

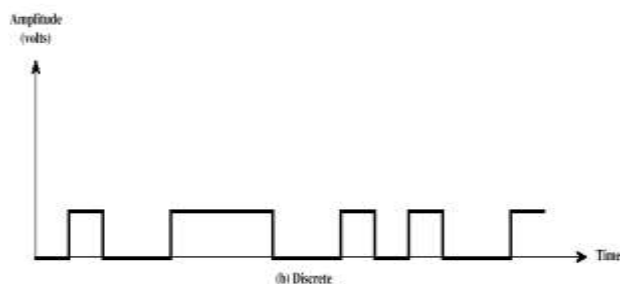
## TRANSMISSÃO DE DADOS

O sucesso de uma transmissão de dados depende principalmente de dois fatores: (i) da qualidade do sinal sendo transmitido e, (ii) das características do meio de transmissão. O objetivo deste capítulo é prover o leitor com um sentimento intuitivo da natureza desses dois fatores.

### 1. Base Teórica sobre Sinais de Transmissão

Toda transmissão de dados ocorre pela inserção de sinais, em geral na forma de ondas eletromagnéticas, em algum sistema de transmissão. O sucesso dessa transmissão depende principalmente de dois fatores: (i) a qualidade do sinal transmitido e (ii) as características do meio de transmissão. Por isso, antes de formalizar os conceitos sobre sinais, é preciso primeiro repassar algumas terminologias utilizadas na comunicação de dados:

- ⇒ Sobre os meios de transmissão: Esses podem ser guiados se fizerem uso de cabos (tipo par trançado, coaxial ou fibra), ou podem ser não guiados se usarem meios de transmissão dispersos (como o espaço aéreo, água do mar, vácuo etc.) como caminho para a transmissão. Em ambos os casos a comunicação, em geral ocorre na forma de ondas eletromagnéticas.
- ⇒ Sobre a forma de utilização do meio de transmissão: Se os dispositivos comunicantes se interligam por um enlace direto (sem qualquer dispositivo ativo, a não ser amplificadores ou repetidores de sinais), diz-se que o caminho entre eles é ponto-a-ponto. Numa configuração multi-ponto ou *broadcast* mais de dois dispositivos compartilham o mesmo meio de transmissão.
- ⇒ Sobre a tipo de diálogo entre os dispositivos comunicantes: Uma transmissão pode ser *simplex*, *half-duplex* ou *full-duplex*. No primeiro caso, os sinais são transmitidos apenas em uma direção, configurando uma estação como transmissora e a outra como receptora. No segundo caso, ambas as estações podem transmitir, mas apenas uma por vez. Na transmissão full-duplex ambas as estações podem transmitir ao mesmo tempo.



sinal periódico digital (onda quadrada).

Matematicamente, um sinal  $s(t)$  é definido como periódico se e somente se:

$$s(t + T) = s(t); \quad -\infty < t < +\infty,$$

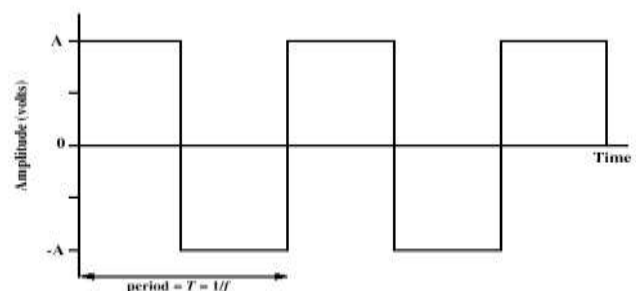
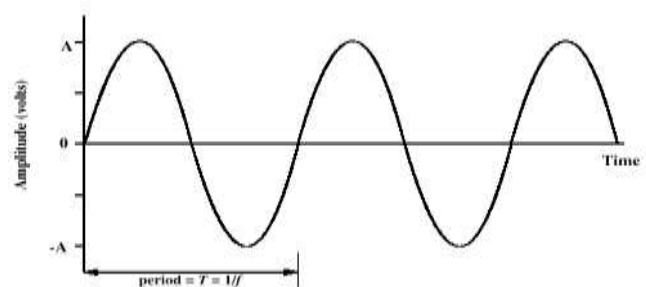
onde a constante  $T$  é o período do sinal, ou seja, é o menor valor que satisfaz a equação. O sinal será considerado aperiódico se não seguir esta formalização.

Dentro do contexto da representação de sinais, a onda senoidal é o sinal contínuo fundamental e, de forma genérica, pode ser representada por três parâmetros:

#### 1.1 Definição de Sinais no Domínio do Tempo

Visualizado como uma função do tempo, um sinal eletromagnético pode ser ou contínuo ou discreto. Um sinal contínuo é aquele no qual a sua intensidade varia suavemente no tempo, ou seja, não existem quebras ou descontinuidades no sinal<sup>1</sup>. Um sinal discreto, por sua vez, é aquele no qual a intensidade desse sinal se mantém num nível constante por um período de tempo e então muda para um outro nível constante. A figura ao lado apresenta ambos os tipos de sinais. O sinal contínuo poderia representar a fala humana e o sinal discreto poderia representar os binários 1s e 0s.

O mais simples tipo de sinal é o sinal periódico, no qual um mesmo padrão de sinal se repete com o tempo. A figura abaixo apresenta um exemplo de um sinal periódico analógico (onda senoidal) e um



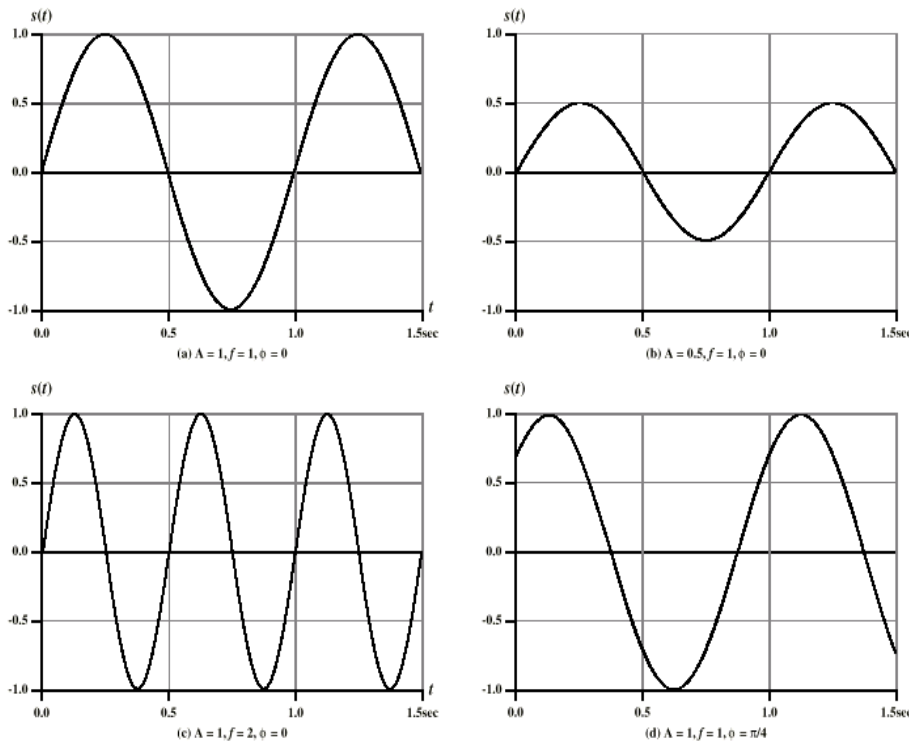
<sup>1</sup> Uma definição matemática: Um sinal  $s(t)$  é contínuo se  $\lim_{a \rightarrow t} s(a) = s(t)$  para todo  $a, t \rightarrow a$ .

- i. Amplitude - é o valor de pico ou a força do sinal em relação ao tempo. Tipicamente, este valor é medido em volts ou watts.
- ii. Frequência ( $f$ ) - é a taxa (em ciclos por segundo ou Hertz) na qual o sinal se repete. Um parâmetro equivalente é o período ( $T$ ) de um sinal, o qual é a quantidade de tempo que leva para que uma repetição ocorra. Portanto,  $T = 1/f$ .
- iii. Fase - é a medida da posição relativa dessa onda no tempo em relação a um único período do sinal, como ilustrado na figura abaixo. Para compreendê-la, devemos entender que uma onda senoidal genérica pode ser escrita da seguinte maneira:

$$s(t) = A \sin(2\pi ft + \phi),$$

onde:  $A$  = Amplitude do sinal;  $f$  = frequência do sinal;  $\phi$  = Fase do sinal

A figura ao lado apresenta o efeito de variar cada um desses três parâmetros. No primeiro quadro, a frequência é 1 Hz e portanto o período é  $T = 1$  segundo. O segundo quadro (da direita) tem a mesma frequência e fase, mas a amplitude é de  $1/2$ . No terceiro quadro, tem-se  $f = 2$ , o qual é equivalente a  $T = 1/2$ . Finalmente, o último quadro apresenta o efeito de uma fase alterada para  $\pi/4$  radianos, a qual é equivalente a 45 graus ( $2\pi$  radianos é igual a 360 graus = 1 período).



### 1.1.1 Ondas Senoidais em Relação ao Espaço

Na figura anterior, houve a apresentação do sinal em função do tempo num dado ponto do espaço. Se, por outro lado, considerarmos o ponto zero do gráfico como sendo a origem do sinal e o eixo horizontal como sendo o eixo do espaço, será possível perceber o comportamento do sinal em relação à sua origem, como uma onda de rádio eletromagnética distante em relação a uma antena de rádio, ou o formato de um sinal sonoro alguma distância em relação à caixa de som que lhe deu origem. Portanto, pode-se concluir que em um instante particular de tempo a intensidade do sinal varia em uma forma senoidal como uma função da distância da fonte.

Sobre o ponto de vista do espaço, pode-se considerar o comprimento de onda,  $\lambda$ , de um sinal como sendo a distância ocupada por um único ciclo, ou, de outra maneira, como a distância entre dois pontos de fase correspondente de dois ciclos consecutivos. Assumindo que o sinal está viajando com uma velocidade  $v$  então o comprimento de onda é relacionado ao período da seguinte maneira:  $\lambda = vT$ , ou, equivalentemente,  $\lambda f = v$ . Essa discussão tem relevância particular para o caso onde  $v = c$ , ou seja, a velocidade da luz no espaço livre, a qual é igual a  $3 \times 10^8$  m/s.

### 1.2 Definição de Sinais no Domínio da Frequência

A informação pode ser transmitida em cabos pela variação de alguma de suas propriedades físicas tais como a voltagem ou a corrente, permitindo formalizar sinais em função do tempo. No entanto, pode-se demonstrar também (usando uma disciplina conhecida como análise de Fourier) que qualquer sinal é formado por componentes de diferentes frequências, na qual cada componente é uma senoidal. Um exemplo disso está apresentado na próxima figura. As componentes deste sinal são exatamente as ondas senoidais de frequências  $f_1$  e  $3f_1$ , nesse caso chamadas de componentes de  $s(t)$ . As partes (a) e (b) da figura a seguir apresentam essas componentes individualmente. A parte (c) é a onda resultante da soma das duas anteriores, representada pela seguinte fórmula:

$$s(t) = \sin(2\pi f_1 t) + 1/3 \sin(2\pi(3f_1)t)$$

Existem vários pontos de vista interessantes que podem ser percebidos nesta figura:

- A segunda frequência é um múltiplo inteiro da primeira frequência. Quando todas as componentes de frequência de um sinal são múltiplos inteiros de uma frequência, essa última é referenciada como sendo a frequência fundamental.
- O período do sinal resultante é igual ao período da frequência fundamental, ou seja, o período da componente  $\sin(2\pi f_1 t)$  é  $T = 1/f_1$  e o período de  $s(t)$  é também  $T$ .

- Seguindo o raciocínio proposto na figura, ao adicionarmos mais um ou dois termos ou componentes<sup>2</sup> à essa onda, as partes alta e baixa da onda resultante tenderá a um sinal discreto, se aproximando cada vez mais de uma onda quadrada. Para mais dois harmônicos a fórmula ficaria da seguinte maneira:

$$s(t) = \sin(2\pi f_1 t) + \frac{1}{3} \sin(2\pi(3f_1)t) + \frac{1}{5} \sin(2\pi(5f_1)t) + \frac{1}{7} \sin(2\pi(7f_1)t)$$

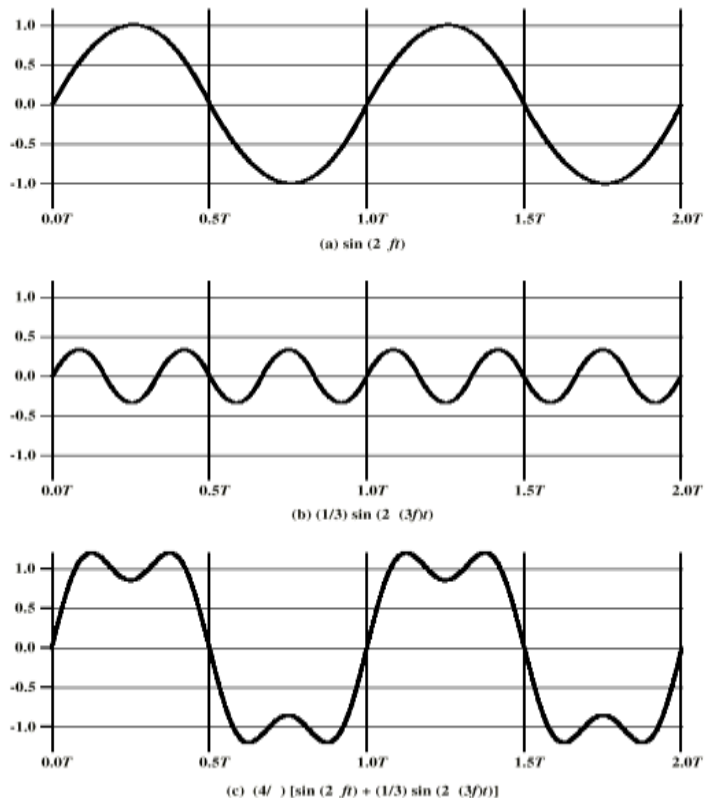
- Uma onda quadrada genérica possui a seguinte fórmula (originada da série de Fourier), onde a amplitude da k-ésima componente de frequência,  $Kf_1$ , é apenas  $1/k$  (nesse caso,  $A$  = Amplitude do sinal e  $f_1$  é a frequência desse sinal):

$$s(t) = A \times \sum_{K=1,3,5,\dots}^{\infty} \frac{1}{K} \sin(2\pi K f_1 t),$$

Portanto, pode-se dizer que para cada sinal, existe uma função no domínio do tempo  $s(t)$  que especifica a amplitude do sinal no instante do tempo e, similarmente, existe uma função  $S(f)$  que especifica as frequências constituintes deste sinal.

Antes de continuarmos nossa discussão sobre as componentes de um sinal, é válido definir alguns conceitos no domínio da frequência cuja compreensão são importantes:

- ⇒ **Espectro** - O espectro de um sinal é o domínio de frequências que ele contém. Para o sinal da figura ao lado, o espectro se estende de  $f_1$  até  $3f_1$ .
- ⇒ **Largura de banda absoluta** de um sinal é a largura do espectro. No caso da figura ao lado, a largura de banda é  $2f_1$ , ou seja, a diferença entre a maior e a menor frequência do sinal. Vale ressaltar que muitos sinais tem uma largura de banda infinita mas a maior parte da energia deles está contida nas primeiras frequências, formando uma banda relativamente estreita de frequências. Esta banda onde a maior parte da energia está concentrada é referenciada como a largura de banda efetiva, ou, simplesmente largura de banda da frequência fundamental.



