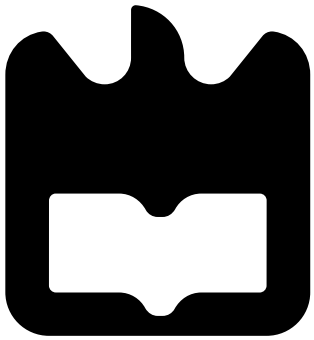
**Universidade de Aveiro** Departamento de Electrónica, Telecomunicações e Informática,
**2011**

**Tiago Rafael Pisco**

# Distribuição eficiente de IPTV



**Tiago Rafael Pisco**

# Distribuição eficiente de IPTV

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Eletrónica e Telecomunicações, realizada sob a orientação científica do Prof. Doutor Armando Nolasco Pinto, Professor do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

**o júri / the jury**

presidente / president

**Prof. Doutor António Luis Jesus Teixeira**
Professor Associado da Universidade de Aveiro (por delegação do Reitor da Universidade de Aveiro)

vogais / examiners committee

**Prof. Doutor Armando Humberto Moreira Nolasco Pinto**
Professor Auxiliar da Universidade de Aveiro (orientador)

**Prof. Doutor João José de Oliveira Pires**
Professor Auxiliar do Instituto Superior Técnico da Universidade Técnica de Lisboa

**agradecimentos / acknowledgements**

Em primeiro lugar quero agradecer ao orientador da presente dissertação de mestrado, Professor Armando Nolasco Pinto. Um muito obrigado pela disponibilidade, interesse e preocupação prestadas, assim como também pelos constantes esclarecimentos e sugestões realizadas ao longo deste trabalho.

Gostaria também de agradecer à minha família, à minha namorada e aos meus amigos pela constante presença na minha vida, sendo estas pessoas em última análise, parte daquilo que sou.

**Resumo**

As redes de transporte de telecomunicações foram inicialmente concebidas para transportar chamadas telefónicas através de tecnologias baseadas em comutação de circuitos. A adaptação destas redes para transportar tráfego de dados apresenta ineficiências que contribuem para que as receitas geradas pelos operadores não acompanhem a largura de banda fornecida aos clientes. Este facto tem vindo agravar-se com o crescimento de serviços como o IPTV, serviço que requer elevada largura de banda, variação acentuada da taxa de transmissão e ligações ponto-multiponto.

A tecnologia Ethernet surge como ponto de partida (e chegada, pois mais de 95% do tráfego de dados é gerado ou tem como destino uma porta Ethernet [1]) para novas tecnologias de rede de transporte, chamadas tecnologias Carrier Ethernet. Neste documento é estudada a tecnologia Ethernet, os requisitos de uma tecnologia Carrier Ethernet e a promissora tecnologia MPLS-TP, ainda em processo de normalização.

Com o propósito de otimizar a distribuição do serviço IPTV é proposto um cenário, onde se verifica que o recurso à transmissão combinada de canais *unicast* e *multicast* pode resultar num uso mais eficiente dos recursos da rede.

**Abstract**

The telecommunication aggregation networks were initially designed to transport phone calls, using technologies based on circuit switching. The adaptation of those networks to transport data traffic has some inefficiencies, which results, that the operator‘s revenues do not follow the bandwidth provided to customers. This fact has become a growing problem with the development of services such as IPTV, a service which requires high bandwidth, high transmission rate variation and point-multipoint connections.

The Ethernet appears as a starting point (and ending as well, since more than 95 % of the data traffic is generated or is destinated to an Ethernet port [1]) for new aggregation network's technologies, called Carrier Ethernet technologies. This paper, presents a study of the Ethernet technology, requirements for Carrier Ethernet technologies and the promising MPLS-TP, still in the standardization process.

In order to optimize IPTV service distribution, there is proposed a scenario, where it was verified that the combined use of unicast and multicast transmission can result in an efficient use of network resources.

# Conteúdo

**Conteúdo** i

**Lista de Figuras** iii

**Lista de Tabelas** v

**Lista de Símbolos** vii

**Acrónimos** ix

**1 Introdução** 1
- 1.1 Rede de transporte tradicional . . . 3
  - 1.1.1 PDH . . . 3
  - 1.1.2 SDH . . . 3
  - 1.1.3 Estado atual . . . 5

**2 Ethernet** 7
- 2.1 Perspetiva histórica . . . 7
  - 2.1.1 Aparecimento . . . 7
  - 2.1.2 CSMA/CD . . . 10
  - 2.1.3 Normas Ethernet . . . 13
  - 2.1.4 OSI . . . 14
  - 2.1.5 Trama . . . 15
- 2.2 Switched Ethernet . . . 16
  - 2.2.1 Funcionamento do swicth . . . 16
  - 2.2.2 Switchs . . . 17
  - 2.2.3 Full-Duplex . . . 18
  - 2.2.4 Controlo de fluxos . . . 19
  - 2.2.5 Auto-Negotiation . . . 21
  - 2.2.6 Spanning Tree Protocol . . . 22
  - 2.2.7 Link Aggregation . . . 24
  - 2.2.8 Virtual LANs . . . 26
- 2.3 Gigabit Ethernet . . . 27
  - 2.3.1 Half-Duplex . . . 28
  - 2.3.2 10 Gigabit Ethernet . . . 31
  - 2.3.3 40 e 100 Gigabit Ethernet . . . 31
- 2.4 Ethernet para a Rede de Transporte . . . 34

2.4.1 Motivação . . . 34
2.4.2 Desafios . . . 34

**3 Carrier Ethernet** **37**
3.1 Atributos segundo o MEF . . . 38
3.1.1 Serviços Normalizados . . . 38
3.1.2 Escalabilidade . . . 38
3.1.3 Fiabilidade . . . 38
3.1.4 Qualidade de serviço . . . 38
3.1.5 Gestão do serviço . . . 39
3.2 Modelo de Serviços . . . 39
3.2.1 UNI, EVC e NNI . . . 40
3.2.2 Caracterização do tráfego . . . 42
3.3 Tipos de Serviço Ethernet . . . 44
3.3.1 E-LINE . . . 44
3.3.2 E-LAN . . . 45
3.3.3 E-TREE . . . 46
3.4 Tecnologias . . . 47

**4 MPLS-TP** **49**
4.1 MPLS . . . 49
4.1.1 Label MPLS . . . 50
4.1.2 Arquitetura . . . 52
4.1.3 Funcionamento . . . 53
4.1.4 Estabelecimento de ligações . . . 53
4.1.5 MPLS para a rede de transporte . . . 56
4.2 MPLS-TP . . . 56
4.2.1 Funcionalidades . . . 57
4.2.2 Plano de Dados . . . 57
4.2.3 Plano de Controlo . . . 58
4.2.4 OAM . . . 60
4.2.5 Sobrevivência . . . 61
4.3 MLPS-TP no IXIA . . . 69

**5 Cenário de distribuição IPTV** **73**
5.1 EVP-Tree sobre MPLS-TP . . . 73
5.2 Modelo de custos . . . 74
5.3 Distribuição da popularidade dos canais . . . 77
5.4 Unicast e Multicast . . . 78
5.5 Aplicação do modelo matemático . . . 80
5.5.1 Abordagem exata . . . 80
5.5.2 Abordagem aproximada . . . 82

**6 Conclusão** **91**
6.1 Trabalho futuro . . . 94

**Bibliografia** **95**

# Lista de Figuras


1.1 Largura de Banda vs Receitas (Adaptado de [2]). . . . . 1
1.2 Trama STM-1 (adaptado de [3]) . . . . . 4
1.3 Ethernet sobre SDH . . . . . 5

2.1 Transmissão de tramas numa rede Aloha . . . . . 8
2.2 Taxa de transferência por tráfego oferecido numa rede ALOHA . . . . . 9
2.3 Rede Ethernet . . . . . 10
2.4 CSMA/CD . . . . . 11
2.5 Camadas de referencia Ethernet . . . . . 14
2.6 Tramas Ethernet . . . . . 15
2.7 Switched Ethernet . . . . . 16
2.8 Operações de *Switching*. É ilustrado o envio de uma trama P do terminal 1 para o terminal 2, e consequente trama de resposta R do terminal 2 para o terminal 1. Inicialmente os *switchs* não possuem informações sobre nenhum dos terminais nas suas tabelas de encaminhamento, desta forma a trama P é encaminhada através de mecanismos de *flooding*. À medida que P é reencaminhada os *switchs* adquirem informação sobre a localização do terminal 1. Este mecanismo de aprendizagem leva a que o posterior encaminhamento da trama R, seja efetuado através de *forwardings*. . . . . 17
2.9 Ligação em *full-duplex* . . . . . 19
2.10 Trama PAUSE . . . . . 20
2.11 *Head-of-Line Blocking* . . . . . 21
2.12 Mensagem de *autonegotiation* . . . . . 22
2.13 Protocolo STP . . . . . 23
2.14 Link Aggregation . . . . . 25
2.15 *Switch* com implementação de VLANs baseadas em MAC . . . . . 26
2.16 *tag* VLAN . . . . . 27
2.17 Trama Ethernet com *Carrier Extension*. . . . . 28
2.18 Tramas enviadas em *frame bursting*. . . . . 29
2.19 Diagrama de fluxos do *frame bursting*. . . . . 29
2.20 Arquitetura da camada física das tecnologias 40GbE e 100GbE . . . . . 32
2.21 Mecanismo Multi-Lane Distribution (MLD) da sub-camada Physical Coding Sublayer (PCS) . . . . . 33
2.22 Pilha protocolar de uma solução para rede de transporte com tecnologias Synchronous Digital Hierarchy (SDH) e Ethernet . . . . . 34

3.1 Modelo de serviços genérico para distribuição de serviços Ethernet (Fonte: MEF) 40
3.2 Algoritmo de atribuição de cores às tramas (Fonte: MEF) . . . . . 43
3.3 Serviço E-Line (Fonte: MEF) . . . . . 44
3.4 Serviço E-Lan (Fonte: MEF) . . . . . 45
3.5 Serviço E-TREE (Fonte: MEF) . . . . . 46
3.6 Serviço EVP-Tree service (Fonte: MEF) . . . . . 46

4.1 Vista geral - IxNetwork 6.0.400.14 (Fonte: Software IxNetwork 6.0.400.14) . . 50
4.2 *Label* MPLS . . . . . 50
4.3 *Label* MPLS (Fonte: Software IxNetwork 6.0.400.14) . . . . . 51
4.4 Rede MPLS . . . . . 52
4.5 Túnel LSP . . . . . 54
4.6 Convergência da tecnologia MPLS-TP (fonte [4]) . . . . . 57
4.7 Formato do pacote G-Ach para um LSP. . . . . 59
4.8 Esquema de proteção 1:1 linear . . . . . 61
4.9 Lógica de funcionamento da geração de mensagens PSC (Fonte [5]) . . . . . 63
4.10 Mecanismo de proteção em anel 1:1 (1) . . . . . 64
4.11 Mecanismo de proteção em anel 1:1 (2) . . . . . 64
4.12 Mecanismo de proteção em anel 1:1 (3) . . . . . 65
4.13 Mecanismo de proteção ponto-multiponto em anel (1:N) . . . . . 66
4.14 Mecanismo de proteção linear ponto-multiponto (1:N) . . . . . 67
4.15 Mecanismo de proteção linear ponto-multiponto $(1:1)^N$ . . . . . 68
4.16 Teste com recurso ao equipamento IXIA XM12 High Performance Chassis . . 69
4.17 Consituição das tramas usadas no teste. . . . . 70
4.18 Número de tramas enviadas e recebidas por fluxo. . . . . 70
4.19 Percentagem de tramas perdidas por fluxo. . . . . 71

5.1 Serviço EVP-Tree sobre MPLS-TP. O servidor de vídeo é a raiz do sistema e está ligado ao *router* interno de distribuição A. O *router* A tem como função distribuir os sinais provenientes do servidor de vídeo para os Label Edge Routers (LERs) que estão ligados à rede de acesso e servem os N clientes. . . . . 74
5.2 Custo de transmissão em *unicast* . . . . . 75
5.3 Custo de transmissão em *multicast* . . . . . 76
5.4 Influência do parametro $\alpha$ numa distribuição Zipf . . . . . 78
5.5 Custo de transmissão dos K canais em *unicast* e *multicast* . . . . . 79
5.6 Número de combinações possíveis entre N e K . . . . . 80
5.7 Função de distribuição de $n_u$ na abordagem exata . . . . . 81
5.8 Comparação do valor exato e aproximado da variável aleatória $n_u$ . . . . . 83
5.9 Custo Rs em função de M . . . . . 84
5.10 Função de distribuição de $n_u$ . . . . . 85
5.11 Probabilidade de bloqueio do sistema . . . . . 86
5.12 Custo Rs em função de M, com $P_{bloq} = 0.0001$ . . . . . 87
5.13 Custo Rs em função do número dos clientes . . . . . 88
5.14 Custo Rs em função dos M canais transmitidos em multicast para *routers* de distribuição com diferente número de portas ativas . . . . . 89
5.15 Custo Rs em função dos M canais transmitidos em multicast para *routers* de distribuição com diferentes $\gamma$ . . . . . 90

# Lista de Tabelas

1.1 Canais PDH . . . . . 3
1.2 Canais SDH . . . . . 4

2.1 Normas Ethernet . . . . . 13
2.2 Parametros do CSMA/CD nas várias tecnologias (Adaptado de [6, pág 57]) . 30

3.1 Atributos da UNI (adaptado Fonte: MEF) . . . . . 41
3.2 Atributos da EVC (adaptado Fonte: MEF) . . . . . 42

# Lista de Símbolos

| | |
|---|---|
| $K$ | Número de canais oferecidos ao cliente |
| $N$ | Número de clientes do sistema |
| $\beta$ | Razão entre o custo de transmitir um canal em *multicast* sobre o custo de transmitir o canal em *unicast* |
| $C_m$ | Custo de transmitir um canal em *multicast* |
| $C_u$ | Custo de transmitir um canal em *unicast* |
| $P_p$ | Custo de processamento de um pacote |
| $P_s$ | Custo de pesquisa na tabela de encaminhamento |
| $\gamma$ | Razão entre o custo de pesquisa na tabela de encaminhamento sobre o custo de processar um pacote |
| $N_A$ | Número de portas ativas do *router* |
| $\pi_k$ | Popularidade do canal k |
| $k$ | Índice de popularidade do canal |
| $d$ | Constante de normalização |
| $\alpha$ | Variável de Zipf |
| $\Pi$ | Vetor de popularidade dos canais |
| $a$ | Probabilidade de um cliente se encontrar ativo |
| $r_m$ | Custo de transmitir todos os canais em *multicast* |
| $r_u$ | Custo de transmitir todos os canais em *unicast* |
| $b_k, c_k$ | Número de clientes sintonizados no canal k |
| $B$ | Vetor de clientes sintonizados em cada k canal |
| $n_a$ | Número de clientes ativos |
| $n_i$ | Número de clientes inativos |

| | |
|---|---|
| $M$ | Canais a transmitir em *multicast* escolhidos pelo operador |
| $P_m$ | Probabilidade de um cliente assistir a um canal transmitido em *multicast* |
| $P_u$ | Probabilidade de um cliente assistir a um canal transmitido em *unicast* |
| $n_u$ | Número de transmissões de canais em *unicast* |
| $n_m$ | Número de canais disponibilizados um *multicast* |
| $\mu$ | Média do número de transmissões *unicast* |
| $\sigma$ | Desvio padrão do número de transmissões *unicast* |
| $r_s$ | Custo de uma transmissão combinada de canais em *unicast* e *multicast* |
| $P_{bloq}$ | Probabilidade de bloqueio do sistema |

# Acrónimos

**ADM** Add/Drop Multiplexer

**ACh** Associated Channel Header

**APP** Application Services Layer

**ASIC** Application Specific Integrated Circuits

**ATM** Asynchronous Transfer Mode

**BFD** Bidirectional Forwarding Detection

**BNC** Bayonet Neill-Concelman

**BPDU** Bridge Protocol Data Unit

**CACM** Communications of the Association for Computing Machinery

**CAPEX** Capital Expenditure

**CBS** Committed Burst Size

**CCDF** Complementary Cumulative Distribution Function

**CE** Customer Equipment

**CES** Circuit Emulation Servivces

**CIR** Committed Information Rate

**CM** Color Marking

**CoS** Class of Service

**CRC** Cyclic Check Sequence

**CSMA/CD** Carrier Sense Multiple Acess with Collision Detection

**DA** Destination Address

**DCS** Digital Crossconnects

**DSAP** Destination Service Access Point

**E-LAN** Ethernet Local Area Network

**E-Line** Ethernet Line

**E-NNI** External Network to Network Interface

**E-Tree** Ethernet Tree

**EBS** Excess Burst Size

**ECMP** Equal Cost Multi-Path

**EIR** Excess Information Rate

**EPL** Ethernet Private Line

**EP-Tree** Ethernet Private Tree Service

**EPLAN** Ethernet Private Local Area Network

**ESM** Ethernet Service Model

**ETH** Ethernet Service Layer

**EVC** Ethernet Virtual Connection

**EVPL** Ethernet Virtual Private Line

**EVP-Tree** Ethernet Virtual Private Tree Service

**EVPLAN** Ehernet Virtual Private Local Area Network

**FCS** Frame Check Sequence

**FEC** Forwarding Equivalence Class

**FLP** Fast Link Pulse

**FPGA** Field Programmable Gate Array

**G-ACh** Generic Associated Channel

**GFP** Generic Framing Procedure

**GMPLS** Generalized Multi-Protocol Label Switching

**HOL Blocking** Head-of-Line Blocking

**HSSG** Higher Speed Study Group

**I-NNI** Internal Network to Network Interface

**IEEE** Institute of Electrical and Electronics Engineers

**IETF** Internet Engineering Task Force

**IFS** Inter Frame Spacing

**IGP** Interior Gateway Protocol

**IP** Internet Protocol

**IPTV** Internet Protocol Television

**ITU-T** International Telecommunication Union - Telecommunication Standardization Setor

**ITU** International Telecommunication Union

**LAN** Local Area Network

**LCAP** Link Control Aggregation Protocol

**LCAS** Link Capacity Adjustment Scheme

**LDP** Label Distribution Protocol

**LER** Label Edge Router

**LFIB** Label Forwarding Information Base

**LLC** Logical Link Control

**LSP** Label Switched Path

**LSR** Label Switching Router

**MAC** Media Acess Control

**MAN** Metro Area Network

**MEF** Metro Ethernet Forum

**MEN** Metro Ethernet Network

**MII** Media Independent Interface

**MLD** Multi-Lane Distribution

**MMF** Multi Mode Fiber

**MPLS-TP** Multi-Protocol Label Switching - Transport Profile

**MPLS** Multi Protocol Label Switching

**MTBF** Mean Time Between Failure

**NG-SDH** New Generation - Synchronous Digital Hierarchy

**NI-NNI** Network Interworking Network to Network Interface

**NIC** Network Interface Controller

**NLP** Normal Link Pulse

**NMS** Network Management System

**NNI** Network to Network Interface

**OAM** Operational, Administration and Maintenance

**OPEX** Operational Expenditure

**OSI** Open Systems Interconnection

**OTN** Optical Transport Network

**OUI** Organizationally Unique Identifier

**PARC** Palo Alto Research Center

**PBB-TE** Provider Backbone Bridge Traffic Engineering

**PCM** Pulse Code Modulation

**PCS** Physical Coding Sublayer

**PDH** Plesiochronous Digital Hierarchy

**PHP** Penultimate Hop Popping

**PMA** Physical Medium Attachment

**PMD** Physical Medium Dependent

**PSC** Protection State Coordination Protocol

**QoS** Quality of Service

**RDI** Remote Defect Indication

**RFC** Request For Comments

**RSVP** Resource Reservation Protocol

**RSVP-TE** Resource Reservation Protocol - Traffic Engineering

**SA** Source Adress

**SDH** Synchronous Digital Hierarchy

**SEN** Service Provider Ethernet Network

**SI-NNI** Service Interworking Network to Network Interface

**SLA** Service Level Agreements

**SMF** Single Mode Fiber

**SONET** Synchronous Optical Networking

**SSAP** Source Service Access Point

**STP** Spanning Tree Protocol

**T-MPLS** Transport Multi-Protocol Label Switching

**TCP** Transmission Control Protocol

**TDM** Time Division Mulplexing

**TPID** Tag Protocol Identifier

**TRAN** Transport Layer

**TTL** Time To Live

**UNI-C** User Network Interface - Client side

**UNI-N** User Network Interface - Network side

**UNI** User Network Interface

**UTP** Unshielded Twisted Pair

**VCAT** Virtual Concatenation

**VLAN** Virtual Local Area Network

**VPN** Virtual Private Networks

**WAN** Wide Area Network

**WDM** Wavelength Division Multiplexing

# Capítulo 1

# Introdução

As redes de transporte de informação são caraterizadas por agregar e transportar o tráfego proveniente das redes de acesso e redes metropolitanas. Antes de surgirem os serviços Internet, o tráfego das redes de transporte era maioritariamente originado em chamadas telefónicas, que eram transportadas com recurso a tecnologias baseadas em comutação de circuitos.

O aparecimento e rápido crescimento dos serviços Internet, levou a que o volume de tráfego de dados transportado, rapidamente ultrapassasse o tráfego de voz. As redes de transporte de telecomunicações, inicialmente desenhadas para transportar chamadas telefónicas apresentam ineficiências ao serem adaptadas para transportar tráfego de dados. Tais ineficiências contribuíram para que as receitas geradas pelos operadores não acompanhassem a largura de banda fornecida aos clientes, tal como ilustra a figura (1.1).

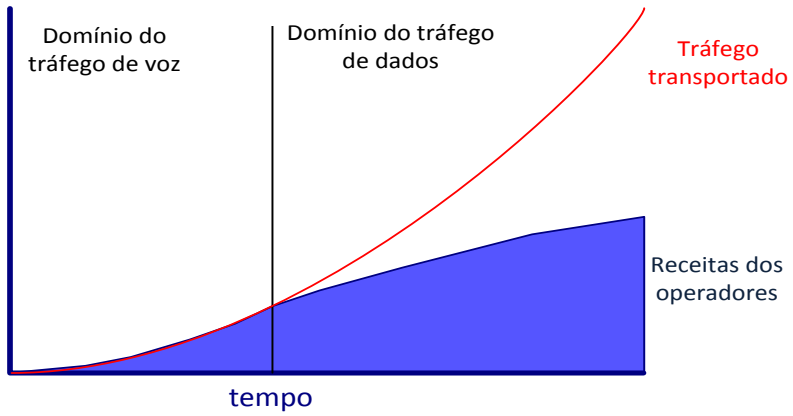



Figura 1.1: Largura de Banda vs Receitas (Adaptado de [2]).

Segundo [7], a largura de banda gerida pelos grandes operadores de telecomunicações nas redes de transporte tem se expandido entre 75% a 125% por ano. O crescimento deve-se sobretudo ao aumento do número de clientes de serviços Internet e ao aparecimento de aplicações de elevado consumo de largura de banda, tal como Internet Protocol Television (IPTV), jogos de vídeo processados *on-demand* ou serviços de videoconferência [7]. O número mundial de clientes IPTV aumentou de 4 milhões em 2004, para mais de 25 milhões em 2010 [8]. É portanto fundamental, uma alteração na arquitetura e funcionamento das redes de telecomunicações

que vá de encontro à mudança da natureza da informação transportada, e que consiga distribuir os serviços mencionados de forma eficiente, para assim reduzir drasticamente o custo de transporte por bit.

O presente documento abordará a temática de uma eficiente distribuição de serviços IPTV. O IPTV é uma tecnologia de transmissão de televisão com recurso ao protocolo Internet Protocol (IP), contudo não deve ser confundido com o serviço Web TV. Uma diferença fundamental reside no facto do IPTV oferecer garantia de qualidade de serviço aos clientes, já que a distribuição dos serviços é realizada através de uma rede IP independente da rede global de internet. Trata-se de uma rede IP totalmente controlada pelo operador, que poderá ser compreendida como uma rede IP privada.

Um sistema de distribuição IPTV poderá disponibilizar vários tipos de serviços. Serviços de *broadcast* como distribuição de canais televisivos ou acesso a grelhas de programação. Serviços a pedido, como por exemplo aluguer de filmes, aluguer de músicas, agendamento de gravações ou jogos interativos. Serviços de publicidade interativa e/ou a pedido. Serviços de interesse público, como comunicações de emergência. Telesserviços de aprendizagem, saúde ou outros. Serviços interativos de comunicação, comércio, informação ou entretenimento[1] [9, pág 27].

Estes serviços possuem diferentes requisitos do ponto de vista da largura de banda. Se por um lado as transmissões de canais televisivos em alta definição, exigem uma elevada largura de banda (4-12 Mbps) que será ocupada durante um tempo considerável. Por outro lado, o aluguer de filmes poderá não necessitar de qualquer largura de banda durante dias, mas após o momento em que o cliente efetua um aluguer é necessário uma taxa de transmissão extremamente elevada, para poder enviar o filme em tempo útil.

Para abordar estas questões serão estudadas as tecnologias candidatas à rede de transporte e os requisitos tecnológicos que uma rede de transporte requer. Posteriormente serão analisadas metodologias que visam distribuir de forma eficiente o serviço de IPTV, através da otimização de transmissões *unicast* e *multicast*.

No primeiro capítulo será estudada a rede de transporte tradicional, onde impera a comutação de circuitos, e as suas principais limitações face à atual realidade do tráfego transportado. O segundo capítulo estudará a tecnologia Ethernet. O aumento da capacidade de transmissão, alcance e flexibilidade da tecnologia Ethernet, fê-la predominar nas redes locais, Local Area Networks (LANs) e ao mesmo tempo candidatou-a a operar também nas redes de área alargada, Wide Area Networks (WANs), podendo segundo alguns autores [10, pág 103], ser a solução global para as redes de telecomunicações. O terceiro capítulo abordará a definição de Carrier Ethernet segundo o consórcio Metro Ethernet Forum (MEF), onde são focados os atributos requeridos por uma tecnologia de rede de transporte. A tecnologia Multi-Protocol Label Switching - Transport Profile (MPLS-TP), ainda em estado de normalização à data de elaboração deste documento, é já assumida como uma forte candidata à implementação massiva em redes de transporte, sendo por isso objeto de estudo no quarto capítulo deste documento. No quinto capítulo, será considerado um cenário de distribuição IPTV, com recurso à tecnologia MPLS-TP, que será analisado à luz do modelo matemático presente em [11] [12] [13].

[1] Os serviços descritos resultam do uso de tecnologia IP, aliada às funcionalidades de armazenamento de uma *set-top-box* na posse dos clientes.

## 1.1 Rede de transporte tradicional

Nas redes de transporte tradicionais europeias a tecnologia encarregue destas funções é o SDH[2]. A tecnologia SDH consiste na multiplexação de canais com baixa taxa de transmissão em canais com uma taxa de transmissão superior. As taxas de transmissão dos canais são fixas e estão organizadas hierarquicamente de modo a possibilitar escalabilidade na agregação dos vários canais.

### 1.1.1 PDH

Para uma melhor perceção do funcionamento da tecnologia SDH é conveniente conhecer a tecnologia que está na base desta, o Plesiochronous Digital Hierarchy (PDH). O PDH foi desenvolvida nos anos 50, e tinha como objetivo encaminhar o tráfego de chamadas telefónicas utilizando multiplexagem Time Division Mulplexing (TDM).

Cada chamada telefónica ocupa uma largura de banda de 64 kbps, já que é realizada uma codificação *Pulse Code Modulation* em 8 bits com uma frequência de amostragem de 8 kHz. Cada 30 chamadas telefónicas são agregadas num mesmo canal PDH de 2.048 Mbps, o canal E1. Vários canais E1 são agrupados em outros canais de capacidade superior e para compensar as ligeiras variações da taxa de transmissão de cada canal, no momento da agregação dos vários canais, são utilizados bits de aposição. A tabela (1.1) mostra os vários canais e respetivas taxas de transmissão usados no PDH.

| **PDH** | **Taxa de transmissão** |
|---|---|
| E1 | 2.048 Mbps |
| E2 | 8.448 Mbps |
| E3 | 34.368 Mbps |
| E4 | 139.264 Mbps |

Tabela 1.1: Canais PDH

Note-se que a taxa de transmissão de cada canal é igual a quatro vezes a taxa de transmissão do canal imediatamente inferior mais a taxa dos bits de aposição e controlo. Apesar da tecnologia colmatar as primeiras necessidades das redes de transporte, possuía algumas limitações. Limitações das taxas de transmissão dos canais, dificuldade em extrair canais com baixa taxa de transmissão aos canais superiores, incompatibilidades entre equipamentos de diferentes fornecedores e funções limitadas de Operational, Administration and Maintenance (OAM) [14, pág 1-2]. Os problemas anteriormente referidos levaram à necessidade de criar uma nova tecnologia, o SDH.

### 1.1.2 SDH

A tecnologia SDH faz parte da primeira geração de redes óticas e opera nas duas camadas inferiores do modelo Open Systems Interconnection (OSI). Fornece ligações do tipo comutação de circuitos extremo a extremo, e tal como o PDH usa um mecanismo de multiplexagem de canais de baixas taxas de transmissão em canais com taxas de transmissão superiores. No entanto é possível retirar com maior facilidade, canais com taxas de transmissão mais baixas a canais com taxas de transmissão superiores. A tabela (1.2) mostra os canais SDH e a

[2]Nos Estados Unidos da América é usada a tecnologia análoga Synchronous Optical Networking (SONET).

capacidade de transmissão que possuem [10, pág 104-106] [3], como se observa o SDH possuí canais com taxas de transmissão bastante superiores às taxas dos canais PDH.

| **SDH** | **Taxa de transmissão** |
|---|---|
| STM-1 | 155.520 Mbps |
| STM-4 | 622.080 Mbps |
| STM-16 | 2488.320 Mbps |
| STM-64 | 9953.280 Mbps |

Tabela 1.2: Canais SDH

A tecnologia possuí funcionalidades de OAM superiores às da tecnologia PDH, o que constitui uma mais valia bastante importante para os operadores. Os cabeçalhos das tramas permitem, por exemplo, determinar se as tramas são recebidas com erros, ou seguir o tráfego através da rede.

Os elementos que constituem a rede são os terminais de linha, os Add/Drop Multiplexers (ADMs), os Digital Crossconnectss (DCSs) e os regeneradores. Os terminais de linha, multiplexam e desmultiplexam os fluxos de tráfego. Os ADMs disponibilizam um método eficiente para extrair parte do tráfego incidente num nó da rede, permitindo a passagem do restante tráfego. Quanto aos DCSs, são usados em nós maiores e permitem comutar um elevado número de fluxos de tráfego. Por fim os regeneradores têm como função regenerar o sinal.

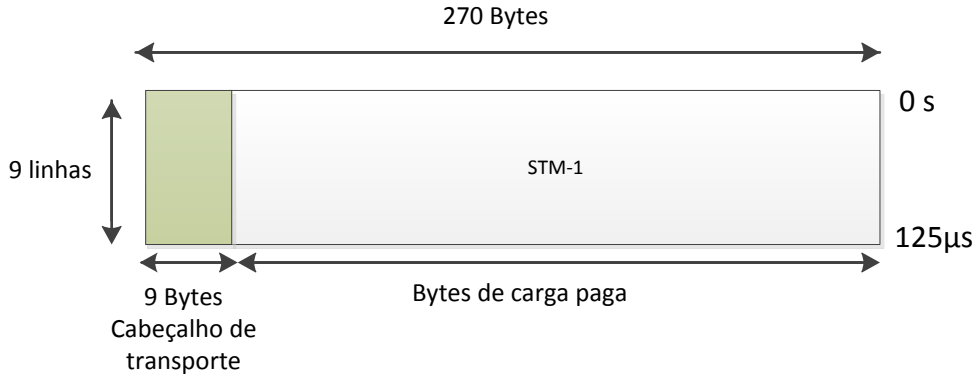



Figura 1.2: Trama STM-1 (adaptado de [3])

O STM-1 é o canal base do SDH, e do ponto de vista lógico pode ser representado por uma estrutura com 270 colunas e 9 linhas tal como é ilustrado na imagem (1.2), onde cada célula representa um byte. As 9 primeiras colunas estão reservadas para o cabeçalho e as restantes 261 estão reservadas para o transporte da carga paga (*payload*). Cada bloco completo é transmitido em 125 $\mu$s, o que resulta numa taxa de transmissão de 155.520 Mbps.

No processo de agregação dos vários fluxos pode acontecer que as tramas cheguem desfasadas, ou que a taxa de transmissão de um fluxo contribuinte tenha variado, devido a alterações nas condições das fibras óticas. Para compensar estas variações das taxas de transmissão a tecnologia recorre a mecanismos de ponteiros que usam os bytes de aposição do cabeçalho para indicar o inicio das tramas. O uso de ponteiros leva a que as cargas pagas não tenham

uma posição fixa nas tramas, mas possam 'flutuar' ligeiramente sobre estas [3].

Um ponto forte da tecnologia é a sua capacidade de sobrevivência. Ou seja a capacidade de recuperar de forma rápida a cortes e falhas nas suas ligações ou nós. Esta capacidade deve-se ao facto da arquitetura usada, em anel ou malha, permitir o estabelecimento de ligações de proteção devido à sua redundância [3]. O SDH garante um tempo de recuperação a falhas inferior a 50 ms [3]. E é esperado que uma rede SDH possua uma disponibilidade entre 99.99% a 99.999% devido aos seus mecanismos de sobrevivência [3].

### 1.1.3 Estado atual

Com o aparecimento e crescimento do tráfego de dados, as redes de telecomunicações inicialmente estruturadas para o transporte de tráfego de voz, começaram também a suportar o tráfego de dados. Tramas Ethernet necessitaram de ser transportadas sobre SDH, o que levou ao aparecimento de redes baseadas em comutação de circuitos adaptadas para comutar pacotes. A figura (1.3) mostra as camadas protocolares destas redes:

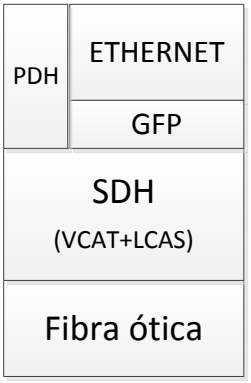



Figura 1.3: Ethernet sobre SDH

No que respeita ao transporte de chamadas telefónicas a tecnologia SDH complementa a tecnologia PDH através da agregação dos canais de baixa taxa de transmissão em canais com taxas superiores, mecanismos de sobrevivência e OAM.

No transporte de tráfego de dados, as tramas Ethernet são encapsuladas nos canais SDH através do Generic Framing Procedure (GFP) [15, pág 40]. Contudo a pouca flexibilidade na largura de banda oferecida nos canais STM (tabela (1.2)) pode levar a elevados desperdícios de recursos. Por exemplo, se for necessária uma ligação Ethernet com 100 Mbps, o canal usado seria o STM-1 com 155 Mbps, e desta forma desperdiçar-se-ia bastante largura de banda [10, pág 106].

A introdução da tecnologia Virtual Concatenation (VCAT) resolve este problema através da criação de contentores virtuais no inicio das ligações. Os contentores virtuais percorrem a rede de forma independente e são novamente agregados no final da ligação [14, pág 11]. Desta forma o VCAT atribui à tecnologia uma maior granularidade na sua oferta de largura de banda, aumentando a eficiência do transporte de tramas Ethernet.

No entanto a largura de banda necessária numa ligação poderá variar durante o período em que a ligação se encontra ativa. Assim é necessário um mecanismo que ajuste a capacidade de uma ligação a possíveis flutuações da fonte de transmissão. A tecnologia Link Capacity

Adjustment Scheme (LCAS) realiza esta função através da alteração do número de contentores virtuais da tecnologia VCAT [15] e garante que não existem perdas de tramas ao surgir um aumento ou diminuição da taxa de transmissão do emissor[3].

O emprego da tecnologia SDH aliada à implementação de tecnologias VCAT, LCAS e GFP nos extremos da rede, é geralmente apelidada de New Generation - Synchronous Digital Hierarchy (NG-SDH) [14].

Na caso da camada física da rede de transporte usar tecnologia Wavelength Division Multiplexing (WDM) a tecnologia de transporte tenderá a ser o Optical Transport Network (OTN). O OTN é uma tecnologia de rede de transporte análoga ao SDH mas pensada para o transporte dos vários comprimentos de onda provenientes do WDM.

Apesar do recurso a várias tecnologias com vista a otimizar as redes para o transporte de tráfego de dados, a atual solução apresenta ainda alguns inconvenientes:

- Falta de eficiência devido, à necessidade da informação atravessar várias camadas protocolares. O que leva a um elevado Capital Expenditure (CAPEX) associado aos equipamentos de interface das ligações dos diferentes protocolos [10, pág 106].
- Elevado Operational Expenditure (OPEX). A multiplicidade de tecnologias leva a que os custos de manutenção sejam também bastante elevados, devido à necessidade de técnicos especializados [10, pág 106] [16, pág 5].
- É orientado para ligações extremo-a-extremo, não suportando assim originalmente tráfego ponto-multiponto, ou multiponto-multiponto [16, pág 5], o que não a torna apelativa para o transporte de serviços IPTV.
- Desadequado para transmitir tráfego do tipo *burst* devido à rigidez das taxas de transmissão do SDH [16, pág 5].

[3]Uma descrição mais detalhada sobre as tecnologia LCAS e VCAT poderá ser encontrada em [15] e [14].

# Capítulo 2

# Ethernet

Atualmente a Ethernet é a tecnologia de LANs mais usada e difundida a nível mundial. Foi inventada por Bob Metcalfe em 1973 na Xerox Palo Alto Research Center (PARC), Califórnia, num projeto que pretendia interligar computadores e impressoras a laser de alta velocidade [17, pág 4].

A necessidade de partilha de impressoras, servidores e mais tarde o acesso à Internet, levou à proliferação de redes locais onde a Ethernet se assumiu como tecnologia dominante. A Ethernet opera nas duas camadas OSI mais baixas, camada física e camada de ligação e baseia-se na comutação de tramas para transmitir informação [16].

Este capítulo pretende estudar o atual estado da tecnologia. Neste sentido a secção seguinte, para além de conter uma breve resenha histórica da tecnologia, analisa alguns conceitos intrinsecamente ligados à Ethernet, tal como o acesso ao meio, a estrutura da trama usada e a referência às principais normas.

Na secção *Switched Ethernet* será descrito o atual funcionamento de uma rede Ethernet, os dispositivos que estão na base deste funcionamento. Também serão estudados os modos de operação, protocolos usados e funcionalidades avançadas da tecnologia. Na última secção serão abordas as alterações que a tecnologia sofreu nas implementações Gigabit Ethernet, sendo foco de especial atenção as tecnologias de 40 e 100 Gibabit Ethernet.

## 2.1 Perspetiva histórica

### 2.1.1 Aparecimento

A 22 de maio de 1973, Bob Metcalfe escreveu um memorando que descrevia o funcionamento da tecnologia que usara para ligar computadores e impressoras. Esta nova tecnologia tinha por base anteriores experiências com redes, nomeadamente a experiência efetuada no final dos anos 60 por Norman Abramson e os seus colegas com uma rede de comunicações rádio que ligava ilhas do Hawaii, a rede Aloha [17, pág 5-7].

A rede Aloha possuía um mecanismo de partilha do canal de comunicação onde qualquer terminal da rede pode enviar tramas quando quiser. Após o envio da trama, o terminal emissor recebe um sinal de *acknowledgment* (ACK) que verifica se a trama foi corretamente enviada. O terminal poderá detetar uma colisão ou assumir que esta acontece se não receber o sinal de *acknowledgment* num intervalo de tempo pré-definido (*timeout*). Ao ocorrer uma colisão a trama é retransmitida após um intervalo de tempo aleatório. Note-se o facto do

*timeout* necessitar de ser maior que o intervalo de propagação de ida e volta da mensagem (*round-trip delay*). A figura (2.1) ilustra três terminais a enviarem tramas para a rede.

