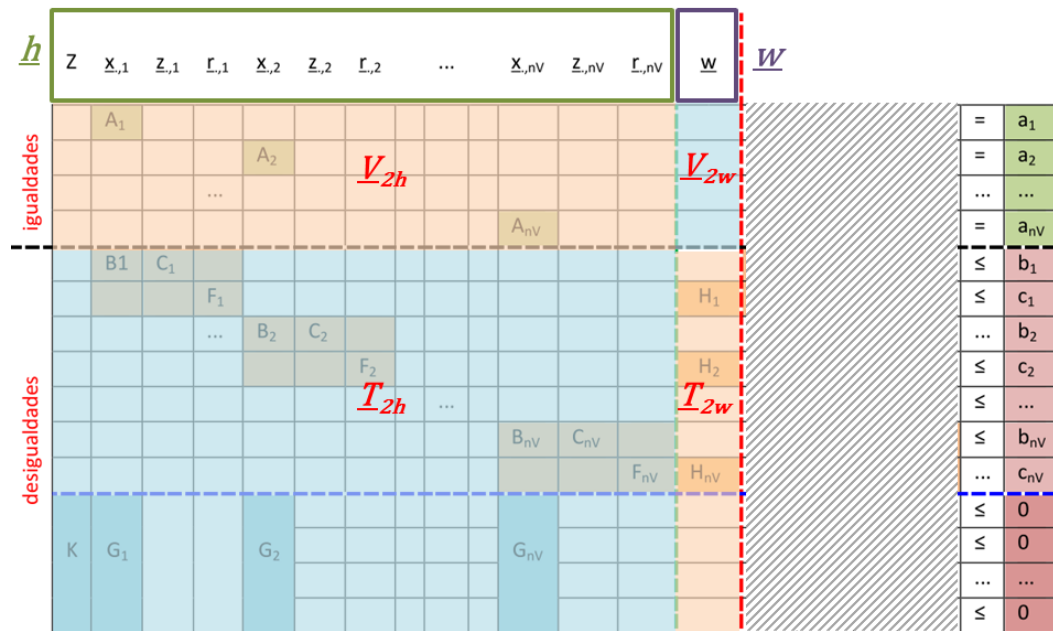


Figura 4.16 – Relação entre os elementos da matriz de restrições e as variáveis no **Passo 1** do **BENDERS-0**, viabilizando o **BENDERS-1** - Foco em eficiência

Fonte: Elaborada pela autora

	A ₁									
			A ₂							
		...								
								A _{nV}		

Figura 4.17 – Representação esquemática da matriz \underline{V}_{2h} - Foco em eficiência

Fonte: Elaborada pela autora

Reescrevendo, então, o problema de otimização representado pela eq. (4-21) em função de \underline{h} e de \underline{w} resulta em:

$$\underline{w}_L \leq \underline{w} \leq \underline{w}_H$$

H_1
H_2
H_{nV}

Figura 4.19 – Representação esquemática da matriz \underline{T}_{2h} - Foco em eficiência
Fonte: Elaborada pela autora

[illegible]

Figura 4.19 – Representação esquemática da matriz \underline{T}_{2h} - Foco em eficiência
Fonte: Elaborada pela autora

Como a componente \underline{V}_{2w} é nula, conforme evidenciado na Figura 4.16, resulta em:

$$\begin{aligned}
\min_{\underline{h}, \underline{w}} \quad & z = \underline{g}_{2h}^T \cdot \underline{h} + \underline{g}_{2w}^T \cdot \underline{w} \\
\text{sujeito a:} \quad & \\
& \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\
& \underline{T}_{2h} \cdot \underline{h} + \underline{T}_{2w} \cdot \underline{w} \leq \underline{s}_0^{(v)} \\
& \underline{h}_L \leq \underline{h} \leq \underline{h}_H \\
& \underline{w}_L \leq \underline{w} \leq \underline{w}_H
\end{aligned} \tag{4-23}$$

É possível observar que o problema descrito pela eq. (4-23) pode ser resolvido por meio da aplicação de um novo algoritmo de Benders (aqui chamado de **BENDERS-1**) em que uma nova “variável complicadora” pode ser separada, sendo \underline{w} a candidata ideal para assumir este papel. Fixando-se, então, a variável \underline{w} como $\underline{w}^{(x)}$, o problema de otimização correspondente para o **Passo 1** do **BENDERS-1** resulta em:

$$\begin{aligned}
\min_{\underline{h}} \quad & z = \underline{g}_{2h}^T \cdot \underline{h} \\
\text{sujeito a:} \quad & \\
& \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\
& \underline{T}_{2h} \cdot \underline{h} \leq \underline{s}_0^{(v)} - \underline{T}_{2w} \cdot \underline{w}^{(x)} \\
& \underline{h}_L \leq \underline{h} \leq \underline{h}_H
\end{aligned} \tag{4-24}$$

Substituindo o termo $\left(\underline{s}_0^{(v)} - \underline{T}_{2w} \cdot \underline{w}^{(x)}\right)$ por $\underline{s}_0^{(v, x)}$, tem-se:

$$\begin{aligned}
\min_{\underline{h}} \quad & z = \underline{g}_{2h}^T \cdot \underline{h} \\
\text{sujeito a:} \quad & \\
& \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\
& \underline{T}_{2h} \cdot \underline{h} \leq \underline{s}_0^{(v, x)} \\
& \underline{h}_L \leq \underline{h} \leq \underline{h}_H
\end{aligned} \tag{4-25}$$

Analisando-se a matriz de restrições associada ao problema descrito pela eq. (4-25), observa-se claramente, como representado na Figura 4.20, que o mesmo envolve “restrições complicadoras”.

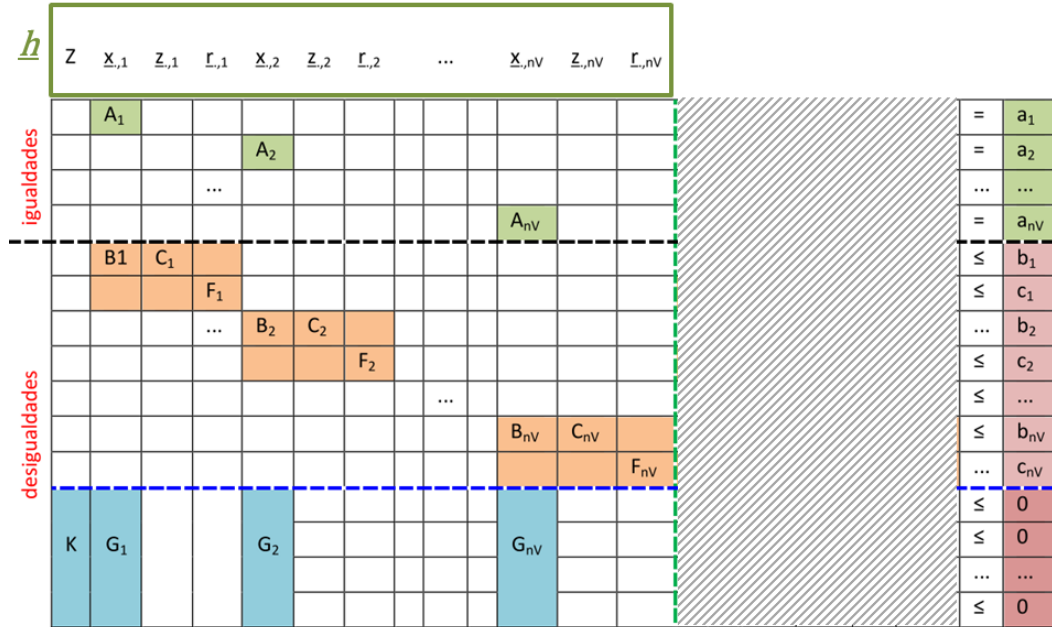


Figura 4.20 – Relação entre os elementos da matriz de restrições e as variáveis no **Passo 1** do **BENDERS-1** - Foco em eficiência

Fonte: Elaborada pela autora

Poder-se-ia resolver o problema por Dantzig-Wolfe, mas, por conveniência e com base no exposto no Capítulo 3 e no Apêndice B no tocante à relação de dualidade existente entre variáveis e restrições complicadoras, opta-se pela resolução deste problema não em sua forma primal, mas em sua forma dual, na qual as restrições são transformadas em variáveis complicadoras.

Para a obtenção do dual do problema descrito pela eq. (4-25) ele é reescrito na forma padrão, empregando as variáveis de folga $\underline{\rho}$, $\underline{\phi}$ e $\underline{\gamma}$, o que resulta em:

$$\min_{\underline{h}, \underline{\rho}, \underline{\phi}, \underline{\gamma}} \quad z = \underline{g}_{2h}^T \cdot \underline{h} + \underline{0}^T \cdot \underline{\rho} + \underline{0} \cdot \underline{\phi} + \underline{0}^T \cdot \underline{\gamma}$$

sujeito a:

$$\begin{aligned}
& \begin{bmatrix} \underline{V}_{2h} & 0 & 0 & 0 \\ \underline{T}_{2h} & I & 0 & 0 \\ \underline{c}_h^T & 0 & 1 & 0 \\ E_h & 0 & 0 & -I \end{bmatrix} \cdot \begin{bmatrix} \underline{h} \\ \underline{\rho} \\ \underline{\phi} \\ \underline{\gamma} \end{bmatrix} = \begin{bmatrix} \underline{r}_0 \\ \underline{s}_0^{(v,\chi)} \\ 1 \\ \underline{1} \end{bmatrix} \\
& \begin{bmatrix} \underline{h} \\ \underline{\rho} \\ \underline{\phi} \\ \underline{\gamma} \end{bmatrix} \geq \underline{0}
\end{aligned} \tag{4-26}$$

Onde:

$$\underline{c}_h^T = \begin{bmatrix} 1 & \underline{0}_{[1,nV \cdot (nE+2 \cdot nV)]} \end{bmatrix}$$

$$E_h = \left[\begin{array}{c|c|c|c|c|c|c} 0_{[(nV)^2,1]} & 0_{[(nV)^2,nE]} & 0_{[(nV)^2,nV]} & U_1 & \dots & 0_{[(nV)^2,nE]} & 0_{[(nV)^2,nV]} & U_{nV} \end{array} \right]$$

$$U_k = \begin{bmatrix} 0_{nV} \\ \vdots \\ 0_{nV} \\ I_{nV} \\ 0_{nV} \\ \vdots \\ 0_{nV} \end{bmatrix} \rightarrow \text{posição } k, \ k = 1, \dots, nV$$

O dual, ajustado para um problema de minimização, então, é dado por:

$$\min_{\underline{\eta}, \underline{\theta}, \xi, \underline{\varphi}} \quad z = -[\underline{r}_0]^T \cdot \underline{\eta} + \left[\underline{s}_1^{(v, \chi)} \right]^T \cdot \underline{\theta}_1 + \left[\underline{s}_2^{(v, \chi)} \right]^T \cdot \underline{\theta}_2 + \xi - \underline{1}^T \cdot \underline{\varphi}$$

sujeito a:

$$\underline{V}_{2h}^T \cdot \underline{\eta} - \underline{T}_{2h_1}^T \cdot \underline{\theta}_1 - \underline{T}_{2h_2}^T \cdot \underline{\theta}_2 - \underline{c}_h \cdot \xi + \underline{E}_h^T \cdot \underline{\varphi} \leq \underline{g}_{2h} \quad (4-27)$$

$\underline{\eta}$ irrestrito

$$\underline{\theta}_1, \underline{\theta}_2, \xi, \underline{\varphi} \geq 0$$

Reescrevendo o problema dual de forma simplificada, empregando as variáveis auxiliares $\hat{\underline{p}}, \hat{\underline{q}}, \hat{\underline{a}}, \hat{\underline{b}}, \hat{\underline{A}}$ e $\hat{\underline{B}}$ resulta em:

$$\min_{\hat{\underline{p}}, \hat{\underline{q}}} \quad z = [\hat{\underline{a}}]^T \cdot \hat{\underline{p}} + [\hat{\underline{b}}]^T \cdot \hat{\underline{q}}$$

sujeito a:

$$\hat{\underline{A}} \cdot \hat{\underline{p}} + \hat{\underline{B}} \cdot \hat{\underline{q}} \leq \underline{g}_{2h} \quad (4-28)$$

$$\hat{\underline{p}} \geq \hat{\underline{p}}_L$$

$$\hat{\underline{q}} \geq 0$$

O problema representado pela eq. (4-28) é claramente decomponível, composto por variáveis complicadoras, permitindo, como desejado, a aplicação, então, de um novo algoritmo de Benders (aqui chamado de **BENDERS-2**). A variável complicadora neste caso é $\hat{\underline{q}}$ e fixando-a como $\hat{\underline{q}}^{(\epsilon)}$, o problema de otimização correspondente para o **Passo 1** do **BENDERS-2** resulta em:

$$\min_{\hat{\underline{p}}} \quad z = [\hat{\underline{a}}]^T \cdot \hat{\underline{p}}$$

sujeito a:

$$\hat{\underline{A}} \cdot \hat{\underline{p}} \leq \underline{g}_{2h} - \hat{\underline{B}} \cdot \hat{\underline{q}}^{(\epsilon)} \quad (4-29)$$

$$\hat{\underline{p}} \geq \hat{\underline{p}}_L$$

Felizmente, o problema descrito pela eq. (4-29) é composto por variáveis separáveis, uma vez que a estrutura das matrizes correspondentes, depois de algum rearranjo de linhas e colunas, é claramente bloco-diagonal, como

ilustrado na Figura 4.21.

A_1^T	B_1^T	0								
0	C_1^T	0								
0	0	F_1^T								
			A_2^T	B_2^T	0					
			0	C_2^T	0					
			0	0	F_2^T					
						...				
							...			
								A_{nV}^T	B_{nV}^T	0
								0	C_{nV}^T	0
								0	0	F_{nV}^T

Figura 4.21 – Matriz de restrições no **Passo 1** do **BENDERS-2** - Foco em eficiência
Fonte: Elaborada pela autora

O **Passo 1** do **BENDERS-2** dá origem ao que se denomina como **OPTIM-0**, onde nV problemas de otimização independentes de dimensões extremamente pequenas devem ser resolvidos.

O encadeamento de todos estes algoritmos de Benders está detalhado no Apêndice F e pode ser visualizado na Figura 4.22, onde o **Passo 1** do algoritmo de **BENDERS- i** dá origem ao algoritmo de **BENDERS- $i + 1$** .

A formulação original da Solução do Subproblema que compõe o **Passo 1** pode ser observada no Apêndice D, onde explicita-se que uma variável dual $\underline{\lambda}$ precisa ser calculada de modo a ser utilizada no **Passo 3** do algoritmo.

Esta limitação, contudo, pode ser facilmente contornada levando-se em consideração que a solução ótima para ambos os problemas é a mesma e conhecida e, então, as condições de Karush-Kuhn-Tucker, descritas no Apêndice C, podem ser usadas para obter-se a variável dual.

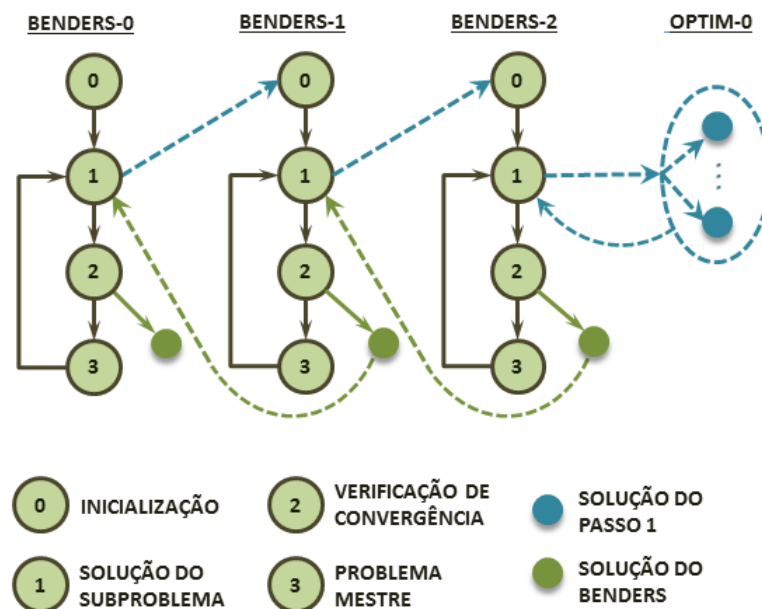


Figura 4.22 – Aplicação encadeada do método de decomposição de Benders

Fonte: Elaborada pela autora

4.5

Decomposição Aplicada ao Problema de Otimização com Foco em Economia de Energia

Neste contexto ressalta-se que, conforme exposto na seção 4.3, a variável Z é eliminada. Rearrumbando linhas e colunas na matriz de restrições (Figura 4.10), obtém-se a estrutura apresentada na Figura 4.23, a partir da qual é possível identificar de maneira mais clara a separação de variáveis necessária ao processo de decomposição.

Em função da estrutura da matriz de restrições representada na Figura 4.23 é fácil perceber que a solução do problema em questão pode ser obtida empregando-se exatamente os mesmos conceitos detalhados para a resolução do problema de otimização com foco em eficiência, conforme demonstrado a seguir.

Figura 4.23 – Matriz de restrições para o problema com foco em economia de energia, após rearranjos

Fonte: Elaborada pela autora

Como também pode ser observado na Figura 4.24, com relação aos termos \underline{r}_0 e \underline{s}_0 , presentes nas eqs. (4-16) e (4-17), tem-se:

$$\underline{r}_0 = \begin{bmatrix} \underline{a}_1^T & \underline{a}_2^T & \dots & \underline{a}_{n_V}^T \end{bmatrix}^T$$

$$\underline{s}_0 = \begin{bmatrix} b_1^T & \dots & b_{n_V}^T & \underline{c}_1^T & \dots & \underline{c}_{n_V}^T & 0^T & \dots & 0^T \end{bmatrix}^T$$

y_2	Variáveis Reais											Variáveis Binárias						y_1		
	$x_{.,1}$	$z_{.,1}$	$f_{.,1}$	$x_{.,2}$	$z_{.,2}$	$f_{.,2}$...	$x_{.,nV}$	$z_{.,nV}$	$f_{.,nV}$	w	$u_{.,1}$	$u_{.,2}$...	$u_{.,nV}$	α	β			
igualdades	A_1																	=	a_1	
				A_2														=	a_2	
			...															=	...	
									A_{nV}									=	a_{nV}	
desigualdades	B_1	C_1											D_1					\leq	b_1	
			F_1									H_1	E_1					\leq	c_1	
			...	B_2	C_2									D_2				\leq	b_2	
						F_2						H_2	E_2					\leq	c_2	
							...											\leq	...	
								B_{nV}	C_{nV}						D_{nV}			\leq	b_{nV}	
										F_{nV}	H_{nV}				E_{nV}			\leq	c_{nV}	
	G_1																	\leq	0	
				G_2													K		\leq	0
																		
																M	N	\leq	0	
															R	S		\leq	0	

Figura 4.24 – Relação entre os elementos da matriz de restrições e as variáveis do problema de otimização simplificado, com foco em economia de energia

Fonte: Elaborada pela autora

Analogamente ao problema da seção anterior, a componente \underline{V}_1 é nula e a função objetivo depende somente das variáveis reais \underline{y}_2 (ou seja, a parcela $\underline{g}_1^T \cdot \underline{y}_1$ da função objetivo é nula). Assim, o problema original pode ser reescrito de forma bem simplificada, conforme representado nas eqs. de (4-15) a (4-19).