### 1.2.1 Relação entre Taxa de Transmissão e Largura de Banda

É sabido que nenhum meio de transmissão consegue transmitir sinais sem perdas porque cada meio possui uma largura de banda inerente às suas propriedades físicas, o que limita a quantidade de frequências do sinal que podem ser transportadas. Por outro lado, baseado no que foi dito na seção anterior, a maior parte da energia do sinal se situa nas primeiras frequências ou componentes deste sinal. Portanto, é possível transportar um sinal digital, não como uma onda quadrada, mas como uma aproximação dessa onda quadrada (considerando somente as frequências que armazenam uma quantidade maior de energia e desconsiderando as demais) cujas bases superior e inferior serão mais retas ou abauladas em função das capacidades de percepção do receptor.

Em função disso, pode-se dizer que a taxa de transmissão de um sinal depende basicamente de dois fatores:

- i. Do método de codificação que é usado para representar os bits. Existem vários métodos de representação de bits, sendo que os pioneiros são aqueles em que os bits são representados por voltagens fixas pré-estabelecidas. Esse assunto será visto com detalhes mais adiante e vamos, por agora, considerar apenas a situação na qual o bit zero é representado pela parte baixa de uma onda quadrada (ou aproximada) e o bit um é representado pela parte alta dessa mesma onda quadrada.
- ii. Da quantidade de alterações que o sinal sofre no tempo<sup>3</sup>, ou seja, quantas vezes ele oscila entre as voltagens alta e baixa no tempo. No caso da Figura anterior, considerando que o pulso positivo representa o binário 1 e o pulso negativo representa o binário zero tem-se a representação da sequência 1010.... A duração de cada pulso é igual a  $1/2f_1$ ; portanto, a taxa de dados é  $2f_1$  bits por segundo (bps). Em outras palavras, a taxa de transmissão está diretamente relacionada à frequência fundamental do sinal, que é a região de maior energia do sinal.

Portanto, um meio de transmissão com determinada largura de banda pode produzir diferentes taxas de transmissão em função dos seguintes fatores:

- i. Da qualidade do sinal implícita no número de componentes que o sinal necessita para que o receptor possa identificar os bits. Significa dizer que, para dada uma largura de banda, quanto maior o número de componentes, menor deve ser a frequência fundamental e, consequentemente, a taxa de transmissão;

<sup>2</sup> Os termos são relacionados às funções seno e co-seno da Série de Fourier. Portanto cada termo possui dois harmônicos.

<sup>3</sup> Essa definição é conhecida como baud.

- ii. Da frequência fundamental necessária para gerar os bits embutidos nas voltagens do sinal. Ou seja, quanto maior for a frequência fundamental maior a taxa de transmissão. Em contrapartida, é possível provar que a frequência fundamental não pode ser mais que a metade da largura de banda do meio.

Em outras palavras, se o sinal contiver muitas componentes, a qualidade do sinal é boa, mas a taxa de transmissão pode ser baixa. Em contrapartida, se o número de componentes for menor, a qualidade do sinal diminui mas a taxa de transmissão pode aumentar<sup>4</sup>. Vejamos alguns exemplos:

1) Vamos supor que estamos utilizando um sistema de transmissão digital que é capaz de transmitir sinais com uma largura de banda de 4MHz e vamos transmitir uma sequência alternada de 1s e 0s como uma onda quadrada considerando apenas duas componentes. Qual a taxa de dados que pode ser alcançada? Embora esta forma de onda seja distorcida de uma onda quadrada, ela é suficiente para que o receptor possa identificar um bit zero de um bit 1. Agora, se tomarmos a frequência fundamental  $f_1 = 10^6$  ciclos/segundo = 1 MHz, então a largura de banda do sinal:

$$s(t) = \sin(2\pi \times 10^6 t) + \frac{1}{3} \sin(2\pi(3 \times 10^6) t) + \frac{1}{5} \sin(2\pi(5 \times 10^6) t)$$

é  $(5 \times 10^6) - 10^6 = 4\text{MHz}$ . Note que para  $f_1 = 1\text{MHz}$  o período da frequência fundamental é  $T = 1/10^6 = 1\mu\text{segundo}$  (microsegundo). Se tratarmos este comprimento de onda como uma sequência de 1s e 0s, um bit ocorre a cada  $0,5 \mu\text{s}$ , formando uma taxa de dados de  $2 \times 10^6 = 2\text{Mbps}$ . Portanto, para uma largura de banda de 4MHz, uma taxa de dados de 2 Mbps pode ser alcançada.

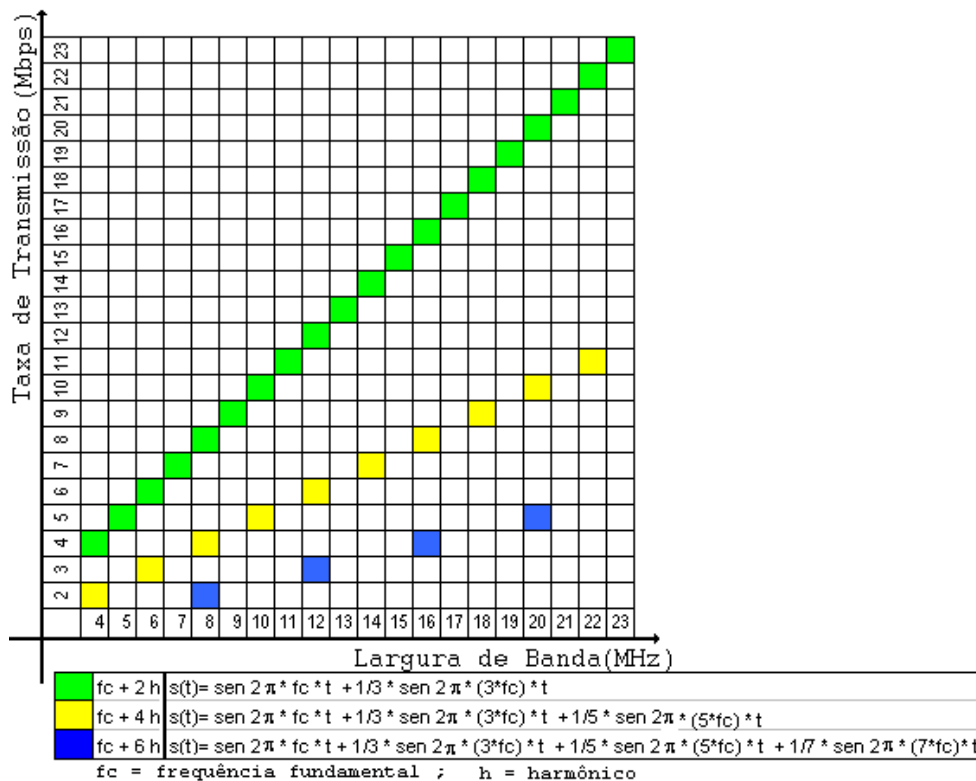
2) Considere o exemplo anterior, mas com largura de banda igual a 8MHz e frequência fundamental  $f_1 = 2\text{MHz}$ . Usando o mesmo raciocínio (número de componentes), a largura de banda do sinal é  $(5 \times 2 \times 10^6) - 2 \times 10^6 = 8\text{MHz}$ . Mas, neste caso  $T = 1/f_1 = 0,5\mu\text{s}$ . Como resultado disto, um bit ocorre a cada  $0,25\mu\text{s}$  para uma taxa de dados de 4Mbps. A partir deste resultado, considerando constante os demais fatores, pode-se concluir que, ao dobrar a largura de banda do meio, dobra-se também a potencial taxa de dados.

3) Considerando um sistema de transmissão com capacidade de transmitir sinais com largura de banda de 4MHz, se fizermos a frequência fundamental  $f_1 = 2\text{MHz}$  e com apenas uma componente extra (mínimo aproximado de onda quadrada para que os bits sejam perceptíveis), usando a mesma linha de raciocínio, a largura de banda do sinal é  $(3 \times 2 \times 10^6) - 2 \times 10^6 = 4\text{MHz}$ . Mas, neste caso,  $T = 1/f_1 = 0,5\mu\text{s}$ . Como resultado, um bit ocorre a cada  $0,25\mu\text{s}$ , totalizando uma taxa de dados na ordem de 4Mbps.

Pode-se desenhar as conclusões gerais a partir do que foi exposto baseado nos resultados dos exemplos:

- Uma dada largura de banda pode suportar várias taxas de dados dependendo dos requerimentos do receptor, de acordo com o número de componentes do sinal e da largura de banda (frequência fundamental) definida para o mesmo. O gráfico adiante apresenta esta constatação.
- Qualquer forma de onda digital terá largura de banda infinita. Portanto, se de um lado as razões práticas e econômicas apontam para que a informação digital seja aproximada por um sinal com largura de banda limitada, por outro lado, limitar a largura de banda cria distorções, as quais tornam a tarefa de interpretar o sinal recebido mais difícil. Quanto mais limitada a largura de banda, maior a distorção e maior o potencial de erros pelo receptor.
- Se a taxa de dados do sinal digital é  $W$  bps, então uma boa representação pode ser alcançada com uma largura de banda em torno de  $2W\text{Hz}$ . Contudo, se o padrão de ruídos não for muito severo, é possível recuperar o padrão de bits com uma largura de banda menor do que essa.
- Existe uma relação direta entre taxa de dados e largura de banda: quanto maior a taxa de dados de um sinal, maior a largura de banda efetiva. Olhando de uma outra maneira, quanto maior a largura de banda de um sistema de transmissão, maior é a taxa de dados que pode ser transmitida sobre este sistema.
- Uma outra observação que pode ser feita é: se pensarmos que a largura de banda de um sinal está centrada sobre alguma frequência, referenciada como frequência central, então quanto maior a frequência central, maior a largura de banda potencial e, portanto, maior a potencial taxa de dados. Considerando que se um sinal está centrado em 2MHz, sua largura de banda máxima é 4MHz.

<sup>4</sup> Essa diminuição, no entanto, deve ser controlada de tal modo que haja um número mínimo de componentes para que o sinal possa ser perceptível pelo receptor.



## 2. Imperfeições na Transmissão

Com qualquer sistema de comunicação deve estar claro para o leitor que o sinal recebido irá diferir do sinal transmitido devido às várias imperfeições na transmissão. Para sinais analógicos, essas imperfeições introduzem várias modificações randômicas que degradam a qualidade do sinal. Para sinais digitais, os erros são introduzidos pela inversão de bits, ou seja, bit zero se transforma em bit 1 e vice-versa. Nesta seção examinar-se-á as várias imperfeições e far-se-ão comentários sobre seus efeitos na capacidade de transportar informações de um link de comunicação. As imperfeições mais significantes são a atenuação, a distorção por atraso e os ruídos.

**a) Atenuação** - A força de um sinal cai com a distância sobre qualquer meio de transmissão. Para meios guiados, esta redução na força, ou atenuação, é geralmente logarítmica e é tipicamente expressa como um número constante de decibéis por unidade de distância. Para meios não guiados (meio aéreo), a atenuação é uma função mais complexa de distância e de makeup da atmosfera. A atenuação introduz três considerações para os engenheiros de transmissão. A primeira delas é que o sinal recebido deve ser suficientemente forte para de modo que o circuito eletrônico no receptor possa detectar e interpretar o sinal. Segundo, o sinal deve manter um nível suficientemente maior que o ruído inerente ao meio para que possa ser recebido sem erros. Terceiro, a atenuação é uma função crescente em relação à frequência. O primeiro e o segundo problemas são tratados pelo controle da força do sinal e pela utilização de amplificadores ou repetidores. Para um enlace ponto-a-ponto, a força do sinal do transmissor deve ser forte o suficiente para que a inteligibilidade do sinal recebido, mas não tão forte para gerar uma sobrecarga no circuito do transmissor, o qual poderia causar uma distorção do sinal a ser gerado. Em distâncias muito grandes, a atenuação é muito grande e, nesse caso, os repetidores e amplificadores são usados para reforçar o sinal de tempos em tempos. Esses problemas são mais complexos para enlaces multiponto onde a distância do transmissor para o receptor é variável, como é o caso das redes Cable Modem, por exemplo. O terceiro problema é particularmente notável para sinais analógicos. Por que a atenuação varia em função da frequência, o sinal recebido é distorcido, reduzindo a sua inteligibilidade. Para contornar este problema, técnicas estão disponíveis para equalizar a atenuação através de uma banda de frequências. Isto é comumente feito para linhas telefônicas de grade-voz por usar bobinas de carga enroladas que mudam as propriedades elétricas da linha; o resultado é uma diminuição dos efeitos da atenuação. Uma outra abordagem é usar amplificadores capazes de amplificar mais as frequências maiores do que as menores.

**b) Distorção por atraso** - É um fenômeno peculiar a meios de transmissão guiados. A distorção é causada pelo fato de que a velocidade de propagação de um sinal através do meio guiado varia com a frequência. Para um sinal com largura de banda limitada, a velocidade tende a ser a maior próxima da frequência central e menor fora das duas arestas da banda. Portanto, várias componentes de frequência de um sinal chegarão no receptor em tempos diferentes. Este efeito é referenciado como uma distorção por atraso, por que o sinal recebido é distorcido devido ao atraso variável de suas componentes. Distorção por atraso é particularmente crítico para dados digitais. Considerando o caso que uma sequência de bits está sendo transmitida, usando sinais analógicos ou digitais. Por causa da distorção por atraso, alguns dos componentes do sinal de uma posição de bit irão pular sobre outras posições de bits causando interferência intersimbólica, a qual é a maior limitação para a taxa máxima de bits sobre um controle de transmissão.

**c) Ruídos** - Em qualquer transmissão de dados, o sinal recebido consistirá do sinal transmitido, modificado pelas várias distorções impostas pelo sistema de transmissão, mais os sinais indesejados que são inseridos durante a transmissão. Alguns tipos de ruídos podem ser mencionados:

- ⇒ Ruído Termal - Ocorre devido à agitação de elétrons no condutor. Está presente em todos os dispositivos e meios de transmissão e é uma função da temperatura. É uniformemente distribuído através do espectro de frequências e é normalmente referenciado como ruído branco. Não pode ser eliminado e impõe limites superiores à performance dos sistemas de comunicação.
- ⇒ Ruído de Intermodulação - Ocorre quando sinais de diferentes frequências compartilham o mesmo meio de transmissão. O efeito é produzir sinais numa frequência que é a soma, a diferença ou é um múltiplo dos dois sinais originais.
- ⇒ Crosstalk - É experimentado por qualquer um que, enquanto usando o telefone, consegue ouvir uma segunda conversação (linha cruzada). Isto é, na verdade, um acoplamento não desejado entre caminhos de sinais. Pode ocorrer por acoplamento elétrico, geralmente em par trançado ou cabo coaxial.
- ⇒ Ruído Impulsivo - São alterações de energia no sinal que ocorrem de forma aleatória e podem provocar uma alteração considerável no valor da informação contida no sinal, seja ele analógico ou digital. Em particular, os sinais digitais sofrem mais com esse tipo de ruído. Por exemplo, a transmissão de voz pode ser corrompida por pequenas alterações mas as entidades comunicantes ainda conseguem compreender o diálogo. Já numa transmissão de bits, seja em sinal analógico ou digital não há como ignorar estes surtos de voltagem/ruído na linha.