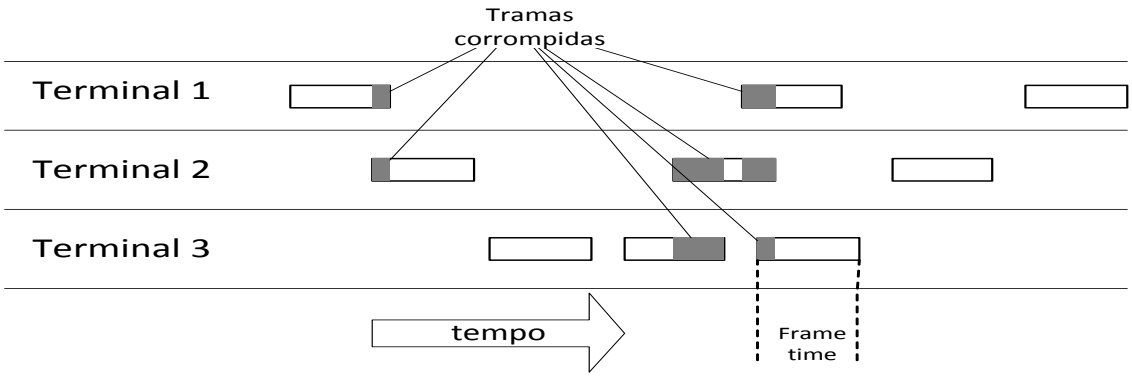



Figura 2.1: Transmissão de tramas numa rede Aloha

Quando mais do que uma trama ocupa o canal ao mesmo tempo ocorre uma colisão e as tramas são ignoradas. Como se pode observar em (2.1) prevê-se o aparecimento de muitas tramas corrompidas como resultado das colisões. A razão entre o número de tramas que atravessa a rede e o número de tramas total geradas pelos terminais poderá ser usada como métrica de eficiência da tecnologia. A questão da eficiência da rede Aloha foi abordada em [18][pág 286-292] e [19, pág 247-250], e será aqui apresentada.

Comecemos por definir o *frame time* como o tempo necessário para transmitir uma trama de tamanho fixo[1]. O *frame time* será igual ao tamanho da trama dividido pela taxa de transmissão. Assume-se que o número de tramas que atravessam a rede com sucesso, num *frame time* é modelado por uma distribuição de Poisson com média S. Assume-se também que a rede possuí infinitos terminais e desta forma o aumento do número de colisões na rede, não afetará a média S. No caso de existir apenas um terminal, e este enviar uma trama a cada *frame time*, S será 1, ou seja a rede terá uma eficiência de 100

O parâmetro S representa assim um parâmetro de utilização da rede, que poderá ser interpretado como a taxa de transmissão da rede, medido pelo número médio da tramas transmitidas num *frame time*.

Será também necessário considerar o evento de retransmissão de tramas devido a colisões. Assumindo que a probabilidade de existirem k tentativas de retransmissão por *frame time* é também uma variável de Poisson com média G, observa-se que $G \geq S$. Se os terminais gerarem poucas tramas $G \approx S$ e se o número de tramas geradas durante um *frame time* aumentar, $G > S$.

Para qualquer situação, S será a carga oferecida G vezes a probabilidade de uma trama não sofrer colisão. Ou seja $S = GP_0$. Para que uma trama não sofra colisões, é necessário que mais nenhum terminal transmita uma trama num intervalo de dois *frame time*. A probabilidade de k tentativas de transmissão de tramas serem geradas durante dois *frame time* é dado por:

$$Pr[k] = \frac{(2G)^k e^{-2G}}{k!} \tag{2.1}$$

[1]Visto tratar-se de uma pequena rede, o tempo de propagação é desprezado.

Note-se que (2.1) é uma distribuição de Poisson com média 2G, já que é necessário considerar dois *frame time* para garantir que não há colisões. Assim a probabilidade de não ocorrer nenhuma tentativa é $e^{-2G}$, logo $P_0 = e^{-2G}$ . A função taxa de transferência por *frame time* será:

$$S = Ge^{-2G} \tag{2.2}$$

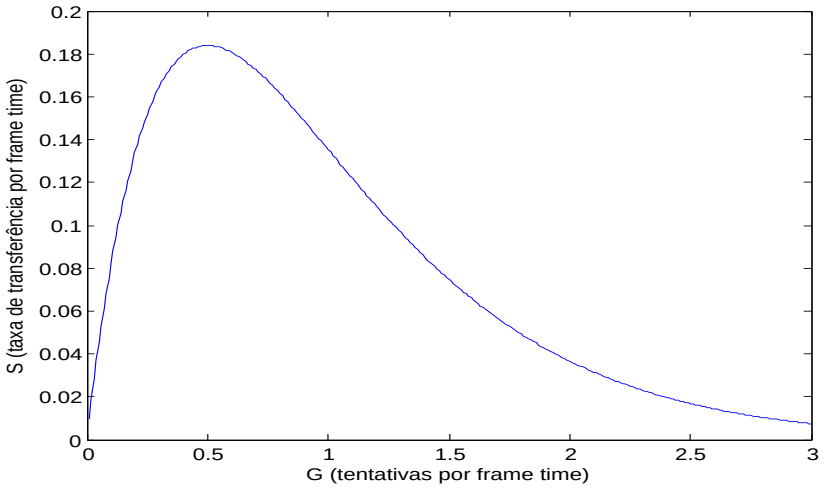



Figura 2.2: Taxa de transferência por tráfego oferecido numa rede ALOHA

A função possui um máximo quando G=0.5, de $S = \frac{1}{2e} \approx 0.184$, assim a taxa máxima de ocupação do meio, será de aproximadamente 18% e é obtida quando é gerada em média uma trama em dois *frame time*. Como é previsível um aumento de tráfego numa rede deste tipo, leva a um rápido aumento do número de colisões e consequente decréscimo da taxa de transmissão.

Bob Metcalfe percebeu que poderia melhorar este mecanismo de acesso ao meio através de uma deteção prévia do estado do canal por parte dos terminais (*Carrier Sense*), diminuindo assim o número de colisões. Foi ainda adicionada a funcionalidade de *Collision Detection* que permite a deteção de colisões, antes das tramas serem totalmente enviadas, ou recebidas. Assim aparece o Carrier Sense Multiple Acess with Collision Detection (CSMA/CD), o mecanismo de acesso ao meio tradicional da Ethernet, que possibilita taxas de transferência por *frame time* próximas de 100% [19, pág 251-252] [20].

A primeira experiência ocorre no final de 1972 na Xerox PARC e regista uma taxa de transmissão de 2,94 Mbps. É a seguir a esta experiência que Bob decide mudar o nome da tecnologia de 'Alto Aloha Network' para 'Ethernet' deixando claro que a tecnologia poderia ser usada por qualquer computador e não apenas pelos computadores da Xerox Parc. O prefixo 'Ether' provém da analogia percecionada por Bob Metcalfe entre o cabo Ethernet, que é partilhado por todos os terminais, e o éter luminífero que se pensava ser o meio comum de propagação das ondas eletromagnéticas.

Em julho de 1976 Bob Metcalfe e David Boggs publicam *Ethernet: Distributed Packet Switching for Local Computer Networks* no Communications of the Association for Computing

Machinery (CACM) e no final de 1977 juntamente com Charles P. Tracker e Butler W. Lampson recebem a patente Norte Americana número 4,063,220 pelo seu trabalho *Multipoint Data Communication System With Collision Detection.* Foi a oficialização do Ethernet.

### 2.1.2 CSMA/CD

Uma rede de telecomunicações que usa um meio de comunicação partilhado por todos os terminais, necessita obrigatoriamente de possuir um mecanismo que regule o acesso ao meio. A figura (2.3) mostra uma arquitetura de rede Ethernet usando uma impressora e um servidor, partilhados por quatro terminais.

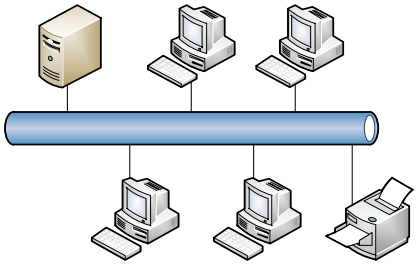


Figura 2.3: Rede Ethernet

Com o uso do protocolo CSMA/CD a gestão do acesso ao meio é distribuída por todos os seus terminais. Todos os computadores recebem toda a informação que atravessa a rede, contudo apenas processam esta se lhes for dirigida. Sabem que lhes é dirigida através da comparação do seu endereço com o endereço destino da trama. Quando um terminal quer enviar informação, a primeira ação que toma é escutar o meio (*Carrier Sense*), contudo este protocolo de escuta pode ter algumas variações. Quando o protocolo é 1-persistente, o terminal após detetar o meio ocupado, escuta este constantemente até o detetar livre, e após isto envia a sua trama imediatamente (com probabilidade de 1). No caso de protocolos de escuta p-persistentes, o terminal após detetar o meio livre, envia a sua trama com uma probabilidade de p e atrasa o envio desta para o próximo intervalo de tempo com uma probabilidade de 1-p. Para além destes existe ainda o protocolo de *Carrier Sense* não-persistente, neste caso quando o terminal deteta o meio ocupado após uma primeira escuta pontual, espera um tempo aleatório até voltar a efetuar uma nova escuta pontual. Em [19, pág 251-252] pode ser encontrada comparações de eficiência entre a tecnologia Aloha e a tecnologia CSMA com diferentes valores da persistência p. O mecanismo de Carrier Sense 0.01-persistente foi o que apresentou melhores resultados, com uma eficiência próxima de 100% para valores de G superiores a 3, ou seja mais de 3 tramas geradas num *frame time.*

Como o tempo de transmissão de uma trama não é nulo e existe uma considerável distancia entre os terminais da rede, existe ainda a possibilidade de ocorrer uma colisão, ao fim da transmissão ser iniciada. A colisão é detetada por parte do terminal emissor, se este verificar que a informação que está no meio não corresponde à trama que está a enviar. Após a deteção de colisão, o terminal emissor aborta o envio da trama, envia uma sinal *jam* para que todos os terminais saibam que ocorreu uma colisão, e espera um tempo aleatório até retransmitir

a informação, definido por um algoritmo de recuo binário truncado[2]. O sinal de *jam* enviado pelo terminal emissor é um sinal de 32 bits, que pode assumir alguns padrões diferentes e tem como objetivo garantir que todos os terminais descartarão a atual trama devido ao seu campo Cyclic Check Sequence (CRC)[3]. Esta funcionalidade de *Colision Detection* permite que todos os terminais detetem a colisão ainda antes do final do envio da trama, havendo assim uma poupança de tempo e largura de banda.

Para garantir que todas os terminais emissores detetam a colisão é necessário que o tempo mínimo de transmissão de uma trama seja maior que o *roun-trip delay*, sendo este, duas vezes o tempo que uma trama demora a efetuar o percurso mais longo da rede (entre os dois terminais mais afastadas). É esta dependência que restringe a distancia máxima da rede e o tamanho mínimo das tramas para uma dada taxa de transmissão (*slot time*).

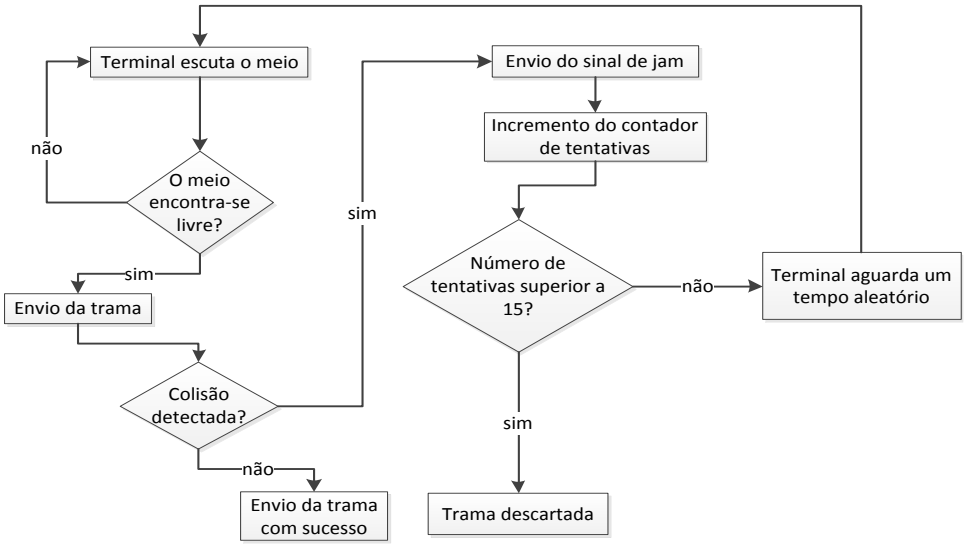



Figura 2.4: CSMA/CD

Quando acontecem 16 colisões consecutivas para a mesma transmissão, a trama é descartada. Isto pode acontecer se a rede estiver sobrecarregada por um considerável intervalo de tempo. No caso de não ocorrer colisão e a transmissão ser efetuada com sucesso, é forçado um intervalo mínimo de tempo para ser iniciada uma nova transmissão, o Inter Frame Spacing (IFS), este tempo é necessário para que as placas de rede retornem ao seu estado inicial, e permite um justo acesso ao meio, já que impossibilita que um terminal envie tramas de forma contínua [21, pág 402]. A figura (2.4) ilustra o diagrama de operações do protocolo CSMA/CD.

Em [20] é provado que a eficiência do CSMA/CD tende para 100% à medida que o tamanho das tramas usadas aumenta. Para esta prova, assume-se que o meio está dividido em blocos

[2]Este algoritmo causa o efeito de captura de canal, e foi melhorado com o Institute of Electrical and Electronics Engineers (IEEE) 802.3w, contudo a alteração não chegou a ser completamente implementada [17, pág 67-70].

[3]O CRC será descrito mais tarde.

de tempo e que uma trama demora k blocos para ser transmitida[4]. Assim a eficiência é dada por,

$$S_{th} = \frac{k}{k+x}. \tag{2.3}$$

A variável x representa o número esperado de blocos de tempo não usados na transmissão de uma trama. Sendo Y uma variável aleatória, que representa o número de blocos necessário para transmitir uma trama com sucesso,

$$P(Y = m) = \beta(1-\beta)^{m-1} \tag{2.4}$$

Onde $\beta$ representa a probabilidade de sucesso para um bloco de tempo. Neste caso a sequência de probabilidades de $P(Y = m)$ será uma progressão geométrica. Assumindo que N representa o número de terminais da rede e p representa a probabilidade de cada terminal iniciar uma transmissão num dado bloco de tempo,

$$\beta = Np(1-p)^{N-1}. \tag{2.5}$$

Visto a variável aleatória Y possuir uma distribuição geométrica, o seu valor esperado será,

$$E[Y] = \sum_{m=0}^{\infty} m\beta(1-\beta)^{m-1} = \frac{\beta}{1-\beta}\sum_{m=0}^{\infty} m(1-\beta)^m = \frac{\beta}{1-\beta}\frac{1-\beta}{\beta^2} = \frac{1}{\beta}. \tag{2.6}$$

Definindo X como a variável aleatória que representa o número consecutivo de slots desperdiçados, o valor médio de X será,

$$E[X] = x = E[Y] - 1 = \frac{1-\beta}{\beta} = \frac{1-Np(1-p)^{N-1}}{Np(1-p)^{N-1}} \tag{2.7}$$

Logo a eficiência da transmissão fica:

$$S_{th} = \frac{k}{k+x} = \frac{k}{k + \dfrac{1-Np(1-p)^{N-1}}{Np(1-p)^{N-1}}} \tag{2.8}$$

Aumentar a eficiência equivale a minimizar x, e consequentemente aumentar $\beta$, ou seja, a probabilidade de sucesso para um bloco de tempo. Igualando a zero a derivada de $\beta$ em ordem a p:

$$\frac{\partial\beta}{\partial p} = N(1-p)^{N-1} - \frac{Np(1-p)^{N-1}(N-1)}{1-p} = 0 \tag{2.9}$$

A equação(2.9) indica que o máximo de de $\beta$ acontece quando $p = \dfrac{1}{N}$. Sabendo que $(1 - \dfrac{1}{N-1})^{N-1}$ tende para $\dfrac{1}{e}$, quando $N \longrightarrow \infty$, a taxa de transferência fica:

$$S_{th} = \frac{k}{k+e-1} \tag{2.10}$$

[4]Note-se que esta prova usa novos parâmetros independentes dos usados em 2.1.1.

A eficiência aproxima-se de 100% à medida que $k \to \infty$, ou seja à medida que uma trama necessita de mais blocos de tempo para ser transmitida. Para aplicar o resultado obtido, assume-se k como o tempo de transmissão de 1 byte, e desta forma, obtém-se uma eficiência máxima de 99,9% para as tramas Ethernet II de tamanho máximo (1500 bytes de dados), e uma eficiência mínima de 97,4% para as tramas de tamanho mínimo (46 bytes de dados).

### 2.1.3 Normas Ethernet

Em 1983 o IEEE publica a norma IEEE 802.3, com o título *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications* [22, pág 48]. Desde esse evento até aos dias de hoje a tecnologia sofreu consideráveis alterações, que foram traduzidas em expansões da norma original 802.3. A tabela que segue salienta algumas expansões da norma.

| | Ano | Norma | Meio físico | Taxa de transmissão |
|---|---|---|---|---|
| 10Base5 | 1983 | 802.3 | Cabo coaxial grosso | 10 Mbps |
| 10Base2 | 1985 | 802.3a | Cabo coaxial fino | 10 Mbps |
| 10BaseT | 1990 | 802.3i | Pares de cobre | 10 Mbps |
| 100BaseT | 1995 | 802.3u | Pares de cobre | 100 Mbps |
| 1000BaseT | 1999 | 802.3ab | Pares de cobre | 1 Gbps |

Tabela 2.1: Normas Ethernet

A primeira coluna da tabela (2.1) indica o nome comum usado para definir as normas, e indica algumas das características físicas da tecnologia. O primeiro número define a taxa máxima de transmissão em Mega bits por segundo, Base indica que a a frequência de operação se encontra em banda base, e o ultimo algarismo, no caso de ser um algarismo, indica o comprimento máximo de um segmento da rede usado como meio de transmissão, o cabo coaxial (5 indica 500 m), no caso de ser uma letra indica o meio de transmissão (T significa *twisted pair*).

O 10Base5 usa segmentos de cabo coaxial, com comprimentos máximos de 500 m. Visto apenas permitir o uso de 4 hubs/repetidores, a extensão máxima da rede atinge os 2500 m. O *round-trip delay* é de 51,2 $\mu$s, usando tramas com tamanho mínimo de 512 bits [23]. Nesta norma é necessário o uso de um *transceiver* para ligar os terminais ao cabo coaxial partilhado.

A norma 10Base2 dispensa o uso do *transceiver* e assim o cabo coaxial pode ser diretamente ligado aos terminais através de um Bayonet Neill-Concelman (BNC). Neste caso o comprimento máximo dos canais coaxiais é reduzido de 500 m para 185 m.

Na norma 10BaseT é usado um cabo Unshielded Twisted Pair (UTP) e conectores RJ-45. Nesta norma os terminais encontram-se, geralmente, ligados diretamente a *hubs*, este facto não acarreta mudanças ao nível do funcionamento da rede. O *hub* é um dispositivo com múltiplas portas que opera apenas no nível físico. Recebe uma trama através de uma porta e reencaminha-a para todas as outras portas. O sinal recebido para além de repetido sofre também uma regeneração. Com a utilização de *hubs* o domínio de colisão continua inalterado.

Antes de surgir o Fast Ethernet, assistiu-se à introdução no mercado de *bridges/switchs*[5], que levaram a uma significativa mudança na topologia lógica das redes. Também a introdução

[5]Os equipamentos *switchs* serão abordados mais tarde.

do modo de operação *full-duplex* acarretou algumas alterações ao funcionamento do Ethernet. Neste modo de operação é possível estabelecer ligações ponto-a-ponto entre dois terminais, aumentando a largura de banda agregada da comunicação para o dobro, já que cada terminal pode enviar e receber tramas simultaneamente

No Fast Ethernet ou 100BaseT a extensão máxima do cabo foi diminuída para 100 m e foi adicionada à tecnologia a funcionalidade de *autonegotiation*. Esta funcionalidade permite a adaptação automática dos terminais, à máxima taxa de transmissão que ambas suportam.

A Gigabit Ethernet ou 1000BaseT foi sobretudo desenhada para operar nas redes de transporte em modo *full-duplex*, não ficando assim sujeita às restrições provenientes do protocolo CSMA/CD. O uso desta tecnologia sobre fibra ótica (1000Base-LX ou 1000Base-SX) limita a funcionalidade de *autonegotiation* a negociar apenas entre configurações de funcionamento [17, pág 85-96].

### 2.1.4 OSI

A tecnologia Ethernet opera sobre as duas camadas OSI mais baixas, a camada física e a camada de ligação. E subdivide estas em três subcamadas: física, controlo de acesso ao meio e controlo das ligações lógicas. A figura seguinte representa esta divisão.

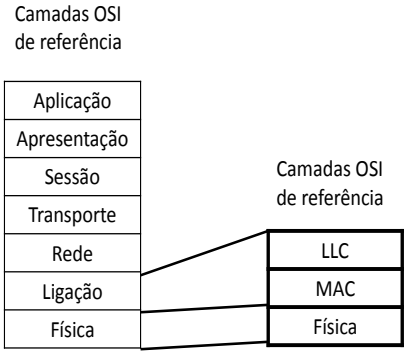



Figura 2.5: Camadas de referencia Ethernet

A subcamada física define a cablagem e os sinais elétricos, sendo responsável pela transmissão e receção dos sinais da rede. Estas funcionalidades aparecem implementadas na placa de rede ou Network Interface Controller (NIC). Cada placa de rede possui um endereço Media Acess Control (MAC) de 6 bytes.

A subcamada MAC define as funções de troca de informação e deteção de erros, e é independente do meio físico. As principais funções desta subcamada são o encapsulamento de informação, construção das tramas a enviar, deteção de erros nas tramas recebidas e controlo de acesso através do protocolo CSMA/CD. Por fim, a subcamada Logical Link Control (LLC) é responsável pela comunicação entre a camada MAC e as camadas superiores e engloba mecanismos de controlo, tal como o Destination Service Access Point (DSAP) que identifica o serviço no terminal destino da trama, o Source Service Access Point (SSAP) que identifica o serviço no terminal origem da trama e o CTL [23] que é um byte de controlo.

### 2.1.5 Trama

Nesta tecnologia a informação trocada está organizada na forma de tramas. A figura (2.6) representa as duas tramas mais usadas na tecnologia Ethernet. Note-se que o facto, da trama IEEE 802.3 representar uma norma impõe o fabrico de equipamentos compatíveis com esta, e torna a versão Ethernet II uma opção [22, pág 156].

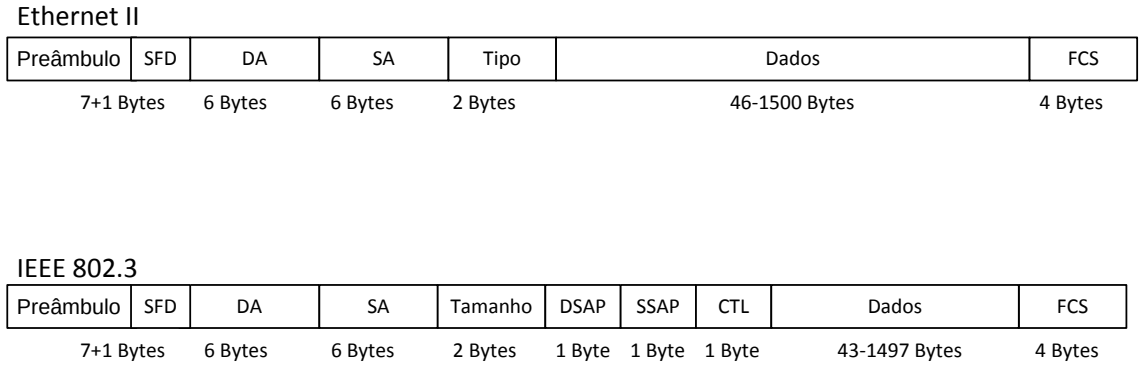



Figura 2.6: Tramas Ethernet

O campo do preâmbulo consiste em sete bytes, que contêm bits 1s e 0s alternados, com o objetivo de anunciar que será transmitida uma trama, permitindo aos recetores sincronizarem-se. O campo *Start Of Frame* aparece no final do preâmbulo e consiste num byte com a sequência 10101011 que marca o inicio da trama [22, pág 156].

O Destination Address (DA) e Source Adress (SA) são os endereços físicos ou endereços MAC dos terminais da rede. Possuem 6 bytes e quem atribui os endereços MAC aos equipamentos são os seus fabricantes, sendo estes endereços únicos. Para organizar esta distribuição de endereços, o IEEE-Standards Association (IEEE-SA) atribui a cada empresa um Organizationally Unique Identifier (OUI), que corresponde aos três bytes mais significativos do endereço.

Uma assinalável diferença entre estas duas tramas reside no campo que segue o SA. Enquanto que na trama Ethernet II, este campo possuí um valor superior a 1500 e representa o protocolo e camada OSI três ao qual o campo de dados pertence, na trama IEEE 802.3 possuí um valor inferior a 1500, e indica o tamanho do campo de dados em bytes. A seguir a este campo a trama IEEE 802.3 implementa as funções LLC, já referidas anteriormente, através dos campos DSAP, SSAP e CTL.

O campo de dados contém a *payload* , ou seja a informação fundamental que se pretende enviar e o seu tamanho pode variar entre 46 a 1500 ou 1497 bytes. O tamanho mínimo de 46 bytes juntamente com a distancia mínima entre quaisquer dois terminais, assegura a deteção de colisões. Se a carga a transportar for menor que 46 bytes, então o campo é preenchido com bits de *padding* até atingir os 46 bytes.

O Frame Check Sequence (FCS) é o ultimo campo e garante a integridade de toda trama, através do seu CRC que é um número único gerado por um algoritmo polinomial que usa todos os bits da trama. Se o CRC não estiver correto a trama contém anomalias, pelo que será descartada.

## 2.2 Switched Ethernet

A massificação do uso das primeiras tecnologias Ethernet a 10 Mbps rapidamente fez baixar os custos dos equipamentos e aumentar a necessidade de largura de banda, o que levou fabricantes de equipamentos a procurar soluções que colmatassem esta necessidade. Assim, apareceram as *bridges*. Estas permitiam ligações lógicas independentes entre diferentes segmentos da rede, funcionando cada segmento como uma ligação Ethernet independente, com um domínio de colisão próprio.

As *bridges*, apesar de satisfazerem as referidas necessidades das redes Ethernet, possuíam alguma limitações, das quais se salienta a impossibilidade de reencaminhar mais que uma trama em simultâneo e a limitação do número de portas do dispositivo. Estas limitações resultavam do facto da *bridges* operar sobretudo no domínio do software. Mais tarde são introduzidos os *Intelligent Switching Hub*, que já permitem o reencaminhamento de mais que uma trama em simultâneo (*parallel switching*) [24, pág 131] e é aqui que se assiste à mudança de termo *bridge* para *switch*. Uma das empresas que mais contribuiu para esta mudança foi a Kalpana (mais tarde adquirida pela CISCO) no inicio dos anos 90 [25, pág 130].

A figura (2.7) mostra a arquitetura de rede Switched Ethernet. Embora a utilização do *hub* apresente ligações físicas semelhantes, o *swicth* garante ligações dedicadas a cada terminal também ao nível lógico, o que não acontece nos *hubs*.

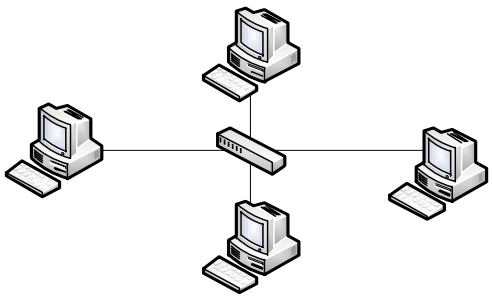


Figura 2.7: Switched Ethernet

### 2.2.1 Funcionamento do swicth

O *switch*, tal como o *hub* possui múltiplas portas, contudo opera na camada de ligação (segunda camada OSI). Funciona registando os endereços MAC alcançáveis por cada uma das suas portas numa tabela de encaminhamento. Quando um *switch* recebe uma trama numa porta de entrada, regista na sua tabela de encaminhamento a porta que recebeu a trama e o endereço MAC origem da trama e procura o endereço MAC destino da trama na sua tabela. Se o endereço MAC destino da trama existe na tabela de encaminhamento, o *switch* envia a trama pela porta associada a esse endereço de destino (mecanismo de *forwarding*). No caso de o endereço destino da trama não constar na tabela de encaminhamento, o *switch* envia a trama por todas as portas, exceto pela porta pela qual a trama foi recebida (mecanismo de *flooding*). A imagem (2.8) ilustra o envio de uma trama P do terminal 1 para o terminal 2, e consequente trama de resposta R do terminal 2 para o terminal 1. Observe-se o facto de inicialmente os *switchs* não possuírem informações sobre nenhum dos terminais nas suas

tabelas de encaminhamento.

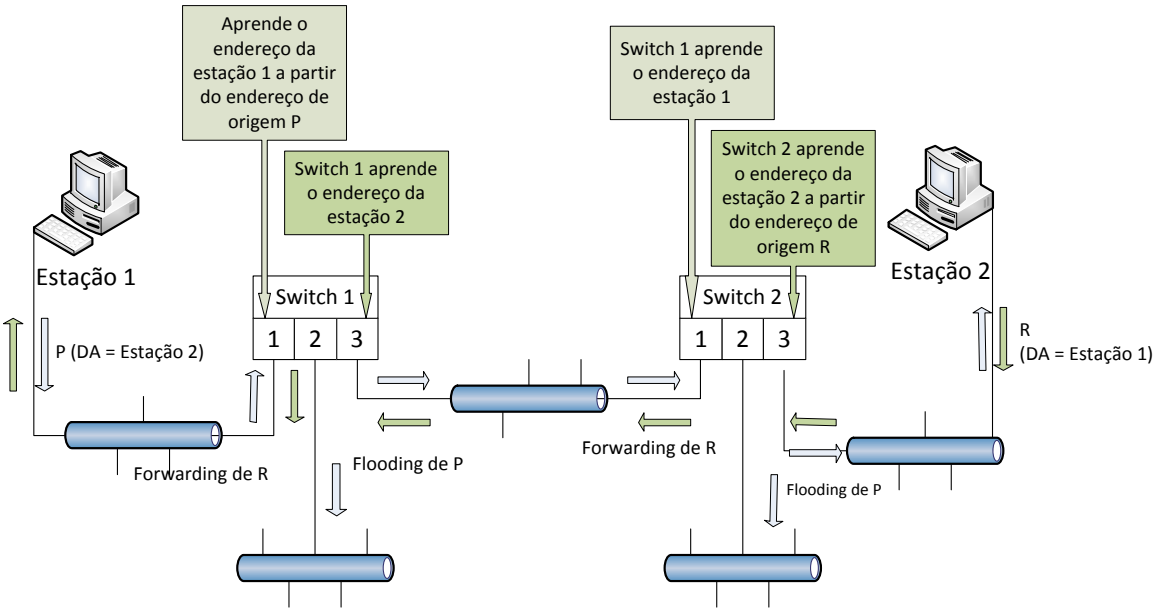



Figura 2.8: Operações de *Switching*. É ilustrado o envio de uma trama P do terminal 1 para o terminal 2, e consequente trama de resposta R do terminal 2 para o terminal 1. Inicialmente os *switchs* não possuem informações sobre nenhum dos terminais nas suas tabelas de encaminhamento, desta forma a trama P é encaminhada através de mecanismos de *flooding*. À medida que P é reencaminhada os *switchs* adquirem informação sobre a localização do terminal 1. Este mecanismo de aprendizagem leva a que o posterior encaminhamento da trama R, seja efetuado através de *forwardings*.

### 2.2.2 Switchs

O facto do *switch* efetuar operações no domínio do hardware, leva a que este apresente naturais vantagens sobre a *bridge*. Geralmente o *switch* tem um maior número de portas e logo um custo por porta inferior, possui taxas de transferência mais elevadas, atrasos inferiores e ainda possibilita *parallel switching* [25, pág 137]. A performance do *switching* por software depende do número de portas por micro-processador, da frequência de processamento e da qualidade do código. Devido às limitações de processamento de tramas por segundo o *switching* por software limita o número de portas entre 6 a 8 [25, pág 140]. A vantagem obvia deste tipo de operação reside na facilidade de efetuar alterações/atualizações no dispositivo através de software.

O *switching* por hardware pode ser efetuado através de um processador indiferenciado ou através de Application Specific Integrated Circuitss (ASICs). Os ASICs são chips desenhados para efetuar as operações pretendidas em hardware e estão aptos a processar tramas à taxa máxima a que são transmitidas na rede [25, pág 141]. Embora o seu desenvolvimento seja dispendioso, a produção destes chips em massa torna-os económicos e apelativos para o mercado. São produzidos, geralmente, em duas etapas, na primeira o código que se pretende

implementar é integrado num Field Programmable Gate Array (FPGA), e após a otimização do código, os FPGAs são substituídos por ASICs. A produção destes chips acarreta um elevado risco para os fabricantes e o modo de fabrico varia entre o uso de um ASIC por porta ou um ASIC único central. Comparativamente ao *switching* usando um processador indiferenciado, o uso de ASICs fundamentalmente, aumenta a taxa de transmissão e o Mean Time Between Failure (MTBF), porque reduz o número de componentes em hardware do *switch*. As desvantagens são o custo, o risco de produção e a dificuldade em implementar modificações no modo de operação.

Os *switchs* podem também ser classificados quanto ao modo de funcionamento, em *cut-through switchs* ou *store-and-forward switchs*. A diferença reside no facto dos *cut-through* iniciarem o reencaminhamento da trama após terem recebido os seis bytes correspondentes ao DA, enquanto os *store-and-forward* apenas reenviarem a trama após a terem recebido na totalidade. Naturalmente os primeiros apresentam uma maior taxa de reencaminhamento, contudo não efetuam uma filtragem total, podendo reencaminhar tramas com erros, desperdiçando largura de banda e prejudicando o desempenho da rede. Uma das soluções utilizadas para resolver este problema, é avaliar a validade da trama, enquanto está a ser transmitida e no caso de uma elevada quantidade de tramas provenientes da mesma porta conterem erros, o *switch* bloqueia a porta ou altera o modo de operação da porta para *store-and-forward* [25, pág 144-145].

Outra classificação, que pode ser usada para caraterizar os *switchs*, relaciona estes com a camada OSI ou o protocolo da camada OSI, em que operam. Existem os *Sinlge MAC layer switchs*, que operam apenas numa tecnologia da segunda camada OSI, tal como os Ethernet *switchs*, Token Ring *switchs*, ou FDDI *switchs*. Os *Multiple MAC layer switches*, que operam com tecnologias diferentes dentro da segunda camada OSI e no caso de reencaminharem uma trama para uma porta com tecnologia diferente da porta origem, efetuam a conversão da trama[6]. Por fim existem ainda os *Multilayer switchs*, que operam na segunda e terceira camada do modelo OSI, são assim a junção de um *swicth* e um *router*. Dentro da mesma rede Ethernet efetuam o reencaminhamento através do seu endereço de camada dois do modelo OSI e quando o pacote tem como destino um terminal fora da rede, o reencaminhamento é efetuado com base no endereço de terceira camada, geralmente IP.

### 2.2.3 Full-Duplex

O modo de operação em *full-duplex* é apenas possível em ligações ponto-a-ponto, já que ambos os terminais necessitam de enviar e receber tramas simultaneamente. Neste cenário as ligações são dedicadas e o conceito de domínio de colisões deixa de fazer sentido, já que as colisões deixam de existir, o que torna desnecessário o uso do protocolo CSMA/CD. A capacidade total agregada da comunicação passa a ser o dobro da prevista na norma da tecnologia. A figura (2.9) mostra a ligação física entre os terminais de receção e envio Ethernet, de modo a ser possível a operação em *full-duplex*:

Neste modo de operação o comprimento máximo de uma ligação deixa de ser limitado pelos requisitos de tempo (*round trip delay*) presentes na Ethernet em meio partilhado. Assim o comprimento das ligações é apenas limitado pelas capacidades do meio físico, o que representa uma vantagem importante no uso de fibra ótica como meio físico

O *full-duplex* encontra-se especificado na norma 802.3x aprovada em março de 1997 e, para

[6] As *bridges* com esta funcionalidade são apelidadas de *translating bridges* [24, pág 127-128]

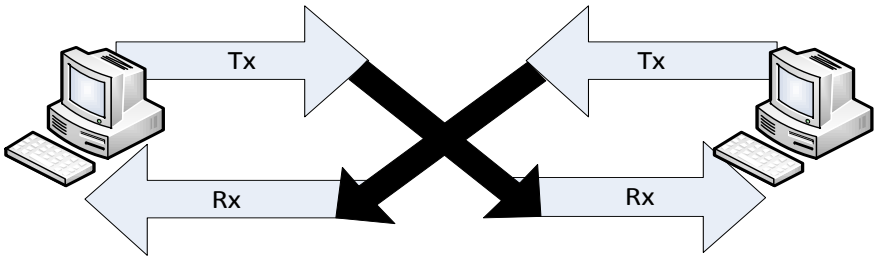



Figura 2.9: Ligação em *full-duplex*

além de descrever o modo de funcionamento, aborda também os mecanismos de controlo de ligações MAC *Control* e *PAUSE*. Apesar do protocolo CSMA/CD não ser usado, os terminais continuam a cumprir o tempo de IFS entre transmissões, pelo mesmo motivo que era usado na primeira Ethernet, devido à eletrónica usada.

Embora o uso de *full-duplex* tenha a capacidade de dobrar a largura de banda das ligações Ethernet, geralmente não aumenta significativamente a performance de computadores pessoais, já que a maior parte dos protocolos de redes estão desenhados de modo a enviar informação e esperar por um sinal de reconhecimento [17]. Isto leva a comunicações com tráfego assimétrico onde quase toda a informação é enviada por um canal, e o outro limita-se a transportar pouco mais que sinais de reconhecimento. Contudo, esta funcionalidade é bastante útil quando aplicada numa rede de transporte, já que neste caso a rede agrega várias comunicações e o tráfego transportado é bastante mais simétrico e homogéneo.

Relativamente à configuração dos terminais para operar em *full-duplex*, é aconselhável que estes usem sempre que possível a funcionalidade de *autonegotiation*[7] para este fim. Em [17, pág 80] são abordadas algumas questões relativas à configuração desta funcionalidade.

### 2.2.4 Controlo de fluxos

O *MAC Control Protocol*, presente na norma IEEE 802.3x [26], é responsável pelo controlo em tempo real dos fluxos de tráfego sendo um dos seus objetivos evitar congestionamentos na rede. Este protocolo atua através do uso de tramas de controlo, que são caraterizadas pelo valor 0x8808 em hexadecimal, no campo "tipo"da trama. Um terminal ao receber uma trama que contenha 0x8808 no campo tipo, vai procurar o código da operação a efetuar, contido no campo de dados. Se o campo tipo não contiver este valor, a trama é reencaminhada normalmente. Os *opcodes* (*operational codes*) localizam-se nos primeiros dois bytes do campo de dados, no entanto o campo necessita de cumprir o comprimento mínimo de 46 bytes [17, pág 83-84].

#### Full-Duplex

A PAUSE é uma trama de controlo importante que está presente no *MAC Control Protocol* e tem como função pausar a transmissão de um terminal, por um tempo determinado, sendo apenas usada em cenários de *full-duplex* Switched Ethernet. O conceito de funcionamento é

[7] Esta funcionalidade será abordada mais tarde.

simples, após ser detetado um congestionamento no terminal receptor, este envia uma trama de controlo com o comando PAUSE, para o terminal emissor.

O *opcode* do comando PAUSE é o 0x0001 e o campo DA da trama de controlo contém o endereço multicast reservado 01-80-C2-00-00-01. Este endereço foi selecionado dentro de uma gama de endereços reservados na norma IEEE 802.3D (*Spanning Tree*), que especifica operações com *switchs*. Desta maneira, o endereço não será interpretado como multicast por parte do *swicth*. Uma trama PAUSE para além do *opcode* respetivo, contém também o período da pausa que está a pedir. Este tempo é medido em quantidades de pausa (*quanta*) onde cada unidade é igual ao tempo de propagação de 512 bits. A gama de valores varia entre 0 e 65.535 unidades, já que usa 2 bytes do campo de dados [17]. A imagem (2.10) ilustra a trama PAUSE. Apesar do terminal que recebe esta trama inibir as suas transmissões num dado período de tempo, esta pode ainda enviar tramas de controlo PAUSE caso também sofra um congestionamento [17].