A_1																			
				A_2															
			...																
										A_{nV}									

Figura 4.25 – Representação esquemática da matriz \underline{V}_2 - Foco em economia de energia

Fonte: Elaborada pela autora

Aplicando-se, então, o método de decomposição de Benders, conforme algoritmo descrito no Apêndice D, ao problema de otimização de interesse, representado pelas eqs. de (4-15) a (4-19), transcritas na eq.(4-30) (onde $\underline{y}_1 \in \mathbb{B}^{nY_1}$ e $\underline{y}_2 \in \mathbb{R}^{nY_2}$) para facilitar o acompanhamento da descrição do método, resulta que o **Passo 3** está relacionado a um problema de otimização contendo apenas variáveis binárias enquanto o **Passo 1** envolve apenas variáveis reais, identicamente ao ocorrido na seção anterior. Assim, esta primeira aplicação do

O problema de otimização correspondente ao **Passo 1** do **BENDERS-0** é dado por:

sujeito a:

$$\underline{T}_1 \cdot \underline{y}_1 + \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 \quad (4-30)$$

$$\underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H}$$

D ₁							
E ₁							
	D ₂						
	E ₂						
		...					
			D _{nV}				
			E _{nV}				
					K		
					M	N	
R					S		

Figura 4.26 – Representação esquemática da matriz \underline{T}_1 - Foco em economia de energia
Fonte: Elaborada pela autora

O problema de otimização correspondente ao **Passo 1** do **BENDERS-0** é dado por:

sujeito a:

$$\underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0^{(v)}$$

$$\underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H}$$

Novamente a resolução deste problema não é trivial em virtude das dimensões das variáveis y_2 envolvidas. Mas, analisando-se a matriz de restrições

associada, observa-se que o problema pode ser reescrito de forma decomponível, conforme representado graficamente na Figura 4.28, considerando-se

$$\underline{y}_2 = \begin{bmatrix} \underline{h} \\ \underline{w} \end{bmatrix}$$

B ₁	C ₁											
		F ₁										H ₁
		...	B ₂	C ₂								
					F ₂							H ₂
							...					
								B _{nV}	C _{nV}			
										F _{nV}	H _{nV}	
G ₁			G ₂					G _{nV}				

Figura 4.27 – Representação esquemática da matriz \underline{T}_2 - Foco em economia de energia
Fonte: Elaborada pela autora

Reescrevendo, então, o problema de otimização representado pela eq. (4-31) em função de \underline{h} e de \underline{w} e sabendo que a componente \underline{V}_{2w} é nula, conforme evidenciado na Figura 4.28, resulta em:

$$\begin{aligned} \min_{\underline{h}, \underline{w}} \quad & z = \underline{g}_{2h}^T \cdot \underline{h} + \underline{g}_{2w}^T \cdot \underline{w} \\ \text{sujeito a:} \quad & \\ & \underline{V}_{2h} \cdot \underline{h} = r_0 \\ & \underline{T}_{2h} \cdot \underline{h} + \underline{T}_{2w} \cdot \underline{w} \leq \underline{s}_0^{(v)} \\ & \underline{h}_L \leq \underline{h} \leq \underline{h}_H \\ & \underline{w}_L \leq \underline{w} \leq \underline{w}_H \end{aligned} \tag{4-32}$$

É possível observar que o problema descrito pela eq. (4-32) pode ser resolvido por meio da aplicação de um novo algoritmo de Benders (aqui chamado de **BENDERS-1**) em que a variável complicadora é \underline{w} . Assim, o problema de otimização correspondente para o **Passo 1** do **BENDERS-1**

resulta em:

$$\begin{aligned}
 \min_{\underline{h}} \quad & z = \underline{g}_{2h}^T \cdot \underline{h} \\
 \text{sujeito a:} \quad & \\
 & \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\
 & \underline{T}_{2h} \cdot \underline{h} \leq \underline{s}_0^{(v, \chi)} \\
 & \underline{h}_L \leq \underline{h} \leq \underline{h}_H
 \end{aligned} \tag{4-33}$$

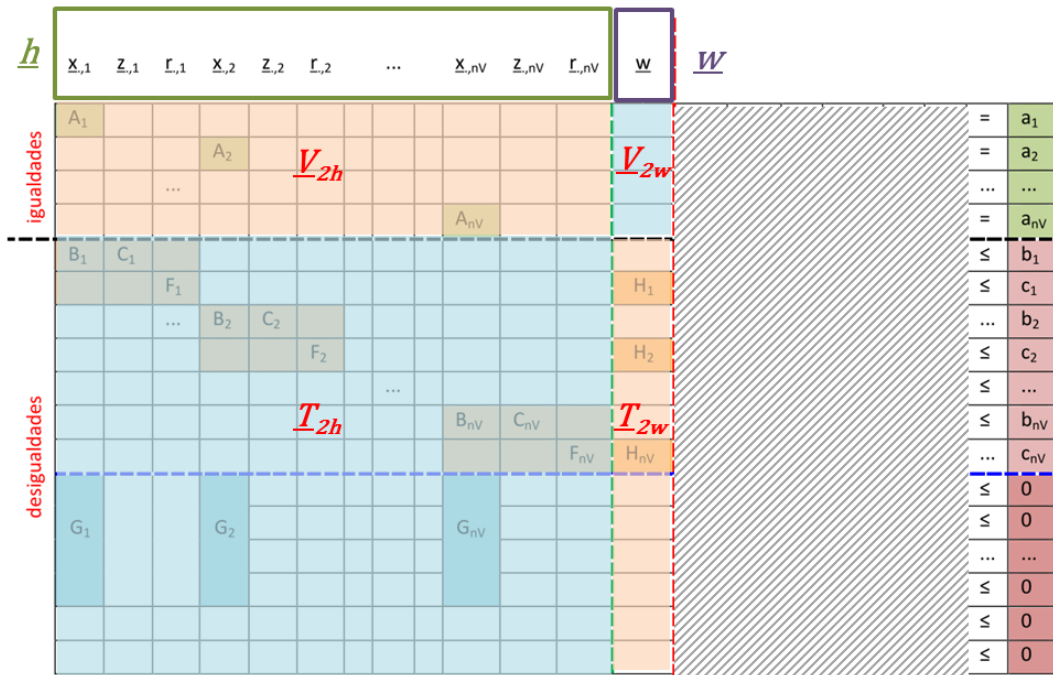


Figura 4.28 – Relação entre os elementos da matriz de restrições e as variáveis no **Passo 1** do **BENDERS-0**, viabilizando o **BENDERS-1** - Foco em economia de energia
Fonte: Elaborada pela autora

Analisando-se a matriz de restrições associada ao problema descrito pela eq. (4-33), observa-se, conforme representado na Figura 4.29, que o mesmo envolve “restrições complicadoras”. No entanto, optando-se por resolver o dual deste problema, as restrições são transformadas em “variáveis complicadoras”.

O problema dual escrito de forma simplificada, empregando as variáveis auxiliares $\underline{\hat{p}}$, $\underline{\hat{q}}$, $\underline{\hat{a}}$, $\underline{\hat{b}}$, $\underline{\hat{A}}$ e $\underline{\hat{B}}$ resulta em:

$$\min_{\hat{\underline{p}}, \hat{\underline{q}}} \quad z = [\hat{\underline{a}}]^T \cdot \hat{\underline{p}} + [\hat{\underline{b}}]^T \cdot \hat{\underline{q}}$$

sujeito a:

$$\hat{\underline{A}} \cdot \hat{\underline{p}} + \hat{\underline{B}} \cdot \hat{\underline{q}} \leq \underline{g}_{2h} \quad (4-34)$$

$$\hat{\underline{p}} \geq \hat{\underline{p}}_L$$

$$\hat{\underline{q}} \geq \underline{0}$$

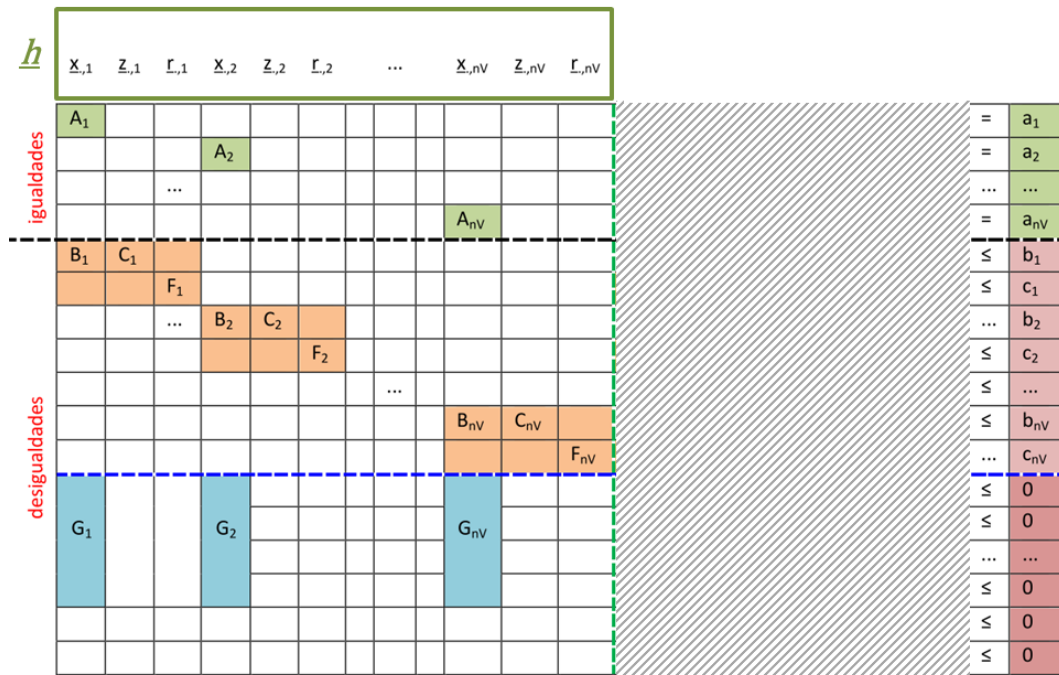


Figura 4.29 – Relação entre os elementos da matriz de restrições e as variáveis no **Passo 1** do **BENDERS-1** - Foco em economia de energia

Fonte: Elaborada pela autora

O problema representado pela eq. (4-34) é claramente decomponível, composto por variáveis complicadoras, permitindo, como desejado, a aplicação, então, de um novo algoritmo de Benders (aqui chamado de **BENDERS-2**). A variável complicadora neste caso é $\hat{\underline{q}}$ e o problema de otimização correspondente para o **Passo 1** do **BENDERS-2** resulta em:

$$\min_{\underline{\hat{p}}} \quad z = [\underline{\hat{a}}]^T \cdot \underline{\hat{p}}$$

sujeito a:

$$\begin{aligned} \underline{\hat{A}} \cdot \underline{\hat{p}} &\leq \underline{g}_{2h} - \underline{\hat{B}} \cdot \underline{\hat{q}}^{(\epsilon)} \\ \underline{\hat{p}} &\geq \underline{\hat{p}}_L \end{aligned} \tag{4-35}$$

Finalmente, o problema descrito pela eq. (4-35) é composto por variáveis separáveis, uma vez que a estrutura das matrizes correspondentes, depois de algum rearranjo de linhas e colunas, é claramente bloco-diagonal. O **Passo 1** do **BENDERS-2**, por conseguinte, dá origem ao que se denomina como **OPTIM-0**, onde nV problemas de otimização independentes de dimensões extremamente pequenas devem ser resolvidos.

Salienta-se que as mesmas observações feitas na seção anterior com relação à obtenção da variável dual $\underline{\lambda}$ aqui se aplicam, haja vista que o método empregado de encadeamento dos algoritmos de Benders é exatamente o mesmo.

4.6

Resultados Computacionais

Com o intuito de avaliar a consistência da modelagem matemática adotada na presente tese e visando a criar um referencial para posterior comparação foram efetuados testes computacionais empregando o programa de otimização Xpress-Optimizer, presente no pacote FICO® Xpress Optimization Suite [8], na resolução direta do problema MILP, com foco em economia de energia, em um computador com processador Intel Core i7 com CPU de 2,20GHz e equipado com 8GB de memória RAM.

Nos testes foram usados dois conjuntos de topologias de redes. O primeiro, elaborado pela autora, contempla as redes que foram denominadas de Nano, Spider e Penta, cujas representações gráficas são apresentadas respectivamente nas Figuras 4.30, 4.31 e 4.32 e cujas dimensões diminutas, descritas no Quadro 4.5, foram consideradas ideais para a visualização da eficácia das soluções obtidas. O segundo, composto por uma seleção extraída do banco de

dados SNDlib [73], contempla as redes Abilene, Atlanta, India35, Newyork, Pdh, Polska e Sun, cujas composições são descritas também no Quadro 4.5, no qual a quantidade de circuitos considera que estes são bidirecionais.

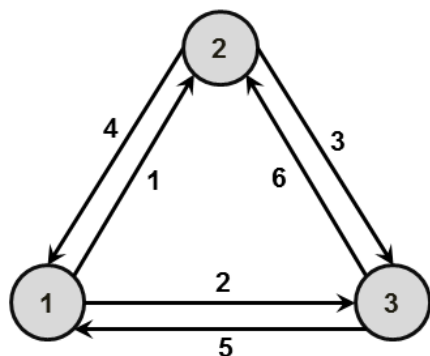


Figura 4.30 – Rede Nano
Fonte: Elaborada pela autora

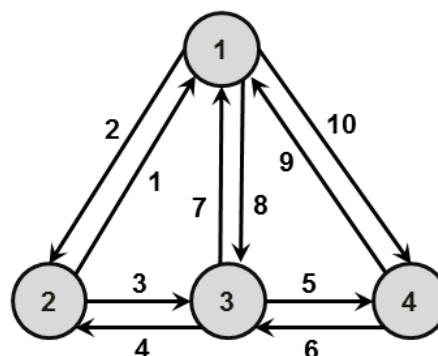


Figura 4.31 – Rede Spider
Fonte: Elaborada pela autora

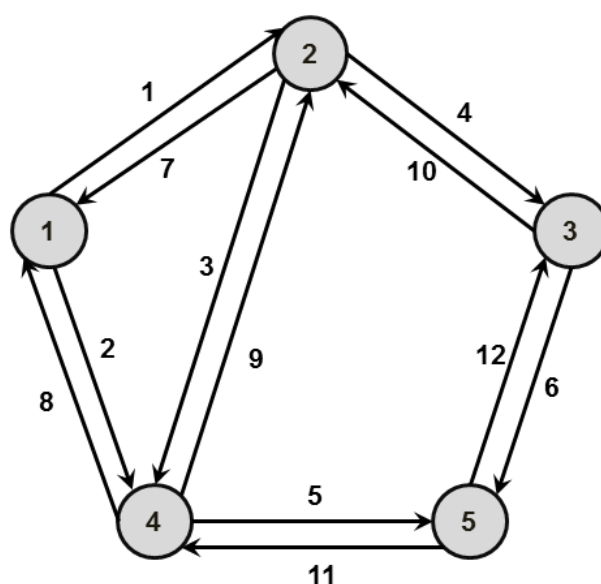


Figura 4.32 – Rede Penta
Fonte: Elaborada pela autora

Cabe salientar que para o cálculo do consumo de potência total em cada uma das redes escolhidas para os testes considerou-se que todas são compostas pelo mesmo modelo de roteador cujo consumo de energia é igual a 500 W e que os circuitos estão conectados ao mesmo tipo de interface, possuindo tais interfaces um consumo de energia equivalente a 10 W. Definiu-se ainda como Potência Plena o consumo de energia total da rede na situação em que todos os nós e circuitos estão ativos, ou seja, quando não há economia de energia.

Quadro 4.5 – Composição das redes empregadas nos testes

Rede	Nós	Circuitos	Potência Plena
Nano	3	6	1560 W
Spider	4	10	2100 W
Penta	5	12	2620 W
Abilene	12	30	6300 W
Atlanta	15	44	7940 W
India35	35	160	19100 W
Newyork	16	98	8980 W
Pdh	11	68	6180 W
Polska	12	36	6360 W
Sun	27	204	15540 W

Fonte: Elaborado pela autora

Ressalta-se que para os testes foram considerados dois cenários, o primeiro, denominado **Cenário 1**, em que todos os circuitos encontravam-se extremamente ociosos apresentando nível de utilização máximo inferior a 20% e no qual, portanto, a economia de energia deveria ser máxima. No segundo cenário, denominado **Cenário 2**, foi considerada a situação em que pelo menos um dos circuitos apresentava ocupação superior a 50%, o que deveria resultar, conseqüentemente, numa solução em que a economia de energia não seria máxima, haja vista que uma quantidade maior de circuitos deveria ser mantida ativa de tal forma a acomodar todas as demandas de tráfego.

Os resultados obtidos por meio do Xpress-Optimizer para os **Cenários 1** e **2** são apresentados respectivamente no Quadro 4.6 e no Quadro 4.7, ratificando as expectativas e validando a modelagem matemática descrita na presente tese. Além disso, a análise do conteúdo destes quadros permite constatar que, como era esperado, o tempo de convergência no **Cenário 1** e inferior ao do **Cenário 2**, no qual, inclusive, houve situações em que mesmo após mais de um dia de processamento não foi possível obter a solução ótima. Cabe esclarecer também que a porcentagem de energia consumida apresentada nos

referidos quadros foi calculada em relação à situação de Potência Plena.

Quadro 4.6 – Soluções para o **Cenário 1**

Rede	Potência	Tempo	Energia Consumida
Nano	1530 W	< 0,1s	98,08%
Spider	2040 W	< 0,1s	97,14%
Penta	2550 W	0,1s	97,33%
Abilene	6130 W	0,5s	97,30%
Atlanta	7660 W	17,4s	96,47%
India35	17850 W	5.908,3s	93,46%
Newyork	8160 W	303,8s	90,87%
Pdh	5610 W	61s	90,78%
Polska	6120 W	11,1s	96,23%
Sun	12240 W	806,1s	78,76%

Fonte: Elaborado pela autora

Quadro 4.7 – Soluções para o **Cenário 2**

Rede	Potência	Tempo	Energia Consumida
Nano	1540 W	< 0,1s	98,72%
Spider	2090 W	0,1s	99,52%
Penta	2570 W	0,2s	98,09%
Abilene	6220 W	1.313,5s	98,73%
Atlanta	Sem solução	> 92.402,6s	—
India35	Sem solução	> 93.503,6s	—
Newyork	Sem solução	> 107.748,2s	—
Pdh	Sem solução	> 109.553,1s	—
Polska	Sem solução	> 122.478,3s	—
Sun	Sem solução	> 133.551,3s	—

Fonte: Elaborado pela autora

Destaca-se que, com o intuito de exemplificar graficamente soluções ótimas obtidas, facilitando a visualização da eficácia na resolução do problema

de otimização de interesse, apresentam-se na Figura 4.33 as topologias resultantes no **Cenário 1** e no **Cenário 2** para a rede Spider, as quais evidenciam que em situação de grande ociosidade (Figura 4.33(a)) é possível economizar o máximo de energia, mantendo ativa apenas a quantidade mínima de circuitos necessária para garantir a conectividade de todos os nós da rede (neste caso, os circuitos 1, 4, 6 e 10). Na situação em que os circuitos não estão tão ociosos a porcentagem de economia possível dependerá do seu grau de utilização (por meio da Figura 4.33(b) pode-se observar que neste caso a economia de energia foi mínima, mas ainda foi possível a partir do desligamento do circuito 8).

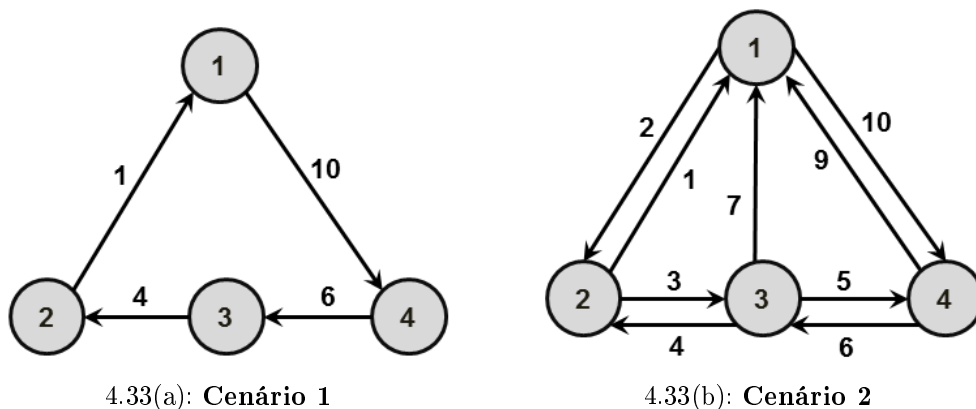


Figura 4.33 – Topologias resultantes para a solução ótima obtida para a rede Spider nos dois diferentes cenários testados

Fonte: Elaborada pela autora

Uma vez comprovada a eficácia da modelagem matemática proposta na presente tese, conforme descrito anteriormente, foram efetuados testes computacionais empregando a rede Penta para comprovar a eficácia do método proposto de aplicação encadeada de algoritmos de decomposição de Benders. Para tanto o problema de otimização com foco em desempenho para a referida rede foi resolvido de forma direta e empregando o método aqui apresentado. No primeiro caso o problema foi resolvido em 0,15s gerando como resultado $Z = 21\%$. No segundo caso o problema convergiu em 0,21s gerando como resultado $Z = 23\%$. E por meio de tais resultados foi possível atestar a viabilidade da metodologia proposta.

5

Conclusão

No presente trabalho é apresentada uma abordagem de Engenharia de Tráfego em **Redes Verdes**. Salienta-se, contudo, que a modelagem matemática empregada permitiu a análise de dois problemas de otimização, a saber, um com foco em Eficiência ou Desempenho e outro com foco em Economia de Energia. Destaca-se ainda que na formulação do modelo foram consideradas restrições bastante realistas e adequadas ao cenário de Engenharia de Tráfego, resultando num viés mais pragmático e menos acadêmico. A imposição de uso de protocolo OSPF com suporte a balanceamento de carga por meio de técnicas ECMP, por exemplo, é padrão na grande maioria das redes de telecomunicações existentes no Mercado.

Os dois problemas de otimização supracitados apesar de serem aplicados a cenários distintos possuem em comum vantagens relacionadas ao aspecto financeiro. O primeiro problema, cujo objetivo é definir a melhor distribuição dos fluxos de tráfego pela rede de tal forma a reduzir o nível de utilização máxima dos circuitos, viabiliza a redução de custos a partir da postergação da ampliação de circuitos. Já o segundo, cujo objetivo é definir a quantidade mínima de roteadores e circuitos que devem ser mantidos ativos de tal modo a assegurar o perfeito funcionamento das aplicações suportadas pela rede, assegura a economia financeira uma vez que o desligamento de elementos reduz o consumo energético e consequentemente os gastos a ele associado.

