“INTERNET BASEADA EM REDES ÓPTICAS”

Eduardo Reigada de Toledo e Silva

Trabalho Final de Mestrado Profissional em Computação

Instituto de Computação
Universidade Estadual de Campinas

"INTERNET BASEADA EM REDES ÓPTICAS"

Eduardo Reigada de Toledo e Silva

20 de dezembro de 2004

Banca Examinadora:

- **Prof. Dr. Nelson L. S. da Fonseca**
  Instituto de Computação - Unicamp
- **Prof. Dr. Edmundo R. M. Madeira**
  Instituto de Computação - Unicamp
- **Prof. Dr. Omar Branquinho**
  Instituto de Informática – PUC-Campinas
- **Prof. Dr. Célio Cardoso Guimarães**
  Instituto de Computação – Unicamp

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues - CRB8a / 2116

Reigada, Eduardo
R272i Internet baseada em redes ópticas / Eduardo Reigada de Toledo e Silva -- Campinas, [S.P. :s.n.], 2004.

Orientador : Nelson Luis Saldanha da Fonseca
Trabalho final (mestrado profissional) - Universidade Estadual de Campinas, Instituto de Computação.

1. Redes de computadores. 2. Comunicações óticas. 3. Internet. I. Fonseca, Nelson Luis Saldanha da. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Título em inglês: Internet upon optical networks.

Palavras-chave em inglês (Keywords): 1. Computer networks. 2. Optical communications. 3. Internet

Área de concentração: Sistemas de Computação (Redes de Computadores)

Titulação: Mestre em Computação

Banca examinadora: Prof. Dr. Edmundo R.M. Madeira (IC/UNICAMP)
Prof. Dr. Omar Branquinho (II-PUC-CAMPINAS)
Prof. Dr. Célio Cardoso Guimarães (IC-UNICAMP)

Data da defesa: 20/12/2004

"INTERNET BASEADA EM REDES ÓPTICAS"

Este exemplar corresponde à redação final do Trabalho Final devidamente corrigido e defendido por Eduardo Reigada de Toledo e Silva e aprovado pela Banca Examinadora.

Campinas, 20 de dezembro de 2004.

Prof. Dr. Nelson L. S. Fonseca
(Orientador)

Trabalho Final apresentado ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Computação na área de Redes de Computadores.

TERMO DE APROVAÇÃO

Trabalho Final Escrito defendido e aprovado em 20 de dezembro de 2004, pela Banca Examinadora composta pelos Professores Doutores:

Prof. Dr. Omar Carvalho Branquinho
PUCCAMP

Prof. Dr. Edmundo Roberto Mauro Madeira
IC - UNICAMP

Prof. Dr. Nelson Luis Saldanha da Fonseca
IC - UNICAMP

Agradecimentos

Agradeço, primeiramente, a Deus por ter me concedido saúde, disposição e perseverança para concluir este trabalho.

Ao apoio, suporte, compreensão e amor irrestritos que recebi da minha esposa Cristina, das minhas filhas Gabriela e Milena, bem como da minha mãe Zélia Reigada, que já descansa nos braços eternos do Senhor, pelas inúmeras horas de convívio familiar que deixaram de existir, tendo em vista o tempo requerido para os estudos e para a elaboração desta dissertação.

Ao meu orientador, Professor Nelson Fonseca, pela sua paciência, dedicação e pelo seu espírito imbatível, sempre me incentivando e mostrando o melhor caminho a seguir.

Resumo

A Internet está cada vez mais se concretizando como o novo meio universal de telecomunicação. Dentro de todos os meios existentes hoje, o IP sobre WDM (Wavelength Division Multiplexing), assim como sua variação o DWDM (Dense Wavelength Division Multiplexing), tem sido visualizado como a arquitetura mais promissora para o novo paradigma da Internet. Conseqüentemente o projeto de redes IP, utilizando-se como meio de transporte fibras ópticas, é um fator crucial neste novo paradigma, pois a capacidade de uma rápida recuperação da rede é a base na qual se dá sustentabilidade ao modelo. Neste trabalho, analisar-se-ão as propostas existentes hoje para a interconexão de roteadores IP com o núcleo de redes ópticas, tanto do Internet Engineering Task Force (IETF), quanto de outras organizações internacionais. Mostrar-se-ão algumas alternativas de arquitetura para integração do IP sobre DWDM. Tratar-se-ão também de questões como problemas de roteamento, sinalização, controle e capacidade de recuperação da rede. Finalmente, analisar-se-ão quais seriam os próximos passos vislumbrados hoje que as redes IP sobre DWDM deveriam seguir, através de uma análise do modelo adotado pela CANARIE.

Abstract

Internet does become the "*de facto*" universal communication carrier. IP over WDM, as well as its variation named DWDM, has been considered as the most promising architecture for the new paradigm of the Internet. IP network, using fiber optics as transmission medium, is of paramount importance for this new paradigm, given its capacity of fast network recovery. In this work, we analyze the existent proposals for IP network based on an Optical core, standardized in the IETF, as well as others international organizations. We show some alternatives to integrate IP over DWDM. We discuss also topics as routing, signaling, network recovery control and capacity. The main idea is that the physical layer can provide a fast protection and the network layer can provide a more intelligent recovery. Finally, we analyze which would be the steps visualized today for the deployment of the IP over DWDM networks, through an analysis of the model adopted by CANARIE.

Índice

CAPÍTULO 1

IP SOBRE WDM

Nos dias atuais, tanto a indústria de telecomunicações, quanto a indústria de informática, quando querem dar uma idéia dos avanços tecnológicos nas suas respectivas áreas, usam como parâmetro a velocidade. Apesar de utilizarem o mesmo parâmetro, cada uma delas tem suas especificidades; por exemplo, a indústria de informática ou computação refere-se à velocidade de processamento, enquanto a indústria de telecomunicações refere-se à taxa de transmissão. Aqueles que trabalham na vanguarda da tecnologia utilizam-se de inovação, do conhecimento e da formação de novos conceitos e paradigmas; como se podem observar nas áreas de circuitos integrados, roteamento, comutação, rádio freqüência e transporte através de fibras ópticas.

Enquanto que a taxa de transmissão de dados (*"bandwidth"*) em circuitos, ou o poder computacional, tem seguido, nos últimos anos, a lei de Moore, ou seja, a cada dezoito (18) meses dobra-se a capacidade de transmissão enquanto o preço decresce pela metade, os limitantes desta lei, têm sido ultrapassados na arena das redes ópticas. Nas redes ópticas, a capacidade de transmissão de dados ("*bandwidth"*) tem dobrado a cada ano. Atualmente é comum ter-se quarenta (40) comprimentos de onda ("*wavelength"*), tendo cada comprimento de onda a capacidade de transmissão de 2,5 a 10 Gbps, por par de fibras ópticas. Para se ter uma idéia do que isso significa, em um único par de fibras, com trinta e dois (32) comprimentos de onda poderiam falar simultaneamente ao telefone cinco (5) milhões de pessoas. Sistemas mais novos têm a capacidade de transmitir até cento e sessenta (160) comprimentos de onda em canais de dez (10) Gbps, ou seja, 1.6 Tbps sobre um único par de fibras ópticas. Sistemas testados em laboratórios já conseguiram uma taxa de transmissão de dados de oitenta (80) Gbps por canal. Nesta taxa de transmissão, cento e sessenta (160) comprimentos de onda por canal de oitenta (80) Gbps resultam em uma taxa efetiva de transmissão de dados de 12.8 Tbps em um único par de fibras, o que é muitas vezes superior à demanda de toda a rede de voz existente hoje no mundo [1].

A grande questão que surge no tocante a essa expansão tecnológica é em relação ao impacto que esta capacidade terá na infra-estrutura de rede no futuro, i.e., qual o impacto na Qualidade de Serviço (QoS), nos "*digital loop carriers"*, nas pilhas de protocolos etc. Para que se possa tentar responder a esta questão, deve-se perguntar também se o protocolo IP sobre WDM será o protocolo *de facto* para se tornar o padrão na infra-estrutura de rede mundial. Outra questão de igual importância a ser feita é a razão de se ter diversos níveis de camadas de protocolos. O cerne de todo o questionamento que a indústria deve fazer hoje é se fossemos recomeçar a construir as pilhas de protocolos, será que todos esses níveis existentes hoje ainda iriam permanecer? Caso a resposta seja negativa, a indústria deve se preparar para esta nova infra-estrutura e os desafios naturais que se apresentarão para que este objetivo seja atingido. Desafios como Qualidade de Serviço (QoS) e robustez da rede, tempo para aprovisionamento (para habilitar conexões sob demanda), granularidade de taxa de transmissão de dados, custos e gerência devem ser analisados antes que esta visão se torne realidade. Antes de se tentar responder a esta questão, a infra-estrutura atual representada pela Figura 1.1 será descrita.

Utilizar-se-á, neste trabalho, o termo WDM ("*Wavelength Division Multiplexing"*) tanto para WDW propriamente dito, quanto para DWDM ("*Dense Wavelength Division Multiplexing"*) de forma intercambiável.

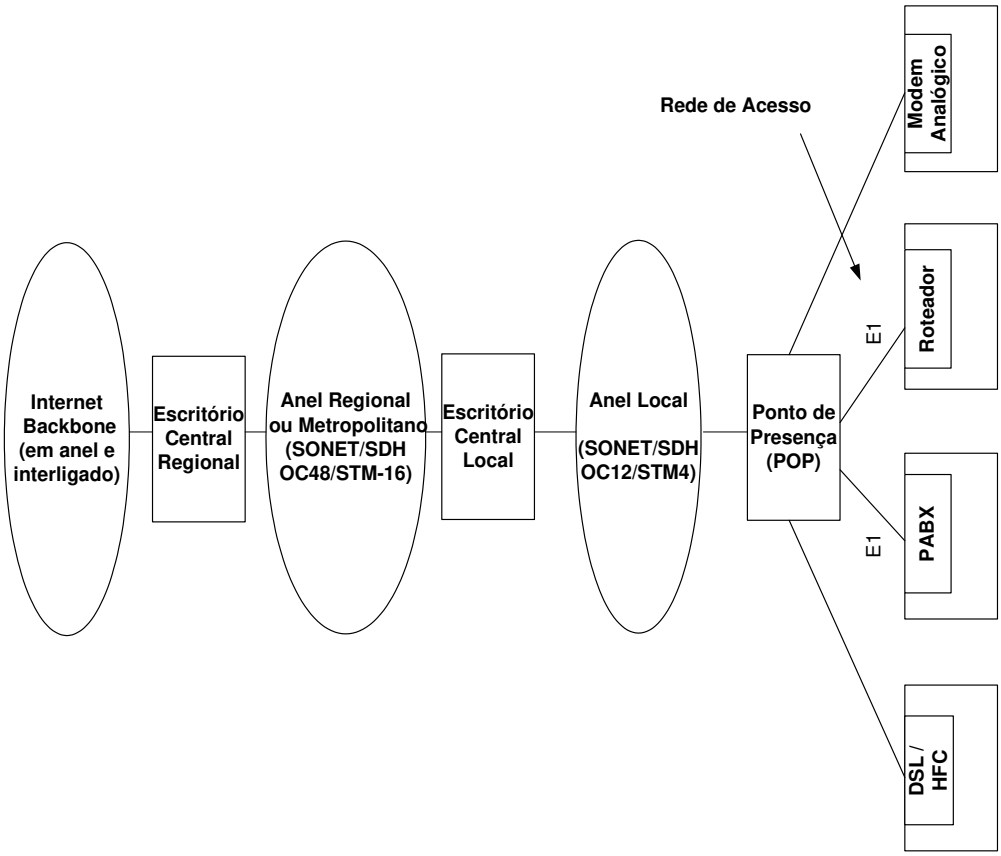


Figura 1.1 – Infra-estrutura da rede IP

A infra-estrutura atual utiliza três (03) níveis de redes em anéis: local, regional ou metropolitano e núcleo ou *backbone* ("*Long Haul"*). Os anéis se interconectam através de conexões ópticas cruzadas ("*Optical Cross-Connects"* – OXCs) em pontos de junções de alta tro10ca de tráfego ou através de multiplexadores ADM ("*Add-Drop Multiplexers"*) em pontos de junção de baixa troca de tráfego (até STM-1 ou OC-3). A rede de acesso consiste de fios de cobre usando canais analógicos, xDSL, RDSI, fibra óptica e cabos coaxiais (HFC – "*Hybrid Fiber and Coax*") e acessos via rádio freqüência em alguns casos. Nos equipamentos OXCs e ADMs, a arquitetura se baseia na conversão O/E/O (Óptica/Elétrica/Óptica), sendo que internamente todo o processamento do nó é feito no domínio elétrico. Nos anéis metropolitanos e locais utiliza-se como transporte SONET ("*Synchrounous Optical Network"*) ou SDH ("*Synchrounous Digital Hierarchy"*). Os anéis do núcleo da rede ("*backbone"*) usam WDM com múltiplos comprimentos de onda em cada par de fibras. A regeneração do sinal é feita entre duzentos (200) Km a quinhentos (500) Km na rede terrestre, dependendo do tipo de fibra utilizado, e entre cinqüenta e seis (56) Km a cento e dez (110) Km nas redes submarinas. Este modelo de

arquitetura de redes foi desenvolvido para atender a demanda da rede de telefonia na utilização de canais de voz de 64 Kbps. Esta arquitetura sofre com os problemas de alto custo existente na conversão O/E/O, que acontece nos equipamentos regeneradores, OXCs e ADMs. A atualização de uma rede como esta, bem como o aprovisionamento de circuitos e serviços através de múltiplos anéis torna-se muito custosa e trabalhosa para as operadoras de telefonia, pois a grande maioria destas atividades ainda tem de ser feitas através de processos manuais.

Assim, há uma tendência no uso de redes interligadas, tanto quanto possível, através de roteadores IP/Ópticos (interconectados via DWDM), configurando os caminhos de luz ("*ligthpath"*) dinamicamente sob demanda, dentro dos quais toda e qualquer operação para regenerar/amplificar os sinais e roteamento são feitos no domínio óptico sem a necessidade de qualquer conversão O/E/O. Configurações dinâmicas de LSPs ("*Label Switch Path"*) ao nível de rede IP até a camada física de fibras ópticas, proporcionam diminuições no aprovisionamento de dias e meses para milisegundos e/ou segundos apenas. Os custos serão reduzidos drasticamente devido à eliminação das conversões O/E/O e pela redução da regeneração dos sinais, com exceção das redes de backbone ("*Long Haul"*).

Os atributos-chave deste paradigma para as redes Internet podem ser resumidos nos seguintes aspectos:

1. O DWDM é utilizado em toda a infra-estrutura, incluindo o acesso. As redes ópticas passivas (PONs – "*Passive Optical Networks"*), especificamente as redes ópticas passivas Ethernet , já prove uma alternativa para servir inúmeros usuários a um custo muito baixo devido ao uso compartilhado dos recursos da banda passante. As redes de acesso serão baseadas mais em anéis do que ponto-a-ponto aumentando desta maneira a confiabilidade e robustez da rede como um todo.
2. Os equipamentos de conexão cruzada óptica (OXCs) e de multiplexação (ADMs) serão substituídos por comutadores ("*switches"*) ópticos e roteadores que configurariam caminhos ópticos ("*ligthpath"*) em resposta ao pedido dos clientes nas pontas. Assim eliminar-se-iam as custosas conversões O/E/O e as regenerações dos sinais.
3. Devido aos desperdícios existentes na utilização da banda passante na arquitetura em anel, essa arquitetura seria substituída por redes interligadas ("*mesh"*) nas áreas locais, metropolitanas e nos backbones.
4. As redes baseadas em tecnologia Gigabit Ethernet serão usadas nas operadoras de longa distância, substituindo o SONET/SDH e o ATM onde for possível.
5. A maioria do tráfego será através da rede de pacotes e da diferenciação de QoS usando técnicas como o serviço IP Diffserv e do MPLS ("*Multiprotocol label Switching"*).

Algumas dessas tecnologias para tornar viável o novo paradigma da arquitetura de redes já estão disponíveis, embora em uso de forma limitada, outras tecnologias ainda estão em fase de desenvolvimento.

1.1 O porquê do IP e do WDM?

Uma transição da rede de circuitos para a rede de comutação de pacotes já está ocorrendo e a idéia de uma rede de pacotes vem crescendo e sendo implementada por um número cada vez maior de operadoras. Em termos de volume, hoje, o tráfego de dados já superou o tráfego de voz. Em um horizonte de cinco (05) anos a tendência é que o volume aumente de cinco (05) a dez (10) vezes. Os investimentos em soluções baseadas em IP é um caminho natural, tendo em vista a independência do protocolo de rede IP em relação aos protocolos das camadas de nível de enlace e das camadas no nível físico; apesar da falta de suporte existente na camada IP para serviços de tempo real e que necessitam de garantia de qualidade de serviço (QoS). Enquanto muitas soluções baseadas em IP esperam uma maior robustez e confiabilidade para ser tornarem produtos de fato, o aumento e disponibilidade de altas taxas de transmissão de dados fornecidas pelas redes ópticas têm aliviado e muito as deficiências do IP, fazendo com que se possa até mesmo oferecer serviços de tempo real baseados no modelo do melhor esforço ("*best effort*") da Internet.

Redes públicas de pacotes requerem robustez, confiabilidade e escalabilidade em termos de roteadores, ADMs, OXCs etc. Se as taxas de transmissão continuarem aumentando nas fibras ópticas, os elementos de rede acima citados deverão ter a capacidade de processarem os dados na mesma velocidade que a rede oferece. Para que isso aconteça, é necessário que se mude a arquitetura de rede existente de tal maneira que a rede pública possa oferecer qualidade de serviço e ser escalável, com o menor custo possível. Felizmente, as redes ópticas oferecem o menor custo de bit possível se comparado aos outros meios de transmissão de dados hoje existentes. Deve-se juntar a isso também o fato de que há um excedente considerável de banda nas redes de fibras ópticas, o que faz com que o custo também diminua.

Com o aumento da eficiência da rede, irão diminuir tanto a necessidade de vários níveis de rede, como a integração de suporte para a rede IP. Neste cenário, faz-se necessário o desenvolvimento de uma camada de adaptação, ou um controle de acesso ao meio físico integrado ao IP, fazendo a ligação entre a camada de rede IP e a camada física WDM.

1.2 O que o WDM oferece?

O WDM oferece um aumento significativo de banda na infra-estrutura óptica existente através da multiplicação de canais ópticos em um único par de fibras. Com isso, problemas de qualidade de serviço em redes de melhor esforço são atenuados na utilização do IP nas redes ópticas. Enquanto os padrões de QoS (e.g. MPLS, DiffServ) são implementados nas redes já em produção, existe uma enorme capacidade de banda nas redes ópticas que potencialmente resolveriam o problema da Qualidade de Serviço através do aprovisionamento de mais banda por canal. As redes ópticas utilizando o WDM podem ser vistas como uma grande rodovia, na qual as faixas são expansivas, onde se pode simplesmente ligar uma cor diferente dentro de uma mesma fibra e conseguir muito mais banda passante. É como se em uma rodovia, nós pudéssemos ir acrescentando faixas a mais, de acordo com a necessidade de tráfego. A atualização de redes WDM é menos custosa que a atualização necessária em uma rede SONET/SDH. O WDM oferece a mesma solução a um custo três vezes menor, aumentando em trinta (30) vezes a

capacidade de transmissão [1]. Ele oferece uma rede segura e com baixo custo de manutenção.

O WDM provê também a possibilidade de serem criados caminhos de luz ("*ligthpath"*), que possibilitam a minimização do número de pulos ("*hops"*) da rede. Inúmeras aplicações que requerem QoS poderiam se beneficiar de um único pulo ("*hop"*) existente entre roteadores de entrada/saída ("*edge routers"*) ou entre duas grandes cidades ligadas por um caminho de luz ("*ligthpath"*) sem a necessidade de conversão O/E/O. Ele também oferece uma melhoria no roteamento, proteção das rotas e na transparência do sinal. Ele pode carregar protocolos digitais e analógicos ao mesmo tempo, pois ele é um protocolo transparente ao meio, ou seja, o WDM não se importa com o que é carregado na carga útil ("*payload"*) de um caminho de luz. Apesar do WDM no início ser usado com uma rede ponto-a-ponto, atualmente ele está sendo utilizado como uma rede interligada ("*mesh"*) com um aumento de inteligência da rede na camada óptica. Com isso, podemos dizer que o WDM evoluiu de uma rede de transporte tática ponto-a-ponto para uma camada de rede inteligente, estratégica e reconfigurável sob demanda.

1.3 Capacidade, tipos de interface e protocolos.

O grande propulsor ou catalisador da necessidade de altas taxas de tráfego de dados é o protocolo IP. A demanda hoje por portas de 155 Mbps (OC-3 SONET / STM-1 SDH) é relativamente usual para as grandes operadoras e grandes provedores de serviços de rede (NSP – "*Network Service Providers"*). As necessidades de portas existentes hoje já chega a 2,5 Gbps (OC-48 SONET / STM-16 SDH) para grandes operadoras em interconexões de longa distância terrestre ou submarina e algumas já possuem portas de 10 Gbps (OC-192 SONET / STM-64 SDH). Hoje a negociação de portas com velocidade menores do que um (1) STM-1, entre operadoras de longa distância e grandes NSP, torna-se pouco atraente devido ao custo/benefício em relação à interface, equipamento, manutenção e gerenciamento. Velocidades consideradas pequenas hoje, como 1,5 Mbps (T1) ou 2 Mbps (E1), são mais caras do que velocidades de 45 Mbps (T3) ou 34 Mbps (E3). A tendência tem mostrado que a utilização de velocidades menores, tais como 2 Mbps (E1) tendem a diminuir abruptamente no futuro próximo.

Tipicamente a rede de comunicação hoje oferece múltiplas camadas ou níveis, indo do IP ao ATM, do ATM do SONET/SDH, do SONET/SDH à camada óptica e esta ao duto físico de transporte da fibra.

Cada uma das camadas de rede possui flexibilidade, vantagens e características próprias, assim como desvantagens também. O objetivo é otimizar a rede de tal maneira que possamos obter vantagens de cada componente em sua respectiva camada em cada uma das partes da rede. Como por exemplo, imagine uma grande operadora interligando São Paulo ao Rio Grande do Sul e que nesta rede há tráfego de IP sensível à velocidade ("*forwarding"*). Não seria razoável o tráfego ter que passar por três (03) ou quatro (04) roteadores com retardo de processamento, retardo de transmissão e de enfileiramento em cada um dos nós. Uma alternativa ideal seria criar um único caminho de luz no domínio óptico com um único pulo ("*hop"*). Isto proporcionaria um melhor desempenho, uma minimização de custos e um protocolo transparente.

No entanto, existem diversas camadas na pilha de protocolos. O IP hoje é encapsulado no Frame Relay e/ou no ATM, e estes são encapsulados no SONET/SDH, que são predominantes nas redes de longa distância das operadoras, que, por sua vez, são encapsulados no WDM. Isso aconteceu, principalmente, por força de custos de investimentos feitos anteriormente e necessidade de se compatibilizar as tecnologias mais novas com as redes existentes e operacionais. Se as redes fossem começar do zero hoje, provavelmente não existiriam as camadas ATM e SONET/SDH [1].

1.4 Por que IP sobre WDM?

Enquanto o WDM oferece uma grande capacidade de banda, o IP oferece uma grande convergência, assim a combinação de ambos representa o caminho natural. A camada de rede IP é o padrão *de facto* para protocolos e a camada WDM oferece uma quantidade enorme de banda a um custo muito pequeno. Tanto o ATM quanto o SONET/SDH não adicionam muito valor ao transporte de dados, porque foram projetados para serem usados no modelo da rede orientada a circuito, na qual a granulidade básica da banda era 64 Kbps e onde o transporte de dados não predominava. O valor do ATM e do SONET/SDH tende a diminuir na medida em que o IP passar a oferecer QoS e o WDM oferecer as características importantíssimas que a rede SONET/SDH hoje oferecem. É importante salientar que, ainda hoje, estas tecnologias continuam sendo implementadas e serviços estão sendo disponibilizados utilizando-se ATM, SONET/SDH no núcleo da Internet e nas redes metropolitanas devido aos níveis de granularidade, de proteção/ restauração e de capacidade que estas tecnologias oferecem.

O IP sobre WDM não é ainda uma solução *de facto*, porém o questionamento atual é quanto tempo demorará para que ele seja implementado em larga escala. Diretamente ligada a esta tendência, existe o GMPLS ("*Generalized Multiprotocol Label Switching"*), o qual está se confirmando como um protocolo de sinalização e controle comum para gerenciar/controlar o roteamento e a comutação nas fibras, nos comprimentos de onda, nos pacotes e até mesmo ao nível de bastidores ("*slots"*). Enquanto o GMPLS não evolui, as redes SONET/SDH e o tráfego IP irão coexistir com as redes WDM em comprimentos de onda, de tal maneira que estes comprimentos de onda possam ser adicionados ou retirados dos nós das redes ópticas.

O IP sobre WDM necessita vencer os desafios existentes hoje na integração dos protocolos. Um dos desafios existentes hoje é desenvolver técnicas para o roteamento e mapeamento dos comprimentos de onda ("*wavelength"*) a serem construídos em caminhos de luz ("*ligthpath"*) em um backbone óptico. Acredita-se que algum tipo de protocolo baseado nas medidas e modelo do OSPF ("*Open Shortest Path First"*) possa ser necessário. Um outro ponto está relacionado ao mapeamento e roteamento dos caminhos de luz em caso de ruptura. É possível simular a confiabilidade e proteção comutada existente hoje nas redes SONET/SDH em um caminho de luz? Estes são os maiores desafios que as redes WDM deverão vencer para ganhar a confiança das operadoras que estão satisfeitas e acostumadas ao modelo de proteção e robustez das redes SONET/SDH. Outro grande desafio é como acessar os canais. As redes SONET/SDH podem ser usadas para acessar canais, mas o objetivo principal é minimizar as conversões O-E (Óptica – Elétrica) e E-O (Elétrica-Óptica). Existem inúmeras

propostas para a criação de uma interface de controle de acesso ao meio WDM MAC. A abrangência da transparência óptica também não está clara, em relação às necessidades de conversão O-E e E-O. Uma coisa no entanto está clara: qual seja o custo da conversão e a regeneração/amplificação do sinal devido a perdas de potência acrescenta, de forma significativa, custos na rede e reduz de uma forma geral o desempenho.

Outro ponto questionável está relacionado à gerência ou não da banda passante. Seria necessário gerenciar a banda passante, a camada óptica e a camada WDM, ou estas camadas poderiam ficar sem gerência devido à imensa capacidade de banda disponível e o custo inexpressivo por bit? Para responder , ou tentar responder a essas questões, um número de especificações de interoperabilidade tem sido desenvolvidas e implementadas pelo Optical Internetworking Forum (OIF), o Internet Engineering Task Force (IETF) e o International Telecommunications Union (ITU).

O restante deste trabalho irá apresentar detalhes do que foi apresentado nesta introdução, mostrando tendências e o que se tem conseguido por em prática.

CAPÍTULO 2

CONCEITOS DE PROJETOS DE PROTOCOLS: TCP/IP E A CAMADA DE REDES

A idéia central de um Projeto de Protocolos de rede é que o mesmo seja feito ou desenvolvido em camadas ou níveis visando estruturar e definir melhor as funções e atividades de cada um dos níveis ou camadas. O princípio guia do protocolo Internet é o modelo fim-a-fim. Como podemos ver, há uma aparente discrepância entre o modelo adotado para o desenvolvimento de protocolos de rede através de camada e o protocolo "*de facto"* dominante na atualidade. Neste capítulo revisar-se-ão estes modelos e descrever-se-á o funcionamento das camadas de rede e de transporte na pilha do protocolo Internet. Iremos iniciar com uma visão geral sobre os princípios de protocolos, modelo de camadas e modelo fim-a-fim. Abordar-se-á, também, em um certo nível de detalhes o protocolo de transporte TCP, o protocolo de rede IP e os algoritmos de roteamento. Em seguida descrever-se-á a comutação de rede IP, MPLS ("*Multi Protocol Label Switching"*), arquiteturas de QoS ("*Quality of Services"*) na Internet utilizando-se o IntServ e DiffServ. O propósito é dar uma visão geral contextualizada do modelo de redes de comunicações interligadas ("*internetworking"*) dentro da qual as redes de comunicações óticas se encaixam.

Protocolos são complexos "pedaços" distribuídos de software. Uma das técnicas padrão utilizadas pelos projetistas de software para lidar com a complexidade é a abstração e o projeto modular. Como abstração entende-se um subconjunto de funções cuidadosamente escolhidas e projetadas para atuar como uma caixa preta ou um módulo.

O módulo possui uma interface descrevendo o seu comportamento na entrada e na saída; pode ser construído e mantido por entidades diferentes e é utilizado em um projeto de software como blocos construtores. A interface tem vida independente da implementação do módulo no sentido de que a tecnologia usada na interface pode ser alterada, contudo a interface tende a permanecer constante.

Os protocolos têm uma outra restrição importantíssima que é a necessidade de serem distribuídos. No entanto, eles têm que se comunicar entre si e com outros módulos que podem estar fisicamente muito distantes e isto é feito através das interfaces. Essas interfaces são denominadas interfaces par-a-par (*"peer-to-peer interfaces*"). Considerando-se que os protocolos tendem a ser uma seqüência de funções e os mesmos são organizados de uma forma seqüencial chamada de camadas (*"layers"*). O conjunto de módulos organizado em camadas é também conhecido como uma pilha de protocolos ("*protocol stack*") (Figura 2.1)

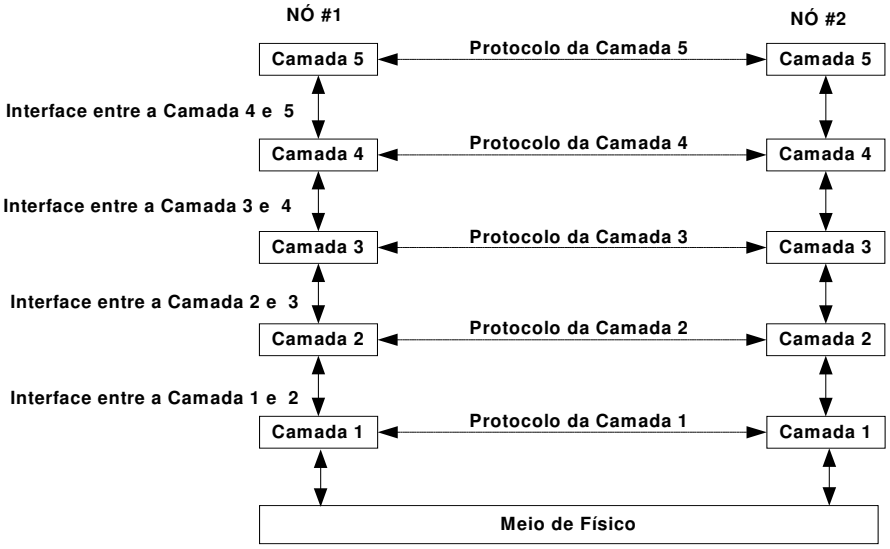


Figura 2.1 Modelo de Pilha de Protocolo [12].

Com o passar dos anos alguns modelos de camadas têm se tornado padrão, independentemente se são os mais adequados ou não, o fato é que existe a necessidade de se criar padrões para que comunicações possam ser feitas. Dois modelos se formaram como padrões: o modelo de sete (7) camadas desenvolvido por um conjunto de comitês sob a liderança do *International Standard Organization (ISO)* chamado de *Open Software Interconnection (OSI)* e o modelo de quatro (4) camadas TCP/IP (*Transport Control Protocol/Internet Protocol*) que se tornou o padrão *de facto* da Internet. Esses modelos estão representados na Figura 2.2.

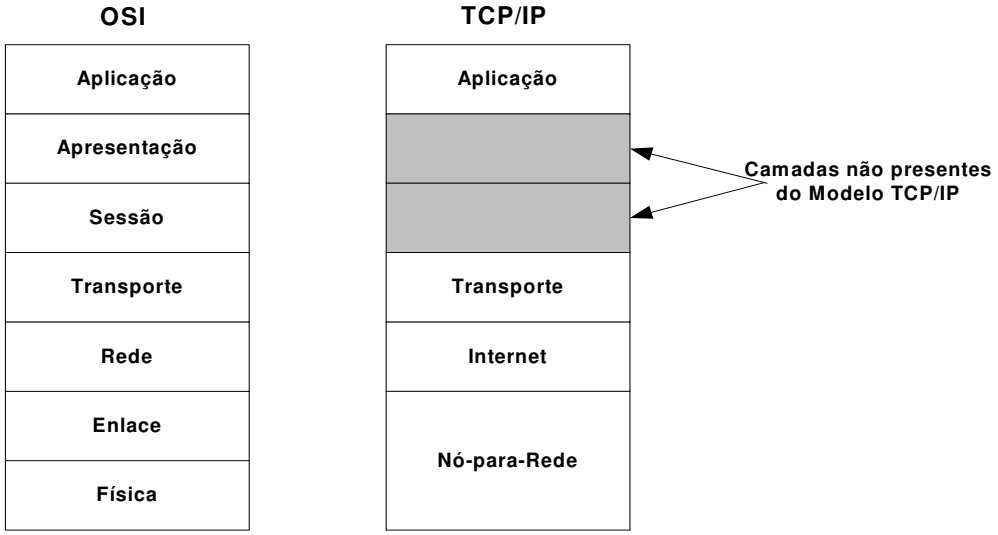


Figura 2.2 Modelo ISO/OSI e TCP/IP [12].

A camada física (*"physical layer"*) tem como função principal tornar a transmissão de bits, que trafega em um meio físico (fibra, cobre, cabo coaxial, ar etc.),

viável. A camada de enlace ("*link layer*") converte essa transmissão de bits em transmissão de quadros. Os quadros são usados para permitir a multiplexação de vários fluxos de dados e definem a unidade de transmissão usada na detecção de erro e no controle de fluxo. Na eventualidade do enlace ser um meio compartilhado, a camada de enlace define também o protocolo de controle de acesso ao meio. (MAC). A camada de rede (*"network* layer") trata da transmissão dos pacotes, os quais vieram dos quadros, através de múltiplos enlaces do nó transmissor para o nó destino. As funções nesta camada incluem roteamento, sinalização e mecanismos que lidam com camadas de enlaces heterogêneas em cada enlace.

A camada de transporte ("*transport layer*") provê serviços de comunicação fim-a-fim, ou seja, permite que aplicações sejam incorporadas no serviço de rede podendo adicionar outras capacidades como controle de conexão, confiabilidade e controle de congestionamento e de fluxo. Alguns exemplos de abstração na comunicação providos pela camada de transporte são o serviço de fluxo de bits confiável (TCP) e o serviço de datagrama não confiável (UDP). Essas abstrações são possíveis através da utilização de Interfaces de Programação de Aplicações (*"Applications Program Interfaces"*) como os programas conhecidos como a interface socket disponibilizada através do BSD ("*Berkeley Software Development*"). As camadas de aplicações (*"applications layers – session, presentation and applications")* usam as abstrações providas pela camada de transporte para criar as bases para aplicações como e-mail (através do protocolo SMTP), Web (através do protocolo HTTP), transferência de arquivos (através do protocolo FTP), teleconferência (através do protocolo H.323) e aplicações ponta-a-ponto (através do protocolo SIP).

2.1 Protocolo Internet: a origem do modelo fim-a-fim

O projeto do Protocolo Internet teve sua origem baseada no modelo fim-a-fim ("*end-to-end"*) [2], o qual serviu como guia e balizador no desenvolvimento das funcionalidades de um sistema distribuído complexo, no qual se insere o Protocolo Internet. Este princípio sugere que "… as funções inseridas nas camadas mais baixas podem ser redundantes ou de pouco valor quando comparadas aos custos (esforços) em mantê-las e provê-las nas camadas mais baixas...". Em outras palavras, pode-se dizer que cada sistema (ou subsistema) deve considerar apenas as funções que ele mesmo pode executar de forma completa e correta, deixando que as outras funções sejam direcionadas para outros sistemas, de tal maneira que estas funções também possam ser executadas de forma correta e completa.

No contexto da Internet, significa que várias funções, tais como confiabilidade, controle de congestionamento e gerência de conexão e sessão são direcionadas para os sistemas finais (i.e, executam o princípio básico do modelo fim-a-fim) e a camada de rede pode se concentrar nas funções que ela deve implementar totalmente (i.e., roteamento e entrega de datagramas). Como resultado, os aspectos de entrega de pacotes feitos pela rede são mantidos simples enquanto os sistemas destinatários são inteligentes e têm o controle da comunicação. Pode-se perceber que este modelo é diametralmente oposto ao modelo utilizado nas redes de telefonia, nas quais todos os sistemas são robustos, executando todas as funções localmente, e ainda têm que manter uma rede inteligente para a comunicação entre os sistemas.