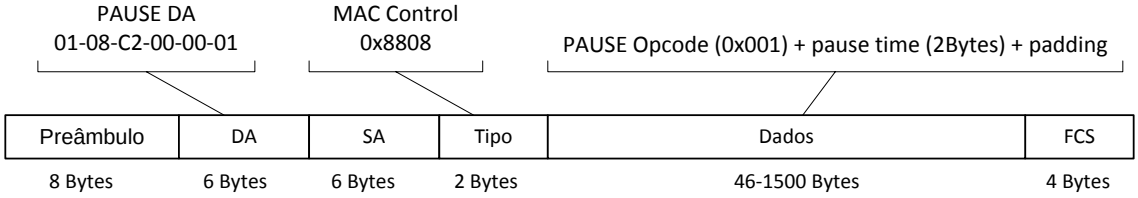



Figura 2.10: Trama PAUSE

### Half-Duplex

No caso de uma rede *half-duplex* Switched Ethernet, ou mais especificamente no caso de portas *switch* a operarem em *half-duplex*, não existe o mecanismo de PAUSE, já que este é apenas usado em ligações ponto-a-ponto. Desta forma o controlo de fluxos será efetuado com recurso a funcionalidades do protocolo CSMA/CD. Existem dois mecanismos para controlo de fluxos em *half-duplex*, o *Backpressure Flow Control* e o *Carrier Extension Flow Control* [27, pág 186-191].

***Backpressure Flow Control:***

Quando o congestionamento ocorre numa porta de entrada de um *switch* a porta induz uma colisão, o que leva os terminais emissores a abortar o envio e esperar um tempo de *backoff* até tentarem retransmitir novamente. Se ocorrer um novo congestionamento o mecanismo repete-se. Se o congestionamento ocorrer numa porta de saída do *switch*, as portas de entrada ao verificarem que o endereço de destino das tramas que estão a receber está associado ao porto congestionado, vão sinalizar colisões nas tramas destinadas à porta congestionada. O terminais emissores esperarão um tempo de *backoff* até tentarem reenviar as tramas. Se for detetado novo congestionamento, o método repete-se.

Neste método o tempo de *backoff* pode aumentar exponencialmente com o aumento do número colisões de acordo com o algoritmo de recuo binário exponencial truncado. Isto leva a períodos, de menor congestionamento onde os terminais não podem enviar tramas devido ao elevado tempo de *backoff*. No caso de serem induzidas 16 colisões, a trama é descartada,

o que acarreta consequências mais nefastas para o desempenho da rede.

***Carrier Extension Flow Control:***
Neste método o *switch* informa os terminais de congestionamento nas comunicações, através da sinalização do meio como ocupado. Cada terminal antes de enviar uma trama, verificará se o meio está ocupado através do seu mecanismo de *Carrier Sense*. Este método não necessita de forçar colisões, pelo que se evita o problema dos tempos de *backoff* [27, pág 190].

### Bloqueio Head-of-Line

O recurso a mecanismos de controlo de fluxos ao mesmo tempo que diminui o congestionamento de uma rede, também agrava o problema conhecido como Head-of-Line Blocking (HOL Blocking) [27, pág 188]. A imagem (2.11) ilustra este problema.

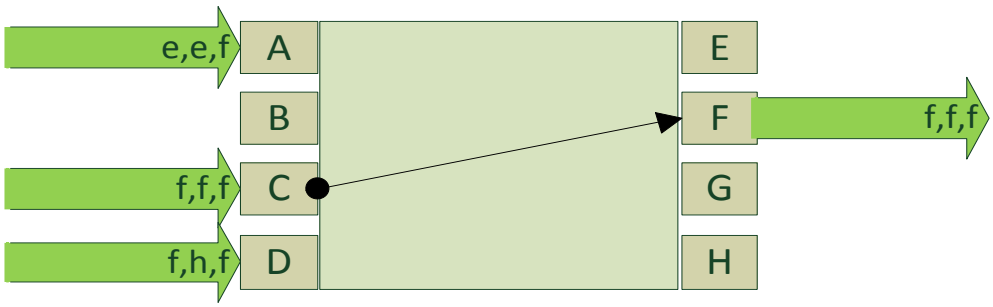



Figura 2.11: *Head-of-Line Blocking*

A porta (de saída) F está congestionada, pelo que todos as portas de entrada que possuem tramas destinadas à porta F vão acionar os mecanismos de controlo descritos anteriormente. Contudo algumas das portas de entrada possuem tramas destinadas a portas de saída livres que também não poderão ser enviadas. Por exemplo o envio de tramas destinadas à portas livres E e H pelas portas A e D, respetivamente, também sofrerá atrasos de transmissão devido ao congestionamento da porta F. Este tipo de bloqueio prejudica bastante o desempenho da rede.

## 2.2.5 Auto-Negotiation

A funcionalidade de *autonegotiation* tem como principal objetivo o reconhecimento automático da taxa máxima de transmissão a que os terminais de uma ligação Ethernet operam, assim é possível sincronizar ambos os terminais para funcionar à mesma taxa de transmissão. Além desta função a *autonegotiation* implementa também outros mecanismos, tal como o reconhecimento do modo de operação em half-duplex ou full-duplex, ou a indicação de falhas.

A funcionalidade foi publicada pela primeira vez em 1995 como parte da norma 802.3u que define o Fast Ethernet. A tecnologia teve como destino as tecnologias Ethernet sobre pares de cobre, pelo que nas atuais implementações sobre fibra ótica a funcionalidade, quando usada, apenas implementa parte das suas funcionalidades

A *autonegotiation* atua apenas sobre ligações ponto-a-ponto e é iniciada imediatamente após ocorrer a ligação física, tendo prioridade sobre qualquer transmissão de dados e sendo independente de qualquer tecnologia Ethernet sobre cobre, já que utiliza um sistema de sinalização próprio [17, pág 85-96].

O sistema de sinalização usa sinais Fast Link Pulse (FLP) baseados nos Normal Link Pulse (NLP) que são usados na verificação da integridade da ligação, na tecnologia 10Base-T. O sinal consiste no envio de 33 bits, onde os 17 bits ímpares contêm a informação de *clock* e os 16 bits pares contêm dados. A figura que segue mostra os 16 bits de dados, numerados de D0 a D15.

| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D11 | D12 | D13 | D14 | D15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Selector Field | | | | | Tecnology Ability Field | | | | | | | | RF | Ack | NP |

Figura 2.12: Mensagem de *autonegotiation*

O *Seletor Field*, ocupa os cinco bits mais significativos e representa a tecnologia LAN que é usada por aquele equipamento, tornando assim a funcionalidade apta a ser implementada em outras tecnologias para além do Ethernet. Os oito bits seguintes contêm o *Tecnology Ability Field* e representam todas as tecnologias/funcionalidades que o dispositivo possui, tal como *full-duplex* ou operação de PAUSE. A existência destas funcionalidades é assinalada através da presença do nível lógico '1' no bit correspondente à funcionalidade.

O bit D13 é o *Remote Fault Indicator* e tem como objetivo indicar uma falha na ligação. Se por exemplo um dispositivo não receber informação no seu Rx, envia o bit D13 a '1' para relatar este evento. O bit D14 é o *Acknowledge* que é enviado quando a mensagem de 16 bits da *autonegotiation* é recebida corretamente três vezes consecutivas. Esta implementação visa diminuir possíveis erros na mensagem de *autonegotiation*. Após a receção deste sinal, a *autonegotiation* está completa. O ultimo bit é o *Next Page* e permite complementar a mensagem atual com mensagens adicionais, podendo assim serem transmitidas outras informações relativas a protocolos específicos do fabricante.

A tecnologia 10BaseT e outras mais antigas, que não suportam FLP, interpretam a mensagem como uma mensagem NLP. E o dispositivo que suporta FLP identificará a tecnologia de 10Base com quem está a comunicar através do mecanismo de *Parallel Detection*. Este mecanismo é ativado quando a funcionalidade de *autonegotiation* não deteta FLPs provenientes do terminal com que o primeiro se encontra a comunicar. O *Parallel Detection* garante assim a interoperabilidade da *autonegotiation* com tecnologias mais antigas que 100BaseT.

### 2.2.6 Spanning Tree Protocol

Um dos fatores de avaliação de uma rede de telecomunicações, é a sua capacidade para recuperar a falhas de forma autónoma. Para que uma rede possa recuperar de possíveis cortes nas ligações ou avarias nos nós, necessita de possuir uma configuração redundante, onde existe mais do que um caminho para a mesma ligação entre dois terminais. Ao ser usada uma configuração redundante numa rede Switched Ethernet, o problema imediato que surge é o *broadcast storm* proveniente dos sucessivos *floodings* de tramas com endereço de

*broadcast*. Isto leva a um desperdício de recursos e declínio de performance da rede, já que é consumida largura da banda da rede por um tempo indefinido. Para resolver este problema, de forma a garantir capacidade de restauro à rede, a tecnologia Ethernet usa o método Spanning Tree Protocol (STP), onde através do encerramento de portas *bridge*[8] é garantido apenas um caminho de trabalho entre quaisquer dois terminais da rede.

Este mecanismo está especificado na norma 802.1D do IEEE, e para além dos requisitos já referidos anteriormente a norma possibilita ao administrador da rede a otimização do caminho de trabalho a usar. O mecanismo deve ser transparente para todos os terminais e deve necessitar do mínimo de memória e processamento por parte das *bridges* de modo a não prejudicar a performance da rede.

### Funcionamento

A figura (2.13) tem como objetivo ilustrar o funcionamento do protocolo STP. Este funcionamento pode ser dividido em três etapas. Primeira etapa. O *root bridge* é escolhido entre todas as *bridges* que pertencem à rede, sendo esta a *bridge* que estará na raiz da configuração de rede em árvore criada pelo protocolo. Se não houver uma intervenção externa por parte do administrador da rede, a *bridge* escolhida será a que possuir o menor endereço MAC. Muitas vezes é aconselhado que o administrador defina a *root bridge*, já que a esta terá uma posição central na arquitetura da rede [21, pág 551].

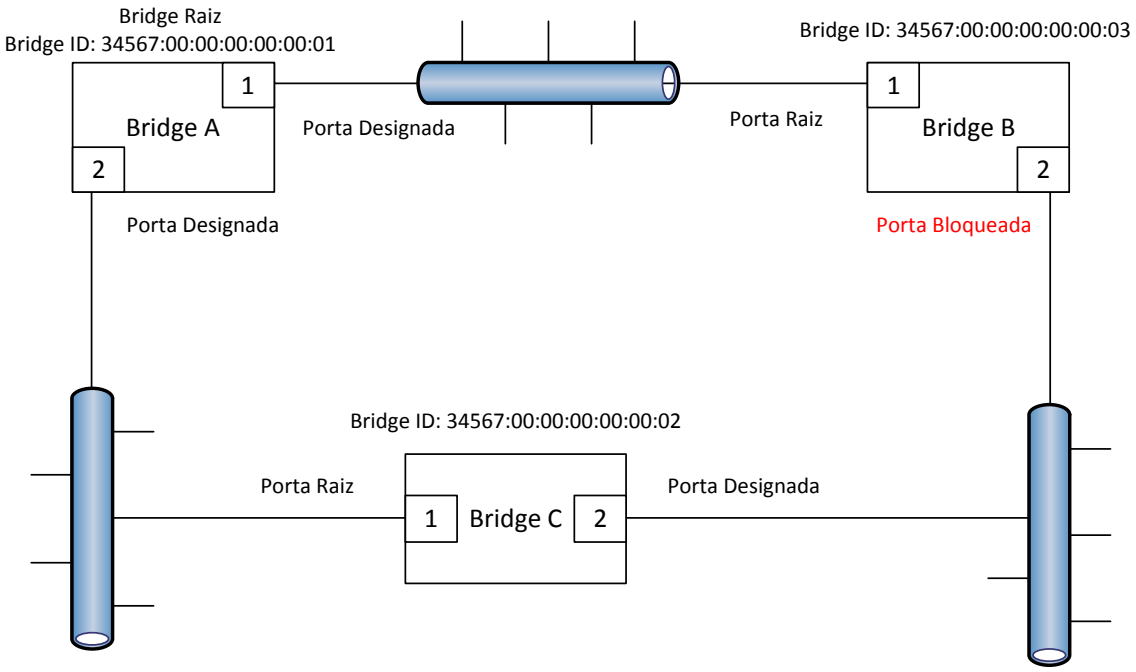



Figura 2.13: Protocolo STP

Para escolher a *root bridge*, o administrador pode configurar os dois octetos mais signi-

[8]Nesta subseção é usado o termo *bridge* para manter a concordância com a bibliografia do tema e respetiva norma definidora do STP.

ficativos da *bridge ID*, sendo o campo restante preenchido com o endereço MAC da *bridge*. Inicialmente todas as *bridges* assumem serem a *root bridge* e então enviam uma Bridge Protocol Data Unit (BPDU), com o seu endereço para todas as suas portas. Após as *bridges* compararem os endereços recebidos com o seu próprio endereço, assumem que a *root bridge* é a que possui um endereço mais baixo. Após todas reconhecerem a mesma *bridge* como *root* termina a primeira etapa. Neste caso a *root bridge* será A.

Segunda etapa. Agora cada *bridge* não raiz vai usar o algoritmo de Bellman-Ford [23], assíncrono e distribuído para calcular qual a *bridge* vizinha no percurso de custo mínimo para a *root bridge*. O conjunto dos percursos de custo mínimo definirão a estrutura da *Spanning Tree*. Cada *bridge* terá assim uma *root port*, que é responsável pelo envio/receção de informação proveniente da *root bridge*. As restantes portas da rede pertencentes ao conjunto dos percursos mínimos, mas que não são *root ports*, são designadas por *designated ports*.

No exemplo da figura, A é a *root bridge*, a porta 1 de B e a porta 1 de C são as *root ports*. As portas da *root bridge*, juntamente com a porta 2 da *bridge* C, serão as *designated ports*. Assumindo que o custo do percurso da porta 2 *bridge* C e porta 2 *bridge* B, à *root bridge* é o mesmo, o critério de escolha para a *designated port* daquele segmento de rede prende-se com o menor valor de *bridge ID* de C.

Terceira etapa. As portas que estiverem definidas como *root ports* ou *designated ports*, ficarão em modo *forwarding*, e as portas que não possuírem nenhuma destas designações ficarão em modo *blocking*. Neste ponto o protocolo STP convergiu totalmente, e a rede deixou de ser redundante. Uma porta que se encontra em modo *blocking*, passa ao estado *listening* quando recebe uma trama que notifica a necessidade de uma reconfiguração da rede [25, pág 451-452]. No estado *listening* se não receber uma BPDU que obriga a porta a retomar ao estado de *blocking* passado um tempo de *forward delay*, a porta passa ao estado *learning*. Neste estado o processo volta ao descrito na primeira etapa desta subsecção, pelo que a porta poderá passar ao estado de *forwarding* ou blocking. Para além destes estados a porta pode-se encontrar no estado *disable* onde não permite qualquer comunicação, nem mesmo mensagens de configuração, contudo só poderá assumir este estado através da intervenção do administrador da rede.

### 2.2.7 Link Aggregation

Uma funcionalidade mais recente nas redes *switched* Ethernet é a capacidade de ser agregada mais do que uma ligação física Ethernet para possibilitar ligações lógicas cuja taxa de transmissão será a soma das capacidades individuais das ligações agregadas. O *link aggregation* representa outra forma de usar redes redundantes, mas não se apresenta como alternativa ao STP, já que a funcionalidade apenas controla as ligações entre dois terminais, e por isso deve ser implementado juntamente com o STP. O STP deve olhar para um agregado de ligações (*trunk*) e interpreta-lo como uma única ligação [21, pág 555-557].

A funcionalidade adiciona à tecnologia Ethernet a possibilidade de fazer atualizações à capacidade da rede de forma linear isto é, se forem necessários 250 Mbps de largura de banda, numa ligação Fast Ethernet, esta não necessita de ser atualizada para Gigabit Ethernet, bastando apenas agregar outras ligações à atual.

A primeira norma desta funcionalidade foi a IEEE 802.3ad lançada em 2000. A norma resume e generaliza as aproximações implementadas pelos principais fabricantes de equipa-

mentos para LANs[9]. Mais tarde, em 2008, a norma foi movida para a família de normas IEEE 802.1 e foi lhe atribuída a extensão AX.

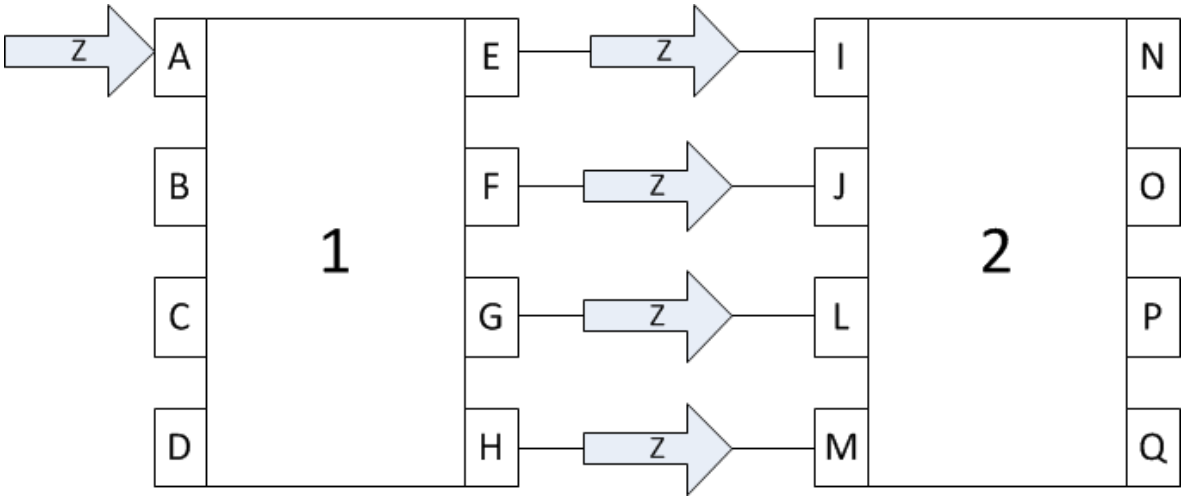



Figura 2.14: Link Aggregation

A imagem (2.14) ilustra a implementação desta funcionalidade entre dois *switchs*. Uma rede com esta topologia apresenta problemas se surgir nas portas do *switch* 1 uma trama com endereço de *broadcast* ou com endereço desconhecido, já que o *switch* vai realizar a operação de *flooding*, levando ao aparecimento de uma *broadcast storm* devido à existência de ligações paralelas.

Para resolver este problema o *switch* necessita de considerar todas as ligações físicas do agregado como uma única ligação lógica (*trunk*), e selecionar nesta ligação lógica uma ligação física para reencaminhar as tramas de endereço desconhecido e endereço *broadcast*. A funcionalidade da norma IEEE 802.3ad que descreve a criação de portas lógicas é a Link Control Aggregation Protocol (LCAP) e permite controlar topologias de *link aggregation* em mais do que dois *switchs*.

Para que se faça pleno uso da capacidade total do *link aggregation* é necessário que a distribuição de carga entre as varias ligações de um *trunk* seja equilibrada. Para isto as tramas deveriam ser reencaminhadas pela ligação que possui menor carga (menor fila de espera) à medida que chegassem (*dynamic frame distribution*). Contudo alguns protocolos de camadas superiores sofrem um forte decréscimo de performance se receberem as tramas por uma ordem diferente da ordem em que foram enviadas e esta distribuição de tramas torna real esta possibilidade, já que as tramas de um fluxo podem sofrer diferentes atrasos de propagação se forem transmitidas por diferentes ligações. Desta forma o mecanismo mais usado para a distribuição de carga é o *static frame distribution*, que reserva cada porta do *link aggregation* para o mesmo fluxo de informação estabelecido entre dois terminais. Todas as tramas de uma dada sessão passam assim pelas mesmas filas de espera e asseguram uma ordem de chegada igual à ordem de envio[10].

[9] EtherChannel da Cisco, MultiLink Trunking da Nortel e Adaptive Load Balancing da Intel.

[10] Existem ainda soluções alternativas mais rudimentares, tal como o CISCO Fast Ethernet Channel [21, pág 561]. Nesta solução é efetuado uma operação *xor* aos últimos dois bits do endereço MAC fonte, ou destino. Esta operação indicará qual a ligação que reencaminhará a trama (quatro possibilidades). Esta solução é completamente aleatória, desperdiça largura de banda e não garante uniformidade na distribuição de fluxos.

### 2.2.8 Virtual LANs

A possibilidade de se criarem redes logicamente independentes dentro de uma rede Ethernet, constitui outra das funcionalidades avançadas do *switched* Ethernet. Essa funcionalidade provém da divisão que o *switch* efetua da rede física em mais que uma rede virtual (lógica). Inicialmente a funcionalidade apareceu com o objetivo de reduzir os domínios de colisão, o que veio a tornar-se obsoleto com o aparecimento dos *switch*. No entanto traz bastantes benefícios à rede, dos quais se salienta a redução dos domínios de *broadcast*, ou a reconfiguração de redes sem necessitar de intervir na sua estrutura física.

As Virtual Local Area Networks (VLANs) podem ser configuradas através de dois métodos, *implicit tagging* ou *explicit tagging*. No *implicit tagging* o *switch* dedica a cada uma das suas portas, uma Virtual LAN, através do uso de diferentes endereços MAC para cada porta [24, pág 143-146], como se pode observar na figura que segue.

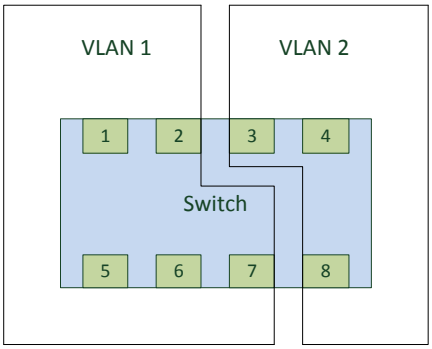



Figura 2.15: *Switch* com implementação de VLANs baseadas em MAC

A atribuição de VLANs às portas pode ser configurada pelo administrador da rede, sem alteração da topologia física. Contudo esta implementação apresenta o problema de ser necessário usar uma ligação para cada VLAN. No cenário de existirem várias VLANs num *switch* é necessário usar uma cabo para ligar cada uma dessas VLANs o que representa um desperdício de recursos.

No *explicit tagging* é necessário introduzir uma *tag* VLAN nas tramas. Apesar de existirem alguns métodos alternativos anteriores ao protocolo IEEE 802.1Q, este é atualmente o método dominante para configuração de VLANs oferecido pelos fabricantes [24, pág 145].

Neste método uma ligação pode transportar várias tramas pertencentes a diferentes VLANs, o que leva a uma melhor gestão dos recursos físicos. A *tag* VLAN é colocada ou retirada da trama Ethernet apenas pelos dispositivos *switch*, o que torna o mecanismo transparente aos terminais da rede. No caso de um terminal receber uma trama com tag VLAN é provável que esta seja descartada.

A *tag* inserida pelo protocolo IEEE 802.1Q é ilustrada em seguida:

A *tag* é inserida entre o SA e o campo tipo/tamanho.

- Os 2 bytes mais significativos da *tag* representam o Tag Protocol Identifier (TPID), que no caso de suportar o protocolo IEEE 802.1Q/802.1P, apresenta o valor hexadecimal 0x8100.

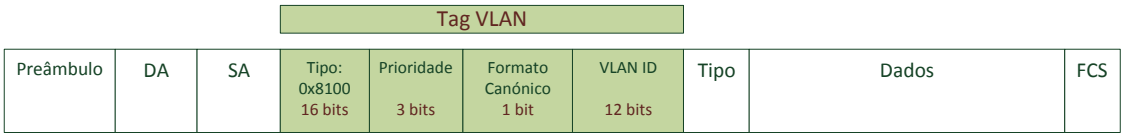



Figura 2.16: *tag* VLAN

- O campo seguinte, indica a prioridade da informação (User Priority) e tem um tamanho de 3 bits, pode assim definir até 8 níveis de prioridade.

- O Formato canónico (Canonical Format Indicator) com 1 bit, indica se a trama é compatível com tecnologia Token-Ring.

- O campo VLAN ID, contém 12 bits para identificar a VLAN da trama. Pode definir até 4094 tramas já que dois endereços estão reservados.

## 2.3 Gigabit Ethernet

A passagem do Fast Ethernet para o Gigabit Ethernet implicou alterações na camada física da tecnologia, de modo a possibilitar a operação sobre fibra ótica [27, pág 163]. Para além das alterações na camada física, também foram adicionadas as funcionalidades de *carrier extension* e *frame bursting* ao modo de operação half-duplex.

Existem dois standards na operação para o Gigabit Ethernet sobre fibra, o 1000Base-SX (*short-wavelength*) e o 1000Base-LX (*long-wavelength laser*), que juntamente com 1000Base-CX (cabos UTP) foram normalizados pelo IEEE na norma 802.3z em junho de 2008.

O 1000Base-SX opera em Multi Mode Fiber (MMF), no comprimento de onda de 850 nm e permite distancias de transmissão de 220 metros em fibras 62.5/125 nm [11]. Contudo, o uso de fibras de 50/125 nm pode aumentar a distancia máxima de transmissão para os 500 metros.

O 1000Base-LX pode operar em MMF ou Single Mode Fiber (SMF). No modo SMF usa fibras com núcleos de 9 $\mu$m e opera no comprimento de onda de 1300 nm. Embora a distancia máxima especificada seja de 2 km, muitos fabricantes conseguem atingir distancias de 20 km. Ao ser usada em modo *multimode* a distância máxima de um segmento é de 550 m [24, pág 75-77].

No Gigabit Ethernet são usados 8 bits para cada canal de transmissão/receção, sendo assim necessários, apenas processamentos de 125 MHz para atingir taxas de 1 Gbps. É também implementada uma nova codificação de canal, onde cada 8 bits de informação são codificados em 10 bits de código, incorporando 2 bits para controlo de erros. A primeira versão da tecnologia apresentava taxas de transmissão de informação de 850 Mbps [27, pág 163-164] [10, pág 50].

Apesar de inicialmente a funcionalidade de *autonegotiation* ter sido desenvolvida para operar nas tecnologias Ethernet para taxas de 10 Mbps ou 100 Mbps sobre cobre, o Gigabit Ethernet adotou, ainda que de forma parcial, esta funcionalidade para permitir a troca de

[11] Estas medidas dizem respeito ao tamanho do núcleo da fibra e do encapsulamento.

algumas informações de configuração. A estrutura da mensagem de 16 bits enviada mantém-se igual (secção (2.2.5)).

### 2.3.1 Half-Duplex

A operação em half-duplex do Gigabit Ethernet necessitou de ser reformulada devido às restrições do tamanho da rede, que este modo impõe. Se não fossem realizadas alterações ao funcionamento half-duplex, o tamanho máximo de uma rede seria de 20 m [27, pág 168-169], o que é naturalmente impraticável.

#### Carrier Extension

Para resolver o problema da distancia máxima da rede, optou-se por aumentar o *slot time* na transmissão de tramas. O *slot time* é o intervalo de tempo definido para transmitir uma trama, que garante que todos os terminais da rede detetam a ocorrência de uma possível colisão.

Na Ethernet a 10 Mbps e 100 Mbps este intervalo é igual ao tempo de transmissão de 64 bytes. Contudo, no Gigabit Ethernet, este valor foi aumentado para o tempo de transmissão de 512 bytes [27, pág 170-172]. A imagem que segue mostra o aumento que a trama sofre, com esta alteração .

| Preâmbulo | DA | SA | Tipo/ Tamanho | Dados | FCS | Extensão |
|---|---|---|---|---|---|---|
| 8 Bytes | 6 Bytes | 6 Bytes | 2 Bytes | 46 - 1500 Bytes | 4 Bytes | 0 - 448 Bytes |

Figura 2.17: Trama Ethernet com *Carrier Extension*.

Na operação de *carrier extension*, quando é transmitida uma trama de tamanho inferior a 512 bytes o terminal emissor continua ocupado e é enviado pela camada física uma extensão de bits até ser completado o tamanho mínimo de 512 bytes do *slot time*. Desta maneira o campo Extensão da trama pode variar entre 0 e 448 bytes. Note-se que o FCS não cobre o campo Extensão, e este campo não contém informação. Desta maneira o *carrier extension* não aumenta o tamanho mínimo de uma trama, apenas aumenta o tempo de transmissão para tramas com tamanho inferior a 512 bytes. No cenário de transmissões de tramas com 64 bytes, o tempo de transmissão vai aumentar 8 vezes, o que representa um grande desperdício de recursos da rede.

A implementação descrita anteriormente, permitiu manter a distancia máxima da rede na passagem para o Gigabit Ethernet, mas comprometeu a eficiência de transmissões com tramas mais pequenas. Para resolver esta falta de eficiência estudaram-se algumas soluções, das quais se salienta o *frame bursting* que será estudado mais tarde, e o *packet packing* que consiste em inserir mais do que uma trama num *slot time*. A segunda aproximação revelou-se bastante complexa de implementar, pelo que foi abandonada [27, pág 172].

#### Frame Bursting

No *frame bursting* quando um terminal inicia a transmissão de uma trama, ativa ao mesmo tempo um temporizador chamado *burst timer*. A trama enviada usará o *carrier extension*

se o seu tamanho for inferior a 512 bytes. Se a trama for transmitida com sucesso e o *burst timer* ainda não tiver expirado, o terminal poderá enviar mais uma trama sem uso de *carrier extension*. Este processo pode voltar a ser repetido, enquanto o *burst timer* não expirar e o terminal possuir tramas para enviar. A figura (2.19) ilustra o diagrama de fluxos deste processo:

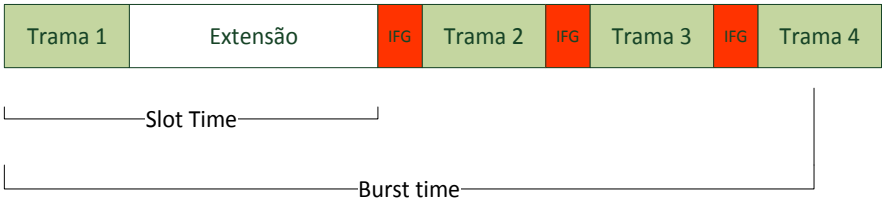



Figura 2.18: Tramas enviadas em *frame bursting*.

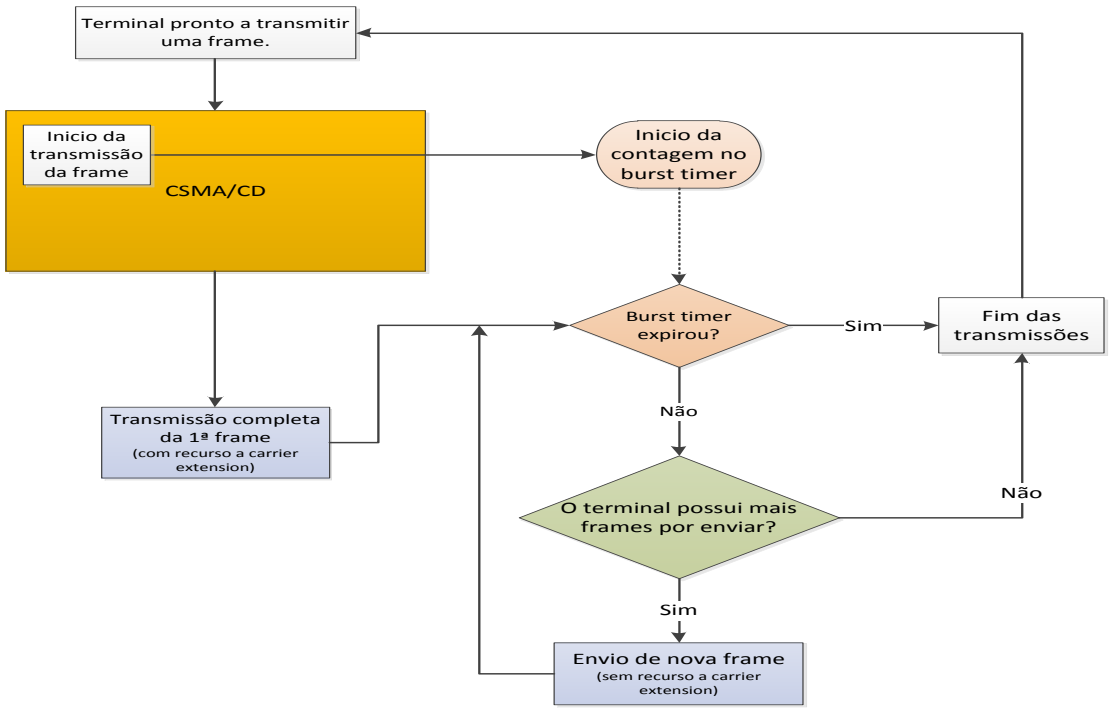



Figura 2.19: Diagrama de fluxos do *frame bursting*.

A figura (2.18) ilustra a operação de *frame bursting* no tempo, como se observa o *Interframe gap* continua a ser necessário nesta implementação. O *frame bursting* mantém um elevado respeito pelo CSMA/CD enquanto aumenta a eficiência na transmissão de tramas mais curtas, e possui uma implementação relativamente simples.

**Eficiência**

Comparando o tamanho da informação útil transmitida em cada trama com o tamanho total da trama necessário para a transmitir, obtemos uma medida de eficiência. Seguem as equações desta medida, para os modos de operação em *half-duplex*:

Sem recurso a *carrier extension*:

$$Eficiencia = \frac{F}{F + I + P} \tag{2.11}$$

Com *carrier extension*:

$$Eficiencia = \frac{F}{max(S, F) + I + P} \tag{2.12}$$

Com *carrier extension* e *frame bursting*:

$$Eficiencia = \frac{F(1) + \sum_{i=2}^{x} F(i)}{max(S, F(1)) + \sum_{i=2}^{x}(F(i) + P + I)} \tag{2.13}$$

$F$ = tamanho da trama (bytes)
$S$ = slot time (512 bytes)
$I$ = interframe gap (96 bits)
$P$ = tamanho do preâmbulo (64 bits)
$x$ = número de tramas consecutivas transmitidas no *frame burst*

No extremo, ao ser usada uma trama com 64 bytes, e ao serem transmitidas 93 tramas consecutivas na operação de *frame burst*, obtemos uma eficiência de 76% sem recurso a *carrier extension*, 12% com *carrier extension*, e 72% com recurso a *carrier extension* e *frame burst*. O uso de *frame burst* compensa parte da eficiência que é perdida na utilização de *carrier extension* em tramas de tamanho muito inferior ao tamanho do *slot time*.

A tabela seguinte resume os parâmetros característicos das tecnologias Ethernet em modo *half-duplex*, ou seja com recurso ao uso de CSMA/CD [22, pág 57]:

| | **10Mbps Ethernet** | **Fast Ethernet** | **Gigabit Ethernet** |
|---|---|---|---|
| Slot time (bytes) | 64 | 64 | 512 |
| Bit time($\mu$s) | 0.1 | 0.01 | 0.001 |
| Interframe gap ($\mu$s) | 9.6 | 0.96 | 0.096 |
| Tentativas de retransmição | 16 | 16 | 16 |
| Tamanho do jam (bits) | 32 | 32 | 32 |
| Tamanho máximo da trama (bytes) | 1518 | 1518 | 1518 |
| Tamanho mínimo da trama (bytes) | 64 | 64 | 64 |
| Tamanho do *burst* (bytes) | | | 8192 |

Tabela 2.2: Parametros do CSMA/CD nas várias tecnologias (Adaptado de [6, pág 57])

### 2.3.2 10 Gigabit Ethernet

O 10 Gigabit Ethernet, especificado na norma IEEE 802.3ae, funciona de modo bastante semelhante ao Gigabit Ethernet, ficando as suas diferenças resumidas a modificações na camada protocolar física de modo a suportar uma maior taxa de transmissão. A tecnologia está definida para operar sobre cobre e fibra ótica apenas em modo *full-duplex*. Não necessita por isso, do recurso ao CSMA/CD e o seu limite não fica restringido pelas limitações do uso deste protocolo.

### 2.3.3 40 e 100 Gigabit Ethernet

A necessidade de largura de banda forçou operadores de telecomunicações a usar nas suas redes varias ligações de 10 Gigabit Ethernet paralelas. Estas soluções necessitam de ser acompanhadas de *switchs* e interfaces adicionais que tornam esta prática bastante dispendiosa.

O IEEE atento às necessidades dos operadores de telecomunicações, criou em 2006 o Higher Speed Study Group (HSSG), que se propôs analisar a necessidade de largura de banda Ethernet, num espaço entre três a sete anos [28] [7]. Verificou-se que as taxas de transmissão necessárias nos centros de processamento de dados (*datacenters*) duplicava a cada 24 meses, e assim previu-se que por volta de 2014 seriam necessárias taxas de transmissão de 40 Gbps. Quanto às redes de transporte, verificou-se que o aumento de largura de banda usada acompanhava aproximadamente o crescimento da Internet e das telecomunicações em geral, ou seja duplicando entre 12 a 18 meses. Desta forma, previu-se que seria necessário taxas de transmissão de 100 Gbps para a rede de transporte no ano de 2014 [28] [7].

Pela primeira vez na história da tecnologia Ethernet, foi necessário normalizar duas taxas de transmissão diferentes. Foi com este objetivo que em Janeiro de 2008 foi formado o *IEEE 802.3ba 40 Gb/s and 100 Gb/s Task Force*. O propósito do grupo consistiu no desenvolvimento de especificações para a camada física de modo a habilitar a Ethernet a operar com as taxas referidas. Segue em detalhe os objetivos a que o *IEEE 802.3ba 40 Gb/s and 100 Gb/s Task Force* se comprometeu [28] [7]:

- Operar apenas em modo *full-duplex*
- Preservar o formato de trama 802.3.
- Preservar o tamanho mínimo e máximo das tramas Ethernet 802.3.
- Possuir um Bit Error Rate (BER) igual ou inferior a $10^{-12}$.
- Garantir o suporte e compatibilidade com tecnologia OTN.
- Garantir taxas de transmissão de 40 Gb/s na camada MAC, através de especificações na camada física que suportem:
  - Pelo menos 10 Km em SMF.
  - Pelo menos 100m MMF.
  - Pelo menos 10 m em cobre.
- Garantir taxas de transmissão de 100 Gb/s na camada MAC, através de especificações na camada física que suportem:

– Pelo menos 40km em SMF.
– Pelo menos 100m em MMF
– Pelo menos 10 m sobre cobre.

**Arquitetura**

O *IEEE 802.3ba 40 Gb/s and 100 Gb/s Task Force* especificou uma única arquitetura para ambas as tecnologias, tal como ilustra a figura (2.20). A camada MAC, correspondente à segunda camada do modelo OSI, encontra-se ligada ao meio de transmissão ótico ou de cobre, através do Ethernet PHY. Esta camada física compreende a sub-camada Physical Medium Dependent (PMD), a sub-camada Physical Medium Attachment (PMA) e a sub-camada PCS. No caso do meio físico ser o cobr, esta arquitetura possuirá ainda uma sub-camada de Auto-Negotiation (AN) e uma sub-camada Forward Error Correction (FEC).

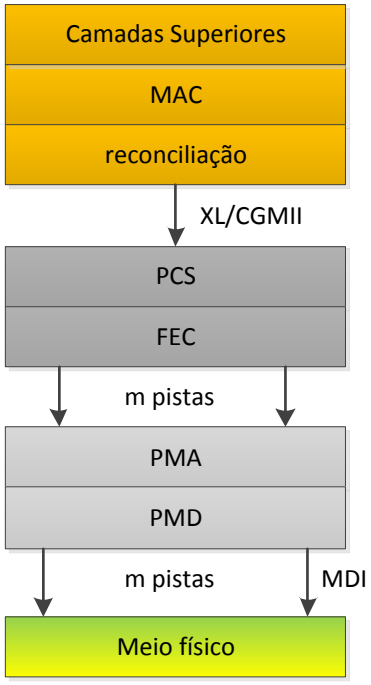



Figura 2.20: Arquitetura da camada física das tecnologias 40GbE e 100GbE

A sub-camada de reconciliação mapeia os fluxos de bits provenientes da camada MAC, para o Media Independent Interface (MII) correspondente. CGMII para 100Gbps e XLGMII para 40 Gbps. A sub-camada PCS possuí funções de multiplexagem e codificação dos sinais. O tipo de codificação usada é 64/66b, ou seja 64 bits de informação codificados em 66 bits de código. A sub-camada opcional FEC fornece codificação de modo a corrigir erros de transmissão, enquanto que a sub-camada PMA será responsável por dispor em série, os fluxos de bits codificados, para posterior envio pelo PMD. O PMD sinalizará o envio do fluxo através do Media Dependent Interface (MDI).