Até onde a autora desta tese teve oportunidade de investigar na literatura científica corrente, estes dois problemas de otimização, quando aplicados a redes existentes em empresas ou companhias de telecomunicações, são indiscutivelmente de grande porte, ou seja, envolvem um número elevado de variáveis de decisão e restrições operacionais. Tentativas de solução destes

problemas por meio estrito de aplicação de Programação Linear ao problema completo revelaram ser este de elevadíssimo custo computacional, sendo que em alguns casos, o seu tempo de execução atinge valores proibitivos sob o ponto de vista operacional. Não é por outra razão que a maioria dos pesquisadores nesta área tem proposto métodos de natureza heurística para contornar esta severa limitação.

A abordagem adotada na presente tese é, novamente até onde a autora teve oportunidade de investigar na literatura corrente, original. Trata-se de uma metodologia onde os conceitos clássicos de decomposição são aplicados numa forma original e particular: os dois problemas de interesse são resolvidos por uma sequência de três etapas de decomposição, onde ao final delas vários problemas de pequeno porte são resolvidos. Estes problemas de otimização linear envolvem variáveis de decisão em quantidade da ordem do número de nós da rede de comunicações em análise. A título de ilustração, uma rede de tamanho considerado “moderado” composta por 40 nós e 50 enlaces produz um problema completo de programação linear com cerca de 8000 variáveis de decisão e 13000 restrições. A aplicação do método aqui discutido a este problema implica em que os diversos problemas de programação linear a serem resolvidos são de dimensão da ordem de 50, valor extremamente factível mesmo para programas de otimização de porte moderado.

Finalmente, enfatiza-se que, uma vez que o método proposto baseia-se na otimização de custos dos enlaces utilizados no cálculo do caminho mais curto, são mantidas as habilidades inerentes ao protocolo de roteamento de se adaptar rapidamente às variações de tráfego e de reagir às mudanças na rede, de forma análoga ao que ocorre com abordagens tradicionais baseadas em métricas fixas de roteamento.

A viabilidade do método proposto foi comprovada a partir da sua aplicação a uma rede pequena composta por 5 nós e 12 circuitos. No entanto, considera-se extremamente importante que futuros trabalhos apliquem-no a redes existentes, por exemplo, no banco de dados SNDlib [73] de tal a modo a permitir a análise comparativa com outros métodos propostos na literatura.

Outrossim, o método apresentado considera que a matriz de tráfego é

conhecida sendo aplicável a períodos específicos do dia. Recomenda-se que em trabalhos futuros o método seja expandido para ser aplicado com base no comportamento em tempo real da rede e não apenas tendo como base uma matriz de tráfego pré-estabelecida.

Ainda como proposta para trabalhos futuros considera-se que seria adequada a análise do comportamento mediante a flexibilização de algumas restrições como o balanceamento de carga por ECMP.

Por fim, sugere-se o tratamento multi-objetivo do problema de otimização considerando como objetivos a redução do consumo de energia e a minimização do nível de utilização máxima dos circuitos, assegurando, contudo, o perfeito desempenho das aplicações suportadas pela rede.

- [1] **The Climate Group: Website.** Disponível em: <<http://www.theclimategroup.org>>. Acesso em: nov. 2012.
- [2] **Quanto custa a energia elétrica para a indústria no Brasil?: Website.** Disponível em: <<http://www.quantocustaenergia.com.br>>. Acesso em: mar. 2015.
- [3] **Internet2 Network NOC: Website.** Disponível em: <<https://noc.net.internet2.edu/i2network/research-and-education-network.html>>. Acesso em: jun. 2015.
- [4] **DIP (Decomposition for Integer Programming): Website.** Disponível em: <<https://projects.coin-or.org/Dip>>. Acesso em: jan. 2015.
- [5] **Consortium GARR – La Rete Italiana dell’Università e della Ricerca: Website.** Disponível em: <<http://www.garr.it/a/rete/infrastruttura-di-rete/mappa-della-rete>>. Acesso em: jun. 2015.
- [6] **Juniper Networks: Website.** Disponível em: <<http://www.juniper.net/us/en/products-services/routing>>. Acesso em: jun. 2015.
- [7] **Cisco: Website.** Disponível em: <<http://www.cisco.com/c/en/us/products/routers/index.html>>. Acesso em: jun. 2015.
- [8] **FICO Xpress Optimization Suite.** Fair Isaac Corporation 2015. Disponível em: <<https://community.fico.com/download.jspa>>. Acesso em: abr. 2015.
- [9] SHAPLEY, L. S.. **A value for n-person games.** In: Kuhn, H. W.; Tucker, A., editors, CONTRIBUTIONS TO THE THEORY OF GAMES,

- volumen 2, p. 307–317. Annals of Mathematical Studies v. 28, Princeton University Press, 1953.
- [10] BENDERS, J. F.. **Partitioning methods for solving mixed variables programming problems**. Numerische Mathematik, 4:238–252, 1962.
 - [11] FIEDLER, M.. **Algebraic connectivity of graphs**. Czechoslovak Mathematical Journal, 23(2):298–305, 1973.
 - [12] BRADLEY, S. P.; HAX, A. C. ; MAGNANTI, T. L.. **Applied Mathematical Programming**. Addison-Wesley, 1977.
 - [13] ROTH, A. E.. **The Shapley value: essays in honor of Lloyd S. Shapley**. Cambridge University Press, 1988.
 - [14] BERTSIMAS, D.; TSITSIKLIS, J. N.. **Introduction to Linear Optimization**. Massachusetts Institute of Technology, 1997.
 - [15] CORNUEJOLS, G.; TRICK, M.. **Quantitative Methods for the Management Sciences**. Carnegie Mellon University, 1998. Disponível em: <<http://mat.gsia.cmu.edu/classes/QUANT/NOTES/chap4.pdf>>. Acesso em: ago. 2015.
 - [16] THALER, D.; HOPPS, C.. **Multipath issues in unicast and multicast next-hop selection**. In: INTERNET ENGINEERING TASK FORCE REQUEST FOR COMMENTS 2991 (RFC 2991), nov. 2000.
 - [17] HOPPS, C.. **Analysis of an Equal-Cost Multi-Path Algorithm**. In: INTERNET ENGINEERING TASK FORCE REQUEST FOR COMMENTS 2992 (RFC 2992), nov. 2000.
 - [18] MEDINA, A.; MATTA, I. ; BYERS, J.. **On the origins of power laws in Internet topologies**. Technical report, 39th Annual Allerton Conference on Communication, Control, and Computing, 2001.
 - [19] GODSIL, C.; ROYLE, G. F.. **Algebraic Graph Theory**. Graduate Texts in Mathematics. Springer, 2001.

- [20] CHONG, E. K. P.; ZAK, S. H.. **An Introduction to Optimization**. Wiley, 2001.
- [21] CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L. ; STEIN, C.. **Introduction to Algorithms**. MIT Press, second edition, 2001. Disponível em: <<http://www.mif.vu.lt/~valdas/ALGORITMAI/LITERATURA/Cormen/Cormen.pdf>>. Acesso em: jan. 2015.
- [22] LIU, G.; RAMAKRISHNAN, K. G.. **A*Prune: an algorithm for finding K shortest paths subject to multiple constraints**. In: PROCEEDINGS OF 20TH JOINT CONFERENCE IEEE COMPUTER & COMMUNICATIONS SOCIETIES (INFOCOM 2001), p. 743–749, 2001.
- [23] YUAN, X.; LIU, X.. **Heuristics algorithms for multi-constrained quality of service routing**. In: PROCEEDINGS OF IEEE INFOCOM, 2001.
- [24] PUC-Rio. **Programação Matemática - Princípios Básicos**, 2002. Disponível em: <http://www.maxwell.vrac.puc-rio.br/3332/3332_3.PDF>. Acesso em: jan. 2015.
- [25] WANG, B.; SU, X. ; CHEN, C. L. P.. **A new bandwidth guaranteed routing algorithm for MPLS traffic engineering**. In: PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, ICC 2002, volumen 2, p. 1001–1005, 2002.
- [26] GUPTA, M.; SINGH, S.. **Greening of the internet**. In: PROCEEDINGS OF THE 2003 CONFERENCE ON APPLICATIONS, TECHNOLOGIES, ARCHITECTURES, AND PROTOCOLS FOR COMPUTER COMMUNICATIONS, SER. SIGCOMM'03, ACM, p. 19–26, 2003.
- [27] ODLYZKO, A.. **Data networks are lightly utilized, and will stay that way**. Review of Network Economics, 2(3), 2003.

- [28] CHRISTENSEN, K. J.; GUNARATNE, C.; NORDMAN, B. ; GEORGE, A. D.. **The next frontier for communications networks: power management**. Computer Communications, 27:1758–1770, 2004.
- [29] GUPTA, M.; GROVER, S. ; SINGH, S.. **A feasibility study for power management in LAN switches**. In: PROCEEDINGS OF THE 12TH IEEE INTERNATIONAL CONFERENCE ON NETWORK PROTOCOLS, p. 361–371, 2004.
- [30] WALLACE, B.. **Constrained Optimization: Kuhn-Tucker conditions**. Royal Holloway, 2004. Disponível em: <<http://amber.feld.cvut.cz/bio/konopka/file/5.pdf>>. Acesso em: ago. 2015.
- [31] NANDA, S.; CKER CHIUEH, T.. **A survey on virtualization technologies**. Technical report, Department of Computer Science, SUNY at Stony Brook, 2005. Disponível em: <<http://www.ecsl.cs.sunysb.edu/tr/TR179.pdf>>. Acesso em: nov. 2013.
- [32] GUNARATNE, C.; CHRISTENSEN, K. ; NORDMAN, B.. **Managing energy consumption costs in desktop PCs and LAN switches with proxying, split TCP connections, and scaling of link speed**. International Journal of Network Management, 15:297–310, 2005.
- [33] NUCCI, A.; SRIDHARAN, A. ; TAFT, N.. **The problem of synthetically generating IP traffic matrices: initial recommendations**. Computer Communication Review, 35:19–32, 2005.
- [34] RALPHS, T. K.; GALATI, M. V.. **Decomposition in Integer Linear Programming (*in* Integer Programming: Theory and Practice)**, chapter 4, p. 57–110. CRC Press, 2005.
- [35] COSTA, A. M.. **A survey on Benders decomposition applied to fixed-charge network design problems**. Computers & Operations Research, 32:1429–1450, 2005.

- [36] GUNARATNE, C.; CHRISTENSEN, K. ; SUEN, S. W.. **Ethernet adaptive link rate (ALR): analysis of a buffer threshold policy**. In: PROCEEDINGS OF IEEE GLOBECOM, 2006.
- [37] CONEJO, A. J.; CASTILLO, E.; MÍNGUEZ, R. ; GARCÍA-BERTRAND, R.. **Decomposition Techniques in Mathematical Programming: Engineering and Science Applications**. Springer, 2006.
- [38] AVALONE, S.; KUIPERS, F.; VENTRE, G. ; MIEGHEM, P. V.. **Dynamic routing in QoS-aware traffic engineered networks**, p. 45–58. EUNICE 2005: Networks and Applications Towards a Ubiquitously Connected World, 2006.
- [39] BARROSO, L. A.; HÖLZLE, U.. **The case for energy-proportional computing**. IEEE Computer, 40(12):33–37, dec. 2007.
- [40] ALLMAN, M.; CHRISTENSEN, K.; NORDMAN, B. ; PAXSON, V.. **Enabling an energy-efficient future Internet through selectively connected end systems**. In: PROCEEDINGS OF THE SIXTH ACM SIGCOMM WORKSHOP ON HOT TOPICS IN NETWORKS (HOTNETS-VI), 2007.
- [41] KOHL, B.. **10GBase-T power budget summary**. IEEE 802.3az Task Force presentation, 2007.
- [42] GUPTA, M.; SINGH, S.. **Using low-power modes for energy conservation in Ethernet LANs**. In: PROCEEDINGS OF IEEE INFOCOM, p. 2451–2455, 2007.
- [43] The Climate Group. **SMART 2020: Enabling the low carbon economy in the information age**, 2008. Disponível em: <http://www.smart2020.org/_assets/files/02_smart2020Report.pdf>. Acesso em: nov. 2014.
- [44] MURUGESAN, S.. **Harnessing green IT: principles and practices**. IEEE Computer Society, p. 24–33, 2008.

- [45] PICKAVET, M.; VEREECKEN, W.; DEMEYER, S.; AUDENAERT, P.; VERMEULEN, B.; DEVELDER, C.; COLLE, D.; DHOEDT, B. ; DEMEESTER, P.. **Worldwide energy needs for ICT: the rise of power-aware networking**. In: 2ND INTERNATIONAL SYMPOSIUM ON ADVANCED NETWORKS AND TELECOMMUNICATION SYSTEMS, 2008 (ANTS' 08), p. 1–3, 2008.
- [46] MOSHARAF, N. M.; CHOWDHURY, K. ; BOUTABA, R.. **A survey of network virtualization**. Computer Networks, 2008.
- [47] ZHANG, B.; SABHANATARAJAN, K.; GORDON-ROSS, A. ; GEORGE, A.. **Real-time performance analysis of adaptive link rate**. In: 33RD IEEE CONFERENCE ON LOCAL COMPUTER NETWORKS, (LCN' 08), p. 282–288, 2008.
- [48] GUNARATNE, C.; CHRISTENSEN, K.; NORDMAN, B. ; SUEN, S.. **Reducing energy consumption of Ethernet with an adaptive link rate (ALR)**. IEEE Transactions on Computers, 57(4):448–461, 2008.
- [49] CHABAREK, J.; SOMMERS, J.; BARFORD, P.; ESTAN, C.; TSIANG, D. ; WRIGHT, S.. **Power awareness in network design and routing**. In: PROCEEDINGS OF THE 27TH IEEE COMMUNICATIONS SOCIETY CONFERENCE ON COMPUTER COMMUNICATIONS (INFOCOM '08), p. 457–465, 2008.
- [50] NORDMAN, B.. **EEE savings estimate**. IEEE 802.3az Task Force Information, mar. 2008.
- [51] TUCKER, R. S.. **Modelling Energy Consumption in IP Networks**. ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN), University of Melbourne, 2008. Disponível em: <http://www.ing.unitn.it/~fontana/GreenInternet/Old%20Papers/cgrs_energy_consumption_ip.pdf>. Acesso em: jun. 2015.

- [52] DE OLIVEIRA, L. K.. **Pesquisa Operacional II - Dualidade**. Fundação Santo André, 2008. Disponível em: <<http://www3.fsa.br/localuser/Producao/arquivos/aula4-Dualidade.pdf>>. Acesso em: jan. 2015.
- [53] CHRISTENSEN, K.. **Green Networks: Opportunities and Challenges**, 2009. Disponível em: <http://www.ieeeln.org/prior/LCN34/2009_Keynote_Christensen.pdf>. Acesso em: nov. 2012.
- [54] INTERNATIONAL ENERGY AGENCY. **Gadgets and Gigawatts – Policies for Energy Efficient Electronics**, 2009. Disponível em: <<http://www.iea.org/publications/freepublications/publication/gigawatts2009.pdf>>. Acesso em: dez. 2012.
- [55] CHIARAVIGLIO, L.; MELLIA, M. ; NERI, F.. **Reducing power consumption in backbone networks**. IEEE International Communications Conference (ICC 2009), p. 1–6, 2009. Disponível em: <<http://porto.polito.it/2285430>>. Acesso em: nov. 2012.
- [56] CHIARAVIGLIO, L.; MELLIA, M. ; NERI, F.. **Energy-aware backbone networks: a case-study**. IEEE International Conference on Communications Workshops (ICC Workshops 2009), p. 1–5, 2009.
- [57] REVIRIEGO, P.; HERNADEZ, J.-A.; LARRABEITI, D. ; MAESTRO, J. A.. **Performance evaluation of energy efficient Ethernet**. IEEE Communications Letters, 13(9):697–699, 2009.
- [58] RODRÍGUEZ-PÉREZ, M.; HERRERÍA-ALONSO, S.; FERNÁNDEZ-VEIGA, M. ; LÓPEZ-GARCÍA, C.. **Improved opportunistic sleeping algorithms for LAN switches**. In: PROCEEDINGS OF IEEE GLOBECOM, 2009.
- [59] VISHWANATH, A.; SIVARAMAN, V. ; OSTRY, D.. **How Poisson is TCP traffic at short time-scales in a small buffer core network?** In: IEEE 3RD INTERNATIONAL SYMPOSIUM ON ADVANCED NETWORKS AND TELECOMMUNICATION SYSTEMS (ANTS 2009), 2009.

- [60] MARTÍNEZ, J. M.. **Otimização Prática Usando o Lagrangiano Aumentado**. UNICAMP, 2009. Disponível em: <<http://www.ime.unicamp.br/~martinez/lagraum.pdf>>. Acesso em: ago. 2015.
- [61] CIANFRANI, A.; ERAMO, V.; LISTANTI, M.; MARAZZA, M. ; VITTORINI, E.. **An energy saving routing algorithm for a green OSPF protocol**. IEEE INFOCOM, 2010.
- [62] Network World. **Redes “verdes”**, 2010. Disponível em: <www.networkworld.es/networking/redes-verdes>. Acesso em: mar. 2015.
- [63] LANZISERA, S.; NORDMAN, B. ; BROWN, R. E.. **Data network equipment energy use and savings potencial in buildings**. Technical report, Lawrence Berkeley National Laboratory, jun. 2010.
- [64] FISHER, W.; SUCHARA, M. ; REXFORD, J.. **Greening backbone networks: reducing energy consumption by shutting off cables in bundled links**. In: PROCEEDINGS OF THE FIRST ACM SIGCOMM WORKSHOP ON GREEN NETWORKING, p. 29–34, 2010.
- [65] ZHANG, M.; YI, C.; LIU, B. ; ZHANG, B.. **GreenTE: Power-aware traffic engineering**,. In: IEEE INTERNATIONAL CONFERENCE ON NETWORK PROTOCOLS (ICNP), 2010.
- [66] BIANZINO, A. P.; CHAUDET, C.; LARROCA, F.; ROSSI, D. ; ROUGIER, J.-L.. **Energy-aware routing: a reality check**. In: GLOBECOM WORKSHOPS (GC WKSHPS), IEEE, p. 1422–1427, 2010.
- [67] DE AQUINO, A. R.. **Otimização em grafos: árvores geradoras com restrições**. Universidade Federal de Pernambuco, 2010.
- [68] IEEE Standards Association. P802.3az Energy Efficient Ethernet Task Force. **IEEE Std 802.3az-2010**, set. 2010. Disponível em: <<http://www.ieee802.org/3/az>>. Acesso em: nov. 2013.

- [69] REVIRIEGO, P.; HERNADEZ, J.-A.; LARRABEITI, D. ; MAESTRO, J. A.. **Burst transmission for energy-efficient Ethernet**. IEEE Internet Computing, 14(4):50–57, 2010.
- [70] CHRISTENSEN, K.; REVIRIEGO, P.; NORDMAN, B.; MOSTOWFI, M. ; MAESTRO, J. A.. **IEEE 802.3az: the road to energy efficient Ethernet**. IEEE Communications Magazine, p. 50–56, 2010.
- [71] UNICAMP. **Modelagem Matemática**, 2010. Disponível em: <http://www.ime.unicamp.br/~moretti/ms428/2sem2010/aula_slides.pdf>. Acesso em: jan. 2015.
- [72] KOSTER, A. M. C. A.; MUÑOZ, X.. **Graphs and Algorithms in Communication Networks: Studies in Broadband, Optical, Wireless and Ad Hoc Networks**. Springer, 2010.
- [73] ORLOWSKI, S.; PIÓRO, M.; TOMASZEWSKI, A. ; WESSÄLY, R.. **SNDlib 1.0 – Survivable Network Design library**. Networks, 55(3):276–286, 2010. Disponível em: <<http://sndlib.zib.de/home.action>>. Acesso em: jul. 2014.
- [74] BAZARAA, M. S.; JARVIS, J. J. ; SHERALI, H. D.. **Linear Programming and Network Flows**. Wiley, 2010.
- [75] CUOMO, F.; ABBAGNALE, A. ; PAPAGNA, S.. **Esol: Energy saving in the Internet based on occurrence of links in routing paths**. World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE Computer Society, p. 1–6, 2011.
- [76] BOLLA, R.; DAVOLI, F.; BRUSCHI, R.; CHRISTENSEN, K.; CUCCHIETTI, F. ; SINGH, S.. **The potential impact of green technologies in next-generation wireline networks: is there room for energy saving optimization?** IEEE Communications Magazine, 2011.
- [77] GOMA, E.; CANINI, M.; TOLEDO, A. L.; LAOUTARIS, N.; KOSTIĆ, D.; RODRIGUEZ, P.; STANOJEVIĆ, R. ; VALENTÍN, P. Y.. **Insomnia**

- in the access or how to curb access network related energy consumption.** In: PROCEEDINGS OF ACM SIGCOMM, p. 338–349, 2011.
- [78] BOLLA, R.; BRUSCHI, R.; DAVOLI, F. ; CUCCHIETTI, F.. **Energy efficiency in the future Internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures.** IEEE Communications Surveys and Tutorials, 13(2), 2011.
- [79] MIERCOM. **Lab Testing Summary Report: Plug and Play Switches**, 2011. Disponível em: <<http://www.miercom.com/pdf/reports/20110829.pdf>>. Acesso em: jun. 2015.
- [80] LANGE, C.; KOSIANKOWSKI, D.; WEIDMANN, R. ; GLADISCH, A.. **Energy consumption of telecommunication networks and related improvement options.** IEEE Journal of Selected Topics in Quantum Electronics (IEEE J SEL TOP QUANTUM ELECTR), 17(2):285–295, 2011.
- [81] CUOMO, F.; ABBAGNALE, A.; CIANFRANI, A. ; POLVERINI, M.. **Keeping the connectivity and saving the energy in the Internet.** IEEE INFOCOM 2011 Workshop on Green Communications and Networking, p. 10–15, 2011.
- [82] BOLLA, R.; BRUSCHI, R.; CIANFRANI, A. ; LISTANTI, M.. **Enabling backbone networks to sleep.** IEEE Network Magazine, 25(2):26–31, 2011.
- [83] GELENBE, E.; MAHMOODI, T.. **Energy-aware routing in the cognitive packet network.** In: INTERNATIONAL CONFERENCE ON SMART GRIDS, GREEN COMMUNICATIONS, AND IT ENERGY-AWARE TECHNOLOGIES (ENERGY 2011), 2011.
- [84] AMALDI, E.; CAPONE, A.; GIANOLI, L. ; MASCETTI, L.. **A MILP-based heuristic for energy-aware traffic engineering with shortest path routing.** Network Optimization, p. 464–477, 2011.