## 2.1 Capacidade do Canal

Em função das imperfeições naturais que podem distorcer ou corromper um sinal. Para dados digitais, a questão que então surge é que a extensão dessas imperfeições limita a taxa de dados que podem ser alcançados.

A taxa na qual o dado pode ser transmitido sobre um meio de comunicação qualquer, ou canal, sobre determinadas condições, é chamada capacidade do canal. Na prática um sinal não pode ter uma taxa de transmissão infinita, por que existem as imperfeições do meio de transmissão que a impedem. Dentro deste contexto, existem quatro conceitos relacionados entre si:

- ⇒ Taxa de Dados - Isto é a taxa de dados em bits por segundo na qual um dado pode ser comunicado
- ⇒ Largura de banda - É a largura do sinal transmitido como restringido pelo transmissor e pela natureza do meio de transmissão, expresso em ciclos por segundo ou Hertz.
- ⇒ Ruído - O nível médio de interferências não desejadas no caminho ou meio de transmissão.
- ⇒ Taxa de erros - A taxa na qual os erros ocorrem, onde um erro é definido como sendo a transformação indesejada de um bit 1 transmitido em zero e vice-versa.

## CODIFICAÇÃO DE DADOS

Na transmissão de dados de uma origem para um destino, uma questão importante é entender qual a natureza desse dado, o meio usado para propagá-lo e quais processamentos ou ajustes que devem ser feitos ao longo da transmissão para garantir a sua recepção de forma inteligível. Para responder a todos estes requisitos, é imperativo compreender a aplicação dos termos analógico e digital no contexto da (i) definição dos dados, (ii) da sinalização, e (iii) da transmissão propriamente dita. Os primeiros dois itens são tratados na seção 1 enquanto o terceiro item é tratado na seção 2.

### 1. Tipos de Dados e Sinais

Conceitualmente, dados podem ser considerados como entidades associadas a um ou mais significados e que, quando disponibilizadas ao usuário de forma correta, podem se transformar em informação ou conhecimento. No contexto da comunicação de dados, os dados podem ser: (i) analógicos, quando a fonte desse dado é analógica por natureza, como é o caso da fala humana, ou (ii) digitais, quando a fonte desse dado é um equipamento digital, como o caso de um texto produzido em um computador. Sinais, por sua vez, são a codificação elétrica ou eletrônica de dados e podem ser caracterizadas como: (i) ondas senoidais contínuas quando geradas por um modulador, provocando a chamada transmissão analógica, ou (ii) ondas quadradas quando geradas por um codificador, provocando a chamada transmissão digital.

Do ponto de vista da performance da sinalização (analógica ou digital), mantendo constantes os demais fatores que influenciam na taxa de transmissão, os seguintes pontos são verdadeiros:

- ⇒ Um aumento na taxa de dados aumenta a taxa de erros dos bits (ocorre um aumento da probabilidade de recepção errada de um bit)
- ⇒ Um aumento na relação S/N (Sinal/Ruído) diminui a taxa de erros na recepção dos bits
- ⇒ Um aumento na largura de banda permite um aumento na taxa de dados

As sub-seções que se seguem abordam as possíveis combinações de dado e sinal.

#### 1.1 Sinais digitais transmitindo dados digitais

Um sinal digital é uma sequência de pulsos de voltagens descontínuas e discretas e são chamados de elementos de sinal. Codificar bits num sinal digital significa associar determinados comportamentos do sinal digital ao bit 1 (um) e outros comportamentos ao bit 0 (zero). Para que o esquema de codificação proposto funcione corretamente, a tarefa envolvida na interpretação do sinal pelo receptor pode ser sumarizada da seguinte maneira:

- (i) O receptor deve conhecer o tempo de cada bit, ou seja, deve saber com alguma precisão quando um bit começa e termina. Esse é o referencial mínimo para permitir a realização das amostragens (leitura das voltagens) do sinal no momento certo pelo receptor<sup>1</sup>. Nessa questão está implícita a necessidade de um relógio de referência entre as partes comunicantes. Existem basicamente três alternativas de sincronização de relógio: (a) Relógio único paralelo fazendo chegar pulsos de sincronização em cada uma das partes ao mesmo tempo; (b) Cada uma das partes contendo um relógio de alta precisão previamente acertado; (c) O receptor possuindo um Circuito Recuperador de Relógio (RLL) que pode ser sensibilizado pela alternância de voltagens do sinal de dados que chega. Dessas três opções, a última é uma das mais utilizadas devido à sua praticidade.
- (ii) O receptor deve conhecer a relação entre as voltagens e a representação dos bits zero e um. Aqui existem duas possibilidades mais genéricas: (a) cada bit é representado pela aplicação de uma determinada voltagem durante o tempo de validade do bit, e (b) cada bit é representado pela alternância de voltagens do sinal, ou seja, durante o tempo de validade do bit, ocorre uma mudança de voltagens e a maneira como essa mudança ocorre representa os bits zero e um.

Existem diversas formas de codificação de bits, mas no contexto deste documento, apenas 04 (quatro) tipos básicos serão mostrados, de acordo com a tabela abaixo:

Método de codificação	Representação dos bits
ON-OFF	⇒ bit 1 corresponde à emissão de um pulso ⇒ bit 0 corresponde à não emissão de pulso
NRZ-L	⇒ bit 1 emissão de voltagem positiva no intervalo do bit ⇒ bit 0 emissão de voltagem negativa no intervalo do bit
Manchester (usada no IEEE802.3)	⇒ bit 0 = transição de alto para baixo no meio do intervalo do bit ⇒ bit 1 = transição de baixo para alto no meio do intervalo do bit
Manchester Diferencial (usada no padrão IEEE802.5)	Existe sempre uma transição no meio do intervalo, mas para diferenciação dos bits: ⇒ bit 0 = transição no começo do intervalo do bit ⇒ bit 1 = não há transição no começo do intervalo do bit

<sup>1</sup> Em geral essas amostragens são feitas próximas do meio do intervalo de validade do bit.

Dentre os métodos apresentados, pode-se dizer que os sinais ON-OFF (unipolar) e NRZ-L (polar)<sup>2</sup> são os mais simples de transmitir, pois existe uma correspondência de um para um entre os bits e os elementos de sinal. Em métodos que usam esse tipo de convenção é possível perceber as seguintes características:

- (a) Capacidade de aproveitar melhor a largura de banda do canal
- (b) Facilidade de implementação do método de codificação e a respectiva percepção dos bits pelo receptor
- (c) Dificuldade de representar seqüências contínuas de zeros ou uns. Em alguns casos são necessárias técnicas de embaralhamento dos bits (acordadas entre transmissor e receptor) para evitar rajadas de bits iguais consecutivos.
- (d) Incapacidade de transportar transições regulares (oscilações) suficientes para excitar o circuito recuperador de relógio (RLL) na recepção.
- (e) Grande sujeição ao fenômeno da atenuação e, é, portanto, mais apropriado para distâncias curtas.

Os métodos Manchester e Manchester Diferencial estão na categoria de esquemas que representam os bits pela alternância de voltagens. Como características desses métodos de codificação pode-se citar os seguintes:

- (a) As alternâncias de voltagens inerentes a métodos dessa natureza produzem muito mais flexibilidade para sensibilização e recuperação de relógio (RLL) no receptor;
- (b) Alternâncias para representação dos bits possuem menos sujeição à atenuação do que os métodos de representação com voltagens fixas.
- (c) A largura de banda necessária para o transporte dos bits é muito maior do que no esquema de voltagens fixas (vide sub-seção a seguir).

### 1.1.1 Relação entre o esquema de codificação e a taxa de transmissão

Para a sinalização digital existem diversas técnicas de codificação e essa escolha é feita visando otimizar o uso do meio de transmissão, em função de parâmetros tais como a largura de banda disponível ou a taxa de erros permitida na transmissão. Portanto, a performance da transmissão depende do esquema de codificação utilizado entre as partes, mas esse não é o único elemento que deve ser analisado para a tomada de decisão. Quanto ao mapeamento dos bits de dados nos elementos de sinal, algumas definições podem ser dadas:

- A taxa de sinalização de dados (ou apenas taxa de dados de um sinal) é a taxa em bits por segundo, que os dados são transmitidos. A duração ou comprimento de um bit é a quantidade de tempo que o transmissor demora em transmitir um bit. Por exemplo, para uma taxa de dados  $R$ , a duração de um bit é  $1/R$ .
- A taxa de modulação é a taxa na qual o nível de sinal é alterado. Isto depende da natureza da codificação digital. A taxa de modulação é medida em bauds, que diz a quantidade de elementos de sinal representados num segundo.

Para entender a influência do esquema de codificação na taxa de transmissão, considere um canal com uma taxa de 10 Mbaud, ou seja, um canal capaz de emitir 10 milhões de elementos de sinal (ou voltagens) por segundo. Se o esquema de codificação for o NRZ-L, onde um elemento de sinal corresponde a um bit, a taxa de transmissão é igual a 10Mbps. Por outro lado, se o esquema de codificação for o Manchester ou Manchester Diferencial, nos quais um bit é representado pela alternância dos bits, a taxa de transmissão desse canal será igual a 5 Mbps. Em outras palavras, um canal com largura de banda de 5MHz consegue transmitir 10Mbps numa codificação NRZ-L, e apenas 5 Mbps se a codificação adotada for Manchester ou Manchester Diferencial.

É importante observar que a largura de banda não é o único fator que determina a escolha por um método de codificação. Cada tipo de codificação possui características que o projetista deve levar em consideração para escolha. Dentre eles é possível citar alguns, tais como:

- Custo e complexidade dos circuitos - Embora os investimentos em tecnologia digital estejam contribuindo para a redução na queda dos preços dos circuitos, a questão de custo não pode ser desconsiderada. Em outras palavras, quanto mais alta a taxa de sinalização (bauds) necessária para alcançar uma determinada taxa de dados, maior o custo.
- Forma de detecção de erros - Existem várias técnicas de detecção de erros, e uma delas é relacionada à percepção de voltagens consecutivas que não representam sinal algum. Por exemplo, numa codificação Manchester, onde o bit 0 é obtido pela alternância de voltagens Alto-Baixo e o bit 1 pelas voltagens Baixo-Alto, ao perceber seqüências como Baixo-Baixo ou Alto-Alto, o receptor pode entendê-las como erros de transmissão.
- Imunidade a ruídos - Certos códigos exibem uma performance superior na presença de ruídos do que outros. Tem a ver com uma série de fatores, incluindo a quantidade de componentes que esse sinal possui.

### 1.2 Transmissão por sinais analógicos

<sup>2</sup> Se os elementos de sinal possuem todos o mesmo sinal algébrico, isto é, positivo ou negativo, então esse sinal é unipolar, como é o caso da sinalização ON-OFF. Na sinalização polar um estado lógico representa por um nível de voltagem positivo e o outro estado é representado por um nível de voltagem negativo, como é o caso da sinalização NRZ-L.



A sinalização analógica é uma técnica popular com utilização garantida nos sistemas de transmissão, devido a diversos fatores, dentre eles, alguns bastante consideráveis:

- (i) A fibra ótica é um meio de transmissão que ocupa uma posição de destaque na transmissão de dados pelas suas qualidades (alta capacidade de transmissão, baixa taxa de erros, etc.) é um meio que consegue propagar apenas sinais de luz que são por natureza, analógicos;
- (ii) Meios não guiados como é o caso do espaço aéreo trabalham melhor a sinalização analógica – Nesse caso a modulação é necessária para alterar o espectro de freqüências da fonte para freqüências mais altas a fim de que esse sinal seja perceptível por antenas convencionais<sup>3, 4</sup>;
- (iii) O sistema de telefonia é ainda a estrutura de comunicação mais capilarizada que se conhece, permitindo, de uma forma ou de outra, a transmissão de dados entre usuários, em qualquer lugar no planeta.

A base para a sinalização analógica é um sinal contínuo conhecido como sinal portador (ou onda portadora) cuja freqüência deve ser constante ( $f_c$ ) e compatível com o meio de transmissão que está sendo usado. O dado de entrada (analógico ou digital) nesse contexto é denominado sinal modulante ou sinal em banda básica (não alterado). O resultado de modular a portadora de acordo com esse sinal de entrada gera um sinal modulado com largura de banda limitada e espectro compatível com a freqüência  $f_c$  da portadora.

O equipamento utilizado para injetar dados no sistema de transmissão analógico é denominado MODEM (Modulador/Demodulador), que possui diferentes níveis de complexidade, dependendo da sua aplicabilidade. No caso da fala humana para a rede de telefonia, por exemplo, os modems são usados para alterar o espectro de freqüências do sinal de voz de modo a acomodar diversos canais num mesmo meio de transmissão, cada um ocupando uma determinada faixa de freqüências, técnica essa conhecida como FDM. A figura a seguir representa, do lado esquerdo a freqüência normal de um sinal de voz, e do lado direito, três sinais de voz (de igual freqüência) ajustados num espectro diferente do original (300 a 3400 Hz, que é a faixa de freqüências correspondente à fala humana).



Nesse caso a voz humana é transportada nos contornos da portadora e esse método é conhecido como sinalização em banda básica, pois não há alterações no sinal original. As mesmas técnicas básicas são usadas para modems que produzem sinais a altas freqüências (exemplo: microondas).

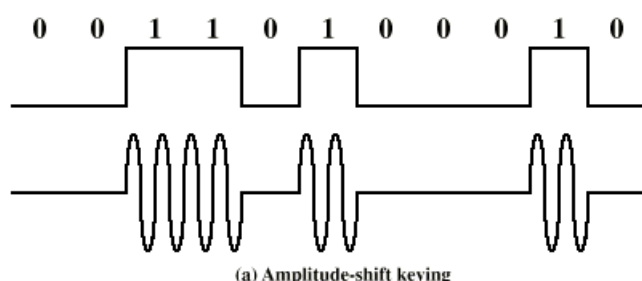
Para dados digitais, os modems promovem uma atividade extra que é a conversão de dados digitais na forma analógica pela alteração das características de amplitude, freqüência ou fase do sinal portador. Nesse ponto vale lembrar que a rede de telefonia foi projetada para receber, comutar e transmitir sinais analógicos com freqüências em torno de 4KHz, suficiente para acomodar a faixa de freqüência da voz humana e acomodar uma sinalização máxima em torno de 2400 baud. Embora muitas mudanças estejam ocorrendo no chamado *loop* local (contato da operadora com o assinante)<sup>5</sup>, as fontes digitais precisam ter seus dados adaptados a essas limitações para viabilizar a transmissão de dados.

As técnicas de codificação de bits em sinais analógicos podem ser classificadas da seguinte maneira:

- ASK - Amplitude-shift Keying
- FSK - Frequency-shift Keying
- PSK - Phase-shift Keying

Em todos os casos, o sinal resultante ocupa uma largura de banda centrada na freqüência da portadora.