O modelo do protocolo fim-a-fim argumenta que mesmo que a camada de rede ofereça funções em questão (i.e., gerenciamento da conexão e confiabilidade), as camadas de transportes teriam que adicionar confiabilidade para uma melhor interação entre as camadas de transportes entre sistemas ou caso a camada de transporte necessite de mais confiabilidade do que a camada de rede possa oferecer.

É importante salientar que o modelo fim-a-fim enfatiza que as funções têm que , *vis-à-*vis, oferecer um grau de correção, têm que ser completas e orientadas a custo. Este argumento do modelo diz que "algumas vezes uma versão incompleta de uma função provida pelos sistemas de comunicação pode ser útil para uma sensível melhoria no desempenho...". É possível notar claramente que este princípio leva em conta a relação custo-benefício e incorpora conceitos econômicos no projeto de protocolos. É desnecessário dizer que o conceito de "versões incompletas de funções" deve ser tratado com prudência.

Um problema no tocante às funções de camada de rede incompleta é a manutenção do controle de estado da rede. Como se sabe, a falta de controle de estado na rede impossibilita que qualquer nó da rede possa notificar os nós destinatários se as conexões estão ativas e/ou foram derrubadas ou estão em algum estado desconhecido ("*zoombie"*). Além disso, os nós destinatários não têm que conhecer, e nem precisam, qualquer dos componentes da rede além do seu destino, o primeiro nó roteador e um nó opcional para resolução de nomes (DNS – "*Domain Name Service"*), pois a integridade do pacote é preservada através da rede e os verificadores (*"checksums"*) de transporte e qualquer outra função de segurança são válido fim-a-fim. Se o controle de estado fosse inserido apenas nos nós destinatários, forçando desta maneira que os estados pudessem ser perdidos apenas quando o nó estivesse fora do ar, isso implicaria em um aumento significativo do tamanho da rede e a probabilidade de falhas de componentes da rede afetar a conexão, iria aumentar substancialmente. Por exemplo, se há uma perda de comunicação, porque a chave de estado se perdeu, a rede se torna frágil e sua utilização se degrada muito rapidamente. Entretanto se um nó destinatário falha, não existe sequer a esperança de um restabelecimento de conexão com o mesmo. Apesar disso, o modelo fim-a-fim define que apenas os nós destinatários é que devem armazenar as chaves de estado crítico.

No caso de um ISP que provê Qualidade de Serviço (*QoS*) e cobra pela mesma, a confiança e robustez da rede devem ser colocadas na própria rede e não apenas nos nós destinatários. Com isso, pode-se dizer que uma parte da rede tem que participar nas decisões do uso em comum de recursos e da bilhetagem, os quais não podem ser endereçados apenas aos nós destinatários. A rede não deve ser e ter uma visão cartesiana sobre isso, a melhor aplicação para o princípio fim-a-fim é o cenário no qual as aplicações poderiam estar participando do processo de decisões na rede, mas o controle da rede pertence a ela mesma, não aos nós destinatários.

O modelo fim-a-fim tem sido o guia para a inserção de várias funções de controle na Internet e permanece relevante ainda hoje, apesar das decisões de projetos de redes estarem casadas com decisões econômicas complexas relativas aos interesses de grandes grupos controladores de ISP e também dos fabricantes de equipamentos. Um recente estudo analisa o modelo fim-a-fim no contexto do projeto de interconexão de redes dentro desta complexidade econômica [3].

2.2 A Camada de Transporte: os protocolos UDP e TCP

A camada de transporte provê um mecanismo de comunicação fim-a-fim entre aplicações que são executadas em diferentes nós. Ela permite que diferentes aplicações se comuniquem sobre redes heterogêneas sem se preocupar com as diferentes interfaces de redes existentes, as tecnologias utilizadas etc, e isola as aplicações dos detalhes das características da rede. A camada de transporte também oferece serviços específicos para as camadas superiores. Existe uma diferença sutil entre os serviços oferecidos pela camada de transporte e o protocolo que permite que esses serviços possam ser oferecidos. Por serviços entende-se como sendo um conjunto de funções que são oferecidas às camadas superiores pela camada de transporte e o protocolo se refere aos detalhes de como o nó transmissor e o nó receptor interagem para poder prover esses serviços.

Existem dois serviços básicos oferecidos pela camada de transporte: serviço orientado à conexão ("*connection oriented service"*) e o serviço sem conexão ("*conectionless service"*). O serviço orientado à conexão é um serviço mais lento, porém altamente confiável e sem limite de tamanho das mensagens. Neste serviço, a camada de transporte é responsável pela entrega ordenada das mensagens à camada de aplicação. É função também desta camada o controle do fluxo de dados entre dois processos comunicantes.

O serviço orientado à conexão possui três (3) fases de operação: o estabelecimento da conexão, a transferência de dados e o término da conexão. Quando um sistema final ("*end-system"*) deseja enviar dados para outro sistema final através de um serviço orientado à conexão, a camada de transporte do sistema final emissor, de forma explícita, faz uma conexão com a camada de transporte do sistema final receptor. Uma vez que esta conexão esteja estabelecida (fase 1), os sistemas finais trocam informações e dados (fase 2) e ao final da transferência, a camada de transporte do sistema final emissor pede à camada de transporte do sistema final receptor que termine a conexão (fase 3). Existem duas (2) variações: serviço orientado a mensagens (*"message-oriented services*") e serviço de fluxo de bytes (*"byte-stream services*"). No serviço orientado a mensagem, a mensagem enviada pelo nó emissor tem um tamanho máximo e os limites da mensagem são preservados. Por exemplo, se um sistema final emissor tem 2 mensagens de 1 Kbyte de dados cada uma, eles são entregues como sendo dois (2) canais de transmissão distintos; a camada de transporte não irá combiná-los em uma mensagem de 2 Kbytes, nem em quatro (4) mensagens de 500 bytes. No serviço orientado a fluxo de bytes, os dados transferidos do sistema final emissor são vistos como uma seqüência não estruturada de bytes, os quais são transmitidos na mesma ordem na qual eles chegam. O dado neste caso não é tratado como mensagem e qualquer dado enviado à camada de transporte é inserido no final do fluxo de bytes. Um exemplo de serviço orientado à conexão é o TCP ("*Transmission Control Protocol*").

O serviço sem conexão, por outro lado, possui apenas uma fase de operação (transferência de dados) e não existe nenhuma fase de estabelecimento e nem de término de conexão.Todos os dados a serem transmitidos são entregues diretamente à camada de transporte, na qual um serviço orientado a mensagem é utilizado para transferir os dados.Um exemplo deste serviço é o UDP ("*User Datagram Protocol"*).

É importante, antes de prosseguirmos na breve discussão sobre a camada de transporte, falarmos sobre o conceito de serviço confiável e não confiável. A

transferência de dados confiável envolve quatro (4) diferentes características: devem ter notificação de perdas (*"no loss"*), devem ser sem duplicações ("*no duplicates*"), ordenadas ("*ordered form*") e devem manter a integridade dos dados ("*data integrity"*).

O serviço sem perdas garante que todo o dado entregue ao nó receptor ou ao nó emissor será notificado caso haja alguma perda de dados durante a transmissão. Isto assegura que o nó emissor não tenha dúvidas se um dado foi entregue ao nó receptor ou não. Esta característica está presente no protocolo TCP, mas não no protocolo UDP.

O serviço sem duplicações garante que todo o dado que tenha sido entregue à camada de transporte do nó emissor seja entregue ao nó receptor no máximo uma vez. Qualquer pacote duplicado que chegar ao nó receptor é descartado. Esta característica está presente apenas no protocolo TCP, mas não no UDP.

O serviço de transporte ordenado garante que a entrega dos pacotes ao nó receptor seja feita na mesma ordem na qual os pacotes saíram do nó emissor. Caso algum pacote chegue fora de ordem devido a algum retardo existente na rede de transmissão, o nó receptor ordena os pacotes primeiro para depois entregá-los às camadas superiores. Esta característica está presente apenas no protocolo TCP, mas não no UDP.

O serviço de integridade dos dados tem como função garantir que os bits dos dados entregues ao nó receptor sejam idênticos aos bits de dados enviados pelo nó emissor. Esta característica está presente tanto no protocolo TCP, quanto no protocolo UDP.

É importante notar que as quatro características de um serviço confiável são ortogonais, no sentido de que a presença de uma das características não implica na presença das outras três.

Um outro conjunto crítico de serviços que têm que ser providos pela camada de transporte é a multiplexação e a demultiplexação das aplicações. Esta característica permite que múltiplas aplicações usem a rede simultaneamente e garante que a camada de transporte possa diferenciar os dados recebidos das camadas inferiores de acordo com a aplicação e o processo a que o dado pertença. A funcionalidade na camada de transporte do receptor que entrega os dados corretamente à aplicação à qual o mesmo pertence é chamado de demultiplexação. O processo no emissor onde a informação sobre os vários processos ativos é coletada é chamado de multiplexação.

Ambos os protocolos possuem multiplexação e demultiplexação utilizando dois campos especiais no cabeçalho do segmento de dado: o número da porta origem ("*source port number field*") e o número da porta destino ("*destination port number field"*). Toda aplicação que é executada em um nó possui um único número de porta o qual deve estar situado entre 0 e 65535. A identificação biunívoca de um processo que está sendo executado em um nó é feita quando pegamos os números da porta origem e da porta destino. (Figura 2.3)

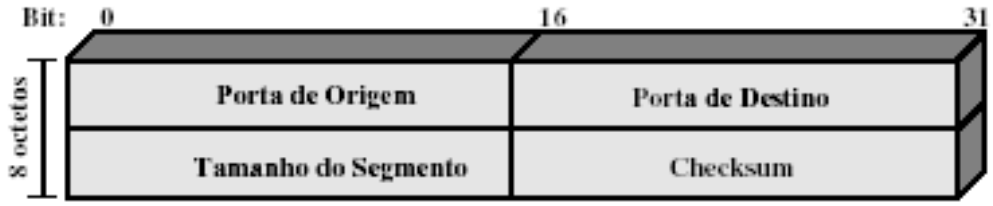


Figura 2.3 Cabeçalho do Protocolo UDP [46].

2.3 UDP ("User Datagram Protocol") – Transporte não orientado à conexão

O UDP é o protocolo mais básico da camada de transporte e as únicas características que possui são os serviços de multiplexação/demultiplexação e integridade dos dados. Por ser um serviço não orientado à conexão, uma transferência de dados feita via UDP é leve e objetiva no sentido de que recebe as mensagens vindas das camadas superiores, insere os números das portas dos nós emissor e destinatário junto com alguns outros campos do cabeçalho e envia o segmento montado para a camada de rede.

Apesar dele não ser um protocolo que ofereça garantia de entrega ele é muito útil em várias aplicações por ser um protocolo leve. Primeiramente, a ausência do estabelecimento da fase de conexão permite que o UDP inicie a transmissão sem nenhum retardo. Isto o torna ideal para aplicações que transferem uma quantidade pequena de dados, como no caso do DNS (*"Domain Name Service"*) no qual o retardo para o estabelecimento da conexão poderia ser significativamente maior do que a própria transferência dos dados (é claro que em algumas situações o DNS se utiliza também do TCP, como, por exemplo, quando da transferência e sincronização das bases de dados, na qual há necessidade de uma transmissão confiável e com garantia). Diferentemente do TCP, o UDP por não manter nenhuma informação relativa aos parâmetros de estado da conexão como os buffers de envio e recebimento de dados, os parâmetros de controle de congestionamento e dos números de seqüência e de reconhecimento, faz com que o UDP possa suportar mais clientes ativos para uma mesma aplicação que o TCP. Um outro ponto interessante é que enquanto o TCP tenta adequar sua taxa de transmissão de acordo com o seu mecanismo de controle de congestionamento na presença de perdas de pacotes da rede, a taxa que o UDP transmite é limitada apenas pela velocidade que ele pega/recebe as informações da aplicação. A taxa de controle de fluxo no TCP pode ter um impacto significante na qualidade de aplicações de tempo real, enquanto o UDP evita este tipo de problema, razão esta pela qual é utilizado em muitas aplicações multimídia, tais como, Telefonia via Internet, fluxo de vídeo e áudio e conferência.

Vale ressaltar, no entanto, que a falta de reação do UDP ao congestionamento existente na rede não é uma característica desejável. O controle de congestionamento é importante para que a rede se proteja de um estado no qual pouquíssimos pacotes chegam ao seu destino, fazendo com que a rede possa entrar em colapso. Para evitar que isto aconteça, alguns pesquisadores propuseram mecanismos de controle para o UDP [4, 5]. Algo aparentemente contraditório é que apesar do UDP não prover confiabilidade na entrega de pacotes, a aplicação propriamente dita pode ter/oferecer esta característica. Como exemplo disso citamos algumas aplicações de fluxo ("*streaming*").

Além da estrutura de campos do protocolo UDP mostrada na Figura 2.3, o cabeçalho do UDP tem dois campos adicionais: o tamanho ("*length*") e o verificador ("*checksum*"). O campo tamanho especifica o tamanho em bytes do segmento UDP incluindo os cabeçalhos. O verificador é usado para que o nó receptor verifique se houve algum bit corrompido durante a transmissão e abrange tanto os cabeçalhos, quanto os dados.

O verificador é calculado no nó emissor através do complemento de 1 da soma de todas as palavras de 16 bits no segmento. O resultado deste cálculo é um número de 16 bits que é inserido no campo verificador no segmento UDP. Quando chega no nó receptor,

o verificador é adicionado a todas as palavras de 16 bits no segmento. Caso nenhum erro tenha sido introduzido durante a transmissão dos dados, o resultado desta operação deveria ser 1111111111111111; no entanto se algum erro for encontrado, o segmento todo é descartado e nenhuma mensagem de erro é enviada ao nó transmissor.

2.4 TCP ("Transport Control Protocol") – Transporte orientado à conexão

Apesar do TCP prover multiplexação/demultiplexação e detecção de erros de uma maneira similar à utilizada pelo UDP, a diferença fundamental entre os dois protocolos está no fato do TCP ser orientado à conexão e prover mecanismos que o torna confiável. Antes de um nó começar a transmitir dados via TCP para outro nó, primeiramente há a necessidade de se configurar a conexão através de um mecanismo de "*handshaking*" e ambos os nós iniciam várias variáveis de estado associadas com a conexão TCP.

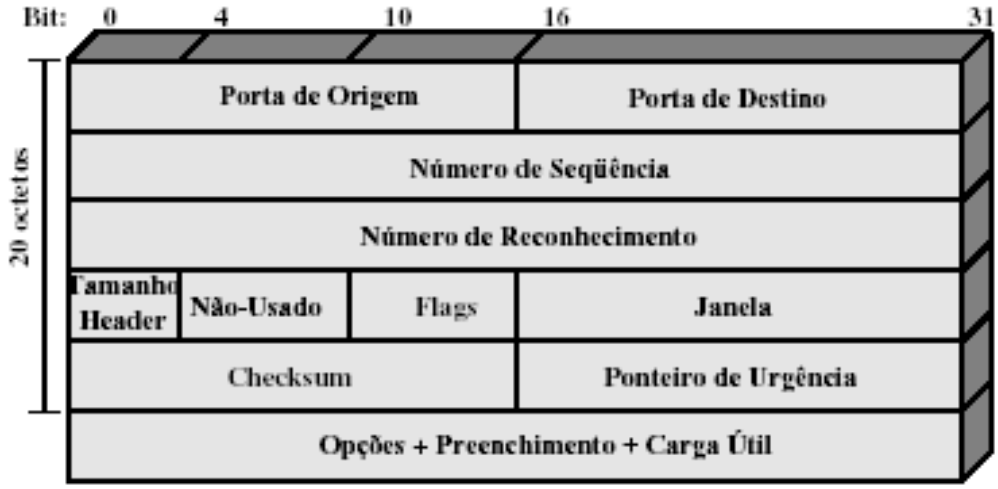


Figura 2.4 Estrutura do TCP [46].

Na Figura 2.4 pode-se ver a estrutura do segmento do TCP. Comparando esta estrutura com a estrutura do UDP, nota-se que enquanto o UDP possui 8 bytes de cabeçalho, o TCP possui 20 bytes As portas origem e destino são similares ao UDP e possuem 2 bytes cada em ambos os protocolos. Os campos número de seqüência ("*sequence number*") e número de reconhecimento/confirmação ("*acknowledgement number*") são utilizados pelo TCP emissor e pelo TCP receptor para implementar serviços de transporte confiável O número de seqüência identifica o número dentro do fluxo de bytes do primeiro byte de dados do segmento e é um número sem sinal de 32 bits que retorna ao 0 depois de ter chegado ao número 2^{32} -1. Como o TCP é um protocolo de fluxo de dados e cada byte é numerado, o número de reconhecimento ("*acknowledgement number*") representa o próximo número de seqüência que está sendo esperado pelo receptor. O número de seqüência é iniciado durante o estabelecimento da conexão e é chamado de número de seqüência inicial ("*initial sequence number").* O campo de 4 bits chamado de tamanho do cabeçalho ("*header length"*) informa o tamanho do cabeçalho em palavras de 32 bits. O uso de apenas 4 bits limita o tamanho máximo do cabeçalho do TCP em 60 bytes. Os próximos 6 bits do cabeçalho são reservados para uso futuro e os outros 6 bits são chamados de campos de sinalização ("*flag field"*). O bit

ACK é usado para sinalizar que o valor existente no campo de reconhecimento/confirmação é válido. O bit SYN é ligado para indicar que o emissor do segmento deseja iniciar uma conexão e o bit FYN é ligado para indicar que o emissor terminou de enviar os seus dados. O bit RST é ligado para indicar que o TCP emissor quer cancelar a conexão e o bit PSH indica que o TCP receptor deverá enviar os dados imediatamente para as camadas superiores. Finalmente, o bit URG é ligado para indicar que existe alguma informação que as camadas superiores do TCP emissor classificou como urgente. A localização do último byte do dado urgente é indicada no campo ponteiro de urgência ("*urgent pointer"*). O conteúdo deste campo só é válido se o bit URG estiver ligado. O campo tamanho da janela receptora ("receiver *window size"*) é usado para o controle de fluxo e indica o número de bytes que o TCP receptor está esperando receber. Este campo tem a função de limitar a taxa de transmissão do TCP emissor para evitar um estrangulamento do TCP receptor. O campo de verificação ("*checksum"*) é similar ao utilizado no UDP e é usado para preservar a integridade do segmento.

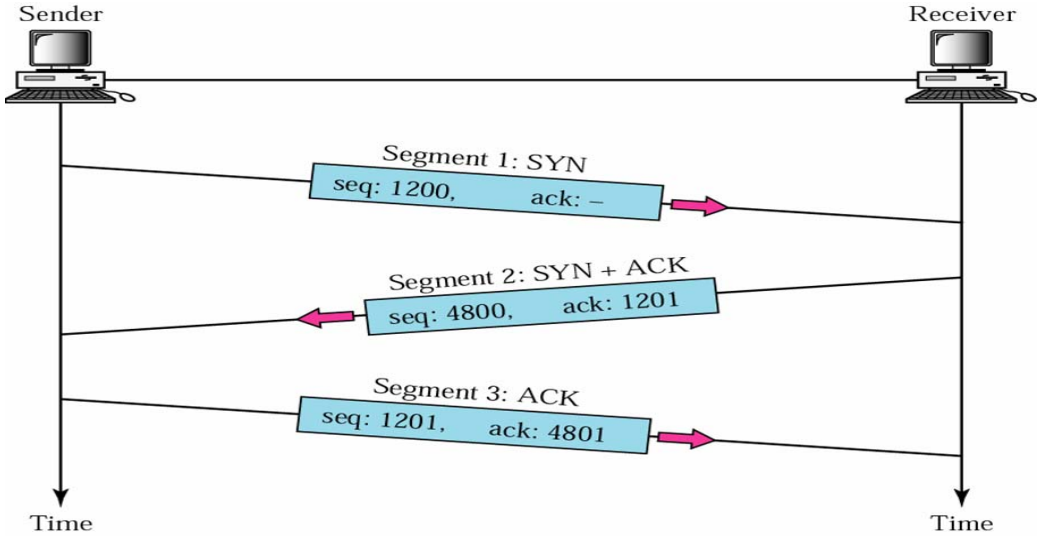


Figura 2.5 Three-Way Handshake (McGraw-Hill 2000)

Uma conexão TCP é estabelecida através de um protocolo chamado aperto de mão a três vias ("*Three-way handshake*"). Na Figura 2.5, mostra-se o processo do "*three-way handshake"*. No primeiro passo para uma conexão ser estabelecida, o cliente (TCP emissor) envia um segmento SYN especificando o número da porta do servidor (TCP receptor) com o qual ele deseja se conectar junto com o seu número inicial de seqüência. No segundo passo, o servidor então responde à mensagem enviando o seu próprio segmento SYN contendo o número inicial de seqüência do servidor juntamente com o reconhecimento do SYN enviado pelo cliente adicionando o valor 1 ao número inicial de seqüência que o cliente enviou. No terceiro passo, o cliente envia um reconhecimento do SYN enviado pelo servidor adicionando o valor 1 ao número inicial de seqüência enviado pelo servidor. A troca deste três segmentos constitui o processo de estabelecimento de uma conexão TCP.

Enquanto que na fase de conexão do TCP envolve a troca de três pacotes, na fase de término de conexão são necessários quatro pacotes. Isto acontece porque o TCP é "*full*

duplex" (os dados podem trafegar nos canais de transmissão e recepção independentemente). Quando o cliente quer terminar a conexão, ele envia um pacote FIN indicando que ele quer terminar a conexão. O servidor então informa a aplicação que não existem mais dados para chegar e envia ao cliente um pacote de ACK acrescentando o valor 1 ao número de seqüência recebido. O recebimento de um pacote FIN significa simplesmente que não há mais nenhum dado a ser transmitido nesta direção. O TCP pode continuar enviando dados na outra direção apesar de ter recebido um pacote FIN. Uma vez que o outro lado terminou de transmitir os seus dados, ele também envia um pacote FIN para o outro lado, o qual responde também com um ACK, formalmente terminando a conexão em ambas as direções. Na linguagem do TCP, o nó que envia o primeiro FIN executa um término ativo ("*active close"*) e o outro nó executa um término passivo ("*passive close"*).

Apesar do fato do TCP ser orientado à conexão, a principal diferença entre ele e o UDP é a questão da entrega confiável. Para entendermos como o TCP gerencia a questão, é preciso antes analisar em profundidade como atuam os campos número de seqüência e número de reconhecimento.

O TCP quebra os dados enviados da camada de aplicação em tamanho de segmento que ele considera ser o ideal. Devemos observar que este mecanismo também é diferente do UDP no qual o tamanho do datagrama é determinado pela aplicação. Os dados enviados pelas camadas superiores são vistos como um fluxo ordenado de bytes e os números de seqüência são utilizados para contar o número de bytes que ele transmite ao invés do número de segmentos. Como já mencionado, o número de seqüência de um segmento TCP é o número do fluxo de byte do primeiro byte do segmento. Enquanto envia os dados que recebe das camadas superiores, o TCP conta o número de seqüência em termos de bytes. No recebimento de um segmento o TCP envia um pacote de reconhecimento informando ao emissor sobre a quantidade de dados que ele recebeu corretamente e o número de seqüência do próximo byte de dados que ele está esperando. Como exemplo, se o receptor recebeu um segmento com os bytes 0 a 1023, ele enviará uma mensagem ao emissor dizendo que espera receber agora o byte 1024. Portanto o segmento que o receptor envia ao emissor contém o byte 1024 no campo de reconhecimento. Por exemplo, suponha que o receptor tenha recebido os bytes 0 a 1023 e que agora ele esteja esperando o segmento que contenha os bytes 1024 a 2047. No entanto, o segmento que ele recebeu do emissor contém os bytes 2048 a 3071. Isto pode ter acontecido devido à perda do segmento que contém os bytes 1024 a 2047. Como o receptor está esperando o segmento que contém os bytes 1024 a 2047, ele continua enviando ao emissor, no campo de reconhecimento, a informação de que ele está esperando o segmento começando com o byte 1024. Pelo fato do TCP reconhecer apenas até o primeiro byte perdido do fluxo, diz-se que ele provê reconhecimento cumulativo ("*cumulative acknowledgement"*). O TCP pode tomar duas atitudes quando recebe segmentos fora de ordem: a primeira seria simplesmente descartar o segmento e a segunda seria mantê-lo em uma fila de espera/armazenamento aguardando o recebimento do segmento faltante antes de entregar os dados para as camadas superiores (normalmente este é o padrão utilizado). Como o TCP opera em um modo *"full duplex",* os segmentos de reconhecimento podem ser enviados tanto no canal cliente-servidor, quanto no canal servidor-cliente, não necessitando nenhuma outra abertura de canal para

que esses segmentos sejam enviados. Este mecanismo é conhecido como reconhecimento "*piggybacked"*.

Quando o TCP envia um segmento, ele mantém um contador de tempo (temporizador) e espera que o receptor envie um reconhecimento de que recebeu o pacote enviado. Este reconhecimento não necessariamente é enviando imediatamente; normalmente existe um retardo de uma fração de segundos antes dele ser enviado. Caso nenhum reconhecimento tenha sido recebido pelo emissor antes do término do temporizador, o segmento é retransmitido. Este mecanismo é conhecido pelo nome de intervalo de tempo ("*timeout*") e é o primeiro mecanismo de recuperação de erro.

Um outro mecanismo de recuperação de erro é através dos reconhecimentos duplicados ("*duplicate acknowledgement").* Isto acontece devido ao mecanismo existente no TCP chamado de reconhecimento cumulativo. O reconhecimento duplicado ocorre para aqueles segmentos os quais recebem mais de um pacote de reconhecimento para um determinado segmento. Como regra geral, o emissor sabe que os dados foram entregues ao receptor através do primeiro pacote de reconhecimento recebido do receptor. Quando o TCP recebe um segmento com o número de seqüência maior do que o que era esperado, ele identifica uma falha no fluxo de dados (ou seja, a falta de um segmento). Como o TCP usa o mecanismo de reconhecimento cumulativo, ele não pode enviar um reconhecimento falso ou negativo para o emissor dizendo que o pacote se perdeu, ao invés disso o receptor envia um reconhecimento duplicado confirmando o último segmento recebido e reforçando que está esperando receber o próximo segmento. Caso o emissor receba três reconhecimentos duplicados para o mesmo segmento, ele reenvia o segmento, mesmo que o temporizador do referido segmento não tenha ainda sido zerado. Este mecanismo de recuperação é conhecido como retransmissão rápida ("*fast retransmit"*).

O TCP possui controles internos os quais permitem que ele, através da "percepção" da capacidade da rede atual, possa modular sua taxa de transmissão. Tendo em vista que o IP não oferece nenhuma informação confiável sobre o congestionamento da rede, o TCP utiliza os fluxos de pacotes de reconhecimentos e pacotes perdidos para inferir o congestionamento na rede e tomar ações preventivas.

O mecanismo de controle de congestionamento do TCP é baseado em se limitar o número de pacotes que podem ser transmitidos em um intervalo de tempo sem que o cliente tenha recebido pacotes de reconhecimentos do servidor. O número desses pacotes não-reconhecidos é chamado de tamanho de janela ("*window size"*). Quando o cliente TCP inicia a transmissão ele não tem idéia sobre o congestionamento da rede, por conta disso, ele inicia a transmissão de uma forma conservativa com um tamanho de janela pequeno pesquisando a rede para descobrir sua capacidade. Como o cliente controla os pacotes de reconhecimentos (ACK) que ele recebeu daqueles que foram enviados, ele incrementa o tamanho da janela e passa a enviar mais e mais pacotes para a rede até que o cliente não receba um reconhecimento indicando que nesta taxa de transmissão há perda de pacotes e, de forma prudente, reduz o tamanho da janela. Feito isso o cliente começa o processo de testar a disponibilidade da rede novamente.

Para implementar o mecanismo de controle de congestionamento, tanto o cliente quanto o servidor têm de manter e gerenciar algumas variáveis adicionais: a janela de congestionamento ("*congestion window*"), cwnd, e o valor limitante ("*threshold*"), ssthresh. A janela de congestionamento juntamente com o aviso enviado pelo servidor no

campo tamanho da janela do receptor ("*receiver window size*"), rwin, no cabeçalho do TCP serve para limitar a quantidade de informações não-reconhecidas que o cliente pode transmitir em um determinado período. A quantidade de informações não-reconhecidas deve ser menor que o mínimo de cwnd e rwin. O rwin é informado pelo servidor e é determinado por fatores locais ao mesmo, como por exemplo, velocidade do processador e tamanho disponível de armazenamento. As variações que afetam o cwnd são mais dinâmicas e se refletem nas condições da rede e nas perdas que o cliente experimenta. A variável ssthresh também controla a maneira através da qual o cwnd aumenta ou diminui. Como exemplo, suponha que um cliente TCP tenha recebido informações das camadas superiores e as tenha quebrado em segmentos para serem transmitidos ao servidor; desconsiderando-se a limitação de tamanho que é imposta através do campo rwin. O tamanho dos pacotes é determinado pelo tamanho máximo permitido para um segmento no meio e é chamado de tamanho máximo do segmento ("*maximum segment size*" – MSS). Sendo um protocolo orientado a fluxo de bytes, o TCP mantém o valor de cwnd em bytes e uma vez que a fase de estabelecimento é concretizada, ele começa a transmitir os dados com um valor inicial de cwnd igual a um MSS (valor inicial de cwnd = 1 MSS) e o ssthresh é iniciado com um valor default (65.535 bytes). Quando o pacote é transmitido, o cliente estabelece o tempo de retransmissão para o mesmo. Caso o cliente receba um reconhecimento para este segmento antes do término do tempo de retransmissão, o TCP "entende que a rede tem capacidade para transmitir mais do que 1 segmento sem perdas (sem congestionamento) e então incremente o cwnd em 1 MSS, fazendo com que o cwnd = 2 MSS. O cliente envia agora os dois segmentos e se chegarem os reconhecimentos antes do término do tempo de retransmissão, o cwnd incrementado de 2 MSS, resultando assim que o cwnd = 4 MSS. O cliente envia agora os 4 segmentos e caso receba os 4 reconhecimentos ele incrementa de 4 MSS o valor de cwnd, resultando que o cwnd = 8 MSS. Neste processo de aumento do cwnd de 1 MSS para cada pacote reconhecido é um processo exponencial e continua até que cwnd < ssthresh e os reconhecimentos continuem chegando antes do término do tempo de retransmissão dos pacotes. Este processo é chamado de partida lenta (*"slow start"*).

Quando o valor de cwnd é igual ao valor de ssthresh, a fase de partida lenta ("slow start") termina e ao invés da janela ser incrementada de 1 MSS para cada reconhecimento recebido, ela é agora incrementa de 1/cwnd. Conseqüentemente, o cwnd aumenta em 1 MSS apenas quando o reconhecimento de todos os pacotes ainda não reconhecidos tiver chegado. O resultado disso é uma curva linear ao invés da curva exponencial existente na fase da partida lenta (Figura 2.6). Esta fase é chamada de "*congestion avoidance"*.

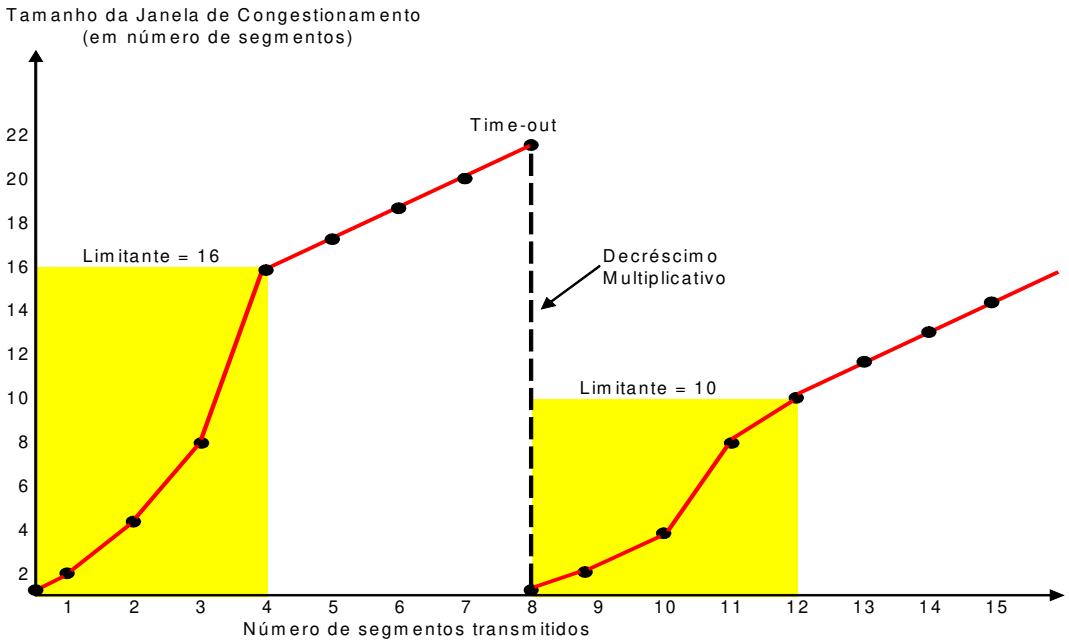


Figura 2.6 “Congestion Avoidance” (McGraw-Hill 2000)

Enquanto a janela aumenta, o cliente continua inserindo mais pacotes na rede e em algum momento haverá alguma perda. Quando isso ocorre, o valor de ssthresh é reduzido para o valor Max {cwnd/2, 2MSS} (ou seja, ele assume a metade do valor atual do cwnd ou 2 MSS o que for maior). Feito isso, o valor de cwnd é reduzido para um valor menor o qual dependerá da versão de TCP. No TCP Tahoe, quando uma perda for detectada, o cwnd é reduzido para 1 MSS e o servidor iniciava o processo de partida lenta. No TCP Reno, uma das versões mais recentes e mais utilizadas hoje, o valor do cwnd é tratado de forma diferenciada dependendo do fato gerador. Por exemplo, se houver o término do temporazidor o valor de cwnd é reduzido para 1 MSS e a partida lenta é iniciada (idêntico ao TCP Tahoe). Quando a perda tiver sido detectada devido ao recebimento de três (3) confirmações duplicadas, o TCP Reno faz uma retransmissão rápida (“*fast retransmit”*) e entra em um processo chamado de restauração rápida (“*fast recovery”*).Quando tiver uma retransmissão rápida, o TCP Reno reduz o valor do cwnd para a metade do valor corrente (i.e., cwnd/2) e ao invés de entrar na partida rápida ele vai diretamente para o “*congestion avoidance”*. Em ambos os modelos, a detecção de uma perda faz com que o valor do ssthresh seja reduzido para o valor Max {cwnd/2, 2MSS}. É importante notar que se o fluxo está na partida lenta, ele entra no “*congestion avoidance”* muito mais rápido que antes, pois o valor de ssthresh foi diminuído pela metade. Este mecanismo, na presença de perdas, é uma tentativa de modular o padrão de aumento da janela tendo como base o mecanismo do TCP que previne congestionamento (“*congestion avoidance”).* [7]

2.5 A Camada de Rede – Internet Protocol

A camada de rede da pilha TCP/IP trata de roteamento e de interconexões entre as redes. O problema principal na interconexão de redes é a heterogeneidade e a escala.

Heterogeneidade está relacionada às diferentes camadas de enlace existentes nas quais o IP é encapsulado e os detalhes específicos de endereçamento e envio necessários para que um pacote IP seja entregue no seu destino. Escala é a propriedade de um sistema crescer de acordo com a demanda sem que este restrinja o seu próprio crescimento.

2.5.1 O modelo de serviços de rede

Uma maneira de se lidar com a heterogeneidade é fornecer serviços de tradução entre entidades heterogêneas quando há necessidade de se usá-las para entregar algum pacote IP. As pontes e os roteadores multiprotocolos são um exemplo disso. Uma maneira de se fazer isso é através de um modelo chamado de modelo de sobreposição ("*overlay model*").