- [85] PREM, A.; CHAUDET, C.; ROSSI, D.; MORETTI, S. ; ROUGIER, J.-L.. **The green-game: striking a balance between QoS and energy saving.** In: PROCEEDINGS OF THE 23RD INTERNATIONAL TELETRAFFIC CONGRESS (ITC23), p. 262–269, 2011.
- [86] CIANFRANI, A.; ERAMO, V.; LISTANTI, M. ; POLVERINI, M.. **An OSPF enhancement for energy saving in IP networks.** IEEE INFOCOM 2011 Workshop on Green Communications and Networking, p. 10–15, 2011.
- [87] BIANZINO, A. P.; CHIARAVIGLIO, L. ; MELLIA, M.. **Grida: a green distributed algorithm for backbone networks.** In: ONLINE CONFERENCE ON GREEN COMMUNICATIONS (GREENCOM), IEEE, p. 113–119, 2011.
- [88] LEE, S. S. W.; TSENG, P.-K. ; CHEN, A.. **Link weight assignment and loop-free routing table update for link state routing protocols in energy-aware Internet.** Future Generation Computer Systems, 28:437–445, 2011.
- [89] HERRERÍA-ALONSO, S.; RODRÍGUEZ-PÉREZ, M.; FERNÁNDEZ-VEIGA, M. ; LÓPEZ-GARCÍA, C.. **A power saving model for burst transmission in energy-efficient Ethernet.** IEEE Communications Letters, 15(5):584–586, 2011.
- [90] REVIRIEGO, P.; CHRISTENSEN, K.; RABANILLO, J. ; MAESTRO, J. A.. **An initial evaluation of energy efficient Ethernet.** IEEE Communications Letters, 15(5):578–580, 2011.
- [91] MARSAN, M. A.; ANTA, A. F.; MANCUSO, V.; RENGARAJAN, B.; VASALLO, P. R. ; RIZZO, G.. **A simple analytical model for energy efficient Ethernet.** IEEE COMMUNICATIONS LETTERS, p. 1–3, 2011.
- [92] HERRERÍA-ALONSO, S.; RODRÍGUEZ-PÉREZ, M.; FERNÁNDEZ-VEIGA, M. ; LÓPEZ-GARCÍA, C.. **How efficient is energy efficient**

Ethernet? In: 3RD INTERNATIONAL CONGRESS ON ULTRA MODERN TELECOMMUNICATIONS AND CONTROL SYSTEMS AND WORKSHOPS (ICUMT), 2011.

- [93] HERRERÍA-ALONSO, S.; RODRÍGUEZ-PÉREZ, M.; FERNÁNDEZ-VEIGA, M. ; LÓPEZ-GARCÍA, C.. **Opportunistic power saving algorithms for Ethernet devices**. Computer Networks, 55(9):2051–2064, 2011.
- [94] LARRABEITI, D.; REVIRIEGO, P.; HERNÁNDEZ, J. A.; MAESTRO, J. A. ; URUEÑA, M.. **Towards an energy efficient 10 Gb/s optical Ethernet: performance analysis and viability**. Optical Switching and Networking, 8(3):131–138, 2011.
- [95] CHU, H.-W.; CHEUNG, C.-C.; HO, K.-H. ; WANG, N.. **Green MPLS traffic engineering**. In: AUSTRALASIAN TELECOMMUNICATION NETWORKS AND APPLICATIONS CONFERENCE (ATNAC), p. 1–4. IEEE, 2011.
- [96] CARDOSO, A.. **Fundamentos da Pesquisa Operacional**. UNIFAL-MG, 2011.
- [97] MARINS, F. A. S.. **Introdução à Pesquisa Operacional**. UNESP, 2011.
- [98] LAPPETELÄINEN, A.. **Equal cost multipath routing in IP networks**. Master's thesis, Aalto University School of Science and Technology, 2011.
- [99] AVALLONE, S.; VENTRE, G.. **Energy efficient online routing of flows with additive constraints**. Computer Networks, 56(10):2368–2382, 2012.
- [100] BIANZINO, A. P.; CHAUDET, C.; ROSSI, D. ; ROUGIER, J.-L.. **A survey of green networking research**. IEEE Communications Surveys and Tutorials, 14(1):3–20, 2012.

- [101] CAPONE, A.; KILPER, D. ; NIU, Z.. **Editorial for computer networks special issue on “green communication networks”**. Computer Networks, 56(10):2317–2319, jul. 2012.
- [102] Real-Time Systems Lab. **Energy-efficient Networking**, 2012. Disponível em: <<http://www.ida.liu.se/labs/rtslab/courses/greennw>>. Acesso em: nov. 2012.
- [103] LAMBERT, S.; HEDDEGHEM, W. V.; VEREECKEN, W.; LANNOO, B.; COLLE, D. ; PICKAVET, M.. **Worldwide electricity consumption of communication networks**. Optics Express, 20(26):B513–B514, 2012.
- [104] CHIARAVIGLIO, L.; MELLIA, M. ; NERI, F.. **Minimizing ISP network energy cost: formulation and solutions**. IEEE/ACM Transactions on Networking, 20(2):463–476, 2012.
- [105] RIZZELLI, G.; MOREA, A.; TORNATORE, M. ; RIVAL, O.. **Energy efficient traffic-aware design of on-off multi-layer translucent optical networks**. Computer Networks, 56(10):2443–2455, 2012.
- [106] KRIEF, F.. **Green Networking**. Wiley, 2012.
- [107] BOLLA, R.; BRUSCHI, R.; CARREGA, A.; DAVOLI, F.; SUINO, D.; VASSILAKIS, C. ; ZAFEIROPOULOS, A.. **Cutting the energy bills of Internet Service Providers and telecoms through power management: an impact analysis**. Computer Networks, 56:2320–2342, 2012.
- [108] KIM, Y.-M.; LEE, E.-J.; PARK, H.-S.; CHOI, J.-K. ; PARK, H.-S.. **Ant colony based self-adaptive energy saving routing for energy efficient Internet**. Computer Networks, 56(10):2343–2354, 2012.
- [109] CUOMO, F.; CIANFRANI, A.; POLVERINI, M. ; MANGIONE, D.. **Network pruning for energy saving in the Internet**. Computer Networks, 56:2355–2367, 2012.

- [110] HOU, W.; GUO, L.; WEI, X. ; GONG, X..
Multi-granularity and robust grooming in power- and port-cost-efficient IP over WDM networks. Computer Networks, 56(10):2383–2399, 2012.
- [111] VIZCAÍNO, J. L.; YE, Y. ; MONROY, I. T..
Energy efficiency analysis for flexible-grid OFDM-based optical networks. Computer Networks, 56(10):2400–2419, 2012.
- [112] RICCIARDI, S.; PALMIERI, F.; FIORE, U.; CAREGLIO, D.; SANTOS-BOADA, G. ; SOLÉ-PARETA, J..
An energy-aware dynamic RWA framework for next-generation wavelength-routed networks. Computer Networks, 56(10):2420–2442, 2012.
- [113] CHATZIPAPAS, A..
Using energy efficient ethernet (802.3az) in web hosting centers. Master's thesis, University Carlos III of Madrid, 2012.
- [114] WOLKERSTORFER, M.; STATOVCI, D. ; NORDSTRÖM, T..
Energy-saving by low-power modes in ADSL2. Computer Networks, 56(10):2468–2480, 2012.
- [115] MANCUSO, V.; ALOUF, S..
Analysis of power saving with continuous connectivity. Computer Networks, 56:2481–2493, 2012.
- [116] TURCK, K. D.; VUYST, S. D.; FIEMS, D.; WITTEVRONGEL, S. ; BRUNEEL, H..
Performance analysis of sleep mode mechanisms in the presence of bidirectional traffic. Computer Networks, 56(10):2494–2505, 2012.
- [117] GOMEZ, K.; BORU, D.; RIGGIO, R.; RASHEED, T.; MIORANDI, D. ; GRANELLI, F..
Measurement-based modelling of power consumption at wireless access network gateways. Computer Networks, 56(10):2506–2521, 2012.
- [118] DA SILVA, A. P. C.; MEO, M. ; MARSAN, M. A..
Energy-performance trade-off in dense WLANs: a queuing study. Computer Networks, 56:2522–2537, 2012.

- [119] NGUYEN, K. K.; CHERIET, M.; LEMAY, M.; REIJS, V.; MACKAREL, A. ; PASTRAMA, A.. **Environmental-aware virtual data center network**. Computer Networks, 2012.
- [120] DE LIMA, A. A.. **Análise e Implementação de um Algoritmo de Busca Dispersa para o Planejamento da Expansão de Sistemas de Transmissão**. UNESP, 2012.
- [121] CIANFRANI, A.; ERAMO, V.; LISTANTI, M.; POLVERINI, M. ; VASILAKOS, A. V.. **An OSPF-Integrated Routing Strategy for QoS-Aware Energy Saving in IP Backbone Networks**. IEEE Transactions on Network and Service Management, 9(3):254–267, 2012.
- [122] LIN, G.; SOH, S.; CHIN, K.-W. ; LAZARESCU, M.. **Efficient heuristics for energy-aware routing in networks with bundled links**. Computer Networks, 57(8):1774–1788, 2013.
- [123] ALECRIM, E.. **O que é virtualização e para que serve?** InfoWester, 2013. Disponível em: <<http://www.infowester.com/virtualizacao.php>>. Acesso em: jun. 2015.
- [124] AMALDI, E.; CAPONE, A. ; GIANOLI, L. G.. **Energy-aware IP traffic engineering with shortest path routing**. Computer Networks, 57(6):1503–1517, 2013.
- [125] HILLIER, F. S.; LIEBERMAN, G. J.. **Introdução à Pesquisa Operacional**. McGraw-Hill, 2013.
- [126] VAN DEN BRIEL, M.. **Combining Linear Programming Based Decomposition Techniques with Constraint Programming**. NICTA and ANU, 2013. Disponível em: <www.cse.unsw.edu.au/~tw/school2013/ColGen.pptx>. Acesso em: jun. 2014.
- [127] FURTADO, M.. **Bytes eficientes – Data Centers já respondem por 2% da eletricidade consumida no mundo e viram foco de projetos de racionalização**. Brasil Energia, 2013.

- [128] KHAN, S.; MAURI, J. L.. **Green Networking and Communications: ICT for Sustainability**. CRC Press, 2014.
- [129] ADDIS, B.; CAPONE, A.; CARELLO, G.; GIANOLI, L. G. ; SANSÒ, B.. **On the energy cost of robustness and resiliency in IP networks**. Computer Networks, 75(Part A):239–259, 2014.
- [130] ADDIS, B.; CAPONE, A.; CARELLO, G.; GIANOLI, L. G. ; SANSÒ, B.. **Energy management through optimized routing and device powering for greener communication networks**. IEEE/ACM Transactions on Networking, 22(1):313–325, 2014.
- [131] JÚNIOR, C.. **Crise Energética: Quem paga a conta?** InteliAgro - Inovações e Tendências do Agronegócio, 2015. Disponível em: <<http://inteliagro.com.br/crise-energetica-quem-paga-a-conta>>. Acesso em: fev. 2015.
- [132] ILUMINA - Instituto de Desenvolvimento Estratégico do Setor Energético. **Imagens dizem mais do que mil palavras**, 2015. Disponível em: <<http://ilumina.org.br/imagens-dizem-mais-do-que-mil-palavras>>. Acesso em: fev. 2015.
- [133] LOUREIRO, A. A. F.; GOUSSEVSKAIA, O. N.. **Grafos**. UFMG, 2015. Disponível em: <http://homepages.dcc.ufmg.br/~loureiro/md/md_9Grafos.pdf>. Acesso em: ago. 2015.

A

Técnicas ECMP

A sigla ECMP é derivada do nome em inglês *Equal-Cost Multi-Path* e consiste em uma técnica de roteamento para o encaminhamento de pacotes ao longo de vários caminhos de mesmo custo [17]. Este recurso apresenta duas grandes vantagens. A primeira e mais óbvia está relacionada ao desempenho da rede, pois possibilita o melhor aproveitamento dos recursos disponíveis a partir do balanceamento de carga. A segunda é inerente a mecanismos de proteção, que são característicos em sistemas com tolerância a falhas, pois permite a redução no tempo de estabilização após uma falha, uma vez que em situação de queda de um ou mais circuitos que compõem os caminhos de mesmo custo, o tráfego pode ser acomodado nos circuitos restantes numa fração de tempo pequena e sem perda significativa de dados.

Cabe esclarecer que entende-se por tolerância a falhas como a propriedade que permite que um sistema continue a funcionar corretamente mesmo em situações de falha de um ou mais de seus componentes.

As técnicas ECMP podem ser adotadas em cenários com roteamento estático ou dinâmico. Neste segundo caso é mais natural o seu emprego com protocolos de roteamento do tipo *estado do enlace* (em inglês, *link-state*), como o OSPF, pois tais protocolos baseiam-se no algoritmo SPF para a definição dos caminhos a serem escolhidos pelos roteadores que compõem a rede e, uma vez que neste algoritmo já está prevista a possibilidade de cálculo de custos iguais para mais de um caminho, não é necessária nenhuma configuração especial.

Vale citar, contudo, que já foram publicados relatos na literatura de implementações de ECMP com protocolos de roteamento do tipo *distance vector*, como, por exemplo, com o IGRP (*Interior Gateway Routing Protocol*) e o EIGRP (*Enhanced Interior Gateway Routing Protocol*) [98].

Um exemplo de balanceamento de carga obtido por meio de ECMP é mostrado na Figura A.1, onde o tráfego é distribuído de forma equilibrada pelos caminhos de mesmo custo.

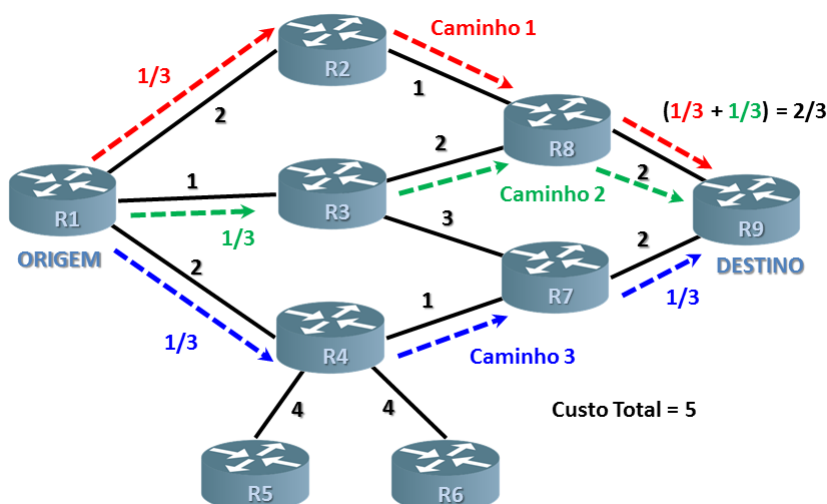


Figura A.1 – ECMP como mecanismo de balanceamento de carga

Fonte: Elaborada pela autora

Salienta-se que os três caminhos destacados na Figura A.1 funcionam como *backups* entre si. Assim, caso um deles fique indisponível, como representado na Figura A.2, o tráfego é dividido entre os outros dois remanescentes após a detecção da falha, sem que ocorra uma grande perda de pacotes, o que caracteriza o emprego do protocolo ECMP como mecanismo de proteção.

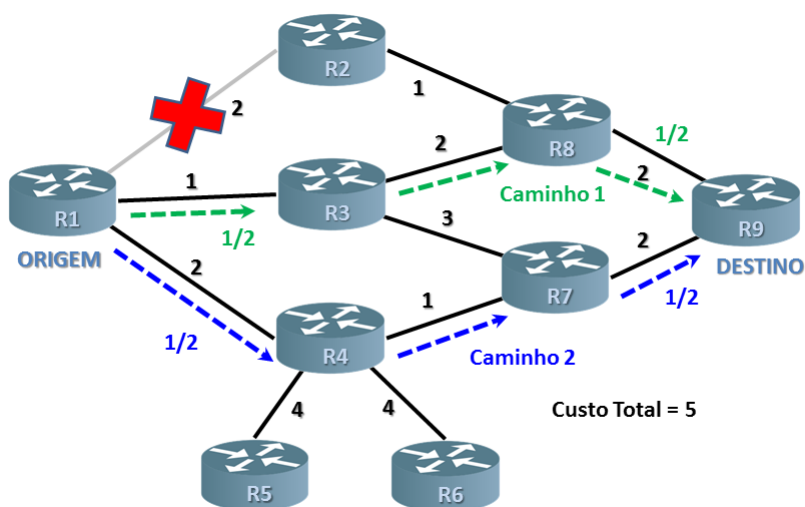


Figura A.2 – ECMP como mecanismo de proteção

Fonte: Elaborada pela autora

Apesar, no entanto, de todos os benefícios, a implementação de mecanismos ECMP requer bastante cuidado, principalmente em grandes redes que percorrem grandes distâncias, pois pode acarretar problemas no funcionamento de muitos protocolos IP, mais especificamente o TCP e o de descoberta de MTU (*Maximum Transmission Unit*), em função dos impactos provocados por alguns fatores como variações bruscas de latência, problemas na reordenação dos pacotes nos elementos de destino e problemas na negociação de MTU ao longo do caminho.

Diversos trabalhos já foram publicados acerca do protocolo ECMP, dentre os quais pode-se citar a RFC 2991 [16], que discute o roteamento *multipath* de forma geral e a RFC 2992 [17], que apresenta uma estratégia de roteamento *multipath* específica, denominada *hash-threshold*, segunda a qual o roteador seleciona, primeiramente, uma chave efetuando uma função *hash* (por exemplo, CRC16) sobre os campos do cabeçalho do pacote que identificam um fluxo. A todos os “N” *next-hops* são atribuídas regiões únicas no espaço da chave, a qual é utilizada pelo roteador para determinar qual região e, portanto, qual *next-hop* usar para cada fluxo.

Além disso, foram desenvolvidos também outros algoritmos para tratar os problemas inerentes ao ECMP citados anteriormente. Alguns exemplos desses algoritmos, cujo detalhamento pode ser encontrado em [98], são: *Table-based Hashing*, *Table-based Hashing with Reassignment* (THR), *Fast Switching* (FS), *LRU-based Caching with Counting* (LCC), *Dual hash table and counters* (DHTC), *Flowlet Aware Routing Engine* (FLARE) e *Link-Criticality-based ECMP routing* (LCER).

B

Teoria da Dualidade

Um dos conceitos mais importantes em Programação Linear, tanto do ponto de vista teórico como prático, é a **Teoria da Dualidade** [14, 74].

O termo dualidade refere-se ao fato de que cada modelo de PL consiste de duas formas que possuem propriedades intrinsecamente relacionadas. A primeira, ou original, é chamada de **Primal**¹, e a segunda forma do modelo é chamada de **Dual**.

A formulação do problema dual segue algumas regras. Assim, considere-se um problema de PL, de maximização, na forma canônica:

$$\begin{aligned} \text{maximize} \quad & z = \underline{c}^T \cdot \underline{x} \\ \text{sujeito a:} \quad & \\ & \underline{A} \cdot \underline{x} \leq \underline{b} \\ & \underline{x} \geq \underline{0} \end{aligned} \tag{B-1}$$

Nesta formulação, $\underline{A} \in \mathbb{R}^{m \times n}$, $\underline{x} \in \mathbb{R}^n$, $\underline{b} \in \mathbb{R}^m$ e $\underline{c} \in \mathbb{R}^n$. Como já sabido, a matriz \underline{A} é a matriz das restrições ou matriz de coeficientes, e contém os coeficientes tecnológicos que compõem as restrições. O vetor \underline{x} é o vetor das incógnitas ou vetor de decisão, já que contém a lista das variáveis de decisão consideradas no problema. O vetor \underline{b} é conhecido como lado direito das restrições, vetor das necessidades ou ainda termo independente das restrições, e indica a disponibilidade de recursos associados à cada restrição. Por fim, o vetor \underline{c} é conhecido como vetor de custos do problema, e contém os coeficientes de custo que compõem a função objetivo.