Na técnica ASK os dois valores binários são representados por dois diferentes níveis de amplitude da portadora. Comumente uma das amplitudes é zero, isto é, um dígito binário é representado pela presença da portadora, numa amplitude constante, e a outra pela ausência da portadora. O sinal resultante é, portanto  $s(t) = A \cos(2\pi f_1 t)$  ► binário 1, e  $s(t) = 0$  ► binário zero, onde  $A \cos(2\pi f_1 t)$  é a portadora do sinal. ASK é suscetível a súbitos ganhos e é uma técnica de modulação ineficiente. Em linhas de voz, é tipicamente usada para transmissão na faixa de 1200bps, para uma taxa de sinalização de 2400 baud.



<sup>3</sup> As chamadas antenas convencionais se referem às antenas comerciais e que só conseguem interceptar sinais com freqüências altas. Para conseguir captar sinais analógicos centrados em baixas freqüências, uma antena deve ter um diâmetro imenso, na ordem de quilômetros.

<sup>4</sup> Um outro motivo evidente para a modulação é o fato de que o espectro de freqüências de meios não guiados como o ar está totalmente mapeado para necessidades específicas e não é recomendado utilizá-lo sem se atentar para este fato.

<sup>5</sup> Atualmente existem alternativas de comunicação tais como as tecnologias xDSL, Cable Modem e WLL (Wireless Local Loop), dentre outras em estudo.

No FSK, os dois valores binários são representados por diferentes frequências próximas da frequência da portadora. O sinal resultante é:

$$\Rightarrow s(t) = A \cos(2\pi f_1 t), \text{ para o binário 1, e}$$

$$\Rightarrow s(t) = A \cos(2\pi f_2 t) \text{ para o binário zero,}$$

onde  $f_1$  e  $f_2$  são variações da frequência da portadora  $f_c$  por quantidades iguais, porém opostas.

A figura ao lado apresenta um exemplo do uso de FSK para operação *full-duplex* sobre uma linha de voz. Nessa figura, deve-se perceber que o domínio de 300 a 3400Hz foi dividido em duas frequências, com faixas de 1170Hz cada uma, para permitir a transmissão *full-duplex* (transmissão nos dois sentidos ao mesmo tempo). Em uma direção as frequências usadas para representar o zero e o um são centradas em 1170Hz com uma folga de 100Hz em cada lado. O efeito de alternar entre essas duas frequências produz um sinal cujo espectro é indicado com a área cinza. Similarmente, para a outra direção, o modem usa frequências centradas em 2125Hz com folgas de 100Hz em cada lado. Este sinal está apresentado na parte cinza da direita da figura.

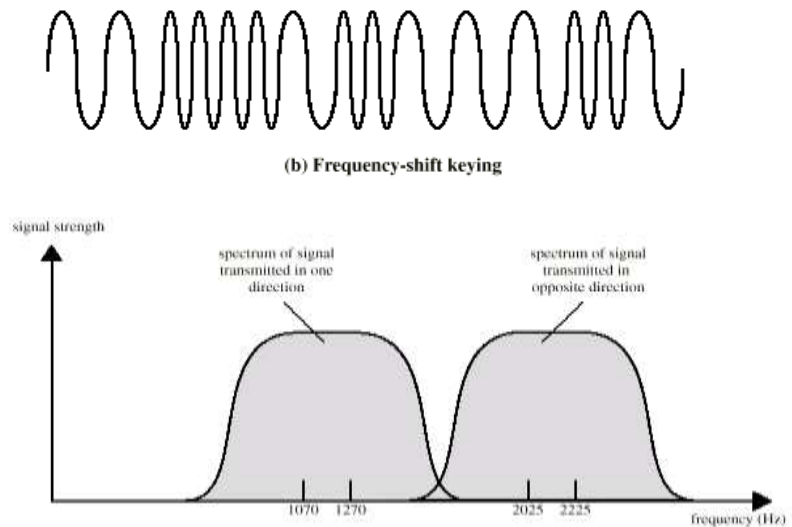
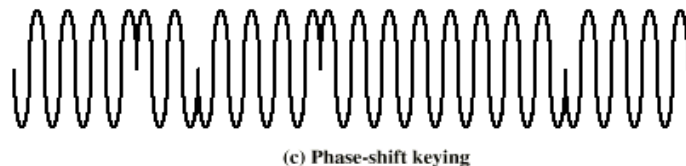


Figure 5.8 Full-Duplex FSK Transmission on a Voice-Grade Line

Algumas características do FSK:

- É menos suscetível a ruídos do que o ASK.
- Em linhas de voz, a sinalização em 2400 bauds consegue taxas similares ao do ASK que é em torno de 1200bps.
- É comumente usado para transmissão de rádio a altas frequências (3 a 30MHz).
- Usado para separar frequências altas em alguns tipos de redes locais que usam cabo coaxial.

No PSK, a fase da portadora é alterada para poder representar os dados. Significa dizer que o sinal sofre um ou mais atrasos (medido em graus) em relação ao valor de referência da portadora. A figura ao lado é um exemplo de um sistema de duas fases representando a sequência de bits 00110100010. Nesse caso um



binário zero é representado por enviar uma rajada de sinal na mesma fase da rajada de sinal anterior e o binário um é representado por enviar uma rajada de sinal de fase oposta à rajada anterior. Isto é conhecido como PSK diferencial, uma vez que a mudança na fase é feita com relação à fase anterior. O sinal resultante possui a seguinte fórmula:  $s(t) = A \cos(2\pi f_c t + \pi)$  para o binário 1 (um), e  $s(t) = A \cos(2\pi f_c t)$  para o binário 0 (zero), com a fase relativa medida para o intervalo anterior do bit.

Uma utilização mais eficiente pode ser alcançada se cada elemento de sinal representar mais do que um bit. Por exemplo, ao invés de alterar a fase em  $180^\circ$  como foi apresentado, pode-se utilizar alterações de fase por múltiplos de  $90^\circ$ , e nesse caso, cada elemento de sinal representa dois bits ao invés de um, conforme pode ser exemplificado abaixo:

$$\Rightarrow s(t) = A \cos(2\pi f_c t + 45^\circ) = \text{bit 11} \quad \Rightarrow s(t) = A \cos(2\pi f_c t + 225^\circ) = \text{bit 00}$$

$$\Rightarrow s(t) = A \cos(2\pi f_c t + 135^\circ) = \text{bit 10} \quad \Rightarrow s(t) = A \cos(2\pi f_c t + 315^\circ) = \text{bit 01}$$

Portanto, mesmo com limitação do sinal em 2400 baud, é possível alcançar uma taxa de dados maior, uma vez que vários tipos de atrasos podem ser gerados, sem alterar as características básicas de amplitude e frequência do sinal (que são mais afetadas pela limitação da largura de banda). Então, quanto mais atrasos um sinal contiver, mais bits por elemento de sinal será possível representar. Por exemplo, é possível transmitir três bits por vez usando um esquema de oito diferentes fases e quatro bits por elemento de sinal se forem usados 16 fases diferentes.

Na comunicação de dados por modem, existe uma terminologia que diz que uma constelação de um modem é definida como um conjunto de pontos, onde cada ponto é formado por diferentes fases, amplitudes e frequências ou uma combinação dessas técnicas. Portanto, quanto maior for a constelação de pontos de um modem, maior a quantidade de bits por ponto e maior a taxa de transmissão.

Um padrão de codificação comum nos modems é conhecido como V.32 e é uma constelação de pontos formada por doze fases, mas quatro delas possuem duas amplitudes, totalizando 16 pontos e quatro bits por ponto para uma taxa de modulação de 2400 baud. Essa técnica, conhecida como QAM (*Quadrature Amplitude Modulation* ou modulação por quadratura de amplitude), consegue gerar uma taxa de 9600 bps.

Neste exemplo, fica muito clara a diferença entre a taxa de dados  $R$  (em bps) e a taxa de modulação  $D$  (em bauds) de um sinal. Assumindo que este esquema está sendo empregado numa entrada NRZ-L (entrada digital), a taxa de modulação pode ser vista como  $R/4$ , na medida em que cada mudança de elemento de sinal comunica 4 bits. Esta é a razão pela qual os modems modernos usam técnicas avançadas para representar mais bits por elemento de sinal num canal de voz, uma vez que a largura de banda (4KHz) é limitada. A tabela abaixo apresenta os padrões adotados pelos principais fabricantes de modem:

Padrão	Pontos na constelação	Bits por ponto	Taxa (em 2400 baud)
V.32	16	4	9600bps
V.32 bis	64	6	14400 bps
V.34	4096	12	28800 bps
V.90	16.777.216	24	57600 bps

É interessante perceber que, para taxas de dados maiores, o número de pontos na constelação cresce exponencialmente. No caso do padrão V.90, por exemplo, o número de pontos é tão grande que existem limitações de distância e número de amplificadores entre a origem e o destino para que essa taxa seja alcançada<sup>6</sup>. Métodos complementares como os padrões MNP-5 e V.42bis conseguem uma taxa de dados maior pelo uso de algoritmos de compressão antes de envio dos dados.

### 1.3 Sinais digitais transmitindo dados analógicos

Nessa seção é examinado o processo de converter dados analógicos em sinais digitais, processo esse conhecido como digitalização. Uma vez que um dado analógico é convertido em dado digital, uma série de coisas podem acontecer, sendo que as três coisas mais importantes são as que estão listadas abaixo:

O dado digital pode ser transmitido usando NRZ-L. Neste caso

Apresenta dado de voz que são digitalizados e então convertidos para um sinal analógico ASK. Isto permite transmissão digital no sentido do que já foi visto anteriormente. O dado de voz, por que tem sido digitalizado, pode ser tratado como dado digital,

O dispositivo usado para converter dado analógico na forma digital para transmissão e, subsequentemente, é capaz de recuperar dados analógicos originais a partir de um sinal digital é conhecido como CODEC (coder/decoder). Existem duas técnicas mais conhecidas e utilizadas em codecs. A primeira é a DM (Delta Modulation) que .... A segunda é a técnica PCM (Pulse Code Modulation) que ...

A técnica PCM é baseada no teorema da amostragem que diz o seguinte:

*"Se um sinal  $f(t)$  é amostrado a intervalos regulares de tempo e a taxa é maior do que duas vezes a frequência de sinal mais significativa, então as amostras contêm toda a informação do sinal original. A função  $f(t)$  pode ser reconstruída dessas amostragens pelo uso de filtros passa-baixa."*

Portanto, se o dado de voz é limitado a frequências abaixo de 4000Hz, um procedimento conservador para a inteligibilidade, 8000 amostras por segundo serão suficientes para caracterizar completamente este sinal de voz, mesmo que essas sejam amostras analógicas.

Em outras palavras, um sinal pode ter largura de banda  $B$ , serão necessárias amostras a uma taxa de  $2B$ , uma vez a cada  $1/2B$  segundos. Essas amostras são representadas como barras estreitas cuja amplitude é proporcional ao valor do sinal original. Este processo é conhecido como PAM ou *Pulse Amplitude Modulation*. Para produzir dados PCM, as amostras PAM são quantizadas, isto é, a amplitude de cada pulso PAM aproximado por um inteiro de  $n$ -bits.

Uma boa reprodução de voz via PCM pode ser alcançada com 128 níveis de quantização, ou codificação em 7 bits. Um sinal de voz, conservadoramente, ocupa uma largura de banda de 4KHz. Portanto, de acordo com o teorema da amostragem, amostras devem ser feitas a uma taxa de 8000 vezes por segundo para recuperação do sinal. Isto implica em uma taxa de  $8000 \times 7 = 56\text{Kbps}$  para dados digitais codificados na forma PCM.

Vamos considerar o que isto significa do ponto de vista de utilização da largura de banda. Um sinal de voz analógico ocupa 4KHz. Um sinal digital de 56Kbps requer uma largura de banda de pelo menos 28KHz. Diferenças mais drásticas têm sido percebidas à medida em que se necessita de larguras de bandas mais altas. Por exemplo, um esquema PCM comum para televisão colorida usa codificação de 10bits, o qual trabalha a uma taxa de 92Mbps para um sinal de largura de banda de 4,6MHz. O interessante é que, indiferente a esses números, as técnicas digitais continuam a crescer em popularidade para transmissão de dados analógicos. As principais razões para essa escolha são:

- ⇒ Por que repetidores são usados ao invés de amplificadores, não existem ruídos adicionais.
- ⇒ A técnica de multiplexação por divisão de tempo (TDM) é usada para sinais digitais ao invés de usar FDM (multiplexação por divisão da frequência). Como a TDM não existe ruído de intermodulação, conforme será visto mais adiante.
- ⇒ A conversão de sinalização digital permite o uso mais eficiente das técnicas de comutação digital.

<sup>6</sup> Em geral a distância entre os modems comunicantes não deve ser mais do que dois ou três quilômetros, e a linha de transmissão deve ter boas características de imunidade a ruídos.

Além disso as técnicas estão sendo desenvolvidas para prover códigos mais eficientes. No caso da voz, um objetivo razoável aparece na vizinhança de 4Kbps. Com vídeo, a vantagem pode ser feita quadro a quadro, a maior parte dos elementos da figura não irão mudar. Técnicas de codificação interquadros permitiriam os requerimentos de vídeo serem reduzidos em até 15Mbps, e para as cenas que mudam vagarosamente, tais como as encontradas em vídeo conferência, serão necessários 64Kbps ou menos.

Como um ponto final, é importante mencionar que, em muitos casos, o uso do sistema de telecomunicações resultarão em processamento tanto analógico-para-digital quanto digital-para-analógico. A maior parte das terminações locais das redes de telecomunicação são analógicas e a rede usa uma mistura de técnicas digital e analógica. Como um resultado, dados digitais no terminal do usuário podem ser convertidos para analógico por um modem, subsequentemente digitalizado por um codec, e talvez sofra repetidas conversões antes de alcançar o seu destino.

Por este motivo, os recursos de telecomunicações manuseiam sinais analógicos que representam dados de voz e digitais. As características das formas de onda são muito diferentes. ...

## 2. Transmissão Analógica x Transmissão Digital

A transmissão de dados por sua vez, está relacionada aos recursos utilizados para garantir a entrega dos dados no destino. Enquanto que na transmissão analógica a estrutura de comunicação, incluindo amplificadores, não se preocupa com a natureza ou o tipo de informação que está sendo transferido, na transmissão digital toda a infra-estrutura de transporte do sinal está intimamente relacionada com o tipo dos dados (analógico ou digital) que estão sendo transportados.

Enquanto que na transmissão analógica é o meio de transmitir sinais analógicos sem se preocupar com o seu conteúdo. Esses sinais podem representar dados analógicos ou digitais. Nesse caso, o sinal analógico se tornará mais fraco (atenuado) com a distância, e a maneira

Transmissão digital em contrapartida, está relacionada ao conteúdo do sinal. Um sinal digital pode ser transmitido apenas a uma distância limitada antes que a atenuação comprometa a integridade dos dados. Para grandes distâncias, repetidores devem ser usados. Um repetidor recebe o sinal digital, recupera o padrão de 0s e 1s e retransmite o novo sinal.

A mesma técnica pode ser usada com um sinal analógico se é assumido que este sinal carrega dado digital. O sistema de transmissão tem repetidores ao invés de amplificadores. O repetidor recupera o dado digital do sinal analógico e gera um novo sinal analógico mais limpo. Portanto, o ruído não é acumulativo.