**Multi-Lane Distribution (MLD)**

Como referido anteriormente a sub-camada PCS realiza a comunicação entre o MII e a sub-camada PMA e tem como funções, a codificação de bits e multiplexagem dos fluxos de bits provenientes da camada MAC em várias pistas de transmissão com taxas inferiores. As pistas serão posteriormente enviadas separadamente. A operação de multiplexagem é assim de extrema importância na presente tecnologia já que através do recurso a alguma engenharia é possível obter taxas de transmissão que superam as taxas de transmissão máximas das fibras óticas.

O mecanismo de multiplexagem dos fluxos chama-se MLD e foi desenhado para suportar a transmissão de 40Gb e 100Gb, sendo escalável e flexível ao ponto de poder operar noutras sub-camadas, e de no futuro poder suportar taxas de transmissão superiores.

A informação proveniente de camadas superiores (MAC), será tal como referido, codificada em blocos de 64/66b. Após esta operação os blocos de 66 bits, serão distribuídos por várias pistas num algoritmo *round robin*. A figura (2.21) ilustra o conceito de distribuição dos blocos em várias pistas PCS.

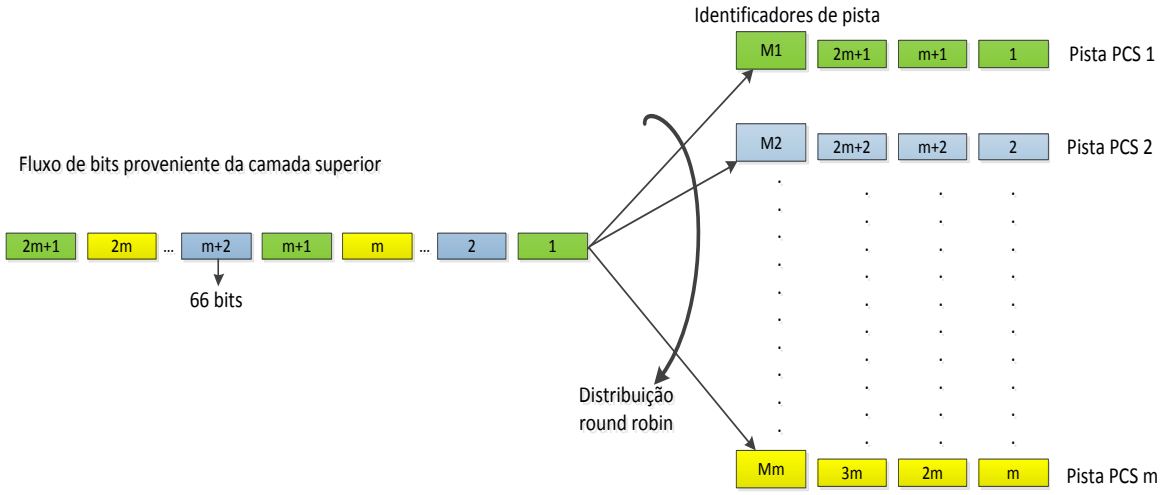



Figura 2.21: Mecanismo MLD da sub-camada PCS

O número de pistas necessárias m será o mínimo múltiplo comum do número de interfaces óticas e elétricas. Para 100 Gbps foram escolhidas 20 pistas, e para 40 Gbps foram escolhidas 4 pistas. Desta forma, a transmissão em 100 Gbps, poderá ser multiplexada em 1, 2, 4, 5, 10 ou 20 pistas, enquanto a transmissão em 40 Gbps poderá ser multiplexada em 1, 2 ou 4 pistas.

Cada pista possui um identificador único que é inserido periodicamente. Este tipo de multiplexagem *round-robin* poderá levar a que múltiplas pistas PCS fiquem alojadas na mesma interface física, contudo todos os bits da mesma pista seguirão a mesmo interface física.

Tome-se como exemplo o funcionamento da tecnologia 100GBASE-LR4, onde são usados 4 comprimentos de onda que operam a 25 Gbps em SMF, para transmitir 100 Gbps [28]. Neste exemplo, a sub-camada PCS efetua a distribuição do sinal proveniente da camada MAC em 20 pistas PCS, que são depois enviadas para a sub-camada PMA. Nesta implementação a sub-camada PMA encontra-se subdividida em 2 camadas interligadas por um Second Attachment

Unit Interface (CAUI), que consiste em 10 interfaces de 10 Gbps cada [28]. A sub-camada PMA superior realiza uma multiplexagem das 20 pistas, em 10 interfaces de 10 Gbps, e em seguida a sub-camada PMA inferior converte os sinais provenientes das 10 interfaces em 20 pistas PCS, que serão multiplexadas novamente através do mecanismo MLD em 4 comprimentos de onda, para serem posteriormente enviados.

Quanto à sub-camada PCS recetora, esta recebe as várias pistas PCS, e alinha estas com recurso aos identificadores, agregando o sinal original. O esquema MLD permite realizar as funções de tratamento/processamento de informação em tecnologia eletrónica de baixo custo, sendo que a eletrónica de alta velocidade será apenas usada nas interfaces com os meios óticos. Esta metodologia para além de escalável e flexível, maximiza o uso dos recursos disponíveis, minimizando os custos.

## 2.4 Ethernet para a Rede de Transporte

### 2.4.1 Motivação

Como estudado anteriormente a tecnologia Ethernet teve o seu desenvolvimento orientado para as redes locais, onde acabou por prosperar e se tornar na tecnologia dominante. Atualmente 95%, do tráfego que atravessa as redes de transporte tem como origem ou destino uma porta Ethernet [1]. Foi mostrado também no primeiro capítulo, que as atuais soluções existentes para a rede de transporte, baseadas em comutação de circuitos, são ineficientes para transportar tráfego de dados, tráfego este que atualmente representam a grande maioria do tráfego que circula nas redes.

Estas duas premissas atribuem assim à tecnologia Ethernet a condição de candidata às redes de transporte. Contudo, o facto da Ethernet ter sido desenhada para operar em LANs, cria alguns obstáculos.

### 2.4.2 Desafios

O desfio inicial da implementação da tecnologia nas redes de transporte, deveu-se ao seu limitado alcance. Como estudado na secção (2.3.3) o alcance máximo da tecnologia 100 Gbps Ethernet é de 40 Km, o que não é propriamente o alcance desejado numa tecnologia de rede de transporte. Para superar esta limitação as implementações Ethernet na rede de transporte são suportadas por tecnologia OTN ou SDH, e desta forma o alcance de 40 Km é superado. A figura (2.22) ilustra a pilha protocolar desta solução.

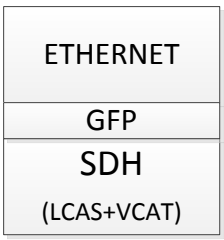



Figura 2.22: Pilha protocolar de uma solução para rede de transporte com tecnologias SDH e Ethernet

Resolvida a questão do alcance, persistem ainda outras limitações técnicas, que impossibilitam as tecnologias Ethernet, a operar de forma independente na rede de transporte:

- O tempo de recuperação a falhas na rede Ethernet variam entre 2 e 120 s [16], que é aproximadamente o tempo que os *switch* demoram a convergir para uma solução, usando o algoritmo de *Spanning Tree*. Para uma rede de transporte estes tempos de recuperação são inaceitáveis, devido à quantidade de tráfego considerado crítico que lá circula. O tempo máximo de recuperação a falhas de uma rede SDH é de 50 ms [3].
- O protocolo *Spanning Tree* limita também o tamanho máximo da rede a 7 *switchs*.
- O número de utilizadores também é limitado, já que cada dispositivo da rede deve conhecer os endereços MAC de todos os outros dispositivos. Esta limitação deve-se aos atrasos introduzidos pelos *switchs*, quando necessitam de efetuar pesquisas em grandes tabelas de encaminhamento.
- A tecnologia não suporta nativamente Quality of Service (QoS). O encaminhamento de tramas é apenas definido pelo seu endereço destino, e desta maneira a tecnologia não faz distinção entre tramas de prioridade crítica e outras tramas de baixa prioridade. Esta falha é colmatada com a introdução da funcionalidade Virtual LANs, contudo deve ainda ser melhorada para o uso em redes de transporte.
- Não existe na tecnologia suporte nativo para serviços TDM, que apesar de tudo, são ainda um fonte importante de receitas para os prestadores de serviço [16, pág 6].
- A operação de *flooding* após receção de uma trama com endereço de *broadcast*, *multicast* ou com endereço desconhecido, deve também ser evitada, devido ao desperdício de largura de banda que introduz.

# Capítulo 3

# Carrier Ethernet

O conceito de Carrier Ethernet apresenta duas definições que não convergem totalmente. O presente capítulo estudará a tecnologia segundo os atributos que a mesma deve incorporar e embora tal definição deva ser alheia à tecnologia usada, esta contudo baseia-se na anterior tecnologia Ethernet, tal como indica o nome Carrier Ethernet.

Uma definição alternativa de Carrier Ethernet consiste na caraterização da tecnologia com base nas implementações reais já existentes ou ainda em desenvolvimento. Estas implementações baseiam-se em tecnologias, que apesar de tentarem responder aos atributos da primeira definição, podem não orientar o seu desenvolvimento a partir destes [1, pág 9].

**MEF**

Em 2001 foi criado o consórcio internacional sem fins lucrativos MEF. O MEF tem como objetivo promover a adoção a nível mundial de uma tecnologia para as redes de transporte, que tenha por base a tecnologia Ethernet. Esta promoção é feita através da elaboração de especificações técnicas destinadas a definir os serviços, e consequentes atributos e requisitos de uma tecnologia de rede de transporte. A certificação destas especificações, fornece um standard de interoperabilidade, globalmente reconhecido que melhora os processos de aprovação e implementação da tecnologia [29]. A título de exemplo, a PT Inovação possui certificação MEF para a especificação MEF 9 e MEF 14 [30].

Tendo em consideração a importância do MEF para a industria de telecomunicações, o presente documento seguirá a definição de Carrier Ethernet sugerida por este. A definição de Carrier Ethernet segundo o MEF é a seguinte: "Rede e Serviços normalizados, universais de rede de transporte, definidos pelos cinco atributos que distinguem a tecnologia Carrier Ethernet da tecnologia Ethernet" [29]. Os cinco atributos, a que a definição se refere são os seguintes:

- Serviços normalizados,
- Escalabilidade,
- Fiabilidade,
- Qualidade de serviço,
- Gestão de serviços.

## 3.1 Atributos segundo o MEF

### 3.1.1 Serviços Normalizados

O operador deve fornecer serviços baseados em comutação de tramas e serviços TDM de forma eficiente e determinística, através de equipamentos normalizados. Os serviços TDM devem ser suportados através da emulação de circuitos virtuais. Note-se que este tipo de serviço representa ainda uma considerável fonte de receitas por parte dos operadores [31, pág 50], pelo que não deve ser desconsiderado.

A normalização de equipamentos deve aptar a tecnologia a operar em qualquer parte do mundo, sem qualquer restrição, à semelhança do que acontece com a tecnologia Ethernet nas LANs [31, pág 50].

Visto os clientes possuírem diferentes necessidades de largura de banda, o MEF propõe a oferta de uma largura de banda entre 1 Mbps até 10 Gbps, com incrementos de 1 Mbps.

O Carrier Ethernet deve suportar os seguintes serviços Ethernet: Ethernet Line (E-Line), Ethernet Local Area Network (E-LAN) e Ethernet Tree (E-Tree). Estes serão abordados com maior detalhe na secção (3.3).

### 3.1.2 Escalabilidade

Uma diferença fundamental entre a LAN e a rede de transporte do operador é a escala. A tecnologia Carrier Ethernet deve ser capaz de variar o número de utilizadores, o alcance geográfico e a largura de banda oferecida sem degradar o desempenho geral da rede. O aparecimento de novas aplicações provenientes do mundo do entretenimento, do mundo empresarial, ou de qualquer outra necessidade/oportunidade que surja, deve também ser suportada pelo Carrier Ethernet.

### 3.1.3 Fiabilidade

O suporte de aplicações críticas é outro dos atributos que a tecnologia Carrier Ethernet deve possuir. Para poder suportar estes serviços a rede deve ser capaz de se proteger e recuperar a falhas de modo eficiente. O sistema de saúde, ou o sistema de educação não podem ficar comprometidos por uma falha da rede. Esta responsabilidade acrescida da rede de transporte não está presente nas LANs. Os mecanismo de sobrevivência devem ser aplicados entre extremos da rede, para que cada terminal esteja protegido contra falhas físicas, ou falhas de tecnologias de camadas protocolares inferiores. O tempo de restauro da rede deve ser igual ou inferior ao tempo máximo de restauro a falhas da tecnologia SDH (50 ms).

### 3.1.4 Qualidade de serviço

A tecnologia Ethernet não possuía mecanismos de QoS, contudo estes são um atributo que deve estar presente nas redes de transporte. Os operadores devem fornecer QoS a milhares aplicações e utilizadores através do respeito por Service Level Agreementss (SLAs) rigorosos, que devem ser definidos através de um conjunto de parâmetros que garantam o comprimento dos valores de QoS acordados.

### 3.1.5 Gestão do serviço

A gestão de uma rede de telecomunicações tenderá a ser mais complexa, à medida que aumenta o número de clientes e a área geográfica que ocupa. Desta forma são necessários mecanismos de OAM sofisticados. Um dos requisitos deste mecanismo será possibilidade de alterar os serviços oferecidos aos clientes de forma remota e célere, o que deve constituir uma vantagem face às tradicionais tecnologias de transporte.

O Carrier Ethernet deverá ter mecanismos de monitorização, diagnóstico e gestão da rede independentes do equipamento usado. A existência de uma gestão unificada independente do equipamento do operador, representaria uma rutura com o estado atual da gestão de redes de transporte, onde os equipamentos usados definem o modo de gestão da rede [1, pág 18].

## 3.2 Modelo de Serviços

O MEF para além da definição dos anteriores atributos, introduziu alguns conceitos com o objetivo de responder a estes. Um desses conceitos é o modelo de serviços da tecnologia.

Quando a tecnologia Ethernet apareceu, foi vista como um produto/solução, logo nunca foi alvo de expectativas em relação à oferta de serviços [31, pág 56]. Mas este cenário é diferente para o Carrier Ethernet, onde é efetivamente esperado que a tecnologia possa suportar uma larga variedade de serviços, tal como aludido anteriormente.

Antes de se estudar os serviços suportados pelo Carrier Ethernet é importante definir o contexto em que estes serviços estarão inseridos, ou seja a arquitetura de serviços. Neste sentido o MEF (com o contributo do International Telecommunication Union (ITU)) desenvolveu um modelo genérico de arquitetura de serviços, apelidado de Ethernet Service Model (ESM). O ESM está ilustrado na figura (3.1).

Este modelo tem duas componentes principais, o Customer Equipment (CE) que é o equipamento localizado nas instalações do cliente e o Service Provider Ethernet Network (SEN) [1] que é a rede que o operador possuí, ou opera. O equipamento do cliente CE está ligado à rede do operador SEN, através da User Network Interface (UNI). O SEN consiste num conjunto de componentes físicos e elementos lógicos que têm como função ligar todas as LANs da rede através das Metro Area Networks (MANs) e WANs e para o realizar o MEF introduziu o Ethernet Virtual Connection (EVC).

No modelo são também formalmente definidas 3 camadas. A Application Services Layer (APP) que suporta as aplicações finais do cliente, e tem como base a ligação Ethernet, ou seja a camada Ethernet Service Layer (ETH). A camada ETH compreende os serviços Ethernet e encontra-se sobre a camada Transport Layer (TRAN) que será uma tecnologia de camada de transporte. O esforço do MEF está principalmente direcionado para a camada ETH, que supostamente é onde estará definido o Carrier Ethernet. Contudo esta distinção de camadas pode não ser completamente rígida, já que, por exemplo, alguns dos atributos definidos na secção anterior podem ser delegados à camada de transporte TRAN.

As 3 camadas possuem 3 planos de operação, o plano de Dados, o plano de Controlo e o plano de Gestão. Estes planos são responsáveis tal como o nome sugere pela entrega, controlo, e gestão do tráfego de informação. Se geralmente, este 3 domínios estão bem definidos para a camada TRAN, já para a camada ETH apenas recentemente começaram a ser realizados

[1] É usado o termo SEN, já que esta rede para além de englobar as redes Metro Ethernet Network (MEN) alcança também as redes WAN.

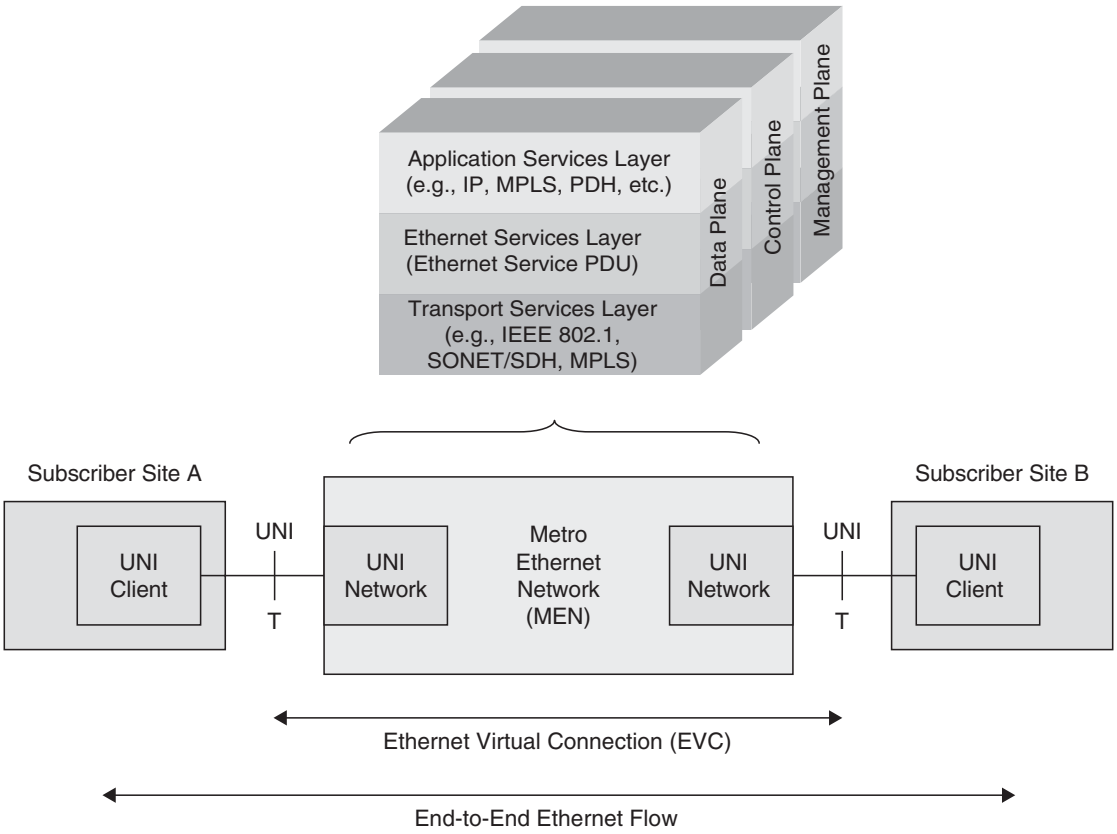



Figura 3.1: Modelo de serviços genérico para distribuição de serviços Ethernet (Fonte: MEF)

esforços nesse sentido [31, pág 59].

As caraterísticas da UNI, EVC e Network to Network Interface (NNI) são fundamentais nas operações de controlo e gestão da camada ETH, já que é através destes componentes que são entregues os serviços Ethernet. As definições de UNI, EVC e NNI são de elevada importância para o MEF pelo que serão abordadas em seguida.

### 3.2.1 UNI, EVC e NNI

#### UNI

A **UNI** é a interface que separa a rede do cliente da rede do operador de telecomunicações. Esta interface pode assim ser dividida em duas, a User Network Interface - Client side (UNI-C), e a User Network Interface - Network side (UNI-N).

A UNI-C contém todas as funções necessárias para ligar um cliente à rede, enquanto que a UNI-N contém todas as funções necessárias para ligar a rede ao cliente. A tabela (3.1) contém os atributos que descrevem a operação de uma UNI. Note-se o facto de uma UNI poder possuir várias ligações EVC e a possibilidade de atribuir um perfil de largura de banda a cada EVC, desta forma é possível que uma UNI possua vários serviços com diferentes perfis de largura de banda.

| **Atributos** | **Descrição** |
|---|---|
| Identificador UNI | *String* usada para identificar a UNI |
| Meio físico | Meio físico Ethernet |
| Taxa de transmissão | 10 Mbps, 100 Mbps, 1 Gbps or 10 Gbps |
| Modo | *full duplex* ou *half duplex* |
| Camada MAC | IEEE 802.3-2002 |
| Multiplexagem do serviço | Sim ou Não. Define se a UNI pode suportar vários serviços |
| UNI EVC ID | *String* usada para identificar um EVC |
| Tabela CE-VLAN ID/EVC | Tabela de correspondência entre VLANs ID e EVCs |
| Número máximo de EVCs | Número máximos de EVCs que a UNI suporta |
| Agregação de VLANs | Sim ou Não. Indica se existem mais do que um VLAN ID para um dado EVC |
| Agregação completa | Sim ou Não. Indica se todas os VLAN ID estão mapeadas para o mesmo EVC |
| Perfil de largura de banda da UNI | Nada ou <CIR, CBS, EIR, EBS>. |
| Perfil de largura de banda do EVC | Nada ou <CIR, CBS, EIR, EBS>. |
| Perfil de largura de banda de CoS ID | Nada ou <CIR, CBS, EIR, EBS>. |

Tabela 3.1: Atributos da UNI (adaptado Fonte: MEF)

### EVC

O **EVC** é um construtor de ligações virtuais, possui a identificação de todas as UNIs que participam na ligação e entrega os fluxo de tráfego Ethernet apenas aos clientes identificados, prevenindo assim a transferência de informação para UNIs que não fazem parte da EVC. Os atributos do EVC estão especificados na tabela(3.2).

A entrega de tramas Ethernet através de um EVC deve respeitar algumas regras. Uma trama nunca deve ser reencaminhada para a UNI que a originou, o conteúdo da trama Ethernet deve permanecer inalterado, isto inclui o seu endereço MAC. O EVC pode definir ligações ponto-a-ponto ou multiponto-multiponto [31, pág 60], do ponto de vista das ligações lógicas, as ligações multiponto-multiponto serão semelhante à funcionalidade de Virtual LANs estudada no capítulo anterior.

### NNI

O **NNI** define a arquitetura de serviços entre operadores de Carrier Ethernet. É assim a interface que separa os domínios de responsabilidades de diferentes operadores. O MEF definiu os seguintes tipos de NNIs:

- External Network to Network Interface (E-NNI) - É uma interface aberta para ligar duas redes pertencentes a dois operadores.

- Internal Network to Network Interface (I-NNI) - É uma interface para ligar vários segmentos do mesmo operador.

| Atributos | Descrição |
|---|---|
| Tipo de EVC | Ponto-a-ponto ou multiponto-multiponto |
| Lista de UNIs | Lista de UNIs usadas pelo EVC, identificadas pelo identificador UNI |
| Guardar CE-VLAN ID | Sim ou não. Especifica quando um VLAN ID é salvaguardado |
| Guardar CE-VLAN CoS | Sim ou não. Especifica quando o VLAN Class of Service (CoS) (802.1p) é salvaguardado |
| Serviço *unicast* | Indica quando tramas *unicast* devem ser descartadas, entregues incondicionalmente, ou entregues sobre determinadas condições |
| Serviço *multicast* | Indica quando tramas *multicast* devem ser descartadas, entregues incondicionalmente, ou entregues sobre determinadas condições |
| Serviço *broadcast* | Indica quando tramas *broadcast* devem ser descartadas, entregues incondicionalmente, ou entregues sobre determinadas condições |
| Performance do serviço | Especifica o atraso das tramas, a variância do atraso e tramas perdidas por um EVC, ou por tramas de um EVC, identificadas através dos seu valor CE-VLAN CoS (802.1p) |

Tabela 3.2: Atributos da EVC (adaptado Fonte: MEF)

- Network Interworking Network to Network Interface (NI-NNI) - É uma interface que permite a extensão das infraestruturas de transporte, através do uso de redes de transporte externas que não pertencem ao operador prestador do serviço.
- Service Interworking Network to Network Interface (SI-NNI) - É uma interface que permite a interoperação entre os serviços Carrier Ethernet e serviços prestados por outras tecnologias.

### 3.2.2 Caracterização do tráfego

O perfil de largura de banda definido pelo MEF, tal como se observa na tabela (3.1), tem a forma: <CIR, CBS, EIR, EBS, CM>. Segue a descrição de cada um destes parâmetros:

- Committed Information Rate (CIR), é a taxa média de transmissão de tramas atendendo à performance desejada (perdas e atrasos) para o serviço associado. Se vários serviços estiverem a ser entregues a uma UNI, a soma dos CIRs associados a cada serviço individual deverá ser inferior ou igual à taxa máxima de transmissão da UNI (Taxa da UNI, na tabela (3.1)).
- Committed Burst Size (CBS), é o limite máximo da taxa de transmissão para o qual as tramas ainda são CIR-*Conformant*, ou seja a sua entrega ainda é garantida.
- Excess Information Rate (EIR), é uma taxa média de transmissão que pode assumir um valor igual ou superior à CIR, e ao qual as tramas são consideradas EIR-*Conformant*, ou seja são aceites na rede, contudo não é garantida a sua entrega.

- Excess Burst Size (EBS), é o limite máximo da taxa de transmissão para que as tramas ainda sejam consideradas EIR-*Conformant*. A partir deste valor as tramas são descartadas imediatamente.

- Color Marking (CM), é um parâmetro adicional, que atribui uma cor à trama, consoante o seu respeito pelos parâmetros anteriores (CIR, CBS, EIR, EBS) .

  – Verde - respeita a CIR e a CBS, logo a sua entrega é garantida.

  – Amarelo - respeita a EIR, não tem uma entrega garantida, poderá ser entregue se a rede não estiver congestionada.

  – Vermelho - não respeita nenhum dos parâmetros, pelo que é descartada.

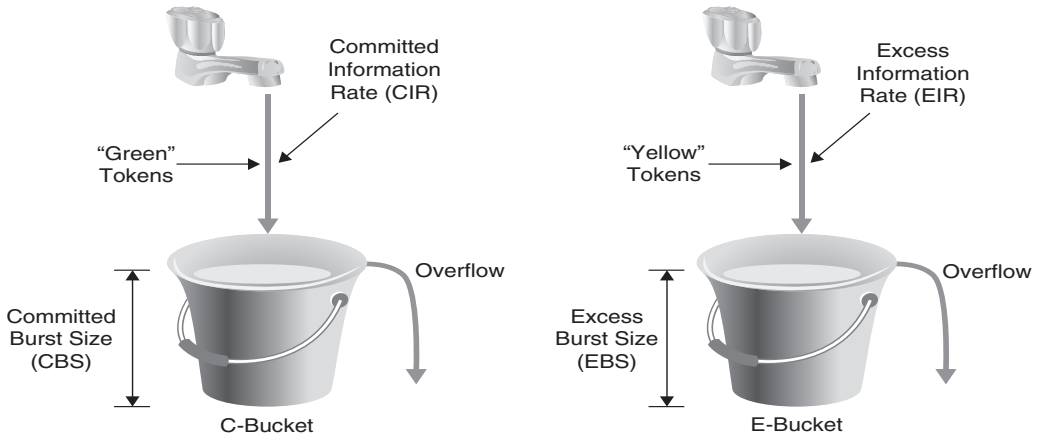



Figura 3.2: Algoritmo de atribuição de cores às tramas (Fonte: MEF)

A figura (3.2) ilustra o algoritmo usado na atribuição de cores às tramas, assim como o papel de cada parâmetro do perfil de largura de banda nesta atribuição. Este algoritmo é muitas vezes referido como trTCM (*two-rate (committed or excess), Three-Color Marker*), e é implementado com recurso a dois baldes virtuais [31, pág 64], tal como mostra a figura (3.2).

Quando as tramas entram da rede do operador através de uma UNI, o mesmo número de *tokens* é removido do *C-Bucket*. Se após isto ainda existirem *tokens* verdes neste balde, as tramas inseridas na rede consideram-se CIR-*conformant*, é lhes atribuída a cor verde e são aceites na rede. No caso de não existirem mais *tokens* verdes no *C-Bucket*, é verificado se existem *tokens* amarelos no *E-Bucket*. Caso existam, as tramas consideram-se EIR-*Conformant*, é lhes atribuída a cor amarela, e são aceites na rede. Caso não existam *tokens* amarelos no *E-Bucket*, é lhes atribuída a cor vermelha e são descartadas.

Existe a possibilidade de adicionar os *tokens* verdes não usados ao balde *E-Bucket* como *tokens* amarelos, quando se verifica se o tráfego enviado para a rede é EIR-*Conformant*. Desta forma dar-se-á um aumento do tráfego "amarelo"nas redes de transporte dos operadores.

## 3.3 Tipos de Serviço Ethernet

O modelo de serviços ESM ilustrado na figura (3.1) mostra que os serviços Ethernet são entregues através de uma ligação EVC controlada pelo operador, e ligada à rede do cliente através de uma UNI. Dependendo do modo como estas ligações virtuais são configuradas é possível obter 3 serviços de Ethernet distintos.

- E-Line
- E-Lan
- E-Tree

Estes 3 tipos de serviço Ethernet são completamente especificados através dos atributos das UNIs e dos EVCs intervenientes (Tabela (3.1) e (3.2)). A seguir serão descrito, com mais detalhe, os três tipos de serviços.

### 3.3.1 E-LINE

Serviço E-Line usa uma EVC ponto-a-ponto para ligar dois UNIs tal como é ilustrado na figura (3.3). Embora o serviço ligue apenas dois cliente, a ligação pode suportar um elevado número de serviços, com diferentes perfis de largura de banda definidos pelos parâmetros <CIR, CBS, EIR, EBS, CM>.

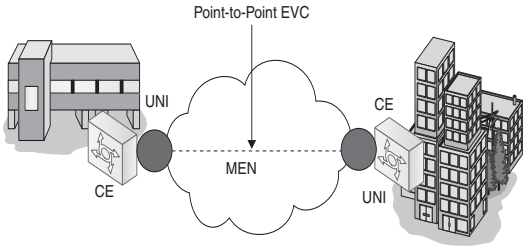



Figura 3.3: Serviço E-Line (Fonte: MEF)

O serviço de ligação E-LINE pode ser direcionado à porta UNI, ou à *tag* da VLAN sendo que, o funcionamento em ambos os casos se assemelha ao funcionamento de Virtual LANs usando *implicit* e *explicit tagging*, respetivamente[2].

A ligação direcionada à porta UNI é apelidada **Ethernet Private Line (EPL)**. Nesta situação toda a largura de banda da ligação é dedicada ao serviço da EVC. A transferência de tramas entre as UNIs é aqui completamente transparente, e espera-se que exista um atraso de transmissão de tramas bastante reduzido, com uma variação de atraso reduzida, e uma taxa de perdas aproximadamente nula. Neste caso não é possível multiplexar serviços, já que a interface física (a porta da UNI) apenas serve esta ligação. É de salientar o facto de, neste caso, não ser necessário usar parâmetros derivados do endereçamento baseado na *tag* VLAN

[2]A funcionalidade de Virtual LAN foi estudada em detalhe no capítulo Ethernet, secção (2.2.8).

nas tabelas (3.1) e (3.2). Esta solução será aconselhada a casos onde é necessário substituir linhas privadas TDM.

Quando a E-LINE faz uso da tag VLAN a ligação denomina-se **Ethernet Virtual Private Line (EVPL)**. Nesta ligação o campo VLAN ID é usado para identificar a EVC, e desta forma a UNI possuirá um mapa que associa cada EVC a um CE-VLAN ID (tabela (3.1). A EVPL possibilita multiplexar vários serviços na mesma UNI, usando várias EVCs para a mesma porta física. Esta é uma ligação bastante útil para o operador, já que pode oferecer vários serviços a um cliente usando apenas uma ligação física. Segundo o MEF pode substituir serviços de Frame Relay e ATM [31, pág 72]. Em ambos os serviços E-LINE é ainda possível configurar outros parâmetros de performance, para além da largura de banda, tal como o atraso por trama, variação do atraso por trama, percentagem de tramas perdidas e disponibilidade da rede (tabela (3.2)) [31, pág 70].

### 3.3.2 E-LAN

Os serviços E-LAN usam ligações EVC multiponto-multiponto tal como ilustra a figura (3.4).

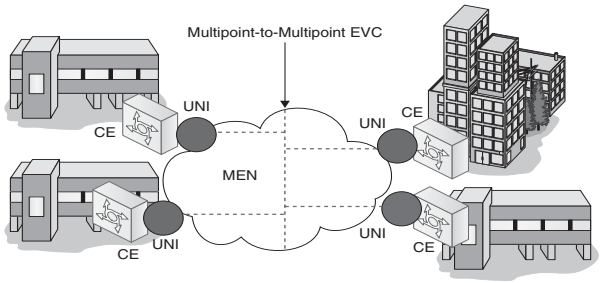



Figura 3.4: Serviço E-Lan (Fonte: MEF)

Tal como acontece nas ligações E-LINE, também as E-LAN poderão ter a sua conectividade orientada à porta física da UNI, ou à tag do campo VLAN da trama. No caso em que a ligação é orientada à interface física, estamos perante uma **Ethernet Private Local Area Network (EPLAN)**, onde a transmissão de tramas é totalmente transparente, e a rede resultante é essencialmente uma LAN Ethernet à escala de uma rede de transporte.

Quando a ligação é orientada à tag VLAN o serviço denomina-se **Ehernet Virtual Private Local Area Network (EVPLAN)**, e o resultado será uma LAN Ethernet à escala de uma rede de transporte, que tenderá a ser vista como um serviço partilhado. Um exemplo da aplicação da EVPLAN seria por exemplo, o alojamento da rede privada de uma empresa dentro da rede de um operador de telecomunicações.

Espera-se que as E-LAN tragam a vantagem de efetuar ligações multiponto-multiponto com menos complexidade que tecnologias como Frame Relay ou ATM. Deste modo, lidar com serviços de *multicast* que exigem uma largura de banda e QoS consideráveis, como por exemplo vídeo-conferencias, tornar-se ia bastante mais fácil. A possibilidade de coexistirem mais do que uma EVPLAN e EPL na mesma UNI está presente, sendo a multiplexagem realizada através das VLANs da trama Ethernet.

### 3.3.3 E-TREE

Os serviços E-Tree estão definidos na especificação MEF 6.1, e sofreram um adiamento da sua aprovação por serem considerados uma derivação dos serviços E-LAN [16, pág 10]. No E-Tree é definida uma UNI raiz, e mais do que uma UNI folha. Esta distinção destina-se a um melhor controlo do tráfego, sendo possível existir tráfego bidirecional entre a UNI raiz e as UNI folhas, mas sendo impossível tráfego entre UNI folhas. Em cenários mais sofisticados é possível existir mais do que uma UNI raiz e neste caso, poderá existir comunicação entre UNI raízes e ambas as UNI raízes poderão comunicar com as UNI folhas. A figura (3.5) ilustra este tipo de serviço.

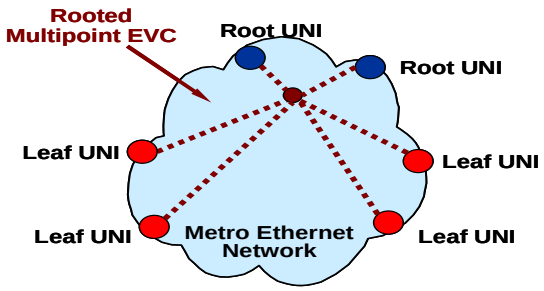



Figura 3.5: Serviço E-TREE (Fonte: MEF)

No serviço E-Tree, as ligações poderão também ser orientadas à porta física da UNI, ou à VLAN *tag*. No caso das ligações serem orientadas à porta física, o serviço é chamado **Ethernet Private Tree Service (EP-Tree)**. Este serviço foi desenhado para ser usado por clientes que queiram ligar vários locais que partilhem o mesmo serviço distribuído. O EP-Tree prevê a preservação da *tag* VLAN, para assim o cliente poder configurar livremente Virtual LANs, sem ocorrerem conflitos com a ligação ao operador [29].

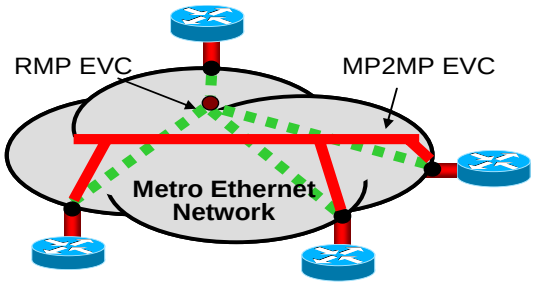



Figura 3.6: Serviço EVP-Tree service (Fonte: MEF)

Os serviços E-Tree com ligações orientadas à *tag* VLAN são denominados **Ethernet Virtual Private Tree Service (EVP-Tree)**. Este é um dos serviços mais interessantes para distribuição de serviços de vídeo, devido à potencial poupança de recursos de largura de banda no fornecimento de serviços *multicast*. O facto das ligações EVC serem orientadas à *tag* VLAN levam o operador a poder configurar remotamente a topologia lógica da EVP-Tree

e poderem existir vários serviços Ethernet orientados à *tag* VLAN na mesma rede. Estas características atribuem ao EVP-Tree uma elevada flexibilidade.

Na figura (3.6) é ilustrado um cenário onde três nós partilham um serviço de EVPLAN (a vermelho, traço contínuo) e um serviço de EVP-Tree (a verde, tracejado). Neste cenário o nó raiz da EVP-Tree poderá ser por exemplo, um servidor de telecomunicações que está a transmitir vídeo em *multicast*, para todos os clientes da ligação.

## 3.4 Tecnologias

Existem essencialmente duas tecnologias que tentam ir de encontro aos atributos definidos pelo MEF e desta maneira, candidatam-se à implementação massiva nas redes de transporte de telecomunicações. A tecnologia MPLS-TP suportada pelo Internet Engineering Task Force (IETF) e ITU, que será abordada em maior detalhe no próximo capítulo. E a tecnologia Provider Backbone Bridge Traffic Engineering (PBB-TE) apoiada pelo IEEE, entidade também responsável pela tecnologia Ethernet.

Atualmente, muitos dos grandes operadores de telecomunicações anunciaram o MPLS-TP como a sua escolha para a tecnologia Carrier Ethernet. Desta forma o presente documento estudará esta solução com maior detalhe. Quanto à tecnologia PBB-TE, esta ainda não cessou o seu desenvolvimento contudo não está a captar o mesmo nível de atenção que a tecnologia concorrente.

# Capítulo 4

# MPLS-TP

O presente capítulo estudará a tecnologia MPLS-TP. A tecnologia é um perfil de funcionamento da tecnologia Multi Protocol Label Switching (MPLS) para a rede de transporte, cuja normalização está a ser desenvolvida através de um esforço conjunto entre o IETF e o International Telecommunication Union - Telecommunication Standardization Setor (ITU-T) desde 2008 [4]. Deste modo, será importante um estudo dos princípios básicos da tecnologia MPLS para uma melhor compreensão da tecnologia MPLS-TP. As soluções baseadas neste perfil constituem atualmente a principal aposta dos operadores de telecomunicações para a rede de transporte, soluções estas, que vão ao encontro dos atributos MEF, estudados no capítulo anterior.

O estudo desta tecnologia, será acompanhado com demonstrações do seu emprego, sendo usado para isso o equipamento IXIA, XM12 High Performance Chassis, com cartas Ethernet. O equipamento é controlado pelo software profissional IxNetwork 6.0.400.14, a imagem (4.1) é uma captura do painel de vista geral do software, onde se pode ver uma ligação Ethernet a 1 Gbps entre duas portas.