¹Embora o termo “dual” seja oriundo da Álgebra Linear, o termo “primal”, na realidade, foi sugerido pelo pai de Dantzig, Tobias Dantzig (que era um matemático), como um antônimo latino apropriado para o termo “dual”, para evitar que todas as vezes que se fosse referir ao primal tivesse que ser dito: “o problema original do qual este é o dual”.

O dual do PL descrito anteriormente é dado por:

$$\begin{aligned} \text{minimize} \quad & w = \underline{b}^T \cdot \underline{y} \\ \text{sujeito a:} \quad & \\ & \underline{A}^T \cdot \underline{y} \geq \underline{c} \\ & \underline{y} \geq \underline{0} \end{aligned}$$

De forma similar, dado um problema de PL, de minimização, na forma típica:

$$\begin{aligned} \text{minimize} \quad & z = \underline{c}^T \cdot \underline{x} \\ \text{sujeito a:} \quad & \\ & \underline{A} \cdot \underline{x} \geq \underline{b} \\ & \underline{x} \geq \underline{0} \end{aligned}$$

existe um outro problema de PL que lhe está associado, o dual, que consiste em:

$$\begin{aligned} \text{maximize} \quad & w = \underline{b}^T \cdot \underline{y} \\ \text{sujeito a:} \quad & \\ & \underline{A}^T \cdot \underline{y} \leq \underline{c} \\ & \underline{y} \geq \underline{0} \end{aligned}$$

Destaca-se que existe exatamente uma variável dual para cada restrição primal e exatamente uma restrição dual para cada variável primal.

As regras de transformação que se aplicam são as seguintes:

- (i) A cada variável do **primal** corresponde uma restrição no **dual**, conforme representado esquematicamente na Figura B.1. Assim, o número de restrições do **dual** é igual ao número de variáveis do **primal**;
- (ii) A cada restrição do **primal** corresponde uma variável do **dual**, conforme representado esquematicamente na Figura B.2. Assim, o número de variáveis do **dual** é igual ao número de restrições do **primal**;



Figura B.1 – Relação entre as variáveis do Problema Primal e as restrições do Problema Dual

Fonte: Elaborada pela autora



Figura B.2 – Relação entre as restrições do Problema Primal e as variáveis do Problema Dual

Fonte: Elaborada pela autora

- (iii) Os coeficientes da função objetivo do **primal** correspondem aos termos independentes das restrições do **dual**;
- (iv) Os termos independentes das restrições do **primal** correspondem aos coeficientes da função objetivo do **dual**;
- (v) A transposta da matriz de restrições do **primal** é a matriz de restrições do **dual**;
- (vi) Se o **primal** for um problema de maximização na forma típica, então, o problema **dual** será um problema de minimização na forma típica;
- (vii) analogamente, se o **primal** for um problema de minimização, então, o problema **dual** será um problema de maximização.

Alguns requisitos para que a conversão seja possível são:

- ★ Todas as variáveis devem ser não-negativas;
- ★ Todas as restrições devem ser desigualdades com:
 - Modelo de maximização \Rightarrow restrições do tipo “ \leq ”
 - Modelo de minimização \Rightarrow restrições do tipo “ \geq ”

As propriedades que devem ser respeitadas são as descritas a seguir e sintetizadas no Quadro B.1.

– **Propriedade 1**

Se a restrição do primal é de igualdade, então, a variável do dual é irrestrita (livre).

– **Propriedade 2**

Se a restrição do primal é do tipo menor ou igual, então a variável do dual é não-positiva.

– **Propriedade 3**

Se a variável do primal é não-positiva, então, a restrição do dual é do tipo maior ou igual.

– **Propriedade 4**

Se a variável do primal é sem restrição de sinal (livre), então a restrição do dual é de igualdade.

Quadro B.1 – Relações de conversão entre os problemas Primal e Dual

		→				
		PRIMAL		DUAL		
MIN	Restrição	\leq	\leq	Variável	MAX	
		$=$	qq.			
		\geq	\geq			
	Variável	\leq	\geq	Restrição		
		qq.	$=$			
		\geq	\leq			
		DUAL		PRIMAL		
		←				

Fonte: De Oliveira [52]

Vale destacar que as soluções de qualquer par de problemas duais, em particular as ótimas (ou seja, aquelas que otimizam a função objetivo), relacionam-se entre si pelas propriedades da Dualidade, as quais serão descritas a seguir.

– **Teorema 1** (*Auto-inversibilidade da Dualidade*)

O dual do dual é o primal, conforme representado esquematicamente na Figura B.3. Os termos dual e primal, portanto, dependem do referencial.



Figura B.3 – Auto-inversibilidade da dualidade

Fonte: Elaborada pela autora

– **Teorema 2** (*Dualidade Fraca*)

Sabendo-se que solução viável ou possível refere-se a um ponto que atenda às restrições do problema, estando incluídas as restrições, mas não incluída a função a ser otimizada, isto é, as soluções possíveis precisam atender a todas as restrições, mas não precisam otimizar a função objetivo, o valor da função objetivo de qualquer solução viável para o problema de minimização é sempre maior ou igual ao valor da função objetivo de qualquer solução viável para o problema de maximização.

Em outras palavras, o valor da função objetivo de qualquer solução viável para o problema de minimização fornece um limitante superior para a função objetivo ótima do problema de maximização; analogamente, o valor da função objetivo de qualquer solução viável para o problema de maximização fornece um limitante inferior para a função objetivo ótima do problema de minimização.

Ou seja, se o primal é um problema de maximização, o valor da função objetivo do dual é sempre maior ou igual ao valor da função objetivo do primal.

Considerando o problema de maximização descrito pela eq. (B-1), o valor da função objetivo, z , de qualquer solução admissível do problema primal, $\underline{x} = (x_1, x_2, \dots, x_n)$, não excede o valor da função objetivo, w , de qualquer solução admissível do problema dual, $\underline{y} = (y_1, y_2, \dots, y_n)$.

Corolário 1: *para soluções viáveis:* todo valor da função objetivo do primal é um limitante inferior para a função objetivo do dual.

Corolário 2: para soluções viáveis: todo valor da função objetivo do dual é um limitante superior para a função objetivo do primal.

Corolário 3: se o primal é viável e o valor de $z \rightarrow \infty$ então o dual é inviável (ou seja, não existe solução possível. Em outras palavras, não há qualquer ponto que atenda ao conjunto de restrições).

Corolário 4: se o dual é viável e $w \rightarrow -\infty$ então o primal é inviável.

Corolário 5: se o primal é viável e o dual inviável então o primal é ilimitado ($z \rightarrow \infty$). Cabe esclarecer que se define como solução ilimitada aquela em que a função objetivo pode crescer (caso da maximização) ou decrescer (caso da minimização), indefinidamente, atendendo a todas as restrições do problema.

Corolário 6: se o dual é viável e o primal inviável então o dual é ilimitado ($w \rightarrow -\infty$).

Nos Quadros B.2 e B.3 são apresentados, respectivamente, os corolários da Teoria da Dualidade e as propriedades intrínsecas entre o Problema Primal (PP) e o Problema Dual (PD).

Quadro B.2 – Resumo dos corolários da Teoria da Dualidade

Dual \ Primal	Tem solução viável	Não tem solução viável
Tem solução viável	$\min z = \max w$	$w \rightarrow -\infty$
Não tem solução viável	$z \rightarrow \infty$	Ambos impossíveis

Fonte: De Oliveira [52]

– **Teorema 3 (Critério de Otimalidade)**

Se \underline{x}^* e \underline{y}^* são soluções viáveis, respectivamente, para o primal e o dual na forma simétrica, de tal modo que os valores das funções objetivo são iguais ($z^* = w^*$); então tem-se que \underline{x}^* é solução ótima para o primal e \underline{y}^* é solução ótima para o dual.

Quadro B.3 – Propriedades intrínsecas entre o PP e o PD

PP Ótimo Finito	\Leftrightarrow	PD Ótimo Finito
PP Ilimitado	\Rightarrow	PD Infactível
PD Ilimitado	\Rightarrow	PP Infactível
PP Infactível	\Rightarrow	PD Ilimitado/Infactível
PD Infactível	\Rightarrow	PP Ilimitado/Infactível

Fonte: De Oliveira [52]

– **Teorema 4 (*Dualidade Forte*)**

Se os problemas primal e dual são viáveis, então, ambos têm soluções ótimas e os valores ótimos de suas funções objetivo são iguais ($z^* = w^*$).

– **Teorema 5 (*Condições de Folgas Complementares - CFC*)**

Além dos valores ótimos das funções objetivo dos problemas primal e dual serem iguais, suas soluções básicas também estão interligadas.

O teorema das folgas complementares permite a determinação da solução ótima do dual, a partir da solução ótima do primal e vice-versa.

Considerando-se \underline{x}^* e \underline{y}^* , respectivamente soluções viáveis para os problemas primal e dual, então \underline{x}^* e \underline{y}^* são soluções ótimas para seus problemas se, e somente se, a seguinte condição, denominada Condição de Folga Complementar, for respeitada: $(\underline{y}^* \cdot \underline{A} - \underline{c}) \cdot \underline{x}^* + \underline{y}^* \cdot (\underline{b} - \underline{A} \cdot \underline{x}^*) = 0$.

– **Teorema 6**

Se para algum dos problemas existir solução não limitada, então, o outro não possui soluções admissíveis.

Em outras palavras, se o primal é ilimitado, então, o dual é impossível. Analogamente, se o dual é ilimitado, então, o primal é impossível.

– **Teorema 7**

Para qualquer par de problemas duais, a existência de solução ótima (finita) para um deles garante a existência de solução ótima (finita) ao outro.

Para exemplificar, supondo que o primal fosse representado pelo problema de maximização descrito pelo sistema a seguir:

$$\text{maximize } z = 2x_1 + x_2$$

sujeito a:

$$x_1 + x_2 = 5$$

$$x_1 \leq 3$$

$$x_2 \geq 3$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

seu dual seria dado por:

$$\text{minimize } z = 5y_1 + 3y_2 + 3y_3$$

sujeito a:

$$y_1 + y_2 \geq 2$$

$$y_1 + y_3 \geq 1$$

$$y_1 \in \mathbb{R}$$

$$y_2 \geq 0$$

$$y_3 \leq 0$$

Aplicando-se o conceito a um problema envolvendo variáveis complicadoras, como o representado a seguir, onde x_1 é a variável complicadora [37]:

PROBLEMA PRIMAL (COM VARIÁVEIS COMPLICADORAS)

$$\text{maximize}_{y_1, y_2, y_3, y_4, y_5, x_1} \quad 4y_1 + 3y_2 + 2y_3 + 3y_4 + 2y_5 + 3x_1$$

sujeito a:

$$\begin{array}{rclcl}
y_1 & + & 2y_2 & & + & 2x_1 & \leq & 3 \\
2y_1 & + & y_2 & & + & x_1 & \leq & 3 \\
-2y_1 & + & 3y_2 & & + & x_1 & \leq & 7 \\
& & y_3 & & + & 3x_1 & \leq & 4 \\
& & 2y_3 & & - & x_1 & \leq & 3 \\
& & y_4 & & & & \leq & 1 \\
& & 2y_4 & + & 4y_5 & + & 3x_1 & \leq & 5 \\
& & 3y_4 & + & y_5 & - & x_1 & \leq & 4
\end{array}$$

o dual obtido seria um problema com restrições complicadoras, como mostrado abaixo:

PROBLEMA DUAL
(COM RESTRIÇÕES COMPLICADORAS)

$$\underset{u_1, u_2, \dots, u_8}{\text{minimize}} \quad 3u_1 + 3u_2 + 7u_3 + 4u_4 + 3u_5 + u_6 + 5u_7 + 4u_8$$

sujeito a:

$$\begin{array}{rclcl}
u_1 & + & 2u_2 & - & 2u_3 & & & = & 4 \\
2u_1 & + & u_2 & + & 3u_3 & & & = & 3 \\
& & & & u_4 & + & 2u_5 & & = & 2 \\
& & & & & & u_6 & + & 2u_7 & + & 3u_8 & = & 3 \\
& & & & & & & & 4u_7 & + & u_8 & = & 2 \\
2u_1 & + & u_2 & + & u_3 & + & 3u_4 & - & u_5 & & + & 3u_7 & - & u_8 & = & 3 \\
& & & & & & & & & & u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8 & \leq & 0
\end{array}$$

C

Condições KKT

Em problemas de otimização matemática, as condições de Karush-Kuhn-Tucker (também conhecidas por Teorema KKT, ou condições KKT ou simplesmente Kuhn-Tucker) são condições de primeira ordem necessárias, mas não suficientes, para que uma solução seja ótima, desde que atendidas algumas condições de regularidade. As condições de Kuhn-Tucker generalizam o método dos multiplicadores de Lagrange, pois, diferentemente deste que permite apenas restrições de igualdade, permitem também restrições de desigualdade [15, 60].

O nome do teorema advém do fato de tais condições terem sido inicialmente publicadas na Tese de Mestrado de William Karush, em 1939, e depois apresentadas por Harold W. Kuhn e Albert W. Tucker, em 1951, quando, então, tornaram-se realmente conhecidas. Sua formulação é apresentada sucintamente a seguir.

Considere-se o problema de otimização representado pela eq. (C-1).

$$\begin{aligned} \max_{\underline{x}} \quad & f(\underline{x}) \\ \text{sujeito a:} \end{aligned} \tag{C-1}$$

$$g_j(\underline{x}) \leq c_j \quad j = 1, \dots, m$$

$$x_i \geq 0 \quad i = 1, \dots, n$$

Deve-se obter, então, o Lagrangiano do referido problema em função da variável \underline{x} e de um conjunto de variáveis $\underline{\lambda}$ definidas como multiplicadores para as restrições de desigualdade, ou seja:

$$\mathcal{L}(\underline{x}, \underline{\lambda}) = f(\underline{x}) + \sum_{j=1}^m \lambda_j \cdot (c_j - g_j(\underline{x})) \tag{C-2}$$

As condições de Kuhn-Tucker obtidas com base no Lagrangiano necessárias (mas não suficientes) para que uma solução para o problema de otimização representado pela eq. (C-1) seja ótima são dadas por:

$$\begin{array}{llll} \frac{\partial \mathcal{L}}{\partial x_i} \leq 0 & x_i \geq 0 & x_i \cdot \frac{\partial \mathcal{L}}{\partial x_i} = 0 & \forall i = 1, \dots, n \\ g_j(\underline{x}) - c_j \leq 0 & \lambda_j \geq 0 & \lambda_j \cdot (c_j - g_j(\underline{x})) = 0 & \forall j = 1, \dots, m \end{array}$$

Assim, para cada variável x_i existem três condições que devem ser atendidas. De forma análoga, para cada restrição (e conseqüentemente cada λ_j) há também três condições que devem ser igualmente cumpridas. Cada ponto que representa uma solução para este sistema de equações é um possível candidato a máximo. Uma vez que todos estes pontos tenham sido obtidos é necessário, então, testá-los individualmente para confirmar sua otimalidade e definir efetivamente a solução ótima.

Para exemplificar o emprego das condições KKT, seja o problema de otimização de interesse representado pela eq. (C-3) [30]:

$$\begin{array}{ll} \max_{x_1, x_2} & f(x_1, x_2) = 4x_1 + 3x_2 \\ & \text{sujeito a:} \\ & 2x_1 + x_2 \leq 10 \\ & x_1, x_2 \geq 0 \end{array} \quad (\text{C-3})$$

Determinando seu Lagrangiano, tem-se:

$$\mathcal{L} = 4x_1 + 3x_2 + \lambda(10 - 2x_1 - x_2)$$

As condições KKT associadas são dadas por:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 4 - 2\lambda \leq 0 \quad x_1 \geq 0 \quad x_1(4 - 2\lambda) = 0 \quad (\text{C-4})$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 3 - \lambda \leq 0 \quad x_2 \geq 0 \quad x_2(3 - \lambda) = 0 \quad (\text{C-5})$$

$$g_1(\underline{x}) - c_1 \leq 0 = 2x_1 + x_2 - 10 \leq 0 \quad \lambda \geq 0 \quad \lambda(10 - 2x_1 - x_2) = 0 \quad (\text{C-6})$$

Resolve-se, então, este conjunto de inequações e equações a fim de obter

os valores que podem maximizar a função objetivo de interesse.

Resolvendo a condição de igualdade presente na eq. (C-4) resulta em:

$$x_1(4 - 2\lambda) = 0 \Rightarrow x_1 = 0 \text{ ou } \lambda = 2 \quad (\text{C-7})$$

Supondo $\lambda = 2$, então:

$$3 - \lambda \leq 0 \Rightarrow 1 \leq 0$$

que é obviamente falso. Consequentemente, a opção factível a partir da eq. (C-7) é $x_1 = 0$. Então:

$$x_2(3 - \lambda) = 0 \Rightarrow x_2 = 0 \text{ ou } \lambda = 3$$

Observa-se que, se $x_2 = 0$ (juntamente com $x_1 = 0$), então:

$$10\lambda = 0 \Rightarrow \lambda = 0$$

Mas tal valor não atenderia à condição $4 - 2\lambda \leq 0$. Assim, tem-se que $x_1 = 0$ e $\lambda = 3$. Consequentemente:

$$3(10 - x_2) = 0 \Rightarrow x_2 = 10$$

Desta forma, consegue-se resolver o sistema, sendo obtida a única solução, que é $x_1 = 0$, $x_2 = 10$, $\lambda = 3$ e, por conseguinte, $f(x_1, x_2) = 30$.

D

Método de Decomposição de Benders

O método de decomposição proposto por Benders em 1962 [10] é um algoritmo que permite resolver certos problemas de otimização de grande porte, desde que o sistema apresente estrutura bloco-angular dual como o representado na Figura 3.5(b) ou, em outras palavras, que seja composto por variáveis complicadoras [34, 35, 74, 126]. Em vez de considerar todas as restrições e variáveis de decisão simultaneamente, que é a abordagem tradicional mas que muitas vezes inviabiliza a solução, o método de decomposição de Benders propõe uma abordagem multiestágios, particionando o problema original em diversos outros menores e, conseqüentemente, dividindo o processo de tomada de decisões em várias etapas. Uma vez que os requisitos de memória e a complexidade computacional dos problemas de otimização crescem significativamente com o número de variáveis e restrições, resolver tais problemas menores iterativamente tende a ser mais eficiente do que resolver um único problema grande.

Na decomposição de Benders, em um primeiro estágio um Problema Mestre é resolvido para um subconjunto de variáveis. Os valores das variáveis restantes são determinados por um Subproblema em um segundo estágio no qual são levados em consideração os valores das variáveis obtidos no estágio anterior. Se o Subproblema determina que as decisões propostas no primeiro estágio são inviáveis, então, uma ou mais restrições são geradas e adicionadas ao Problema Mestre, o qual é, então, novamente resolvido.

Fica evidente, portanto, que o requisito inicial para que o método de decomposição de Benders funcione adequadamente é a possibilidade de fixação temporária de algumas variáveis do sistema.

Representando esquematicamente, considere-se o seguinte exemplo:

$$\underset{\alpha_1, \alpha_2, \beta_1, \gamma_1, \gamma_2, \Upsilon_1}{\text{maximize}} \quad e_1\alpha_1 + e_2\alpha_2 + f_1\beta_1 + g_1\gamma_1 + g_2\gamma_2 + h_1\Upsilon_1$$

sujeito a:

$$a_{11}\alpha_1 + a_{12}\alpha_2 + d_{11}\Upsilon_1 \leq a_1$$

$$a_{21}\alpha_1 + a_{22}\alpha_2 + d_{21}\Upsilon_1 \leq a_2$$

$$a_{31}\alpha_1 + a_{32}\alpha_2 + d_{31}\Upsilon_1 \leq a_3$$

$$b_{11}\beta_1 + d_{41}\Upsilon_1 \leq b_1$$

$$b_{21}\beta_1 + d_{51}\Upsilon_1 \leq b_2$$

$$c_{11}\gamma_1 + c_{12}\gamma_2 + d_{61}\Upsilon_1 \leq c_1$$

$$c_{21}\gamma_1 + c_{22}\gamma_2 + d_{71}\Upsilon_1 \leq c_2$$

$$c_{31}\gamma_1 + c_{32}\gamma_2 + d_{81}\Upsilon_1 \leq c_3$$

Depreende-se facilmente que, caso a variável Υ_1 possua valor conhecido Υ_1^{fixo} , o problema pode ser automaticamente decomposto em três problemas independentes, como mostrado a seguir [14].

$$\underset{\alpha_1, \alpha_2}{\text{maximize}} \quad e_1\alpha_1 + e_2\alpha_2$$

sujeito a:

$$a_{11}\alpha_1 + a_{12}\alpha_2 \leq a_1 - d_{11}\Upsilon_1^{fixo} + c_{11}\gamma_1 + c_{12}\gamma_2 \leq c_1 - d_{61}\Upsilon_1^{fixo} +$$

$$a_{21}\alpha_1 + a_{22}\alpha_2 \leq a_2 - d_{21}\Upsilon_1^{fixo} c_{21}\gamma_1 + c_{22}\gamma_2 \leq c_2 - d_{71}\Upsilon_1^{fixo}$$

$$a_{31}\alpha_1 + a_{32}\alpha_2 \leq a_3 - d_{31}\Upsilon_1^{fixo} c_{31}\gamma_1 + c_{32}\gamma_2 \leq c_3 - d_{81}\Upsilon_1^{fixo}$$

$$\underset{\beta_1}{\text{maximize}} \quad f_1\beta_1$$

+ sujeito a:

$$b_{11}\beta_1 \leq b_1 - d_{41}\Upsilon_1^{fixo}$$

$$b_{21}\beta_1 \leq b_2 - d_{51}\Upsilon_1^{fixo}$$

A dificuldade reside justamente em como obter o valor Υ_1^{fixo} de tal forma que seja possível resolver o problema de maneira decomposta e garantir que a solução obtida seja a ótima.