A questão que surge é por que a preferência pela transmissão digital. A indústria de telecomunicações tem investido muito tempo e dinheiro para a conversão gradual do sistema de transmissão de analógico para digital, devido a diversos fatores. Abaixo estão citados alguns.

- ⇒ Tecnologia digital - O advento das tecnologias LSI e VLSI tem causado uma contínua redução no custo e no tamanho dos circuitos digitais. Equipamentos analógicos não estão sofrendo uma progressão similar.
- ⇒ Integridade dos dados - Como o uso de repetidores ao invés de amplificadores os efeitos dos ruídos e outras imperfeições do sinal são não cumulativas. Por isso, é possível transmitir dados sobre grandes distâncias e sobre linhas de qualidade mais baixas, podendo, ainda por cima, preservar a integridade dos dados.
- ⇒ Utilização da capacidade - Tem se tornado bastante econômico construir linhas de transmissão de alta largura de banda, incluindo canais de satélite e conexões envolvendo fibra ótica. Um alto grau de multiplexação é necessário para efetivamente utilizar tal capacidade, e é mais fácil e barato alcançar isto por meio de técnicas digitais (divisão do tempo) do que com técnicas analógicas (divisão de frequências).
- ⇒ Segurança e Privacidade - Técnicas de encriptação pode ser prontamente aplicadas à dados digitais e também para dados analógicos, desde que sejam digitalizados.
- ⇒ Integração - Por tratar tanto os dados digitais quanto os dados analógicos na forma digital, todos os sinais tem a mesma forma e podem ser tratados de forma similar. Portanto, economia de escala e conveniências podem ser alcançadas por integrar voz, vídeo e dados digitais num mesmo meio.

## MULTIPLEXAÇÃO

No processo de transmissão de dados o ideal é que as partes comunicantes consigam fazer uso de toda a capacidade de transmissão que o meio possa disponibilizar. Nesse sentido, toda a teoria de controle de fluxo, formação de quadros e controle de erros existe considerando as limitações do meio físico e as suas capacidades de transmissão. No caso de uma comunicação ponto-a-ponto, por exemplo, o desejável é que o transmissor envie o máximo número de quadros possível para o receptor de tal maneira que a capacidade de transmissão do meio seja bem aproveitada.

No entanto, é provado que, em muitos casos, mesmo que as partes comunicantes atuem no máximo de suas capacidades, o meio de transmissão fica ainda sub-utilizado. Existem situações clássicas onde, por mais que o meio de transmissão seja veloz, as partes comunicantes possuem uma limitação no uso desses recursos. É o caso, por exemplo, dos terminais e microcomputadores que podem funcionar razoavelmente bem numa transmissão remota a taxas girando na faixa entre 9,6Kbps e 64kpbs. Por outro lado, é sabido que o custo do meio de transmissão cresce proporcionalmente com a capacidade de transmissão desse meio. Uma forma de redução desse custo é compartilhar ou multiplexar esse meio por vários usuários para que possa haver redução de custos, como mostra a fig. Abaixo:



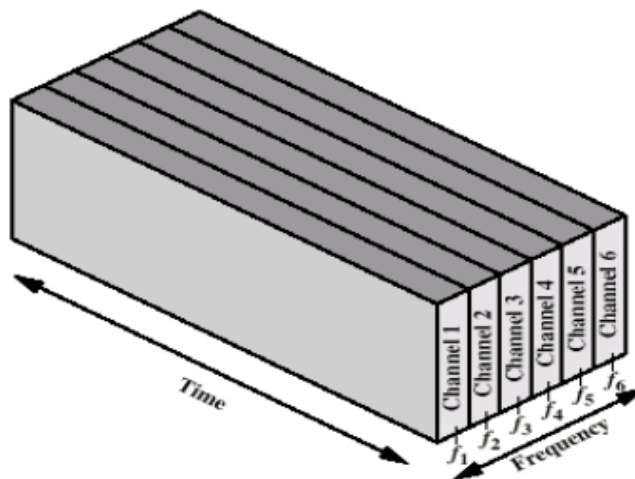
Dentro dessa visão existem três tipos de técnicas de multiplexação. A primeira, multiplexação por divisão de frequência (FDM), é a mais utilizada e é bastante familiar a quem costuma ver televisão ou ouvir rádio. A segunda técnica é um caso particular de multiplexação por divisão de tempo (TDM), conhecido como TDM síncrono. Essa técnica é normalmente utilizada para multiplexar sinais de voz digitalizados em conjunto com textos digitais. A terceira técnica visa aumentar a eficiência do TDM síncrono por adicionar complexidade ao multiplexador e é conhecido mais comumente como TDM Estatístico. As seções que se seguem, abordam cada um desses métodos.

### 1. FDM – Multiplexação por Divisão de Frequência

Essa técnica é possível quando a largura de banda útil do meio de transmissão excede a largura de banda requerida por cada um dos sinais a serem transmitidos. Dessa forma, vários sinais podem ser transportados simultaneamente se cada sinal for modulado em uma frequência diferente de tal maneira que não haja sobreposição entre elas. Em outras palavras, pode-se dizer que a largura de banda do meio é dividida em sub-canais, cada um deles alocados previamente para cada uma das fontes durante todo o tempo. Na verdade apenas dois grandes problemas devem ser tratados para que o FDM possa ser utilizado adequadamente:

- *Crosstalk* – Ocorre quando componentes de sinais adjacentes se sobrepõem de forma significativa, como quando ocorre linha cruzada numa ligação telefônica.
- Ruído de Intermodulação – Quando as energias de frequências diferentes se somam gerando sinais numa frequência que é a soma ou a diferença das frequências originais. Esse efeito normalmente ocorre nos amplificadores que podem atuar de forma não linear nas componentes de um canal alterando-as de forma que acabem invadindo outros canais.

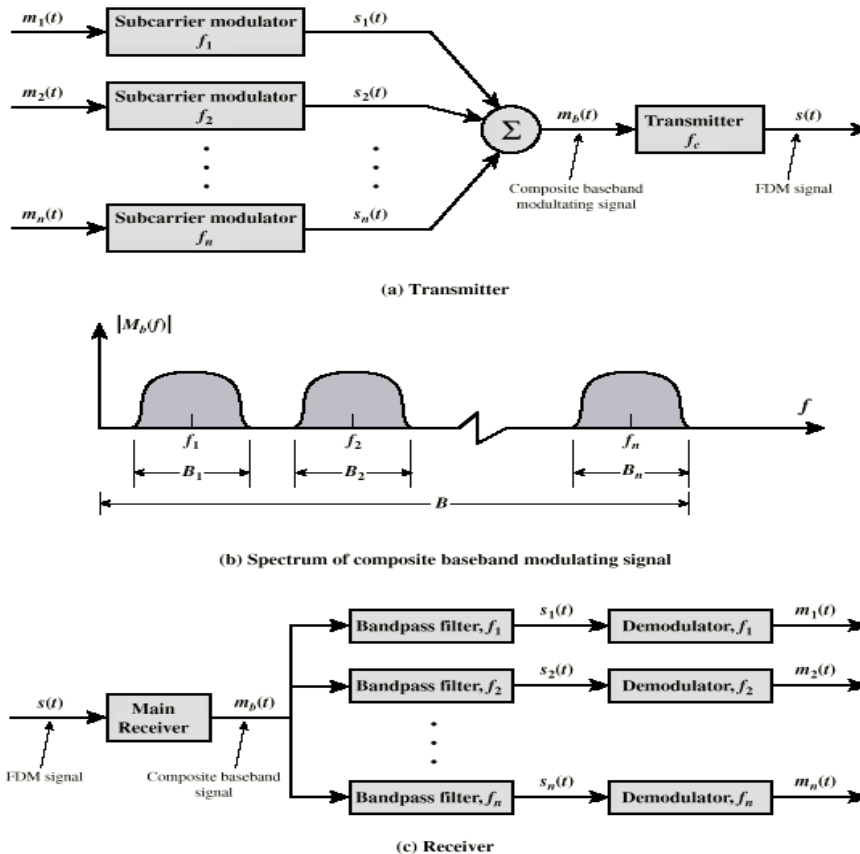
A figura a seguir apresenta um caso onde seis fontes de sinal são colocadas juntas num multiplexador, cada qual modulado em uma frequência diferente ( $f_1, f_2, \dots, f_6$ ).



Cada sinal modulado requer uma certa largura de banda centrada em torno de uma frequência portadora, normalmente referenciada como um canal. Para prevenir interferências, os canais são separados por bandas de guarda ("guard bands"), as quais são porções não usadas do espectro.

O sinal composto transmitido através do meio é analógico. No entanto os sinais que dão origem ao sinal composto podem ser analógicos ou digitais. No caso de sinal de entrada digital os dados devem passar por modems para serem convertidos para a forma analógica e serem modulados na posição correta do espectro. No caso de sinais de entrada analógicos, esses devem apenas ser adaptados para a frequência correta a ser utilizada no espectro do meio de transmissão.

Como exemplo de FDM é possível citar os canais de televisão captados por antenas ou por cabos de TV coaxiais. Nesse caso cada canal possui uma largura de banda na ordem de 6MHz. Uma descrição genérica de um sistema FDM é apresentada na figura a seguir.



Um número de sinais (analógicos e/ou digitais)  $[m_i(t), i = 1, N]$  deve ser multiplexados no mesmo meio de transmissão. Cada sinal  $m_i(t)$  é modulado numa portadora  $f_{sci}$ . Pelo fato de que existem muitas portadoras a serem usadas, essas são conhecidas individualmente como sub-portadoras. Qualquer tipo de modulação pode ser usado em qualquer das  $f_{sci}$ 's. Os sinais modulados resultantes são então somados para produzir um sinal composto  $m_c(t)$ . A parte b da figura abaixo reflete o resultado. O espectro do sinal  $m_i(t)$  é alterado para estar centrado em  $f_{sci}$ . Para este esquema de trabalho,  $f_{sci}$  deve ser escolhido de modo que as larguras de banda de vários sinais não se sobreponham, caso contrário, será impossível recuperar os sinais originais no destino.

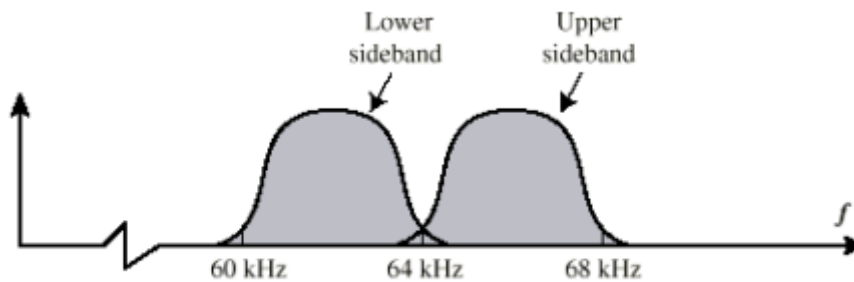
Esse sinal composto pode ser alterado como um todo para uma outra frequência mais alta, dando origem aos sistemas hierárquicos analógicos, os quais serão vistos mais adiante. É importante ressaltar que esse passo adicional de modulação não precisa usar a mesma técnica de modulação já utilizada nos passos anteriores.

Do que foi exposto é possível deduzir que a largura de banda do sinal composto é maior do que a soma das larguras de banda de todos os sinais que o formaram. No lado receptor, esse sinal composto é filtrado por  $N$  filtros passa-banda, cada um deles centrado nas mesmas frequências das sub-portadoras usadas na origem. Dessa forma o sinal é dividido em  $N$  partes e cada uma dessas partes é demodulada para recuperar o sinal original.

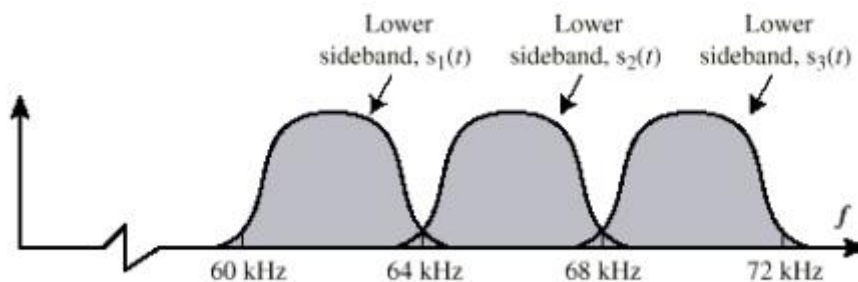
Para ilustrar um pouco mais, é possível apresentar mais um exemplo onde três sinais de voz são enviados simultaneamente sobre um meio de transmissão. Antes disso é preciso lembrar que a largura de banda de um sinal de voz é gerada em torno de 4KHz, com um espectro efetivo entre 300 até 3400Hz, conforme ilustra a figura abaixo:

(a) Spectrum of  $m_1(t)$ , positive  $f$ 

A figura abaixo mostra um sinal de voz transportado para a faixa de frequências de 64KHz disponível num meio de transmissão qualquer. Percebe-se que o sinal modulado possui uma largura de banda de 8KHz, estendendo-se de 60 até 68KHz. Para tornar o uso da largura de banda mais eficiente, é possível eleger a transmissão apenas da faixa menor ("lower sideband") do sinal de voz.

(b) Spectrum of  $s_1(t)$  for  $f_1 = 64$  kHz

Finalmente, se forem utilizadas apenas a faixa menor de cada um dos três sinais de voz, as faixas de 64 até 72KHz podem ser usadas, conforme demonstra a figura seguir:



(c) Spectrum of composite signal using subcarriers at 64 kHz, 68 kHz, and 72 kHz

## 1.1 Hierarquia de Transmissão Analógica

Existe um sistema de portadora de longa distância usado em todo o mundo para transmitir sinais de voz sobre enlaces de alta velocidade (ex.: cabos coaxiais, sistemas de micro-ondas etc.). Esse esquema usa a tecnologia FDM mas existem divergências nas técnicas adotadas por diferentes países.

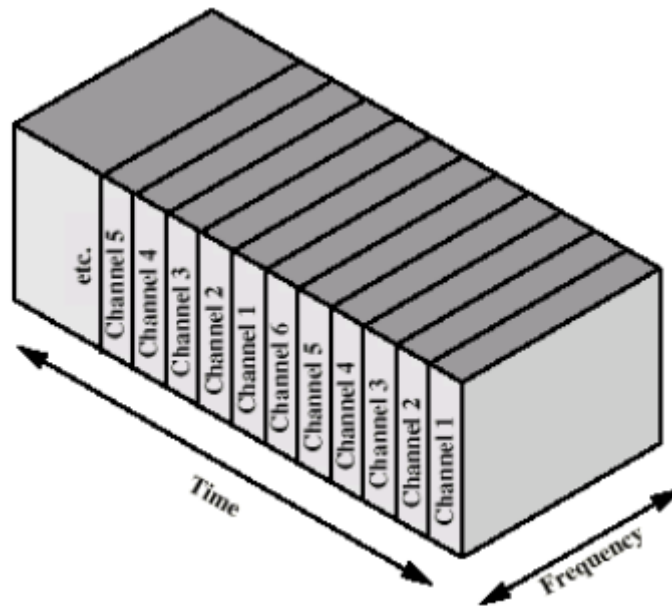
É importante perceber que o sinal de voz original pode ser modulado diversas vezes. Por exemplo, um sinal de dados pode ser codificado usando QPSK para formar um sinal de voz analógico. Este sinal é então modulado para a portadora de 76KHz para formar um componente de um grupo. Este sinal de grupo é então modulado para a portadora na ordem de 516KHz para formar um componente de um supergrupo. Em cada um desses estágios, o sinal original pode ser distorcido em função dos equipamentos envolvidos (moduladores, amplificadores e multiplexadores). Esses equipamentos podem interpretar o sinal de forma não linear ou mesmo introduzir ruídos.