## 4.1 MPLS

O aparecimento da tecnologia MPLS, remonta ao inicio dos anos 90 quando surgiram várias tecnologias baseadas em comutação de pacotes IP, que tinham como propósito resolver os problemas de escalabilidade e taxa de encaminhamento baseado em IP [32, pág 1-2]. Estes problemas provêm do aumento das tabelas de encaminhamento IP, que levam a tempos de pesquisa mais elevados e, logo a maiores atrasos de encaminhamento. Em 1996, o IETF juntou-se ao desenvolvimento desta tecnologia, e em 1997 foi criado o MPLS Working Group com o objetivo de normalizar a tecnologia [32, pág 1]. Este esforço teve como produto, o documento RFC 3031 [33], finalizado em 2001, com o título *Multiprotocol Label Switching Architecture*. O documento RFC 3209 *RSVP-TE: Extensions to RSVP for LSP Tunnels*, lançado em 2001, assume também alguma importância nesta tecnologia, e ambos constituem os pilares definidores da tecnologia MPLS.

O MPLS opera entre a camada de ligação de dados e a camada de rede, do modelo OSI, é por isso muitas vezes referida como uma tecnologia de camada OSI 2.5. A tecnologia suporta o transporte de tecnologias baseadas em comutação de circuitos, através de emulação de circuitos de virtuais, podendo assim transportar serviços baseados em comutação de pacotes, ou circuitos de forma unificada e transparente. As siglas *Multi Protocol*, devem-se à sua

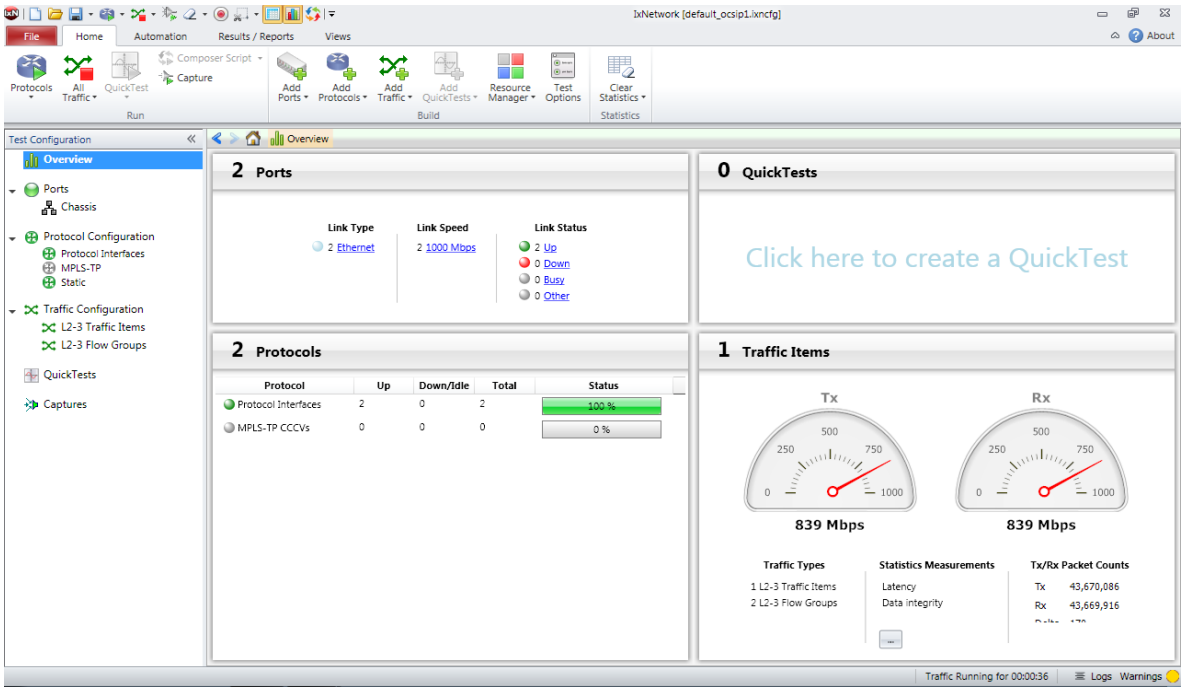



Figura 4.1: Vista geral - IxNetwork 6.0.400.14 (Fonte: Software IxNetwork 6.0.400.14)

capacidade de poder encapsular várias tecnologias, tal como, tramas Ethernet, tramas Frame Relay ou células Asynchronous Transfer Mode (ATM). A operação base da tecnologia, à qual se devem as siglas de *Label Switching*, consiste em atribuir a cada pacote uma ou mais *labels* de tamanho fixo, que definem o caminho que este pacote tomará ao atravessar a rede.

### 4.1.1 Label MPLS

A figura (4.2) ilustra a *label* MPLS, e a sua localização no modelo OSI. É importante assinalar o facto de a tecnologia permitir encapsular mais do que uma *label* em cada pacote, criando desta maneira uma pilha de *labels*.

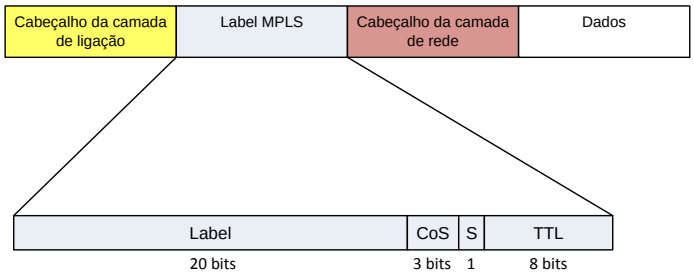



Figura 4.2: *Label* MPLS

O pacote possuí 4 campos:

- O campo *label* possuí 20 bits, e é onde está definida a *label* de encaminhamento do pacote, dentro da rede MPLS. Os 20 bits permitem assim criar cerca de 1 milhão de diferentes *labels*.

- O campo QoS, possuí 3 bits e é usado para implementação de qualidade de serviço, através da atribuição de prioridades. Inicialmente o campo não possuía uma função específica, era um campo experimental (EXP).

- O campo S, possuí apenas um bit, e tem como objetivo indicar se a pilha de *labels* está vazia, de modo que, apenas será '1' na *label* que se encontra no fundo na pilha.

- o campo Time To Live (TTL) possuí 8 bits, que têm como objetivo impedir que os pacotes circulem indefinidamente na rede, através da limitação do número se saltos que um pacote pode realizar. Para além desta função o campo TTL pode também ser um mecanismo de endereçamento de pacotes de OAM, veja-se, ao atravessar um *router* o valor TTL da *label* que se encontra no topo da pilha de *labels* é decrementado, e no caso deste valor atingir zero, é verificado se existe uma *label* de OAM imediatamente a seguir à *label* que atingiu o valor 0 de TTL. No caso de possuir, o *router* que efetuou o decremento do TTL de 1 para 0 sofre uma operação de OAM. No caso de não existir nenhuma *label* de OAM, o *router* descarta o pacote.

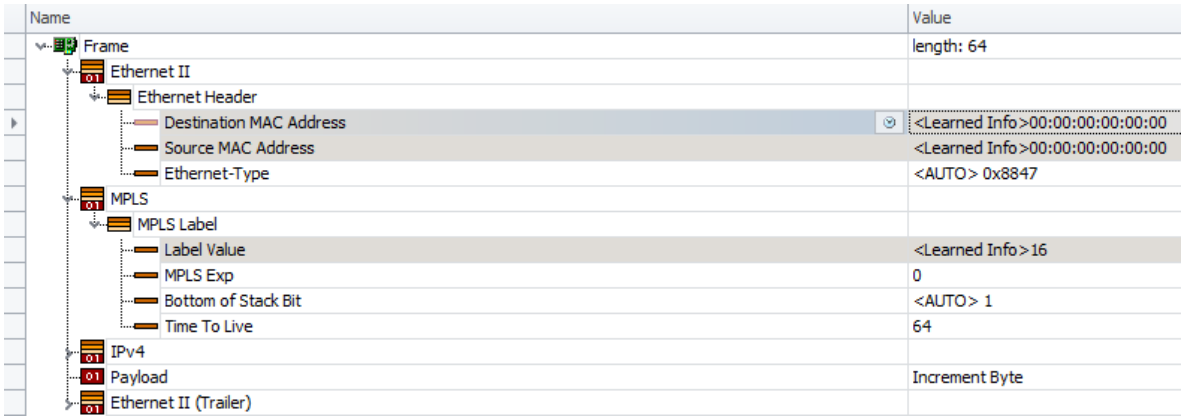



Figura 4.3: *Label* MPLS (Fonte: Software IxNetwork 6.0.400.14)

A figura (4.3), mostra a utilização de uma *label* MPLS entre a tecnologia Ethernet, e a tecnologia IP. São ilustrado os campos do cabeçalho da tecnologia Ethernet II, estudados no capítulo Ethernet, e os campos da *label* MPLS, descritos anteriormente. Como se pode observar, o campo tipo da trama Ethernet II possuí o valor 0x8847, o que indica o transporte de um pacote MPLS com destino *unicast*. A *label* possui o valor 16, não possui classe de serviço, o campo S tem o valor '1', indicando que é a última *label* da pilha, e o seu campo TTL indica que aquela *label* atravessará no máximo 64 *routers*. A imagem foi retirada do software IxNetwork 6.0.400.14.

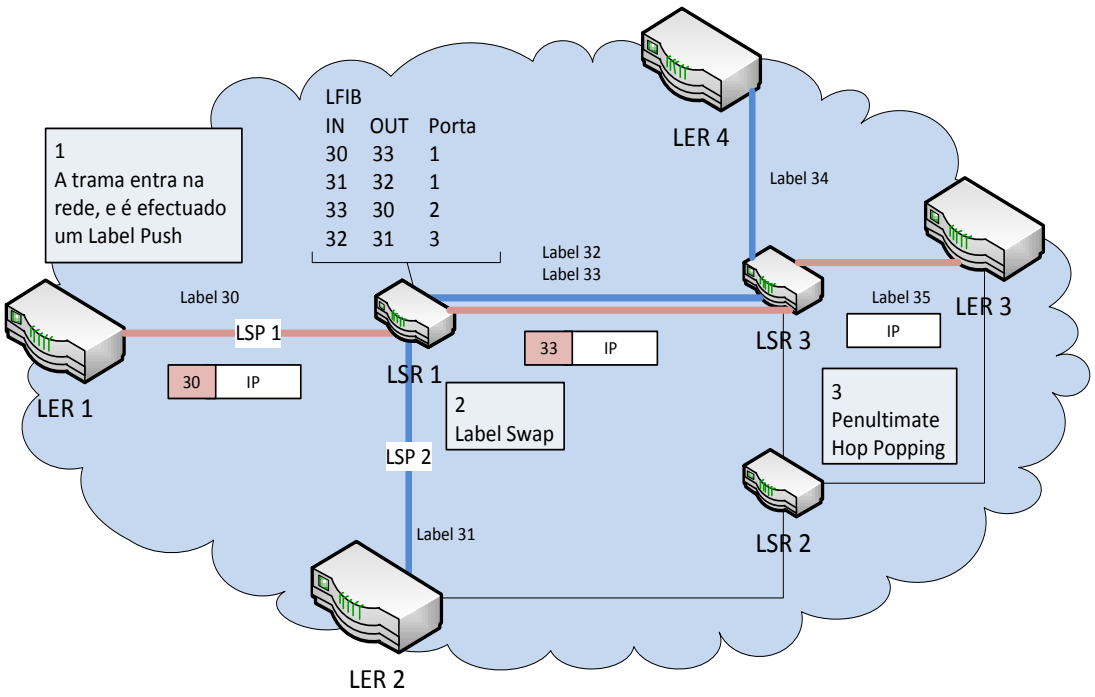



Figura 4.4: Rede MPLS

---

### 4.1.2 Arquitetura

A figura (4.4) ilustra uma rede MPLS. Os caminhos que os pacotes tomam dentro da rede são chamados de Label Switched Paths (LSPs) e podem ser configurados com recurso a diferentes protocolos. Note-se que, um LSP representa uma ligação virtual, que está sobre uma rede não orientada à ligação. Os *routers* responsáveis por encaminhar os pacotes, são classificados quanto ao local que ocupam no LSP, e desta forma temos:

- Os Label Edge Router (LER) que se situam na fronteira da rede, ou seja no inicio ou final de um LSP. Estes equipamentos são responsáveis pelo encapsulamento e posterior encaminhamento, de pacotes que chegam à rede. E são responsáveis por retirar todas as *labels* e encaminhar os pacotes que vão abandonar a rede, contudo a funcionalidade de Penultimate Hop Popping (PHP), permite que seja o ultimo *router* antes do LER, a efetuar esta tarefa, de modo a não sobrecarregar os LERs [16, pág 60]. O encaminhamento de pacotes é realizado com recurso à Label Forwarding Information Base (LFIB).

- Os Label Switching Router (LSR), são os *routers* intermédios de um LSP, e a sua função é encaminhar os pacotes MPLS, com base na *label* situada no topo da pilha de *labels*. Para o realizar recorrem também à LFIB. O mesmo *router*, pode atuar como um LER para um LSP, e como LSR para um LSP diferente.

### 4.1.3 Funcionamento

As operações fundamentais que os *routers* anteriormente descritos, podem realizar no plano de dados, são as seguintes:

- O*Label Push* consiste na adição de *labels* ao pacote. Um pacote ao entrar na rede é encapsulada com a *label* MPLS descrita na figura (4.2), e à medida que avança na rede, poderão ser adicionadas mais labels, criando assim uma pilha de labels. As operações a serem efetuadas sobre o pacote, ocorrerão na *label* que se encontra no topo da pilha.
- O *Label Pop* retira a *label* situada no topo da pilha. Uma variação desta operação será o retirar de toda a pilha de *labels* em simultâneo, que acontece no PHP.
- O *Label Swap* é a troca da label que se encontra no topo da pilha, pode ser vista como um *label pop* seguido de um *label push*, e é uma operação frequente no encaminhamento interno dos pacotes efetuado pelos LSRs.

A tabela de encaminhamento de *labels* a que os *routers* recorrem para encaminhar os pacotes, é designada por LFIB, e cada entrada desta tabela possuí obrigatoriamente uma *label* de chegada, uma *label* de saída e a interface de saída. Podendo também ter outra informação, tal como identificador de interface e/ou endereços IP. Na figura (4.4), encontra-se uma proposta de LFIB para o LSR 1.

O mecanismo de encaminhamento de pacotes, compreende a operação de *label pop* ao pacote que chega ao *router*, a pesquisa da *label* de entrada na LFIB, e o *label push* com a *label* de saída correspondente. Após isto o pacote é enviada para a interface de saída correspondente.

No caso da figura (4.4), o LSR 1 ao receber o pacote proveniente do LER 1, retira a *label* 30 do pacote, e procura na sua tabela esta entrada. Após ser encontrada a entrada correspondente o router, faz um *label push* da *label* 33, e reencaminha o pacote, para a porta 1. Também o LSR 3 fará uma operação semelhante a esta, contudo a sua entrada na LFIB, para a *label* de chegada com o valor 33, indicará a necessidade de efetuar a operação de PHP, que encaminhará o pacote para o LER 3, e este por sua vez enviará o pacote para fora da rede MPLS, com recurso por exemplo a um protocolo IP. Note-se que o exemplo da figura poderá transportar uma trama Ethernet ou IP, e que o caminho de transporte é o LSP 1, mas esta escolha do LSP para cada pacote, pode ser realizada pelo LER e ter por base qualquer parâmetro, tal como endereço MAC, IP, QoS ou tipo de aplicação do pacote.

Uma outra funcionalidade com bastante importância na tecnologia MPLS, é a possibilidade de se criarem túneis LSP. A figura (4.5) ilustra um túnel LSP, onde fluxos proveniente de diferentes *routers* são agregados no mesmo LSP. Após atravessarem o túnel a escolha do caminho a dar a cada pacote é efetuada com base na *label* que o pacote possuía antes de entrar no túnel. Pelo que o *router* que agrega os fluxos, apenas efetua a operação de *label push*, e o *router* que no final do túnel encaminha os pacotes apenas efetua a operação de *label pop*. Desta maneira o túnel LSP, faz uso da possibilidade de empilhar várias labels no mesmo pacote, e permite uma elevada poupança de memória e processamento, assim como a possibilidade de estabelecer Virtual Private Networkss (VPNs) [34, pág 6].

### 4.1.4 Estabelecimento de ligações

As ligações poderão ser estabelecidas manualmente, ou com recurso a protocolos de sinalização, tal como Label Distribution Protocol (LDP) ou Resource Reservation Protocol -

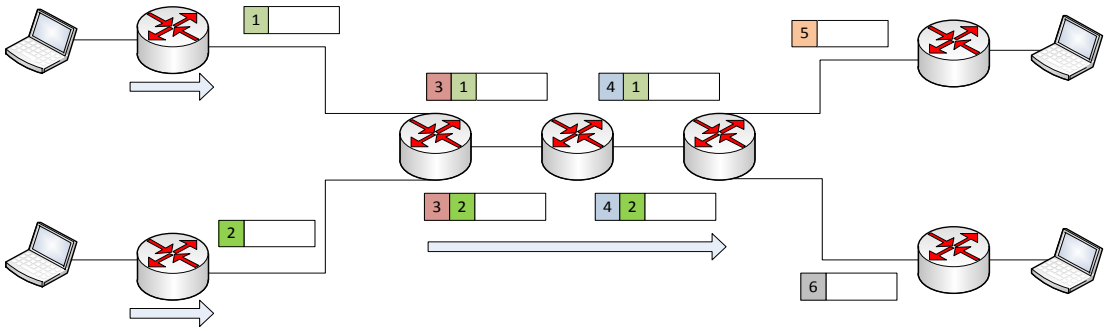



Figura 4.5: Túnel LSP

---

Traffic Engineering (RSVP-TE).

### LDP

O LDP, definido no RFC 5036 configura os LSPs da rede, com recurso a um mapeamento das opções de encaminhamento provenientes de protocolos IP (tal como Interior Gateway Protocol (IGP)). Após este mapeamento são trocadas mensagens entre LSRs adjacentes onde é transmitida informação sobre as opções de encaminhamento que cada um toma (com base no mapa de encaminhamento proveniente de protocolos IP). Os LSPs resultantes desta comunicação podem apenas possuir dois LSRs (à semelhança do encaminhamento IP, salto a salto) ou podem resultar LSPs que atravessam toda a rede. Para um eficiente uso dos recursos é associado a cada LSP um Forwarding Equivalence Class (FEC), onde são especificados todas os pacotes que poderão usar aquele LSP.

Os dois LSRs que trocam mensagens de controlo para encaminhamento de pacotes chamam-se *LDP peers*, e na sua sessão LDP é possível que ambos adquiram conhecimento sobre o encaminhamento de *labels* um do outro em simultâneo, devido à bidireccionalidade do protocolo LDP. Os 4 tipos de mensagens LDP são as seguintes:

- Mensagens de descoberta. São usadas para anunciar e manter a presença do LSR na rede. São mensagens de *"Hello"* enviadas periodicamente por UDP, contudo se enviadas para o estabelecimento de novas sessões usam o protocolo Transmission Control Protocol (TCP).

- Mensagens de sessão. Para dois *LDP peers* iniciarem a troca de mensagens necessitam de estabelecer uma sessão LDP. As mensagens de sessão destinam-se a estabelecer, manter e terminar sessões LDP entre *LDP* peers, e usam sempre o protocolo TCP.

- Mensagens de anúncio. Estas mensagens são usadas para criar, modificar e apagar os vínculos de *labels* aos FECs. Usam também o protocolo TCP.

- Mensagens de notificação. São mensagens de aconselhamento e podem conter informação sobre erros. São também enviadas através de TCP [15, pág 149-157].

### RSVP-TE

O RSVP-TE resulta da adição de funcionalidades de engenharia de tráfego ao protocolo Resource Reservation Protocol (RSVP) sobre TCP. O protocolo RSVP é caraterizado por permitir estabelecer ligações nas redes IP com reserva de largura de banda.

Quando é definido um RSVP com recurso ao MPLS o fluxo de tráfego resultante possuí uma elevada flexibilidade e robustez, sendo o produto final deste protocolo um túnel LSP. A *label* atribuída aos pacotes pelo *router* de acesso ao túnel LSP, definem completamente o fluxo neste túnel, e o túnel cumpre assim os requisitos de ligação exigidos pelo protocolo RSVP-TE.

O túnel é configurado da seguinte maneira: Inicialmente o *router* de acesso ao túnel envia uma mensagem *PATH* na direção do caminho da rede desejado. Cada LSR intermédio confirma se possuí a largura de banda pedida pelo *router* anterior, e encaminha a mensagem até ao último *router* do LSP pretendido. O último *router* envia uma mensagem de resposta *RESV* que atravessará todos os *routers* intermédios, reservando desde logo a largura de banda necessária e as *labels* a serem usadas. Após isto o túnel será mantido e atualizado através da troca de mensagens entre *routers* adjacentes. O protocolo RSVP-TE engloba as mensagens do anterior protocolo RSVP e adiciona novos tipos de mensagem [15, pág 167-179].

Em algumas aplicações é útil associar conjuntos de túneis LSPs, por exemplo um túnel LSP poderá ser dividido em vários túneis, ou alguns túneis poderão ser agregados, o que é útil no transporte de protocolos que comutação de circuitos. Estes túneis são conhecidos como TE Tunnels (*Traffic Engineered Tunnels*).

Uma outra funcionalidade interessante do RSVP-TE é o mecanismo de proteção local *Fast Reroute*. Esta funcionalidade permite a proteção de ligações ou LSPs, através de caminhos pré-estabelecidos. Mais detalhe sobre este mecanismo pode ser encontrado no documento RFC 4090 (*Fast Reroute Extensions to RSVP-TE for LSP Tunnels*).

### BFD, LSP Ping e LSP traceroute

Estes três recursos usados pela tecnologia MPLS têm como função monitorizar o estado das ligações e dos LSPs. Embora não sejam funcionalidades originais desta tecnologia foram adaptadas e como se verá adiante, constituem a base de várias funcionalidades de OAM adicionadas à tecnologia MPLS-TP.

O Bidirectional Forwarding Detection (BFD) é um protocolo usado para detetar falhas de ligação entre dois *routers*. Para isso estabelece uma sessão entre os *routers* adjacentes que consiste num *handshake* de três mensagens. Após estabelecer uma sessão, são enviadas periodicamente mensagens de *Hello* e no caso destas mensagens não obterem uma resposta, considera-se que a ligação possui uma falha. Este mecanismo relativamente simples informa automaticamente sobre uma falha no plano de dados da tecnologia MPLS. Este mecanismo é descrito com maior detalhe no documento RFC 5880 *Bidirectional Forwarding Detection*.

O LSP Ping é um pedido de resposta usado para validar caminhos LSP, análogo ao Ping usado na tecnologia IP. A utilização do LSP Ping pode validar LSPs configurados tanto através do protocolo LDP como do RSVP-TE, e indica que tanto o plano de dados como o plano de controlo da tecnologia MPLS estão a funcionar. O uso do LSP Ping em conjunto com o BFD permite por exemplo, apurar se a falha é de origem física (ambos os testes falham), ou se a falha tem origem no plano de controlo (apenas o LSP Ping falha). O emprego conjunto deste recursos está descrito com maior detalhe no RFC 5884 *BFD for MPLS LSPs*.

Uma outra funcionalidade que também se mostra de grande utilidade é o LSP *traceroute*, especialmente em túneis LSP. O LSP *traceroute* permite obter informações sobre todos os *routers* que o LSP em análise atravessa, através da aplicação de sucessivos LSP Pings ao mesmo LSP. Mais informações sobre esta funcionalidade podem ser encontradas no *draft LSP-Ping extensions for MPLS-TP* (draft-nitinb-mpls-tp-lsp-ping-extensions01).

### 4.1.5 MPLS para a rede de transporte

A tecnologia MPLS, muitas vezes chamada de IP/MPLS devido à frequência com que encapsula pacotes IP é atualmente uma tecnologia madura, que tem já algum peso nas redes de transporte. Apresenta como vantagens para este contexto de operação a possibilidade de agregar facilmente tráfego, elevada taxa de encaminhamento dos *routers* devido à redução das entradas das tabelas de encaminhamento, capacidade de suportar engenharia de tráfego, através do protocolo RSVP-TE, eficientes mecanismos de recuperação, suporte a VPNs, e o suporte a vários tipos de tecnologia (Generalized Multi-Protocol Label Switching (GMPLS) [32]).

Apesar das vantagens enumeradas anteriormente, o MPLS possui ainda alguns aspetos que desagradam aos operadores tanto no encaminhamento de tráfego de pacotes, como na comutação de tráfego de voz. Quanto ao encaminhamento de dados é necessário tecnologia adicional sob a tecnologia IP, é necessária uma complexa configuração da rede devido ao recurso do MPLS a muitas outras tecnologias (LDP, OSPF, EBGP, ...). A estreita relação entre a tecnologia MPLS e a tecnologia IP é algo que desagrada também aos operadores.

Quando se trata da possibilidade da tecnologia substituir as robustas tecnologias baseadas em TDM, tal como o SDH, esta é considerada inapta devido ao uso de protocolos de encaminhamento dinâmicos que aumentam os tempos de recuperação a falhas, e devido à inconsistência das suas ferramentas de OAM para as redes de transporte. Os operadores necessitam de uma tecnologia que possua caminhos de recuperação pré-determinados, ferramentas avançadas de OAM e planos de controlo o mais simples possível.

## 4.2 MPLS-TP

A tecnologia MPLS, apesar de ter crescido e amadurecido no mercado, não preenchia todos os requisitos necessários para operar numa rede de transporte. E foi com esta premissa, que em 2006, o ITU-T definiu uma tecnologia de rede de transporte baseada no MPLS, o Transport Multi-Protocol Label Switching (T-MPLS). A arquitetura da tecnologia foi definida no documento ITU-T G.8110.1, a interface, no documento G.8112, o equipamento no documento G.8121 e os mecanismos de proteção no documento G.8131 [35, pág 8]. A nova tecnologia pretendia ainda atribuir escalabilidade, aprovisionamento estático, e engenharia de tráfego de transporte através da implementação de novas funcionalidades de operação, administração e gestão (OAM)(G.8113, G.8114, G.7710, G.7712, G.8151)[1].

Em 2008 o IETF juntou-se à normalização da tecnologia, contudo não concordou com algumas das especificações existentes e tentou impedir a normalização da proteção em anel (G.8132) e das funcionalidades de OAM (G.8114), alegando incompatibilidades com a anterior tecnologia IP/MPLS.

[1]Implementações estas, que tendem a ir de encontro aos atributos MEF para um rede de transporte, estudados anteriormente.

De modo a solucionar as referidas incompatibilidades, as duas entidades decidiram cooperar através de uma *Joint Working Team*, onde o IETF ficou responsável por publicar os Request For Commentss (RFCs) da nova tecnologia, e o ITU-T ficou encarregue de rever as anteriores recomendações do T-MPLS, para que estas fossem de encontro ao pretendido para a nova tecnologia. Foi decidido também alterar a designação para MPLS-TP, de modo a assinalar a rutura com o anterior T-MPLS. Mais detalhe relativo ao papel de cada entidade na *Joint Working Team*, pode ser encontrada no RFC 5317 (JWT Report on MPLS Architectural Considerations for a Transport Profile) [35].

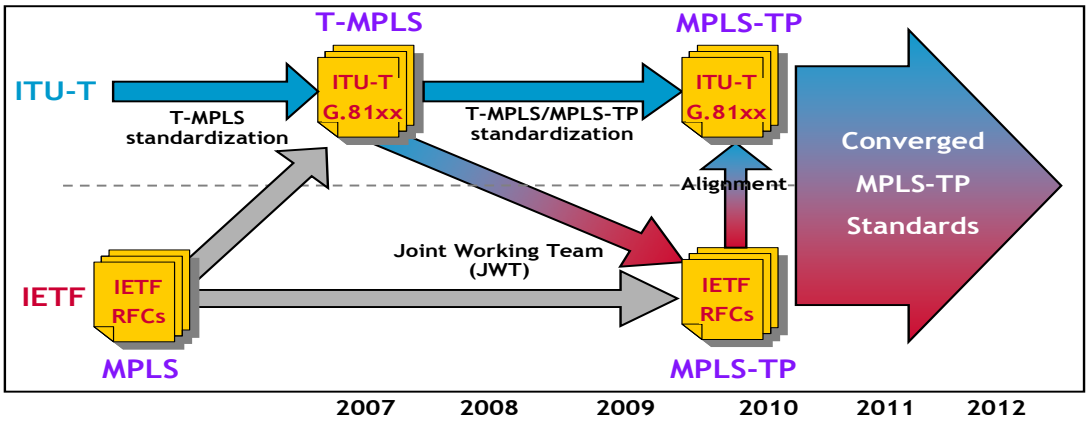



Figura 4.6: Convergência da tecnologia MPLS-TP (fonte [4])

A imagem (4.6) retirada do documento [4], ilustra o esforço de ambas as entidades acima mencionadas, para a convergência da tecnologia. Atualmente a tecnologia é alvo de elevado interesse por parte dos operadores de telecomunicações, e é apontada como a próxima tecnologia dominante na rede de transporte, contudo não se encontra ainda totalmente normalizada.

### 4.2.1 Funcionalidades

Visto a tecnologia MPLS-TP resultar da implementação de novas funcionalidades de transporte à anterior tecnologia MPLS, é importante estudar quais as novas funcionalidades adicionadas até ao momento e o que implicam no funcionamento de toda a rede. Serão abordadas também as funcionalidades que foram retiradas, ou que se tornaram opcionais e o motivo dessa alteração. A tecnologia pode ser dividida em quatro grandes áreas: o plano de dados, o plano de controlo, OAM e os mecanismos de sobrevivência.

### 4.2.2 Plano de Dados

- O MPLS-TP faz uso das anteriores arquiteturas e mecanismos de encaminhamento de pacotes MPLS. Contudo deve operar sem recurso à tecnologia IP e para isso não pode recorrer a mecanismos de encaminhamento de pacotes, provenientes de protocolos superiores (tal como acontece no LDP), pelo que deve ser capaz de possuir estas funcionalidades nos seus planos de controlo e OAM [35] [33, pág 5].

- Os LSPs que anteriormente eram apenas unidirecionais, passam a poder ser também bidirecionais e nesse caso congruentes [16] [36, pág 7] [35, pág 28]. No MPLS os pacotes que circulavam numa ligação ponto-a-ponto em diferentes direções poderiam percorrer diferentes caminhos, o que tornava mais difícil monitorizar e controlar o tráfego. Ora para que isto não aconteça, as ligações bidirecionais do MPLS-TP obrigam a que o caminho percorrido em ambos os sentidos da ligação seja o mesmo e assim todo o tráfego das ligações bidirecionais atravessará os mesmos LSRs.

- A funcionalidade de PHP (*Penultimate Hop Popping*) tornou-se opcional. Esta opção foi largamente discutida já que ao ser mantida, poderão ser retiradas *labels* de OAM a pacotes que teriam como função atuar no ultimo *router* de um LSP. Este facto tornaria impossível aplicar as funções normais de OAM aos últimos *routers* de um LSP [35, pág 74]. Foram elaboradas soluções para a possibilidade do PHP estar ativo ou desativo e atualmente a funcionalidade encontra-se desativa por defeito. A funcionalidade de agregar vários LSPs foi também desativada, pelos mesmos motivos que o PHP.

- A funcionalidade de Equal Cost Multi-Path (ECMP) foi desativada. Esta permite que o tráfego entre dois *routers* seja dividido entre dois ou mais caminhos, se estes apresentarem igual custo (*Loud Balancing*). Ora para além de dificultar a monitorização do tráfego [36], também implica que os pacotes de OAM, que circulam nos mesmos LSPs que os pacotes de dados possam seguir um percurso diferente. Entende-se que o tráfego de dados e o tráfego OAM devem partilhar sempre o mesmo percurso [35, pág 70].

- As ligações multiponto-a-multiponto foram suprimidas, já que este tipo de ligação não faz sentido numa rede de transporte. Desta forma a tecnologia MPLS-TP poderá apenas operar em ligações ponto-a-ponto unidirecionais e bidirecionais, e ligações ponto-a-multiponto unidirecionais.

### 4.2.3 Plano de Controlo

No plano de controlo foram definidos duas opções, o controlo estático e o controlo dinâmico:

- O controlo estático recorre a Network Management System (NMS) para o estabelecimento de LSPs estáticos.

- O controlo dinâmico usa protocolos de controlo independentes (OSPF-TE ou RSVP-TE) para total automação da rede [36].

Foram também definidas *labels* próprias para implementar na tecnologia funções de OAM transportadas no mesmo canal de comunicação que os dados. Esta implementação será estudada com maior detalhe em seguida. Note-se que tanto o controlo dinâmico como o controlo estático devem estabelecer uma clara divisão entre os planos de dados e de controlo [36].

#### Generic Associated Channel

Com o propósito de assegurar coerência entre os mecanismos de OAM e os LSPs, os pacotes de OAM circulam nas mesmas ligações que os demais pacotes. Para isto os pacotes de OAM possuem *labels* adicionais que os identificam e acionam os mecanismos de OAM nos *routers* pretendidos.

Na tecnologia MPLS os mecanismos de OAM dos LSPs são realizados pelo protocolo IP, contudo um dos requisitos da tecnologia MPLS-TP é a independência de tecnologias de camada superior, nomeadamente IP. Desta forma, e para manter a interoperabilidade com a tecnologia IP/MPLS, os pacotes IP de OAM continuam a ser aceites no MPLS-TP, no entanto será o mecanismo de Generic Associated Channel o principal meio de gerir e controlar as redes MPLS-TP.

O conjunto de mensagens adicionais chama-se Generic Associated Channel (G-ACh) e pretende ser a generalização do mecanismo Associated Channel Header (ACh), usado inicialmente para assinalar pacotes com funções de OAM no contexto dos *pseudowires*[2]. A G-ACh, no caso de possuir informações de OAM destinas a *pseudowires*, é em tudo igual à mensagem ACH. No entanto quando é direcionada aos LSPs do MPLS-TP, o G-ACh possui uma *label* que informa sobre este facto,chamada G-ACh Label (GAL). A GAL é uma *label* semelhante às *labels* usadas para encaminhamento, contudo possuí um valor (13) que está reservado para esta finalidade, e é por isso interpretada de forma diferente pelos *routers* [37].

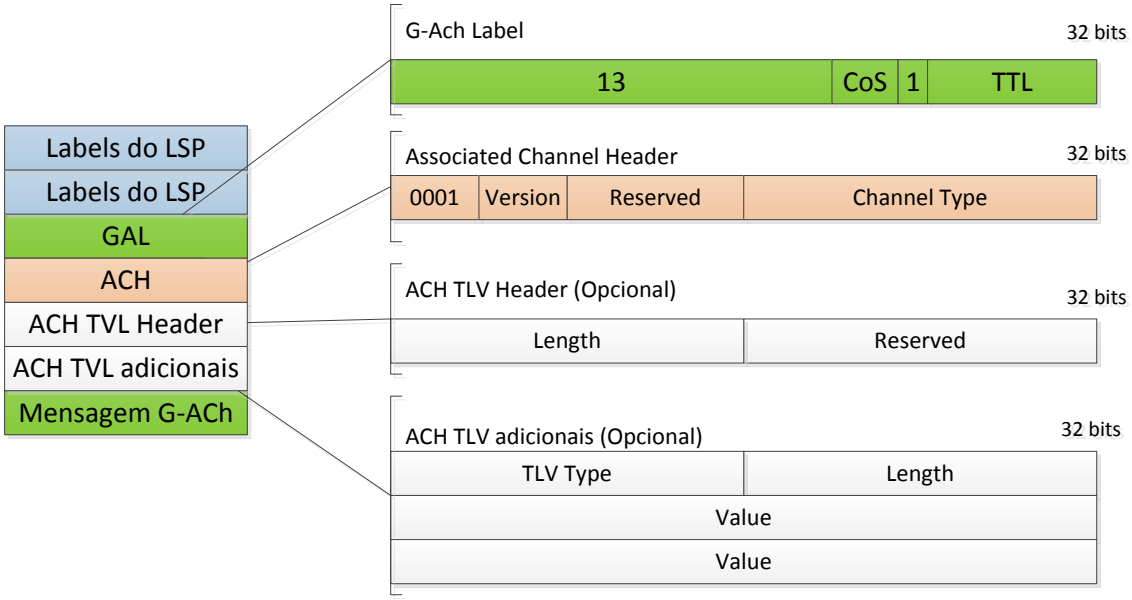



Figura 4.7: Formato do pacote G-Ach para um LSP.

A figura (4.7) retrata um pacote G-ACh MPLS-TP. Após as *labels* LSP de encaminhamento aparece a *label* GAL, com o número reservado 13, significando que o atual pacote possui informações de OAM. Seguidamente encontra-se a *label* ACH, definida antes do aparecimento do MPLS-TP nos mecanismos de PW. Os campos *Version* e *Reserved* desta *label* devem estar a 0, enquanto o campo *Channel Type* deverá conter informar sobre o protocolo que está a ser alvo de controlo.

As *labels* ACH TVL Header e ACH TVL são opcionais e têm como função contextualizar as mensagens de G-ACh, podendo por exemplo possuir informação sobre o destino ou fonte

[2]Um *pseudowire* (PW) é a emulação de um ligação física dedicada sobre uma rede de comutação de pacotes. Na tecnologia MPLS o *pseudowire* é aproximado por um LSP com requisitos e funcionalidades específicas (RFC 4385).

da mensagem [37, pág 7]. O campo *length* do ACH TLV Header indica o tamanho em bytes do restante conjunto de TLV. O campo *Reserved* está reservado para uso futuro, devendo atualmente encontrar-se a 0 e ser ignorado pelo *routers* recetor. Quanto aos ACH TLV adicionais, o campo *TLV Type* define o formato e função do campo *Value*, e o campo *length* o seu tamanho. Os campos *Value* poderão conter sub-TLVs [37].

Uma descrição mais completa da G-Ach poderá ser encontrada no documento RFC 5586 (MPLS Generic Associated Channel) [37]. Note-se que o documento não faz restrições quanto à localização destas *labels* na pilha de *labels* em outros ambientes MPLS contudo, quando usada no MPLS-TP estas *labels* deverão encontrar-se sempre no fundo da pilha de *labels* e assim sendo, o bit S do GAL deve ser sempre 1 e o GAL deverá ser sempre seguido da *label* ACH [37, pág 10].

### 4.2.4 OAM

As funcionalidades OAM foram claramente as que mais enriqueceram com a elaboração desta nova tecnologia. Foram implementados mecanismos pró-ativos de deteção de falhas unidirecionais e bidireccionais, gestão de falhas com mecanismos de recuperação [36] e implementação de gestão de performance com recurso a medições de parâmetros, como perdas e atrasos de pacotes. De seguida são resumidas as novas funcionalidades de OAM do MPLS-TP em duas funcionalidades principais, deteção e localização de falhas, e monitorização de performance. Note-se que estas possuem sub-funcionalidades que fazem uso de anteriores recursos do MPLS, nomeadamente o BFD, o LSP *Ping* e o LSP *trace*.

#### Deteção e Localização de Falhas

O MPLS-TP para detetar e localizar falhas usa as seguintes sub-funcionalidades:

- Continuity Check (CC). Identifica falhas de modo rápido e pró-ativo. Recorre as anteriores funções de BFD e LSP *Ping*.
- Connectivity Verification (CV). A pedido, permite localizar a falha após ser detetada. Recorre às anteriores funções de BFD e LSP *Ping*.
- Loopback. Permite ao operador colocar um LSP em modo *Loopback* (Os pacotes do LSP retornaram à fonte que os originou). Útil para testes e medições. Recorre a *Labels* G-ACh e LSP Ping.
- Lock. Permite ao operador retirar de serviço um LSP. Após iniciada esta funcionalidade somente tráfego de OAM poderá ser enviado. Os LSPs necessitam de entrar em modo *Lock* para posteriormente serem colocados em modo *Loopback* para medições. Recorre a *Labels* G-ACh ou LSP Ping.
- Remote Defect Indication (RDI). É usado pelos *LER* para comunicar defeitos. Recorre ao BFD.

#### Monitorização de Performance

- Atraso. Permite medir o atraso numa ligação. Usa uma nova ferramenta, o *Delay Measurement* (DM).