Portanto, buscando-se de forma iterativa um valor para as variáveis complicadoras que possa ser fixado o problema original pode ser resolvido de forma decomposta, sendo particionado em dois mais simples, a saber:

- (i) Problema Mestre, que consiste em uma versão relaxada do problema original, contendo um subconjunto eleito de variáveis, juntamente com suas respectivas restrições; e
- (ii) Problema Escravo ou Subproblema, que consiste na versão original do problema com os valores do subconjunto de variáveis eleitos para o Problema Mestre temporariamente fixados [10].

O método resolve cada um dos dois problemas mais simples iterativamente, um de cada vez. A cada ciclo, uma nova restrição, conhecida como corte de Benders, é adicionada ao Problema Mestre. Quando o problema original é de minimização, a resolução do novo Problema Mestre originado a partir da resolução do Subproblema permite estimar limites inferiores para o problema original.

O algoritmo cicla até que as funções objetivo do Problema Mestre e do Subproblema sejam iguais ou até que o limitante superior do problema seja igual ao seu limitante inferior. A solução do Problema Mestre é o limitante inferior. Já o limitante superior é obtido pela melhor solução completa do problema, a partir da junção dos conjuntos de variáveis do Mestre e de uma solução factível do Subproblema a ele associado. Ao final do processo, tem-se a solução ótima do problema linear inteiro misto original.

Formalmente o método de decomposição de Benders pode ser descrito como exposto a seguir.

Considere-se o problema de otimização original descrito por meio da função objetivo representada pela eq. (D-1) e pelo conjunto de restrições a ela associado.

$$\min_{\underline{x}, \underline{y}} z = \underline{a}^T \cdot \underline{x} + \underline{b}^T \cdot \underline{y} \quad (\text{D-1})$$

sujeito a:

$$\underline{A} \cdot \underline{x} + \underline{B} \cdot \underline{y} = \underline{r} \quad (\text{D-2})$$

$$\underline{C} \cdot \underline{x} + \underline{D} \cdot \underline{y} \leq \underline{s} \quad (\text{D-3})$$

$$\underline{x}_L \leq \underline{x} \leq \underline{x}_H \quad (\text{D-4})$$

$$\underline{y}_L \leq \underline{y} \leq \underline{y}_H \quad (\text{D-5})$$

onde $\underline{A} \in \mathbb{R}^{p \times m}$, $\underline{B} \in \mathbb{R}^{p \times m}$, $\underline{C} \in \mathbb{R}^{q \times n}$, $\underline{D} \in \mathbb{R}^{q \times n}$, $\underline{r} \in \mathbb{R}^p$, $\underline{s} \in \mathbb{R}^q$, $\underline{x} \in \mathbb{R}^m$ e $\underline{y} \in \mathbb{R}^n$.

Supondo as variáveis \underline{x} como sendo as variáveis complicadoras, pois se elas fossem fixadas tornariam o problema muito mais simples de ser resolvido, o algoritmo de Benders decompõe o problema representada pela eq. (D-1) e pelo conjunto de restrições em dois: um Problema Mestre que contém as variáveis \underline{x} e um Problema Escravo ou Subproblema, que contém as variáveis \underline{y} , e resolve-os de forma iterativa.

Salienta-se que a essência da decomposição de Benders reside justamente em determinar esta partição que em várias situações pode não ser nem clara nem imediata e, muitas vezes, exige dispor de conhecimento específico do problema.

Visando a facilitar a sua implementação, o algoritmo de Benders pode ser descrito por um processo constituído de quatro etapas, representadas esquematicamente na Figura D.1 e detalhadas a seguir.

Passo 0: Inicialização

1. Defina $v = 1$ como sendo o contador de iterações;
2. Defina os limites inferior \underline{x}_L e superior \underline{x}_H para o conjunto de variáveis complicadoras \underline{x} ;
3. Defina uma variável adicional σ ;
4. Defina o limite inferior σ_L para a variável σ ;

5. Resolva o problema inicial cuja solução é representada pelo par $(\underline{x}^{(v)}, \sigma^{(v)})$.

Passo 1: Problema Escravo ou Solução do Subproblema

1. Determine uma variável dual $\underline{\lambda}$;
2. Resolva o problema denominado de Escravo, que consiste na versão original do problema considerando fixos os valores das variáveis complicadoras, obtendo o valor $\underline{\lambda}^{(v)}$, que corresponde à variável dual associada à restrição de igualdade nesta iteração. Analogamente ao passo anterior, a solução deste problema é representada pelo par $(\underline{y}^{(v)}, \underline{\lambda}^{(v)})$.

Passo 2: Teste ou Verificação de Convergência

Se a diferença $z_H^{(v)} - z_L^{(v)}$ for menor do que uma tolerância prefixada ε , conclui-se que o algoritmo convergiu e a solução final do problema é o par $(\underline{x}^{(v)}, \underline{y}^{(v)})$. Caso contrário, deve-se continuar com o passo seguinte.

Passo 3: Problema Mestre

1. Incremente o contador v de uma unidade;
2. Resolva o problema denominado de Mestre, retornando em seguida ao **Passo 1**.

Para exemplificar, aplicando-se o algoritmo de Benders ao problema original descrito pela eq. (D-1) e pelo conjunto de restrições a ela associado, onde \underline{x} corresponde ao conjunto de variáveis complicadoras, ter-se-ia o seguinte em cada Passo:

Passo 0: Inicialização

$$\begin{array}{l}
 v = 1 \\
 \left. \begin{array}{l}
 \min_{\underline{x}, \sigma} z = \underline{a}^T \cdot \underline{x} + \sigma \\
 \underline{x}_L \leq \underline{x} \leq \underline{x}_H \\
 \sigma \geq \sigma_L
 \end{array} \right\} \Rightarrow \underline{x}^{(v)}, \sigma^{(v)}
 \end{array}$$

Passo 1: Problema Escravo ou Solução do Subproblema

$$\left. \begin{array}{l} \min_{\underline{y}} z = \underline{b}^T \cdot \underline{y} \\ \text{sujeito a:} \\ \underline{A} \cdot \underline{x} + \underline{B} \cdot \underline{y} = \underline{r} \\ \underline{C} \cdot \underline{x} + \underline{D} \cdot \underline{y} \leq \underline{s} \\ \underline{x} = \underline{x}^{(v)} : \underline{\lambda} \\ \underline{y}_L \leq \underline{y} \leq \underline{y}_H \end{array} \right\} \Rightarrow \underline{y}^{(v)}, \underline{\lambda}^{(v)}$$

Passo 2: Teste ou Verificação de Convergência

$$z_H^{(v)} = \underline{a}^T \cdot \underline{x}^{(v)} + \underline{b}^T \cdot \underline{y}^{(v)}$$

$$z_L^{(v)} = \underline{a}^T \cdot \underline{x}^{(v)} + \sigma^{(v)}$$

Se $z_H^{(v)} - z_L^{(v)} < \varepsilon$ então PARE. Solução ótima encontrada.

Caso Contrário $v = v + 1$

Passo 3: Problema Mestre

$$\left. \begin{array}{l} \min_{\underline{x}, \sigma} z = \underline{a}^T \cdot \underline{x} + \sigma \\ \text{sujeito a:} \\ \underline{b}^T \cdot \underline{y}^{(k)} + \left[\underline{\lambda}^{(k)} \right]^T \cdot [\underline{x} - \underline{x}^{(k)}] \geq \sigma \quad k = 1, \dots, v-1 \\ 0 \leq \underline{x} \leq \underline{x}_H \\ \sigma \geq \sigma_L \end{array} \right\} \Rightarrow \underline{x}^{(v)}, \sigma^{(v)}$$

Vá para o **Passo 1**.

Com relação ao método de decomposição de Benders cabem, portanto, as seguintes importantes observações:

- como o Problema Mestre é um problema aproximado, sua solução ótima é sempre um limite inferior para o valor da solução real;

- o Problema Mestre possui somente parte das restrições (ditas cortes) que definem a região viável do Problema Escravo;
- se a solução do Problema Mestre é avaliada no problema original, obtém-se um limite superior;
- originalmente o algoritmo deveria ciclar até que os limites inferior e superior fossem iguais, mas, na prática, adota-se a estratégia de definir-se uma determinada precisão desejada (ε) e o algoritmo é encerrado quando a diferença entre estes limites é inferior a este valor de precisão;
- em linhas gerais, o que o método propõe é retirar do problema a parte referente à variável complicadora, para depois inseri-la aos poucos em termos de sua função de custo; e
- a eficiência computacional do algoritmo de Benders depende principalmente de três quesitos:
 - (i) o número de ciclos necessários para a convergência global;
 - (ii) o tempo gasto na resolução do subproblema em cada iteração;
 - (iii) o tempo e o esforço computacional demandados para resolução do Problema Mestre.



Figura D.1 – Representação esquemática do algoritmo de Benders

Fonte: Elaborada pela autora

E

Problemas Mestre e Escravo

Considerando as etapas do algoritmo de Benders representadas na Figura D.1, cabem algumas observações importantes com a relação ao **Passo 1** (denominado Problema Escravo ou Solução do Subproblema) e ao **Passo 3** (denominado Problema Mestre), as quais serão detalhadas a seguir.

Para tanto, admita-se o problema de otimização original descrito por meio função objetivo representada pela eq. (E-1) e pelo conjunto de restrições a ela associado.

$$\min_{\underline{x}, \underline{y}} z = \underline{a}^T \cdot \underline{x} + \underline{b}^T \cdot \underline{y} \quad (\text{E-1})$$

sujeito a:

$$\underline{A} \cdot \underline{x} + \underline{B} \cdot \underline{y} = \underline{r} \quad (\text{E-2})$$

$$\underline{C} \cdot \underline{x} + \underline{D} \cdot \underline{y} \leq \underline{s} \quad (\text{E-3})$$

$$\underline{x}_L \leq \underline{x} \leq \underline{x}_H \quad (\text{E-4})$$

$$\underline{y}_L \leq \underline{y} \leq \underline{y}_H \quad (\text{E-5})$$

Para os casos em que \underline{a} e \underline{A} são respectivamente um vetor nulo e uma matriz nula o problema de otimização acima pode ser reescrito como segue:

$$\min_{\underline{x}, \underline{y}} z = \underline{b}^T \cdot \underline{y}$$

sujeito a:

$$\underline{B} \cdot \underline{y} = \underline{r} \quad (\text{E-6})$$

$$\underline{C} \cdot \underline{x} + \underline{D} \cdot \underline{y} \leq \underline{s}$$

$$\underline{x}_L \leq \underline{x} \leq \underline{x}_H$$

$$\underline{y}_L \leq \underline{y} \leq \underline{y}_H$$

O **Passo 1** do algoritmo de decomposição de Benders aplicado a este problema resulta em:

$$\begin{aligned}
 \min_{\underline{y}} z &= \underline{b}^T \cdot \underline{y} \\
 \text{sujeito a:} \\
 \underline{B} \cdot \underline{y} &= \underline{r} \\
 \underline{C} \cdot \underline{x} + \underline{D} \cdot \underline{y} &\leq \underline{s} \\
 \underline{y}_L &\leq \underline{y} \leq \underline{y}_H
 \end{aligned} \tag{E-7}$$

Considerando que a variável \underline{y} só pode assumir valores não-negativos e que \underline{x} é a variável complicadora e que nesta etapa seus valores são fixados como $\underline{x}^{(v)}$, o problema de otimização representado pela eq. (E-7) pode ser reescrito como segue:

$$\begin{aligned}
 \min_{\underline{y}} z &= \underline{b}^T \cdot \underline{y} \\
 \text{sujeito a:} \\
 \underline{B} \cdot \underline{y} &= \underline{r} \\
 \underline{D} \cdot \underline{y} &\leq \underline{s} - \underline{C} \cdot \underline{x}^{(v)} \\
 \underline{0} &\leq \underline{y} \leq \underline{y}_H
 \end{aligned} \tag{E-8}$$

A partir da resolução do problema descrito na eq. (E-8) obtém-se, além de $\underline{y}^{(v)}$ (solução ótima \underline{y} na iteração v), o valor da variável dual $\underline{\lambda}$ nesta iteração ($\underline{\lambda}^{(v)}$), o qual é necessário para a resolução do **Passo 3** do algoritmo.

Caso o **Passo 1** não seja resolvido de forma convencional, como ocorre no presente trabalho, perde-se a capacidade de determinar a variável dual $\underline{\lambda}$, o que poderia inviabilizar a finalização do algoritmo.

Esta limitação, no entanto, pode ser contornada levando-se em consideração que a solução ótima para ambos os problemas é a mesma e conhecida. Assim, as condições de Karush-Kuhn-Tucker (KKT), descritas no Apêndice C, podem ser aplicadas para obter-se $\underline{\lambda}$.

Aplicando-se, portanto, as condições KKT a ambos os problemas, tem-se:

$$\text{Problema Original} \left\{ \begin{array}{l} \underline{\mu}_1^T \cdot \underline{C} + \underline{\lambda}^T = 0 \\ \underline{b}^T + \underline{\lambda}_1^T \cdot \underline{B} + \underline{\mu}_1^T \cdot \underline{D} - \underline{\mu}_2^T + \underline{\mu}_3^T = 0 \\ \underline{\mu}_1, \underline{\mu}_2, \underline{\mu}_3 \geq 0 \\ \underline{\lambda}, \underline{\lambda}_1 \text{ irrestritas} \end{array} \right.$$

$$\text{Problema Alternativo} \left\{ \begin{array}{l} \underline{b}^T + \underline{\lambda}_1^T \cdot \underline{B} + \underline{\mu}_1^T \cdot \underline{D} - \underline{\mu}_2^T + \underline{\mu}_3^T = 0 \\ \underline{\mu}_1, \underline{\mu}_2, \underline{\mu}_3 \geq 0 \\ \underline{\lambda}_1 \text{ irrestrita} \end{array} \right.$$

Com as variáveis duais fornecidas pelo problema alternativo, é possível calcular $\underline{\lambda} = -\underline{C}^T \cdot \underline{\mu}_1$.

Já no tocante ao **Passo 3** do algoritmo de decomposição de Benders aplicado ao mesmo problema de otimização descrito por meio função objetivo representada pela eq. (E-1) e pelo conjunto de restrições a ela associado tem-se que este resulta em:

$$\begin{aligned} \min_{\sigma} z &= \sigma \\ \text{sujeito a:} \\ \underline{b}^T \cdot \underline{y}^{(k)} + \left[\underline{\lambda}^{(k)} \right]^T \cdot \left[\underline{x} - \underline{x}^{(k)} \right] &\geq \sigma \quad k = 1, \dots, v-1 \\ 0 &\leq \underline{x} \leq \underline{x}_H \\ \sigma &\geq \sigma_L \end{aligned}$$

cuja solução ótima é dada claramente por:

$$\sigma_{OPT} = \max \left\{ \sigma_L, \min_{\underline{x}} \max_{k=1, \dots, v-1} \left\{ \underbrace{\underline{b}^T \underline{y}^{(k)} - \left[\underline{\lambda}^{(k)} \right]^T \cdot \underline{x}^{(k)}}_{\rho^{(k)}} + \left[\underline{\lambda}^{(k)} \right]^T \cdot \underline{x} \right\} \right\}$$

Sabe-se que o **MATLAB**[®], por exemplo, tem uma função, chamada “fminimax”, que resolve problemas minimax restritos, mas somente para os casos com variáveis reais, não sendo aplicável diretamente àqueles com variáveis binárias.

Para se empregar esta função em situações com variáveis binárias, generalizando a solução do problema, pode-se adotar a seguinte estratégia: introduzir um termo de penalidade como $\gamma \cdot \underline{x}^T \cdot (\underline{1} - \underline{x})$ e se γ for grande o suficiente, os valores binários podem ser forçados a \underline{x} .

F

Aplicação Encadeada do Método de Decomposição de Benders

O artifício empregado no presente trabalho de aplicação de forma encadeada do algoritmo de Benders é descrito detalhadamente a seguir. Para tanto, considera-se uma rede representada pelo grafo $G = (V, E)$. Define-se também nV como o número de vértices e nE como o número de circuitos e o problema de otimização MILP que se deseja resolver é descrito por:

$$\min_{\underline{y}_1, \underline{y}_2} \quad z = \underline{g}_2^T \cdot \underline{y}_2$$

sujeito a:

$$\underline{V}_2 \cdot \underline{y}_2 = \underline{r}_0$$

$$\underline{T}_1 \cdot \underline{y}_1 + \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 \quad (\text{F-1})$$

$$\underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H}$$

$$\underline{y}_1 \in \mathbb{B}^{nY_1}$$

$$\underline{y}_2 \in \mathbb{R}^{nY_2}$$

Salienta-se ainda que as dimensões das matrizes e vetores envolvidos são dadas por:

$$\dim(\underline{y}_2) = \dim(\underline{g}_2) \Rightarrow [1 + (nE + 2 \cdot nV) \cdot nV + nE] \times 1$$

$$\dim(\underline{y}_1) \Rightarrow [nE \cdot nV] \times 1$$

$$\dim(\underline{r}_0) \Rightarrow [(nV)^2] \times 1$$

$$\dim(\underline{s}_0) \Rightarrow [5 \cdot nE \cdot (nV - 1) + nE] \times 1$$

$$\dim(\underline{V}_2) \Rightarrow [(nV)^2] \times [1 + (nE + 2 \cdot nV) \cdot nV + nE]$$

$$\dim(\underline{T}_1) \Rightarrow [5 \cdot nE \cdot (nV - 1) + nE] \times [nE \cdot nV]$$

$$\dim(\underline{T}_2) \Rightarrow [5 \cdot nE \cdot (nV - 1) + nE] \times [1 + (nE + 2 \cdot nV) \cdot nV + nE]$$

Considerando-se que a matriz de restrições seja representada esquematicamente pela Figura F.1 e que os vetores \underline{y}_1 e \underline{y}_2 sejam formados, respectivamente, pelo agrupamento das variáveis binárias (\underline{u}) e reais (Z , \underline{x} , \underline{z} , \underline{r} e \underline{w}), resulta que \underline{y}_1 terá dimensão $nE \times nV$, enquanto, \underline{y}_2 será computacionalmente descrito por uma estrutura \underline{Y} com nY elementos da forma:

$$\underline{Y} \cdot Z \leftrightarrow \dim = 1 \text{ e valores no intervalo } [0, 1]$$

$$\underline{Y} \cdot \underline{x} \leftrightarrow \dim = nE \times nV \text{ e valores no intervalo } [0, \infty)$$

$$\underline{Y} \cdot \underline{z} \leftrightarrow \dim = nV \times nV \text{ e valores no intervalo } [0, \infty)$$

$$\underline{Y} \cdot \underline{r} \leftrightarrow \dim = nV \times nV \text{ e valores no intervalo } [1, \infty)$$

$$\underline{Y} \cdot \underline{w} \leftrightarrow \dim = nE \times 1 \text{ e valores no intervalo } [1, \kappa]$$

Observa-se, então, com base na representação esquemática apresentada na Figura F.1 que:

$$\underline{V}_2 \cdot \underline{y}_2 = \begin{bmatrix} A_1 \cdot [Y \cdot x_{.,1}] \\ \vdots \\ A_{nV} \cdot [Y \cdot x_{.,nV}] \end{bmatrix}$$

$$\underline{r}_0 = \begin{bmatrix} \underline{a}_1 \\ \vdots \\ \underline{a}_{nV} \end{bmatrix}$$

$$\underline{s}_0 = \begin{bmatrix} \underline{b}_1 \\ \underline{c}_1 \\ \vdots \\ \underline{b}_{nV} \\ \underline{c}_{nV} \\ \underline{0}_{nE} \end{bmatrix}$$

$$\underline{T}_1 \cdot \underline{y}_1 = \left[\begin{array}{c} D_1 \cdot u_{.,1} \\ E_1 \cdot u_{.,1} \\ \hline \vdots \\ D_{nV} \cdot u_{.,nV} \\ E_{nV} \cdot u_{.,nV} \\ \hline \underline{0}_{nE} \end{array} \right]$$

$$\underline{T}_2 \cdot \underline{y}_2 = \left[\begin{array}{c} B_1 \cdot [Y \cdot x_{.,1}] + C_1 \cdot [Y \cdot z_{.,1}] \\ F_1 \cdot [Y \cdot r_{.,1}] + H_1 \cdot [Y \cdot w] \\ \hline \vdots \\ B_{nV} \cdot [Y \cdot x_{.,nV}] + C_{nV} \cdot [Y \cdot z_{.,nV}] \\ F_{nV} \cdot [Y \cdot r_{.,nV}] + H_{nV} \cdot [Y \cdot w] \\ \hline K \cdot Z + G_1 \cdot [Y \cdot x_{.,1}] + \dots + G_{nV} \cdot [Y \cdot x_{.,nV}] \end{array} \right]$$

$$\underline{y}_{2_L} = \left[\begin{array}{c} 0 \\ \underline{0}_{nE} \\ \underline{0}_{nV} \\ \underline{1}_{nV} \\ \vdots \\ \underline{0}_{nE} \\ \underline{0}_{nV} \\ \underline{1}_{nV} \\ \underline{1}_{nE} \end{array} \right] \qquad \underline{y}_{2_H} = \left[\begin{array}{c} 1 \\ \underline{\infty}_{nE} \\ \underline{\infty}_{nV} \\ \underline{\infty}_{nV} \\ \vdots \\ \underline{\infty}_{nE} \\ \underline{\infty}_{nV} \\ \underline{\infty}_{nV} \\ \underline{\kappa}_{nE} \end{array} \right]$$

Variáveis Reais															Variáveis Binárias					
Z	<u>x</u> ,1	<u>z</u> ,1	<u>r</u> ,1	<u>x</u> ,2	<u>z</u> ,2	<u>r</u> ,2	...	<u>x</u> ,nV	<u>z</u> ,nV	<u>r</u> ,nV	<u>w</u>	<u>u</u> ,1	<u>u</u> ,2	...	<u>u</u> ,nV					
igualdades	A ₁			A ₂												=	a ₁			
			...													=	a ₂			
								A _{nV}										
																=	a _{nV}			
desigualdades	B ₁	C ₁	F ₁									D ₁	E ₁			≤	b ₁			
				B ₂	C ₂							H ₁		D ₂		≤	c ₁			
			b ₂			
						F ₂						H ₂	E ₂			≤	c ₂			
							≤	...			
								B _{nV}	C _{nV}							D _{nV}	≤	b _{nV}		
										F _{nV}		H _{nV}			E _{nV}	...	c _{nV}			
	K	G ₁		G ₂				G _{nV}								≤	0			
																≤	0			
																		
																≤	0			

Figura F.1 – Representação esquemática da matriz de restrições

Fonte: Elaborada pela autora

Neste contexto, considerando que i varia de 1 a nV , as dimensões das matrizes envolvidas e suas respectivas composições são dadas por:

$\dim(A_i) \Rightarrow nV \times nE$ e qualquer elemento da matriz \underline{A} pertence ao intervalo $\{-1, 0, +1\}$.