## 2. TDM – Multiplexação por Divisão de Tempo

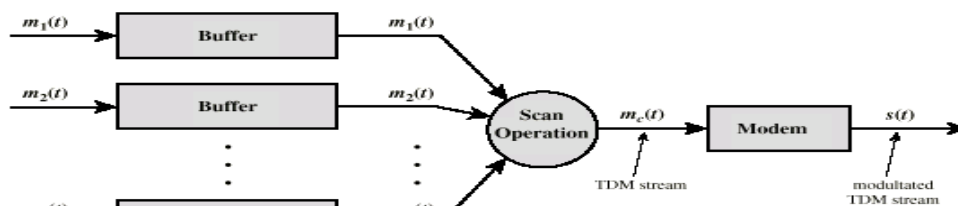
É um esquema utilizado quando a taxa de transmissão de dados do meio excede a taxa de dados dos sinais digitais a serem transmitidos. Nesse caso, múltiplos sinais digitais são alternados no tempo. É importante não misturar aqui o conceito de taxa de transmissão do meio com a largura de banda do meio para não confundir com o FDM.

Múltiplos sinais digitais (ou sinais analógicos carregando sinais digitais) podem ser transportados em um único caminho pela interpolação (alternância) de porções de sinais de cada uma das fontes. Em outras palavras, a largura de banda do meio é completamente cedida a cada uma das fontes por um período pré-definido no tempo. O tempo de interpolação pode ser o tempo de validade de um bit, de um caractere ou de um bloco de caracteres. Por exemplo, o multiplexador da figura abaixo tem seis entradas. Se cada uma delas for de 9.6Kbps então o canal multiplexado deve ser no mínimo de 57.6 Kbps (mais uma quantidade relacionada ao overhead) para poder acomodar todas as seis entradas.

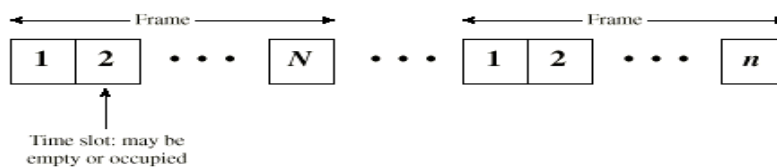




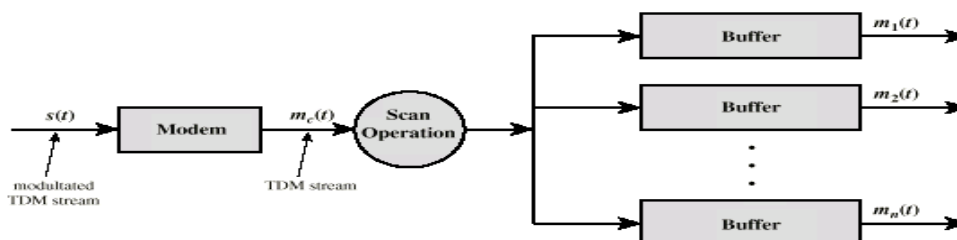
Uma descrição genérica de um TDM síncrono é provida na figura a seguir onde os sinais  $[m_i(t), i = 1, N]$  devem ser multiplexados no mesmo meio de transmissão. Os sinais de entrada transportam dados digitais e são geralmente sinais digitais. Os dados de cada fonte são armazenados em pequenos buffers, normalmente do tamanho de um bit ou um caracter. Esses buffers são pesquisados sequencialmente para formar um sinal digital composto denominado  $m_c(t)$ . A operação de pesquisa deve ser suficientemente rápida de modo que o buffer é esvaziado antes que os dados sejam sobrepostos. O sinal digital pode ser transmitido diretamente ou passado através de um modem de modo que um sinal analógico seja transmitido. Independente da forma, a transmissão é tipicamente síncrona. Esses dados são organizados em quadros que contém um ciclo de "slot times", ou seja, são formados por um tempo pré-definido e dedicado a cada uma das fontes de entrada. É importante lembrar que uma mesma fonte pode ter mais de um "slot time" por quadro. A soma de todos os slots de tempo de uma determinada fonte em todos os quadros compõe o que se chama de canal.



(a) Transmitter



(b) TDM Frames



(c) Receiver

No receptor, o dado interpolado é demultiplexado e roteado para o buffer de destino correto, ou seja, para cada entrada  $m_i(t)$ , existe uma saída correspondente que irá receber o dado de entrada na mesma taxa na qual ele foi gerado.

O TDM é considerado síncrono não porque a transmissão síncrona é usada, mas por que os slots de tempo são pré-definidos e fixos para cada fonte. Os slots de tempo de cada fonte são transmitidos de qualquer maneira, mesmo que a fonte não tenha dados a transmitir (idêntico ao que ocorre no FDM). Em ambos os casos, a capacidade do meio físico é desperdiçada em nome de uma simplicidade de implementação dos multiplexadores. Mesmo com uma atribuição de tempo fixa é possível para o TDM manusear fontes de diferentes taxas



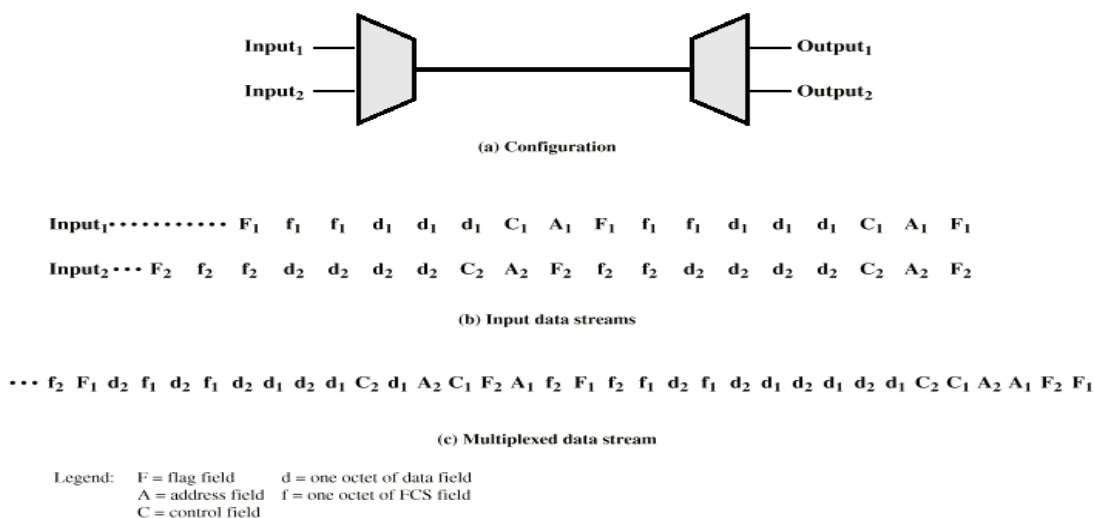
de dados. Por exemplo, a um dispositivo de entrada mais lento pode ser atribuído um slot de tempo por ciclo, enquanto que um dispositivo de entrada mais rápido pode ganhar mais de um slot de tempo.

Em termos de aplicabilidade, a técnica de interpolação de caracteres é normalmente usada com fontes assíncronas, enquanto que a interpolação de bits pode ser feita com fontes síncronas e assíncronas. No primeiro caso, cada slot de tempo contém um caractere e os bits de start/stop são eliminados antes de serem colocados no buffer. No destino esses bits são reinseridos novamente e essa estratégia resulta num aumento de eficiência da transmissão como um todo. A seção a seguir apresenta os mecanismos de controle necessários para viabilizar enlaces TDM.

## 2.1 Controle de Enlace TDM

Uma das características do TDM é não possuir caracteres de controle nem no início (cabeçalho ou “header”) nem no fim dos quadros (“trailer”). Mecanismos de controle de fluxo e de erro não tem sentido para o canal TDM por que o canal é dividido entre muitos pares origem-destino (canais) e um problema de um par ou canal individual não pode interferir na comunicação dos demais. Significa dizer que o controle do TDM é o próprio tempo alocado a cada fonte e nada mais.

No entanto essa constatação não impede que pares individuais implementem entre si algum mecanismo de controle de erros e de fluxo, usando protocolos específicos como o HDLC por exemplo que resolve problemas dessa ordem. A figura abaixo exemplifica dois pares comunicantes, cada um usando um protocolo com caracteres de controle entre si, mas compartilhando uma linha TDM.



Perceba que a sequência apresentada na parte c da figura é um tanto confuso e sem nexo quando vista integralmente. No entanto, considerando que o TDM consegue rotear esses caracteres para os destinos específicos, a sequência colocada na origem será remontada no destino e, dessa forma, será possível manter controles entre os pares. Como última observação sobre esse exemplo, seria interessante pensar que em cada lado existe uma combinação de um multiplexador e um demultiplexador para formar uma comunicação full-duplex.

Por outro lado, apesar de não precisar de controle de fluxo e nem de erros, a TDM precisa lançar mão de duas técnicas de enquadramento e sincronização, conforme mostra as sub-seções que se seguem.

### a) Enquadramento (“Framing”)

Conforme já foi visto, o TDM não precisa de protocolos de controle para funcionar adequadamente. Apesar disso, uma de suas necessidades é tornar visível a formação dos seus quadros. Como não existem caracteres de sincronização ou delimitadores de início e fim de quadros, é preciso lançar mão de alguns mecanismos para prover a sincronização de modo que origem e destino mantenham o mesmo compasso.

Um dos mecanismos de enquadramento mais conhecidos é a técnica de adição de dígito de enquadramento. Neste esquema um bit de controle é adicionado para cada quadro TDM, gerando um “canal de controle”. Um padrão identificável de bits combinado entre origem e destino deve ser inserido em cada quadro. O exemplo mais comum é a sequência formada pela alternância de zeros e uns (101010...) entre quadros consecutivos. O receptor, sabendo da posição desse bit extra e da lei de formação para inserção desses bits, procura por um bit de framing com valor zero se no quadro anterior esse valor foi igual a um e vice-versa.

Se o padrão de bits não é encontrado, caberá ao receptor, analisar posições sucessivas dos bits que chegam até encontrar a sequência desejada. Uma vez que a sincronização está estabelecida, o receptor continua a monitorar o bit do canal de framing. Se o padrão se desfaz, o receptor reinicia o processo de análise sucessiva novamente. Como exemplo típico, pode-se citar o padrão norte americano que multiplexa 24 canais de 8 bits cada e usa um bit extra para framing, totalizando um quadro de 193 bits.

### b) Inserção de Pulsos (“Pulse Stuffing”)

Talvez um dos mais difíceis problemas que o TDM deve tratar é a questão da sincronização de várias fontes de dados. Se cada fonte possuir um relógio diferente, qualquer variação entre esses relógios causará a perda de sincronismo. Dessa forma é importante pensar que o multiplexador possua um relógio próprio.

Uma outra questão é com relação à taxa de dados. O ideal no TDM é que cada fonte possua a mesma taxa de dados ou que, sendo diferentes, essas taxas sejam múltiplas entre si. Por exemplo, se uma fonte deseja transmitir à taxa de 9.6 Kbps o ideal é que as demais fontes transmitam em alguma taxa múltipla dessa (ex.: 4.8Kbps, 28.8Kbps, 57.6Kbps, etc.). Como essa nem sempre é uma realidade entre as fontes de dados, algum mecanismo deve ser utilizado para equalizar essas taxas.

Neste contexto, uma técnica conhecida como inserção de pulso (*"pulse stuffing"*) é bastante utilizada. Nessa técnica define-se a taxa de dados de saída do multiplexador, excluindo os bits de framing, com um valor maior que a soma das taxas instantâneas de entrada (vazão de entrada). Essa capacidade extra é usada para bits ou pulsos extras de modo que a taxa de dados fique compatível com um sinal de relógio gerado localmente. Os pulsos extras são inseridos em posições fixas no quadro pelo multiplexador de tal forma que o demultiplexador consiga removê-los da sequência.

Como exemplo, pode-se considerar uma situação de multiplexação de duas fontes, uma com taxa de 16Kbps e outra com taxa 6.5Kbps. Nesse caso, a fonte de 6.5 Kbps deve sofrer inserção de pulsos para chegar a 8Kbps e ser racionalmente compatível com a taxa de dados da outra fonte.

## 2.2 Hierarquia de Sistemas Digitais

Assim como no sistema analógico, no sistema digital existe uma hierarquia para transmissão de sinais de voz na forma digital sobre links de alta capacidade de transmissão tais como fibras óticas, cabos coaxiais e microondas. Parte da evolução dessas redes de telecomunicações usam as estruturas de transmissão TDM.

Nos EUA e no Japão existe uma hierarquia TDM cujo formato é o chamado DS-1 com quadros de 193 bits (24 sinais de voz mais um bit de framing). Considerando 8000 amostras por segundo, é possível chegar a uma taxa de 1.544 Mbps. Para cada cinco de seis quadros enviados usa-se a codificação PCM de 8 bits, enquanto que no sexto quadro usa-se a PCM com 7 bits. Nesse último caso, o oitavo bit serve para controle de sinalização contendo informações de controle da rede (ex.: estabelecimento e encerramento de chamada) e de roteamento do sinal de voz.

Esse mesmo formato DS-1 é usado para transmissão de dados digitais, só que preservando a mesma taxa final de 1.544Mbps para compatibilização com sinais de voz. Portanto, é possível gerar uma mistura de canais de dados e voz num mesmo enlace físico. No caso da transmissão de dados propriamente dita, 23 canais são providos para dados enquanto o 24º canal é reservado para controle. Esse byte extra é usado para permitir um re-enquadramento mais confiável e mais rápido no caso de um erro de framing. Dentro de cada canal, sete bits são usados para dados enquanto que o oitavo bit é usado para indicar se naquele canal os dados contidos são do usuário ou se são dados de controle. Dessa forma é possível alcançar a taxa de 56Kbps considerando 7 bits por quadro, 8000 vezes por segundo.

Para tratamento de taxas acima e abaixo da taxa DS-1 as seguintes regras são seguidas:

- Para taxas menores – Um canal pode transportar bits a uma taxa menor do que 56Kbps. Nesse caso um bit extra é usado para indicar que aquele canal carrega uma taxa menor e, portanto, sobram 6 bits gerando um total de 48Kbps. Pode-se concluir que essa capacidade pode multiplexar 5 canais de 9.6Kbps ou 10 canais de 4.8Kbps ou 20 canais de 2.4Kbps. Por exemplo, se o canal 2 é usado para prover 9.6Kbps, então 5 sub-canais de dados compartilham esse canal e o dado para cada sub-canal aparece como blocos de 6 bits no canal 2 a cada quinto quadro.
- Para taxas maiores – Acima da taxa 1.544Mbps, as altas velocidades são obtidas pela multiplexação de bits de vários canais DS-1. Por exemplo, o sistema de transmissão DS-2 combina quatro canais DS-1 (12 bits de cada fonte) totalizando uma sequência de 6.312Mbps ( $1.544 \times 4 = 6.176$  Mbps mais os bits extras para controle e *framing*).