- Perda de pacotes. Permite contar os pacotes perdidas numa ligação. Usa também uma nova ferramenta, o *Loss Measurement* (LM).
- Taxa de Transferência. Permite medir a taxa de transferência de uma ligação. Recorre ao LM.
- Variância do Atraso. Permite medir a variância do atraso (*jitter*) de pacotes nas ligações. Recorre à ferramenta DM.

### 4.2.5 Sobrevivência

O conceito de sobrevivência no contexto das redes de telecomunicações consiste na sua capacidade para continuar a oferecer serviços quando ocorrem falhas, ou degradação dos seus recursos. Os mecanismos de proteção para estas falhas no MPLS-TP, usam caminhos determinísticos e esquemas de proteção 1:1, 1+1, 1:N, tanto em arquiteturas lineares como em anel e devem permitir à rede tempos de recuperação a falhas inferiores a 50 ms. A rede deve obedecer a mensagens de OAM provenientes do operador e todas as proteções devem poder ser aplicadas a ligações ponto-a-ponto e ponto-multiponto [35, pág 88-89].

Atualmente os mecanismos de proteção estão a ser alvo de elevada atenção por parte da *Joint Work Team* mas não existem ainda RFCs que normalizem estes mecanismos. O recente RFC 6372 [38] indica os requisitos esperados para a sobrevivência de redes MPLS-TP e o RFC 5317 [35] apresenta alguns mecanismos baseados nas anteriores recomendações do ITU-T para o T-MPLS.

Nesta secção serão apresentados e discutidos algumas soluções baseadas nos requisitos e mecanismos apresentados no RFC 5317 [35] e RFC 5317 [35]. Para além destes, ter-se-à em consideração também os *drafts* [5] e [39] relativos a este tema.

#### Proteção linear

Os mecanismos de proteção linear podem ser subdivididos nos esquemas de proteção 1+1, e 1:N que engloba o esquema 1:1. Ao ser usada uma proteção 1:1, um caminho de proteção é salvaguardado para proteger o caminho de trabalho e deve possuir toda a capacidade de largura de banda que o caminho de trabalho suporta [38].

A figura (4.8) retrata o esquema de proteção 1:1. Em condições normais a informação é transmitida através do caminho de trabalho, enquanto o caminho de proteção se encontra inativo, podendo ainda assim transmitir pacotes de OAM ou informação de baixa prioridade [38, pág 26-27].

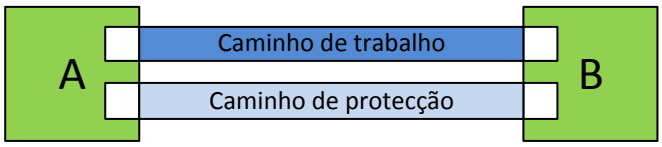



Figura 4.8: Esquema de proteção 1:1 linear

Quando é detetada uma falha no caminho de trabalho, a transmissão de informação passa a ser realizada através do caminho de proteção. Ambos os *routers* necessitam de coordenar

qual o caminho que se encontra em funcionamento, pelo que é necessário um protocolo de coordenação[3].

No caso da proteção linear 1:N, um caminho de proteção é reservado para proteger N caminhos de trabalho. Nesta situação é natural que o caminho de proteção não possua largura de banda suficiente para proteger simultaneamente os N caminhos de trabalho, pelo que este problema deve ser minimizado através de funcionalidades de priorização de falhas no protocolo de coordenação, de modo a dar prioridade a tráfego crítico. Deverá ser possível também, reverter o tráfego novamente para o caminho de trabalho que anteriormente sofreu a falha, de forma automática, para assim poderem ser atendidas outras falhas da rede [38, pág 28].

O caso da proteção linear 1+1 é em parte idêntico ao descrito anteriormente para a proteção 1:1, contudo nesta situação tanto o caminho de trabalho como o caminho de proteção transportam a mesma informação e desta forma o *router* que recebe a informação dos dois caminhos poderá escolher qual a informação a processar, com base na qualidade do sinal. Este cenário, apresenta um elevado desperdício de recursos, contudo torna desnecessário um protocolo de coordenação de estados entre os *routers* [38, pág 28], o que poderá representar uma vantagem, especialmente enquanto os mecanismos de coordenação de estados não se encontrarem suficientemente estáveis.

Anteriormente, foi referida a necessidade da existência de um protocolo de coordenação de estados entre os *routers* para arquiteturas de proteção lineares 1:N. Ora o *draft* [5], vem ao encontro desta necessidade e propõe o Protection State Coordination Protocol (PSC).

**Protection State Coordination Protocol (PSC)**

Para abordar esta questão o protocolo PSC considerou o domínio de proteção como a distancia total de um LSP ou seja, todo o percurso entre os seus dois LERs [5, pág 4]. O PSC define que após um LER tomar conhecimento da falha deve alterar imediatamente o seu estado para usar o caminho de proteção e enviar 3 mensagens consecutivas PSC através do caminho de proteção, para que o outro *router* altere também o seu estado e use o caminho de proteção. Note-se que as mensagens PSC apenas devem ser transmitidas sobre os caminhos de proteção. O mecanismo PSC, tal como foi aqui apresentado, apenas possuí uma fase de coordenação, não se tratando por isso de um mecanismo de *handshake*, o que vai de encontro aos requisitos do documento [38].

Após o envio das 3 mensagens PSC consecutivas, devem ser enviadas mensagens PSC em intervalos de 5 segundos, com o objetivo de verificar que a sessão PSC continua ativa. A frequência do envio das 3 primeiras mensagens assim como o período das mensagens PSC que sucedem a estas, devem poder ser configuradas pelo operador, contudo é recomendado que o período de envio das 3 primeiras mensagens não ultrapasse os 3.3ms, para que o tempo de recuperação a falhas de 50ms não seja ultrapassado.

A figura (4.9) que se segue foi retirada do documento [5] e retrata a lógica de funcionamento da geração de mensagens PSC.

Para cumprir os requisitos de prioridade das mensagens PSC, é necessário que a geração destas mensagens possua um algoritmo de priorização que considere *triggers* gerados por várias fontes. As ferramentas de OAM de deteção e localização de falhas, assim como ferramentas de monitorização de performance devem naturalmente constituir entradas do gerador

[3]O protocolo Protection State Coordination Protocol (PSC) em fase de elaboração, tenta responder a esta questão e será estudado mais à frente.

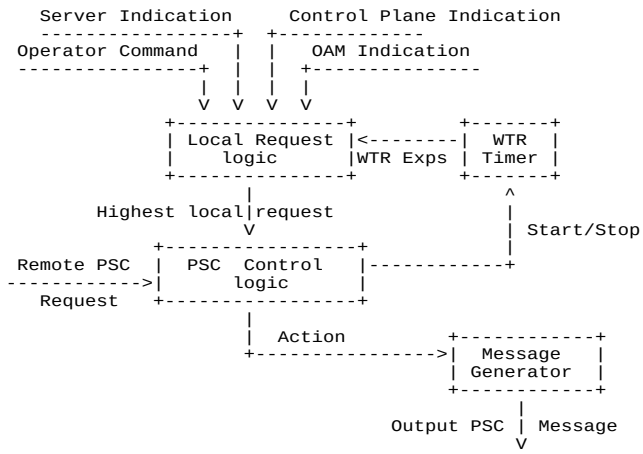



Figura 4.9: Lógica de funcionamento da geração de mensagens PSC (Fonte [5])

---

de mensagens PSC, já que são a principal ferramenta do MPLS-TP para gerar estas mensagens. Os comandos introduzidos pelo operador da rede devem acionar também esta mensagem, para que possa existir um maior controlo do operador sobre a rede. As tecnologias de camadas inferiores, poderão também possuir mecanismos de deteção de falhas, pelo que devem poder comunicar ao MPLS-TP este acontecimento. Um plano de controlo dinâmico, pode possuir funcionalidades que lhe permitam acionar mecanismos de proteção, e assim deve também ser uma das entradas do gerador de mensagens. Ao serem usados mecanismos que revertam a transmissão para o caminho de trabalho em redes instáveis, poderá acontecer que a transmissão esteja constantemente a "saltar"entre o caminho de trabalho e o caminho de proteção, pelo que a introdução do *timer* WTR (Wait To Restore), tem como finalidade impedir esta situação [5, pág 8-9].

A mensagem PSC, tal como referido anteriormente será uma mensagem de OAM transmitida através do G-ACh. O formato desta mensagem, a descrição extensiva de cada campo e a descrição dos estados possíveis que cada *router* pode assumir no protocolo PSC, podem ser encontradas em [5, pág 16-32].

### Proteção em anel

Muitos operadores mostraram um elevado interesse na operação do MPLS-TP em topologias em anel, já que esperam que este modo de operação seja bastante robusto contra falhas, tal como acontece na tecnologia SDH. No documento [40, pág 24-27] consta uma lista de requisitos baseados nas expectativas dos operadores, para o funcionamento do MPLS-TP nesta topologia. Deve ser possível implementar mecanismos de proteção linear à proteção em anel, contudo espera-se que os mecanismos de proteção em anel sejam um caso particular (otimizado) dos mecanismos de proteção linear da tecnologia [40, pág 24-27].

O requisito 96 do documento [38] especifica que o tempo de recuperação a falhas não deve exceder 50 ms para uma topologia em anel com 16 nós e com uma extensão inferior a 1200 Km de fibra. Em seguida ilustrar-se-à um exemplo do processo de proteção 1:1 em anel, com base nos exemplos do RFC 5317 [35] e nos mecanismos presentes no documento [5].

A figura (4.10) ilustra os LSPs de trabalho e proteção do mecanismo de proteção em anel

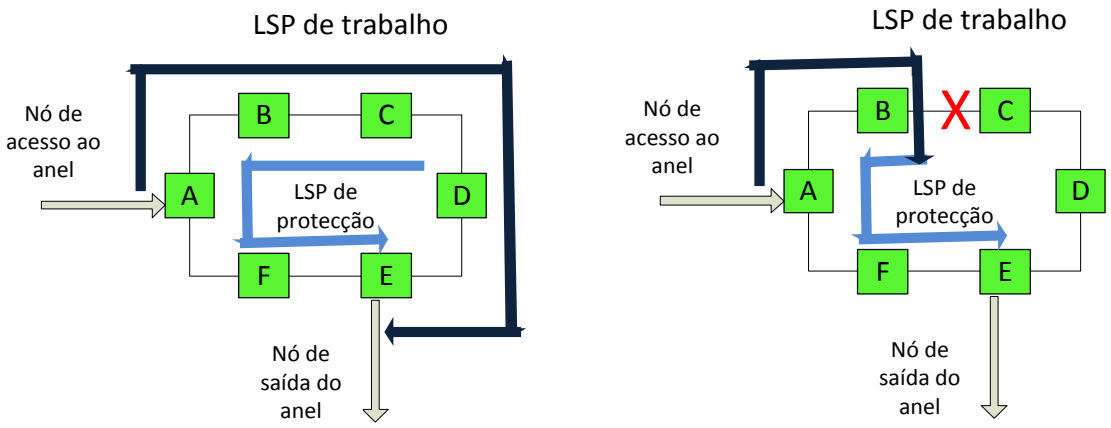



Figura 4.10: Mecanismo de proteção em anel 1:1 (1)

---

1:1. Neste exemplo pretende-se proteger um LSP unidirecional, e tal é realizado através de um único LSP de proteção. A escolha deste LSP deve minimizar o percurso dos pacotes após o restauro à falha e cobrir todas as ligações do LSP de trabalho. Por exemplo, se o LSP de proteção fosse CBAFED, após o corte ilustrado na imagem (4.10) o caminho de restauro necessitaria de passar pelo nó D, o que em (4.10) é evitado. A imagem (4.11) ilustra o anel usando o seu LSP de trabalho, e mostra as *labels* de ambos os LSPs.

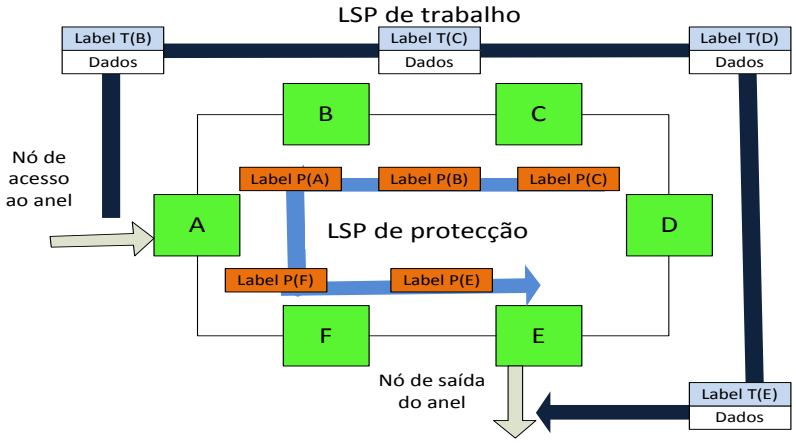



Figura 4.11: Mecanismo de proteção em anel 1:1 (2)

A monitorização de falhas no anel, funciona através do envio de mensagens de Connectivity Verification (CV), com uma periodicidade de 3.3ms. E é assumida uma falha numa secção do anel, após a ausência de três mensagens CV consecutivas (10ms) [35, pág 101]. É informado em [35] que quando é detetada uma falha, os nós adjacentes a esta alteram o seu modo de encaminhamento e enviam mensagens de OAM para que os *routers* vizinhos alterem também o seu modo de encaminhamento. Estes procedem á alteração e reencaminham a mensagem de OAM para os restantes *routers* sem a alterarem e assim consecutivamente [35, pág 102]. O

modo de operação descrito no documento [35] será uma resposta à falha baseada na mudança de estados dos *routers*, o que está também de acordo com o documento [5] para proteções lineares[4]. A mudança de estado de todos os *routers* do anel poderá contudo não ser necessária, como será discutido a seguir.

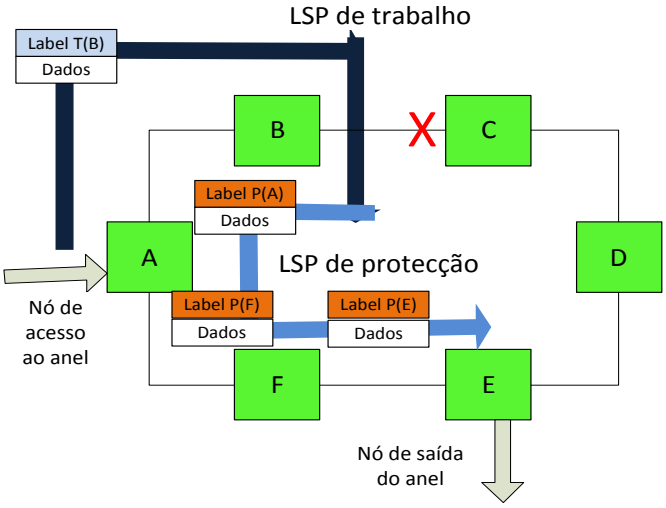



Figura 4.12: Mecanismo de proteção em anel 1:1 (3)

A alteração do estado no *router* B resumir-se-á a modificar a sua tabela LFIB de modo a que os pacotes que chegam com a *label* T(B), sofram a operação de *label swap* para a *label* P(A) (em vez de T(C)). No caso do *router* A possuir na sua tabela LFIB uma entrada para reencaminhar a *label* P(A) não necessitará de efetuar qualquer alteração do estado ou LFIB, já que reencaminhará estes pacotes do mesmo modo que reencaminha qualquer outro tipo de pacotes. Desta forma uma proteção em anel, poderia abdicar do protocolo PSC, em prejuízo do uso por parte dos *routers* de LFIBs mais extensas. Contudo a ausência de estados torna mais difícil identificar se está a ser usado o caminho de trabalho ou proteção. Neste cenário apenas os *routers* adjacentes à falha indicariam claramente que estão a usar o caminho de proteção.

Este mecanismo, tal como um mecanismo análogo que tenha por base o protocolo PSC e que altere o estado de todos os *routers*, poderão ser aplicados a ligações ponto-a-multiponto e serão capazes de recuperar uma rede que sofra múltiplas falhas, incluindo falhas na fibra ou no nó [35, pág 103-104].

O exemplo anterior é um LSP unidirecional, e o seu LSP de proteção foi otimizado. Contudo, quando de trata de um LSP bidireccional a escolha do LSP de proteção não é tão linear, já que se numa direção a escolha do LSP de proteção minimiza o percurso de restauro, na outra direção irá maximizá-lo. Deste modo, a escolha do LSP de proteção em ligações bidireccionais necessitará de ter em conta outras questões, tal como a geografia da rede, ou os serviços críticos que os LSPs transportam.

As proteções entre anéis são também um importante tópico para os operadores, já que o tráfego que atravessa vários anéis pode ser perdido se ocorrer uma falha na interseção

[4]No entanto nas proteções lineares apenas os LERs do LSP alteram o seu estado.

dos anéis. Com o objetivo de precaver esta situação e de aumentar de forma controlada a escalabilidade da rede o documento [40, pág 24-27] possuí alguns requisitos para a tecnologia MPLS-TP relativos à proteção das interseções de anéis.

### Proteções partilhadas ponto-multiponto

Num serviço IPTV o tipo de ligação mais apelativo será o ponto-multiponto. É portanto útil abordar de forma detalhada possíveis mecanismos de proteção em ligações ponto-multiponto. Para estas ligações o requisito 67 do RFC 5654 [40] apenas refere a proteção ponto-multiponto unidirecional.

A imagem (4.13) ilustra o mecanismo de proteção em anel estudado anteriormente, mas agora aplicado a ligações ponto-multiponto. Neste caso a proteção é 1:N ou seja, um caminho de proteção partilhado por várias ligações.

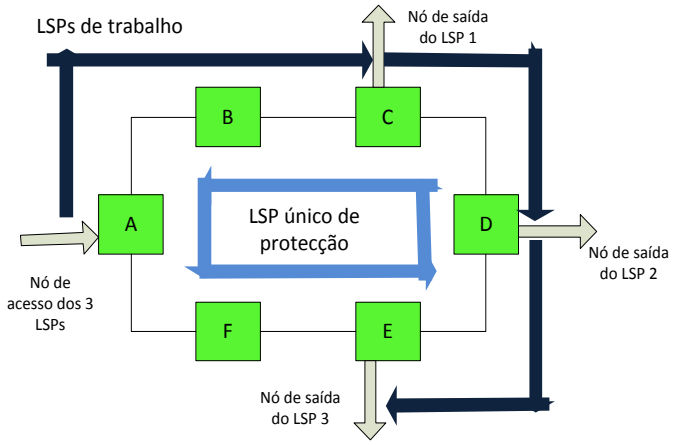



Figura 4.13: Mecanismo de proteção ponto-multiponto em anel (1:N)

Para poder proteger todas as ligações o LSP de proteção necessita de abranger todos os segmentos da rede. O funcionamento deste tipo de mecanismo é em tudo semelhante aos descrito anteriormente para ligações ponto-a-ponto [35, pág 100].

Quanto aos mecanismos de proteção lineares em ligações ponto-multiponto serão aqui apresentadas possíveis implementações, baseadas no documento [39] que por sua vez, considera já o algoritmo PSC definido no documento [5]. Uma arquitetura ponto-multiponto linear será pois, uma arquitetura em árvore pelo que será usada uma nomenclatura similar à descrição nos serviços Ethernet E-Tree, onde o nó fonte das mensagens é a raiz, e os nós destino, as folhas.

#### 1:N

Numa proteção do tipo 1:N, vários caminhos de trabalho são protegidos por um único caminho de proteção ponto-multiponto. O exemplo a seguir ilustrado é uma generalização da proteção 1:1 apresentada anteriormente, onde é implementado o protocolo PSC que selecionará o caminho de trabalho a proteger com base na sua prioridade. O mecanismo de proteção funcionará da seguinte maneira: os LERs folhas ao detetarem um defeito na receção de pacotes, enviam uma mensagem Remote Defect Indication (RDI) ao LER raiz responsável.

o nó raiz após tomar conhecimento de uma ou mais falhas nas folhas destino, gera um pedido de proteção que envia ao *router* responsável pelo caminho de proteção. A mensagem enviada constituirá uma entrada do algoritmo PSC do *router* de proteção. Com base nesta mensagem o *router* de proteção originará mensagens PSC. As mensagens geradas são mensagens G-ACH PSC, contendo um TLV adicional que indica qual o LSP ou LSPs ponto-multiponto que serão alvo de proteção. As mensagens PSC serão enviadas através do caminho de proteção aos vários LER folhas e os LER que identificarem aqueles LSPs como seus, passarão a aceitar pacotes provenientes do caminho de proteção.

A imagem (4.14) mostra uma arquitetura de proteção 1:N onde será aplicada esta solução[5]. É simulado que o caminho 1 sofre uma falha na ligação ao LER folha F1, e que o caminho 2 sofre uma falha na ligação ao LER folha F2.

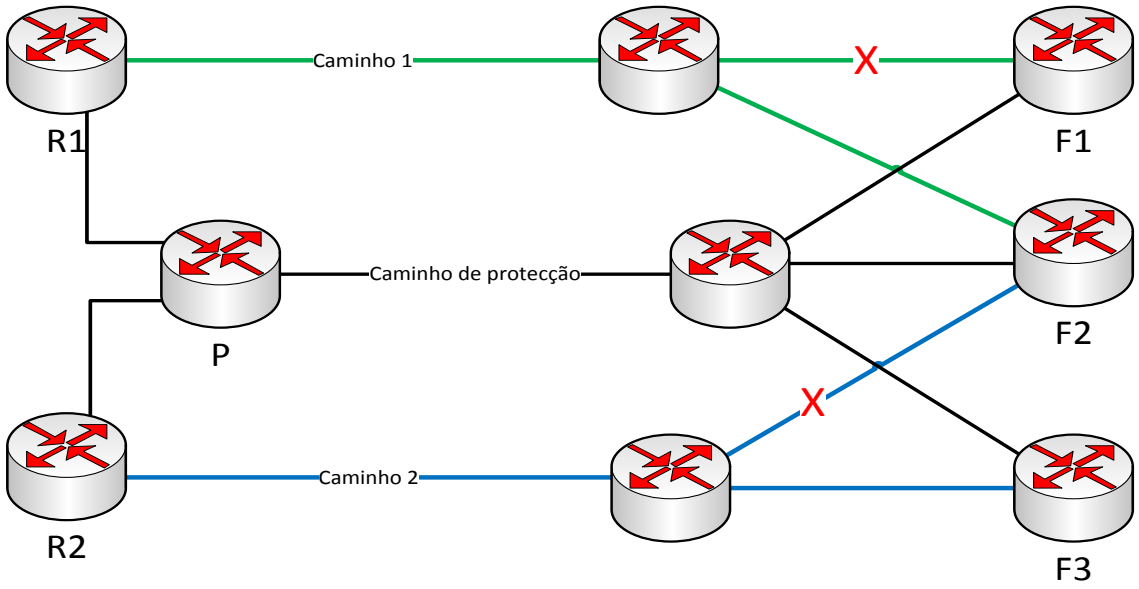



Figura 4.14: Mecanismo de proteção linear ponto-multiponto (1:N)

Assuma-se que o caminho 1 é mais prioritário que o caminho 2 e que as falhas ocorrem ao mesmo tempo. Após a falha ser detetada ambos os LER enviam mensagens RDI de forma independente, aos LER raiz R1 e R2 através do caminho de proteção. Após os pacotes RDI serem recebidas e processadas, ambos os *routers* raiz enviarão para o *router* de proteção P um pacote de controlo, anunciando a falha. O *router* P executará o protocolo PSC e visto o caminho 1 ser mais prioritário que o caminho 2, o *router* P atribuirá o caminho de proteção ao caminho 1. Em seguida o *router* P envia mensagens *PSC* para os *routers* folha informando que o caminho de proteção está a encaminhar o caminho 1, e envia mensagens de controlo aos *routers* raiz informando qual o caminho que está atualmente a ser protegido. Após isto R1 altera o envio de pacotes para o caminho de proteção e os *routers* folha F1 e F2 passam a aceitar estes pacotes. O *router* F3 descarta os pacotes provenientes do caminho de proteção, já que verifica nas mensagens PSC que o caminho de proteção está a transmitir pacotes de um LSP ou conjunto de LSPs a que não pertence. Quanto ao *router* raiz R2, este foi informado

[5]Caminho 1, caminho 2 e caminho de proteção designam três LSPs ponto-multiponto

pela mensagem de controlo proveniente do *router* P que o caminho de proteção está a ser usado para transmitir pacotes do caminho 1 pelo que, os pacotes do LSP 2 continuam a circular no caminho de trabalho 2. Nesta situação os pacotes do caminho 2, destinadas ao *router* L2 são perdidas.

Este mecanismo de proteção levanta algumas questões, nomeadamente sobre a real necessidade de todos as comunicações mencionados anteriormente para o restauro da falha. Note-se que para restaurar uma falha existe uma comunicação entre uma folha e a raiz, seguida de uma comunicação entre a raiz e o *router* de proteção e uma comunicação entre o *router* de proteção e as raízes e folhas daquele caminho. Ora, se a comunicação inicial poder ser encurtada de modo a que as mensagens de alerta de falha provenientes dos *routers* folha, possam imediatamente ativar o mecanismo PSC do *router* de proteção, isto resultaria em tempos de recuperação mais rápidos e em recuperações mais eficientes. Contudo a mensagem de alerta proveniente do *router* folha, deveria conter desde logo informação sobre o *router* raiz origem da comunicação com falha.

Outro aspeto que deve ser debatido é qual a escalabilidade da rede em número de *routers* raiz, e de que modo estes se encontrariam ligados ao *router* de proteção, já que no caso de se encontrarem diretamente ligados ao *router* de proteção o número de ligações possíveis estará limitado pelo número de portas do *router* de proteção. No caso de se encontrarem ligados ao *router* de proteção através de outros *routers*, os primeiros serão prejudicados ao ser necessário usar o caminho de proteção para proteger os seus caminhos, já que se encontram mais afastados do *router* de proteção.

$(1:1)^N$

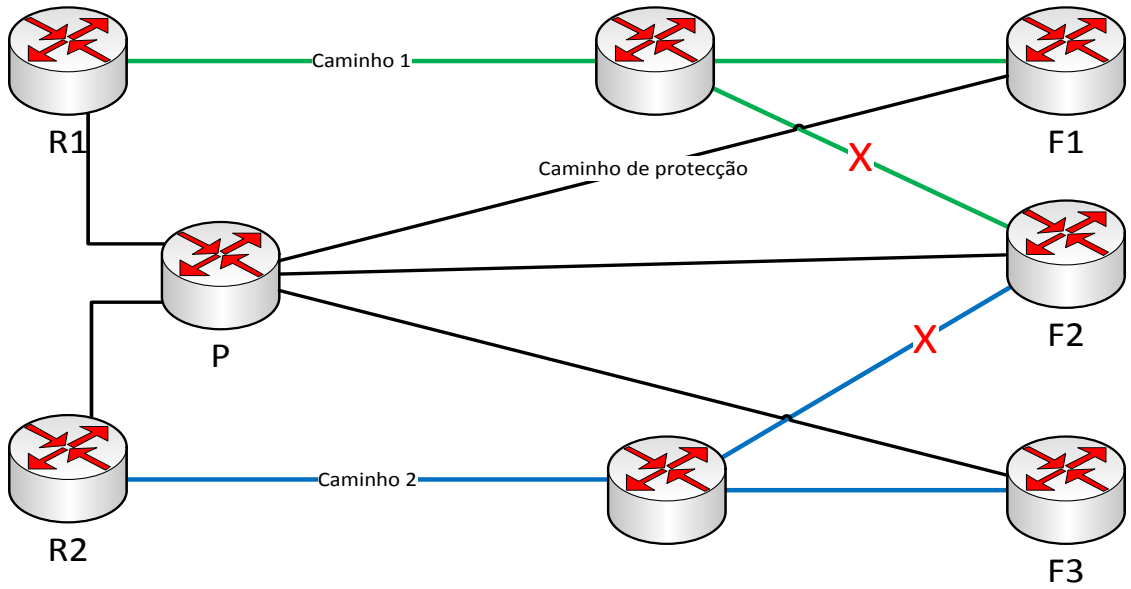



Figura 4.15: Mecanismo de proteção linear ponto-multiponto $(1:1)^N$

Outro exemplo abordado no documento [39] é a proteção $(1:1)^N$, onde cada LER folha é protegido por uma ligação ponto-a-ponto que o liga ao *router* de proteção P. Neste caso as

mensagens PSC serão imediatamente geradas pelo *router* folha após detetar uma falha. Esta mudança deve-se ao facto de existir um caminho de proteção para cada ligação, retirando a necessidade do *router* de proteção P escolher qual o caminho que deve proteger. Contudo o algoritmo PSC deve ser realizado pelos *routers* folha, já que o mesmo *router* folha pode ser o destino de vários LSPs.

A imagem (4.15) ilustra este tipo de arquitetura. Se acontecerem duas falhas simultâneas, nas ligações ao *router* F2, este executará o algoritmo PSC, gerando mensagens onde informa o *router* P que o caminho 1, que é o caminho de maior prioridade, deve ser protegido através do caminho de proteção. Após isto o *router* P passará a encaminhar os pacotes provenientes de R1, com destino a F2 através do caminho de proteção e F2 passará a aceitar estes pacotes. Quanto aos pacotes do caminho 2 com destino a F2, estes serão perdidos.

A proteção ponto-multiponto partilhada não se encontra ainda totalmente definida. Existem questões que não foram abordadas com a profundidade exigida, tal como qual a função dos caminhos de proteção quando não ocorrem falhas na rede, quais os tempos recomendados para as operações descritas de modo a que o tempo total de recuperação da rede não exceda os 50ms, qual o tipo de mensagem trocada entre os *routers* raiz e o *router* de proteção, qual o número máximo de *routers* e alcance da rede de modo a respeitar os 50ms de proteção. Outra questão que não foi abordada em detalhe, é o mecanismo de retorno ao caminho de trabalho após a sua falha ter sido corrigida, este mecanismo terá uma especial importância na proteção 1:N, devido à partilha de um único caminho de proteção.

## 4.3 MLPS-TP no IXIA

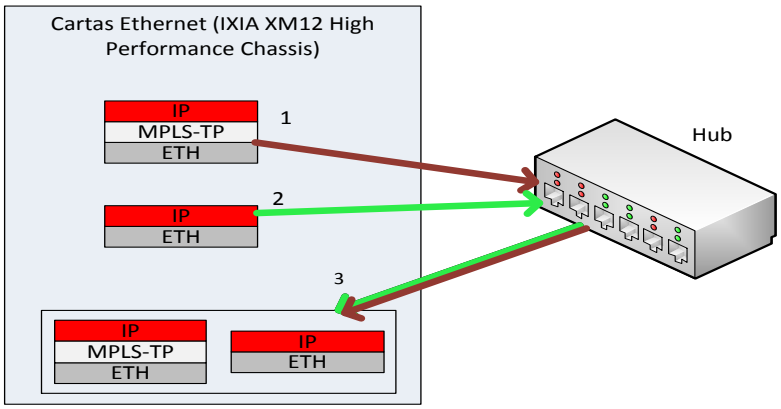



Figura 4.16: Teste com recurso ao equipamento IXIA XM12 High Performance Chassis

Nesta secção pretendem-se usar as cartas Ethernet do gerador de sinais IXIA XM12 High Performance Chassis, para gerar e analisar sinais MPLS-TP. No teste realizado, uma porta Ethernet foi atacada por dois fluxos de tramas Ethernet provenientes de duas portas distintas. Tal como ilustra a figura (4.16) ambos os fluxos incidem num *hub*, onde ficam sujeitos a perdas provenientes de colisões, e são depois reenviados para o gerador IXIA através da mesma porta do *hub*. A utilização do *hub* acarretou a truncagem das taxas máximas de transmissão a 10

Mbps.

Um dos fluxos consiste em tramas Ethernet que contêm um cabeçalho MPLS-TP, que por sua vez contém um pacote IP, o outro fluxo consiste em tramas Ethernet encapsulando diretamente um pacotes IP, tal como ilustra a figura (4.17).

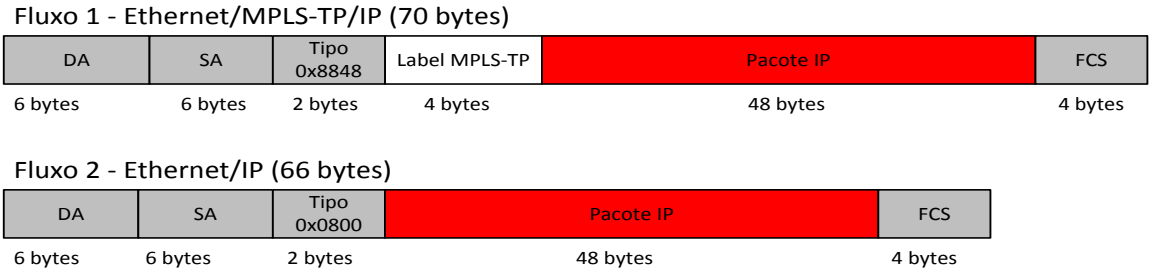



Figura 4.17: Consituição das tramas usadas no teste.

Para a mesma capacidade do pacote IP (48 bytes), são necessários 4 bytes extra no fluxo com tecnologia MPLS-TP correspondentes ao uso de uma *label*. O tamanho total das tramas no fluxo com MPLS-TP são 70 bytes, enquanto o tamanho das tramas no fluxo sem recurso a esta tecnologia são 66 bytes.

### Resultados

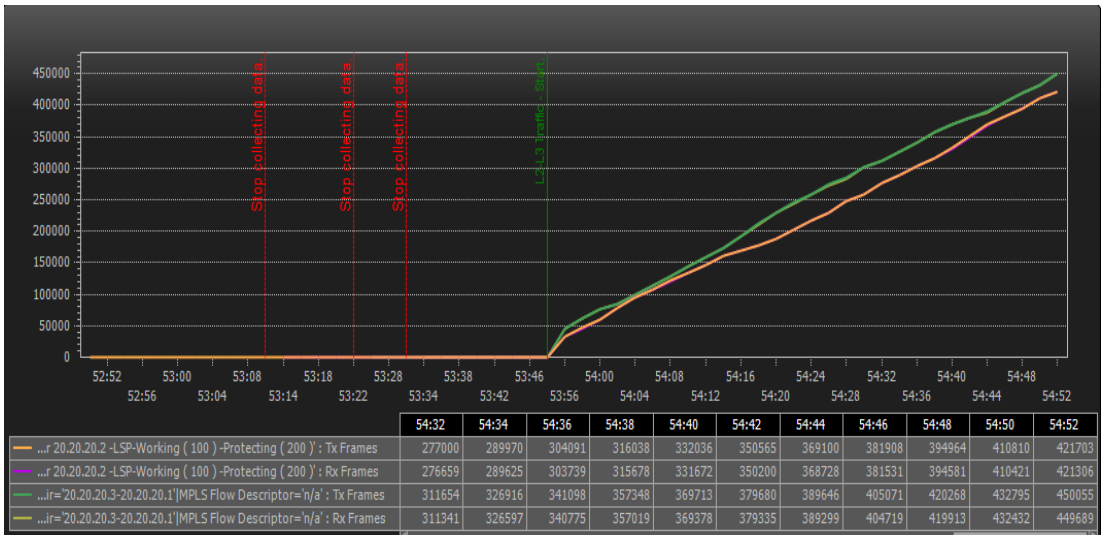



Figura 4.18: Número de tramas enviadas e recebidas por fluxo.

A figura (4.18) ilustra o número de tramas geradas e recebidas em cada fluxo, com períodos de amostragem de dois segundos. A baixa taxa de tramas perdidas torna pouco percetível a distinção entre as tramas enviadas e recebidas de cada fluxo, já que se encontram aproximadamente sobrepostas. O inicio das transmissões iniciou-se sensivelmente aos 53 minutos e 51

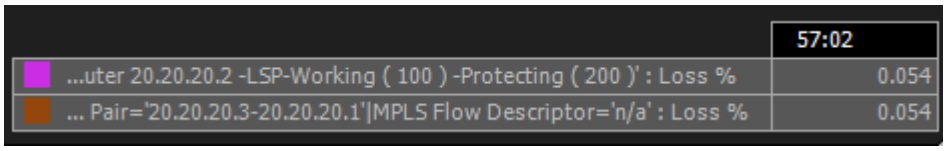



Figura 4.19: Percentagem de tramas perdidas por fluxo.

---

segundos.

Observa-se em (4.18) que o número de pacotes enviados e recebidos dos fluxos tiveram comportamentos semelhantes, contudo o facto dos fluxos possuírem a mesma taxa de transmissão leva a que sejam geradas e recebidas mais tramas no fluxo Ethernet/IP. A relação entre esta diferença é, como seria de esperar, igual à relação entre o tamanho das tramas dos dois fluxos, ou seja,

$$\frac{Tramas\ \ fluxo\ 1}{Tramas\ \ fluxo\ 2} \simeq \frac{tamanho\ Trama\ 2}{tamanho\ Trama\ 1}. \tag{4.1}$$

A taxa de transmissão de ambos os fluxos é igual a,

$$\frac{(Tramas\ \ fluxo\ 1)(tamanho\ Trama\ 1)}{tempo} = \frac{(tramas\ \ fluxo\ 2)(tamanhoTrama2)}{tempo}. \tag{4.2}$$

A imagem (4.19) mostra que a percentagem de perdas dos dois fluxos é igual. Ambos os fluxos possuem a mesma taxa de transmissão, gerando e perdendo tramas à mesma taxa, e estando ambos sujeitos às colisões que ocorrem no *hub*. Este teste ilustrou a forma como a adição de cabeçalhos MPLS-TP, influencia a quantidade de informação útil transportada nas redes MPLS-TP, para a uma dada taxa de transmissão. Note-se que neste exemplo apenas se recorreu ao uso de uma *label*, contudo tal como se viu, os pacotes MPLS-TP não têm limite para o número de *labels* transportadas.

# Capítulo 5

# Cenário de distribuição IPTV

Neste capítulo será modelado um sistema de distribuição IPTV, tendo por base a tecnologia MPLS-TP, um serviço de distribuição EVP-Tree e tendo como referência o modelo matemático apresentado em [11] [12] [13].

Dos serviços de Carrier Ethernet estudados anteriormente, aquele que potencialmente maior benefício trará a uma implementação IPTV é o E-Tree, nomeadamente o EVP-Tree, sendo sobre ele que recairá o cenário de distribuição IPTV. A escolha, prende-se com a arquitetura de distribuição que um serviço de IPTV necessita, onde um servidor raiz deve possuir ligações físicas ponto-multiponto a vários clientes, para desta forma poder optar por enviar os seus conteúdos através de ligações ponto-a-ponto, ou seja numa transmissão *unicast* ou por outro lado, enviar simultaneamente o mesmo conteúdo a vários clientes, através de ligações ponto-multiponto numa transmissão *multicast*, sem necessitar de alterar a topologia física da rede.

A escolha da tecnologia MPLS-TP prende-se com o elevado interesse demonstrado pelos operadores de telecomunicações no emprego desta tecnologia.

Após ser obtida uma solução para a arquitetura de rede e para a tecnologia usada, será desenvolvido um modelo matemático tendo por base o trabalho [11] [12] [13], que visa analisar qual a solução de distribuição que usa de modo mais eficiente os recursos da rede. As soluções possíveis para esta distribuição serão: o envio de todos os canais televisivos em *unicast*, a distribuição de todos os canais em *multicast*, ou uma solução mista onde os canais mais vistos são enviados em *multicast* e os menos vistos em *unicast*.

## 5.1 EVP-Tree sobre MPLS-TP

O serviço EVP-Tree descrito na secção 3.3.3, é um caso particular do E-Tree, onde uma UNI raíz se encontra ligada a várias UNIs folhas e é permitida a comunicação apenas entre UNIs raízes e UNIs folhas, devendo ser impossível qualquer comunicação entre UNIs folhas. O serviço EVP-Tree definido no documento MEF 6.1 é realizado através de ligações EVC orientadas à *tag* VLAN, permitindo assim numa arquitetura física com topologia em árvore, efetuar ligações ponto-multiponto ou ponto-a-ponto virtuais, e criar na mesma rede física várias redes virtuais em árvore.

A possibilidade de realizar LSPs unidirecionais ponto-multiponto em MPLS-TP, possibilita à tecnologia implementar de uma forma relativamente simples o serviço Ethernet EVP-Tree. A imagem (5.1) mostra como a tecnologia MPLS-TP suportaria um serviço

EVP-Tree.