Historicamente e por necessidade, o IP ("*Internet* Protocol") tem que ser um protocolo simples de tal maneira que o mapeamento entre a camada IP e a camada inferior seja feito de forma simples. Deve-se lembrar que ele foi desenvolvido sob o comando dos militares americanos, tendo como visão a Guerra Fria. Para atingir os objetivos a que ele foi projetado, o IP usa o princípio do melhor esforço ("*best effort*"), que é um modelo de serviço de datagrama não confiável, no qual os datagramas são enviados entre a origem e o destino através de redes heterogêneas as quais podem estar distantes milhares de quilômetros. O IP espera um mínimo de enquadramento da camada inferior. O mapeamento do IP para as camadas inferiores envolve problemas de resolução de endereçamento, de fragmentação e remontagem dos pacotes. A experiência tem mostrado que este mapeamento tem sido feito diretamente em várias sub-redes, especialmente naquelas que não são muito grandes e que suportam difusão ("*broadcasting"*) ao nível da LAN ("*Local Area Network*"). Em redes de múltiplo acesso que não suportam difusão ("*NonBroadcast Muliple-Access" – NBMA*), a resolução de endereçamento tem sido um grande desafio. Um dos grandes desafios em redes NBMA está relacionado aos protocolos que estão associados ao IP, como por exemplo o roteamento via BGP ("*Border Gateway Protocol*") que em sub-redes de grande escala encontra problemas de mapeamento entre a camada IP e as camadas inferiores. Algumas tecnologias híbridas, como o MPLS, foram desenvolvidas e estão sendo usadas para minimizar este problema de mapeamento e também prover capacidade de Engenharia de Tráfego para estas redes.

Para várias aplicações, constatou-se que o serviço de melhor esforço ("*best effort*"), juntamente com os protocolos TCP, UDP e o RTP se mostraram muito adequados. Algumas aplicações que necessitam de requisitos específicos para poderem utilizar a rede IP, como por exemplo a telefonia, precisam ou se adaptar às restrições ou utilizar capacidade de QoS provida pela rede. Apesar de terem sido desenvolvidos inúmeros protocolos e mecanismos para oferecer QoS para a rede IP, uma Internet provendo QoS ainda pode ser considerado como algo além do "estado da arte".

2.5.2 O modelo de entrega do protocolo IP ("Forwarding Paradigm")

O núcleo do serviço da rede IP provê a entrega de datagramas através de redes heterogêneas, por si só, isto é um problema não trivial, mas que é feito "naturalmente" pela Internet. O resultado deste serviço de entrega de datagramas é basicamente prover conectividade. As duas grandes abordagens em conectividade são: a direta e a indireta. A

conectividade direta é aquela na qual o destino está em uma conexão física ponto-a-ponto (isto inclui meios de transmissão de uso compartilhado e não compartilhado). A conectividade indireta é aquela na qual o destino é alcançado através de componentes intermediários e de redes intermediárias também. Os componentes intermediários (pontes, comutadores, roteadores, equipamentos para tradução de endereços IP etc.) são dedicados a funções que lidam com problemas de escala e de heterogeneidade das redes. A Internet, na realidade, é uma grande rede de conectividade indireta.

Um fato conhecido e amplamente discutido na rede IP é o problema da escalabilidade. Um equívoco comum é o de se pensar que a escalabilidade de uma rede está relacionada ao número de nós existentes nela. Na realidade, o problema da escalabilidade de uma rede em relação a algum parâmetro (por exemplo o número de nós) está relacionada inversamente à eficiência das características da arquitetura relacionadas a este mesmo parâmetro. Por exemplo, as arquiteturas de conectividade direta não são escaláveis devido à limitação física existente no uso do meio compartilhado, no número de interfaces, ou no alto custo de aprovisionamento de uma conexão totalmente interligada ("*Full Mesh*"). Uma maneira de se lidar com isso é construir uma rede de comutadores, na qual os componentes intermediários ("*switches*") provêem capacidade de filtragem e de entrega para isolar redes distintas, mantendo com isso as características de cada uma das redes enquanto provê uma interconexão escalável. Em geral, quanto mais eficiente a capacidade dos componentes na filtragem e na entrega, mais escalável é a arquitetura da rede. Os hubs de camada 1, fazem apenas difusão ("broadcast"), não têm capacidade de filtragem, mas entregam sinais elétricos que trafegam no segmento. As pontes ("*bridges*") de camada 2 e os comutadores ("*switches*") podem fazer filtros através de tabelas de entrega ("*forwarding*") capturadas por meio de espionagem ("*snooping*"), porém o seu comportamento padrão é o de fazer uma inundação de dados ("*flooding*") através de uma árvore de espalhamento ("*spanning tree*") quando a tabela de entrega não possui o endereço do destinatário. É claro que este comportamento padrão não é nada eficiente e limita e muito a escalabilidade.

Em contraste com os componentes das camadas 1 e 2, os comutadores (roteadores) da camada 3 não fazem difusão através de sub-redes; eles se baseiam em protocolos de roteamento confiáveis e um conjunto local de decisões de transmissão para entregar os pacotes através da Internet. O endereçamento IP é hierárquico e o acesso aos mesmos é feito através de políticas de roteamento. Este modelo permite que os nós intermediários de uma rede consigam determinar de forma clara se o nó destino está diretamente conectado no segmento de rede ou não. No caso do nó destino estar conectado diretamente no mesmo segmento de rede, a Camada 2 é acionada para que a entrega seja feita ao nó destinatário; no caso do nó destino não estar diretamente conectado no mesmo segmento de rede, a Camada 3 toma uma decisão de entrega para o próximo elemento de transmissão ("next *hop*") e aí solicita a Camada 2 para fazer a entrega.

A heterogeneidade é suportada pelo IP porque ele precisa apenas de uma parte mínima do serviço de entrega da Camada 2. Antes de chamar este serviço da Camada 2, o roteador tem que determinar o endereço de Camada 2 do nó destinatário e tem que mapear o datagrama no formato do quadro de transmissão da Camada 2. Se o datagrama é muito grande para ser inserido dentro do formato de um quadro, ele precisa ser quebrado e depois precisa ser remontado novamente no destino final. Este é o segredo do

IP: ele não precisa de nenhuma característica especial dos protocolos da Camada 2 e por isso pode trabalhar com um número muito grande de protocolos de Camada 2.

De uma forma geral, o modelo de entrega do protocolo IP foge do conceito de conectividade direta e indireta. O segredo disso reside no fato de que o endereçamento é feito para tornar possível o acesso quer seja de forma direta ou indireta a um nó destino e aos protocolos de roteamento que ajudam a identificar o próximo elemento de transmissão ("next hop") apropriado caso o nó destino não pertença ao mesmo segmento de rede. A heterogeneidade trata dos problemas de mapeamento, os quais são simplificados devido aos requisitos mínimos que o IP espera das camadas inferiores (apenas o serviço de entrega dos quadros é requerido). Todos os outros detalhes são abstraídos pelo IP.

2.5.3 O formato do pacote IP

Nesta seção iremos analisar o formato do pacote IP na versão 4, como ilustra a Figura 2.7.

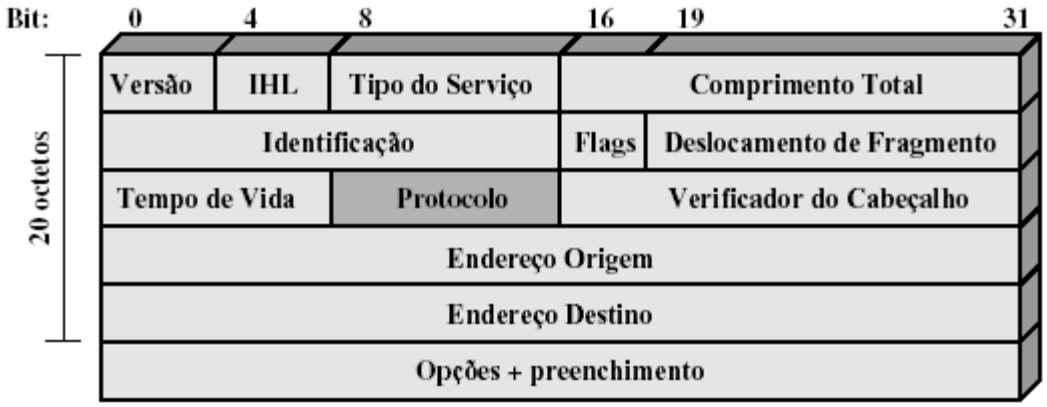


Fig. 2.7 – O Formato do pacote IP versão 4 [46].

Os maiores campos do cabeçalho são os endereços de origem e destino, cada um com 32 bits. A primeira linha de 32 bits contém os campos versão, IHL, tipo de serviço e comprimento total. O campo versão indica a versão do protocolo IP a que o pacote pertence e oferece com isso uma grande flexibilidade em termos de versões do protocolo IP. A versão atual do protocolo IP é a versão 4. A próxima versão do protocolo IP é a de número 6, possui 128 bits de endereçamento contra os 32 bits de endereçamento da versão 4, e deverá ser largamente utilizada até 2010. [S. KALYANARAMAN]. O campo IHL (tamanho do cabeçalho) é utilizado devido ao fato do campo opções poder variar de tamanho, porém este campo é muito pouco usado. O campo tipo de serviço foi projetado para dar suporte a serviços opcionais para entrega diferenciada, mas nunca foi utilizado. Recentemente o Grupo de Trabalho de Serviços Diferenciados do IETF ("*Internet Engineering Task Force*") renomeou este campo para DS byte ("*Differential Services - DiffServ*") para poder suportar serviços diferenciados ("*DiffServ*"). O campo comprimento indica o tamanho de todo o datagrama e é necessário porque o IP aceita

carga útil ("*payload"*) de tamanho variável. A segunda linha de 32 bits que contém os campos identificador, "flags" e deslocamento do fragmento, será descrito mais adiante. O campo verificador do cabeçalho, como o próprio nome diz cobre apenas o cabeçalho e não a carga útil ("*payload*") e é utilizado para detectar qualquer erro no cabeçalho evitando com isso erros de roteamento. A detecção de erros na carga útil do pacote é de responsabilidade da camada de transporte (ambos os protocolos TCP e UDP fazem isso). O campo protocolo permite ao IP fazer a demultiplexação do datagrama e entregá-lo a um protocolo de mais alto nível para ser tratado.

O campo tempo de vida é decrementado em cada roteador através do qual o pacote IP passa. Quando este contador tiver valor 0 o pacote é descartado; com isso previne-se que pacotes fiquem transitando "eternamente" na Internet. Este campo é também usado para determinar o escopo dos pacotes e pode ser utilizado em conjunto com outros protocolos como o ICMP e mutlicast para auxiliar em funções administrativas de rede.

2.5.4 O endereçamento IP e a Alocação dos endereços

Um endereço IP é um identificador único existente nas redes de computadores que falam o protocolo TCP/IP. Um endereço IP deve ter um significado válido na localização fonte, mas pode variar se o destino se desloca. Endereços de tamanho fixo podem ser processados de forma mais rápida. O conceito de endereçamento é fundamental em redes. Não existe nenhuma rede sem endereços. O espaço de endereçamento limita o tamanho da rede. Um grande espaço de endereços permite que, através de uma política de endereçamentos, possa minimizar as configurações e tempo de processamento.

O endereço IP é composto de duas partes: a parte de rede (também chamada de prefixo) e a parte do nó (chamada de sufixo). É importante dizer que os nós intermediários entre o nó A e nó Z, têm que determinar se o endereço destino está diretamente conectado a rede ou não. Examinando a parte de redes do endereço IP nos permite determinar se o endereço IP está diretamente conectado ou não. Se o nó destino estiver diretamente conectado, a parte de rede (prefixo) "casa" com a parte da rede de uma interface de saída do nó intermediário. Esta estrutura de endereçamento hierárquico, a qual é fundamental na escalabilidade da rede IP, não é vista no endereçamento existente na camada 2 (IEEE 802). Esta estrutura tem implicação na alocação dos endereços porque todas as interfaces de uma sub-rede têm que estar assinaladas no mesmo prefixo de rede.

Infelizmente, a alocação de endereços IPs no início da Internet não levou em conta algumas implicações que isso poderia causar no futuro e foi preciso fazer algumas evoluções no processo. Parte desta evolução foi forçada devido ao crescimento exponencial que a Internet teve e que não havia sido previsto. Esta evolução se caracterizou conceitualmente na identificação variável da parte de rede (prefixo) e da parte do nó (sufixo). No início o endereçamento seguia o esquema da classe cheia, onde o espaço de endereçamento era dividido em poucos blocos e demarcadores estáticos eram alocados para cada bloco. A Classe A tinha um demarcador de 8 bits; a Classe B tinha um demarcador de 16 bits; a Classe C tinha um demarcador de 24 bits. A Classe D era reservada para "*Multicast*" e a Classe E para uso futuro. (Figura 2.8)

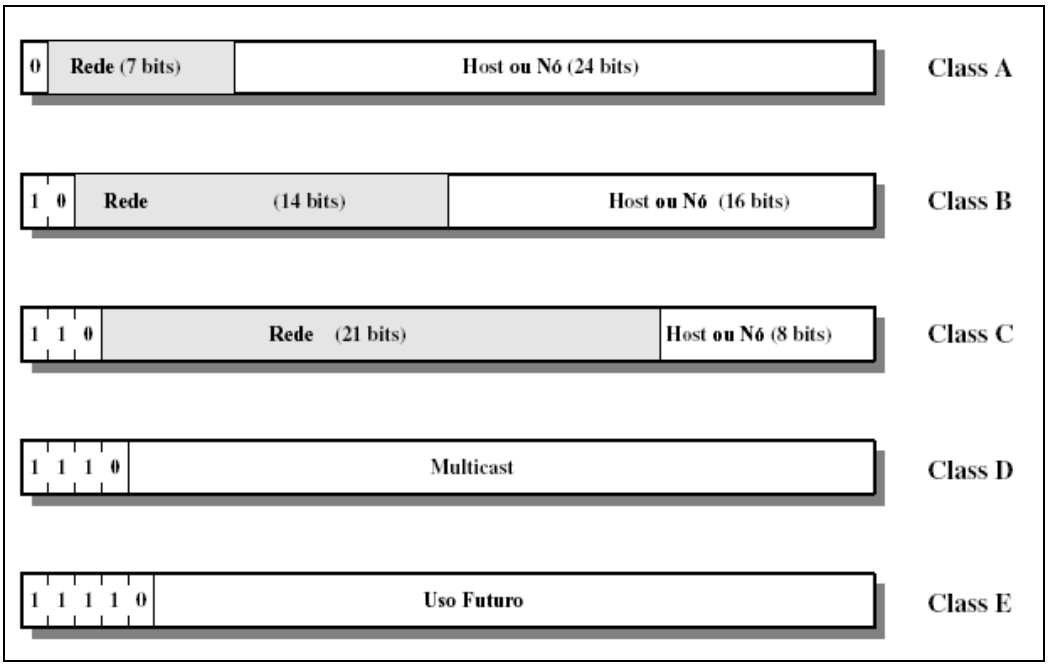


Figura 2.8 Endereçamento inicial de classe cheia [46].

Este modelo entrou em colapso no final dos anos oitenta (80) por duas razões: (1) as Classes B eram mais populares na rede (apesar de haver um grande número de Classes C sem alocação) e (2) os nós existentes na Classe A estavam sem uso, pois não existia uma rede tão grande que pudesse utilizar todos os endereços de nós ("*hosts*") existentes na Classe A. A solução para este problema foi simples e eficaz – permitir que o sufixo pudesse ser dividido mais vezes e permitir que os demarcadores de redes e de nós pudessem ter uma flexibilidade de alocação. Com esta solução, criou-se o conceito de máscara de sub-rede e máscara de super-rede. A máscara é um padrão de 32 bits, no qual os números 1's indicam a que rede os bits pertencem e os 0's indicam o número de nós ("*hosts*"). Por exemplo, a máscara de sub-rede 255.255.255.0 aplicada ao endereço IP 128.113.40.50 indica que a rede foi estendida de 16 bits para 24 bits (tendo em vista se o endereço IP 128.113.40.50 pertencente a Classe B cheia). As máscaras de super-redes são usadas entre sistemas autônomos para indicar alocação de endereços ou redes a serem anunciadas para roteamento. Por exemplo, a notação 198.28.29.0/18 indica um espaço de endereçamento de 18 bits. A máscara de super rede escrita como /18 é escrita como 255.255.192.0. Temos que observar que o endereço IP 198.28.29.0 pertence ao espaço de uma Classe C cheia e a Classe C admite apenas uma rede /24 (i.e., com um espaço para nós ("*hosts"*) de 8 bits).

Uma vez que os limites de classes não são mais usados no modelo de super-redes, este esquema de alocação de endereços é chamado de alocação sem classes e o esquema de roteamento é chamado de roteamento sem classes inter-domínios ("*Classless Inter Domain Routing"*). Um problema do CIDR e da máscara de super-redes é a possibilidade de um endereço IP de destino se enquadrar em múltiplos prefixo de tamanhos diferentes. Para resolver isto, o CIDR define que o maior prefixo "casado" é escolhido pela decisão de envio da camada de rede. Como resultado disso, todos os roteadores na metade dos anos oitenta (80) tiveram que refazer seus algoritmos de roteamento. De uma maneira

similar, quando a máscara de sub-rede foi introduzida, os nós ("*hosts"*) e os roteadores tiveram que ser configurados com máscaras de sub-redes e tiveram que aplicar a máscara no processo de envio dos pacotes para poder identificar a rede a qual pertencia o endereço IP.

Nas redes modernas dois outros esquemas são utilizados para também conservar os espaços de endereços públicos: o DHCP ("*Dynamic Host Configuration Protocol")* e o NAT (*"Network Address Translator"*). O DCHP foi originalmente projetado para ser um protocolo de boot na rede que configurava os parâmetros essências nos nós e nos roteadores. Atualmente ele é utilizado para distribuir um conjunto de endereços IP's públicos que são escassos entre nós que precisam acessar a Internet diretamente. Neste modelo os endereços IP's não pertencem ao nó que momentaneamente alocou o IP para si.

O NAT permite o uso de endereços privados, que não são roteados na Internet, dentro de grandes corporações. O IANA ("*Internet Assigned Numbers Authority"*) reservou os seguintes blocos de endereços IP para uso em Internet privada:

- 10.0.0.0 – 10.255.255.255 (prefixo 10/8)
- 172.16.0.0 – 172.31.255.255 (prefixo 172.16/12)
- 192.168.0.0 – 192.168.255.255 (prefixo 192.168/16)

Os equipamentos que possuem a capacidade de fazer NAT estão localizados nas bordas destas redes privadas para fazer a tradução dos endereços públicos em endereços privados, para todas as sessões que estiverem ativas e que passarem por esses equipamentos. O NAT quebra alguns protocolos de segurança, especificamente o IPSEC. O IPSEC protege (através da criptografia e autenticação de cada pacote) alguns campos do cabeçalho como os endereços IPs, bem como a parte de dados ("*payload*"). Os equipamentos que fazem o NAT não podem alterar os endereços IPs sem quebrar o IPSEC.

A combinação destas técnicas tem proporcionado uma sobrevida ao IPv4 e postergado a entrada do IPv6 no dia-a-dia.

2.5.5 O ARP, a Fragmentação e o Reagrupamento

No modelo de camadas usado pelo IP existem dois problemas de mapeamento: (1) o mapeamento do endereço e (2) o mapeamento do formato do pacote. O mapeamento do endereço é resolvido através de um protocolo chamado ARP ("*Address Resolution Protocol"*) e o mapeamento do pacote é feito através dos procedimentos de Fragmentação e Reagrupamento.

Os problemas de mapeamento de endereços ocorrem, na camada de rede, quando o endereço IP de destino é determinado ou endereço IP do próximo roteador ("hop"). O problema é o seguinte: o nó sabe que o endereço IP do próximo roteador, o qual é por definição conectado diretamente, ou seja, acessível via e entrega na camada de enlace – camada 2. Para se utilizar a entrega existente na camada 2 é necessário saber exatamente o endereço de enlace do próximo nó. Uma vez que os endereços das camadas 2 e 3 são assinalados independentemente, o mapeamento não é uma simples função relacional, i.e. o endereço tem que ser descoberto dinamicamente. O protocolo ARP é usado para fazer esta tarefa de mapeamento.

O protocolo ARP envia uma mensagem de difusão ("broadcast") na camada de rede, pedindo o mapeamento do endereço. Caso o próximo nó ("hop") esteja na mesma porção da rede física, o nó irá responder através de uma mensagem ARP de unicast

informando ao nó que requisitou o mapeamento o seu endereço da camada de rede. Feito isso o nó encapsula o datagrama IP no campo de carga útil ("*payload*") do frame da camada 2 e envia isso para a camada de enlace. Depois disso para evitar novas mensagens de "broadcast" o ARP armazena estas informações em uma tabela de ARP, reduzindo com isso a necessidade de mais mensagens de difusão ("broadcast"). Tendo em vista a volatilidade dos endereços na Camada 3, a tabela de ARP tem um período definido para que ela exista, sendo depois disso "zerada".

O problema do mapeamento do pacote ocorre quando o datagrama IP a ser enviado é maior do que a unidade de transmissão máxima (MTU) possível para a camada de enlace. Cada camada de enlace tem seu MTU específico, por razões como multiplexação justa, eficiência da detecção de erros et cetera. Por exemplo, nas redes Ethernet o MTU é de 1518 bytes. O objetivo é fragmentar o datagrama, de tal maneira que cada fragmento tenha até 1518 bytes de carga útil. Cada fragmento se torna um pacote IP independente; no entanto o cabeçalho do IP é copiado em todos os fragmentos. Além disso, ele precisa indicar o datagrama original, a posição do fragmento no datagrama original e se é o último datagrama ou não. Estas informações são preenchidas no campo de fragmentação do cabeçalho IP (ID, Flag, Deslocamento), respectivamente. O reagrupamento é feito depois na camada de rede do nó destino. Os fragmentos podem chegar fora de order ou atrasados. Uma tabela de reagrupamento de dados e um tempo máximo de espera do pacote é definido no nó destino para implementar esta função. O reagrupamento não ocorre nos nós intermediários, pois os fragmentos podem não ser roteados pelo mesmo caminho.

Apesar da fragmentação ser uma função necessária para a correção, ela possui inúmeras restrições em termos de performance. Isto acontece porque para qualquer um dos fragmentos perdidos, o datagrama inteiro é descartado no destino. Com isso há uma perda enorme de recursos em todos os nós intermediários que tiveram um certo trabalho que foi inútil. Para tentar diminuir esta restrição, os protocolos modernos tentam evitar a fragmentação tanto quanto possível, primeiramente descobrindo o MTU mínimo do caminho a ser seguido. Este procedimento é conhecido como descoberta do MTU mínimo ("*path-MTU discover*"). Normalmente a cada seis (6) segundos, uma sessão ativa irá chamar o procedimento de descoberta do MTU mínimo. O procedimento se inicia com o envio de um datagrama de tamanho máximo com o bit "*do not fragment*" ligado. Quando um roteador é forçado a considerar a fragmentação devido a um MTU menor do que o datagrama, ele descarta o datagrama e envia uma mensagem de ICMP indicando o MTU do link. O nó emissor inicia o procedimento novamente com o novo valor de MTU recebido. Este procedimento é repetido até que um pacote de tamanho apropriado chegue ao nó destino. O valor do MTU utilizado por este datagrama é utilizado para futuras transmissões. Resumindo, podemos dizer que os problemas de mapeamento do IP são resolvidos através do protocolo ARP e os procedimentos de fragmentação e reagrupamento. É de fundamental importância evitar o processo de fragmentação e para isso é usado o procedimento chamado de descoberta do MTU mínimo ("*path-MTU discover*").

CAPÍTULO 3

TRANSMISSÃO EM FIBRAS ÓPTICAS

O primeiro trabalho apresentado sobre a possibilidade de controlar a direção do raio de luz, através de jatos d'água, foi realizado em 1840 em Londres por Daniel Collodon e Jacques Babinet. Noventa anos depois, em 1930, Heinrich Lamm demonstrou a transmissão de imagens através de um conjunto de fibras ópticas de partes inacessíveis do corpo humano em uma aplicação médica.

No entanto, o grande avanço das comunicações ópticas está associado ao desenvolvimento do laser (1960) e da própria fibra (1970). Na Figura 3.0, temos o espectro de freqüência eletromagnético e as respectivas aplicações de cada. Nesta figura pode-se verificar o enorme potencial de transmissão existente na fibra óptica.

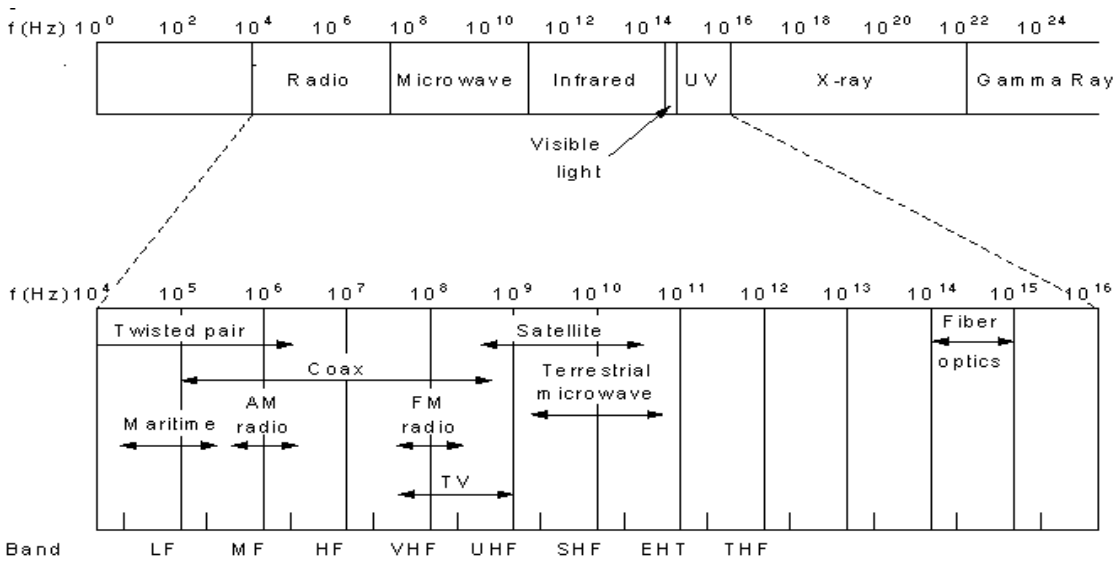


Figura 3.0 - O espectro de freqüência eletromagnético [12].

O primeiro estudo da propagação de luz em fibras ópticas foi publicado em 1966 por Kao e Hockham [8] e seu potencial para sistemas de telecomunicações foi logo percebido. Apesar do potencial uso em sistemas de telecomunicações, apenas em 1977 os primeiros sistemas práticos foram desenvolvidos para transmissões em até 140 Mbps [9].

A característica da banda passante (velocidade efetiva de um meio de transmissão) na fibra óptica pode ser observada na Figura 3.1.

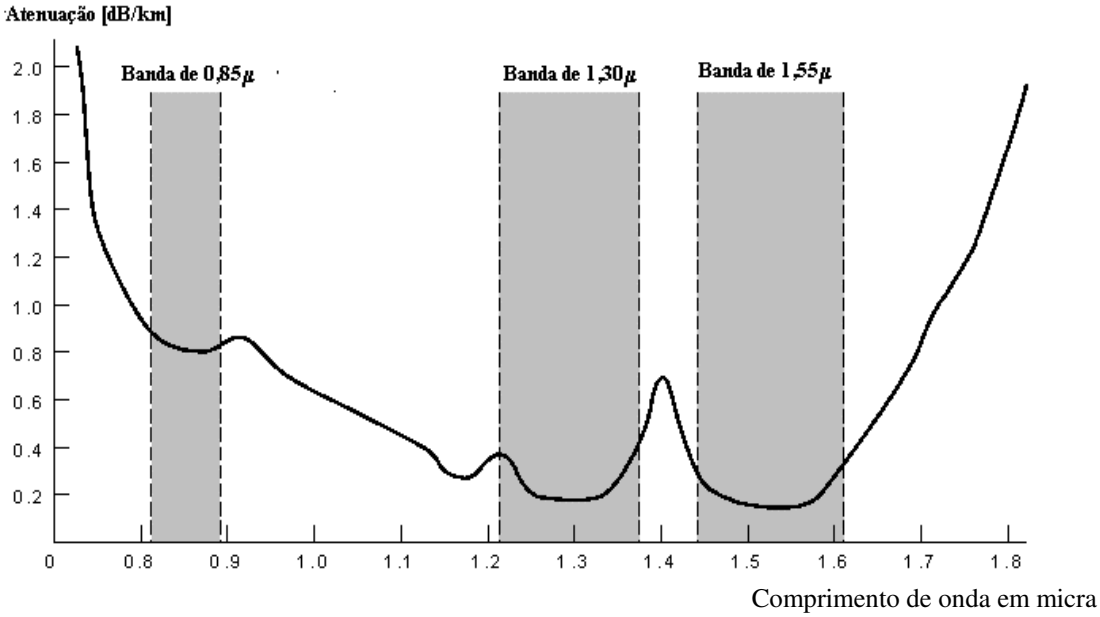


Fig. 3.1 - Atenuação de uma fibra óptica na região do infravermelho [12].

As três bandas assinaladas na Figura 3.1, e que são utilizadas em comunicação de dados, possuem cada uma, uma largura de banda de 26.000 a 30.000 GHz

- Banda de 0,85μ (ou λ=850nm)
  - Nesta banda é utilizada a tecnologia de LED e fotodetectores, com atenuação de 17% por quilômetro.
- Banda de 1,30μ (ou λ =1300nm)
  - Tecnologia de LED ou Laser, mais cara porém a atenuação é de aproximadamente 4% por quilômetro
- Banda de 1,55μ (ou λ =1550nm)
  - Tecnologia de Laser com outras dopagens. A atenuação também é de aproximadamente 4% por quilômetro

Como foi dito na Introdução deste trabalho, hoje existem sistemas que conseguem transmitir até 160 Gbps por comprimento de onda (cor) de luz (lâmbda). Nesta seção, pretende-se apresentar uma visão geral sobre a transmissão em fibras ópticas.

3.1 Propagação da luz em fibras ópticas

A luz, que na Antigüidade se suponha propagar instantaneamente, teve sua velocidade de propagação medida pela primeira vez por Olaus Röemer (1666), na época de Newton, por meios astronômicos. Encontrou-se um valor da ordem de $3x\ 10^8$ m/s e, com o passar do tempo as medições, hoje por meio terrestres, levam ao valor de $(2{,}997924580 \pm 0{,}000\ 000\ 012)\ x10^8$ m/s, média de várias medidas por vários processos no vácuo.

Em 1865 James Clerk Maxwell conseguiu unificar todo o conhecimento existente sobre os campos elétrico e magnético num conjunto de quatro equações fundamentais, chamadas de Equações de Maxwell. Essas equações não só contêm as leis fundamentais, como as de Gauss, Ampère, Faraday e outras, como também nos permitem concluir que a

luz é um campo eletromagnético que se propaga no espaço como uma onda. Sendo uma onda, a luz apresenta características que, embora sejam decorrentes das equações de Maxwell, podem ser descritas sem a necessidade de levarmos em conta sua natureza eletromagnética. Essas características são:

- ✓ Propaga-se em linha reta num meio isotrópico e homogêneo.
- ✓ Apresenta reflexão e refração na interface de meios nos quais a velocidade de propagação é diferente.

O formalismo que descreve esses aspectos da luz é conhecido como Óptica Geométrica. Esse ramo da Ótica, de origem pré Maxwelliana, encontra-se até hoje em constante evolução e tem aplicações que vão desde o desenho de novos equipamentos ópticos até a integração de circuitos.

Sendo um campo vetorial eletromagnético, a luz também apresenta características que só podem ser compreendidas tendo esse fato em mente. Dentre essas, a polarização é uma das mais interessantes, pois, além de ser facilmente compreendida, tem vasta aplicação na confecção de filtros óticos, na transmissão de informação e, até mesmo, na construção de portas lógicas para micro processadores de alta performance.

A luz vem sendo usada como meio de transmissão de sinais há muito tempo. Entretanto a eficácia da transmissão é limitada por várias considerações: é essencial a visada direta, sem ser obstruída por nuvens, fumaça ou semelhantes, e a distância é limitada pela potência do transmissor e a sensibilidade do receptor. Em particular, a potência do sinal cai com o inverso do quadrado da distância, devido ao espalhamento da radiação. Uma maneira de manter concentrada esta energia é usar uma guia de onda, que canaliza em uma única direção toda a energia da transmissão, possibilitando estender o alcance de uma transmissão.

3.1.1 Fibras ópticas

A fibra óptica funciona como um guia de onda para a luz. As fibras ópticas são feitas tanto em vidro (Sílica – SiO_2), como em plástico. As fibras de plástico são mais baratas, mas exibem uma atenuação bem maior. As dimensões físicas do diâmetro variam de 5 a 100 μm (micra) (1 micron = 10^{-6} m).

O mecanismo de propagação da luz pela fibra está baseado num fenômeno da Física chamado de refração de um raio luminoso ao passar entre dois meios com índices de refração distintos (Fig. 3.2).

Índice de refração $n = \frac{c}{v}$ c = velocidade da luz no vácuo

v = velocidade da luz no meio

O índice de refração depende da freqüência pois, $c = \lambda f$ onde:

λ = comprimento de onda
f = freqüência da onda.

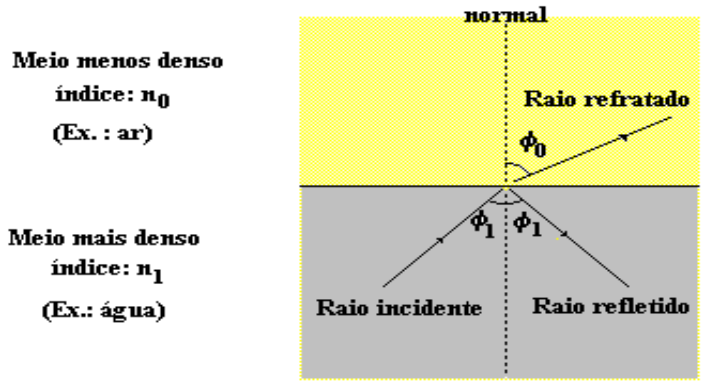


Fig. 3.2 – O fenômeno da refração de um raio luminoso

Existe uma relação entre os índices dos meios e os ângulos dos raios luminosos incidentes e refratados em relação a uma reta normal à superfície de separação conhecida como Lei de Snell.

$$\mathbf{n_0 \text{ sen } \Phi_0 = n_1 \text{ sen } \Phi_1}$$

A Física mostra que existe um ângulo $\mathbf{\Phi_C}$, chamado ângulo crítico, tal que, qualquer ângulo de incidência $\mathbf{\Phi_1}$ com $\mathbf{\Phi_1 \leq \Phi_C}$, não haverá raio refratado, ou seja , o raio será totalmente refletido de volta no limite entre os dois meios. Pode se mostrar que este ângulo crítico $\mathbf{\Phi_C}$ pode ser dado por:

$\Phi_c \cong \text{arcsen}\dfrac{n_0}{n_1}$ $\mathbf{\Phi_C}$= ângulo crítico

$\Phi_c \cong \text{arcsen}\dfrac{n_1}{n_2}$ Dois meios quaisquer com n1< n2 (n1 menos denso)

A fibra óptica é constituída de um núcleo de vidro mais denso, circundado por uma cobertura ("*clading*") menos densa (Figura 3.3).

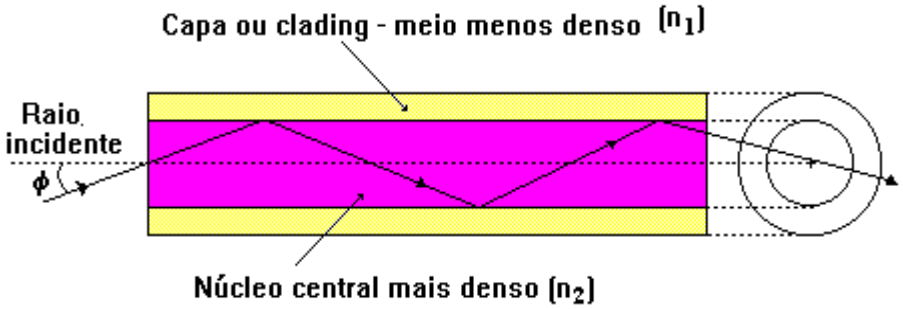


Fig. 3.3 - Mecanismo de propagação de um raio luminoso numa fibra óptica
$\mathbf{\Phi}$ – ângulo máximo no qual a reflexão total ocorre.

Para que o raio luminoso se propague pela fibra através de múltiplas reflexões sem que haja refração (fuga) o angulo de incidência deverá obedecer à condição:

$$\Phi \leq \Phi_c \cong \text{arcsen}\frac{n_1}{n_2}$$

O ângulo crítico no caso da fibra óptica também é chamado de Abertura Numérica (*NA- "Numerical Aperture"*). O ângulo crítico corresponde a uma determinada freqüência de radiação e é chamado de modo de transmissão da fibra para uma determinada radiação. A cada comprimento de onda λ corresponde uma determinada abertura numérica.

A transmissão de uma onda luminosa por uma fibra óptica é limitada quanto ao comprimento da fibra, devido principalmente a dispersão no tempo e a atenuação na amplitude do sinal luminoso. A Figura 3.4 mostra as conseqüências destes dois fenômenos sobre um pulso luminoso.