$\dim(a_i) \Rightarrow nV \times 1$ e os vetores a são não-negativos.

$\dim(B_i) \Rightarrow 3 \cdot \#\delta(i) \times nE$ e qualquer elemento da matriz \underline{B} pertence ao intervalo $\{-1, 0, +1\}$.

$\dim(C_i) \Rightarrow 3 \cdot \#\delta(i) \times nV$ e qualquer elemento da matriz \underline{C} pertence ao intervalo $\{-1, 0, +1\}$.

$\dim(D_i) \Rightarrow 3 \cdot \#\delta(i) \times nE$ e qualquer elemento da matriz \underline{D} é real.

$\dim(b_i) \Rightarrow 3 \cdot \#\delta(i) \times 1$ e os vetores b são não-negativos.

$\dim(E_i) \Rightarrow 2 \cdot \#\delta(i) \times nE$ e qualquer elemento da matriz \underline{E} pertence ao intervalo $\{-1, 0, +M\}$.

$\dim(F_i) \Rightarrow 2 \cdot \#\delta(i) \times nV$ e qualquer elemento da matriz \underline{F} pertence ao intervalo $\{-1, 0, +1\}$.

$\dim(H_i) \Rightarrow 2 \cdot \#\delta(i) \times nE$ e qualquer elemento da matriz \underline{H} pertence ao intervalo $\{-1, 0, +1\}$.

$\dim(c_i) \Rightarrow 2 \cdot \#\delta(i) \times 1$ e os vetores c são reais.

$\dim(G_i) \Rightarrow nE \times nE$ e qualquer elemento da matriz \underline{G} pertence ao intervalo $\{0, +1\}$.

$\dim(\underline{K}) \Rightarrow nE \times 1$ e o vetor \underline{K} é negativo.

Aplicando-se, então, o algoritmo de Benders na resolução do problema descrito na eq. (F-1) resulta nos passos descritos a seguir. Salienta-se que para facilitar o entendimento denomina-se como **BENDERS-0** esta primeira aplicação do método de decomposição de Benders ao problema original.

Passo 0: Inicialização

$$v = 1$$

$$\underline{y}_1^{(v)} = \underline{0}$$

$$\alpha^{(v)} = \alpha_L$$

Passo 1: Solução do Subproblema

$$\left. \begin{array}{l} \min_{\underline{y}_2} z = \underline{g}_2^T \cdot \underline{y}_2 \\ \text{sujeito a:} \\ \underline{V}_2 \cdot \underline{y}_2 = \underline{r}_0 \\ \underline{T}_1 \cdot \underline{y}_1 + \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 \\ \underline{y}_1 = \underline{y}_1^{(v)} : \underline{\lambda} \\ \underline{y}_1 \in \mathbb{B}^{nY_1} \\ \underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H} \end{array} \right\} \Leftrightarrow \left. \begin{array}{l} \min_{\underline{y}_2} z = \underline{g}_2^T \cdot \underline{y}_2 \\ \text{sujeito a:} \\ \underline{V}_2 \cdot \underline{y}_2 = \underline{r}_0 \\ \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0 - \underline{T}_1 \cdot \underline{y}_1^{(v)} \\ \underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H} \end{array} \right\} \Rightarrow \underline{y}_2^{(v)}, \underline{\lambda}^{(v)} \quad (\text{F-2})$$

Passo 2: Verificação de Convergência

Se $\underline{g}_2^T \cdot \underline{y}_2^{(v)} - \alpha^{(v)} < \iota$ então

PARE. Solução ótima encontrada.

Caso Contrário

$$v = v + 1$$

Fim

Passo 3: Problema Mestre

$$\left. \begin{array}{l} \min_{\underline{y}_1, \alpha} z = \alpha \\ \text{sujeito a:} \\ \alpha \geq \underline{g}_2^T \cdot \underline{y}_2^{(k)} + \left[\underline{\lambda}^{(k)} \right]^T \cdot \left[\underline{y}_1 - \underline{y}_1^{(k)} \right] \quad k = 1, \dots, v-1 \\ \underline{y}_1 \in \mathbb{B}^{n_{Y_1}} \\ \alpha \geq \alpha_L \end{array} \right\} \Rightarrow \underline{y}_1^{(v)}, \alpha^{(v)}$$

Vá para o Passo 1

Contudo a resolução do subproblema representado pela eq. (F-2) não é trivial, podendo ser extremamente onerosa computacionalmente. Visando a contornar esta dificuldade decide-se reescrevê-lo de forma a torná-lo visivelmente decomponível, viabilizando a aplicação novamente de mecanismos de decomposição para a sua resolução.

Com este intuito substitui-se o termo constante $\left(\underline{s}_0 - \underline{T}_1 \cdot \underline{y}_1^{(v)} \right)$ na eq. (F-2) por $\underline{s}_0^{(v)}$, o que resulta em:

$$\begin{array}{l} \min_{\underline{y}_2} z = \underline{g}_2^T \cdot \underline{y}_2 \\ \text{sujeito a:} \\ \underline{V}_2 \cdot \underline{y}_2 = r_0 \\ \underline{T}_2 \cdot \underline{y}_2 \leq \underline{s}_0^{(v)} \\ \underline{y}_{2_L} \leq \underline{y}_2 \leq \underline{y}_{2_H} \end{array} \quad (\text{F-3})$$

Além disso, fazendo $\underline{y}_2 = \begin{bmatrix} \underline{h} \\ \underline{w} \end{bmatrix}$ e considerando \underline{w} a variável complicadora, tem-se que o vetor \underline{w} será descrito por $\underline{Y} \cdot \underline{w}$ enquanto o vetor \underline{h} será descrito por uma estrutura \underline{H} como abaixo:

$$\underline{H} \cdot \underline{Z} = \underline{Y} \cdot \underline{Z}; \quad \underline{H} \cdot \underline{x} = \underline{Y} \cdot \underline{x}; \quad \underline{H} \cdot \underline{z} = \underline{Y} \cdot \underline{z} \quad \text{e} \quad \underline{H} \cdot \underline{r} = \underline{Y} \cdot \underline{r}.$$

Reescrevendo, então, o problema de otimização representado pela eq. (F-3) em função de \underline{h} e \underline{w} resulta em:

$$\begin{aligned} \min_{\underline{h}, \underline{w}} z &= \underline{g}_{2h}^T \cdot \underline{h} \\ \text{sujeito a:} \\ \underline{V}_{2h} \cdot \underline{h} &= \underline{r}_0 \\ \underline{T}_{2h} \cdot \underline{h} + \underline{T}_{2w} \cdot \underline{w} &\leq \underline{s}_0^{(v)} \\ \underline{h}_L &\leq \underline{h} \leq \underline{h}_H \\ \underline{w}_L &\leq \underline{w} \leq \underline{w}_H \end{aligned} \tag{F-4}$$

Ressalta-se que na formulação acima, os equivalentes ao vetor \underline{g}_{2w} e à matriz \underline{V}_{2w} são nulos, tendo sido, por conseguinte, omitidos.

Observa-se, ainda, com base na representação esquemática apresentada na Figura F.1 que:

$$\underline{g}_{2h}^T \cdot \underline{h} = \underline{H} \cdot \underline{Z}$$

$$\underline{V}_{2h} \cdot \underline{h} = \underline{V}_2 \cdot \underline{y}_2$$

$$\underline{T}_{2w} \cdot \underline{w} = \begin{bmatrix} \underline{0}_{3 \cdot \# \delta(1)} \\ H_1 \cdot [Y \cdot w] \\ \vdots \\ \underline{0}_{3 \cdot \# \delta(nV)} \\ H_{nV} \cdot [Y \cdot w] \\ \underline{0}_{nE} \end{bmatrix}$$

$$\underline{T}_{2h} \cdot \underline{h} = \begin{bmatrix} B_1 \cdot [Y \cdot x_{.,1}] + C_1 \cdot [Y \cdot z_{.,1}] \\ F_1 \cdot [Y \cdot r_{.,1}] \\ \vdots \\ B_{nV} \cdot [Y \cdot x_{.,nV}] + C_{nV} \cdot [Y \cdot z_{.,nV}] \\ F_{nV} \cdot [Y \cdot r_{.,nV}] \\ K \cdot Z + G_1 \cdot [Y \cdot x_{.,1}] + \dots + G_{nV} \cdot [Y \cdot x_{.,nV}] \end{bmatrix}$$

$$\underline{s}_0^{(v)} = \underline{s}_0 - \underline{T}_1 \cdot \underline{y}_1^{(v)} = \begin{bmatrix} \underline{b}_1 - D_1 \cdot u_{.,1} \\ \underline{c}_1 - E_1 \cdot u_{.,1} \\ \vdots \\ \underline{b}_{nV} - D_{nV} \cdot u_{.,nV} \\ \underline{c}_{nV} - E_{nV} \cdot u_{.,nV} \\ \underline{0}_{nE} \end{bmatrix}$$

$$\underline{h}_L = \begin{bmatrix} 0 \\ \underline{0}_{nE} \\ \underline{0}_{nV} \\ \underline{1}_{nV} \\ \vdots \\ \underline{0}_{nE} \\ \underline{0}_{nV} \\ \underline{1}_{nV} \end{bmatrix} \qquad \underline{h}_H = \begin{bmatrix} 1 \\ \underline{\infty}_{nE} \\ \underline{\infty}_{nV} \\ \underline{\infty}_{nV} \\ \vdots \\ \underline{\infty}_{nE} \\ \underline{\infty}_{nV} \\ \underline{\infty}_{nV} \end{bmatrix}$$

$$\underline{w}_L = \underline{1}_{nE}$$

$$\underline{w}_H = \underline{\kappa}_{nE}$$

As dimensões das matrizes envolvidas são dadas por:

$$\dim(\underline{V}_{2h}) = [(nV)^2] \times [1 + (nE + 2 \cdot nV) \cdot nV]$$

$$\dim(\underline{T}_{2h}) = [5 \cdot nE \cdot (nV - 1) + nE] \times [1 + (nE + 2 \cdot nV) \cdot nV]$$

$$\dim(\underline{T}_{2w}) = [5 \cdot nE \cdot (nV - 1) + nE] \times nE$$

Assim, o problema de otimização descrito pela eq. (F-4) pode, nitidamente, ser resolvido pelo método de decomposição de Benders e a esta segunda aplicação do algoritmo de Benders denomina-se **BENDERS-1**. Seu detalhamento é descrito a seguir.

Passo 0: Inicialização

$$\chi = 1$$

$$\underline{w}^{(\chi)} = \underline{0}$$

$$\beta^{(\chi)} = \beta_L$$

Passo 1: Solução do Subproblema

$$\left. \begin{array}{l} \min_{\underline{h}, \underline{w}} z = \underline{g}_{2h}^T \cdot \underline{h} \\ \text{sujeito a:} \\ \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\ \underline{T}_{2h} \cdot \underline{h} + \underline{T}_{2w} \cdot \underline{w} \leq \underline{s}_0^{(v)} \\ \underline{w} = \underline{w}^{(\chi)} : \underline{v} \\ \underline{h}_L \leq \underline{h} \leq \underline{h}_H \\ \underline{w}_L \leq \underline{w} \leq \underline{w}_H \end{array} \right\} \Leftrightarrow \left. \begin{array}{l} \min_{\underline{h}, \underline{w}} z = \underline{g}_{2h}^T \cdot \underline{h} \\ \text{sujeito a:} \\ \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\ \underline{T}_{2h} \cdot \underline{h} \leq \underline{s}_0^{(v)} - \underline{T}_{2w} \cdot \underline{w}^{(\chi)} \\ \underline{h}_L \leq \underline{h} \leq \underline{h}_H \end{array} \right\} \Rightarrow \underline{h}^{(\chi)}, \underline{v}^{(\chi)} \quad (\text{F-5})$$

Passo 2: Verificação de Convergência

$$\text{Se } \underline{g}_{2h}^T \cdot \underline{h}^{(\chi)} - \beta^{(\chi)} < \iota \quad \text{então}$$

PARE. Solução ótima encontrada.

Caso Contrário

$$\chi = \chi + 1$$

Fim

Passo 3: Problema Mestre

$$\left. \begin{array}{l} \min_{\underline{w}, \beta} z = \beta \\ \text{sujeito a:} \\ \beta \geq \underline{g}_{2h}^T \cdot \underline{h}^{(k)} + [\underline{v}^{(k)}]^T \cdot [\underline{w} - \underline{w}^{(k)}] \quad k = 1, \dots, \chi - 1 \\ \underline{w}_L \leq \underline{w} \leq \underline{w}_H \\ \beta \geq \beta_L \end{array} \right\} \Rightarrow \underline{w}^{(\chi)}, \beta^{(\chi)}$$

Vá para o Passo 1

Analogamente ao ocorrido no **BENDERS-0**, o subproblema representado pela eq. (F-5) não é de fácil resolução, e busca-se reescrevê-lo de tão forma a tornar possível a aplicação mais uma vez de mecanismos de decomposição para a sua resolução.

Desta forma, substituindo-se o termo constante $\left(\underline{s}_0^{(v)} - \underline{T}_{2w} \cdot \underline{w}^{(\chi)}\right)$ na eq. (F-5) por $\underline{s}_0^{(v, \chi)}$ resulta em:

$$\begin{aligned} \min_{\underline{h}} z &= \underline{g}_{2h}^T \cdot \underline{h} \\ \text{sujeito a:} \\ \underline{V}_{2h} \cdot \underline{h} &= r_0 \\ \underline{T}_{2h} \cdot \underline{h} &\leq \underline{s}_0^{(v, \chi)} \\ \underline{h}_L &\leq \underline{h} \leq \underline{h}_H \end{aligned} \tag{F-6}$$

Neste contexto, destaca-se que:

$$\begin{aligned} \underline{h} \leq \underline{h}_H &\Rightarrow \underline{h}_L \leq 1 \\ \underline{h}_L \leq \underline{h} &\Rightarrow \begin{cases} \geq 1 \\ \underline{h}_{not\ related} \geq 0 \end{cases}, \end{aligned}$$

A estrutura assumida pelo problema de otimização representado pela eq. (F-6), no entanto, não permite a aplicação do método de decomposição de

Benders, por ser composto por restrições complicadoras e não por variáveis complicadoras.

Poder-se-ia resolvê-lo empregando o método de decomposição de Dantzig-Wolfe, mas opta-se por aplicar Benders ao seu dual.

Assim, para facilitar o cálculo do dual, reescreve-se a eq. (F-6) na forma padrão empregando as variáveis de folga $\underline{\rho}$, $\underline{\phi}$ e $\underline{\gamma}$, o que resulta em:

$$\left. \begin{array}{l} \min_{\underline{h}} z = \underline{g}_{2h}^T \cdot \underline{h} \\ \text{sujeito a:} \\ \underline{V}_{2h} \cdot \underline{h} = \underline{r}_0 \\ \underline{T}_{2h} \cdot \underline{h} + I \cdot \underline{\rho} = \underline{s}_0^{(v,\chi)} \\ \underline{h}_L + \underline{\phi} = 1 \\ \underline{h}_{related} - \underline{\gamma} = 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \min_{\underline{h}, \underline{\rho}, \underline{\phi}, \underline{\gamma}} z = \underline{g}_{2h}^T \cdot \underline{h} + \underline{0}^T \cdot \underline{\rho} + \underline{0}^T \cdot \underline{\phi} + \underline{0}^T \cdot \underline{\gamma} \\ \text{sujeito a:} \\ \begin{bmatrix} \underline{V}_{2h} & 0 & 0 & 0 \\ \underline{T}_{2h} & I & 0 & 0 \\ \underline{c}_h^T & 0 & 1 & 0 \\ E_h & 0 & 0 & -I \end{bmatrix} \cdot \begin{bmatrix} \underline{h} \\ \underline{\rho} \\ \underline{\phi} \\ \underline{\gamma} \end{bmatrix} = \begin{bmatrix} \underline{r}_0 \\ \underline{s}_0^{(v,\chi)} \\ 1 \\ \underline{1} \end{bmatrix} \\ \begin{bmatrix} \underline{h} \\ \underline{\rho} \\ \underline{\phi} \\ \underline{\gamma} \end{bmatrix} \geq \underline{0} \end{array} \right\}$$

Onde:

$$\underline{c}_h^T = \begin{bmatrix} 1 & \underline{0}_{[1, nV \cdot (nE + 2 \cdot nV)]} \end{bmatrix}$$

$$E_h =$$

$$\left[\begin{array}{c|c|c|c|c|c|c} 0_{[(nV)^2, 1]} & 0_{[(nV)^2, nE]} & 0_{[(nV)^2, nV]} & U_1 & \dots & 0_{[(nV)^2, nE]} & 0_{[(nV)^2, nV]} & U_{nV} \end{array} \right]$$

$$U_k = \begin{bmatrix} 0_{nV} \\ \vdots \\ 0_{nV} \\ I_{nV} \\ 0_{nV} \\ \vdots \\ 0_{nV} \end{bmatrix} \rightarrow \text{posição } k, \quad k = 1, \dots, nV$$

A forma dual é dada, então, por:

$$\max_{\underline{\eta}, \underline{\theta}, \underline{\xi}, \underline{\varphi}} z = [\underline{r}_0]^T \cdot \underline{\eta} + \left[\underline{s}_0^{(v, \chi)} \right]^T \cdot \underline{\theta} + \xi + \underline{1}^T \cdot \underline{\varphi}$$

sujeito a:

$$\begin{bmatrix} \underline{V}_{2h}^T & \underline{T}_{2h}^T & \underline{c}_h & \underline{E}_h^T \\ 0 & I & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -I \end{bmatrix} \cdot \begin{bmatrix} \underline{\eta} \\ \underline{\theta} \\ \underline{\xi} \\ \underline{\varphi} \end{bmatrix} \leq \begin{bmatrix} \underline{g}_{2h} \\ \underline{0} \\ 0 \\ \underline{0} \end{bmatrix} \quad (\text{F-7})$$

Salienta-se que as dimensões de $\underline{\eta}$ e $\underline{\varphi}$ são dadas por $[(nV)^2, 1]$. E as dimensões de $\underline{\theta}$ e $\underline{\xi}$ são dadas, respectivamente por: $[5 \cdot nE \cdot (nV - 1) + nE, 1]$ e 1.