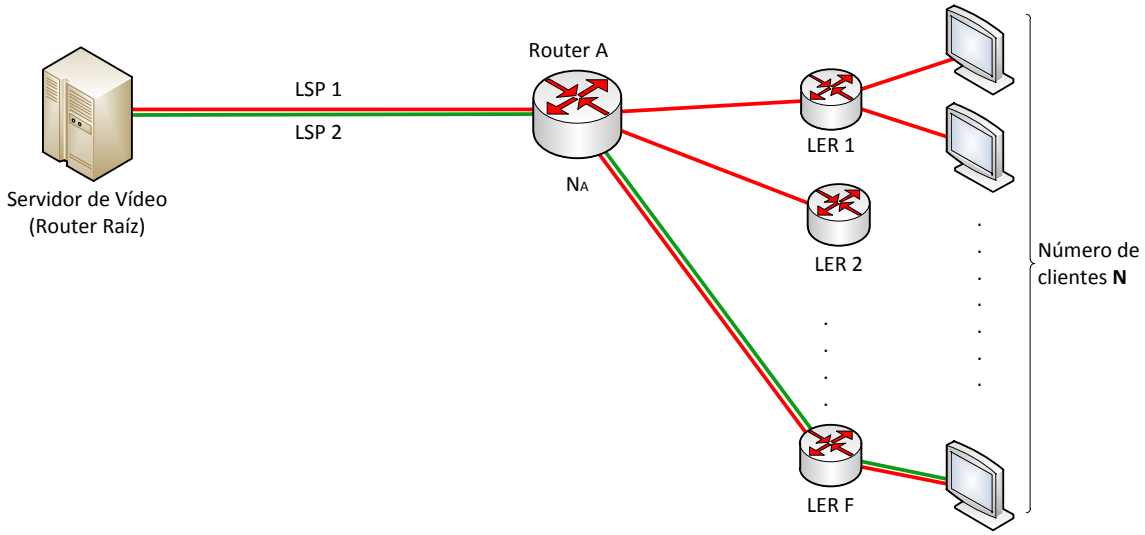



Figura 5.1: Serviço EVP-Tree sobre MPLS-TP. O servidor de vídeo é a raiz do sistema e está ligado ao *router* interno de distribuição A. O *router* A tem como função distribuir os sinais provenientes do servidor de vídeo para os LERs que estão ligados à rede de acesso e servem os N clientes.

Na arquitetura da figura (5.1) está representado um servidor de vídeo que assume o papel de *raíz*, ligado a um *router* A, que tem como função distribuir o sinal proveniente do servidor de vídeo aos *routers* folha da rede MPLS-TP, ou seja aos LER. Os LER encontram-se na fronteira com a rede de acesso, e podem servir vários clientes. Na figura, é ilustrada uma ligação ponto-multiponto, através do LSP 1, e uma ligação ponto-a-ponto através do LSP 2. A arquitetura (5.1) não apresenta os caminhos de proteção que atribuem à rede mecanismos de sobrevivência, contudo poder-se-á assumir que a rede possuí uma proteção linear partilhada, ponto-multiponto 1:N, semelhante à descrita na secção (4.2.5).

Os RFCs [35] [40] que definem os requisitos da tecnologia MPLS-TP não referem LSPs ponto-multiponto bidirecionais, pelo que a comunicação de retorno entre os clientes e o servidor IPTV deverá ser estabelecida através de LSPs ponto-a-ponto unidirecionais. Estes LSPs poderão usar o caminho de trabalho ilustrado na figura (5.1) ou o caminho de proteção omitido. A comunicação dos clientes com o servidor, apesar de necessitar de uma largura de banda bastante inferior à necessária na comunicação entre o servidor IPTV e os clientes, não poderá ser desprezada, já que só assim um cliente de IPTV poderá solicitar serviços ao seu operador, serviços esses que podem ir desde uma mudança do canal de televisão, até ao aluguer de um vídeo, ou outra qualquer interatividade que o operador possibilite ao cliente.

## 5.2 Modelo de custos

No modelo de custos apresentado em [11] é definido o parâmetro $\beta$ como a razão entre o custo de transmitir um canal em *multicast* e o custo de transmitir o mesmo canal em *unicast*.

Ou seja,

$$\beta = \frac{C_m}{C_u}, \tag{5.1}$$

- $C_m$ representa o custo de transmitir um sinal em *multicast*;
- $C_u$ representa o custo de transmitir um sinal em *unicast*.

No presente documento a razão de custos $\beta$ será aproximada através da análise de custos do *router* A. O *router* de distribuição A constitui o nó central da arquitetura de distribuição de canais, sendo por isso o *router* que sofrerá uma maior carga e poderá condicionar de forma crítica o desempenho da rede.

No *router* A o custo de processar um pacote *unicast* será duas vezes o custo de processar o pacote ($P_p$) (já que é necessário processar este à entrada e à saída do *router*), mais o custo de pesquisa na tabela de encaminhamento LFIB ($P_s$), tal como ilustra a figura (5.2).

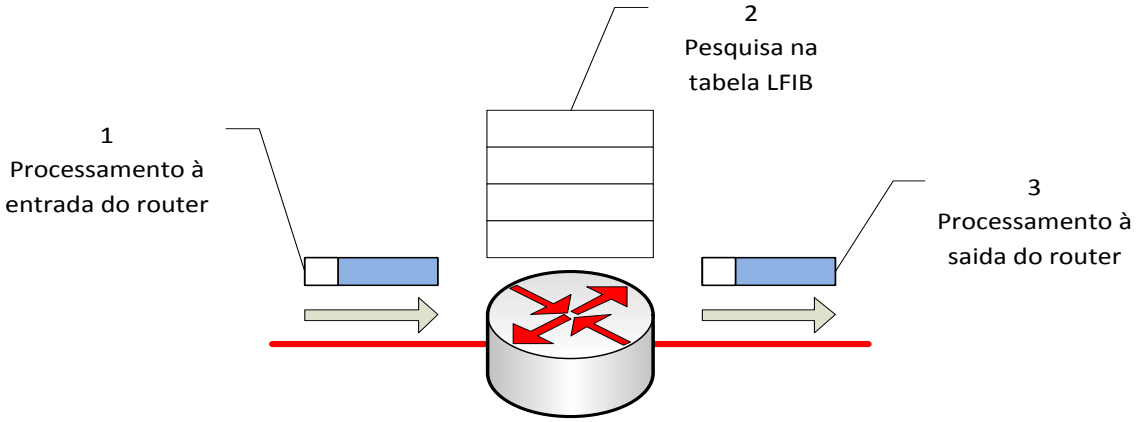



Figura 5.2: Custo de transmissão em *unicast*

Assim sendo $C_u$ será definido como,

$$C_u = 2P_p + P_s, \tag{5.2}$$

- $P_p$ representa o custo de processamento de um pacote;
- $P_s$ representa o custo de pesquisa na tabela de encaminhamento (LFIB) do *router* A.

Se definirmos $\gamma$ como a relação entre o custo de uma pesquisa na tabela de encaminhamento, sobre o custo de processamento de um pacote,

$$\gamma = \frac{P_s}{P_p}. \tag{5.3}$$

Podemos escrever $C_u$ como

$$C_u = P_p(2 + \gamma). \tag{5.4}$$

O custo de transmissão de um canal em *multicast*, tal como ilustra a figura (5.3), será o custo de processar um pacote multiplicado pelo número de portas ativas do *router*, mais o custo de uma pesquisa na tabela de encaminhamento (LFIB), ou seja,

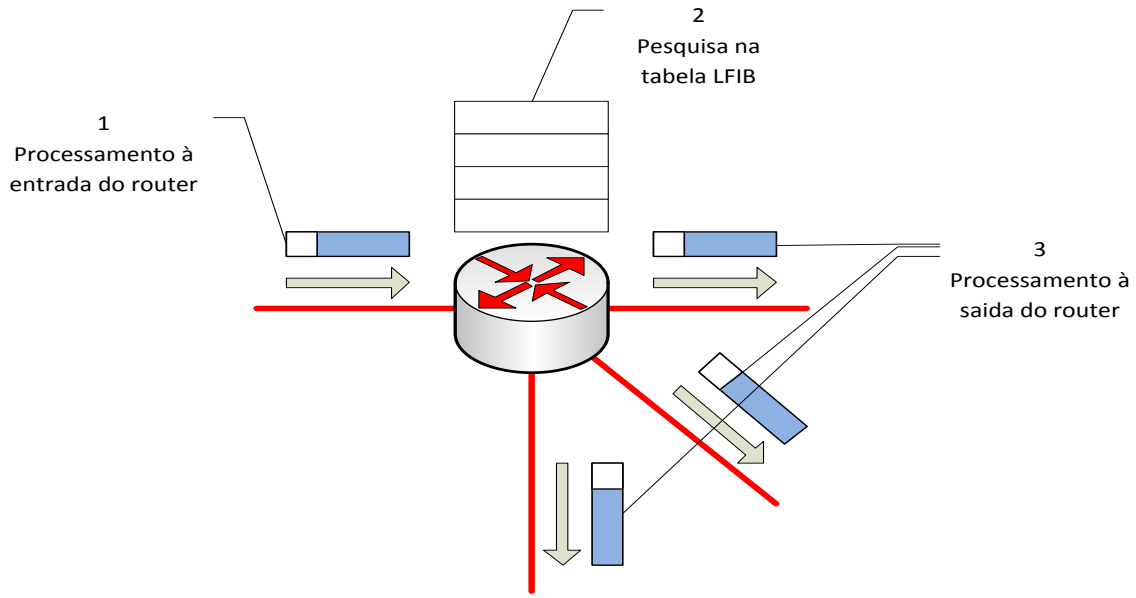



Figura 5.3: Custo de transmissão em *multicast*

$$C_m = N_A P_p + P_s \tag{5.5}$$

Com recurso a (5.3) podemos escrever a equação anterior como

$$C_m = P_p(N_A + \gamma) \tag{5.6}$$

A variável $N_A$ representa o número de portas usadas pelo *router* A. Após estas definições, é agora possível definir $\beta$ em função dos custos de processamento do *router* A;

$$\beta = \frac{N_A + \gamma}{2 + \gamma}. \tag{5.7}$$

A abordagem realizada permite-nos obter um $\beta$ em função do número de portas ativas que o *router* possuí ($N_A$) e de $\gamma$. Assumindo que o custo de processamento de um pacote ($P_p$) é constante, o parâmetro $\gamma$ será função do custo de pesquisa na tabela de encaminhamento LFIB $P_s$, ou seja do tamanho desta tabela de encaminhamento. Desta forma os valores que definem $\beta$ poderão ser relacionados com os valores reais de memoria e processamento de um *router*. Notemos que a aproximação do valor de $\beta$ poderia assumir outros contornos, o que acontece por exemplo no documento [41], onde $\beta$ é analisado em função dos custos das ligações da rede.

## 5.3 Distribuição da popularidade dos canais

Assuma-se que o servidor de IPTV da figura (5.1) pertence a um operador de telecomunicações que oferece aos clientes do seu serviço de IPTV um pacote de $K$ canais fixos e predefinidos. Consoante a popularidade destes canais, é natural que alguns sejam mais requisitados do que outros. Será assim importante para um operador determinar quais os canais mais requisitados e qual a relação entre a popularidade dos vários canais, ou seja qual a distribuição da popularidade dos canais oferecidos. Se os canais forem ordenados desde o mais popular ($k = 1$) até o menos popular ($k = K$), pode assumir-se que a sua distribuição seguirá a lei de Zipf [11]. Assim a popularidade de um canal $k$ será modelada por,

$$\pi_k = \frac{d}{k^\alpha} \tag{5.8}$$

para k= 1,2,...,K. Desta forma:

- $k$ representa o índice de popularidade do canal. Tal como já foi referido $k = 1$ representa o canal mais popular, e $k = K$ o menos popular.

- $\alpha$ é a constante de Zipf. O documento [11] apresenta uma amostragem da popularidade de canais de um operador de IPTV realizada durante 1 mês e meio. Após a análise de resultados é mostrado que o valor deste parâmetro variou entre 0.5 a 0.8.

- $d$ é a constante de normalização. Sabendo que o somatório de todas as probabilidades $\pi_k$ é 1, ($\sum_{k=1}^{K} \pi_k = 1$) a constante $d$ será dada por:

$$d = \frac{1}{\sum_{k=1}^{K} \frac{1}{k^\alpha}} \tag{5.9}$$

A figura (5.4) mostra uma distribuição de Zipf que modela a popularidade de 100 canais. A popularidade destes vem expressa no seu valor de probabilidade $\pi_k$. Para demonstrar a influencia da constante de Zipf usou-se $\alpha = 0.4$, $\alpha = 0.6$ e $\alpha = 0.9$. Como se verifica à medida que $\alpha$ aumenta, os canais mais populares aumentam a sua popularidade em detrimento da popularidade dos canais menos populares.

Para além da modelação da popularidade dos canais, é também necessário modelar a atividade dos clientes. Desta forma assume-se que o comportamento dos clientes é independente entre si e que a probabilidade de um cliente estar ativo é um processo de Bernoulli. Assim um cliente poder-se-á encontrar em dois estados, no estado ativo e inativo. A probabilidade de um cliente se encontrar num estado ativo é $a$, e a probabilidade de se encontrar inativo é $1 - a = \pi_{u,0}$, o índice $k = 0$ representará um cliente inativo. Desta forma, cada utilizador terá associado um conjunto de probabilidades $\pi_{u,k} = a_u \pi_k$. Se $\Pi$, definir um vetor de dimensão K, que contém o conjunto de probabilidades $\pi_k$,

$$\Pi = [\pi_k]. \tag{5.10}$$

O comportamento de cada utilizador poderá ser modelado por $\Pi_u$, com dimensão K+1, tal como mostra a equação (5.11),

$$\Pi_u = [1 - a_u | a_u \Pi]. \tag{5.11}$$

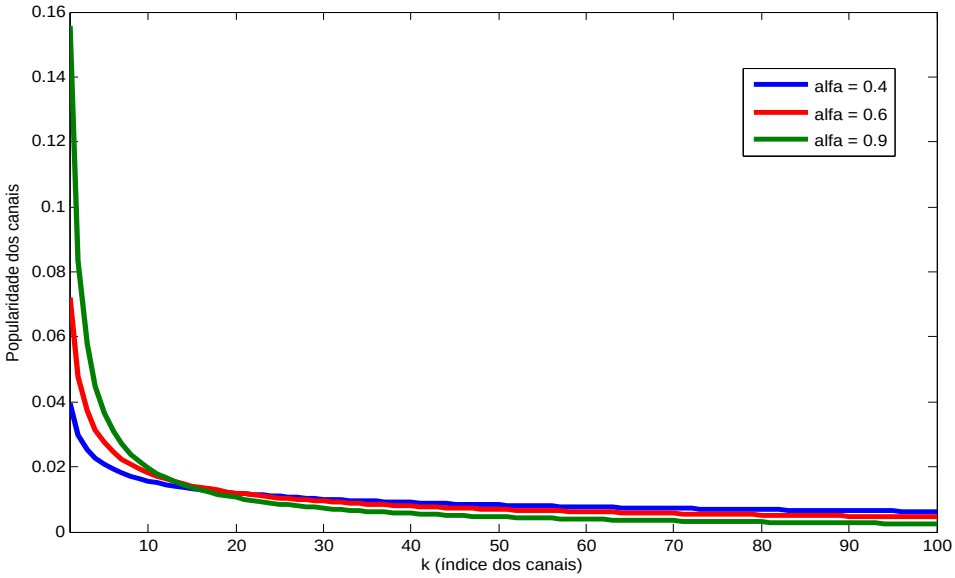



Figura 5.4: Influência do parametro $\alpha$ numa distribuição Zipf

---

O vetor $\Pi_u$ contempla também a possibilidade de um cliente se encontrar inativo, não solicitando qualquer canal, ou seja, $\pi_{u,0} = 1-a$. No caso do cliente estar ativo, a probabilidade de solicitar a transmissão de um canal k, será $\pi_{u,k} = a_u \pi_k$ tal como referido.

Embora o parâmetro $a$ possa variar ao longo de dias, semanas, ou estações, neste documento o parâmetro será tratado como uma constante.

## 5.4 Unicast e Multicast

O servidor de IPTV da figura (5.1) de forma a usar com maior eficiência os recursos de que dispõe, pode optar por diferentes estratégias, no que diz respeito à forma como transmite os seus $K$ canais. O operador pode optar por transmitir todos os seus canais em *multicast*, em *unicast*, ou optar por uma estratégia onde combine os dois tipos de transmissão.

Para determinar a melhor forma de combinar as transmissões de canais, é necessário definir alguns parâmetros do sistema:

- N, representa o número de clientes.
- K, o número de canais que o operador oferece.
- C, um vetor de dimensão N, cujos valores $c_k$ representam o número de clientes sintonizados no canal k.
- $n_a$, é o número de total de clientes ativos. Visto o comportamento dos clientes ser independente, $n_a$ é uma variável aleatória com distribuição binomial.

$$n_a = \sum_{k=1}^{K} c_k \tag{5.12}$$

- $n_i$, é o número de utilizadores inativos, dado por $n_i = N - n_a$.

O custo de transmitir todos os canais em *multicast* será o número de canais K que o operador oferece, multiplicado pelo custo desta transmissão relativa à transmissão em *unicast*, ou seja,

$$r_m = \beta K. \tag{5.13}$$

O custo de transmitir todos os canais um *unicast* poderá ser aproximado pelo valor esperado da variável aleatória $n_a$, ou seja,

$$r_u = Na. \tag{5.14}$$

O gráfico (5.5) ilustra a variação destes dois custos à medida que o número de clientes N aumenta, para um fator de atividade $a = 0.5$, com K=120 canais e $\beta = 2$.

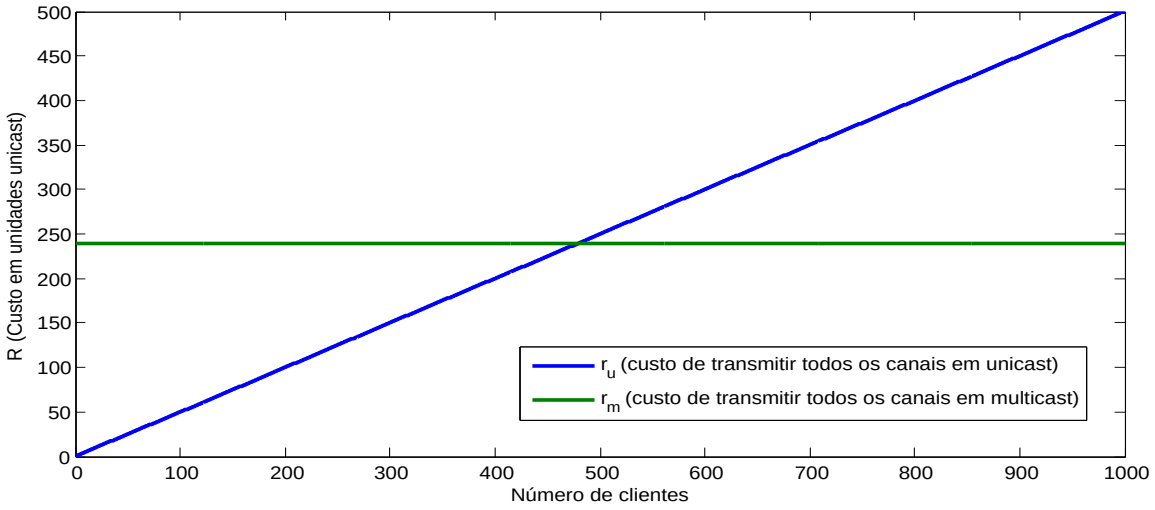



Figura 5.5: Custo de transmissão dos K canais em *unicast* e *multicast*

Como seria de esperar à medida que o número de clientes de um sistema aumenta, torna-se mais eficiente transmitir todos os canais em *multicast*, já que a transmissão em *unicast* aumenta linearmente em função de N. No entanto, o facto dos canais possuírem diferentes níveis de popularidade perante os clientes, aliado à diferença de custos de transmissão de canais em *multicast* e *unicast*, pode possibilitar uma minimização de recursos ao ser usada uma estratégia de transmissão mista, *unicast* e *multicast*.

Com o objetivo de otimizar ao máximo o uso dos recursos que possui, um sistema IPTV, deve saber exatamente quais os canais que deve transmitir em *unicast* e quais os canais que deve transmitir em *multicast* em cada situação. Naturalmente, deverá transmitir os canais mais populares em *multicast* e os menos populares em *unicast*, então dados K canais, como deve ser realizada a escolha?

Uma possível solução, será identificar previamente os M canais mais populares e envia-los por *multicast*, sendo os restantes canais transmitidos em *unicast*, se solicitados pelos clientes. Nesta solução será de extrema importância o conhecimento do número de canais a transmitir em *multicast* que maximiza a eficiência da rede.

## 5.5 Aplicação do modelo matemático

### 5.5.1 Abordagem exata

Com o propósito de determinar quantos clientes se encontram sintonizados em cada canal, é definida a probabilidade $w_B = P[c_1 = b_1, c_2 = b_2, ..., c_k = b_k]$. $b_k$ representa o número de utilizadores sintonizados no canal $k$ e $B$ será o conjunto de $b_k$ com dimensão $K$.

A probabilidade $w_B$ pode ser representada como a interseção entre a probabilidade do sistema possuir $n$ clientes ativos, ou seja $P[n_a = n]$, com a probabilidade de se encontrarem $b_K$ clientes sintonizados em $c_k$ canais, ou seja $P[c_1 = b_1, c_2 = b_2, ..., c_k = b_k]$. Visto $P(A \cap B) = P(B)P(A|B)$, $w_B$ poderá ser calculado da seguinte forma:

$$w_B = P[c_1 = b_1, c_2 = b_2, ..., c_k = b_k], \tag{5.15}$$

$$= P[n_a = n]P[c_1 = b_1, c_2 = b_2, ..., c_k = b_k | n_a = n], \tag{5.16}$$

$$= \frac{N!}{(N-n)!}(1-a)^{N-n}a^n \Pi_{k=1}^{K} \frac{(\pi_k)^{b_k}}{b_k!}. \tag{5.17}$$

Note-se que o número de possíveis vetores $B$ será a o número de possíveis conjuntos de tamanho N (clientes) dentro de um grupo de K+1(canais), logo, o número possível de combinações para B será,

$$C(K+N, N) = \frac{(N+K)!}{N!K!}. \tag{5.18}$$

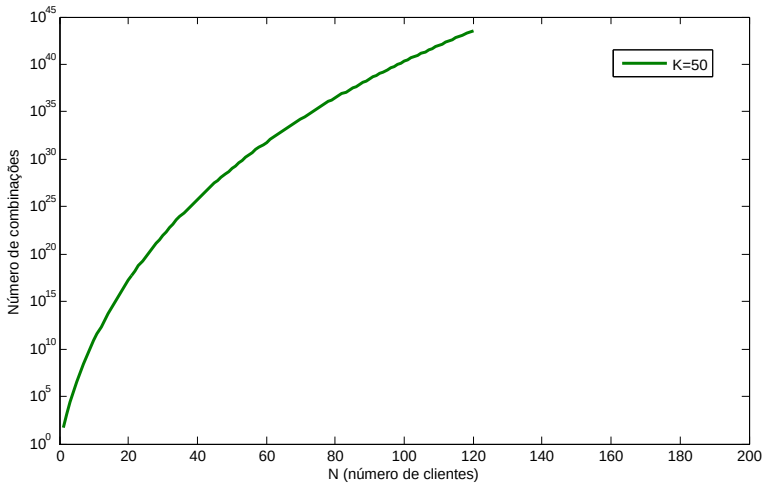



Figura 5.6: Número de combinações possíveis entre N e K

O gráfico (5.6) representa a função(5.18). Considerando um cenário onde um operador oferece um pacote de 50 canais, observa-se que o número de possíveis combinações excede as

capacidades do simulador (Matlab R2009b©) quando N atinge 120 clientes[1].

Num cenário onde o número de canais a transmitir em *multicast* é escolhido pelo operador, a variável aleatória $n_m$ que representa o número de canais transmitidos em *multicast* é definida pelo parâmetro determinístico M. Para determinar o número de canais transmitidos em *unicast*, ou seja $n_u$, é necessário determinar a probabilidade de ter i canais em *unicast* sabendo que o sistema possui n utilizadores ativos, ou seja determinar $P[n_u = i|n_a = n]$. $n_u$ não é simplesmente $K - M$, já que poderão existir canais cuja popularidade leva a que não sejam solicitados por nenhum cliente, $n_u$ é uma variável aleatória, da qual se pretende obter a distribuição.

A probabilidade de um cliente querer ver um canal transmitido em *multicast* é dada por:

$$P_m = \sum_{k=1}^{M} \pi_k \tag{5.19}$$

Logo a probabilidade de um cliente querer assistir a um canal transmitido em *unicast*, será $P_u = 1 - P_m$. Desta forma a probabilidade de existirem i canais a serem requisitados em *unicast*, é:

$$P[n_u = i] = \sum_{n=0}^{N} P[(n_a = n) \cap (n_u = i)] \tag{5.20}$$

$$= \sum_{n=0}^{N} P[(n_a = n)] P[n_u = i|(n_a = n)] \tag{5.21}$$

$$= \sum_{n=0}^{N} \frac{N!}{(N-n)!} (1-a)^{N-n} a^n \frac{(P_u)^i (P_m)^{n-i}}{i!(n-i)!} \tag{5.22}$$

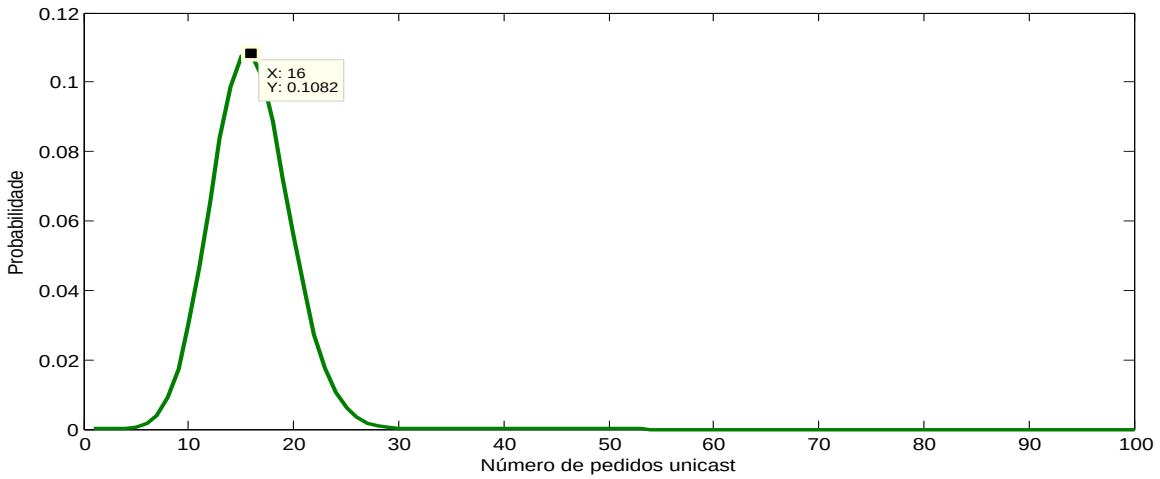



Figura 5.7: Função de distribuição de $n_u$ na abordagem exata

[1]Matlab©é um software desenhado para o cálculo numérico. Na sua versão R2009b a sua capacidade numérica é excedida para valores superiores a 170!, o que acontece em (5.6) quando N=120 clientes.

Realizou-se uma simulação da distribuição da variável $n_u$, considerando os seguintes parâmetros:

- $\alpha = 0.7$, tal como mostrado em [11], $\alpha$ deve ser um valor entre 0.5-0.8;
- $a = 0.5$, este valor é ilustrativo, sendo apenas uma estimativa da probabilidade média de um cliente se encontrar ativo, durante um dia;
- $N = 100$, assume-se que o sistema serve 100 clientes. O número de clientes nesta abordagem é extremamente limitada. Valores de N superiores a 170 levam a $N!$ superiores à capacidade do simulador.
- $K = 50$, assume-se que o operador disponibiliza um pacote de 50 canais;
- $M = 20$, serão transmitidos em *multicast* 20 canais.

O gráfico (5.7) ilustra a probabilidade de pedidos de canais *unicast* num cenário de 100 clientes, onde o operador disponibiliza um pacote de 50 canais e transmite 20 destes em *multicast*. Como se observa existirá cerca de 10.82% de probabilidade de serem solicitadas 16 transmissões que não se encontram no conjunto dos M canais *multicast*.

A manipulação e uso das equações aqui apresentadas, levam a algumas limitações dos parâmetros usados e elevados tempos de simulação. Este facto deve-se sobretudo ao elevado número de combinações a considerar entre $K$ canais e $N$ clientes, tal como mostra a equação (5.18) e o gráfico (5.6). Em [11] são mencionados alguns recursos de programação que podem levar à diminuição destes tempos, contudo ainda assim não são possíveis soluções em tempo útil, para elevados valores de $K$ e $N$. O documento [11] apresenta uma proposta para contornar este problema, através da aproximação de $n_u$ por uma distribuição conhecida.

### 5.5.2 Abordagem aproximada

Na abordagem aproximada assume-se que $n_u$ respeita uma distribuição normal, enquanto que a variável $n_m$ é uma variável determinística, com valor igual a M, tal como definido anteriormente. Esta aproximação de $n_u$ mostra-se bastante próxima do valor exato da variável, nomeadamente para elevados valores de N [11].

Uma distribuição gaussiana é completamente caraterizada pelo valor da sua média ($\mu$) e desvio padrão ($\sigma$), tal como mostra a equação (5.23) [42, pág 104-107],

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{5.23}$$

A média de $n_u$ será,

$$E[n_u] = E_{n_a}[E[n_u|n_a]] = E_{n_a}[n_a P_u] = NaP_u, \tag{5.24}$$

$$\mu = NaP_u. \tag{5.25}$$

Quanto ao desvio padrão, este poderá ser calculado da seguinte forma,

$$E[(n_u)^2] = E_{n_a}[E[(n_u)^2|n_a]] = E_{n_a}[n_a P_u(1 - P_u) + n_a^2 P_u^2], \tag{5.26}$$

$$E[(n_u)^2] = NaP_u(P_m + (1-a)P_u) + (NaP_u)^2, \tag{5.27}$$

Como $E[(n_u)^2] = \sigma^2 + \mu^2$, e $Var(n_u) = \sigma^2$, logo,

$$\sigma = \sqrt{NaP_u(P_m + (1-a)P_u)}. \tag{5.28}$$

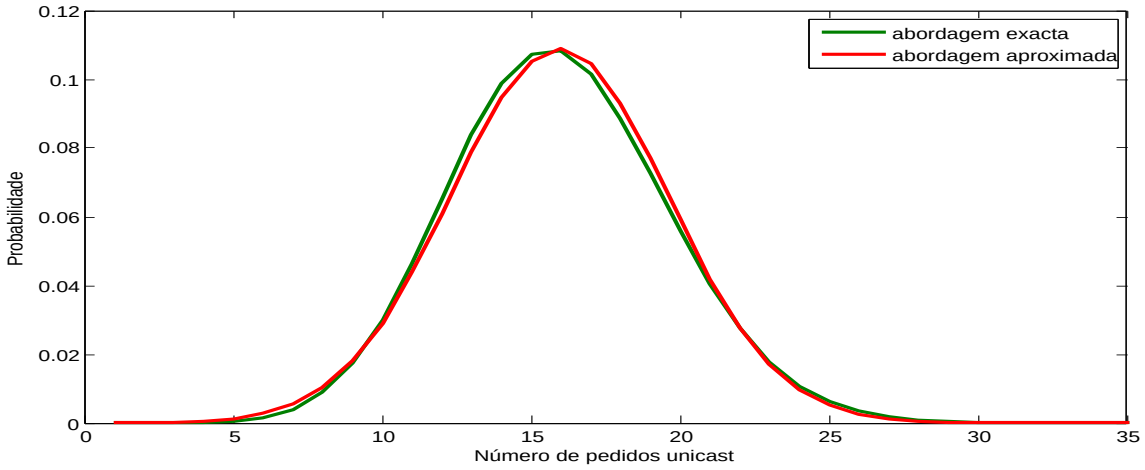



Figura 5.8: Comparação do valor exato e aproximado da variável aleatória $n_u$

A variável $n_u$ encontra-se agora totalmente caracterizada através da aproximação normal, e desta forma poderá ser comparada com o seu valor exato. Realizando uma simulação com recurso aos parâmetros usados em (5.7) é possível comparar as duas abordagens.

O gráfico (5.8) verifica a proximidade entre as duas abordagens. Para obter uma melhor perceção desta diferença foram omitidas as probabilidades de ocorrerem mais de 35 pedidos. A título de curiosidade, a probabilidade de existirem 16 pedidos em *unicast* para a abordagem aproximada, será de 10.9%.

O valor do custo $r_s$ num cenário de distribuição combinada de canais em *unicast* e *multicast*, será obtido em unidades de custo de transmissão *unicast* e por definição será,

$$r_s = n_u + \beta n_m = n_u + \beta M. \tag{5.29}$$

Com base em (5.29) e na aproximação de $n_u$, é agora possível representar o custo $r_s$ de transmissão de todos os canais do sistema, em função do número de canais M distribuídos em *multicast*, da variável aleatória $n_u$ que representa o número de canais requisitados em *unicast*, e do parâmetro $\beta$, que relaciona o custo de ambas as transmissões. O custo $r_s$ será uma variável aleatória com valor esperado,

$$E[r_s] = E[n_u + \beta M] = E[n_u] + \beta M. \tag{5.30}$$

Recorrendo à equação (5.30), e com o auxilio da equação (5.19), obtém-se:

$$E[r_s] = Na(1 - \sum_{k=1}^{M} \pi_k) + \beta M \tag{5.31}$$

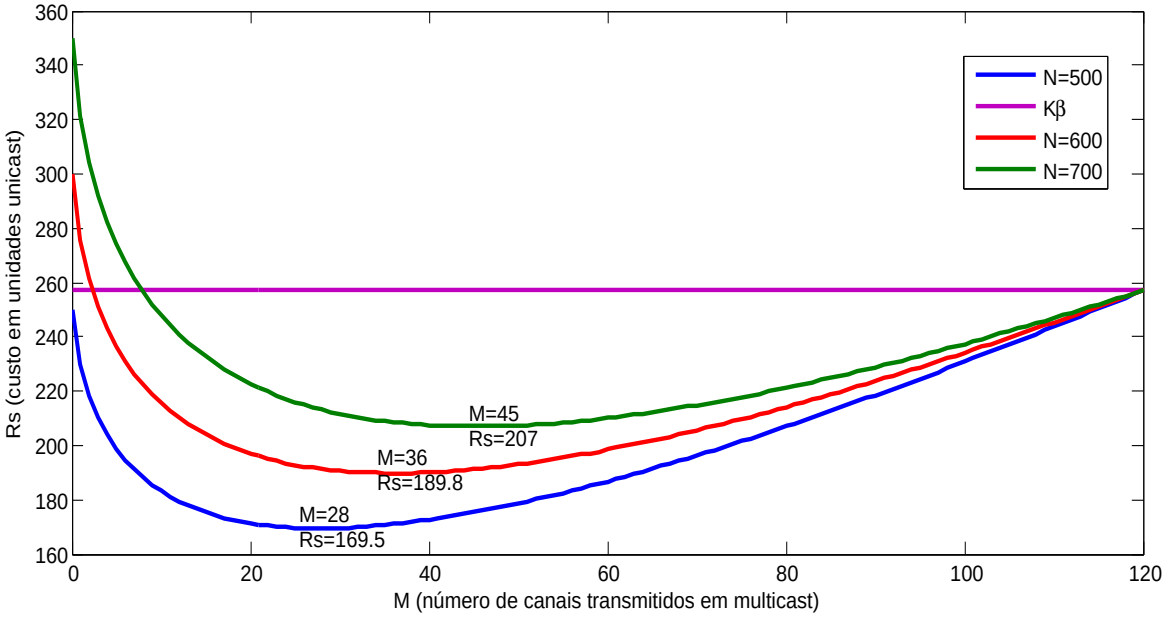



Figura 5.9: Custo Rs em função de M

---

Aplicando agora as equações definidas anteriormente, ao cenário IPTV definido na secção (5.1), é possível determinar os valores ótimos de M, para uma qualquer possibilidade de K canais e N clientes de forma célere. Na simulação (5.9) foram usados os seguintes parâmetros:

- $\alpha = 0.7$, tal como mostrado em [11], $\alpha$ deve ser um valor entre 0.5-0.8;
- $a = 0.5$, este valor será uma estimativa da probabilidade média de um cliente se encontrar ativo;
- $N_A = 10$, assume-se que o *router* $A$ possuí 10 portas ativas.
- $\gamma = 5$, assume-se que o custo de pesquisa da tabela de encaminhamento do *router* é 5 vezes superior ao custo de processar um pacote.
- $\beta = \dfrac{N_A + \gamma}{2 + \gamma} = 2.14$ para o caso de $N_A = 10$ e $\gamma = 5$.
- $N = 500, 600, 700$, assumem-se três valores de N.
- $K = 120$, assume-se que o operador disponibiliza um pacote de 120 canais;

O valor ótimo de M, ou seja aquele que minimiza $R_s$, está assinalado nas três funções de N (5.9). No caso de M=K, todos os canais são transmitidos em *multicast*, o que leva a que o custo $R_s$ seja $\beta K$, e desta forma seja o mesmo para os três conjuntos de clientes. Para o caso de M=0, o custo $R_s$ será $Na(1 - \sum_{k=1}^{M} \pi_k)$.

Em (5.9) é possível observar que o aumento do número de clientes de um sistema, força o aumento do valor ótimo de M. Isto deve-se ao facto, do aumento do número de solicitações

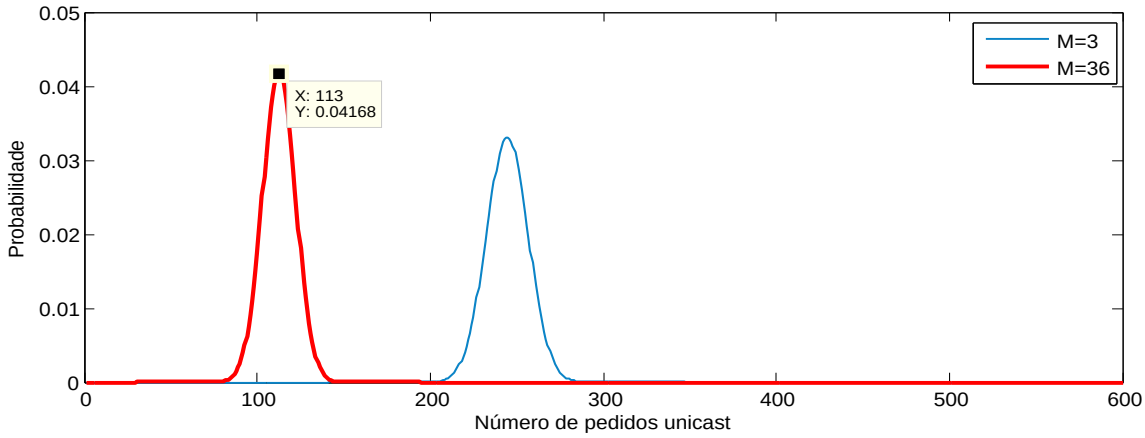



Figura 5.10: Função de distribuição de $n_u$

---

dos canais, levar a que mais canais sejam visualizados por mais pessoas, tornando o envio destes mais rentável em *multicast*. Note-se que a transmissão de um canal em *multicast* passa a ser mais económico para o operador, quando a popularidade desse canal, implica que este possua mais de $\beta$ clientes.

A figura (5.10) ilustra, a distribuição da probabilidade de solicitação de canais *unicast* $n_u$ dada por (5.23), para o cenário de (5.9) com N=600. A função a vermelho ilustra uma situação onde é usado o valor ótimo de M, ou seja M=36, e neste caso o mais provável é existirem cerca de 113 pedidos de canais distribuídos em *unicast*. Como seria de espera, a diminuição do número de canais distribuídos em *multicast*, leva ao aumento do número de pedidos de canais *unicast*, o que acontece na função a azul, com M=3. Recorde-se, que o uso de um fator de atividade $a = 0.5$ leva a que apenas estejam ativos sensivelmente metade dos N clientes, logo no caso de não ser transmitido qualquer canal em *multicast*, apenas existiriam cerca de 300 pedidos de transmissão (*unicast*).