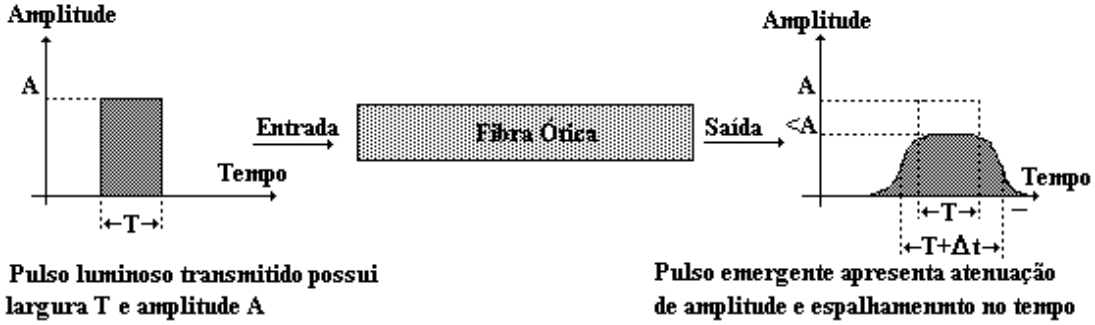


Fig. 3.4 - Atenuação de amplitude e dispersão temporal em fibra óptica

A atenuação é causada principalmente por impurezas no material (transparência) e é de difícil controle na fabricação.

A atenuação reduz gradativamente a energia do sinal, limitando o seu alcance. Para tornar a transmissão óptica eficaz em distâncias maiores, é necessário regenerar o sinal de tempos em tempos. Num regenerador completo, o sinal sofre a chamada regeneração 3R, ou seja, Re-amplificação (restaura a energia), Re-moldagem (recupera a forma "digital") e Re-sincronização (ajuste da taxa de sinalização) (Figura 3.5). Tipicamente a regeneração 3R é feita através de equipamentos eletrônicos.

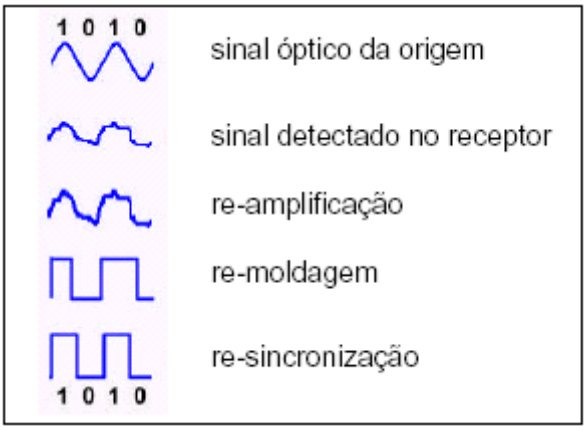


Figura 3.5 - Regeneração 3R de um sinal.

A dispersão no tempo é causada principalmente devido à incidência da luz em vários ângulos na entrada, fazendo com que os caminhos percorridos variem e os tempos de chegada no outro lado também (dispersão modal). Um outro fator que causa dispersão é que a luz na entrada possui diversos comprimentos de onda (cores) o que causa tempos de propagação diferentes e portanto dispersão. As impurezas dentro da fibra óptica também causam o fenômeno de dispersão.

A atenuação de amplitude do pulso luminoso ao passar por uma fibra óptica é principalmente devido às perdas causadas por impurezas dentro do núcleo central. As modernas técnicas de purificação têm conseguido fibras com atenuação menor que 0,1 dB/Km e a cada ano o comprimento do segmento entre repetidores praticamente dobra.

As fibras de vidro são classificadas, basicamente, em dois tipos segundo critérios de construção física e o desempenho associado. Definiu-se um fator de qualidade para as fibras ópticas denominado, Capacidade de Transmissão da fibra, o qual é praticamente constante para cada tipo de fibra.

A Capacidade de Transmissão C_T de uma fibra é por definição, o produto da banda passante (velocidade efetiva de um meio de transmissão) pela distância e é aproximadamente constante para um determinado tipo de fibra.

C_T = Banda Passante x Distância

A capacidade de transmissão das fibras de vidro tem praticamente dobrado a cada ano como pode ser observado na Figura 3.6.

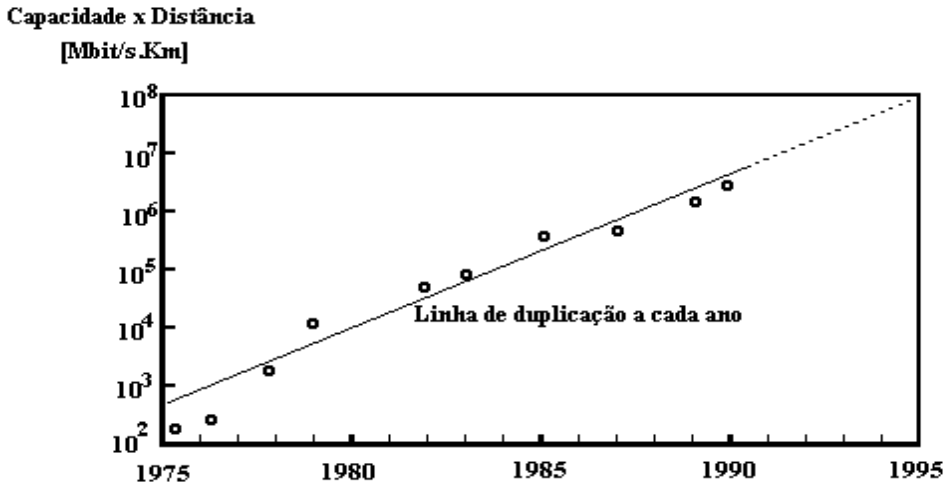


Fig. 3.6 - Crescimento da capacidade de transmissão nas fibras ópticas

3.1.2 Tipos de fibra óptica

De acordo com a tecnologia de construção do núcleo central da fibra pode-se distinguir entre dois tipos de fibra óptica:

a - fibra óptica do tipo multímodo

b - fibra óptica do tipo monomodo.

As fibras tipo multímodo ("*Multi-Mode Fiber*"- MMF) foram as primeiras fibras a surgir. Nestas fibras existem múltiplos modos de propagação em raios fazendo ângulos diferentes ao eixo da fibra e de comprimentos diferentes entre si. Isto significa que a velocidade de propagação longitudinal é ligeiramente diferente para cada modo de propagação, ocasionando a chamada dispersão intermodal do sinal. Esta dispersão intermodal é a principal limitação do uso de fibras tipo multímodo, pois restringe seu uso para taxas de até centenas de Mbps sobre distâncias curtas, tipicamente 2 km. Hoje em dia são largamente empregadas em aplicações de curta distância, como por exemplo em redes locais e automação industrial. A dispersão causa o espalhamento gradativo de cada pulso de uma seqüência, dificultando distingui-los dos seus vizinhos (Figura 3.7).

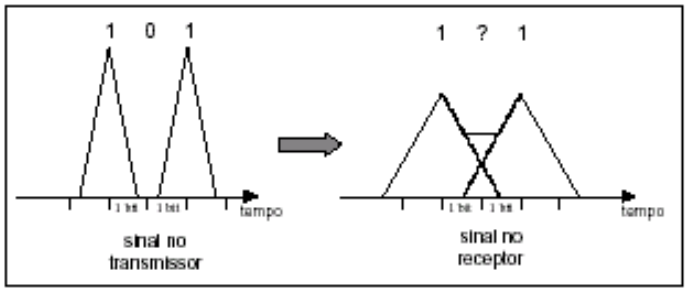


Figura 3.7 - O efeito de dispersão

Nas fibras tipo monomodo ("*Single-Mode Fiber"* – SMF) consegue-se apenas um modo de propagação, paralelo ao eixo da fibra, através do estreitamento do diâmetro do núcleo da fibra, minimizando-se desta forma a dispersão temporal. As fibras monomodo

são atualmente as fibras que apresentam o melhor desempenho e por isso são utilizadas em troncos de fibra óptica de longa distância.

Na Figura 3.8 apresentam-se alguns detalhes dos dois tipos de fibra e o desempenho associado a cada uma em termos de dispersão temporal.

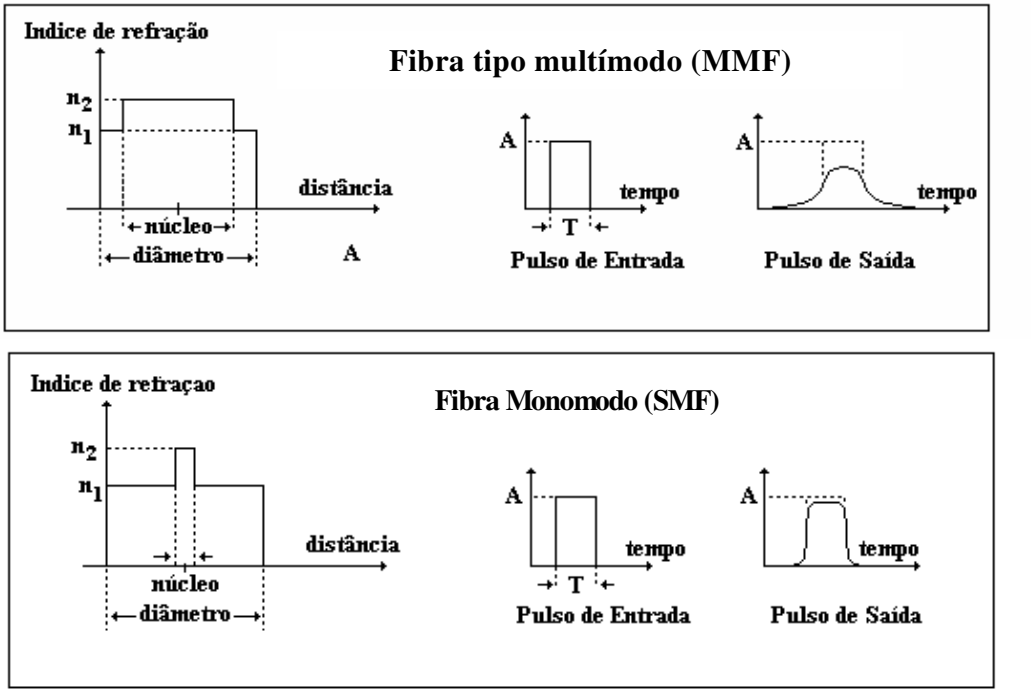


Fig. 3.8 - Detalhes construtivos dos diversos tipos de fibras e performance quanto à dispersão temporal considerando segmentos de mesmo comprimento.

3.2 Lasers

Para entendermos o funcionamento de um laser, vamos tomar um laser a gás (HeNe) de maneira didática onde os números usados são ilusórios para maior visualização dos fenômenos. Um átomo é composto de um núcleo e de elétrons que permanecem girando em torno do mesmo em órbitas bem definidas.Quanto mais afastado do núcleo gira o elétron, menor a sua energia. Quando um elétron ganha energia ele muda de sua órbita para uma órbita mais interna, sendo este um estado não natural para o átomo, mas sim forçado. Como esse estado não é natural, o átomo por qualquer distúrbio tende a voltar a seu estado natural, liberando a energia recebida em forma de ondas eletromagnéticas de comprimento de onda definido em função das órbitas do átomo.

Existem duas condições básicas para que o fenômeno laser aconteça:

- Inversão de população
- Alta concentração de luz

Inversão de população é o estado em que uma grande quantidade de átomos fica com elétrons carregados de energia, girando em órbitas mais internas. É como se o átomo fosse engatilhado para o disparo de ondas eletromagnéticas (os fótons). Esse estado é conseguido através de altas tensões de polarização fornecidas ao laser (200 à 300V).

A alta concentração de luz é a perturbação necessária para que o átomo dispare, ou seja, volte à sua condição natural, liberando a energia armazenada em forma de ondas eletromagnéticas. Se tivermos uma quantidade de átomos suficientes engatilhados e se a concentração de luz for suficiente teremos um efeito multiplicativo onde o fóton gerado gera outros fótons, obtendo-se assim o fenômeno laser (emissão de radiação estimulada amplificada pela luz).

O laser ("*Light Amplification by Stimulated Emission of Radiation"*) é responsável pela geração dos sinais ópticos a serem transmitidos num sistema óptico. (Poucas pessoas se atêm a esse fato, mas o laser foi descoberto por um físico brasileiro chamado Sérgio Porto, na época em que trabalhava nos EUA nos Laboratórios da Bell.).

Isso é feito através da emissão estimulada de fótons, que é o que permite ao laser produzir intensos feixes de alta potência de luz coerente (luz que contém uma ou mais freqüências distintas ou cores). (Figura 3.9)

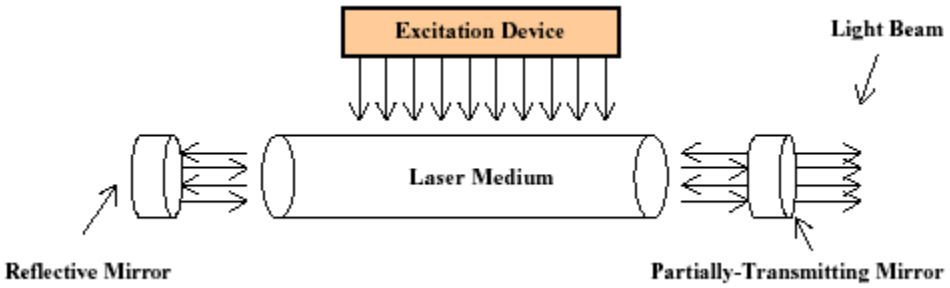


Figura 3.9 – Princípio do funcionamento de um Laser

Algumas das características físicas dos lasers que podem afetar o desempenho do sistema são: a largura de linha do laser, a sua estabilidade em freqüência e o número de modos longitudinais. A largura de linha do laser é a largura espectral da luz (cor) gerada pelo laser. A largura de linha afeta o espaçamento dos canais e também afeta a quantidade de dispersão que ocorre quando a luz está se propagando ao longo da fibra. Esse efeito de dispersão limita a taxa máxima de transmissão de bit. A instabilidade de freqüência nos lasers são variações na freqüência do laser. A fim de evitar grandes deslocamentos em freqüência devem ser utilizados métodos compensativos através de variações na temperatura ou pela injeção de corrente. O número de modo longitudinais em um laser é o número de comprimentos de onda que ele pode amplificar.

3.3 A transmissão em fibra óptica

Para transmitir dados através de uma fibra óptica a informação deve primeiramente ser codificada ou modulada dentro do sinal do laser e a distância desta transmissão é limitada pela potência do transmissor e a sensibilidade do receptor. A potência do sinal diminui com o inverso do quadrado da distância, devido ao espalhamento da radiação. Existem inúmeras técnicas de modulação tanto analógicas, quanto digitais. Dentre as técnicas de modulação existentes o ASK binário é freqüentemente o método de modulação digital preferido em razão da sua simplicidade. No ASK binário, também conhecido como OOK ("*On-Off Keying*"), o sinal é comutado entre dois níveis de potência. O nível mais baixo representa um bit "0", enquanto o nível mais alto representa um bit "1". Nos sistemas empregando OOK, a modulação do sinal

pode ser realizada simplesmente ligando e desligando o laser. Outra forma de modular o sinal é através de um modulador externo que modula a luz que está saindo do laser. O modulador externo bloqueia ou deixa passar a luz dependendo da corrente que está sendo aplicada sobre ele.

A transmissão de sinais digitais pode ser realizada usando LEDs (diodos emissores da luz) ou lasers MLM ("Multi-Longitudinal Mode", ou de "Fabry-Perot"). Os LEDs são de baixa potência, e emitem luz num espectro amplo e contínuo. Os lasers MLM também emitem luz num espectro também amplo, mas concentrado em alguns comprimentos de onda (cores) discretos.(Figura 3.10)

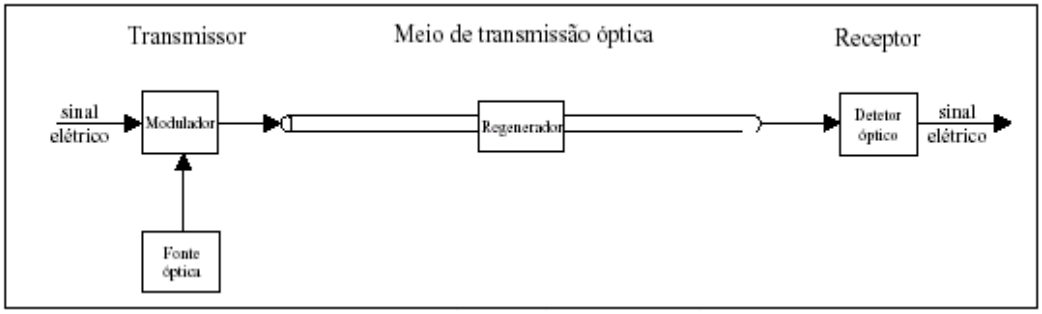


Figura 3.10 - Sistema de transmissão óptica

Como visto anteriormente, as fibras tipo monomodo SMF ("*Single-Mode Fiber*") devido às suas características são utilizadas para transmissões ópticas de longa distância. A capacidade de transmissão destas fibras, usando lasers MLM na faixa de 1300nm, pode chegar a centenas de Mbps em distâncias maiores que 40 km. [10].

Para alcançar distâncias maiores, deve-se passar a transmitir na faixa de 1550 nm, onde a atenuação é ainda menor do que em 1300 nm. Com isto, passou a ser importante o problema de dispersão cromática do sinal gerado pelos lasers MLM. Num laser MLM ("Fabry-Perot"), a energia do sinal transmitido está concentrada em diversos comprimentos de onda em torno do comprimento de onda central. A dependência do comprimento de onda da velocidade de propagação da luz na fibra causa dispersão cromática, como descrita acima. Este efeito pode ser combatido pela introdução de um novo tipo de laser, chamado SLM ("Single Longitudinal Mode", também conhecido como "DFB-Distributed Feedback"), onde a energia transmitida está inteiramente concentrada num só comprimento de onda (Figura 3.11).

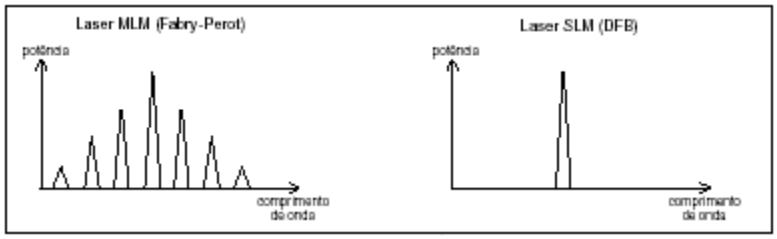


Figura 3.11 - Comparação dos espectros de lasers MLM e SLM

3.3.1 A regeneração puramente óptica

A regeneração 3R (Re-amplificação, Re-moldagem e Re-sincronização), citada anteriormente, é relativamente cara, pois é feita eletronicamente, requerendo a conversão do sinal do domínio óptico para o eletrônico e vice-versa. Se nós pudermos dispensar 2Rs (Re-moldagem e a Re-sincronização) desta regeneração do sinal, ficando apenas com amplificação, temos a regeneração 1R. Nos últimos anos, foram descobertas maneiras de fazer a amplificação puramente óptica dos sinais, eliminando as conversões O-E (óptico-eletrônica) e E-O (eletrônico-óptica).

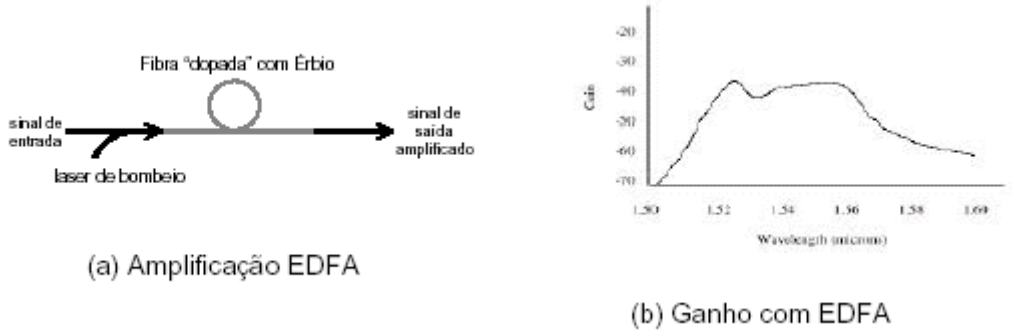


Figura 3.12 - Amplificação usando Fibra Dopada com Érbio (EFDA)

A EDFA ("*Erbium Doped Fibre Amplification*") emprega um trecho (de até algumas dezenas de metros) de fibra "dopada" com o elemento érbio, e um laser de bombeio funcionando em 980 ou 1480 nm. O sinal a ser amplificado, na faixa de 1550 nm, e a saída do laser de bombeio são combinados e passam pelo trecho de fibra dopada. A estimulação dos átomos de érbio causada pela bombeio efetua emissão de fótons na mesma faixa que o sinal incidente, contribuindo para sua amplificação (Figura 3.12(a)). A Figura 3.12(b) mostra que a EDFA amplifica, de forma mais ou menos uniforme, os comprimentos de onda entre 1520 e 1560 nm, os quais, para a alegria geral, coincidem quase exatamente à janela de baixa atenuação na faixa de 1550 nm (Figura 3.1).

Uma segunda técnica de amplificação óptica, chamada de amplificação Raman, também utiliza um laser de bombeio, mas sem a necessidade da fibra especial (dopada). Na amplificação Raman, a amplificação ocorre na fibra padrão e é distribuída sobre vários quilômetros. Ela poderá ser efetuada usando um laser bombeando luz no mesmo sentido que o sinal (co-bombeamento), no sentido contrário (contra-bombeamento), ou em ambas direções simultaneamente (co-contra-bombeamento) (Figura 3.13). Para ampliar diferentes comprimentos de onda (cores) do sinal de forma uniforme, é necessário também realizar o bombeamento em diversos comprimentos de onda.

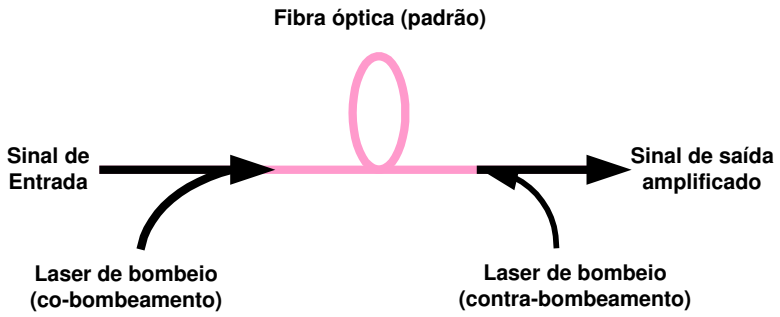


Figura 3.13 - Amplificação Raman

3.4 A multiplexação óptica WDM e a evolução dos sistemas

O limite de transmissão de um sistema óptico está relacionado aos equipamentos eletrônicos terminais utilizados. De forma análoga à multiplexação feita através da multiplexação por divisão de freqüência ("*Frequency Division Multiplexing* - FDM"), criou-se a multiplexação por divisão de comprimento de onda ("*Wavelength Division Multiplexing*"- WDM). Isto é, cada canal representa à transmissão de um sinal por um laser dedicado num determinado comprimento de onda (lâmbda) e os sinais são combinados para propagação na mesma fibra. No destino, os sinais em diferentes lâmbdas são novamente separados e destinados a detectores ópticos próprios. (Figura 3.14).

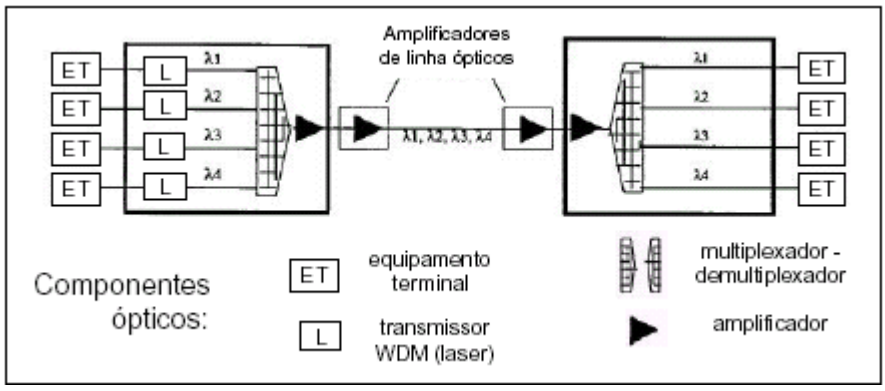


Figura 3.14 - Sistema WDM: multiplexação por comprimento de onda [45]

O WDM é uma tecnologia de multiplexação analógica, pois cada canal pode ser modulado de maneira independente, sem a necessidade dos canais usarem a mesma taxa de sinalização nem de se fazer sincronização entre si. Em princípio, cada canal de transmissão suportar uma taxa máxima de sinalização, determinada pelas características ópticas do enlace. Claramente, a capacidade agregada do enlace será a soma das capacidades de cada canal individual.

A variedade mais importante hoje de WDM é a chamada DWDM – WDM densa, com transmissão na faixa de 1550 nm, e onde é essencial a utilização de lasers SLM para transmissão DWDM, para evitar interferência entre canais vizinhos; pois nesta faixa de 1550nm os canais estão muito próximos.

O desenvolvimento e ampla adoção de DWDM na prática foi devido à descoberta da amplificação óptica usando EDFA, o que permite amplificar simultaneamente todos os sinais sendo multiplexados juntos, com grande economia em termos de equipamentos e um aumento significativo da capacidade transmitida e/ou da distância alcançada. (Figura 3.15) [11].

Configuração	**Capacidade**
1 λ x 40 Gbps até 65 km (Alcatel 1998)	40 Gbps
32 λ x 5 Gbps até 9300 km (1998)	160 Gbps
64 λ x 5 Gbps até 7200 km (Lucent 1997)	320 Gbps
100 λ x 10 Gbps até 400 km (Lucent 1997)	1 Tbps
16 λ x 10 Gbps até 6000 km (1998)	160 Gbps
132 λ x 20 Gbps até 120 km (NEC 1996)	2,64 Tbps
70 λ x 20 Gbps até 60 km (NTT 1997)	1,4 Tbps
80 λ x 40 Gbps até 60 km (Siemesn 2000)	3,2 Tbps
1022 λ em uma única fibra (Lucent 1999)	
64 λ x 40 Gbps até 4000 km (Lucent 2002)	2,56 Tbps

Tabela 3.15 - Marcas de desenvolvimento de transmissão DWDM.

Deve-se notar também que um sistema de transmissão DWDM é altamente modular, sendo possível configurar o número de lambdas desejado, através da colocação de lasers e detectores em números suficientes nas pontas do trecho.

A Figura 3.16 apresenta um sumário da evolução de sistemas de transmissão óptica desde sua introdução até os dias de hoje. Inicialmente utilizou-se uma fibra multímoda com o uso de LEDs para iluminá-las. O curto alcance desta combinação obrigou a utilização de regeneradores em intervalos pequenos (Figura 3.16(a)).

Na segunda fase, passou-se a adotar lasers SLM para iluminar as fibras multímodas, transmitindo na faixa de 1300 nm. Esta combinação permitiu reduzir o número de regeneradores ao longo do enlace, por ter eliminado as limitações devidas à dispersão intermodal (Figura 3.16(b)).

Numa terceira fase, foi estendido o alcance das transmissões sem regeneração através da migração para a faixa de 1550 nm, de menor atenuação, e por eliminar a dispersão cromática com o uso de lasers MLM (Figura 3.16(c)).

A última modificação retrata a adoção de DWDM, com transmissão ainda na faixa de 1550 nm, e com adoção de amplificação EDFA, ao invés da regeneração óptico-eletrônica até então usada (Figura 3.16(d)).

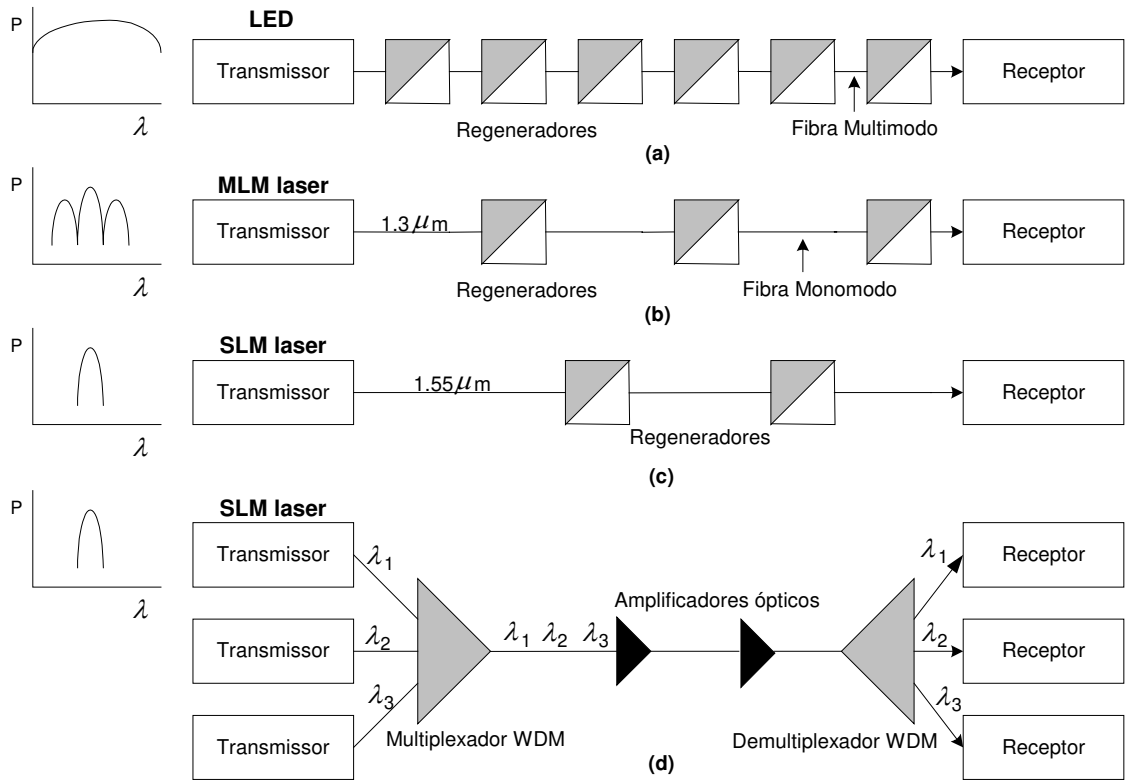


Figura 3.16 - Evolução de sistemas de transmissão óptica: (a) sistema usando LEDs sobre fibra multímodo; (b) uso de lasers MLM com fibra monomodo na faixa de 1300 nm para evitar dispersão intermodal;(c) uso de lasers SLM na faixa de 1550 nm para evitar dispersão cromática; (d) sistema atual com múltiplos comprimentos de onda na faixa de 1550 nm e amplificadores ópticos em lugar de regeneradores [10].

CAPÍTULO 4

TÉCNICAS DE COMUTAÇÃO

A função de comutação, ou chaveamento, em uma rede de comunicação se refere à alocação dos recursos da rede (meios de transmissão, repetidores, sistemas intermediários etc.) para a transmissão pelos diversos dispositivos conectados. As redes utilizam, atualmente, diferentes formas de comutação eletrônica. Com o uso massivo das fibras ópticas nos últimos anos, as redes continuaram usando elementos de comutação eletrônicos, porém com interfaces ópticas, passando a se chamar OEO (interface Óptica – comutador Eletrônico – interface Óptica). O uso de elementos de comutação OEO limita a capacidade da rede, tendo em vista que a capacidade de comutação destes elementos é da ordem de dezenas de Gbps, enquanto que a capacidade nas fibras ópticas chega a algumas dezenas de Tbps. Para se tirar o máximo proveito das fibras ópticas e utilizar todo o seu potencial, o ideal seria utilizar-se apenas de elementos de comutação OOO (interface Óptica – comutador Óptico – interface Óptica), sem a necessidade de se fazer conversões óptico-eletrônicas. Entretanto, o processamento e o armazenamento ópticos de informação não possuem a mesma flexibilidade existente na comutação eletrônica.

Nesta seção, o foco será os comutadores ("*switches"*) de rede os quais possam ser configuráveis. Neste contexto e de forma mais ampla, comutadores podem ser entendidos como sendo roteadores ("*routers"*), equipamentos de conexão cruzada ("*cross-connection"*) e multiplexadores *add-drop* (ADM*s*). No caso específico do mundo óptico, um comutador óptico é aquele que pode comutar um sinal de dados óptico sem convertê-lo do domínio óptico para o domínio eletrônico (conversões O/E/O), embora o comutador possa ser controlado através de sinais eletrônicos. Comutadores são comumente utilizados no núcleo da rede (*backcone* ou *core*) ao invés de serem utilizados na rede de acesso. Em uma rede comutada, nem todos os nós (onde um nó consiste de um comutador e seu controlador) estão ligados diretamente uns aos outros; com isso, para ir de um nó para outro, o circuito poderá passar por vários nós intermediários.

O aumento de demanda para uma infra-estrutura de rede transparente que possa prover serviços integrados tem inspirado muitas pesquisas em redes ópticas. As pesquisas em tecnologias de comunicação óptica, em especial o WDM, têm sido movidas das redes locais em estrela para o uso em redes metropolitanas e em redes de longa distância, principalmente baseada em redes ópticas em anel ou em redes em malha. Apesar de existirem vários estudos sobre as técnicas de comutação entre circuitos e pacotes, no campo óptico, os estudos estão em andamento e se baseiam em técnicas de roteamento de comprimento de onda e em *Optical Burst Switching (OBS)*. Neste capítulo, as técnicas de comutação existente hoje em dia são descritas, bem como os estudos que estão sendo feitos para o WDM.

4.1.1 Tipos de comutação e comutadores

O trabalho realizado por um comutador é bem simples. Tendo em vista as informações de controle, ele encaminha o tráfego recebido por uma porta de entrada, através de uma matriz de comutação, para uma ou mais portas de saída.

As redes de comunicação construídas a partir desses dispositivos de comutação costumam ser chamadas de redes comutadas. Todavia, dependendo do tipo de tráfego (voz, vídeo, ou dados) que essas redes irão transportar, diferentes paradigmas de comutação podem ser empregados. A princípio, existem três alternativas: a comutação de circuitos, a comutação de pacotes e a comutação de rajadas.

A comutação de circuitos utiliza um canal dedicado (chamado circuito) entre as duas estações que desejam se comunicar, e possui três fases distintas: o estabelecimento do circuito, a transmissão dos dados e a liberação do circuito. O estabelecimento do circuito pode ser feito "manualmente", através de um sistema de gerenciamento, ou por meio de um processo automático de sinalização, onde o emissor envia uma solicitação através da rede para estabelecer o circuito, e recebe de volta uma confirmação enviada pelo destino, caso a operação tenha sido bem sucedida. Os dados só podem ser transmitidos após o recebimento da confirmação do estabelecimento da conexão.

Na comutação de pacotes os dados são transmitidos em pequenos pedaços, chamados pacotes, os quais podem ser de comprimento fixo ou variável com um limite mínimo e máximo, no caso de comprimento variável. Não há o processo de estabelecimento de um caminho dedicado para transmissão dos dados. Logo, cada pacote precisa conter informações de controle necessárias para que ele possa ser encaminhado até o destino. Os comutadores de pacotes, exemplificados por comutadores Ethernet, roteadores IP e comutadores ATM, agem no modo armazena e encaminha (“store-and-forward”), ou seja, quando o pacote chega no comutador, a unidade inteira é armazenada temporariamente em um dispositivo apropriado (“buffer” de entrada). Na seqüência, as informações de controle são então analisadas e o pacote é encaminhado, através da matriz de comutação, para a porta de saída adequada.

Na comutação de rajadas, a unidade de transmissão, apesar de poder ter tamanho variável, tem uma granularidade intermediária entre o circuito e o pacote, pois uma rajada corresponde a uma seqüência de pacotes. Como na versão "automática" de comutação de circuitos, uma mensagem de solicitação de reserva de recursos é enviada, através de um canal de controle, antes da transmissão da rajada de dados. Contudo, diferentemente da comutação de circuitos, o processo de reserva dos recursos não aguarda mensagem de confirmação, ou seja, a rajada de dados é enviada em seguida ao envio da solicitação de reserva, sem esperar por qualquer tipo de confirmação sobre a reserva dos recursos. Se, por ventura, a mensagem de solicitação não puder ser atendida em algum nó intermediário ao longo do caminho, a rajada de dados relacionada será sumariamente descartada. Além disso, na comutação de rajadas os canais não ficam dedicados além do necessário: os recursos da rede são utilizados apenas para a propagação da rajada e são liberados tão logo esta termine.