### Exercícios:

- i) Dada uma única fonte com taxa de vazão (capacidade de colocar bits no meio) igual a 2.0 bps e taxa de duto (capacidade de transmissão do meio) = 2 bps, certifique-se de que, pelo fato das capacidades da fonte e do meio de transmissão serem compatíveis, não há qualquer tipo de desperdício. Preencha a tabela abaixo simulando a transferência de bits e verificando quantos bits serão entregues ao final de 3,0 segundos.

Tempo (s)	Origem	Destino
0.5		
1.0		
1.5		
2.0		
2.5		
3,0		

- ii) Considerando uma única fonte com taxa de vazão igual a 2.0 bps e a taxa de duto = 1 bps. ao final de 3,0 segundos quantos bits serão perdidos? Certifique-se de que nessa configuração não é possível compartilhar o meio de transmissão.

Tempo (s)	Origem	Destino
0.5		
1.0		

1.5		
2.0		
2.5		
3.0		

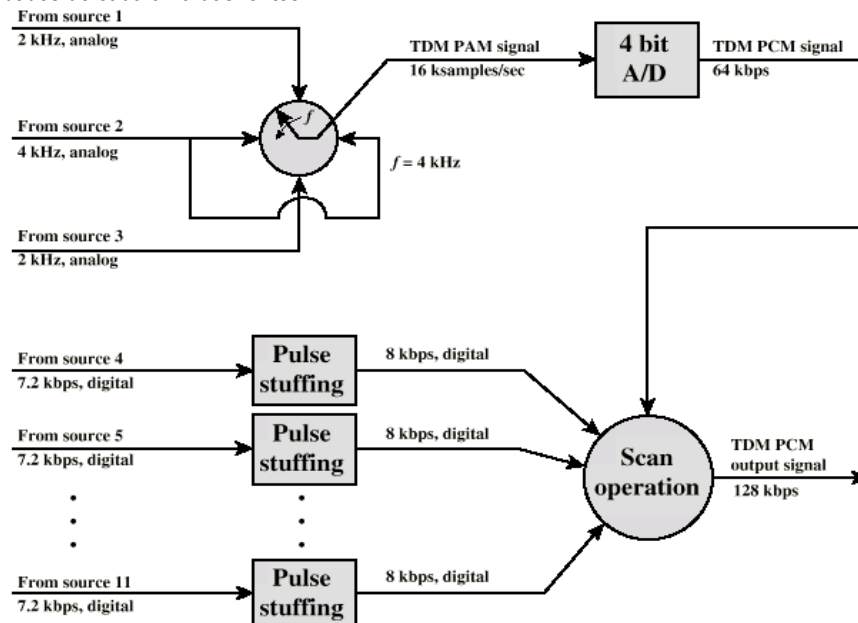
- iii) Se uma única fonte estiver com taxa de vazão igual a 1.0 bps e a taxa de duto for igual a 3 bps, ao final de 3,0 segundos quantos bits terão sido transmitidos. Quantos clientes de vazões iguais poderiam estar utilizando simultaneamente este meio para garantir um melhor aproveitamento?

Tempo (s)	Origem	Destino
0.5		
1.0		
1.5		
2.0		
2.5		
3.0		

- iv) Considere três fontes analógicas  $f_1$ ,  $f_2$  e  $f_3$ , com frequências iguais a 2KHz, 4KHz e 2KHz respectivamente. Considerando 4 bits por amostra, preencha a tabela abaixo com a quantidade de bits recuperados em cada fonte, de acordo com a amostragem apresentada nas linhas.

	$F_1 = 2\text{KHz}$	$F_2 = 3\text{KHz}$	$F_3 = 4\text{KHz}$	TOTAL DE BITS
8000 amostras/s				
6000 amostras/s				
4000 amostras/s				

- v) Dada a figura abaixo, na qual existem 11 fontes de dados, sendo as fontes  $f_1, f_2$  e  $f_3$  analógicas de 2KHz, 4KHz e 2KHz respectivamente, e as demais sendo digitais, síncronas e com taxas na ordem de 7.2Kbps. Considerando o teorema da amostragem para canais analógicos e um ciclo de 32 bits, faça um desenho esquemático de um quadro indicando quantos bits serão obtidos de cada uma das fontes.



### 2.3 SONET/SDH - Synchronous Optical Networks

Quando surgiram os enlaces com fibra ótica, cada companhia telefônica tinha o seu próprio sistema ótico TDM. Após a quebra da AT&T em 1984, as companhias de telefone locais tinham que se conectar com múltiplas portadoras de longa distância (WAN's), todas com diferentes sistemas óticos. Esse fato motivou o processo de padronização TDM para fibras óticas desencadeando o padrão SONET/SDH, sendo que o primeiro foi elaborado pela BellCore, enquanto que o SDH ("Synchronous Digital Hierarchy") ficou sob responsabilidade do CCITT. Esses dois padrões possuem diferenças mínimas. O resultado desse processo foi que praticamente todas as operadoras de telefonia de longa distância nos EUA usam SONET na camada física (uma interface de transmissão óptica).

Na verdade o raciocínio era de que, ao ser adotado como padrão pelos fabricantes, a tecnologia SONET poderia ser vendida a um custo baixo em cartões para PC's. Dessa forma, seria possível ao usuário se ligar diretamente ao coração da rede de telefonia sobre linhas especiais (tipo LPCD's).

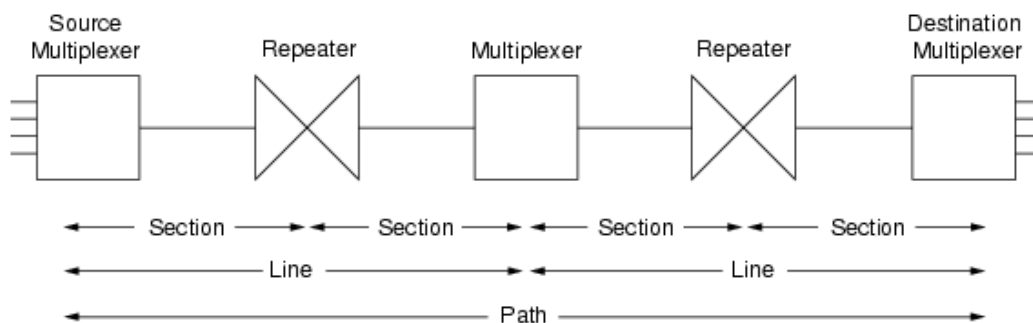
O SONET surgiu com objetivos específicos, os quais podem ser citados:

- Tornar possível a interconexão de diferentes operadoras de telefonia – Requer definir um padrão de inicialização comum que defina o comprimento de onda, a temporização a ser utilizada, estrutura do quadro a ser transmitido etc.
- Unificar os sistemas digitais Americano, Europeu e Japonês – Todos esses foram baseados em canais PCM de 64K, mas com incompatibilidades entre si em itens como formato de quadros, mecanismos de sincronização etc.
- Prover uma maneira para multiplexar vários diferentes tipos de canais digitais – Em outras palavras, continuar o processo de hierarquização de multiplexação na ordem de gigabits e acima. Quando o SONET foi criado, o T3 (canal de 44.736 Mbps) era a maior taxa de transmissão disponível (o T4 havia sido criado mas não estava sendo utilizado). Um outro compromisso era o de padronizar a multiplexação de canais de taxas baixas em um canal SONET.
- Prover suporte para OAM (Operação, Administração e Manutenção) – Os sistemas anteriores não faziam isso muito bem.

Do ponto de vista tecnológico, a primeira decisão foi fazer do SONET um sistema TDM tradicional, no qual a largura de banda completa da fibra seria atribuída para um canal contendo slots de tempo para os vários sub-canais. Portanto pode-se dizer que o SONET é um sistema síncrono controlado por um relógio mestre de alta precisão (os intervalos de envio e duração dos bits são extremamente precisos).

Obs<sub>1</sub>.: Quando a comutação de células foi proposta como base para o B-ISDN (a questão era permitir a chegada de células de forma irregular) essa foi chamada de ATM para contrastar com a operação do SONET.

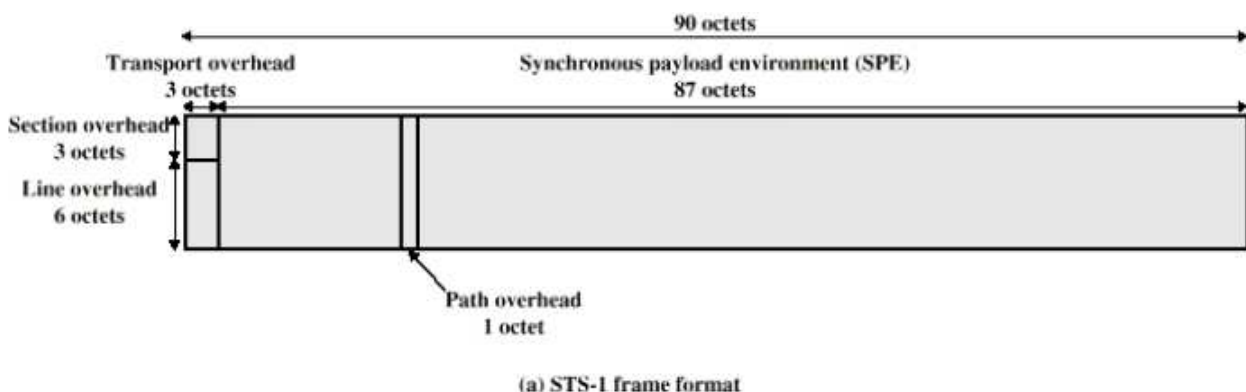
c. Componentes do SONET – Um sistema SONET consiste de switches, multiplexadores e repetidores, todos interligados por enlaces de fibra. Os enlaces entre esses dispositivos possuem denominações distintas, como mostra a figura a seguir:



Obs<sub>2</sub>.: A topologia pode ser uma malha, mas normalmente o que se vê é um anel duplo.

### 2.3.1 Formato do Quadro SONET

O SONET possui um quadro de 810 bytes emitidos a cada 125 microssegundos, visando coincidir com as 8000 amostras do PCM. Por ser um TDM, os quadros são emitidos com ou sem dados. Os quadros SONET são melhor vistos como uma matriz A de 90 colunas por 9 linhas, onde cada elemento  $A_{ij}$  é um octeto do quadro. Portanto a taxa mais básica do sonet, chamada STS-1 ("Synchronous Transport Signal – 1"), é de 51,84Mbps (9linhas x 90 colunas x 8 bytes x 8000 amostras/s). Todas as demais taxas são derivadas dessa.

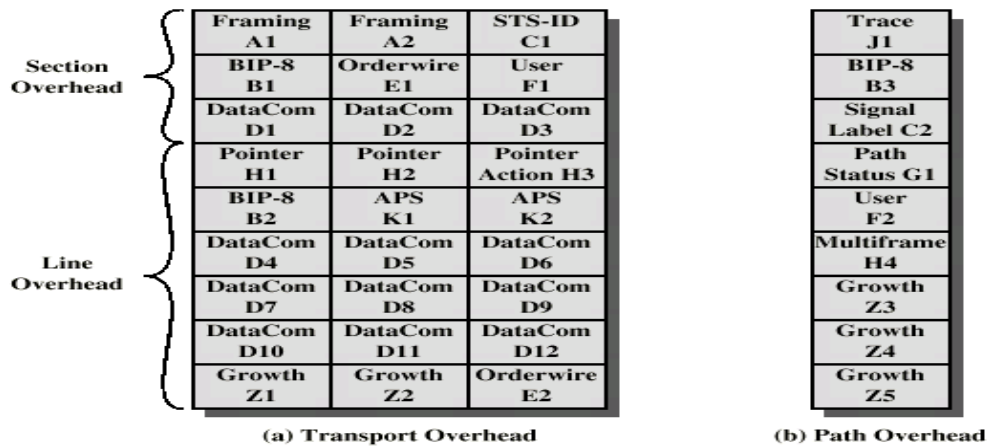


Apesar da técnica TDM tradicionalmente não possuir cabeçalho, o SONET possui 27 bytes de controle, sendo 9 bytes (3 colunas x 3 linhas) para o overhead de seção entre repetidor e multiplexador, e 18 bytes (3 colunas x 6 linhas) para o overhead de linha entre multiplexadores ponto-a-ponto. As próximas 87 colunas x 9 linhas formam o SPE ("Synchronous Payload Envelope"), que contém na primeira coluna o overhead do caminho entre multiplexadores fim-a-fim e o restante são dados do usuário. Dessa forma, pode-se caracterizar o seguinte:

- Dados de overhead = 3 colunas x 9 linhas x 8000/s = 1.728 Mbps
- SPE = 87colunas x 9 linhas x 8000/s = 50.112 Mbps, incluindo 86 colunas para dados e uma coluna para overhead de caminho

Uma das grandes características do SPE é que o mesmo pode começar em qualquer lugar no quadro. Isso porque existe um ponteiro na primeira linha do overhead de linha que diz onde o SPE se inicia, conforme demonstra a figura a seguir. Essa característica dá mais flexibilidade ao SONET que não precisa esperar até o início do próximo quadro para poder inserir dados que chegam do usuário (camada superior).

Além do que foi mencionado, o apontador de início de dados é bastante apropriado quando aquilo que vai ser transportado (ex.: uma célula ATM), possui tamanho diferente do payload SONET. Nesse caso o overhead de linha pode apontar para o início da primeira célula cheia para haver sincronização.



Pela figura acima, percebe-se que os overheads de seção, linha e caminho formam um conjunto de bytes que permitem realizar OAM. Uma vez que cada quadro ocorre 8000 vezes por segundo, esses overheads representam um canal PCM de controle.

### 2.3.2 Hierarquia de Multiplexação no SONET

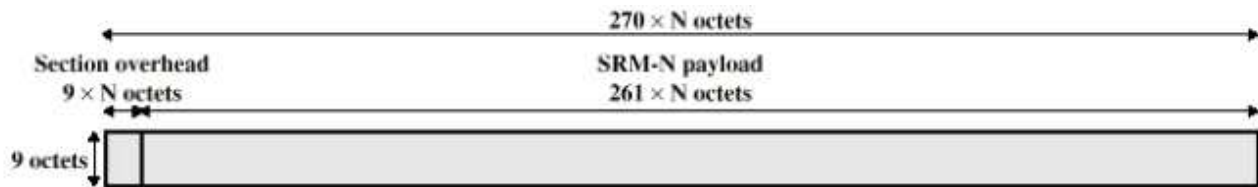
Na terminologia SONET, os canais que sofrem multiplexação são chamados tributários. Uma sequência típica de multiplexação SONET a altas velocidades é feita pela seguinte sequência de passos:

- i. Primeiro, os sistemas de baixa velocidade são convertidos para o STS-1 básico. Se houver necessidade, ocorre preenchimento de quadros vazios para completar os 51,84 Mbps desejados.
- ii. Grupos de três tributários STS-1 são multiplexados em uma saída formando um sinal denominado STS-3 de 155,52Mbps
- iii. Grupos de quatro STS-3 são multiplexados gerando uma saída denominada STS-12. Aqui ocorre um tratamento (mistura) dos bits para evitar longas sequências de zeros e um's que possam interferir no mecanismo de sincronização do relógio.