### Probabilidade de bloqueio

Um dos objetivos de um operador de telecomunicações passa por minimizar os custos de uma rede, salvaguardando certos parâmetros de robustez. A probabilidade de um pedido ser bloqueado ($P_{bloq}$) é um dos indicadores de robustez de uma rede e espera-se que seja bastante baixo (ou nulo).

A probabilidade do sistema IPTV bloquear será a probabilidade dos recursos necessários à implementação pretendida serem superiores aos recursos disponíveis do sistema, ou seja $P[r_s > r]$, onde r representa os recursos do sistema. É assim oportuno determinar a inequação $Pr[r_s > r] \leq P_{bloq}$ ou seja, encontrar os custos mínimos em canais *unicast* de modo a que o sistema possua uma probabilidade de bloqueio inferior ao valor pretendido de ($P_{bloq}$). Desta forma,

$$P[r_s > r] = 1, \qquad se\ r < \beta M \quad e, \tag{5.32}$$

$$P[r_s > r] = P[n_u > r - \beta M], \qquad se\ r \geq \beta M. \tag{5.33}$$

Esta análise poderá ser realizada com recurso às equações (5.32) e (5.33), tendo em vista a aproximação normal da variável $n_u$ realizada em (5.23). Admitindo que Rs é também uma distribuição gaussiana com média $NaP_u + \beta M$ e desvio padrão (5.28), a probabilidade de bloqueio do sistema poderá ser obtida através da função Complementary Cumulative Distribution Function (CCDF) de Rs, tal como é ilustrado no gráfico (5.11), para os parâmetros de (5.9), com N=600 e M=36. Como se pode observar a probabilidade de bloqueio será 0.5118 para valores próximos do mínimo de Rs encontrado em (5.9)(Rs=190). Visto (Rs=190) representar aproximadamente o valor esperado[2] da distribuição normal de Rs, o valor de probabilidade de bloqueio $P_{bloq} = 0.5118$ faz sentido. Observa-se também que o gráfico (5.11) se encontra de acordo com as equações (5.32) e (5.33).

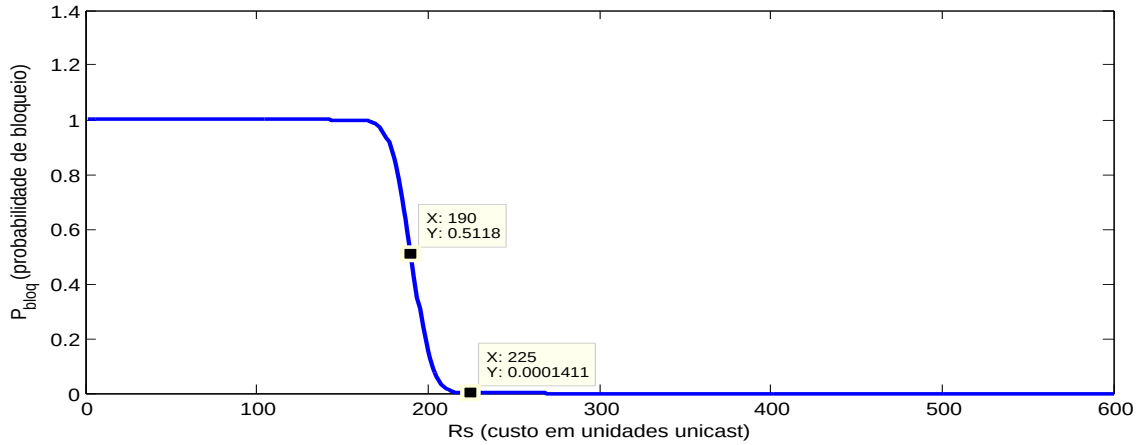



Figura 5.11: Probabilidade de bloqueio do sistema

A CCDF de Rs não é contudo a abordagem mais precisa às equações (5.32) e (5.33), já que por exemplo, não prevê a possibilidade de Rs ser inferior a $\beta K$, o que leva a que em determinados cenários possa existir $P_{bloq}$ inferior a 1 para valores de Rs inferior a $\beta K$. A expressão (5.34) [11] representa uma abordagem mais próxima às equações (5.32) e (5.33).

$$r = \beta M + NaP_u + erfc^{-1}(P_{bloq})\sqrt{NaP_u(1 - aP_u)} \tag{5.34}$$

Em (5.34) a função $erfc^{-1}$ representa a função inversa de uma CCDF com média nula e desvio padrão unitário. Observe-se que (5.34) resulta da adição de um termo à anterior expressão (5.31). Numa análise semelhante a (5.9) poder-se-à obter os recursos ($Rs$) mínimos, que uma distribuição do serviço de IPTV deve possuir para garantir uma probabilidade de bloqueio inferior a $P_{bloq}$. Para determinar o custo Rs em função do número de canais transmitidos em *multicast* M, recorreu-se à equação (5.34) e aos parâmetros de (5.9) com N=600 e uma probabilidade de bloqueio de $P_{bloq} = 0.0001$. A função está representada em (5.12) e como se observa a função apresenta um comportamento semelhante às funções do gráfico (5.9). O valor ótimo de M e o custo Rs aumentaram ligeiramente. A diminuição da probabilidade de bloqueio de um sistema leva à necessidade de aumentar os recursos fornecidos, facto que justifica os aumentos anteriormente descritos[3].

[2] O valor esperado exato será Rs=189.8. Para este valor $P_{bloq} = 0.5$.

[3] Para uma probabilidade de bloqueio de $P_{bloq} = 0.0001$ ao serem transmitidos 36 canais em *multicast* o

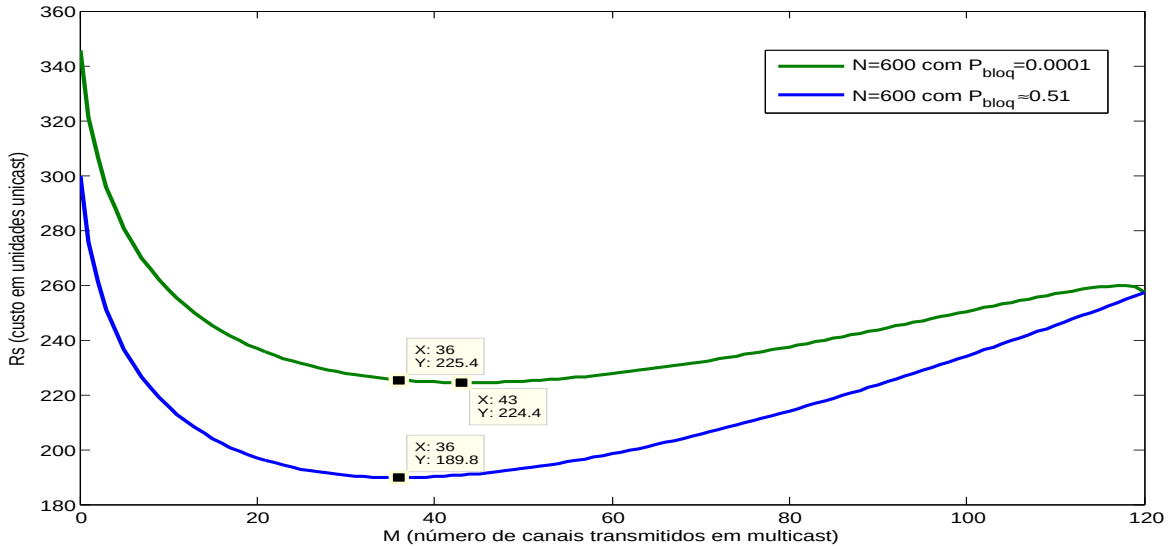



Figura 5.12: Custo Rs em função de M, com $P_{bloq} = 0.0001$

---

Fazendo um breve ponto de situação, neste momento é possível a um operador otimizar os seus recursos de distribuição de IPTV através do calculo do valor ótimo de M e calcular a probabilidade de ocorrer um bloqueio no seu sistema com base nos seus recursos. O valor de M na expressão (5.34) deverá ser aquele que minimiza os recursos (Rs).

### Vantagem do modelo

Neste ponto reside ainda uma questão que seria oportuno resolver. Até que ponto trará mais benefício para o operador aplicar uma estratégia combinada de transmissão de canais em *multicast* e *unicast*, em relação à possibilidade de transmitir todos os canais de que dispõe em *multicast* dado um sistema com N clientes?

Será razoável assumir que para um operador será vantajoso usar o modelo apresentado desde que os recursos ($R_s$) necessários à sua implementação sejam inferiores ao mínimo entre o custo de transmitir todos os K canais em *multicast* ($\beta K$), e o custo de todos os clientes possuírem uma sessão *unicast*(aN).

Sabendo que os recursos ($r_s$) aumentam à medida que o número de clientes aumenta, poder-se-à calcular o limite do número de clientes que leva todos os canais a serem transmitidos em *multicast*. Para realizar esta análise calculou-se o valor ideal de canais a transmitir em *multicast* M para cada valor de N, calculando-se depois o custo $R_s$ para cada N ($r_s(N(M_{optimo}))$), com recurso à expressão (5.31) e à expressão (5.34) com uma probabilidade de $P_{bloq} = 0.0001$[4]. Os parâmetros usados foram os de (5.9). O gráfico (5.13) mostra os resultados obtidos e as funções de custo da transmissão dos K canais em *multicast* e *unicast*.

Observa-se em (5.13) que ao ser usado $P_{bloq} = 0.51$ a função atinge o custo $\beta K$ quando N=1349. Ou seja, para uma rede com a arquitetura definida em (5.1) e para os parâmetros

custo será Rs=225.4, o que esta de acordo com o CCDF da $P_{bloq}$ em (5.11).

[4]O mínimo das funções (5.31) e (5.34) foi determinado de forma independente já que, tal como mostrado no gráfico (5.12) as funções não partilham obrigatoriamente o mesmo mínimo.

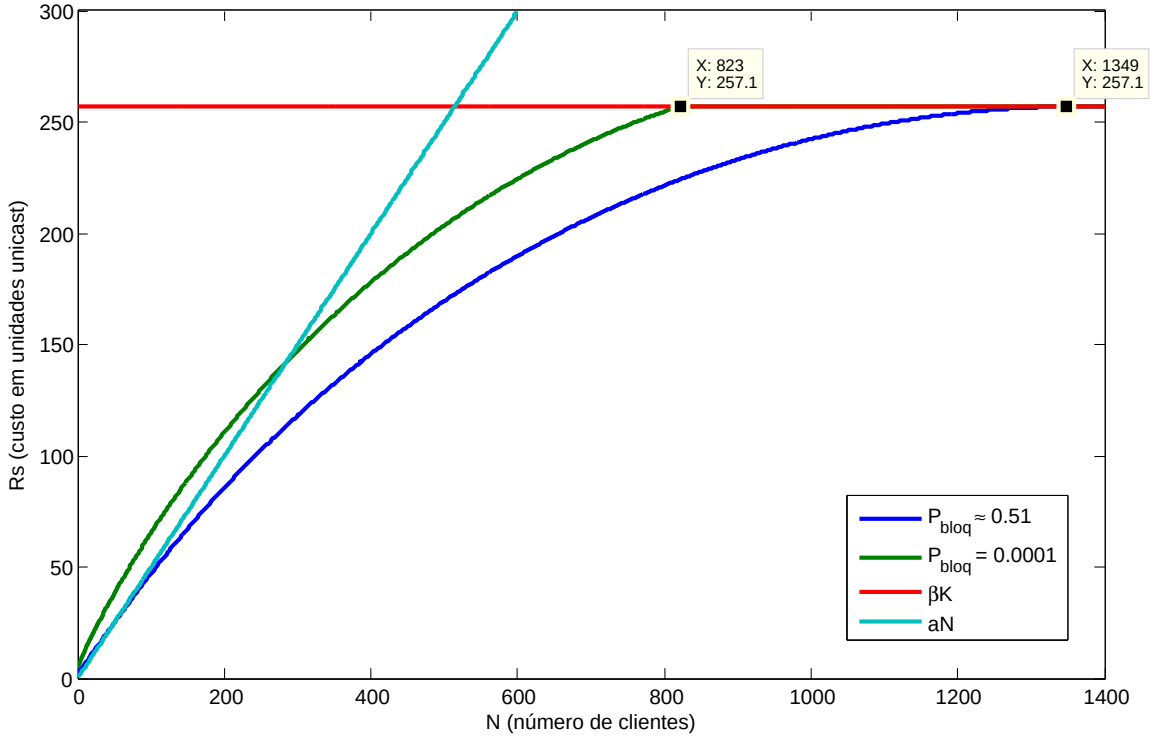



Figura 5.13: Custo Rs em função do número dos clientes

---

anteriormente assumidos uma solução de transmissão combinada entre canais *unicast* e *multicast*, fará sentido para um número de cliente inferior a 1349. Se nesta arquitetura se restringir o número de clientes a 500, haverá uma ganho de 34.27%, em relação à transmissão de todos os canais em *multicast*[5].

A função de custo Rs com probabilidade de bloqueio $P_{bloq} = 0.0001$ possuí um comportamento análogo à função Rs com $P_{bloq} = 0.51$, contudo necessita de mais recursos, alcançando por isso primeiro o valor de $\beta K$ em N=823. Confirmou-se que o custo Rs das funções para N=600 correspondem aos valores determinados em (5.12).

### Considerações

Tal como verificado a aplicação deste modelo matemático possibilita um uso mais eficiente dos recursos de processamento disponíveis numa rede de distribuição de IPTV. Contudo o desenvolvimento desta análise incorreu em algumas aproximações e considerações relativas a aplicações reais que serão importantes discutir.

A distribuição da popularidade dos canais é baseada numa distribuição Zipf que aproxima de forma razoável a realidade, tal como mostra a monitorização de um servidor IPTV em [11]. Contudo não deverá ser esquecido que para além de se tratar de uma aproximação, o valor $\alpha$ poderá apresentar variações instantâneas, o que leva a escolha dos M canais a distribuir em

[5]Nesta situação o ganho é definido como a percentagem de recursos não utilizados pela solução, em relação a $\beta K$.

*multicast* a poder ficar momentaneamente desajustada.

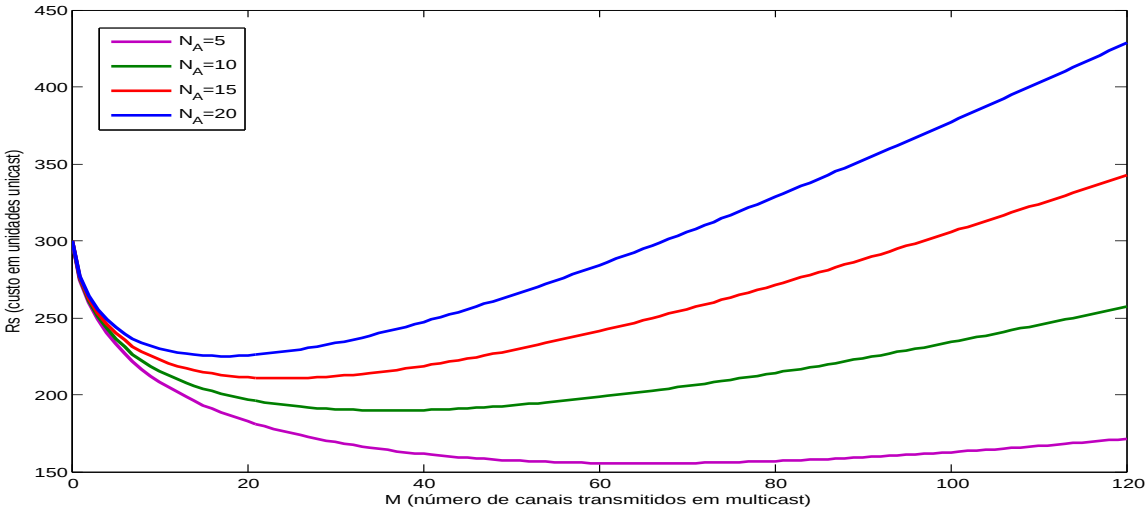



Figura 5.14: Custo Rs em função dos M canais transmitidos em multicast para *routers* de distribuição com diferente número de portas ativas

Nas anteriores simulações, considerou-se um fator de atividade dos utilizadores $a$ de 0.5, contudo este é apenas um valor ilustrativo. Este parâmetro modelador do comportamento humano variará consoante a hora do dia, época do ano ou mesmo tipo de clientes.

Neste documento a diferença de custos entre uma transmissão em *unicast* e *multicast* definida pelo parâmetro $\beta$, compreendeu os custos de processamento do *router* A na arquitetura (5.1). Visto a expressão (5.7) de $\beta$ ser função de $N_A$ e $\gamma$. E visto $\gamma$ ser função de $P_p$ e do tamanho da rede ou seja $P_s$, é aceitável afirmar que $\beta$ nunca possuirá valores excessivamente elevados. Isto porque o número de portas $N_A$ de um *router* é limitado e o aumento do tamanho e complexidade da rede será também limitado por parâmetros de performance. Note-se que um valor excessivamente elevado de $\beta$ leva a que seja sempre mais económico enviar todos os canais em *unicast*, e um valor excessivamente baixo de $\beta$ torna o envio de todos os canais em *multicast* uma solução mais económica, o que tornaria o aplicação do modelo desnecessária.

Nas figuras (5.14) e (5.15) é ilustrada a variação do custo Rs em função do número de canais transmitidos em *multicast*, quando se varia o número de portas do *router* $N_A$ e o tamanho das tabelas de encaminhamento, ou seja $P_s$. Observa-se que o aumento do número de portas ativas aumenta o custo de transmissão de um canal em *multicast*, o que leva à diminuição do número ótimo de canais M a enviar em *multicast*.

O aumento das tabelas de encaminhamento levam ao aumento do custo de pesquisa $P_s$, o que consequentemente faz aumentar $\gamma$. Este aumento faz diminuir $\beta$ e consequentemente aumenta o número de canais ótimo a transmitir em *multicast*.

A necessidade do operador definir manualmente os canais *multicast* é um facto que poderá desagradar aos operadores, nomeadamente quando o comportamento dos clientes é bastante instável. Esta instabilidade poderá ser traduzida por uma elevada variação da constante de Zipf e do fator de atividade dos clientes

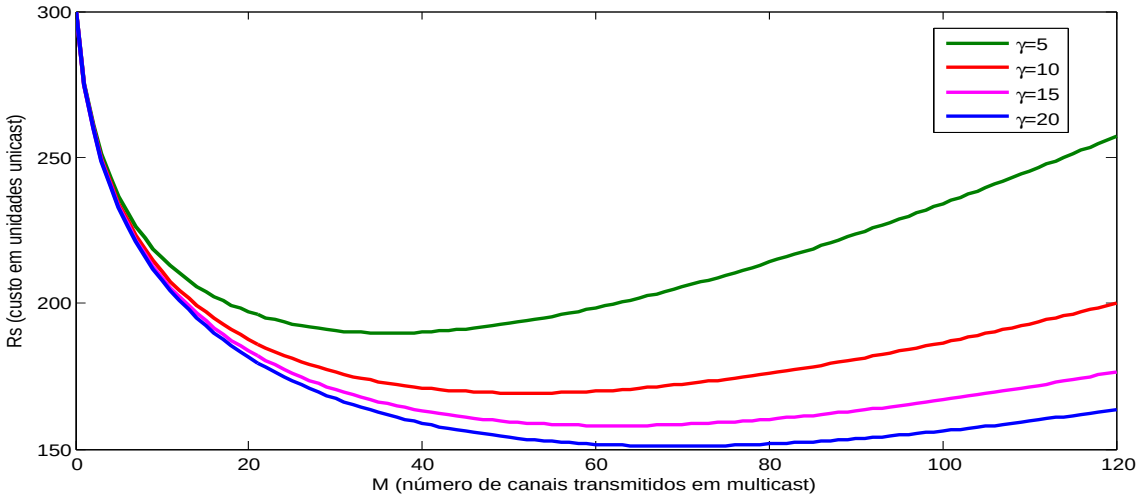



Figura 5.15: Custo Rs em função dos M canais transmitidos em multicast para *routers* de distribuição com diferentes $\gamma$

---

Não foram realizadas considerações sobre outros tipos de serviços suportados simultaneamente pela arquitetura (5.1). Contudo a oferta de serviços IPTV é muitas vezes associado à oferta de serviços de voz e dados, no pacote de serviços chamado *triple play*. Nesta condição é importante ter em atenção a influência dos restantes serviços nas transmissões IPTV. Em [43] é demonstrado que o aumento da carga paga dos blocos TDM num pacote de serviços *triple play* sobre MPLS-T, leva à deterioração das transmissões dos canais televisivos. Foi também mostrado no documento [43] que a disciplina de fila mais adequada à transmissão de canais televisivos é o *Weighted Fair Queuing*. Os valores de prioridade assumidos para cada serviço, assim como a percentagem de largura de banda de cada serviço, pode ser encontrados em [43, pág 1].

# Capítulo 6

# Conclusão

Nos últimos anos o tráfego transportado pelos operadores de telecomunicações, nas suas redes de transporte, tem crescido entre 75% a 125% por ano [7]. Contudo este aumento da capacidade dos operadores de telecomunicações, não foi acompanhada da forma desejada pelo crescimento das receitas, e uma das causas apontadas a esta discrepância, prende-se com o facto das redes de transporte tradicionais transportarem de forma ineficiente o tráfego baseado em pacotes.

O tráfego baseado em pacotes, constitui a grande maioria do tráfego que circula nas redes de transporte atuais, muito devido ao aparecimento e rápido crescimento de serviços Internet e aplicações de vídeo sobre IP, tal como serviços de videoconferência ou IPTV. É neste ponto que incide o problema que esta dissertação procurou estudar, ou seja, a eficiência na distribuição do serviço IPTV nas redes de transporte.

As redes de transporte de telecomunicações surgiram com o sistema de telefonia fixo, onde o tráfego proveniente de várias chamadas telefónicas é agregado e transportado num mesmo canal, nas tecnologias PDH e SDH. Contudo o paradigma das telecomunicações alterou-se e as chamadas telefónicas que atravessam redes através da comutação de circuitos, foram perdendo importância em função do aparecimento e crescimento de serviços Internet. No entanto a rede de transporte baseada na tecnologia SDH, não se encontrou preparada para a mudança, e as adaptações realizadas para o transporte de dados, levaram a que as receitas dos operadores não acompanhassem a sua oferta de largura de banda. As tecnologias PDH e SDH foram estudadas no primeiro capítulo, assim como as suas limitações no atual contexto das telecomunicações.

O crescimento do serviço de Internet, foi acompanhado pelo aparecimento de pequenas redes locais de comunicação de dados em tecnologia Ethernet. Hoje, mais de 95% do tráfego que atravessa as redes de transporte surge ou termina numa porta Ethernet [1]. É por isso natural, que esta tecnologia seja vista como uma forte candidata à rede de transporte, podendo assim ser uma solução completa para toda a rede de telecomunicações. Os fundamentos da tecnologia Ethernet, assim como as funcionalidades avançadas da tecnologia e as atuais implementações de 40 e 100 Gbps, foram estudadas no segundo capítulo.

Apesar da tecnologia Ethernet ser promissora, existem algumas limitações de ordem técnica que impedem a tecnologia de ser implementada em grande escala nas redes de transporte, tal como o seu alcance, os elevados tempos de recuperação do mecanismo STP ou os mecanismos de *flooding* nos *switchs*. Estas limitações foram abordadas no final do capítulo 2.

Atentos à realidade da tecnologia Ethernet operadores de telecomunicações e fabricantes

de equipamentos, formaram em 2001 o consórcio MEF com o propósito de promover a adoção a nível mundial de uma tecnologia para as redes de transporte, que tenha por base a tecnologia Ethernet [29]. Com este objetivo, definiu a tecnologia Carrier Ethernet como uma tecnologia de rede de transporte que se distingue da tradicional Ethernet em cinco atributos chave:

- Serviços normalizadas;
- Escalabilidade;
- Fiabilidade;
- Qualidade de serviço;
- Gestão de serviços.

Estes atributos foram estudados em detalhe no terceiro capítulo, e como se pode observar, encontram-se todos interligados, para assim resultarem numa tecnologia mais coerente e sustentável. Para além dos atributos anteriores, o MEF introduz também conceitos que pretendem dar resposta aos mesmos atributos. Desta forma definiu um modelo de serviços da tecnologia, os tipos de serviços Ethernet que devem ser suportados e como o fazer. Um destes serviços, o E-Tree poderá assumir um importante papel para a distribuição do serviço de IPTV.

Em paralelo com o desenvolvimento conceptual de Carrier Ethernet por parte do MEF, surgem duas tecnologias que se aproximam dos anteriores atributos e despertam o interesse dos operadores de telecomunicações. A tecnologia PBB-TE desenvolvida pelo IEEE, que surge da Ethernet usada em LANs, e a tecnologia MPLS-TP desenvolvida pelo ITU e IETF, que provém da tecnologia MPLS usada nas redes de telecomunicações de transporte.

Visto uma grande parte dos operadores, terem anunciado que as suas redes de transporte, seriam implementadas ou atualizadas, com soluções baseadas na tecnologia MPLS-TP, este documento procurou estudar as origens da tecnologia (MPLS), identificar as alterações realizadas que tornam a tecnologia apta a operar em redes de transporte e apresentar e discutir o atual estado de normalização da tecnologia.

A versatilidade e simplicidade da tecnologia MPLS, permitiu ampliar de forma considerável a escalabilidade da tecnologia IP, nas redes metropolitanas e redes de transporte. O MPLS amadureceu ao longo do tempo e ao surgir a necessidade de uma rede de transporte robusta, baseada em comutação de pacotes, o ITU e IETF uniram esforços para desenvolver um tecnologia baseada no MPLS, que possuísse os requisitos de operação de uma rede de transporte. Para adquirir estes requisitos, foi delineado que a nova tecnologia deveria ser totalmente independente da tecnologia IP, desde o plano de dados, até às funções de OAM e de sobrevivência. Algumas funcionalidades do MPLS foram também retiradas, tal como o PHP ou o ECMP, de modo a atribuir à tecnologia a predictabilidade e o controlo necessários.

As tradicionais redes de transporte baseadas em tecnologia SDH, possuem características de sobrevivência bastante robustas, o facto dos tempos de recuperação a falhas não ultrapassarem os 50ms é prova disso. O MPLS-TP não poderá assim baixar a fasquia de sobrevivência colocada pelas redes tradicionais, sob pena de perder atratividade para os operadores de telecomunicações. Os requisitos de sobrevivência do MPLS-TP estão definidos nos RFCs [38] [40], contudo a implementação destes requisitos encontra-se ainda em discussão.

Neste documento foram apresentadas e discutidas as propostas presentes nos *drafts* [5] [39] para a proteção linear ponto-a-ponto e proteção linear ponto-multiponto, e no RFC [35] para

a proteções em anel ponto-a-ponto e ponto-multiponto. Deve ser recordado que, tal como refere o documento [38], a proteção em anel, deverá ser um caso particular e otimizado da proteção linear. Para a distribuição de IPTV a arquitetura mais apelativa será baseada em ligações ponto-multiponto, e desta forma o esquema de proteção mais adequado para uma distribuição de IPTV deverá ser a proteção (partilhada) ponto-multiponto.

Os operadores de telecomunicações procuram dominar uma tecnologia através de implementações mais consistentes, que respondam de forma eficiente a qualquer realidade, ao mesmo tempo que procuram o aumento dos lucros através da otimização de recursos. É nesta ótica que começam a aparecer estudos que mostram, que o uso de estratégias de transmissão *unicast* e *multicast* combinadas, podem levar a significativas poupanças de recursos.

Estudou-se o modelo matemático de transmissão mista, presente nos documentos [11] [12] [13], num cenário de distribuição de canais IPTV sobre tecnologia MPLS-TP. A arquitetura escolhida foi a E-Tree, implementada através de ligações ponto-multiponto. O ponto chave de discussão deste mecanismo reside na escolha de quais os canais a transmitir em *unicast* e quais os canais a transmitir em *multicast*. No cenário abordado, para a distribuição de IPTV o operador seleciona M canais que transmitirá em *multicast*, sendo os restante transmitidos em *unicast* quando solicitados.

Para iniciar a análise do problema foi estudado inicialmente um modelo de custos que define um parâmetro $\beta$, como a razão entre os custos de transmitir um sinal em *multicast* e *unicast*. Os custos foram definidos em função do processamento do *router* central da arquitetura. Após esta análise foi necessário modelar o comportamento dos clientes no que toca ao modo como assistem aos canais de televisão. A distribuição da popularidade dos canais, foi aproximada por uma distribuição Zipf, com uma constante de Zipf entre 0.5 e 0.8.

Verificou-se que simulações baseadas nas equações exatas do modelo matemático são computacionalmente pesadas, o que poderia levar a simulações não realistas. Desta forma procedeu-se à aproximação das distribuição de pedidos *unicast* ($n_u$) através de uma distribuição gaussiana.

No caso de não se enviar qualquer canal em *multicast*, o custo (Rs) será aproximadamente igual ao número de clientes (N) multiplicado pelo seu fator de atividade (a). No caso de todos os canais serem transmitidos em *multicast* o custo será igual ao número total de canais (K) multiplicado pela razão $\beta$. Contudo, foi possível mostrar que existe sempre um número de canais a enviar em *multicast* (M) que minimiza os recursos necessários (Rs), e que variará entre os extremos de funcionamento anteriores, podendo até assumir estes valores.

O número de canais a enviar em *multicast* (M) que minimiza os recursos necessários (Rs), diminui à medida que $\beta$ e/ou a constante de Zipf ($\alpha$) aumenta, e aumenta à medida que o número de clientes (N) e/ou o fator de atividade (a) aumenta. É importante salientar que se estes parâmetros variarem significativamente, poderá levar a que o valor ótimo de M possa ser 0 ou K, o que significa que transmitir todos os canais em *unicast* ou *multicast* é mais económico, levando à inutilidade do modelo matemático.

Um outro parâmetro de assinalável importância é a probabilidades de bloqueio do sistema ($P_{bloq}$), definida como a probabilidade do sistema de IPTV não possuir os recursos (Rs) necessários. Este parâmetro tomará o valor 1 se os recursos do sistema forem inferiores a $\beta M$, e decresce à medida que os recursos disponíveis ultrapassam este valor.

Para aperfeiçoar o conhecimento sobre um sistema IPTV num cenário estático, é necessário conhecer os seus limites, ou seja calcular os limites dos parâmetros que levam a que a aplicação do modelo não otimize o sistema. Visto o valor da razão $\beta$, fator de atividade (a) e variável de Zipf ($\alpha$) não poderem assumir elevadas variações, e visto a escolha do número de clientes

(N), num dado sistema, ser realizada pelo operador e poder variar significativamente, será importante analisar qual o limite do número de clientes (N) que leva a um custo igual à transmissão de todos os canais em *multicast* ($\beta K$). Esta análise foi realizada no gráfico (5.13) e foi possível concluir que à medida que o número de clientes (N) suportados pelo sistema aumenta, diminui o ganho em relação à transmissão de todos os canais em *multicast*, até atingir o ponto em que deixa de existir ganho. Verificou-se também que a diminuição da probabilidade de bloqueio $P_{bloq}$ da rede, leva a um aumento dos custos de transmissão da solução mista, o que também faz diminuir o ganho em relação à transmissão de todos os canais em *multicast*, atingindo assim o valor de $\beta K$ para valores inferiores do número de clientes. Pelo referido, um operador deverá assumir um compromisso entre o número de clientes e probabilidade de bloqueio do sistema em relação ao ganho face à transmissão total em *multicast*.

## 6.1 Trabalho futuro

Quanto ao conjunto de tecnologias que definirão as próximas redes de transporte, este documento estudou em detalhe a tecnologia Ethernet e MPLS-TP. Contudo seria importante uma análise detalhada à tecnologia WDM, OTN e GFP, já que estas constituirão a base tecnológica das soluções de rede de transporte no que toca às tecnologias de camada OSI inferiores.

Relativamente ao quinto capítulo, seria interessante aprofundar algumas questões. A constante $\beta$, é um parâmetro bastante importante, já que é responsável pela inserção no modelo matemático de variáveis que representam parâmetros físicos, técnicos e económicos da rede. Em [44] é realizado um estudo, em função dos custos de ligação, contudo a generalização deste a custos de nós, e outros custos derivados poderia ser matéria de um trabalho futuro.

As vantagens provenientes do recurso a um cenário onde os canais são transmitidos em *multicast* ou *unicast* consoante a variação do número de clientes que está a assistir ao canal em cada instante, são apresentadas em [11]. Seria importante uma análise dos custos de processamento e monitorização desta solução comparativamente ao cenário anteriormente estudado.

# Bibliografia


[1] Daniel Jorge Lobato de Macedo. O plano de controlo das redes de transporte Ethernet. Master's thesis, Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática, 2010.

[2] Stuart Little, Harris Stratex, and Trevor Manning. *Deployment and Link Planning of Adaptive Coding and Modulation Radio Networks.* Microwave Journal Vol. 52 No. 11 November Supplement 2009 Page 12, 2009.

[3] José Ferreira da Rocha. *Apontamentos da disciplina de Redes Ópticas.* Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática, 2010.

[4] Dieter Beller and Rolf Sperber. *MPLS-TP The New Tecnology for Packet Transport Networks.* Alcatel-Lucent Deutshland AG.

[5] E. Osborne, Stewart Bryant, N. Sprecher, A. Fulignoli, and Y. Weingarten. *MPLS-TP Linear Protection - draft-ietf-mpls-tp-linear-protection-09.* IETF, Agosto, 2011.

[6] Paul Izzo. *Gigabit Networks.* John Wiley and Sons Ltd, 2000.

[7] *An Overview of Next-Generation 100 and 40 Gigabit Ethernet Technologies.* IXIA white paper, 915-0908-01 Rev D, July 2011.

[8] M.N.O. Sadiku and S.R. Nelatury. IPTV: An alternative to traditional cable and satellite television. *Potentials, IEEE*, 30(4):44 –46, july-aug. 2011.

[9] André Miguel Dias Claro. Framework para Personal TV. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, 2008.

[10] Mark Norris. *Gigabit Ethernet - Technology and Applications.* ARTECH HOUSE, INC., 2003.

[11] Zlata Avramova, Danny De Vleeschauwer, Sabine Wittevrongel, and Herwig Bruneel. Capacity gain of mixed multicast/unicast transport schemes in a TV distribution network. *Multimedia, IEEE Transactions on*, 11(5):918 –931, aug. 2009.

[12] Zlata Avramova, Danny De Vleeschauwer, Sabine Wittevrongel, and Herwig Bruneel. *Models to Estimate the Unicast and Multicast Resource, Demand for a Bouquet of IP-Transported TV Channels.* SMACS Research Group, Department of Telecommunications and Information Processing (TELIN), Ghent University.

[13] Zlata Avramova, Danny De Vleeschauwer, Sabine Wittevrongel, and Herwig Bruneel. *Dimensioning Multicast-Enabled Networks for IP-Transport TV Channels*. SMACS Research Group, TELIN, Ghent University, 2007.

[14] Gustavo Campos Sebastião. Comparação de desempenho de redes SDH convencionais e de redes NG-SDH/WDM para o transporte de tráfego IP. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, 2008.

[15] Harry G. Perros. *Connection-oriented Networks, Sonet/SDH, ATM, MPLS and OPTICAL NETWORKS*. 2005.

[16] João Morais Davim. Ethernet para a rede de transporte. Master's thesis, Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática, 2010.

[17] Charles E. Spurgeon. *Ethernet The Definitive Guide*. O'Reilly & Associates, Inc, 2000.

[18] M. Schwartz. *Computer-communication network design and analysis*. Prentice-Hall, 1977.

[19] Andrew S. Tanenbaum. *Computer Networks (third edition)*. Prentice-Hall, Inc, 1996.

[20] Ben Brahim Taha, Jingren Jin, and Arunkumar T J. Comp6340: Network quality assurance and simulation. *Department of CSSE, Auburn University*.

[21] Natalia Olifer and Victor Olifer. *Computer Networks: principles, tecnologies, and protocols for network design*. John Wiley and Sons Ltd, 2006.

[22] Gilbert Held. *Ethernet Networks, 4th Edition*. John Wiley and Sons Ltd, 2003.

[23] Armando Nolasco Pinto and Rui L. Aguiar. *Apontamentos da disciplina de Redes de Telecomunicações*. Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática, 2010.

[24] Gilbert Held. *Carrier Ethernet - Providing the Need for Speed*. Auerbach Publications, 2008.

[25] Mathias Hein and David Griffiths. *Switching Tecnology In The Local Network*. International Thomson Publishing, 1997.

[26] *Standards IEEE for Local and Metropolitan Area Networks: Supplements to Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications - Specification for 802.3 Full Duplex Operation and Physical Layer Specification for 100 Mb/s Operation on Two Pairs of Category 3 Or Better Balanced Twisted Pair Cable (100BASE-T2) - IEEE Std 802.3x-1997 and IEEE Std 802.3y-1997 (Supplement to ISO/IEC 8802-3: 1996; ANSI/IEEE Std 802.3, 1996 Edition)*, 1997.

[27] Jayant Kadambi, Mohan Kalkunte, and Ian Crayford. *Gigabit Ethernet - Migrating to High-Bandwidth LANs*. Prentice Hall PTR, 1998.

[28] John D'Ambrosia, David Law, and Mark Nowell. *40 Gigabit Ethernet and 100 Gigabit Ethernet - Technology Overview*. Ethernet Alliance, 2008.

[29] Metro Ethernet Forum. *http://metroethernetforum.org*. acedido em 2011.

[30] *http://www.inova-ria.pt/noticias/detalhe.asp?id=289*. acedido a 11 de Julho de 2011.

[31] Abdul Kasim. *Delivering Carrier Ethernet Extending Ethernet beyond the LAN*. The McGraw-Hill Companies, 2008.

[32] Adrian Farrel and Igor Bryskin. *GMPLS - Architecture and Applications*. Morgan Kaufmann, 2006.

[33] E. Rosen, A. Viswanathan, and R. Callon. *Multiprotocol Label Switching Architecture*. IETF RFC 3031 (Standards Track), 2001.

[34] Zhuo (Frank) Xu. *Designing and Implementing IP/MPLS-Based Ethernet Layer 2 VPN Services*. Wiley Publishing, Inc, 2010.

[35] Loa Anderson and Stewart Bryant. *Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport Profile*. IETF RFC 5317, 2009.

[36] Luyuan Fang, Dan Frost, Nabil Bitar, Raymond Zhang, Masahiro DAIKOKU, Jian Ping Zhang, Lei Wang, Henry Yu, Mach(Guoyi) Chen, and Nurit Sprecher. *MPLS-TP Use Cases Studies and Design Considerations draft-fang-mpls-tp-use-cases-and-design-03.txt*. 2011.

[37] M. Bocci, M. Vigoureuxm, and S. Bryant. *MPLS Generic Associated Channel*. IETF RFC 5586, 2009.

[38] N. Spencher and A. Farrel. *MPLS-TP Survivability Framework*. IETF RFC 6372, Setembro, 2011.

[39] G. Liu, Y. ji, J. Yang, j. Yu, and Z. Du. *Multiprotocol Label Switching Transport Profile p2mp Shared Protection*. IETF draft-liu-mpls-tp-p2mp-shared-protection-02, Agosto, 2011.

[40] B. Niven-Jenkins, D. Brungard, M. Betts, N. Sprencher, and S. Ueno. *Requirements of an MPLS Transport Profile*. IETF RFC 5654, Setembro, 2009.

[41] Manuel Domínguez Dorado. *Soporte de Garantía de Servicio (GoS) sobre MPLS Mediante Técnicas Activas, Proyecto Final de Carrera*. 2004.

[42] Francisco Vaz. *Probabilidades e Processos Estocásticos para Engenharia Electrotécnica*. Universidade de Aveiro, 2002.

[43] Chang Cao, Yongjun Zhang, Shanguo Huang, Yongli Zhao, Song Yu, Jie Zhang, and Wanyi Gu. Demonstration of transmission performance in MPLS-TP network using streaming media traffic. In *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pages 1 –3, march 2010.

[44] John Chung I. Chuang and Marvin A. Sirbu. *Pricing Multicast Communication: A Cost-Based Approach*. Kluwer Academic Publishers, 2001.