A seguir serão apresentados mais detalhes sobre comutação por Circuito, Pacote, por Rajadas, assim como a comutação óptica.

4.2 Comutação por circuito e por pacote

Existem, basicamente, dois modelos de comutação: por circuito e por pacotes. A comutação por circuito é utilizada em redes tradicionais de comunicação de voz, enquanto a comutação por pacotes é utilizada em redes de comunicação de dados. Neste contexto, o termo dados refere-se à carga útil (“payload”) o qual inclui tanto voz, vídeo ou dados. Tudo o que não for dados será referenciado como sendo controle ou sinalização,

o qual pode incluir endereçamento de rede com propósito de roteamento e códigos de correção de erros e de verificação. O termo conexão é utilizado para a comunicação na camada de aplicação, como por exemplo, o circuito que é estabelecido hoje em uma rede telefônica ou uma sessão de Telnet durante a qual nenhum circuito é estabelecido e as informações são transmitidas em pacotes.

4.2.1 Comutação por circuito

Na comutação por circuito existem três fases distintas: estabelecimento do circuito, transferência de dados e cancelamento do circuito. Na primeira fase, apenas informações de controle são trocadas para estabelecer um circuito entre a origem e o destino. Este circuito é um canal dedicado e possui uma banda fixa. Os comutadores intermediários também são configurados para encadear ou ligar os canais para formar um circuito único. Depois de estabelecido o circuito, inicia-se a segunda fase, na qual apenas os dados são transmitidos enquanto durar a conexão. Finalmente, depois que a transferência dos dados tiver sido feita, a terceira fase se inicia e o circuito é desfeito. A comutação por circuito é adequada para aplicações que requerem uma taxa de transmissão constante compatível com a capacidade de transmissão do canal. Uma vez que não há necessidade de nenhum processamento em nenhum comutador intermediário depois que o circuito é estabelecido, não há necessidade do uso de comutação rápida ("fast switching") (embora o seu uso possa ajudar a reduzir o tempo de estabelecimento do circuito) e também de nenhum mecanismo de armazenamento nos nós intermediários (exceção feita ao mecanismo de retardo necessário para o intercâmbio de "time slots". [13]. Dado que normalmente existem mais do que um caminho entre a origem e o destino, há necessidade de algum tipo de roteamento, o qual basicamente determina o caminho ou rota a ser seguido para ir da origem ao destino. Na comutação por circuito, o roteamento faz parte da fase de estabelecimento do circuito, cujo controle pode ser feito de forma centralizada ou distribuída. Uma variação da comutação por circuitos utilizada em sistema tipo TASI ("Time Assignment Speech Interpolation") é chamada de comutação rápida por circuito ("fast circuit switching"), onde a primeira fase envolve apenas roteamento e não estabelece o circuito propriamente dito. O estabelecimento do circuito, bem como o cancelamento do mesmo, é feito quando o início (ou o término) de uma rajada ("burst") é detectado pelo envio de um sinal de controle especial e é rápido ("fast") desde que o roteamento tenha sido feito.

Controle centralizado versus controle distribuído

No controle centralizado, existe uma entidade chamada de Controlador Central (CC) que mantém uma base de dados global da topologia da rede e das informações da utilização das conexões. Estas informações são fundamentais para a tomada de decisões de roteamento. Para exemplificar, o emissor envia um pedido de conexão contendo o endereço do nó destino (entre outras informações pertinentes) ao CC, o qual através de algum critério pré-definido, escolhe o caminho que o circuito deverá seguir. (e.g. um caminho com o menor número de conexões ou links).

No controle distribuído, cada nó pode ter um controlador semelhante ao CC mencionado no controle centralizado. No entanto, este controlador pode manter apenas informações locais (.e.g. status das conexões que chegam e que vão para os seus

vizinhos), e determinar que conexão de saída usar, tendo em vista que o próximo nó pode também fazer a sua escolha da melhor conexão e que este processo pode continuar até o nó destino ser alcançado.

É importante notar que independentemente do controle ser centralizado ou distribuído, um circuito pode não ser estabelecido na primeira tentativa quando ainda não existem recursos de rede disponíveis naquele momento. Neste caso, o pedido de estabelecimento de conexão pode ser retransmitido para o nó emissor tardiamente, ou pode ser armazenado em algum nó intermediário para então ser enviado para o próximo nó em algum momento oportuno.

4.2.2 Comutação por pacotes

A comutação por pacotes usa o mecanismo de controle distribuído de roteamento. A maior diferença entre a comutação por circuitos e a comutação de pacotes (mesmo que ambas utilizem controle distribuído de roteamento) reside no fato de que, na comutação por pacotes, os dados podem ser enviados sem que o circuito tenha sido estabelecido. Em outras palavras, isto quer dizer que na comutação por circuito, os dados (ou mensagem) são transmitidos em pacotes, no qual cada um deles contém um cabeçalho com alguma informação de controle. Eles são enviados para o nó destino através de nós intermediários via modelo de armazeana-e-envia ("*store-and-forward"*) e quando o pacote chega em um nó, ele é armazenado primeiro e depois que o cabeçalho do pacote é processado, ele é enviado ao próximo nó. Isto implica que na comutação por pacotes, o comutador é configurado depois que os dados (pacotes) chegam. Um pacote pode ter um tamanho fixo (como em pacotes de voz digitalizados), ou um tamanho variável ou um tamanho máximo.

Uma técnica semelhante é conhecida como comutação por mensagens ("*message-switching"*) na qual uma mensagem inteira (de tamanho grande) pode ser enviada com um único cabeçalho, reduzindo com isso o custo de se ter que quebrar a mensagem grande em várias mensagens menores. Entretanto, devido a natureza do encaminhamento existente nos routers ("*store-and-forward"*), seria necessário possuir unidades de armazenamento muito grandes ("*buffers"*) em cada um dos nós, ao contrário do que acontece quando o pacote é quebrado em pacotes menores. O tempo de transmissão também teria um sensível acréscimo, pois em cada nó, a mensagem toda deveria ser recebida para depois ser enviada para o próximo nó.

A comutação por pacotes é mais adequada para tráfico em rajadas ("*burst"*) do que a comutação por circuito, pois permite o uso estatístico do canal entre os pacotes para diferentes emissores e destinatários.

Datagrama e Circuito Virtual (VC)

Existem duas variações básicas na comutação por pacotes, uma baseada em datagrama e a outra baseada em circuitos virtuais (VCs). Na comutação por pacotes baseada em datagrama, o cabeçalho é similar a um pedido de estabelecimento de circuito (usando o mecanismo de controle distribuído) no qual ambos contém informações de controle similares e são processados de maneira semelhante em cada um dos nós intermediários. Entretanto, tanto o cabeçalho, quanto a carga útil do pacote são enviados

no mesmo canal, sem nenhum retardo entre eles, enquanto na comutação por circuito, o pedido de estabelecimento e os dados são enviados em canais distintos com um retardo igual ao tempo de estabelecimento do circuito (no mínimo o retardo do "*round-trip"*). É importante salientar que o protocolo IP usa comutação por pacotes baseado em datagrama.

Na comutação de pacotes baseada em circuito virtual (VC), existem duas fases distintas, uma delas é o estabelecimento do circuito virtual (ou o roteamento), similar ao que ocorre na comutação de circuito rápida ("*fast circuit switching"*), e a outra é o envio propriamente dito dos pacotes via o VC (ou comutação). É importante notar que no estabelecimento do VC não há necessidade de alocação de uma banda dedicada para cada conexão – apenas é criada uma entrada na tabela de comutação de cada nó intermediário no caminho selecionado. Esta tabela consiste basicamente de um mapeamento de um rótulo ("*label"*) de entrada (que na realidade é um número de identificação do VC) para uma porta de saída. Na segunda fase da comutação de pacotes baseada em VC, cada pacote contém um rótulo e quando este é idêntico ao que está mapeado na tabela de comutação, o pacote é direcionado para a porta de saída apropriada e provavelmente recebe um outro rótulo.

MPLS (Multiprotocol Label Switching)

A idéia do MPLS ("*Multiprotocol Label* Switching") [14][15], que atualmente é padrão definido pelo IETF ("*Internet Engineering Task Force"*), é semelhante ao circuito virtual baseado em comutação por pacotes. Ao invés das decisões de roteamento serem feitas para cada pacote, como acontece nas redes de comutação de pacotes orientados a datagrama, um caminho comutado por rótulo, ou LSP ("*Label-Switched Path"*) é criado, de forma análoga ao VC, de tal maneira que todas as decisões de roteamento são feitas uma única vez no emissor através da associação de um rótulo em cada pacote. Isto permite ao MPLS simplificar o envio de pacotes e a suportar roteamento explícito sem a necessidade de cada pacote ter que carregar uma rota explicita dentro dele.

O estabelecimento do LSP pode ser controlado pela rede de acordo com a sua topologia e sua conectividade usando a comutação por etiqueta ("*Tag* Switching") criada pela Cisco [16]. Ele pode também ser controlado por datagrama, por exemplo, em uma comutação IP onde o IP roda em cima de comutadores ATM, ao invés de se estabelecer um VC ATM entre o emissor e o destino como é feito em uma clássica solução de IP sobre ATM, um LSP é criado depois que alguns pacotes IP's de um fluxo tiverem sido enviados [17] (neste contexto, um fluxo pode se referir a todos os pacotes IP de um nó emissor para um nó destinatário. Esmiuçando um pouco mais, ele pode se referir aos pacotes pertencentes à mesma conexão TCP). Como exemplo, os primeiros pacotes IP serão roteados em cada nó intermediário; entretanto, assim que o nó destino reconhece o fluxo (i.e. quando o número de pacotes IP recebidos do nó emissor excede um valor limitante ("*threshold"*)), ele requisita o estabelecimento de um LSP na direção do nó emissor, isto é, cada nó antecessor associa um rótulo para ser usado (i.e. um rótulo de saída) pelos nós posteriores via um protocolo chamado LDP ("*Label Distribuition Protocol"*). Depois que o LSP é estabelecido, o nó emissor quebra cada pacote subseqüente de um fluxo em células cada uma com o seu rótulo apropriado. Estas células são então comutadas nos nós intermediários eliminando desta maneira a necessidade de roteamento em cada um dos nós intermediários. O exemplo dado foi baseado em IP sobre

ATM, mas o MPLS pode suportar múltiplos protocolos de rede sobre múltiplas camadas de enlace.

4.3 Comutação por rajadas

No exemplo ilustrado acima, rodar IP sobre uma rede grande de comutadores ATM causa uma perda aproximada de 10% em cada célula, visto que dos 53 bytes de cada célula ATM, 5 bytes são utilizados para cabeçalho. Atualmente existe uma proposta no ATM Fórum chamada de FAST (Framed ATM over SONET Transport), a qual permitirá que sejam transportados até 64 Kbytes de dados para cada 4 bytes de cabeçalho, minimizando e muito esta deficiência existente hoje.

Para reduzir a alta porcentagem de bytes de controle existentes em células ou pacotes pequenos, criou-se o conceito de comutação por rajadas (as quais podem ser entendidas como um jato ou jarro de voz digitalizada ou uma mensagem de dados).

4.3.1 Reserva em uma direção ("One-Way Reservation")

A comutação por rajadas pode ser enquadrada em três variações básicas, as quais não apenas alocam recursos da rede via uma base definida na rajada-por-rajada, mas também integra elementos da comutação por circuito e por pacotes. Estas variações estabelecem um circuito no qual a duração da rajada é baseada na reserva em uma direção do canal. Neste modelo, o nó emissor envia um pedido de estabelecimento, o qual é seguido por uma rajada antes de ter recebido o pacote de reconhecimento de nó destinatário. Isto reduz o retardo das pré-transmissões da rajada, pois em redes de alta velocidade (ex. Gbps), o tempo de transmissão da rajada pode ser relativamente curto.

As três variações básicas da comutação por rajadas diferem entre si na maneira que a banda (capacidade de transmissão alocada) alocada é liberada. Exemplificando, na variação "*tell-and-go"* (TAG), assim que a rajada é transmitida, o nó emissor envia um sinal explícito de liberação para cortar o circuito, da mesma maneira que é feito na fase 3 da comutação por circuito. Na reserva-com-duração-fixa ("*reserve-a-fixed-duration"* ou RFD), cada pedido de estabelecimento de circuito define também a duração do mesmo. Finalmente existe o terminador-dentro-da-banda ("*in-band-terminator"* ou IBT), no qual a rajada contém um cabeçalho e um terminador para indicar o final da rajada [17][18]. Desta maneira, a liberação da banda alocada em uma comutação por rajada através do TAG e do IBT é similar ao que acontece na comutação por pacotes e por circuitos, enquanto que o RFD é único.

4.3.2 Tempo de recuo e Comutação por corte ("Offset tié and Switch Cut-Through")

Apesar das semelhanças entre as variações TAG e IBT com a comutação por circuitos e por pacotes, é importante deixar claro que a comutação por rajada não é a mesma coisa que a comutação por pacotes nem por circuitos.

Primeiramente, na comutação por circuito, o estabelecimento de um circuito é um processo de duas fases envolvendo o envio de um pedido de estabelecimento de conexão e o recebimento de um reconhecimento desta conexão antes do início da transmissão dos

dados propriamente dita. Na reserva em uma direção_(“*One-way Reservation”*), não existe uma separação clara entre as fases 1 e 2 devido à falta de um pacote de reconhecimento antes do início da transmissão. Tanto no TAG, quanto no RFD, o tempo de recuo entre o pedido de estabelecimento de conexão e os dados correspondentes, representamos por **T**, é menor que o que ocorre na comutação por circuito. De fato, os dados podem ser enviados antes que o circuito todo tenha sido estabelecido (i.e. os últimos canais podem ser perdidos). Na maneira que este tempo de recuo **T** pode ser tão grande que quando os dados chegarem ao comutador, o comutador já estabeleceu a conexão de saída para outro comutador através de um pedido de estabelecimento de conexão, os dados não precisam ser armazenados em nenhum nó intermediário como acontece na comutação por circuito.

É importante notar que se **T** é muito pequeno (e.g. **T = 0**) no TAG e no RFD, ou caso o IBT seja usado, os dados precisaram ser armazenados (ou melhor, dizendo, eles precisaram sofrer um retardo) em um nó intermediário, digamos **X**, onde o circuito parcialmente temina, enquanto aguarda que o processamento do pedido de estabelecimento de conexão seja completado. Esta técnica é chamada de “virtual cut-through” [19], a qual é diferente da técnica armazena-e-envia (“*store-and-forward*”) existente na comutação por pacote ou por mensagem. É claro que no pior caso, no qual o retardo de processamento do pedido de estabelecimento de conexão é grande ou ocorrer congestionamento tal que o canal de saída não estava disponível, a rajada inteira pode ter que ser armazenada em um nó intermediário ou ser descartada, ou ainda ser desviada (i.e. roteada para uma porta de saída alternativa) caso não exista espaço de armazenamento suficiente para a rajada.

4.4 Comutação em redes ópticas

Neste item, iremos descrever as técnicas de comutação correspondentes propostas para as redes ópticas (especificamente as redes WDM) dando um visão das suas características únicas.

4.4.1 Roteamento por comprimento de onda (“Wavelength Routing”)

O roteamento por comprimento de onda é uma forma de comutação por circuito. Especificamente, nas redes roteadas por comprimentos de onda, um caminho de luz (“*lihtpath”*), que é um caminho de dados óptico, no qual não há necessidade de nenhuma conversão dos dados do meio Óptico/Elétrico/Óptico, é estabelecido antes que os dados possam ser enviados [20]. Esses “*lightpaths”* são chamados de “comprimentos de onda roteados” (“*wavelength-routed”*) porque cada um deles usa um canal de comprimento de onda dedicado em cada conexão física entre o nó emissor e o nó destinatário. O que determina como os dados serão roteados nos nós intermediários é a cor de cada comprimento de onda. O roteamento por comprimento de onda é baseado nas seguintes premissas. Primeiro, era esperado que a funcionalidade maior de uma camada WDM seria prover “*lighpaths”* entre dois dispositivos eletrônicos (e.g. SONET ADMs) que não estão fisicamente adjacentes, em outras palavras, estão separados por múltiplos (fibras) links interconectados com comutadores ópticos. Esses “*lighpaths”* além de proverem alta velocidade e alta capacidade de transmissão que é transparente a taxa de bit e formato de

codificação, também reduz o número de equipamentos eletrônicos caríssimos como os SONET ADMs, através da agregação de tráfego e algoritmos de atribuição de comprimento de onda. Segundo, os comutadores ópticos (roteadores de comprimentos de onda) baseados em tecnologia óptico-mecânica, óptico-acústico ou termo-óptico são muito lentos para serem usados na comutação de pacotes de forma eficiente

A única propriedade das redes roteadas por comprimento de onda (assim como redes ópticas tipo TDM) é que, devido ao fato da tecnologia de conversão de comprimento de onda ("*time-slot interchanging"*) não estar madura o suficiente, o "*lightpath"* pode ter que usar o mesmo comprimento de onda ("*time-slot"*) em links diferentes. Esta abordagem é chamada de multiplexação de links ("*link-multiplexing"* ou LM), se contrapondo à abordagem chamada de multiplexação de caminhos ("*path-multiplexing"* ou PM), onde diferentes comprimentos de onda ("*time-slots"*) podem ser utilizados em diferentes links [21].

Vale notar que com o aumento das aplicações de uso intenso de banda que estão aparecendo (e.g. "*High Definition Television Distribution"*), alguns usuários que consomem grandes quantidades de banda, poderão pedir um ou mais "*lightpaths"* por um período curto, digamos alguns minutos ou até mesmos algumas semanas, e com isso a camada WDM deve estabelecer e liberar dinamicamente caminhos de luz roteados por comprimento de onda, semelhantemente ao que ocorre hoje em dia nas redes telefônicas comutadas por circuito. Embora as mais recentes pesquisas no estabelecimento de "*ligthpaths"* assumam um controle centralizado, vêm sendo também investigado protocolo de reserva distribuída [22][23].

O controle distribuído pode melhorar a confiabilidade e escalabilidade das redes roteadas por comprimento de onda. No entanto, existem duas grandes diferenças no estabelecimento de um circuito no controle distribuído entre uma rede eletrônica e uma rede óptica. A primeira diferença já foi mencionada anteriormente, a qual um circuito em uma rede WDM pode ter que ser estabelecido via PM. Assim sendo, reservando-se múltiplos comprimentos de onda em cada enlace simultaneamente, produz uma maior possibilidade de sucesso do que reservando apenas um comprimento de onda de cada vez [22][23].

A segunda maior diferença é que nas redes eletrônicas é possível enviar uma informação de estado em cada enlace para todos os nós de tal maneira que cada nó tem o conhecimento global da rede, semelhantemente ao uso do OSPF ("*Open Shortest-Path First*") utilizado na Internet. Entretanto, isto pode não ser possível em uma rede WDM, especialmente para as que utilizam o PM, devido ao fato de cada comprimento de onda ser gerenciando (alocado e liberado) individualmente como uma unidade; e um link pode possuir inúmeras fibras, cada uma carregando até dez comprimentos de onda. Por conta disso os protocolos de reserva distribuídos baseados em conhecimento local, i.e. a informação de uso do comprimento de onda no link de saída de um nó [22], podem necessitar do uso dos protocolos baseados em conhecimento global [23].

O estabelecimento dinâmico de caminhos de luz pode ser feito no contexto do MPLS. Depois que um fluxo IP é reconhecido, um caminho de luz por ser estabelecido para todos os futuros pacotes deste fluxo [25].

Uma das limitações do roteamento por comprimento de onda é que ele é ineficiente para o tráfico Internet, que opera em rajadas, ou seja ele é estatístico. Além disso, existem poucos recursos para o controle de utilização de banda em um "*ligthpath"*.

Um outro limitante é que dado um certo apenas número pequeno de comprimentos de onda, apenas um número pequeno de caminhos de luz pode ser estabelecido ao mesmo tempo, o que resulta na conectividade de uma topologia virtual, em uma camada WDM, ser potencialmente fraca. Mesmo que os caminhos de luz sejam estabelecidos de forma dinâmica, o tempo de estabelecimento de conexão baseado na técnica reserva de duas vias ("*two-way reservations"*) pode ser muito longo para uma rajada contendo alguns Megabits de dados (e.g. um arquivo pequeno) devido a alta taxa de transmissão (e.g. 2.5 Gbps).

4.4.2 Comutação por rótulo, célula e pacotes ópticos

Devido às previsões que a quantidade de tráfego em rajadas irá em breve ultrapassar o tráfico de voz, e dado ao fenomenal sucesso da Internet (principalmente do WWW), é razoável esperar que a camada WDM irá também utilizar técnicas similares de comutação como as usadas na comutação por pacotes e por rajadas, quando tecnologias como as de "*Semiconductor Optical Amplifiers"* (SOA) e de "*Lithium Niobate"* se tornarem disponíveis em comutadores mais rápidos.

A comutação por pacote óptica é similar à comutação por pacote tradicional, excetuando-se que os dados ("*payload"*) deverão permanecer no ambiente óptico e que o cabeçalho ("*header"*) poderá ser processado eletronicamente ou também no ambiente óptico [26]. Entretanto, apenas um processamento óptico limitado é hoje possível devido a atual tecnologia nesta área, o que torna a comutação óptica de pacotes baseado em circuitos virtuais mais atrativa do que a comutação óptica de pacotes baseada em datagrama.

Na realidade, mesmo no processamento eletrônico (e.g. comutação fotônica), o tempo de processamento do cabeçalho deve ser mantido o menor possível. Adicionalmente, cada pacote deve ter um tamanho fixo (o preenchimento ("*padding"*) é usado em caso de dados insuficientes) e deve ser também muito pequeno. Esses dois quesitos se devem ao fato de não haver equivalência óptica à memória de acesso randômico (RAM), e também pelos seguintes fatos: 1) um sinal de dados óptico pode ser retardado em um limitado período de tempo através do uso de linhas de retardo de fibras ópticas (LDLs) [27] antes que o processamento do cabeçalho tenha se completado, e 2) o tamanho de cada pacote, em termos do produto do seu tempo de transmissão e a velocidade da luz, não pode exceder o que está disponível via FDL para que o pacote óptico possa ser "armazenado". Essas são as razões para a comutação óptica de pacotes usar pacotes com tamanho fixo.

Quando os pacotes são do mesmo tamanho, é natural o uso do TDM onde o eixo do tempo é dividido em pedaços de tempo. Mesmo sem usar o TDM, a sincronização é um grande problema na comutação de pacotes óptica desde que cada nó precisa reconhecer o cabeçalho e o fim do pacote e alinhar os pacotes em diferentes portas de entrada e de saída antes e depois deles serem comutados.

Devemos notar que, tradicionalmente, a comutação por pacotes assume que cada pacote está sendo transmitido na capacidade máxima de transmissão da banda, implicando com isso uma alta taxa de bits (e.g. 40 Gbps em um link de fibra óptica usando TDM) tanto para o cabeçalho ("*header"*) quanto para os dados ("*payload"*). A comutação de pacotes óptica tem sido adaptada para utilizar tecnologia WDM para

transmitir pacotes na capacidade de transmissão que um comprimento de onda ("*wavelength"*) pode oferecer. Na prática para facilitar as implementações, os cabeçalhos podem ser transmitidos em comprimentos de onda separados em subcanais. O uso de controle *"out-of-band"* é uma característica encontrada na comutação por rajada, já descrita anteriormente. Atualmente, as propostas de comutação por rótulo óptica [28], as quais usam SCM para levar os rótulos dos pacotes IP's (de tamanho variável) e entregar os pacotes IP assim que o rótulo é reconhecido (e re-escrito quando necessário), deixa uma névoa na distinção entre a comutação por pacotes óptica e a comutação por rajada óptica.

4.4.3 Comutação por rajada óptica (OBS)

A comutação por rajada óptica ("*Optical Burst Switching"* – OBS) é o caminho para o balanceamento entre o roteamento por comprimento de onda de granulação grossa e a comutação por pacotes de células ópticas. Tendo em vista a dificuldade de opticamente reconhecer o final de cada rajada, a comutação por rajada óptica (OBS) é baseada nas técnicas RFD ("*Reserve-a-Fix*-Duration) e TAG ("*Tell-Ang-Go"*) (ao invés da IBT – "*in-bound terminator"*). Um novo protocolo para comutação por rajada óptica (OBS) baseado em RFD chamado JET ("*Just Enough-Time"*) foi proposto por Yoo, Jeong e Qiao [29-31] e será descrito, resumidamente, abaixo.

A sinalização fora de banda é utilizada na comutação óptica por rajada (OBS). Na comutação óptica por rajada, através do protocolo JET, os comprimentos de onda em um link utilizado para a rajada será desfeito automaticamente de acordo com a reserva feita assim que a rajada passa pelo enlace. Desta maneira rajadas de diferentes emissores para diferentes destinos podem efetivamente utilizar a banda de um mesmo comprimento de onda em um enlace compartilhando o tempo (multiplexação estatística de uma certa maneira). Caso o controle de pacotes falhe na reserva de banda em um nó intermediário, a rajada (a qual é considerada bloqueada neste período) pode ser descartada. A OBS pode suportar tanto transmissões de rajada confiáveis e não-confiáveis na camada óptica. É claro que quando se utiliza protocolo de camada superior que já possua controle de transmissão, deixa-se para este protocolo a responsabilidade de retransmitir os dados perdidos.

Pode-se notar que uma maneira de suporte IP sobre WDM usando comutação óptica por rajada é "rodar" o IP em cima de cada comutador óptico, como parte da interface entre o roteador na camada eletrônica e o comutador WDM na camada óptica. Na camada WDM, um controle dedicado de comprimento de onda é usado para prover a conexão física entre entidades IP. Especificamente, ele é usado para suportar comutação por pacotes entre entidades IP fisicamente adjacentes às quais mantêm a topologia da rede e as tabelas de roteamento. Para enviar os dados, o pacote de controle é roteado do emissor para o destino, o qual estabelece uma conexão através da configuração de todos os comutadores ópticos entre os dois pontos. Feito isso, uma rajada (i.e., um ou mais pacotes de IP ou qualquer mensagem inteira) é entregue sem a necessidade de ir através de nós intermediários de entidades IP, com isso reduzindo a latência assim como o tratamento (carregamento) do processo nos roteadores.

4.5 Resumo conclusivo do Capítulo

Nesta seção foram descritas algumas técnicas de comutação utilizadas atualmente nas redes de voz e de dados, assim como algumas que são promessas para as redes ópticas transparentes do futuro. Estas técnicas de comutação podem ser classificadas em três categorias: comutação por circuito, comutação por rajada e comutação por pacotes. A correspondência destas técnicas na camada WDM é: roteamento por comprimento de onda, comutação por rajada óptica e comutação por pacote óptico, muito embora as distinções entre as duas últimas têm ficado cada vez mais obscuras tendo em vista que a comutação por pacotes óptica começa a utilizar comutação "*cut-through*" e controle fora de banda para comutar pacotes de comprimentos variável. A Figura 4.1 faz um resumo conceitual das técnicas de comutação e a Figura 4.2 mostra os relacionamentos entre os três paradigmas de comutação óptica.

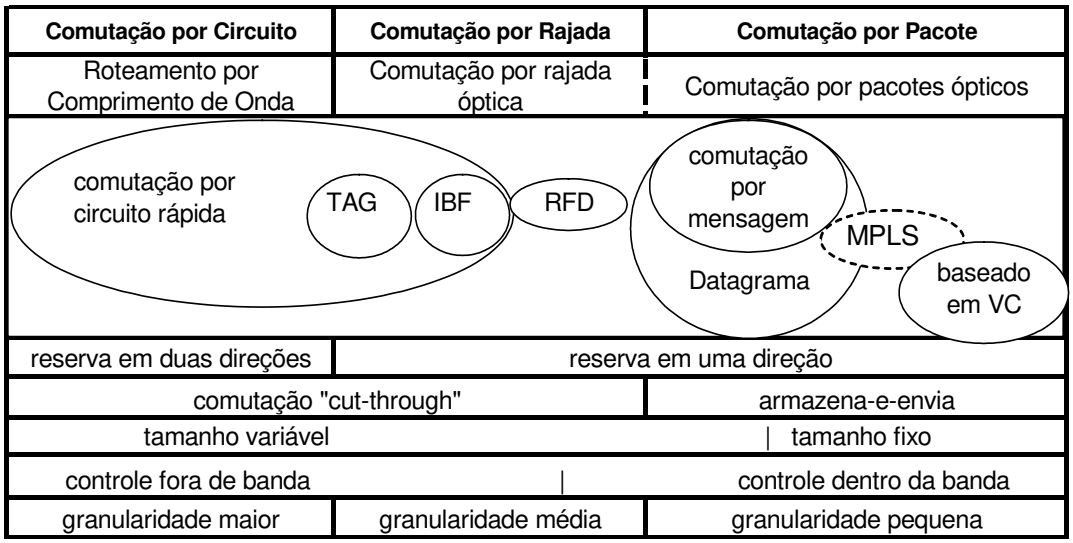


Figura 4.1 – As três maiores técnicas de comutação

Comutação Óptica (paradigma)	Utilização de Banda Passante	Latência (set-up)	Buffer Óptico	Overhead Proc./Sync.	Adaptabilidade (Tráfego & Falha)
Circuito	baixa	alta	não requerido	baixo	baixa
Pacote/Célula	alta	baixa	requerido	alto	alta
Rajada	alta	baixa	não requerido	baixo	alta

Figura 4.2 – Comparação entre os paradigmas da comutação óptica.

Percebe-se que nas aplicações correntes e nas tecnologias que estão sendo testadas, a questão se haverá uma única técnica de comutação (e caso exista, qual delas) em uso na camada eletrônica e/ou na camada óptica (WDM) não está definida ainda; muito pelo contrário, o debate está em aberto.

CAPÍTULO 5

REDES IP SOBRE DWDM – UMA ANÁLISE SOBRE A CA*NET

Em 1998, a CANARIE entregou a rede CA*net 3 (Fig 5.1), a primeira rede internet óptica do mundo de pesquisa de educação. A rede CA*net 3 estava entre as mais avançadas no mundo quando foi construída, e seu projeto tem sido reproduzido desde então por muitas empresas operadoras de rede, tanto para fins comerciais, quanto para as áreas de pesquisa e educação. Devido ao crescimento exponencial do tráfico de rede, ao crescimento das novas aplicações para banda larga e à própria maturação do projeto CA*net3, houve necessidade na criação de uma nova rede que apoiasse as realidade atual do desenvolvimento, assim como apoiasse a pesquisa tecnológica no Canadá. Por conta disso, o Governo de Canadá investiu US$ 110 milhões de dólares junto à CANARIE para o projeto, desenvolvimento e operação da rede CA*net 4.

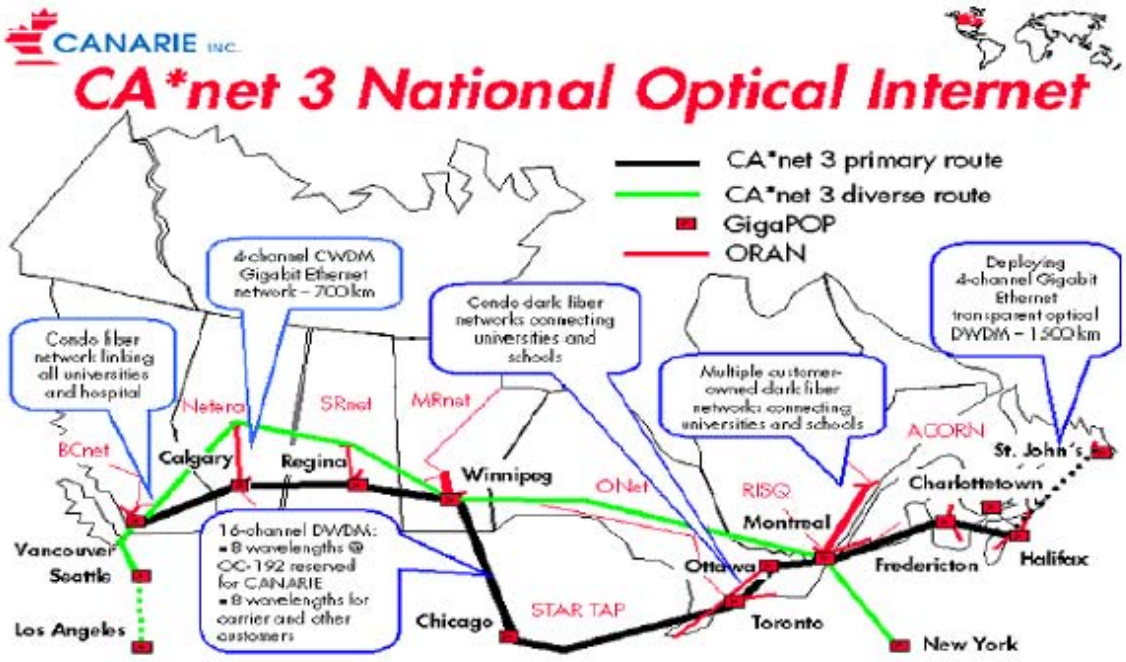


Figura 5.1 - Rede óptica canadense CA*net 3 (1998)

A rede CA*net 4, assim como fez a sua antecessora a rede CA*net 3, interconecta as redes de pesquisa das províncias canadenses (semelhantes aos nossos Estados), e, dentro dela interconecta, as universidades, os centros de pesquisa, os laboratórios de pesquisas do governo, as escolas, e outros locais de interesse, todos entre si, assim como a outras redes internacionais. Através de uma série de comprimentos de onda ópticos

ponto-a-ponto, a maioria dos quais aprovisionados em velocidades de OC-192 (SONET) ou STM-64 (SDH) (10 Gbps) , a rede CA*net 4 tem sua capacidade de rede inicial total posicionada entre quatro a oito vezes maior que a rede CA*net 3.

A rede CA*net 4 incorpora o conceito de uma rede "orientada às necessidades dos usuários" que oferece alocação dinâmica de recursos de rede nas mãos dos usuários o que permite, aos usuários, uma maior inovação no desenvolvimento de aplicações de redes. Estas aplicações, baseado no uso crescente de computadores e das redes como plataforma para pesquisa em muitos campos do conhecimento, são essenciais para a colaboração entre os pesquisadores tanto em âmbito nacional, quanto internacional, tais como: análise e processamento de dados, computação distribuída, e controle remoto de instrumentação. Nas seções a seguir, discutem-se algumas questões levantadas durante o desenvolvimento da CA*net 4. Para tanto, algumas definições sobre os princípios utilizados serão introduzidas para que os resultados sejam apresentados posteriormente. O conceito P2P em redes ópticas e a arquitetura WebServices são descritos e em seguida o roteamento óptico através do BGP será introduzido.

5.1 Rede óptica Par-a-Par (P2P Optical Network)

5.1.1 Arquitetura

A forma de comunicação par-a-par existe em grande escala na Internet. Ao nível do AS ("*Autonomous System*"), a arquitetura de rede hierárquica baseada na rota default co-existe com a arquitetura de rede par-a-par. Foi proposta, recentemente, uma nova arquitetura de computação distribuída, a qual é conhecida por vários nomes: computação par-a-par ("*peer-to-peer computing*") ou redes par-a-par (*"peer-to-peer networking*") ou simplesmente P2P [33].

A evolução das arquiteturas de computação distribuída inspira a idéia da arquitetura de redes ópticas P2P. Antes do advento do PC, e mesmo durante a era do mini-computador, a computação foi dominada por grandes empresas e escritórios de serviços de computação. O escritório de serviço central tinha, normalmente, grandes redes de terminais que se conectavam ao computador central. As arquiteturas de computação naquele tempo eram focadas na construção de grandes computadores centrais em torno de grandes edifícios e escritórios e operados por um time de engenheiros e técnicos para operarem as máquinas e para fazer a administração central de software.

Uma quantidade considerável de pesquisas foi direcionada para pesquisas sobre como resolver problemas de tempo compartilhado, gerenciamento de CPU, priorização de atividades computacionais etc. A introdução dos primeiros mini-computadores nas universidades e nos centros de pesquisa mudou, de forma fundamental, o conceito de computação centralizada. A computação e as telecomunicações se tornaram um ativo no lugar de um serviço. A partir deste momento os pesquisadores puderam adaptar os computadores para novas e inovadoras aplicações e processos sem se preocupar com assuntos complexos relacionados ao compartilhamento de recursos de grandes sistemas computacionais tipo "mainframe".

As arquiteturas de redes de telecomunicação atuais estão construídas em torno de um conceito de uma rede central administrada por uma operadora, muito parecido com a arquitetura utilizada nos sistemas de computação.