Reescrevendo, então, a eq. (F-7) para converter o problema de maximização em um de minimização, tem-se:

$$\min_{\underline{\eta}, \underline{\theta}, \underline{\xi}, \underline{\varphi}} z = -[\underline{r}_0]^T \cdot \underline{\eta} + \left[\underline{s}_0^{(v, \chi)} \right]^T \cdot \underline{\theta} + \xi - \underline{1}^T \cdot \underline{\varphi}$$

sujeito a:

$$\underline{V}_{2h}^T \cdot \underline{\eta} - \underline{T}_{2h}^T \cdot \underline{\theta} - \underline{c}_h \cdot \underline{\xi} + \underline{E}_h^T \cdot \underline{\varphi} \leq \underline{g}_{2h} \quad (\text{F-8})$$

$\underline{\eta}$ irrestrito

$\underline{\theta}, \underline{\xi}, \underline{\varphi} \geq \underline{0}$

Na eq. (F-8) reescrevendo $\underline{\theta}$ em função de $\underline{\theta}_1$ e $\underline{\theta}_2$, onde $\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix}$, e as dimensões de $\underline{\theta}_1$ e $\underline{\theta}_2$ são dadas, respectivamente, por: $[5 \cdot nE \cdot (nV - 1), 1]$ e $[nE, 1]$, resulta em:

$$\begin{aligned} \min_{\underline{\eta}, \underline{\theta}, \xi, \underline{\varphi}} z &= -[\underline{r}_0]^T \cdot \underline{\eta} + [\underline{s}_1^{(v, \chi)}]^T \cdot \underline{\theta}_1 + \underbrace{[\underline{s}_2^{(v, \chi)}]^T}_{\underline{0}} \cdot \underline{\theta}_2 + \xi - \underline{1}^T \cdot \underline{\varphi} \\ \text{sujeito a:} \\ \underline{V}_{2h}^T \cdot \underline{\eta} - \underline{T}_{2h_1}^T \cdot \underline{\theta}_1 - \underline{T}_{2h_2}^T \cdot \underline{\theta}_2 - \underline{c}_h \cdot \xi + \underline{E}_h^T \cdot \underline{\varphi} &\leq \underline{g}_{2h} \\ \underline{\eta} &\text{ irrestrito} \\ \underline{\theta}_1, \underline{\theta}_2, \xi, \underline{\varphi} &\geq \underline{0} \end{aligned} \tag{F-9}$$

Rearrmando a eq. (F-9) tem-se:

$$\begin{aligned} \min_{\underline{\eta}, \underline{\theta}, \xi, \underline{\varphi}} z &= \underbrace{\begin{bmatrix} -[\underline{r}_0]^T & [\underline{s}_1^{(v, \chi)}]^T \end{bmatrix}}_{\hat{\underline{a}}^T} \cdot \underbrace{\begin{bmatrix} \underline{\eta} \\ \underline{\theta}_1 \end{bmatrix}}_{\hat{\underline{p}}} + \underbrace{\begin{bmatrix} \underline{0}^T & 1 & -\underline{1}^T \end{bmatrix}}_{\hat{\underline{b}}^T} \cdot \underbrace{\begin{bmatrix} \underline{\theta}_2 \\ \xi \\ \underline{\varphi} \end{bmatrix}}_{\hat{\underline{q}}} \\ \text{sujeito a:} \\ \underbrace{\begin{bmatrix} \underline{V}_{2h}^T & -\underline{T}_{2h_1}^T \end{bmatrix}}_{\hat{\underline{A}}} \cdot \underbrace{\begin{bmatrix} \underline{\eta} \\ \underline{\theta}_1 \end{bmatrix}}_{\hat{\underline{p}}} + \underbrace{\begin{bmatrix} -\underline{T}_{2h_2}^T & -\underline{c}_h & \underline{E}_h^T \end{bmatrix}}_{\hat{\underline{B}}} \cdot \underbrace{\begin{bmatrix} \underline{\theta}_2 \\ \xi \\ \underline{\varphi} \end{bmatrix}}_{\hat{\underline{q}}} &\leq \underline{g}_{2h} \\ \underbrace{\begin{bmatrix} -\infty \\ 0 \end{bmatrix}}_{\hat{\underline{p}}_L} &\leq \underbrace{\begin{bmatrix} \underline{\eta} \\ \underline{\theta}_1 \end{bmatrix}}_{\hat{\underline{p}}} \\ \underbrace{\begin{bmatrix} \underline{\theta}_2 \\ \xi \\ \underline{\varphi} \end{bmatrix}}_{\hat{\underline{q}}} &\geq \underline{0} \end{aligned} \tag{F-10}$$

Onde:

$$\underline{s}_1^{(v,\chi)} = \begin{bmatrix} \underline{b}_1 - D_1 \cdot u_{.,1} \\ \underline{c}_1 - E_1 \cdot u_{.,1} - H_1 \cdot [Y \cdot w] \\ \vdots \\ \underline{b}_{nV} - D_{nV} \cdot u_{.,nV} \\ \underline{c}_{nV} - E_{nV} \cdot u_{.,nV} - H_{nV} \cdot [Y \cdot w] \end{bmatrix}$$

Reescrevendo, então, a eq. (F-10) empregando as variáveis auxiliares criadas: $\underline{\hat{p}}$, $\underline{\hat{q}}$, $\underline{\hat{a}}$, $\underline{\hat{b}}$, $\underline{\hat{A}}$ e $\underline{\hat{B}}$, resulta em:

$$\min_{\underline{\hat{p}}, \underline{\hat{q}}} z = [\underline{\hat{a}}]^T \cdot \underline{\hat{p}} + [\underline{\hat{b}}]^T \cdot \underline{\hat{q}}$$

sujeito a:

$$\underline{\hat{A}} \cdot \underline{\hat{p}} + \underline{\hat{B}} \cdot \underline{\hat{q}} \leq \underline{g}_{2h} \quad (\text{F-11})$$

$$\underline{\hat{p}} \geq \underline{\hat{p}}_L$$

$$\underline{\hat{q}} \geq \underline{0}$$

Salienta-se que as dimensões de $\underline{\hat{a}}$, $\underline{\hat{b}}$, $\underline{\hat{p}}$, $\underline{\hat{q}}$, $\underline{\hat{A}}$ e $\underline{\hat{B}}$ são dadas por:

$$\dim(\underline{\hat{a}}) = \dim(\underline{\hat{p}}) = (nV)^2 + 5 \cdot nE \cdot (nV - 1)$$

$$\dim(\underline{\hat{b}}) = \dim(\underline{\hat{q}}) = nE + 1 + (nV)^2$$

$$\dim(\underline{\hat{A}}) = \left([1 + (nE + 2 \cdot nV) \cdot nV], [(nV)^2 + 5 \cdot nE \cdot (nV - 1)] \right)$$

$$\dim(\underline{\hat{B}}) = \left([1 + (nE + 2 \cdot nV) \cdot nV], [nE + 1 + (nV)^2] \right)$$

Observa-se ainda que:

$$\underline{\varphi} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{(nV)^2} \end{bmatrix}$$

$$\hat{\underline{A}} = \begin{bmatrix} 0_{1,nV} & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ A_1^T & \dots & 0 & -B_1^T & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & -C_1^T & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & -F_1^T & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & A_{nV}^T & 0 & 0 & \dots & -B_{nV}^T & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & -C_{nV}^T & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & -F_{nV}^T \end{bmatrix}$$

$$\underline{T}_{2h_2}^T = \begin{bmatrix} K^T \\ G_1^T \\ 0 \\ 0 \\ \dots \\ G_{nV}^T \\ 0 \\ 0 \end{bmatrix}$$

$$\hat{\underline{B}} = \begin{bmatrix} -K^T & 1 & 0_{1,(nV)^2} \\ -G_1^T & 0 & 0_{nE,(nV)^2} \\ 0_{nV,nE} & 0 & 0_{nV,(nV)^2} \\ 0_{nV,nE} & 0 & U_1^T \\ \vdots & \vdots & \vdots \\ -G_{nV}^T & 0 & 0_{nE,(nV)^2} \\ 0_{nV,nE} & 0 & 0_{nV,(nV)^2} \\ 0_{nV,nE} & 0 & U_{nV}^T \end{bmatrix}$$

$$\widehat{\underline{p}} = \begin{bmatrix} \widehat{\underline{p}}_{a,1} \\ \dots \\ \widehat{\underline{p}}_{a,nV} \\ \widehat{\underline{p}}_{b,1} \\ \widehat{\underline{p}}_{c,1} \\ \dots \\ \widehat{\underline{p}}_{b,nV} \\ \widehat{\underline{p}}_{c,nV} \end{bmatrix} \quad \widehat{\underline{a}} = \begin{bmatrix} \widehat{\underline{a}}_1 \\ \dots \\ \widehat{\underline{a}}_{nV} \\ \widehat{\underline{b}}_1 \\ \widehat{\underline{c}}_1 \\ \dots \\ \widehat{\underline{b}}_{nV} \\ \widehat{\underline{c}}_{nV} \end{bmatrix}$$

Onde:

$$\widehat{\underline{a}}_k = \underline{a}_k$$

$$\widehat{\underline{b}}_k = \underline{b}_k - D_k \cdot u_{.,k}$$

$$\widehat{\underline{c}}_k = \underline{c}_k - E_k \cdot u_{.,k} - H_k \cdot [Y \cdot w]$$

Destaca-se que as dimensões de $\widehat{\underline{p}}_{a,k}$, $\widehat{\underline{p}}_{b,k}$ e $\widehat{\underline{p}}_{c,k}$ são dadas por:

$$\dim(\widehat{\underline{p}}_{a,k}) = \dim(\widehat{\underline{a}}_k) = (nV, 1) \text{ para todo } k$$

$$\dim(\widehat{\underline{p}}_{b,k}) = \dim(\widehat{\underline{b}}_k) = (3 \cdot \#\delta(k), 1) \text{ para todo } k$$

$$\dim(\widehat{\underline{p}}_{c,k}) = \dim(\widehat{\underline{c}}_k) = (2 \cdot \#\delta(k), 1) \text{ para todo } k$$

Assim, finalmente, o problema de otimização descrito pela eq. (F-11) permite a aplicação do método de decomposição de Benders e a esta terceira aplicação do algoritmo de Benders denomina-se **BENDERS-2**. Seu detalhamento é descrito a seguir.

Passo 0: Inicialização

$$\varepsilon = 1$$

$$\widehat{\underline{q}}_k^{(\varepsilon)} = \begin{cases} 0 & \text{se } \widehat{b}_k \geq 0 \\ \widehat{q}_{max} & \text{se } \widehat{b}_k < 0 \end{cases}$$

$$\gamma^{(\varepsilon)} = \gamma_L$$

Cabe ressaltar que $\widehat{\underline{q}}_k^{(\varepsilon)}$ tem que ser escolhido de modo que:

$$[-K^T \quad 1 \quad 0_{1,(nV)^2}] \cdot \widehat{\underline{q}}_k^{(\varepsilon)} \leq 1$$

Passo 1: Solução do Subproblema

$$\left. \begin{array}{l} \min_{\underline{\hat{p}}} z = [\underline{\hat{a}}]^T \cdot \underline{\hat{p}} \\ \text{sujeito a:} \\ \underline{\hat{A}} \cdot \underline{\hat{p}} \leq \underline{g}_{2h}^{(\varepsilon)} \\ \underline{\hat{q}} = \underline{\hat{q}}^{(\varepsilon)} : \underline{\zeta} \\ \underline{\hat{p}} \geq \underline{\hat{p}}_L \end{array} \right\} \Rightarrow \underline{\hat{p}}^{(\varepsilon)}, \underline{\zeta}^{(\varepsilon)} \quad (\text{F-12})$$

Onde: $\underline{g}_{2h}^{(\varepsilon)} = \underline{g}_{2h} - \underline{\hat{B}} \cdot \underline{\hat{q}}^{(\varepsilon)}$

Passo 2: Verificação de Convergência

Se $[\underline{\hat{a}}]^T \cdot \underline{\hat{p}}^{(\varepsilon)} - \gamma^{(\varepsilon)} < \iota$ então
PARE. Solução ótima encontrada.

Caso Contrário

$$\varepsilon = \varepsilon + 1$$

Fim

Passo 3: Problema Mestre

$$\left. \begin{array}{l} \min_{\underline{\hat{q}}, \gamma} z = \gamma + [\underline{\hat{b}}]^T \cdot \underline{\hat{q}} \\ \text{sujeito a:} \\ \gamma \geq [\underline{\hat{a}}]^T \cdot \underline{\hat{p}}^{(k)} + [\underline{\zeta}^{(k)}]^T \cdot [\underline{\hat{q}} - \underline{\hat{q}}^{(k)}] \quad k = 1, \dots, \varepsilon - 1 \\ \underline{\hat{q}} \geq \underline{0} \\ \gamma \geq \gamma_L \end{array} \right\} \Rightarrow \underline{\hat{q}}^{(\varepsilon)}, \gamma^{(\varepsilon)}$$

Vá para o Passo 1

Neste caso, diferentemente do ocorrido na aplicação do **BENDERS-0** e do **BENDERS-1**, o subproblema representado pela eq. (F-12) pode ser resolvido diretamente. A esta resolução direta denomina-se **OPTIM-0** e seu detalhamento matemático será descrito a seguir.

Ressalta-se que:

$$\underline{g}_{2h}^{(\varepsilon)} = \begin{bmatrix} 1 \\ \underline{0} \\ \underline{0} \\ \underline{0} \\ \vdots \\ \underline{0} \\ \underline{0} \\ \underline{0} \end{bmatrix} - \begin{bmatrix} -K^T & 1 & 0_{1,(nV)^2} \\ -G_1^T & 0 & 0_{nE,(nV)^2} \\ 0_{nV,nE} & 0 & 0_{nV,(nV)^2} \\ 0_{nV,nE} & 0 & U_1^T \\ \vdots & \vdots & \vdots \\ -G_{nV}^T & 0 & 0_{nE,(nV)^2} \\ 0_{nV,nE} & 0 & 0_{nV,(nV)^2} \\ 0_{nV,nE} & 0 & U_{nV}^T \end{bmatrix} \cdot \begin{bmatrix} \underline{\theta}_2 \\ \underline{\xi} \\ \underline{\varphi} \end{bmatrix} = \begin{bmatrix} 1 + K^T \cdot \underline{\theta}_2 - \underline{\xi} \\ G_1^T \cdot \underline{\theta}_2 \\ \underline{0} \\ -U_1^T \cdot \underline{\varphi} \\ \vdots \\ G_{nV}^T \cdot \underline{\theta}_2 \\ \underline{0} \\ -U_{nV}^T \cdot \underline{\varphi} \end{bmatrix}$$

Desta forma, o subproblema representado pela eq. (F-12) transforma-se em nV problemas da forma:

$$\min_{\underline{\hat{p}}} z = \begin{bmatrix} \underline{\hat{a}}_k^T & \underline{\hat{b}}_k^T & \underline{\hat{c}}_k^T \end{bmatrix} \cdot \begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \\ \underline{\hat{p}}_{c,k} \end{bmatrix}$$

sujeito a:

$$\begin{bmatrix} A_k^T & -B_k^T & 0 \\ 0 & -C_k^T & 0 \\ 0 & 0 & -F_k^T \end{bmatrix} \cdot \begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \\ \underline{\hat{p}}_{c,k} \end{bmatrix} \leq \begin{bmatrix} G_k^T \cdot \underline{\theta}_2 \\ \underline{0} \\ -U_k^T \cdot \underline{\varphi} \end{bmatrix} \quad (\text{F-13})$$

$$\begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \\ \underline{\hat{p}}_{c,k} \end{bmatrix} \geq \begin{bmatrix} -\underline{\infty} \\ \underline{0} \\ \underline{0} \end{bmatrix}$$

Os problemas representados pela eq. (F-13), por sua vez, podem ser fracionados em outros dois problemas.

O primeiro representado por:

$$\begin{aligned}
 \min_{\underline{\hat{p}}} z &= \underline{\hat{c}}_k^T \cdot \underline{\hat{p}}_{c,k} \\
 \text{sujeito a:} \\
 \underline{F}_k^T \cdot \underline{\hat{p}}_{c,k} &\geq U_k^T \cdot \underline{\varphi} \\
 \underline{\hat{p}}_{c,k} &\geq \underline{0}
 \end{aligned} \tag{F-14}$$

E o segundo representado por:

$$\begin{aligned}
 \min_{\underline{\hat{p}}} z &= \begin{bmatrix} \underline{\hat{a}}_k^T & \underline{\hat{b}}_k^T \end{bmatrix} \cdot \begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \end{bmatrix} \\
 \text{sujeito a:} \\
 \begin{bmatrix} A_k^T & -B_k^T \\ 0 & -C_k^T \end{bmatrix} \cdot \begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \end{bmatrix} &\leq \begin{bmatrix} G_k^T \cdot \underline{\theta}_2 \\ \underline{0} \end{bmatrix} \\
 \begin{bmatrix} \underline{\hat{p}}_{a,k} \\ \underline{\hat{p}}_{b,k} \end{bmatrix} &\geq \begin{bmatrix} -\underline{\infty} \\ \underline{0} \end{bmatrix}
 \end{aligned} \tag{F-15}$$

Com base no exposto, fica evidente, portanto, que o emprego do método de decomposição de Benders de forma encadeada permite obter de maneira simples e eficiente a solução para o **Passo 1** do algoritmo.

No entanto, apesar desta visível vantagem, com o encadeamento perde-se a capacidade de determinar a variável dual associada, representada por $\underline{\lambda}$ no **BENDERS-0**, por \underline{v} no **BENDERS-1** e por $\underline{\zeta}$ no **BENDERS-2**. Uma vez que tal variável é essencial para a resolução do **Passo 3** do algoritmo, o método proposto poderia inviabilizar a solução do problema de interesse original.

Esta limitação, felizmente, é facilmente sanada graças ao fato de a solução ótima para o **Passo 1** e o **Passo 3** do algoritmo serem a mesma e conhecida. Assim empregam-se as condições de Karush-Kuhn-Tucker para a obtenção de $\underline{\lambda}$, \underline{v} e $\underline{\zeta}$, o que permite concluir a implementação dos três algoritmos de

Benders e, conseqüentemente, torna viável a solução para o problema original de interesse, conforme representado esquematicamente na Figura F.2.

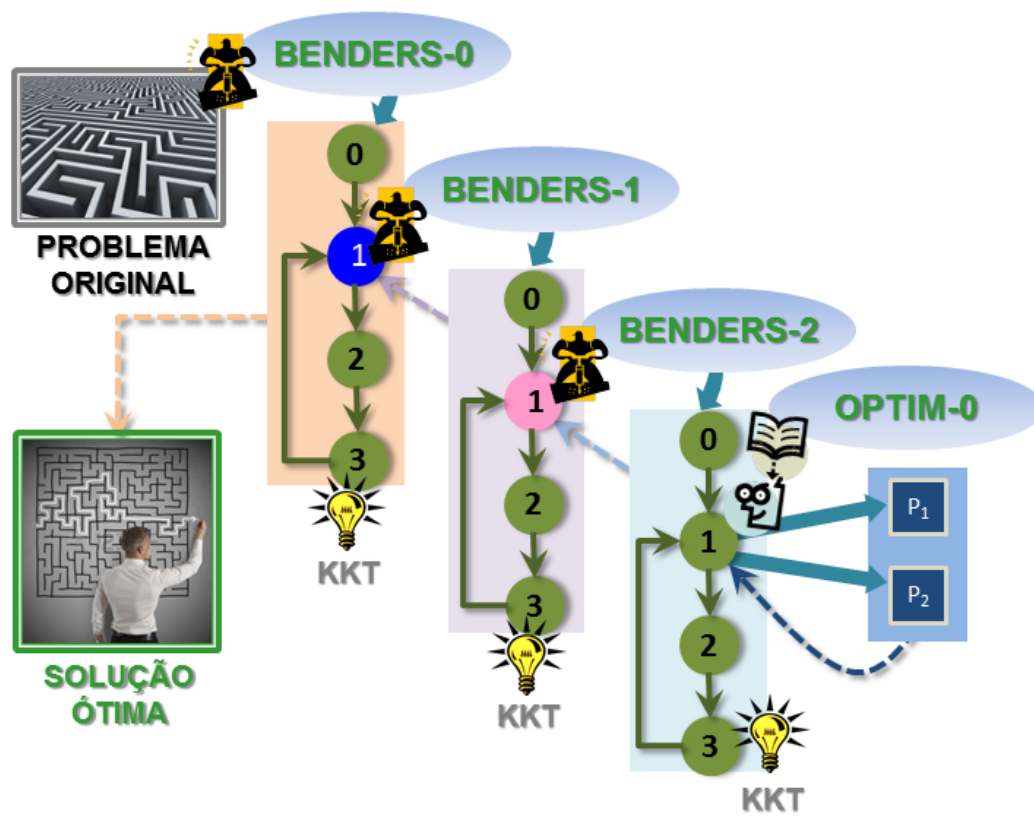


Figura F.2 – Visão geral do método proposto

Fonte: Elaborada pela autora