Na arquitetura de roteamento existente na Internet, os AS's do núcleo da rede, ou seja, dos grandes provedores de IP e nos pontos de interconexão, carregam a tabela completa dos endereços de toda a Internet. Os outros AS usem rotas default interligando-os à hierarquia existente dos provedores, permitindo que trafeguem subconjuntos dos endereços IP atribuídos a cada AS. Quando dois AS se interconectam, eles estabelecem um acordo de troca de tráfego ("*peering agreement*"). Se ambos estão no mesmo nível de hierarquia, este acordo é simplesmente um acordo de troca de informações de roteamento. Entretanto, quando um AS está em uma hierarquia menor ("*downstream*"), normalmente este AS entra em relacionamento com o ISP de maior hierarquia ("*upstream*") para fornecer serviços de trânsito. Um ISP pode instalar um roteador em locais denominados pontos de troca de tráfego Internet ("*Internet eXanges* – IXs"). Alguns IXs oferecem Servidores de Rotas para facilitar a distribuição de informações de rotas entre os ISP conectados. Ao invés de cada ISP estabelecer uma sessão de protocolo de roteamento com cada um dos participantes, ele estabelece uma única sessão com o Servidor de Rotas. O Servidor de Rotas então distribui as informações aprendidas entre os ISP's que estão no IXs.

Os protocolos de roteamento interdomínios permitem que dois AS's diferentes troquem informações de roteamento de tal maneira que os dados possam ser enviados para fora da área do AS através do seu roteador de borda. Tendo em vista esta possibilidade os AS's não mais precisam ser organizados rigidamente em uma forma hierárquica. Dentre os serviços que um AS pode oferecer a outro estão os serviços de trânsito restrito. Estes serviços de trânsito restrito referem-se ao tráfego destinado a um AS específico e/ou o tráfego originado de algum AS, entre outros. Isto demonstra que existe uma co-existência entre a arquitetura hierárquica baseada na rota default e a arquitetura de rede P2P ao nível dos AS's na Intenet.

A computação par-a-par ou redes P2P foi proposta como um novo modelo de computação distribuída. Além das aplicações do dia-a-dia, ele é bastante útil para aplicações científicas e de engenharia. Na Internet atual têm-se dois exemplos de aplicações usando computação distribuída através do uso compartilhado de arquivos, o Napster e o Gnutella. No Napster, um índice (isto é um arquivo de diretórios) provê endereços para recursos disponíveis no momento, de maneira que um nó membro possa iniciar a conexão direta com qualquer outro nó membro que no momento possua as informações requeridas. Os nós individuais acessam o nó central para pesquisar qual o nó que no momento possui a informações que se quer e depois inicia o processo de conexão P2P com este nó para acessar o conteúdo [34]. O Gnutella elimina a necessidade do servidor central de arquivos de diretórios existente no Napster. Cada nó mantém um índice parcial dos nós membros assim como ele aprende endereços de sites conhecidos via a web ou através de listas de e-mails. Uma busca por conteúdo se inicia nestes índices e se propaga nos diretórios encontrados em outros nós. Este sistema é menos eficiente na busca e na recuperação, pois as mensagens com os pedidos se replicam rapidamente e amplamente pela web. Entretanto, ele é muito menos vulnerável a uma falha do nó central e não irá sobrecarregar um único nó excessivamente.

Uma definição para redes P2P pode ser entendida como "Uma arquitetura de rede distribuída pode ser chamada de redes P2P, caso os participantes usem de forma compartilhada parte dos seus recursos de hardware próprios, e.g., poder de processamento, capacidade de armazenamento, capacidade de conexão de rede ("*links*"), impressoras etc. Estes recursos compartilhados são necessários para oferecerem serviços e conteúdo para a rede, e.g., arquivos compartilhados ou áreas de trabalho colaborativas. Eles são acessíveis diretamente a outros nós, sem a necessidade de entidades intermediárias. Os participantes desta rede são provedores de recursos assim como receptores de recursos (serviços e conteúdo)." [35]. Pelo fato do Napster empregar um nó central como um diretório de arquivos, ele é considerado como uma rede P2P híbrida; enquanto o Gnutella com sua natureza de distribuição é considerado uma rede P2P pura. As redes P2P diferenciam-se da tradicional arquitetura cliente-servidor pelo fato de, neste último, o cliente apenas fazer requisições para o servidor, sem compartilhar nenhum dos seus recursos.

As redes de transportes foram criadas, desde o início, tendo como modelo a arquitetura de rede cliente-servidor. Com o desenvolvimento e instalação de novas gerações de equipamentos de redes de transportes e com a desregulamentação das telecomunicações, é possível para alguns domínios clientes construir redes de transporte privadas entre eles através do modelo de arquitetura de redes P2P. Dentro desta arquitetura, os domínios clientes têm possibilidade de trocar tráfego de forma mais eficiente direcionando as conexões na camada física. Comparado ao serviço confiável provido pelas conexões permanentes e semipermanentes utilizados no SONET/SDH ou nos caminhos de luz, eles têm mais controle na conectividade entre eles e portanto mais flexibilidade, o que torna possível a criação de aplicações inovadoras. O uso maciço de conexões P2P provê mecanismos tolerantes a falhas e restauração e proteção dos circuitos de forma autônoma.

Deficiência da arquitetura cliente-servidor nas redes ópticas

Dentro da arquitetura cliente-servidor, existem várias tecnologias para o controle de redes ópticas tais como o MPLS Generalizado (GMPLS) e O-UNI.(Interface Óptica entre o Usuário a Rede). Estes protocolos têm um papel muito importante para as operadoras, especificamente para aquelas nas quais existem milhares de conexões e acordos de prestação de serviço. No entanto, para um número pequeno de conexões, outras estratégias que não requerem uma rede gerenciada centralmente podem ser utilizados. Isto se aplica para as redes de pesquisa e educação as quais normalmente têm o conhecimento e capacitação técnica para gerenciar suas redes ópticas.

A experiência das universidades e dos centros de pesquisa na utilização da arquitetura cliente-servidor em redes ópticas mostrou as deficiências do modelo.

Primeiramente, os clientes não podem mudar a topologia e a largura da banda da sua Rede Virtual Private - VPN ("Virtual Private Network") de forma independente. Para que um cliente possa fazer alguma mudança na sua rede ele tem que liberar os recursos que está utilizando e requisitar outros recursos junto a operadora. No início do uso das VPN's, isto não se caracterizou como um problema, tendo em vista a longa vida útil de uma VPN. Mas para projetos maiores onde uma grande quantidade de dados é transferida para vários locais distintos em forma de rajadas, a capacidade para alterar dinamicamente

a largura de banda e a topologia da rede, sem a necessidade de re-sinalização é um requisito importante.

Secundariamente, clientes de diferentes VPNs não podem ser conectados "em cruz" dentro da nuvem da operadora; tendo em vista que os serviços de VPN ópticos são válidos apenas nos limites entre as extremidades da nuvem. Estas conexões "em cruz", se necessárias, devem ser feitas fora da nuvem.

Em terceiro lugar, as VPNs ópticas não podem estabelecer conexões através de múltiplos domínios administrados independentemente. Existem alguns trabalhos, ainda em desenvolvimento, para criar estruturas padrão para protocolos de Interface Rede-a - Rede – NNI ("Network to Network Interface"). No entanto, a maioria destes desenvolvimentos de protocolos está focado em interconexão de caminhos de luz entre as redes das operadoras. Na realidade hoje, um cliente de domínio não pode ter múltiplas operadoras provendo conexões de forma colaborativa, tendo em vista o foco do negócio de cada uma.

Atualmente, apenas instituições de pesquisa e universidades estão confortáveis em administrar e gerenciar redes P2P e para controlar as conexões BGP multi-homed. A arquitetura cliente-servidor não está apta para apoiar o paradigma P2P em redes ópticas como ele é feito na Internet.

Características da arquitetura de redes ópticas P2P

A arquitetura cliente-servidor é assimétrica em termos de acesso de recurso. Dentro da arquitetura cliente-servidor, um domínio de cliente nunca permite que outros domínios acessem suas conexões internas ou as conexões que o mesmo possua com outros provedores, com exceção do acesso a clientes finais que por ventura estejam dentro da sua rede.

As novas tecnologias abrem a possibilidade para domínios de cliente colaborarem entre si para construir sua rede de transporte usando a rede de arquitetura P2P. Na camada física, a tecnologia SONET permite que os usuários finais utilizem taxa de transmissões variáveis para construir VPNs. Com o desenvolvimento do WDM em redes ópticas, os provedores estão começando a oferecer um novo tipo de serviço de taxa variada de bits baseando-se no serviço de comprimento de onda. Através disso, domínios funcionalmente diferentes estão conectados como se estivessem utilizando fibras dedicadas para sua interconexão. Uma outra situação atual é que em algumas áreas metropolitanas está surgindo uma grande quantidade de fibras apagadas. Domínios podem ser conectados diretamente através de fibras escuras. No plano de gerenciamento e controle, o evento da orientação a objeto ou do protocolo WebService tais como JXTA, J2EE, Common Object Request Broker Architecture(CORBA), Simple Object Access Protocol (SOAP), eXtensible Markup Language (XML) e o JINI abriu novas possibilidades para o projeto e implementação de sistemas de administração de rede.

O modelo P2P é proposto como um dos modelos de controle na arquitetura cliente-servidor especificamente para o protocolo IP rodando em cima de redes ópticas. Ele sugere uma extensão dos protocolos de controle intra-domínio como os protocolos de roteamento e os protocolos de sinalização além dos limites da camada óptica. A idéia é utilizá-los como um método unificado para controlar a camada óptica e a camada de IP.

A arquitetura cliente-servidor é assimétrica em termos de controle de recursos da rede em qualquer modelo de controle utilizado. Na Figura 5.2 pode-se ver que os clientes de domínios utilizam os serviços de transporte do provedor de serviço para falarem entre si. Um cliente de domínio não provê serviços de transportes a seu provedor de serviço para que este possa alcançar um outro domínio parceiro. Na mesma linha de raciocínio, cliente de domínios de rede conectado diretamente a um provedor de serviço não suporta a arquitetura P2P. Comparado à arquitetura de cliente-servidor, a arquitetura P2P tem duas características chave.

Primeiramente, cada domínio não só recebe serviços de transporte de outros participantes, mas também contribui com serviços de transporte novos a outros domínios. Secundariamente, uma ligação entre dois domínios é igualmente controlada por ambos ao invés de ser controlado como uma ligação de acesso onde um provedor de serviço tem um papel ativo da rede enquanto um domínio de cliente tem um papel passivo.

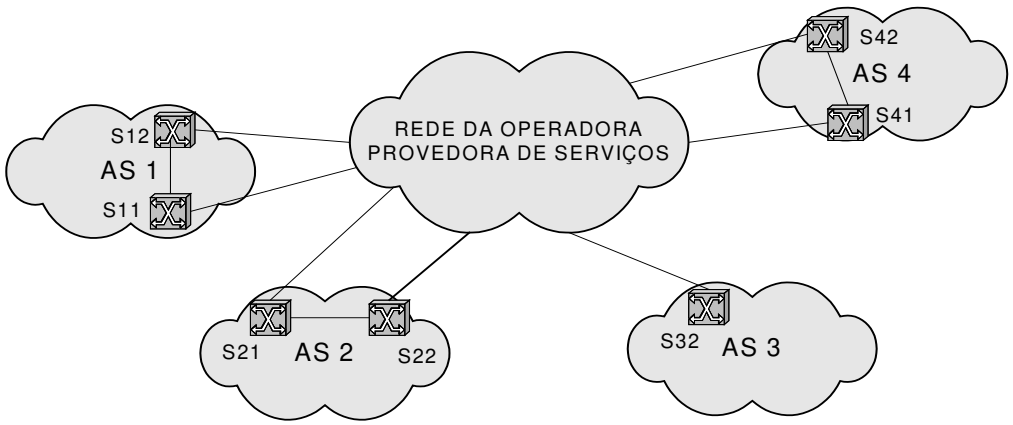


Figura 5.2. Arquitetura Cliente-Servidor Padrão

A arquitetura de redes óptica P2P (Figura 5.3) é àquela na qual múltiplos domínios de rede ópticos controlam as ligações igualmente entre si sem um controle centralizado e mutuamente provêm serviços de trânsito um ao outro baseado em uma política de acesso aberta.

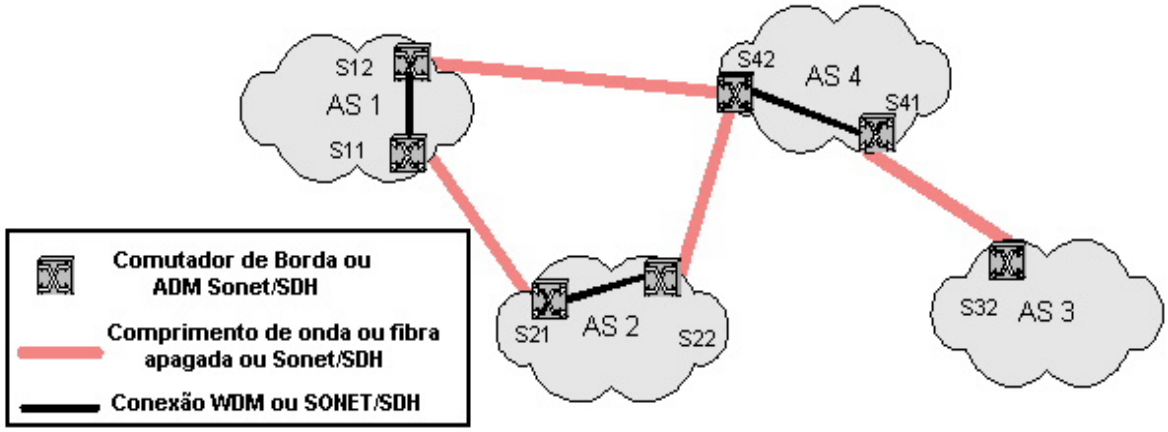


Figure 5.3. Arquitetura de redes óptica P2P.

5.1.2 Requisitos funcionais

As redes de transportes, dentre as quais se incluem as redes de backbone ópticos, foram construídas baseadas no modelo cliente-servidor (arquitetura de redes) desde o início. Com o desenvolvimento e a operacionalização de equipamentos de rede de nova geração e a desregulamentação das comunicações, é possível para alguns clientes de domínios construírem redes privativas de transporte entre estes domínios através do modelo de arquitetura de redes P2P. Com isso, os clientes de domínio podem trocar tráfego de uma forma mais eficiente através da troca direta com outro domínio através da camada física. Tendo mais poder de controle sobre a conectividade, significa ter mais poder de flexibilidade, o que, em última análise, viabiliza o desenvolvimento de novas aplicações.

A arquitetura de redes ópticas P2P conectar domínios e oferece controle igualitário entre as conexões sem um controle centralizado e provê serviço de trânsito entre os domínios baseados em uma política aberta de acesso. A diferença entre o modelo da arquitetura de rede P2P e o modelo de igual posição ("*peer model"*) será detalhada a seguir. O "*peer model*" foi proposto como um modelo cliente-servidor na arquitetura de rede. Ele sugere protocolos de controle intra-domínios como os protocolos de roteamentos e de sinalização os quais ultrapassam os limites de um provedor de serviços e se estende de forma única em toda a rede. Ele funciona paralelamente ao modelo de camadas ("*overlay model*") e o controle dos recursos da rede é assimétrico, independente do modelo utilizado. Comparada ao "*overlay model*" a arquitetura de rede P2P possui duas funções importantes.A primeira é que cada domínio além de receber serviços de transportes de outros domínios, também contribuí com novos serviços para outros domínios. A segunda é que uma conexão entre dois domínios é igualmente controlada por

ambos os domínios e não mais por um provedor de acesso que atuaria de forma ativa e os clientes de domínios de forma passiva. Há uma mudança de paradigma fenomenal.

Durante o estabelecimento de uma conexão fim-a-fim, em uma arquitetura de rede P2P, o requisito único nesta conexão é que cada segmento, entre os domínios participantes, tenha idêntica autoridade em relação ao seu controle. É claro que há necessidade da existência de um centro de inteligência e de resolução de conflitos, mas o gerenciamento do dia-a-dia e o controle das conexões são feitos de forma descentralizada.

Protocolos de descoberta de recursos e de roteamento interdomínio

Diferentemente do roteamento interdomínio existente na Internet onde apenas informações de conectividade lógica são disseminadas, as redes ópticas P2P requerem informações adicionais tais como disponibilidade de canais (i.e. inativo ou ocupado) ou comprimentos de ondas a serem dinamicamente disseminados. Isto pode ser explicado pela diferença entre a natureza da comutação por pacotes da Internet e a natureza de comutação de circuito em redes de transmissão ópticas P2P. Nas redes baseadas em tecnologia SONET, as informações de disponibilidade podem se agregar no número de canais disponíveis a uma taxa de dados específica. Nas redes baseadas em tecnologia WDM, alguns esquemas de assinalamento de comprimentos de ondas precisam de informações de disponibilidade de cada canal de comprimento de onda em todo a conexão ("*link"*).

A capacidade de interação de um protocolo de roteamento inter-domínio com o protocolo de roteamento intra-domínio é um outro requisito funcional. Esta interação é um processo mão dupla. Por um lado, um mecanismo precisa ser introduzido para que informações de roteamento inter-domínio sejam aprendidas fora do domínio, sendo levadas através dos nós. Por outro lado, um mecanismo é requerido para injetar informações de roteamento intra-domínio dentro dos protocolos de roteamento inter-domínio. Assim como na Internet, a disseminação da topologia interna de uma rede óptica para os nós externos é estritamente controlada, porém em um nível mais abstrato.

Mecanismos de sinalização simétricos interdomínios

Como a rede óptica P2P opera em um modo orientado à conexão, é necessária a existência de um mecanismo de sinalização para estabelecer, terminar ou manter as conexões.

Este é um requisito funcional diferente da Internet a qual opera em um modo não orientado à conexão. O MPLS generalizado (GMPLS) atua como um mecanismo de sinalização intradomínio no controle de redes ópticas. A sinalização O-UNI ("*Optical-User Nework Interface*") é feita em modo assimétrico e é utilizada entre o domínio cliente e o provedor do serviço [32]. Como os domínios, nas redes ópticas P2P, podem operar tanto como clientes, quanto como provedores de serviços e que os pedidos de conexão podem ser propagados quando necessário, há necessidade de um novo mecanismo de sinalização simétrico. Quando um domínio faz um pedido de conexão para um segundo domínio, eles estarão participando de uma relação cliente-servidor, na qual o primeiro domínio é o cliente ou o gerador do pedido de conexão. A entidade sinalizadora do primeiro domínio atua como cliente e a entidade sinalizadora do segundo domínio

atua como servidor no sentido de enviar para o cliente uma confirmação ou uma notificação de erro. O mecanismo de sinalização precisa prever também a situação na qual o segundo domínio faz um pedido de conexão, ou seja, passa a atuar como cliente, ao primeiro domínio, que passa a atuar como servidor. Por tudo isso, o mecanismo de sinalização interdomínio em redes ópticas P2P precisa ser integrado/sincronizado com o mecanismo de sinalização intradomínio; além disso mensagens de sinalização precisam ser traduzidas e enviadas entre as entidades interdomínios e intradomínios nos comutadores ópticos de borda ou nos distribuidores SONET/SDH. Em termos de proteção e segurança dos domínios há necessidade de se implementar mecanismos de autenticação.

Controle autônomo de bloqueio

A gerência de recursos de rede usados de forma compartilhada é muito diferente da gerência de recursos em rede não compartilhada. O uso de recursos de rede compartilhados em um modo distribuído precisa resolver os problemas de contenção. Na Internet, as contenções ocorrem quando há problemas de congestionamento de rede e estes problemas são resolvidos pelos mecanismos de controle de fluxo das camadas superiores como o controle de fluxo do TCP. Nas redes ópticas as contenções ocorrem quando um pedido de conexão chega a um link o qual não tem recursos disponíveis.

Tendo em vista o problema da contenção, em redes de recursos compartilhados, há necessidade de um mecanismo de controle de bloqueio distribuído, pois nas redes ópticas P2P cada domínio participante tem igual autoridade para utilização dos recursos existente e o mecanismo de controle de bloqueio tem que operar de forma autônoma. Este mecanismo autônomo deve se basear no princípio de que cada domínio participante pode estabelecer suas políticas na rede como um todo, levando-se em conta a cooperação e a boa intenção de cada um dos membros da comunidade.

Na Internet existem três (3) estratégias principais para efetivamente construir relações cooperativas e que possam gerenciar possíveis contenções. Quando existem recursos suficientes na rede para executar todos os pedidos, não há necessidade da aplicação de nenhum mecanismo de controle de bloqueio, mesmo que não haja muita cooperação entre os usuários destas redes. Na prática este era o "modus operandi" no início das redes ópticas P2P. O ponto crucial era o estímulo à utilização da rede sem se preocupar muito com a Qualidade de Serviço. Quando a rede não consegue atender aos requisitos e a um conjunto de Qualidade de Serviço pedido, o controle de bloqueio precisa ser introduzido. Como não existe um controle central, um mecanismo de controle autônomo de bloqueio precisa ser inserido. Este mecanismo nada mais é do que uma maneira para que entidades individuais possam gerenciar recursos compartilhados sem a necessidade de um coordenador central. Quando acontecem os bloqueios em um enlace, todas as conexões que usam este enlace devem ser notificadas. Feito isso, as conexões devem retroceder no nível de utilização deste enlace. Através do controle de bloqueio autônomo não apenas a utilização de sua banda é otimizada, mas a rede toda opera com mais eficiência. O uso do controle de bloqueio autônomo e a sua implementação em todos os domínios da rede mostram um comportamento educado e bem disciplinado na rede como um todo. Como resultado deste comportamento, todos os domínios têm acesso à Qualidade de Serviço.

Embora nenhum controle centralizado esteja em uso nas redes ópticas P2P, cada domínio tem que auditar e armazenar as operações de rede de seu interesse. Caso aconteça algum mal-comportamento na rede, um comitê administrativo central deve analisar a situação e tomar as devidas providências.

Camada física de interconexão de redes

Em redes ópticas P2P, todos os domínios são considerados iguais, mas eles não podem utilizar ao mesmo tempo as mesmas capacidades na camada física. Os nós de borda nos domínios podem possuir capacidades de comutação e de transmissão diferentes. Alguns domínios são capazes de fazer apenas conexão cruzada; alguns outros têm a capacidade de fazer comutação de comprimento de onda, alguns outros têm capacidade de fazer comutação por TDM e outros têm capacidade de fazer a distribuição de capacidade ("*add/drop*").

Para poder interconectar domínios com tecnologias heterogêneas na camada física e/ou na camada de dados, as informações de controle tanto ao nível de protocolos de roteamento, quanto aos protocolos de sinalização devem ser melhoradas.

Em redes ópticas P2P, múltiplos domínios de rede óptica conectados têm igual controle sobre os recursos da rede, sem um controle centralizado e todos os domínios provê serviço de trânsito entre si, tendo como base uma política de acesso. Este é uma nova arquitetura para construção de redes ópticas entre domínios com acordos mútuos e benefícios. Todo o seu potencial não está explorado ainda. Neste modelo, uma central de inteligência e de julgamento (ou decisão) deve ser criada para resolução de conflitos. No dia-a-dia, todo o gerenciamento e o controle das conexões devem ser descentralizados. Especificamente quatro (4) requisitos funcionais são identificados: a descoberta dos recursos e o protocolo de roteamento interdomínios, mecanismos de sinalização simétrico interdomínio, controle de bloqueio autônomo e a interconexão de redes na camada física e na camada de dados.

5.1.3 Aplicações

A rede CA*net 4 é a rede de "*next generation*" de pesquisa e de educação criada pelo governo do Canadá.(Figura 5.4) [7]. Seu objetivo final é prover aos usuários finais a habilidade, dentro do conceito de rede P2P, aprovisionar, gerenciar e controlar o roteamento e o estabelecimento de conexões com garantidas de banda em toda a rede. Esta rede irá operar sem a necessidade dos usuários finais sinalizarem ou pedirem a qualquer servidor central existente na rede autorização para estas atividades. Os usuários finais da rede CA*net 4 são instituições de pesquisa ou universidades que precisem de grandes capacidades fim-a-fim dedicadas para aplicações que necessitem de muita largura de banda, tais como, Física de Alta Energia, Bio-Informática entre outros. Os pontos de presença nos quais existem os GigaPOPs da rede CA*net 3, que estão atualmente em operação na rede acadêmica e de pesquisa baseada na tecnologia IP, serão conectados como um usuário final na rede CA*net 4.

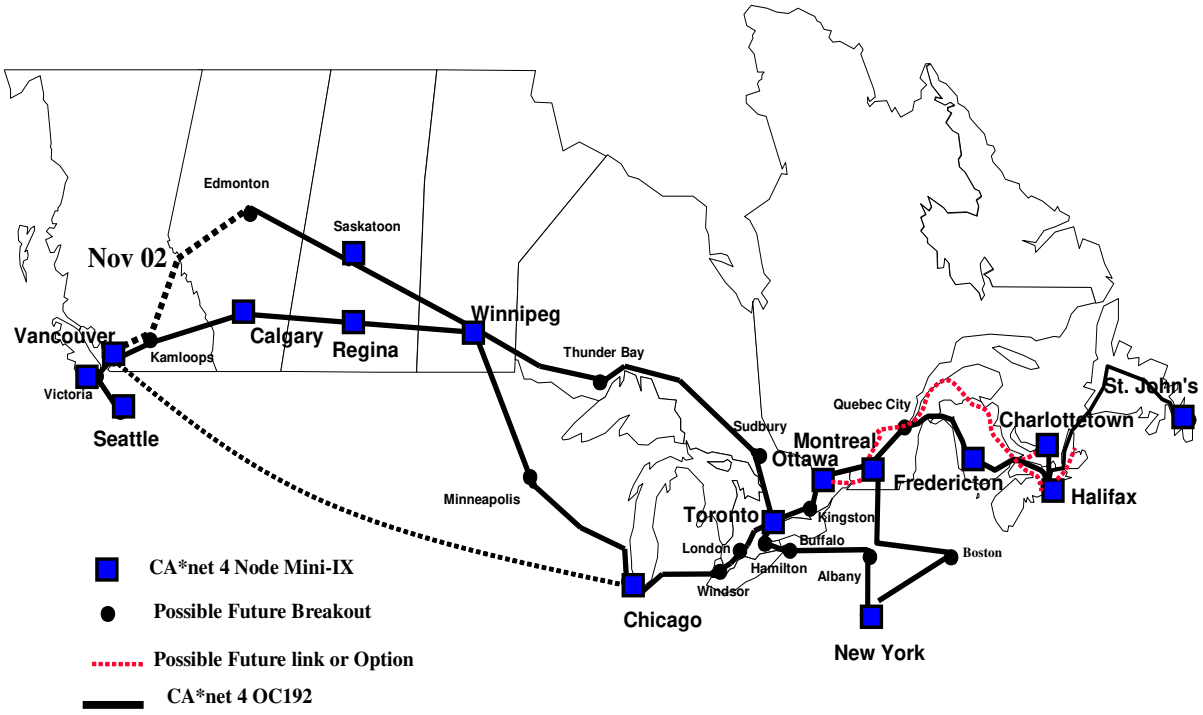


Figura 5.4. Topologia proposta para a rede CA*net 4 [36]

Uma das tecnologias-chave que será entregue no CA*net 4 é o que chamamos de "Míni" Pontos de Troca da Internet (Mini-IXs). Os Míni-IXs são conectados "em cruz" e podem fazer a ponte entre outros usuários. Ao invés de existir uma organização central que gerencie estes Mini-IXs, seu gerenciamento é dividido entre os diferentes usuários. Cada conexão em cruz pode ser administrada independentemente e individualmente. Concatenações e trocas de tráfego entre Mini-IXs independentes devem ser feitos através de acordos bi-lateral entre os "gerentes" de cada conexão individual.

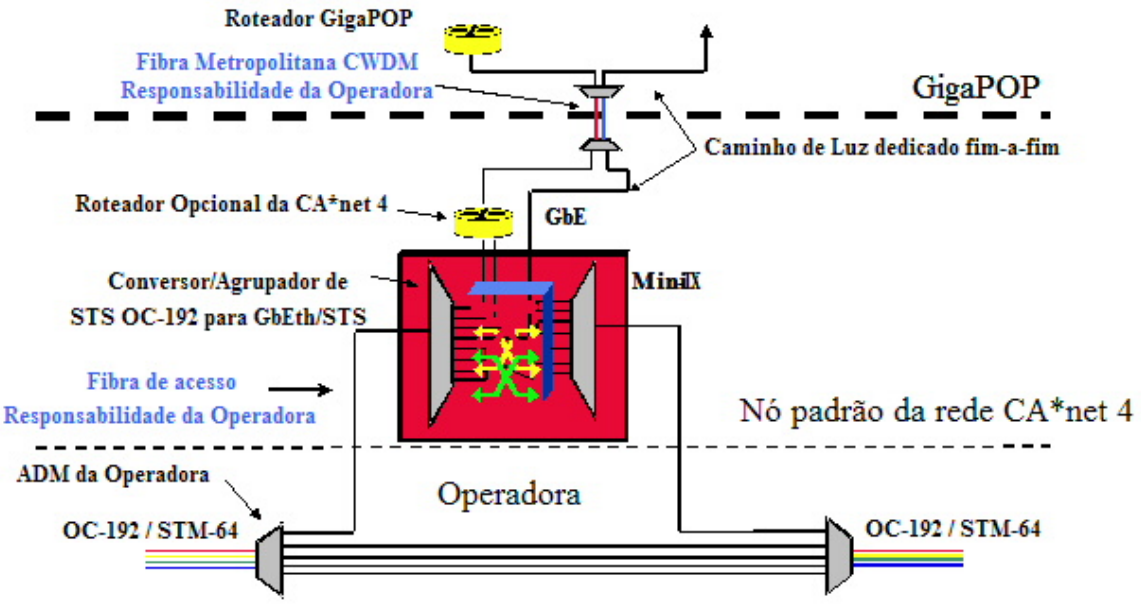
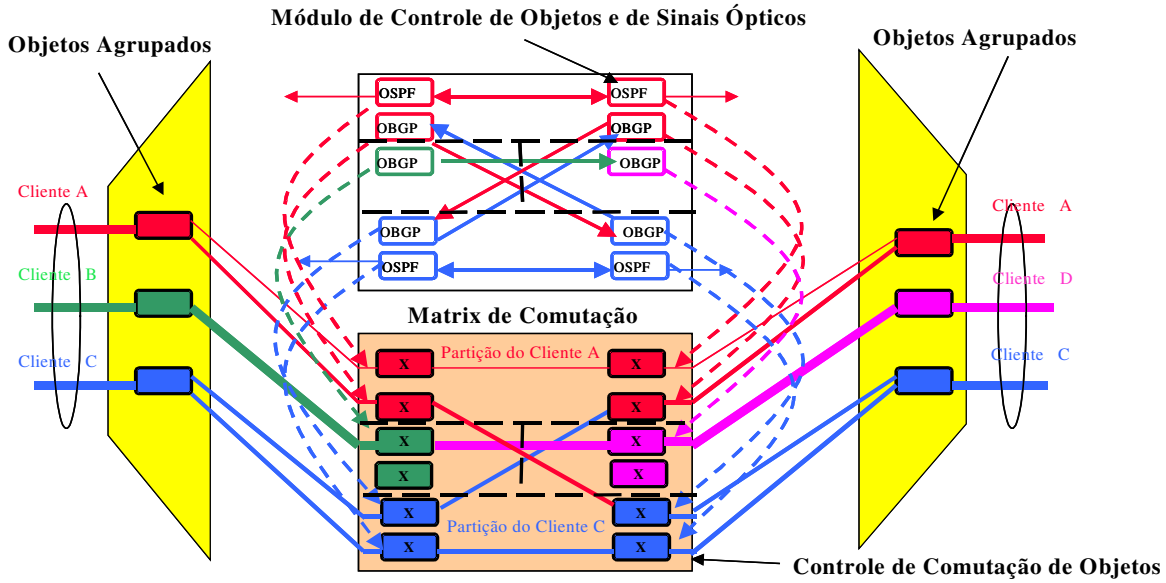


Figura 5.5 Os princípios dos Míni-IXs [36].

Cada cliente possui um canal conectado ao Mini-IX. O canal pode ser um comprimento de onda dedicado como no exemplo acima ou canais SONET/SDH separados. O Míni-IX é dividido em quatro domínios distintos que representam quatro clientes. Cada um dos domínios possui as seguintes funções associadas: grooming, comutação e controle de serviços. Ao invés de se ter uma interface única de administração para todas estas funções como em redes ópticas tradicionais, agentes ou objetos são associados com cada função respectiva para todo cliente que está no Míni-IX.



Obs.: Os Módulos de Sinal e de Comutação são divididos entre os participantes.
Todos os objetos são controlados remoteamente via SOAP.

Figura 5.6. Míni-IX com o cliente controlando as portas e a conexão em cruz [36]

Agentes ou objetos são módulos de software independentes que administram as funções associadas, mas se comunicam através de um protocolo padrão com o respectivo gerente do domínio. Agentes ou objetos tipo CORBA, SOAP ou qualquer outro agente ou objeto baseado em um protocolo de comunicação pode ser usado para a comunicação entre um gerente de domínio (ou de uma partição) e seus agentes ao Míni-IX.

Por exemplo, o cliente A pode se comunicar diretamente com seu agente agrupador ("grooming") e particionar seu comprimento de onda em qualquer direção que desejar. Da mesma forma ele pode se comunicar diretamente com seus agentes de comutação e de conexão em cruz dos circuitos agrupados. Os circuitos agrupados podem ser conectados em cruz internamente para criar a topologia de rede que o usuário desejar e externamente pode se conectado em cruz em outra partição dentro do comutador.
No caso posterior, a aprovação bi-lateral é requerida pelo dono da conexão em cruz destino. OBGP (BGP com extensões ópticas) foi projetado para este propósito [37].

5.2 Arquitetura WebServices

Entende-se por WebService, uma arquitetura de computadores distribuída formada por vários computadores diferentes tentando se comunicar através da rede para formar um sistema único. Eles se baseiam em um conjunto padrão para a criar aplicações que usam uma combinação de módulos de software que são chamados de sistemas distintos e com gerenciamento separado de domínios. Em geral, os WebServices utilizam-se de XML para a descrição, o chamamento e a descoberta de web services. Uma outra característica importante é que o WebService não envolve o uso de programas de aplicações persistentes sendo executados em uma única máquina. Ao invés disso, a máquina só executa um programa, quando ela recebe uma requisição feita através do WebService disponível na máquina.

A tecnologia WebServices tem se tornado uma importante área de pesquisa e desenvolvimento, especialmente para o comércio eletrônico e aplicações para as ciências. Isto é a junção de vários modelos tecnologicamente distintos dentro de uma construção unificada. Cada um destes modelos vem sendo desenvolvido separadamente através dos últimos anos como as linguagens (JAVA, C#, PYTHON), recursos "middleware" e tecnologias de busca/descoberta (CORBA, DCOM, .NET, JINI, JXTA, UDDI), e esquemas de conectividade (cliente-servidor, P2P).

É muito pouco provável que uma única tecnologia de serviço seja adotada e deva continuar existindo várias versões de arquitetura de WebServices. Entretanto, existe uma congruência em torno dos seguintes modelos: WSDL ("WebService Description Language") para descrição dos serviços, UDDI ("Universal Description, Discovery and Integration of Web Services") para descoberta de novos recursos e SOAP ("Simple Object Access Protocol") para transferência de mensagens. Um outro modelo muito promissor é a tecnologia JXTA ("Juxtapose") para redes P2P, formando grupos para compartilhar recursos e serviços baseados em interesses comuns. O JXTA também suporta o conceito de "Tubos P2P" que podem facilmente ser mapeamos dentro do mundo óptico nos "caminhos de luz". Com o uso de JXTA, a comunidade de pesquisas em redes, pode por exemplo, compartilhar a aprovisionar "caminhos de luz" entre eles, dentro de uma arquitetura P2P.

Os desafios de gerenciamento inter-domínio ou no gerenciamento de redes ópticas proprietárias é idêntico aos desafios encontrados na computação distribuída entre sistemas de computadores distintos operando em domínios de gerenciamento diferentes. Alguns problemas a serem resolvidos são comuns tanto para as aplicações da computação distribuída, quanto para as redes ópticas; por exemplo, o gerenciamento de autenticação, de políticas de uso e da descoberta de recursos.

O grande atrativo em se utilizar a tecnologia WebServices em redes ópticas é a possibilidade de se eliminar tarefas longas e tediosas no desenvolvimento de novos padrões para o gerenciamento de redes ópticas. Os padrões atuais, desenvolvidos pelo ITU-T, IETF e/ou OIF, envolvem projetos muito detalhado, consensual e comprometido entre as entidades participantes. As ferramentas utilizadas no WebServices como WSDL, por outro lado, podem eliminar as necessidades de processos tediosos para prover uma linguagem unificada para definição de protocolos de mensagens, descrição dos serviços e tratamentos de exceção. Muito mais importante e interessante é que a descrição de serviços e de processos podem ser colocadas juntas, e tratadas como objetos, de tal maneira que o serviço não tenha que estar ligado com a última versão de um padrão, mas possa ser chamado como um objeto ou um agente por uma outra pilha de serviço ou uma API (“Application Program Interface”)

Os serviços da tecnologia WebServices podem ser facilmente integrados aos API de gerenciamento de redes existentes ou com as interfaces de gerenciamento tais como CLI, TL1 ou SNMP ou ainda com chamadas de serviços existentes como O-UNI. Atualmente muitas empresas de redes fabricantes de equipamentos de comutação e de roteamento estão desenvolvendo descrições gráficas de processos (esquema) em XML e serviços WSDL para interagir com interfaces CLI e TL1. Através destas ferramentas é possível se utilizar ambientes que incluem ricas interfaces gráficas para fazer gerência de rotas, investigações de roteamentos e anomalias ocorridas nos comutadores, editar as listas de acesso e os mapas de roteamento.Uma descrição gráfica de processos feita em XML para interoperar com o CLI permitirá que novos serviços sejam facilmente desenvolvidos e entregues. Como exemplo, cita-se o OBGP.

A Figura 5.7 é um exemplo de como a tecnologia WS pode ser usada para anunciar, descobrir e fazer o estabelecimento de um objeto denominado caminho de luz (“lightpath”).

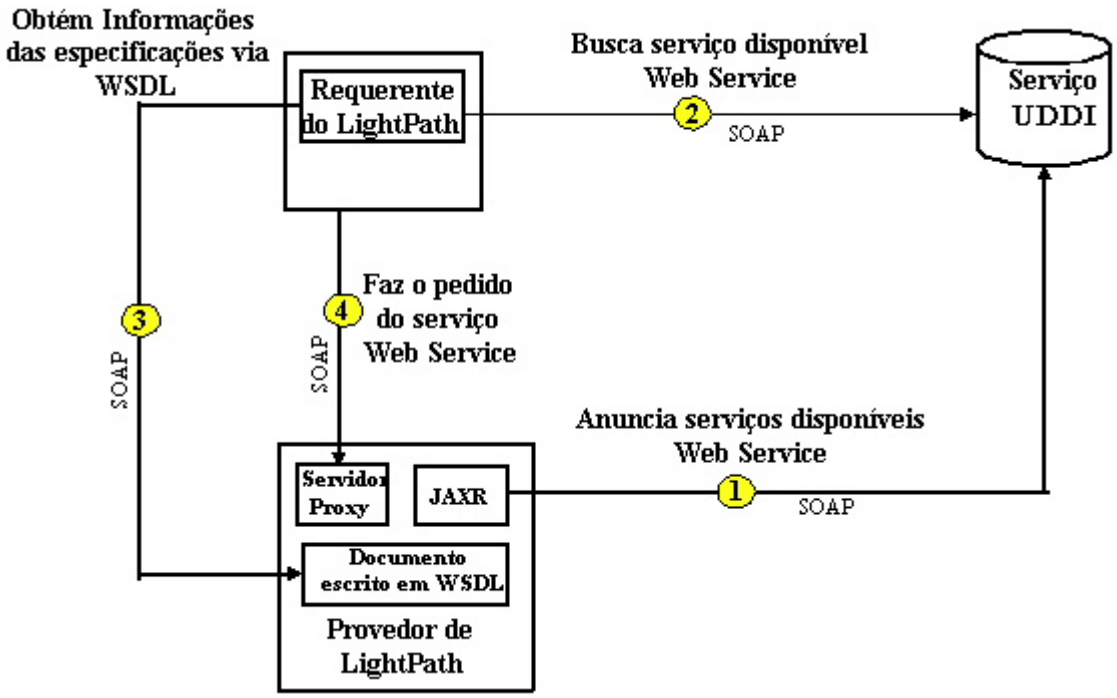


Figura 5.7 Arquitetura WebService para o estabelecimento de um LightPath

Neste exemplo, o provedor de caminhos de luz, que pode ser uma rede de pesquisa com alguns caminhos de luz não utilizados, gostaria de anunciar a disponibilidade de alguns caminhos de luz para outras redes de pesquisas através do servidor de base de dados UDDI ("Universal Description Discovery Integration"). O servidor de UDDI é o local onde as políticas de uso do WebService são anunciadas tais como restrições, custos etc. No exemplo acima, um applet Java (o JAXR) se comunica com a base de dados do servidor UDDI através do protocolo SOAP e uma parte de dados escritos em XML para descrever o serviço do caminho de luz. Alguém que necessita de um caminho de luz pode consultar a base de dados do UDDI para verificar a disponibilidade e a localização deste serviço de caminho de luz.

Feito isso, o requerente pode pedir ao provedor mais detalhes sobre o serviço oferecido através de um descritivo gráfico escrito em XML. Depois, o requerente pode então através de API's que sinalizam diretamente para o O-UNI ou através do uso do protocolo SOAP, com informações escritas em XML, requisitar o serviço deste caminho de luz para a porta do servidor específica.

A Figura 5.8 mostra como este conceito pode ser extrapolado para grandes redes, nas quais existem vários elementos de rede e serviços gerenciados de forma independente.

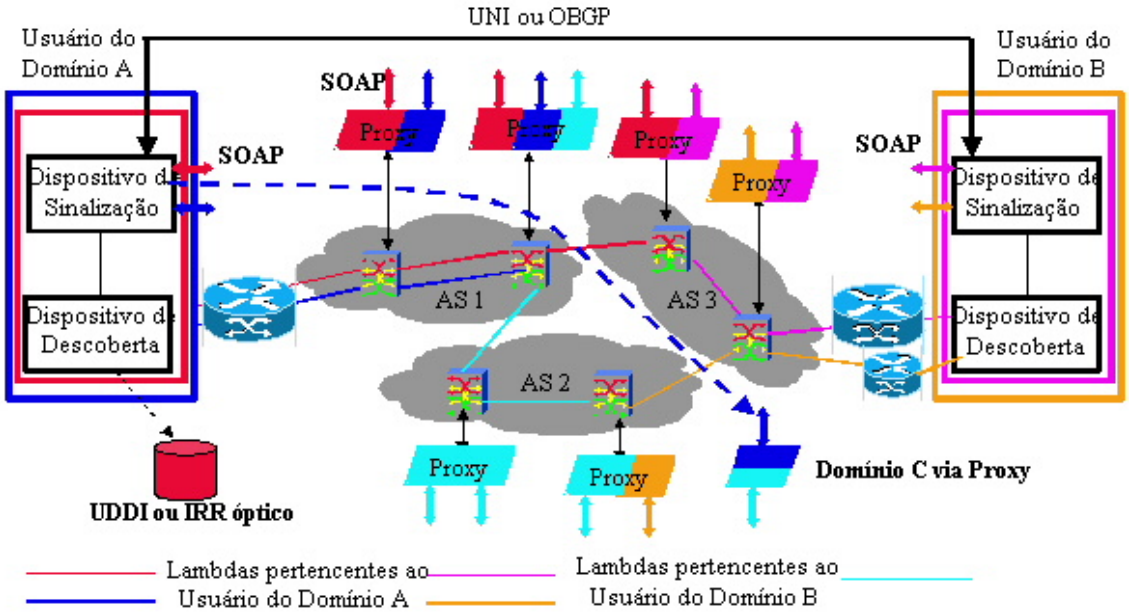


Figura 5.8 Exemplo de um estabelecimento de WebServices em larga escala

Neste exemplo, têm-se dois domínios independentes que desejam estabelecer conexões fim-a-fim de caminho de luz, através de várias nuvens de redes ópticas gerenciadas de forma independentes. Dois cenários são ilustrados neste exemplo: o primeiro deles é o pedido de um Usuário do Domínio A através de um caminho de luz acessar um Usuário do Domínio B e um Usuário do Domínio A pedindo um caminho de luz para um usuário que está depois do Domínio B e que está representado por um router.

As nuvens de redes AS1, AS2 e AS3 aprovisionam, inicialmente, caminhos de luz P2P para três usuários de domínios – A, B e C. É importante notar que embora apenas 6 comutadores são mostrados, na realidade eles representam uma abstração de grandes comutadores. Esta agregação é transparente para o usuário.

Associado com cada um destes comutadores está disponível um servidor proxy de um WebService. Os serviços iniciais necessários para suportar o aprovisionamento de caminhos de luz estão disponibilizados no servidor proxy. Os usuários de domínios se comunicam com seus respectivos WS através do protocolo SOAP juntamente com um descritivo gráfico escrito em XML.

No primeiro cenário no qual um usuário do domínio A quer ter um acesso direto com um domínio B, o usuário A sinaliza o estabelecimento de um caminho de luz com o domínio B através do OBGP ou uma variação P2P do O-UNI. O anúncio e descoberta do Domínio B como um potencial ponto de conexão pode ser determinado através de um Internet Route Registry com extensões criadas para caminhos de luz ou através de um servidor UDDI. O usuário do Domínio B gostaria de ser anunciado ao IRR ou ao servidor de UDDI que ele está interessado em fazer conexões com quem, onde e sob quais condições. O usuário do Domínio A descobriria que um serviço de troca está sendo oferecido pelo usuário B e então ele faria uma pesquisa em base de dados da sua topologia para verificar se existe um caminho através de terceiros para fazer este ponto de troca.

No segundo cenário, assume-se que não existe nenhum ponto de conexão de caminhos de luz diretamente entre o usuário do Domíno A e o usuários do Domínio B. Neste exemplo, um usuário do Domínio C anunciou um conjunto de objetos caminhos de luz alternativos que permitiria que o usuário do Domínio A se conectasse com o usuário do Domíno B. Uma vez que o usuário do Domínio A tenha adquirido estes caminhos do usuário C, ele pode invocar os seus próprios WebServices para fazer a conexão com o usuário do Domínio B. Isto é feito quando o usuário do Domínio C cria uma cópia do WebService que o usuário A está usando. Feito isso, o usuário A pode então se comunicar diretamente com o seu servidor de WS para finalizar a interconexão (linha azul pontilhada no desenho).

5.3 OBGP (Optical Border Gateway Protocol)

O modelo clássico utilizado na Internet de hoje é aquele no qual grandes empresas e ISPs médios têm apenas um caminho default para ISP maiores ou grandes operadoras que agregam o tráfego e fazem a gerência para outras redes. Com o aumento significativo de investimentos ocorrido até início de 2001, houve um grande aumento na quantidade de fibras ópticas instaladas no mundo, bem como um aumento considerável na utilização destas fibras através da tecnologia WDM e suas variantes.

O protocolo de gerenciamento de tráfego entre Sistemas Autônomos, de facto, é o BGP ("*Border Gateway Protocol*"). O gerenciamento do BGP pode se tornar muito complexo, quando existe a necessidade de se gerenciar vários links entre AS ("*Autonomous Systems*") independentes. A essência é do OBGP é propor um mecanismo que, de forma, automática faça a configuração do BGP para suportar diferentes links ópticos entre AS distintos.

Existem algumas possíveis soluções para isso. Uma delas é tratar cada conexão óptica como se fosse um caminho direto entre duas entidades trocando informações de BPG entre si. Este modelo, no entanto, aumenta significantemente a complexidade de qualquer sessão de BGP, por mais simples que seja. O ideal, até o momento, é tratar cada conexão óptica cruzada como sendo um único roteador virtual utilizando o BGP com apenas uma porta de entrada e uma porta de saída. Com isso, cada roteador virtual utilizando BGP pode ser configurado para cada uma das conexões ópticas cruzadas iniciando as sessões de BGP separadamente com seus pares. Desta maneira, cada roteador virtual utilizando o BGP pode ser facilmente replicado de e para outro roteador virtual.

Pode-se dizer que pouco se tem dedicado ao desenvolvimento da automatização da gerência, configuração e estabelecimento de conexão de comprimentos de onda entre domínios ou mesmo de ferramentas que permitam que empresas que estejam na borda da nuvem possam ter gerência sobre seus comprimentos de onda dentro em uma nuvem óptica. A solução convencional utilizada ainda hoje pela maioria das operadoras é que a própria operadora ofereça um caminho de luz gerenciado aos clientes que estão na borda da rede (Figura 5.9). As empresas que estão conectadas nas bordas destas nuvens ópticas, geralmente têm pouca visão da rede e de uma forma virtual têm pouco ou nenhum controle na maneira pela qual os caminhos de luz são roteados.

Internet Óptica Atual – Visão Geral

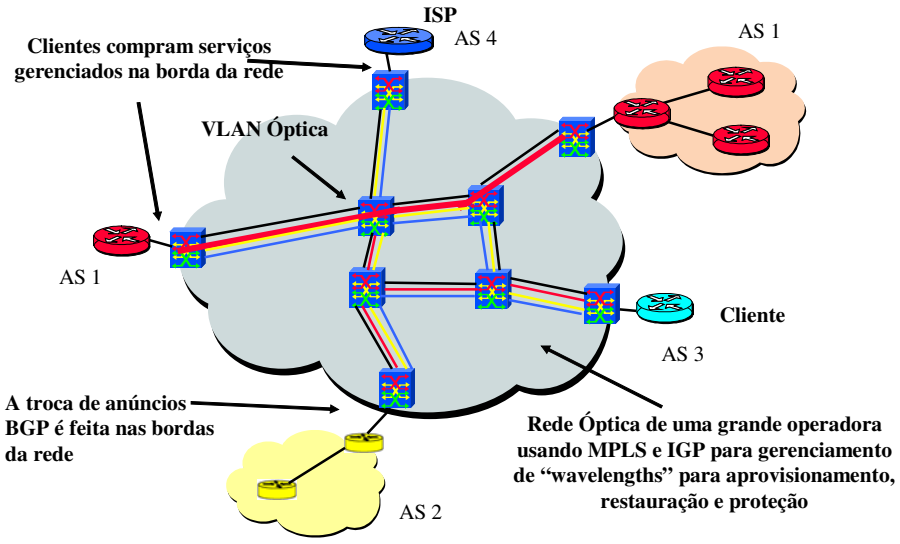


Figura 5.9 Visão Geral da Internet Óptica Atual

Vários mecanismos têm sido propostos para a gerência e o controle de sistemas de nuvens ópticas [38]. A maioria destes sistemas tem sido desenvolvida como variações dos protocolos de roteamento "*link state*" tais como o OSPF, IS-IS e o PNNI ou mesmo extensões complementares do MPLS, tais como o MPLmS [39].

A proposta e o uso do OBGP, como uma extensão óptica ao BPG padrão, para gerência e controle de comprimentos de ondas entre ASs distintos dentro de uma nuvem óptica é o grande objetivo do modelo utilizado no projeto CA*net4. Com a utilização do OBGP, as empresas clientes que se situam na borda da rede, serão capazes de gerenciar e controlar suas conexões e acordos de troca de tráfego e de roteamento com outras redes dentro de uma rede de comprimentos de onda (Figura 5.10).

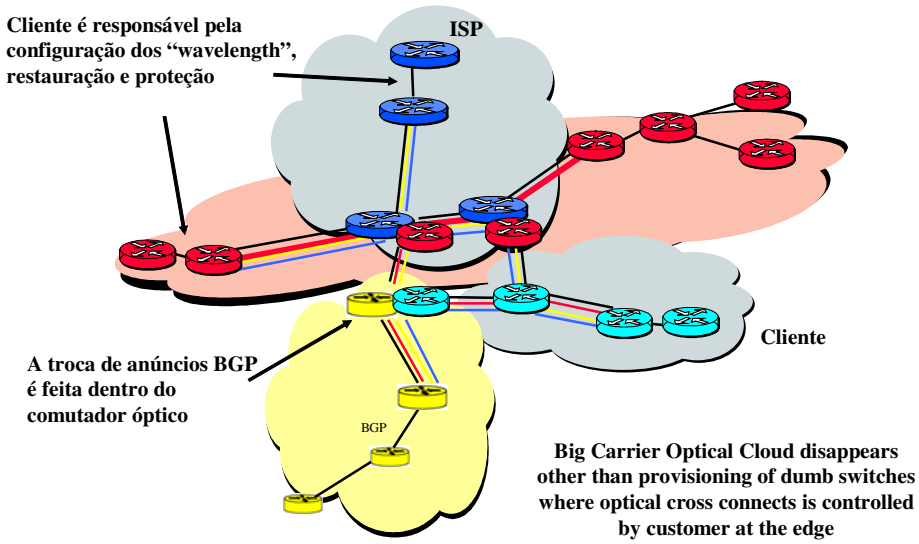


Figura 5.10 Internet Óptica usando o OBGP

5.3.1 Diferenças entre o OBGP e outros protocolos de gerência ópticos

É importante neste momento especificarmos que o propósito do OBGP não é o de substituir nenhum mecanismo proposto para gerenciamento de nuvens ópticas. Ele deve ser visto como um melhoramento destes protocolos. O OBGP provê mecanismos que acessam e que usam todos os mecanismos propostos. Os pontos abaixo esclarecem as diferenças:

1. O OBPG não é um protocolo de sinalização para suportar o estabelecimento de conexões fim-a-fim. Ao invés disso, ele é proposto para o estabelecimento de escalas curtas de caminhos de luz para primeiramente otimizar o fluxo de dados no roteador local e, após isso ter sido feito, permitir que pontos de conexões externos possam estabelecer caminhos de luz para otimizar também seu fluxo de tráfego. O resultado final pode ser um caminho de luz fim-a-fim, mas essa coincidência benéfica não é o objetivo.
2. Nas redes de fibras metropolitanas e regionais, onde as organizações participantes, quando houver, são interconectadas pelo responsável pela fibra. São necessários acordos multilaterais de trânsito e de troca de conteúdo. Muitas destas redes estão planejando o uso de sistemas WDW para suportar diretamente ponto de troca entre si e entre organizações de interesses similares. Sem uma conexão direta entre os pontos de troca, há necessidade de uma organização central que gerencie o trêfego agregado e o roteamento nas bordas da rede. O OBGP foi projetado para minimizar a necessidade de um administrador central ter que ficar fazendo isso, permitindo que os participantes façam suas interconexões entre si.

3. O OBGP dará aos usuários um controle razoável do roteamento dos seus caminhos de luz através de outras entidades dentro de uma rede óptica de comprimentos de onda, talvez até mesmo como uma camada superior ao protocolo de gerenciamento interno de comprimento de onda. Por exemplo, uma operadora pode ter uma grande rede óptica de comprimentos de conda, mas ao invés de não explicitar o roteamento dos comprimentos de onda dos clientes, os clientes podem ter uma visão específica da topologia de rede ou um conjunto de possíveis rotas as quais são um subconjunto de todas as rotas possíveis. Além disso, o usuário pode ter rotas ópticas transitando em duas operadoras distintas e pode querer se interconectar através dessas nuvens distintas em algum ponto central. Como conseqüência deste modelo hipotético, a topologia de comprimentos de onda ideal para este usuário pode ser uma variação da topologia ideal otimizada para a rede específica do usuário. O OBGP permite que a topologia do usuário tenha precedência sobre a topologia preferencial da operadora.
4. Grandes domínios de comprimentos de onda dentro de redes ópticas podem se tornar não gerenciáveis. A solução trivial é dividi-los em vários domínios menores os quais podem ser mais bem gerenciados. O OBGP pode ser utilizado entre esses domínios menores para a melhor utilização dos mesmos.

5.3.2 O uso do BGP no roteamento e configuração de redes ópticas

Muitos dos problemas existentes na configuração e roteamento de protocolos externos de gerenciamento foram solucionados com a criação do BGP. Este mesmo modelo de solução pode ser aplicado em redes ópticas de comprimento de ondas. O BGP tem uma arquitetura básica e uma conjunto de ferramentas que tem como definição o fato de que ele será usado para estabelecer conexões entre domínios distintos, gerenciados de forma autônoma. Embora outros protocolos IGP ("*Interior Gateway Protocol*") pudessem ser utilizados para o gerenciamento de comprimentos de onda, esses mecanismos se aplicam a uma única entidade controladora.

O roteamento BGP apenas transporta informações de alcance. Ele não transporta nenhuma informação sobre a melhor topologia de rede, Qualidade de Serviço ou capacidade de uma rota específica. Os caminhos de luz têm, normalmente, uma capacidade de transmissão fixa , tipicamente de 1 a 10 Gbps para sistemas CWDW ou 2,5 e 10 Gbps para sistemas DWDM. As características físicas de um caminho de luz dá a ele a capacidade intrínseca de ser o melhor caminho com uma Qualidade de Serviço pré-definida.

Parâmetros intrínsecos sofisticados como Qualidade de Serviço, Restauração e Proteção, os quais estão comumente disponíveis em outras tecnologias baseadas em circuitos, tais como MPLS, Frame Relay, SONET/SDH ou ATM, podem não ser necessários em redes utilizando o BGP óptico com comprimentos de ondas DWDM entre todos os nós BGP. Uma rede óptica DWDM poder ter desde o mais simples até o mais complexo ambiente de BGP interdomínios, onde existem múltiplos caminhos de capacidade de transmissão fixa conhecidos entre as redes vizinhas.

Pelo fato dos protocolos "vetor distâncias" ("*distance vector*") listarem os domínios ou os AS nos quais o pacote deve trafegar em uma rota anunciada, a informação deste caminho torna possível para o roteador do cliente executar um tipo de

engenharia de tráfego rudimentar para interdomínios. Este tipo de engenharia de tráfego não é tão rigorosa ou completa quanto a engenharia de tráfego do MPLS. O problema é que o MPLS-TE, até o momento, trabalha apenas dentro de um único domínio. Inter-domínios MPLS, provavelmente, irão permitir enlaces de engenharia de tráfego através de domínios, mas a negociação e transferência de mensagens RSVP e LDP requerida entre os domínios será complexa e sujeita a discussões de trânsito de negócios ("*business*") e não técnicas.

Uma das características úteis do BGP é que ele usa a porta 179 do TCP para todas as comunicações feitas entre pares de roteadores utilizando o BGP. Isto significa que qualquer tipo de canal de comunicação pode ser estabelecido entre estes roteadores. Os roteadores ou comutadores falando BGP não precisam de nenhum caminho de luz específico para enviar suas informações; ele pode utilizar qualquer canal de comunicação, incluindo a Internet pública para enviar suas informações de roteamento entre os pares.
Pelo fato do BGP utilizar o TCP para todas as comunicações, isso faz com que os roteadores e os comutadores que utilizam o BGP não precisam se comunicar diretamente entre si. Em uma topologia complexa BGP toda a informação de atualização de rotas pode ser enviada através de um roteador central, normalmente chamado de roteador árbitro ou roteador refletor. O roteador árbitro elimina a necessidade de estabelecer uma rede de malha (*"mesh")* de conexões BGP. Todas as atualizações e trocas de roteamento BGP podem ser anunciadas de e para o roteador árbitro. Esses roteadores árbitros são comumente utilizados nos grandes pontos de troca de tráfego na Internet.

A outra grande vantagem do BGP é que ele foi projetado para suportar rotas unicast. A Internet é fundamentalmente uma rede unidirecional. As redes tradicionais de telecomunicações, por outro lado, assumem que todos os links são bi-direcionais e a grande maioria dos protocolos de roteamento e configuração nestas redes, não fazem distinção entre caminhos de envio e de retorno [40].

O BGP é normalmente configurado para trabalhar em conexões P2P ou em segmento de redes tipo LAN. Tanto os comprimentos de onda ópticos, quanto os caminhos de luz são essencialmente conexões P2P. Os caminhos de luz ópticos, devido às limitadas propriedades físicas da luz, não têm a complexidade e a flexibilidade existentes em circuitos ATM e SONET/SDH. Por conta disso, os caminhos de luz ópticos podem ser facilmente gerenciados por um protocolo P2P como o BGP.

É importante notar que, normalmente, a simplicidade e o baixo custo sempre vencem a complexidade e o custo de soluções sofisticadas, como tem nos demonstrado o protocolo Ethernet e suas variantes.

5.3.3 Tornando conectores ópticos em cruz em dispositivos BGP

Como o tráfego em um roteador é sempre distribuído entre portas de entrada e portas de saída, há necessidade de um dispositivo de entrega e seleção que possa direcionar os pacotes egressos das portas de entrada para as respectivas portas de saída. A capacidade computacional necessária para entrega e seleção destes pacotes nesta configuração é uma função descrita pela raiz quadrada do número de pontas multiplicada pela velocidade dos dados.

Todavia, em várias configurações de roteadores, é comum que as portas de entrada sejam direcionadas apenas a uma ou duas portas de saída. Raramente existe uma

distribuição par-a-par entre as portas de entrada e as portas de saída. Por exemplo, um provedor de acessos à Internet (ISP) de tamanho médio, gostaria de entregar seu tráfego para as portas que estão conectadas à saída do seu maior provedor. Uma porção pequena seria entregue às redes que estão no mesmo nível hierárquico deste roteador.

Uma rede regional ou metropolitana que tem conexões com várias entidades, quer sejam acadêmicas ou comerciais, e que se conecta a um ISP é um exemplo para este tipo de configuração. Os roteadores do núcleo ("*core*") da rede regional gostariam de ver a grande maioria dos pacotes entregues na porta de saída que se conecta diretamente com o ISP e significantemente menos tráfego de dados entregues entre as universidades e instituições.

Nesta configuração seria muito mais vantajoso o uso de conexão em cruz óptica para suportar o grande fluxo de dados que são direcionados de uma única porta de entrada para uma única porta de saída. As empresas que comercializam roteadores já entregam esta característica em seus roteadores através do mecanismo chamado de "*Fast Forwarding*", direcionando os pacotes para uma porta comum de saída, dentro do domínio eletrônico. Uma outra técnica utilizada é o uso de circuitos virtuais ATM combinado com roteamento virtual para suportar grandes fluxos de dados. Uma das técnicas propostas para a Entrega Rápida ("*Fast Forwading")* entre vários roteadores é o NHRP ("*Next Hop Resolution Protocol"*), entretanto atualmente vem sendo substituído pelas soluções providas pelo MPLS. A maioria do controle de fluxo de dados basedo na técnica "*cut thru"* tem sido abandonada em favor à engenharia de tráfego para gerenciar este tipo de trêfego. O OBGP é uma técnica alternativa à engenharia de tráfego para aplicações interdomínios.

Como um primeiro passo, foi proposto que comutadores de conexão em cruz ópticos fossem integrados dentro de roteadores BGP. O roteamento eletrônico pode ser feito por equipamentos eletrônicos simples que têm o mínimo de capacidade de entrega no domínio elétrico.

Os conectores "em cruz" ópticos incluem multiplexadores e demutiplexadores ópticos e os filtros ópticos podem ser ativos ou passivos. Como conseqüência disso, o detentor do roteador óptico sabe com antecedência qual o comprimento de onda que pode ser conectado "em cruz". Os pares externos não necessitam saber de antemão a identificação física das portas (diferentemente do ATM e do SONET) dos nós ou mesmo se ele pode suportar um certo conjunto de comprimentos de onda.

Embora conectores em cruz ópticos sejam dispositivos muito simples e possam ser gerenciados através de uma simples interface serial, seria interessante que cada interface óptica pudesse ser gerenciada por um dispositivo IP independente. É muito vantajoso ter um roteador externo que faça a gerência de conectores ópticos, pois com isso, um usuário externo pode gerenciar sua própria conexão "em cruz" óptica e direcionar os comprimentos de onda para o parceiro de sua escolha.

5.3.4 Mapeando comprimentos de onda em endereços IP

Caso os comprimentos de onda em um roteador falando OBGP fossem enquadrados dentro das definições do ITU, poder-se-ia fazer um mapeamento entre comprimentos de onda e endereços IP. Tendo em vista que os lasers ajustáveis e os filtros

têm um alcance limitado, sufixos diferentes poderiam ser usados para indicar de forma apropriada o alcance de comprimentos de onda.

Por exemplo, um endereço na banda óptica "C" poderia ser mapeado com o sufixo "x.x.1.x/24"; isso nos permitiria o mapeamento de quase 250 comprimentos de onda. Um outro exemplo poderia ser o dos endereços na banda "L" que poderia ser definido pelo prefixo "x.x.10.x/20" o qual permitiria o mapeamento de uns 16.000 comprimentos de onda e assim por diante.

Não é intuito deste trabalho fazer sugestões para este tipo de mapeamento no momento, apenas indicar que seria um mecanismo muito útil entre os roteadores OBPG.

5.3.5 Configuração de roteadores OBGP

Os roteadores rodando OBPG terão vários caminhos entre si e qualquer caminho óptico terá preferência sobre qualquer outro caminho que venha através de um equipamento elétrico usando técnicas do BGP normal selecionando, por exemplo, caminhos de AS, MEDs e preferências locais.

A Figura 5.10 ilustra um modelo da configuração do BGP em uma rede. Cada roteador está conectado a um par de comprimentos de onda. Os roteadores A,B e C estão interconectados como vizinhos em uma configuração padrão de BGP.

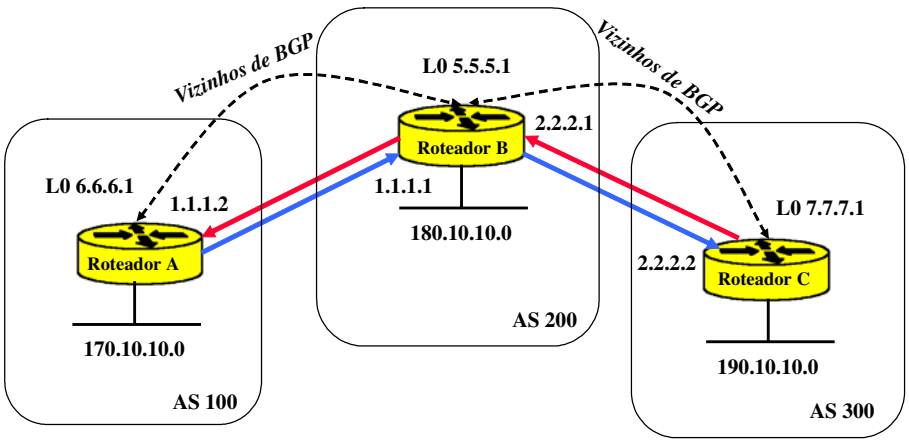


Figura 5.10 Modelo de configuração de BGP em uma rede

Suponha que o roteador B seja uma combinação de um comutador óptico e um roteador tradicional, como na Figura 5.11. Os componentes do comutador óptico são mostrados em detalhes na Figura 5.12.

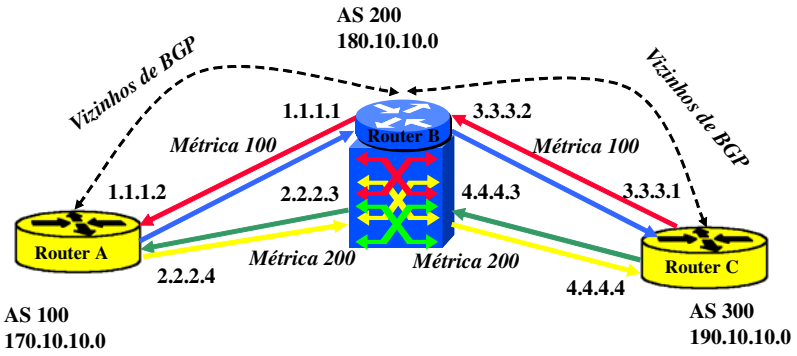


Figura 5.11 Roteador e o Comutador Óptico são uma única entidade

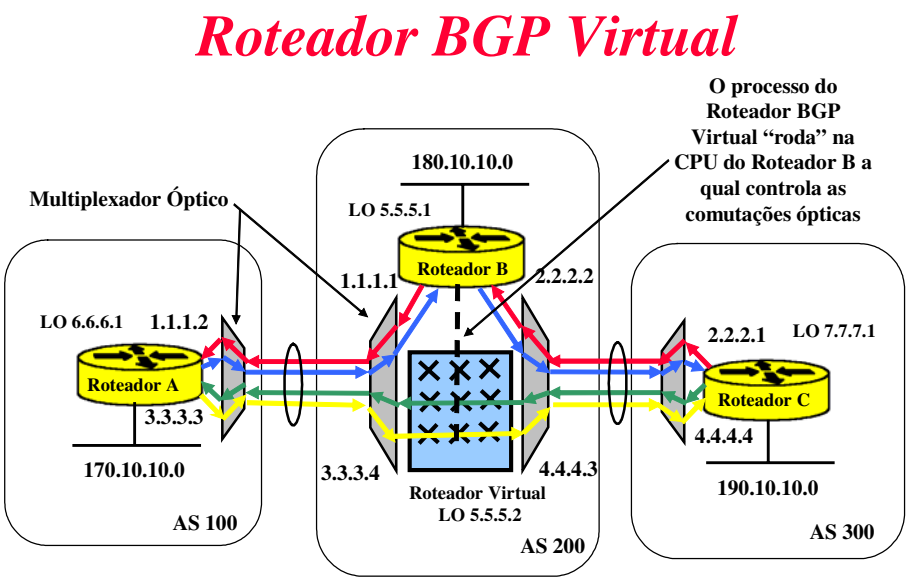


Figura 5.12 Componentes de um comutador óptico

Para multiplexar e demultiplexar os comprimentos de onda o roteador B deve usar filtros ópticos que possam separá-los individualmente. Estes filtros podem ser ajustados ou fixos. Com cada uma das cores da luz um endereço default de porta pode ser

assinalado ou diretamente mapeado entre a pilha do ITU e um endereço IP, como descrito anteriormente.

Através do uso de uma comutação óptica simples, cada uma das portas dos caminhos de luz podem ser tratadas, de fato, como um caminho alternativo para o roteador B. Neste exemplo, um outro conjunto de comprimentos de onda é estabelecido entre os roteadores A e C para o roteador B. Existem dois (2) roteadores entre o roteador A e o roteador B. O roteador B deve agora avisar ao roteador C a melhor rota para o roteador A. Em uma situação normal de roteamento apenas um dos roteadores entre A e B seria avisado; mas devido ao fato de um dos caminhos ser através do comutador óptico sem nenhuma multiplexação, isto poderia de uma forma inadvertida deixar o roteador A sem ser visto ou conhecido pelo roteador C.

Existem duas maneiras de se configurar o roteador B. A primeira e mais complexa é usar a conexão em cruz para estabelecer um caminho direto entre os dois roteadores. Existem várias maneiras conhecidas para se estabelecer caminhos paralelos na configuração do BGP entre dois roteadores e a maioria deles pode suportar até seis (6) caminhos paralelos. Entretanto existem algumas deficiências neste modelo que podem não ser escaláveis e que discutiremos um pouco mais adiante.

A segunda e talvez a maneira mais simples é tratar cada uma das conexões em cruz como se fossem um roteador virtual BGP independente.

Conexões paralelas de BGP

Uma técnica vantajosa sobre as conexões ópticas em cruz é a configuração "multi-homing" do roteador B para o roteador A. Existem várias técnicas para o estabelecimento de caminhos paralelos de BGP entre roteadores para balanceamento de rotas e definições de enlaces de backup [41]. O BGP suporta até seis (6) enlaces paralelos.

Em uma configuração padrão de BGP o roteador B anuncia uma única melhor rota para o roteador A vinda do roteador C. Por outro lado,um roteador utilizando OBGP tem que estar ciente e informar todas as rotas existentes dentro de um caminho de luz definido entre os roteadores.

O roteador C conheceria as rotas do roteador A através do caminho de luz existente entre os dois e também através do roteamento normal via o roteador B. Caso a caminho de luz diretamente ligado entre o roteador B e o roteador A falhe, é condição "sine qua non" que o roteador A sinaliza a falha para a sessão de BGP aberta. O roteador A envia uma mensagem padrão de "update" para o roteador B notificando-o da perda do enlace. Em uma sessão normal de BGP, o roteador B ainda assim assumiria que ambas as conexões ópticas entre os roteadores B e C estariam ainda operacionais.

Tendo em vista isso, alguns campos tiveram que ser adicionados ao banco de dados de roteamento ("*Routing Information Database*") estabelecendo conexões entre os vários caminhos existentes. É importante notar que não existe um protocolo de sinalização para o re-estabelecimento de uma conexão óptica "em cruz", caso a mesma falhe.

A outra desvantagem deste método é que os roteadores que querem ser conectados em cruz devem se conhecer previamente. Este não é um grande problema se a rede for pequena, entretanto vai aumento a sua complexidade em uma rede maior. Para auxiliar na resolução deste problema, novos protocolos de descoberta de topologia

deveriam ser desenvolvidos para poder obter da rede informações da topologia e da configuração da mesma.

Roteadores BGP virtuais

O conceito básico em roteadores BGP virtuais é tratar cada conexão em cruz óptica como sendo um roteador BGP. O roteador virtual teria apenas uma porta de entrada e uma porta de saída. Ele iria também, no exemplo citado, anunciar a si mesmo para o roteador B com o seu endereço de loopback e o conjunto de endereços IP das suas interfaces. Diferentemente de uma configuração padrão de BGP, o roteador BGP virtual não estabelece nenhuma conectividade IBGP ("*Interior Border Gateway Protocol*") mesmo pertencendo ao AS do roteador B. Ele se comporta como um roteador independente, com suas métricas próprias rotas.

O uso de um roteador virtual para cada conexão óptica em cruz permite o uso do roteamento padrão do BGP com quase nenhuma modificação necessária para suportar caminhos de luz ópticos. Na realidade, o roteador BGP virtual pode estar associado a seu próprio AS privado ou público de tal maneira que a métrica do caminho do AS pode ser utilizada para se fazer Engenharia de Tráfego básica.

Através da instanciação do roteador BGP virtual, o dono do OXC pode primeiramente estabelecer conexão óptica em cruz entre vizinhos para reduzir a carga dos seus equipamentos de entrega puramente elétricos. De tempos em tempos, ele pode reconfigurar o roteador BGP virtual para se interconectar com outros vizinhos caso o padrão de tráfego venha a se alterar. Além disso, o dono do OXC pode estabelecer conexão óptica em cruz entre vizinhos sem envolver os vizinhos no processo de configuração e de decisão.

De maneira muito interessante, o roteador BGP virtual pode ser facilmente realocado dentro de outros roteadores de outros AS. Como exemplo, suponha um Ponto de Troca de Tráfego Internet óptica (OIXP). Cada uma das conexões ópticas "em cruz", operando como um roteador BGP virtual pode ser associada aos AS's dos ISP participantes ao invés de fazerem parte de uma administração centralizada de AS.

O propósito principal do BGP em conexão óptica em cruz é fazer o anúncio das rotas, executar os filtros e a classificação e prover a viabilidade de se fazer Engenharia de Tráfego de Rede padrão via BGP para os pares. Como em cada roteador BGP virtual existe apenas uma porta de entrada e uma porta de saída, não existe a necessidade de se fazer tabelas de encaminhamento de um OXC.

De uma forma objetiva, através do OBPG o processo de roteamento do BGP é distribuído através dos roteadores de borda da rede e é separado fisicamente do processo de entrega/envio dos dados. O processo de roteamento da rede tem lugar no primeiro ponto de egresso na rede, enquanto a classificação dos pacotes e a entrega é feita no núcleo da rede. É como se os comprimentos de onda "esticassem" as portas de entrada/saída da localização física do roteador até algum ponto de egresso em um comutador óptico.

A vantagem principal desta abordagem é que dentre vários caminhos existentes, o roteador BGP virtual anuncia apenas o melhor caminho para um determinado roteador baseado nas métricas existentes no BGP. Caso ocorra uma falha no caminho, o roteador

BGP virtual recalcula o próximo melhor caminho e anuncia este novo caminho através de uma mensagem de NLRI ("Network Layer Reachability Information") UPDATE.

Configuração estática entre roteadores virtuais pares

As conexões podem ser estabelecidas durante a configuração de rede, como é feito atualmente em ambientes de redes rondando BGP. Os roteadores podem ser configurados entre os seus pares, caminhos de luz etc, com nenhuma alteração no protocolo BGP hoje existente.

O que se faz necessário para criar roteadores BGP virtuais é a modificação de uma parte do código aberto do BGP de tal maneira que ele possa fazer referência a um comutador de conexão em cruz óptico. Apenas um processador é necessário, mas existem vários processos BGP rodando ao mesmo tempo e cada processo representa um roteador BGP virtual para cada conexão "em cruz" óptica.

Para redes pequenas, o uso da configuração estática tradicional é suficiente. Entretanto, de acordo com o crescimento e da complexidade da rede, é mais vantajoso ter algum grau de configuração automática das conexões "em cruz" óptica para os roteadores BGP virtuais.

Configuração dinâmica entre roteadores virtuais pares

Até o momento, todas as configurações BGP são feitas manualmente. Está claro que desta maneira não se pode ter uma solução escalável para centenas ou até milhares de sessões BGP, quer sejam reais ou virtuais. Um dos requisitos primordiais para grandes redes e sessões de BGP é a possibilidade de se estabelecer e configurar de forma dinâmica as sessões BGP entre os pares.

É importante salientar que para o gerenciamento de um domínio único, existem técnicas melhores para serem aplicadas na configuração e no gerenciamento de comprimentos de onda, como por exemplo, o MPLambdaS; entretanto, quando não existir nenhuma entidade central na rede para o gerenciamento da mesma, como por exemplo, o que ocorre em um domínio autônomo dentro de um ponto de troca de tráfego óptico (IX) ou em uma rede de fibra óptica municipal, é importantíssimo ter um outro modelo.

O BPG, da forma como foi projetado, não está preparado para tratar com entidades tipo caminhos de luz. A proposta abaixo foi apresentada como um possível método de como um roteador BGP virtual poderia dinamicamente estabelecer e fazer a configuração destas conexões:

- Inicialmente dois roteadores reais BGP são configurados manualmente, como é feito atualmente;
- Quando a sessão inicial executa o "BGP OPEN" entre dois roteadores configurados manualmente, o campo de dados de informação opcional existente na mensagem OPEN pode ser usado para trocar informações sobre o número de caminhos de luz existentes entre os dois roteadores, os endereços IPs das portas ópticas (ou a freqüência definida pelo ITU para mapear com os endereços IPs), o protocolo de encapsulamento, o destino preferido e outras informações relevantes.

- Uma vez que os roteadores reais tenham determinado um caminho óptico válido entre cada um deles e dois outros roteadores, ele poderia instanciar um roteador BGP virtual. O roteador virtual pode de maneira independente estabelecer sessões de BGP entre ele mesmo e os outros dois roteadores através do envio de mensagens de BGP OPEN. O roteador virtual tem sua própria interface e endereço IP de loopback (LO), de tal maneira que sessões independentes de BGP podem ser estabelecidas através da porta 179;
- Caso uma sessão virtual de BGP não possa ser estabelecida ou haja uma falha de um enlace, a sessão virtual de BGP pode ser deixada em um estado IDLE esperando que uma outra mensagem de OPEN chegue ou que o roteador virtual BGP termine a conexão;
- O mesmo endereço externo de NLRI poderia fazer o anúncio tanto para os roteadores virtuais, quanto para os roteadores reais. Os endereços que são anunciados pelo roteador real não seriam anunciados pelo roteador virtual, porque não haveria nenhuma conexão iBGP entre os roteadores reais e virtuais;
- Os roteadores vizinhos do roteador virtual podem anunciar suas preferências através do roteador virtual ou o roteador virtual pode associar métricas maiores para os seus enlaces ópticos do que os enlaces ópticos usados pelos roteadores reais.

Conforme ilustrado na Figura 5.11 acima (item 5.3.5), assume-se que os roteadores reais A e B estabelecem uma sessão padrão de BGP entre si, através de algum canal TCP acordado, como por exemplo, um canal de sinalização óptico ou um comprimento de onda default.

A seguir ilustra-se uma configuração padrão de uma sessão de BPG entre os roteadores através de um canal de sinalização ou comprimento de onda padrão:

Configuração do roteador A:

```
router bgp 100
   network 170.10.10.0/24

interface Ethernet 0/1   (conexão default para o roteador B)
   ip address 1.1.1.2/30  (por definição o λ vermelho usa o sufixo x.x.x.2)
   neighbor 1.1.1.1  remote-as  200  (o λ azul usa o  sufixo x.x.x.1)

interface Ethernet 0/2  (conexão ao OXC)
   ip address 3.3.3.3/30 (por definição o λ verde usa o sufixo x.x.x.3)
   neighbor 3.3.3.4 remote-as unknown (o λ amarelo usa o sufixo x.x.x.4)
```

Configuração do roteador B:

```
router bgp 200
    network 180.10.10.0/24
```

interface Ethernet 0/1 (conexão default para o roteador A)
ip address 1.1.1.1/30 (por definição o λ azul usa o sufixo x.x.x.1)
neighbor 1.1.1.2 remote-as 100 (o λ vermelho usa o sufixo x.x.x.2)

interface Ethernet 0/2 (conexão default para o roteador C)
ip address 2.2.2.3/30 (por definição o λ vermelho usa o sufixo x.x.x.2)
neighbor 2.2.2.1 remote-as 300 (o λ azul usa o sufixo x.x.x.1)

Configuração inicial do roteador Virtual
(Este roteador é um potencial OXC para os λs verde e amarelos)

router bgp 200
interface loopback0
ip address 5.5.5.2/32

interface oxc 0/1 (λ verde cross connectado)
ip address x.x.x.3/30 (por definição o λ verde usa o sufixo x.x.x.3)
neighbor x.x.x.4 remote-as unknown

interface oxc 0/2
ip address y.y.y.4/30 (por definição o λ amarelo usa o sufixo x.x.x.4)
neighbor y.y.y.3 remote-as unknown

Configuração do roteador C:

router bgp 300
network 190.10.10.0/24

interface Ethernet 0/1 (conexão default para o roteador B)
ip address 2.2.2.1/30 (por definição o λ azul usa o sufixo x.x.x.1)
neighbor 1.1.1.2 remote-as 200 (o λ vermelho usa o sufixo x.x.x.2)

interface Ethernet 0/2 (conexão para o OXC)
ip address 4.4.4.4/30 (por definição o λ amarelo usa o sufixo x.x.x.4)
neighbor 4.4.4.3 remote-as unknown (o λ verde usa o sufixo x.x.x.3)

Neste exemplo, assume-se que o roteador virtual faz parte do grupo de pares de troca do roteador B e, portanto, pode compartilhar as mesmas políticas de roteamento e de atualzação em termos de anúncio de rotas. Na realidade a instanciação do roteador virtual é feita na mesma plataforma do roteador B, portanto não é de se estranhar que os roteadores virtuais e reais compartilhem estas informações.

Neste exemplo, assume-se que os comprimentos de onda vermelho e azul são os comprimentos de onda default. Na sessão do BGP OPEN o roteador A anuncia ao roteador B que ele tem uma porta IP e um comprimento de onda na cor verde como

receptor e um comprimento de onda na cor amarela como transmissor através de um protocolo de encapsulamento como o PPP over Ethernet, ATM ou SONET/SDH. A mensagem vinda do roteador A teria as seguintes informações:

Do roteador A para o roteador B:

OPEN AS 100 Loopback 6.6.6.1/32
(neste espaço estariam descritas informações dos campos opcionais)
(as informações atuais seriam retiradas dos campos de dados numéricos)
interface Ethernet 0/2 (Connection to OXC)
ip address 3.3.3.3/30
neighbor 3.3.3.4 remote-as unknown

O roteador A não gera uma mensagem de OPEN para o seu vizinho desconhecido neste momento.

De uma maneira similar o roteador C na sua mensagem de BGP OPEN pode anunciar ao roteador B que ele tem uma porta IP e um comprimento de onda na cor amarela como receptor e um comprimento de onda na cor verde como transmissor através de um protocolo de encapsulamento como o PPP over Ethernet, ATM ou SONET/SDH.

Do roteador C para o roteador B:

OPEN AS 300 Loopback 7.7.7.1/32
(neste espaço estariam descritas informações dos campos opcionais)
(as informações atuais seriam retiradas dos campos de dados numéricos)
interface Ethernet 0/2 (Connection to OXC)
ip address 4.4.4.4/30
neighbor 4.4.4.3 remote-as unknown

O roteador C não gera uma mensagem de OPEN para o seu vizinho desconhecido neste momento.

De uma forma assíncrona o roteador B na sua mensagem de BGP OPEN para os roteadores A e C envia informações que são capazes de suportar conexões ópticas em cruz. A mensagem pode carregar informações como as que seguem:

Do roteador B para o roteador A:

OPEN AS 200 Loopback 5.5.5.1/32
(neste espaço estariam descritas informações dos campos opcionais)
(as informações atuais seriam retiradas dos campos de dados numéricos)
interface loopback0 (indica a presença de um roteador virtual)
ip address 5.5.5.2/32

interface oxc 0/1
ip address x.x.x.4 (pode aceitar o λ amarelo)

interface oxc 0/2
neighbour x.x.x.3 update source loopback (procura o receptor λ verde)

Do roteador B para o roteador C:

OPEN AS 200 Loopback 5.5.5.1/32
(neste espaço estariam descritas informações dos campos opcionais)
(as informações atuais seriam retiradas dos campos de dados numéricos)

interface loopback0 (indica a presença de um roteador virtual)
ip address 5.5.5.2/32

interface oxc 0/1
ip address x.x.x.3 (pode aceitar o λ verde)
interface oxc 0/2
neighbour x.x.x.4 update source loopback (procura o receptor λ amarelo)

O roteador B, após receber mensagens de BGP OPEN dos roteadores A e C, pode decidir de forma assíncrona estabelecer uma conexão em cruz óptica entre os dois roteadores assumindo que existe uma compatibilidade entre o comprimento de onda e o protocolo de encapsulamento. Se o comutador óptico possuir comprimento de onda com lasers e filtros ajustáveis, então o roteador B pode decidir criar uma conexão "em cruz" óptica através do laser ou do filtro ajustáveis no comutador óptico, obviamente "casando" os comprimentos de onda apropriados.

Ao invés de se modificar o código existente do BGP no roteador B, o melhor seria que após a verificação da existência de campos adicionais na mensagem de OPEN, o BGP criasse um processo filho que instanciasse o estabelecimento de um roteador virtual e as conexões em cruz necessárias. Chamamos a isso de LRA "*Lightpath Router Arbiter*"

Neste exemplo, o processo de LRA que roda no roteador B "entende" que o roteador A está apto a receber um comprimento de onda na cor verde na interface que possui o endereço IP 3.3.3.3 e que o roteador C está transmitindo um comprimento de onda na cor amarela para um vizinho desconhecido na interface 3.3.3.4. Da mesma maneira, o processo de LRA vê que o roteador C está apto para receber um comprimento de onda na cor amarela na interface com endereço IP 4.4.4.4 e que o roteador C está transmitindo um comprimento de onda na cor verde para um vizinho desconhecido na interface com endereço IP 4.4.4.3 .

O processo de LRA do roteador B faz a instanciação de uma conexão em cruz óptica entre os roteadores A e B através da criação de um processo filho na CPU do roteador B. A isso, dá-se o nome de roteador BGP virtual. Feito isso o LRA cria um arquivo de configuração para o roteador virtual com as informações recebidas na mensagem de OPEN dos roteadores A e C. O arquivo de configuração do roteador virtual ficaria da seguinte maneira:

Configuração do roteador Virtual criado pelo LRA:

```
interface loopback0
   ip address 5.5.5.2/32

interface oxc 0/1
  ip address 4.4.4.3 (o λ verde vindo do roteador C)
  neighbour 4.4.4.4 update source loopback (o λ amarelo para o roteador C)

interface oxc 0/2
     ip address 3.3.3.4 (o λ amarelo vindo do roteador A)
     neighbour 3.3.3.3 update source loopback (o λ verde para o roteador C)
```

Enquanto o roteador B está configurando o novo roteador virtual, o processo de LRA nos roteadores A e B estão re-escrevendo suas configurações usando as informações obtidas nos campos opcionais na mensagem de OPEN recebidas do rotearod B.

A informação crítica nesta troca é o número do AS remoto ("remote-as"). Os roteadores A e C atualizam suas informações com o número do "remote-as" recebido do roteador B e então eles executam um "soft re-boot" no processo BGP. Feito isso o roteador BGP virtual pode então estabelecer sessões de troca de BGP padrão entre os roteadores A e B.

Se a sessão TCP não puder ser estabelecida ou a sessão BGP com um dos roteadores A ou C não puder ser estabelecida, o roteador B pode decidir deixar o roteador B em um estado de IDLE ou terminar o processo que criou o roteador BGP virtual.

Se o estabelecimento da sessão BGP ocorre com sucesso, a mensagem de BGP UPDATE é usada para trocar as atualizações de endereços LNRI. Por exemplo, o roteador C veria apenas o anúncio das rotas do roteador A (170.10.10.0) através do roteador virtual B. Através do roteador B real, ele veria os anúncios das rotas dos roteadores A (170.10.10.0) e B (180.10.10.10). O roteador C poderia decidir a rota de maior preferência recebida do roteador virtual ou o roteador virtual poderia anunciar as rotas do roteador real B com métricas maiores.

Desta maneira, o tráfego de dados do roteador C para o roteador A iria ser feito através do caminho óptico e o tráfego para o roteador B seria feito via o caminho óptico original.

5.4 Resumo conclusivo do Capítulo

Nesta seção foram descritos os conceitos aplicados no desenvolvimento da rede CA*net4, sucessora da rede CA*net3 que foi a primeira rede Internet óptica do mundo de pesquisa e educação. Foi introduzido o conceito de uma rede orientada às necessidades dos usuários que oferece alocação dinâmica de recursos de rede. Dentro deste paradigma descreveu-se os conceitos de P2P em redes ópticas, da arquitetura WebServices e de uma proposta de protocolo de roteamento para redes ópticas chamado OBPG. Esta proposta é uma extensão do protocolo de roteamento existente na Internet chamado de BGP. Ela introduz alguns novos conceitos, entretanto tenta fazer um elo entre o modelo atual em uso na rede e como isso seria utilizado em redes totalmente implementadas em WDM.

Existe um grupo de trabalho no IETF estudando estas propostas, mas ainda está longe de ser definido como um padrão de facto.

CAPÍTULO 6

CONCLUSÕES

No primeiro capítulo deste trabalho, apresentou-se uma visão geral da convergência da rede IP sobre o WDM; mostrando-se a infra-estrutura atual das redes de telecomunicações, as tecnologias utilizadas, a direção centrada no uso de redes ópticas passivas, notadamente na rede Ethernet, a qual já provê alternativa para servir inúmeros usuários a um custo muito baixo. Apresentaram-se também as razões do porquê se utilizar o IP e o WDM, tendo em vista o fato da migração da rede de circuitos para rede de pacotes ser uma tendência e, mais do que isso, ser um fato. As últimas estimativas em relação à rede de telefonia mundial mostram que dez por cento (10%) do tráfego de voz, algo em torno de 20 bilhões de minutos por ano, já trafega via rede de pacotes. Descreveram-se as características oferecidas pelo WDM, notadamente no aumento significativo de banda na infra-estrutura óptica existente através da multiplicação de canais ópticos em um único par de fibras, mostrando, com isso, que problemas de qualidade de serviço em redes de melhor esforço são atenuados. Mostrou-se que enquanto o WDM oferece uma grande capacidade de banda, o IP oferece convergência, sendo que a combinação de ambos representa um caminho natural. Descreveu-se, também, que o IP sobre WDM necessita vencer desafios existentes na integração de protocolos, sendo que um dos desafios é o roteamento e mapeamento de comprimentos de onda, gerência e acesso aos canais.

No segundo capítulo, foram tratados os Conceitos de Projetos de Protocolos, dando enfoque ao funcionamento das Camadas de Rede e de Transporte na pilha de protocolo TCP/IP. Descreveu-se também o princípio do protocolo fim-a-fim que é a base conceitual da camada de transporte. Detalharam-se, além das transmissões orientadas e não orientadas à conexão, os modelos de entrega de pacotes, seu endereçamento, roteamento, desmontagem e remontagem dos pacotes..

No terceiro capítulo, descreveu-se a transmissão em fibras ópticas, conceituando-se a Propagação da Luz em Fibras Ópticas, através da teoria básica ditada pela Lei de Snell, além das definições dos fenômenos de atenuação, dispersão, bem como, os tipos de fibras existentes e a ação do Laser sobre os mesmos. Descreveu-se o mecanismo existente na regeneração do sinal óptico e a multiplexação do sinal óptico através do WDM.

No quarto capítulo, dando continuidade à elaboração descritiva do conceito sobre o assunto proposto nesta dissertação, descreveram-se as Técnicas de Comutação, através da apresentação dos conceitos sobre comutação por circuito e pacote, o dilema sobre o controle centralizado versus o controle distribuído, a comutação por rajadas e, finalmente, descreveu-se a comutação das redes ópticas, detalhando seus elementos.

No quinto capítulo, descreveu-se o projeto de Redes IP sobre DWDM desenvolvido pela CANAIRE, notadamente de cunho acadêmico e de pesquisa, denominado CA*net4, sobre o qual a análise deste trabalho foi centrada. Dentro do amplo espectro de possíveis estudos, decidiu-se focar no novo conceito de rede, introduzido pela CA*net4, que é a rede “orientada às necessidades dos usuários”. Neste novo conceito a rede oferece alocação dinâmica de recursos da rede para os usuários.

Este novo conceito permite uma maior inovação no desenvolvimento de aplicações de rede, tendo em vista o controle que cada usuário tem da própria rede externa. Introduziu-se o conceito de P2P em redes ópticas, a arquitetura WebServices e o roteamento óptico através do OBGP.

Apresentaram-se possíveis soluções para que a proposta conceitual da rede CA*net4 possa ser implementada, entretanto será necessário o desenvolvimento e o aperfeiçoamento de muitas técnicas apresentadas e propostas. Em relação ao P2P, uma área promissora para pesquisas futuras relacionadas ao gerenciamento da rede feita pelos usuários é o ato de disponibiliza banda larga na última milha com as facilidades competitivas existentes nos provedores de serviços. Uma dessas linhas de pesquisa atuais e que se chama RPON ("Reverse Passive Optical Networking") é mostrada na Figura 6.1. Em contraste com a RPON existem as redes ópticas passivas (PON) cuja tecnologia tem sido utilizada até agora. O propósito original da PON foi prover aos usuários a utilização dos equipamentos existentes a um baixo custo para as redes FTTH ("Fiber To The Home"). Um feixe de laser originado da operadora seria dividido e distribuído para várias casas/usuários através de divisores ópticos passivos (em alguns casos podendo atingir a 32 casas/usuários). Este feixe de laser tem que ser cuidadosamente controlado de tal maneira que a informação de cada uma das diferentes residências possa ser incluída em diferentes "times slots" (Time Division Multiplexing).

Dentro do conceito do RPON, o laser ativo seria gerado nos equipamentos dos usuários e os divisores ópticos estariam localizados em um local neutro da operadora. O sinal do usuário é dividido entre os vários provedores de serviço que estão de alguma maneira conectados neste local neutro. Os prestadores de serviços com os quais os usuários tenham feito contratos, incluiriam o sinal do usuário dentro dos seus equipamentos e enviariam um sinal de retorno através de um comprimento de onda dedicado ou através de um "time slot" TDM.

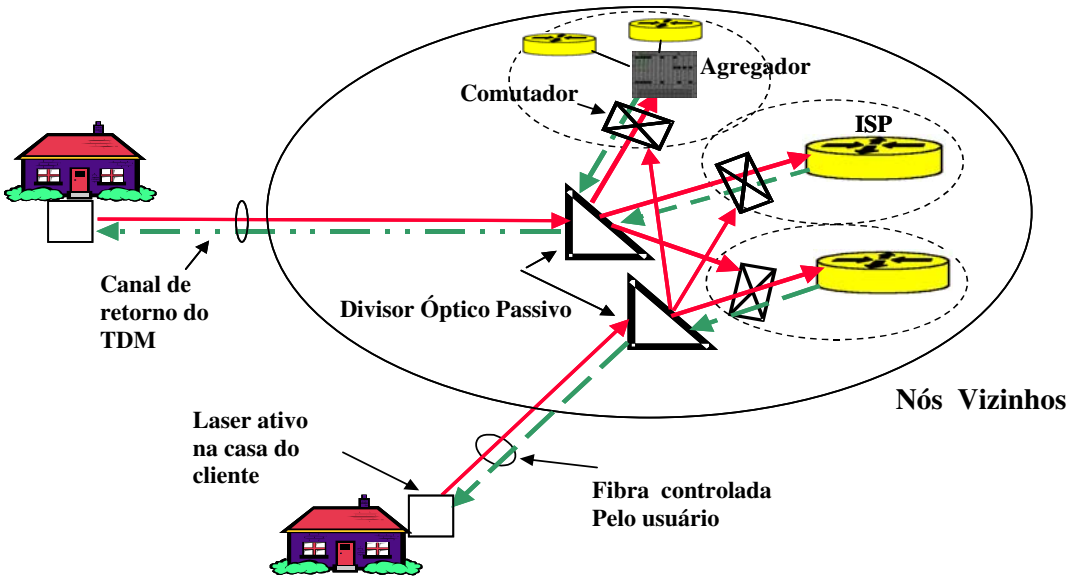


Figura 6.1 Exemplo de Reverse Passive Optical Network (RPON)

O custo dos lasers é atualmente muito alto e, portanto, impeditivo, para que uma solução prática esteja disponível hoje em dia. Entretanto, com o advento de lasers mais baratos e de maior poder, este conceito poderá ser empregado em alguns poucos anos. O ponto alto do RPON é que ele pode controlar e configurar suas próprias conexões comutadas por circuito. Como exemplo, um canal TDM ou de um comprimento de onda originado do equipamento do cliente pode ser utilizado para serviços de pacotes não orientados a conexões, enquanto que os canais adicionais ou comprimento de onda seriam usados para suportar conexões P2P. Um exemplo prático seria um usuário interessado em descarregar um filme em formato DVD de uma rede de conteúdo P2P, como o Morpheus ou o Kazaa. Um "time slot" não orientado à conexão poderia usar uma busca P2P e descobrir a localização do DVD. Uma vez que o DVD está localizado uma aplicação P2P iria configurar uma conexão fim-a-fim através de um "time slot" separado ou um comprimento de onda para poder descarregar múltiplos Gigabytes ou Terabytes de dados relacionados ao DVD.

Uma outra área promissora e desafiadora de pesquisas para se atingir a proposta conceitual da CA*net4 é o OBPG. Apesar de não nos aprofundarmos em detalhes de todos os componentes existentes, o OBGP pode prover uma maneira simples e efetiva para os usuários gerenciarem e operarem as suas próprias redes ópticas, independentemente da infra-estrutura de rede óptica existente nas operadoras. Existem várias perguntas a serem respondidas ainda relativas a escalabilidade e praticidade destes conceitos. Outras questões a serem resolvidas estão relacionadas à definição de serviços, políticas de negociação e mecanismos de controle e medição das políticas acordadas. Talvez estes conceitos só se apliquem para redes pequenas entre entidades que tenham uma relação de confiança bem alicerçada. Isso só saberemos quando os conceitos, aqui apresentados, puderem ser aplicados de uma forma mais ampla e genérica e mais pesquisas tiverem sido feitas.

O DWDM representa um excelente passo em direção a integração das redes atuais, devido a possibilidade de transportar diferentes tecnologias, e ao mesmo tempo dá a possibilidade de transportar mais tráfego sobre uma fibra, do que foi feito antes.Um sistema DWDM aberto é ótimo em fornecer uma solução flexível, considerando o fato de que os provedores de serviço de Internet e as operadoras de telecomunicações no mundo tem vários tipos de equipamentos e fornecedores. A introdução de um sistema DWDM deve ser suave, resolvendo os problemas, de fácil manuseio, permitindo modularidade, que hoje é obrigatório ter, pois a velocidade em que as transformações no mundo são absorvidas, requer capacidade imediata de transação. As operadoras estão diante de um futuro totalmente imprevisível, um rápido crescimento na demanda de capacidade e rápida diminuição dos prazos para alocar estas capacidades serão requeridas.

Estamos caminhando para uma sociedade dos Terabytes de informação. As pessoas necessitarão de comunicação em qualquer lugar do planeta em tempo real. Telefones móveis estão se tornando cada vez menores e os serviços exigidos demandam maior largura de banda. Novos padrões para comunicações móveis exigirão largura de banda para tráfego "on line" de alta capacidade de dados. Conexões de alta velocidade fornecerão acesso a comunicação de dados nas grandes corporações, instituições de ensino e residências. Ainda assim, a necessidade por maiores capacidades de tráfego continuará. Os custos de equipamentos e novas tecnologias são ainda bastante significantes, mas tendência num futuro breve é a queda destes valores, o que tornará

possível a implementação de novas tecnologias bastante acessíveis. Neste contexto o FTTH (Fiber to the Home) será uma realidade. As fibras ópticas chegarão até as residências, que tornarão as comunicações mais ágeis, exigindo banda, velocidade, switching e modularidade capaz de atender toda esta demanda explosiva.

A tecnologia DWDM tornará possível a Super Via de Informação, e em termos simples, representará um grande passo para aumentar a capacidade de responder à rápida explosão na demanda causada pela Internet. A longo prazo, ela fornecerá um meio para a integração das redes, fornecendo um aumento de capacidade, sem um aumento explosivo de custo, o que é impossível de se obter hoje em dia.

Num futuro breve, a sociedade será capaz de obter serviços que demandem alta capacidade de forma integrada, como Internet em alta velocidade, videoconferência, acesso banda larga (incluindo wireless) e multimídia. O uso massivo de serviços de banda larga e suas aplicações serão os propulsores para o desenvolvimento das redes de comunicações em patamares nunca antes imaginados.

REFERÊNCIA BIBLIOGRÁFICA

[1] S. Dixit, IP over DWDM, *Building the Next-Generation Optical Internet,* Edited by S. Dixit, Wiley & Sons publication, ISBN 0-471-21248-2, 2003.

[2] J.Saltzer, D. Reed, and D. Clark*, End-to-end arguments in a system design*, ACM Transactions on Computer Systems, 2(4):195-206, 1984.

[3] M. Blumenthal and D. Clark, *Rethinking the design of the Internet: The end to end arguments versus the brave new world*, ACM Transactions on the Internet Technology, 1(1):70-109, August 2001.

[4] J. Mahdavi and S. Floyd, *TCP-friendly unicast rate-based flow control, unpublished note*, Jan, 1997. http://www.psc.edu/networking/papers/tcp_friendly.html

[5] S. Floyd, M. Handley, J. Padhye, and J. Widmer, *Equation-based congestion control for unicast applications,* Proceedings of ACM SIGCOMM, Stockholm, Sweden, Aug. 2000.

[6] A. E. LIMA, Michele M.; FONSECA, Nelson Luis S. da; *"Controle de Tráfego Internet",* "Livro Texto dos Mini-Cursos SBRC", 04/2002, ed. 1, SBRC, pp. 63, pp.187-249, 2002

[7] W. R. Stevens, *TCP/IP Illustraded Vol. 1*, Addison-Wesley, Reading, MA, 1994.

[8] K. C. Kao e A. G. Hockham , *Dielectric surface waveguides for optical frequencies*, *Proc. IEEE*, vol. 113, pp. 1151–1158, 1966

[9] J.E. Midwinter, *The Start of Optical Fiber Communications as Seen from a U.K.Perspective,* IEEE Journal on Selected Topics in Quantum Electronics, vol. 6, no. 6, p.1307 a 1311. Novembro/Dezembro de 2000.

[10] R. Ramaswami e K. Sivarajan, *Optical Networks: A Practical Perspective*, 2a ed., Morgan-Kaufman, 2001.

[11] Professor Raj Jain's homepage, http://www.cis.ohio-state.edu/~jain

[12] A. Tanenbaum, *Computer Networks*, 3rd ed., Prentice Hall, 1996.

[13] J. Hui., *Switching and Traffic Theory for Integrated Broadband Networks*, Kluwer Academic Publishers, 1990.

[14] R. Calon et al., *A framework for multiprotocol label switching*, IETF Draft, September 1999.

[15] E. Rosen et al., *Multiprotocol label switching architecture*, IEEE Journal on Selected Areas in Communications, August 1999.

[16] G. Rouskas and V. Sivaraman., On the design of optimal TWDM schedules for broadcast WDM networks with arbitrary tranciever tuning latencies., In Proceedings of IEEE Infocom, pages 1217–1224, 1996.

[17] A. Amstutz, Burst switching – an introduction, Communications Magazine, 21:36-42, November 1983.

[18] E. Haselton., *A PCM frame switching concept leading to burst switching network architecture*.,IEEE Communications Magazine, 21:13–19, June 1983.

[19] P. Kermani and L. Kleinrock., *Virtual cut-through : A new computer communication switching technique*., *Computer Networks*, 3:267–286, 1979.

[20] I. Chlamtac, A. Ganz, and G. Karmi. *Lightpath communications: an approach to highbandwidth optical WANs*., Transactions on Communications, 40:1171–1182, July 1992.

[21] C.Qiao and Y. Mei. *On the multiplexing degree required to embed permutations in a class of interconnection networks*,, IEEE Symp. High Performance Computer Architecture, pages 118–129, February 1996.
[22] C.Qiao and Y.Mei., *Wavelength reservation under distributed control*, In IEEE/LEOS Broadband Optical Networks, August 1996.
[23] R.Ramaswami and A. Segall., *Distributed network control for wavelength routed optical networks*., In Proceedings of IEEE Infocom, pages 138–147, March 1996.
[24] A. Sengupta et al., *On an adaptive algorithm for routing in all-optical networks*, In SPIE Proceedings, All optical Communication Systems: Architecture, Control and Network Issues, volume 3230, pages 288–297, November 1997.
[25] J. Bannister et al., *How many wavelengths do we really need in an optical backbone network*, in IEEE Gigabit Networking Workshop (GBN), 1999.
[26] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer., *Photonic packet switches: Architectures and experimental implementations*, in proceedings of the IEEE, 82(11):1650–1667, November 1994.
[27] F.Masetti, P.Gavignet-Morin, D. Chiaroni, and G. Da Loura, *Fiber delay lines optical buffer for ATM photonic switching applications*, in Proceedings of IEEE Infocom, volume 3, pages 935–942, 1993.
[28] G. Chang, *Optical label switching*, in DARPA/ITO Next Generation Internet PI meeting, Oct, 1998.
[29] M. Yoo, M. Jeong, and C. Qiao, *A high speed protocol for bursty traffic in optical networks*, in SPIE Proceedings, All optical Communication Systems: Architecture , Control and Network Issues, volume 3230, pages 79–90, November 1997.
[30] M. Yoo and C. Qiao, *Just-enough-time(JET): a high speed protocol for bursty traffic in optical networks*, In IEEE/LEOS Technologies for a Global Information Infrastructure, pages 26–27, August 1997.
[31] M. Yoo and C. Qiao, *A novel switching paradigm for buffer-less WDM networks*, in Optical Fiber Communications (OFC), pages 177–179, March 1999.
[32] Aboul-Magd, et al., *User Network Interface (UNI) 1.0 Signaling Specification.* Optical Internetworking Forum (OIF), 2001.
[33] Parameswaran M., Susarla A. and Whinston A.B., *P2P Networking: An Information-Sharing Alternative.*, IEEE Computer, 34 (2001), pp. 31-36
[34] Macedonia M., *Distributed File Sharing. Barbarians at the Gates? ,* IEEE Computer, 33 (2001) pp. 99-101.
[35] Schollmeier R., *A Definition of Peer-To-Peer Networking for the Classification of Peer-To-Peer Architectures and Applications.* Proceddings of the 2001 International Conference on Peer-To-Peer Computing Aug. 37-29, 2001, Linkopings University, Sweden, pp. 101-102.
[36] St. Arnaud B., *Proposed CA*net 4 Network Design and Research Program.,*.http://www.canet3.net/library/papers/CAnet4_Design_Document.doc.

[37] Francisco M.J, PEzoulas L., Huang C. and Lambadaris I., , *End-to-End Signalling and Routing for Optical IP Networks.* Proceedings of the 2001 International Conference on Peer-To-Peer Computing Aug. 37-29, 2001, Linkopings University, Sweden, pp. 101-102.

[38] Luciani, J., Rajagopalan, B., Awduche, D., Cain, B., Jamoussi, B., *IP over Optical Networks – A Framework*, draft-ip-optical-framework-00.txt, September 10, 2000.

[39] Soares L., Lemos G., Colcher S., *Redes de Computadores*, 6ª. Edição, Ed. Campus, ISBN 85-7001-998-X.

[40] C.Qiao and Y.Mei., *Wavelength reservation under distributed control*, In sIEEE/LEOS Broadband Optical Networks, August 1996.

[41] Halabi, B., *Internet Routing Architectures,* Cisco Press, Indianapolis, IN, 1997.

[42] Awduche, D., Rekhter, Y., Darke, J., Coltun, R., *Multi-protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects*, draft-awduche-mple-te-optical-01.txt.

[43] Stanton M., Abelem A., *Redes Ópticas,* trabalho apresentado na SBRC em 2002.

[44] Abelem A., *Difusão Seletiva em Inter-Redes IP baseadas em Redes Ópticas*, tese de doutorado, Departamento de Informática, PUC-RJ, Abril de 2003.

[45] S. Mukherjee, *WDM Optical Communications Networks: Progress and Challenges.* IEEE Journal on Selected Areas in Communications, 18(10):1810-1024, Oct. 2000.

[46] W. Stallings, Data and Computer Communications, 5th Edition, Prentice Hall.

[47] J.F. Kurose and K. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*, Addison-Wesley, Reading, MA, 2001.