A multiplexação em todos os passos é feita byte a byte e os dados são convertidos de sinal elétrico para sinal ótico apenas no passo 3. Como um exemplo pode-se pensar na composição de três STS-1 a 51,84 Mbps. Nesse caso, o STS-3 de saída contém um byte do primeiro tributário, depois do segundo e assim por diante.

Em termos de nomenclatura, no SONET existe a convenção STS-n para sinalização elétrica e OC-n para sinalização equivalente em termos de taxa de transmissão, para a forma de representação ótica. Como exemplo, a portadora ótica correspondente à STS-n é a OC-n, ambas idênticas exceto no caso da STS-12 onde ocorre mistura dos bits, conforme já foi dito. Aqui vale mencionar algumas informações interessantes:

- ❑ Algumas taxas de transmissão tais como OC-9, OC-18 e OC-36 existem para cumprir o requisito de compatibilidade com o esquema utilizado no Japão.
- ❑ No SDH a nomenclatura apropriada para essas taxas de sinalização é STM-n. e possuem equivalência apenas a partir do OC-3 por que no SDH não existe taxa próxima de 51,84 Mbps.
- ❑ Quando uma linha não contém multiplexação embutida, ela recebe um "c" de concatenação. Como exemplo, o OC-3 indica uma portadora 155,52 Mbps composta de 3 canais OC-1's separados, enquanto que o OC-3c indica uma portadora de mesma taxa só que carregando dados de um único usuário. A quantidade de dados de um OC-3c é um pouco maior do que no OC-3 devido ao fato de que o overhead de caminho está incluído dentro do SPE apenas uma vez e no OC-3 está incluído três vezes. Em outras palavras, de 270 colunas, o OC-3c aproveita 260, enquanto que o OC-3 aproveita apenas 258. A figura abaixo apresenta um formato de quadro STM de tamanho N:



(b) STM-N frame format

### 3 STDM – Multiplexação por Divisão de Tempo Estatística (ou Assíncrona)

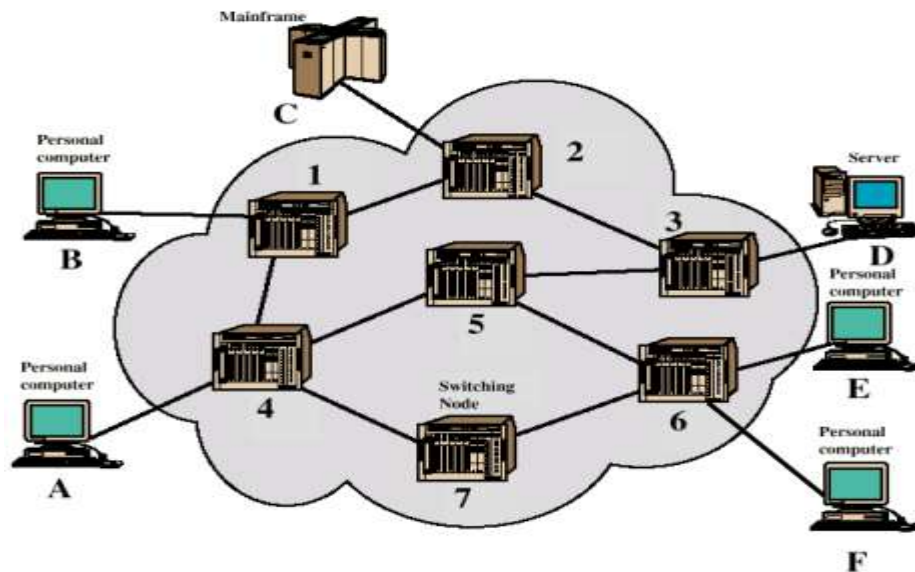
O TDM síncrono é interessante para situações onde as fontes de dados estão sempre aptas a enviar informações periodicamente e, se possível, em tempos cadenciados e fixos. Como essa nem sempre é uma realidade, existe uma alternativa para esse esquema, denominado TDM estatístico, cuja filosofia é disponibilizar o meio físico para quem estiver precisando, independente de uma sequência temporal.

Assim como no TDM normal, o estatístico possui várias linhas de entrada/saída de um lado e um enlace físico de alta velocidade do outro, sendo que as fontes (entradas) possuem buffers associados também. O multiplexador percorre os buffers de todas as fontes até que o quadro esteja preenchido e pronto para ser enviado. Um detalhe importante é que o tamanho do quadro pode ser menor do que o tamanho da soma de buffers que devem ser percorridos, por que se considera que nem todos querem transmitir ao mesmo tempo. Na saída o multiplexador recebe um quadro e distribui os slots de dados para os buffers apropriados.

Portanto, o multiplexador estatístico não envia slots vazios se existem dados a serem enviados. Portanto, durante a primeira passagem apenas os slots de A e B são enviados. Note que a significância posicional dos slots é perdida neste esquema. Portanto não sabido a quantidade de tempo que uma fonte de dados gastará em qualquer slot particular. Por que os dados chegam de forma imprevisível, informações de endereço são necessárias para garantir uma entrega apropriada. Como consequência, existem mais overhead por slot na medida em que cada slot carrega informações de endereço e dados do usuário.

## COMUTAÇÃO DE CIRCUITOS

Sistemas de comutação são formados por um conjunto de nós e enlaces de interconexão cuja estrutura pode ser utilizada para comunicação de longa distância (WAN) para a transmissão de voz, dados etc., conforme demonstra a figura a seguir:



Conceitualmente, os sistemas de comutação podem ser vistos como uma sub-rede de comunicação (representada pela nuvem da figura) com as seguintes características:

- ☐ Normalmente, a rede não é completamente conectada (não existem enlaces cruzados entre todos os nós);
- ☐ Alguns nós possuem apenas a função de repassar adiante as informações que chegam, como é o caso dos nós 5 e 7. Os demais nós, além de repassar as informações adiante, devem atender as conexões com as estações usuárias finais;
- ☐ Os troncos de interligação de nós normalmente usam algum esquema de multiplexação FDM ou TDM.

O sistema de transmissão de dados nas redes de comutação pode ser feito sob dois aspectos: a comutação de circuitos, muito utilizada para telefonia e a comutação de pacotes, mais voltada para a transferência de dados. Nas seções a seguir será apresentada uma abordagem de cada um desses métodos.

### 1. Conceitos Associados à Comutação de Circuitos

Desde a invenção do telefone, a comutação de circuitos tem sido a tecnologia dominante nas operadoras de telecomunicações para transmissão de voz, apesar de, atualmente, existirem várias outras tendências tecnológicas em função das novas necessidades dos usuários. Dentre as características dos sistemas de comutação de circuitos tradicionais, pode-se citar os seguintes:

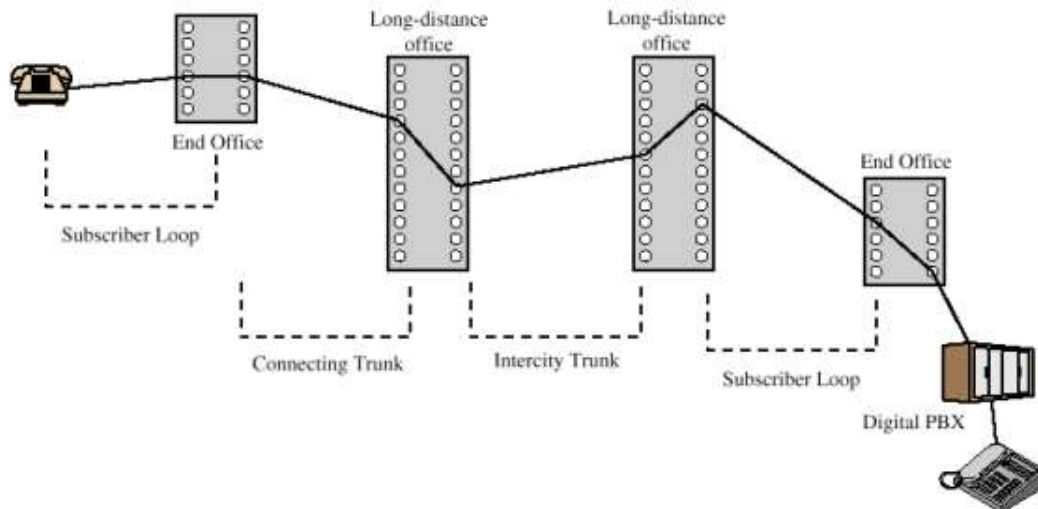
- Existe um caminho dedicado entre origem e destino (o canal é dedicado para quem o solicitou, não importando se vai haver utilização ou não);
- Uma vez conectado com o destinatário, a transferência de dados é transparente (não há congestionamento na rede);
- A comunicação ocorre em três fases bem distintas:
  - Estabelecimento de conexão – Definição do caminho a ser utilizado (quais nós/enlaces farão parte do caminho),
  - Transferência de dados – Diálogo propriamente dito entre as partes,
  - Encerramento de conexão - Liberação dos recursos alocados na fase de estabelecimento;
- Os nós de comutação devem possuir recursos suficientes para atender diversos usuários;
- Além de ser capazes de separar sinais de diferentes usuários, os comutadores devem ser capazes de realizar roteamento desses sinais para o destino correto.

Atualmente, a demanda por comunicação de dados entre computadores passou a ser uma realidade e, em função das características desse tipo de transmissão, a comutação de circuitos passou a ser questionada. Dentre as críticas aos sistemas de comutação de circuitos, é possível mencionar:

- Sistemas de comutação são considerados ineficientes, uma vez que a capacidade do canal não pode ser utilizada por outros usuários, mesmo que os originários da conexão não estejam transmitindo dados.
- O estabelecimento da conexão (identificação de comutadores e enlaces participantes) leva um tempo considerável que algumas aplicações podem não suportar.



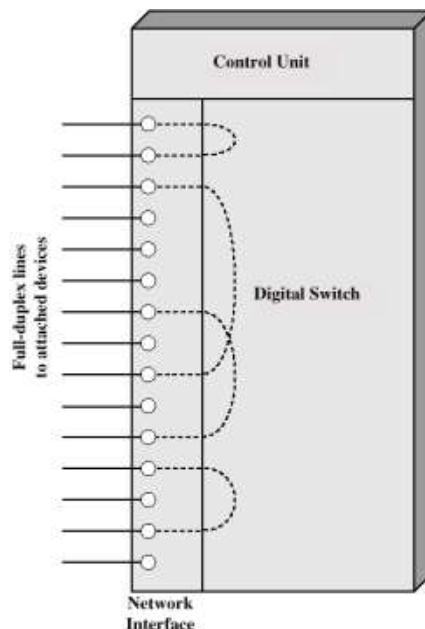
Dentro do jargão utilizado nas operadoras de telefonia, os seguintes elementos podem ser mencionados como participantes de um sistema de comutação de circuitos (vide figura a seguir):



- Assinante – usuário do sistema;
- Loop local – enlace que vai do assinante até o primeiro nó de comutação;
- Nós de comutação – comutadores da sub-rede (*"long distance office"*); possuem funcionalidades extras quando são preparadas para receber conexões de assinantes (*"end office"*).
- Troncos – Ramificações entre nós de comutação que utilizam alguma técnica de multiplexação (FDM ou TDM).

## 2. Tecnologia Utilizada em Comutadores

A tecnologia de comutação de circuitos pode ser vista pela análise individual de um único comutador ou pela análise de vários comutadores interagindo para viabilizar um caminho físico entre dois assinantes. Do ponto de vista didático, é melhor apresentar o sistema de comutação pelo estudo de um comutador individual. A comutação envolvendo vários equipamentos será vista em separado, na seção seguinte, por possuir características um pouco mais complexas. A figura abaixo representa um diagrama esquemático de um comutador digital composto por uma interface de rede, por uma unidade de controle e por uma unidade de comutação.



Nessa figura, existem vários assinantes vinculados ao comutador. As linhas pontilhadas representam caminhos físicos estabelecidos entre dispositivos que fizeram pedido de conexão. Aqui, os componentes apresentados caracterizam um comutador digital, cuja inteligência está inserida numa unidade de controle, similar a uma CPU de um computador qualquer. No entanto, vale lembrar que os primeiros comutadores possuíam unidade de controle não-digital, as entradas eram apropriadas para sinais analógicos e as conexões entre assinantes eram feitas por intermédio de relés. Os componentes de um *switch* podem ser descritos como:

- Interface de rede - representa as funções de hardware necessárias para conectar dispositivos digitais, tais como dispositivos de processamento de dados e telefones digitais para a rede (além de acesso para outros comutadores através de troncos TDM). Os telefones analógicos podem também ser ligados se a interface de rede contiver uma lógica de conversão de sinal analógico para digital.



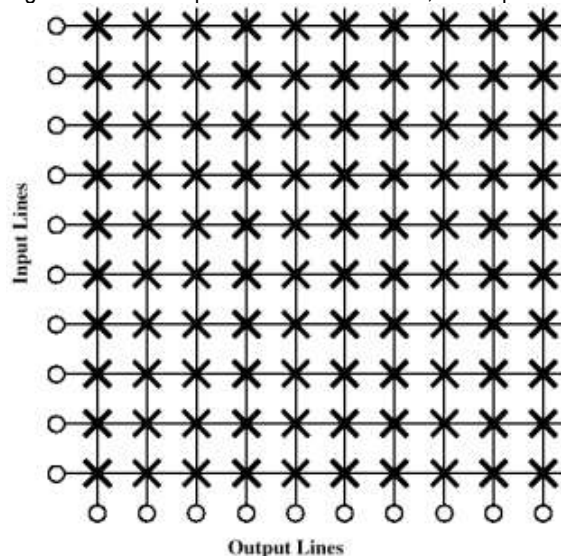
- Unidade de controle - executa três tarefas gerais: (i) estabelecimento de conexão, o qual é feito sob demanda. Para isso, a unidade de controle deve manusear e reconhecer a requisição do usuário, identificar se o destino está livre ou não e, finalmente, construir um caminho entre as partes; (ii) manter a conexão ativa. Em função dos switches digitais trabalharem com princípios de divisão de tempo (TDM), isto pode requerer manipulação dos elementos comutadores. Contudo, os bits de comunicação são transferidos de forma transparente; (iii) deve desfazer a conexão ou em resposta à uma requisição de uma das partes ou por conta própria.
- Unidade de comutação – é a maneira de formar o caminho entre origem e destino. Sob esse ponto de vista, existem dois tipos de análise que podem ser feitas para formação do caminho (veja que esse esquema não tem nada a ver com o fato do destinatário já estar utilizando o sistema):
  - Esquema bloqueante – Uma solicitação de conexão pode não ser atendida devido aos recursos necessários para formar o caminho estarem em uso por outro par de assinantes. Esse esquema foi considerado apropriado para sistemas de telefonia, uma vez que as chamadas telefônicas, em média, são de curta duração.
  - Esquema não bloqueante – Toda solicitação é atendida, haja vista que a estrutura do comutador permite formar um caminho entre origem e destino.

## 2.1 Tipos de Comutadores

Em linhas gerais, os comutadores são os de divisão de espaço e os de divisão de tempo. Os *switches* de divisão de espaço foram originalmente desenvolvidos para o ambiente analógico e têm sido bastante utilizados até hoje, mesmo com o advento da tecnologia digital. Isso é possível porque os princípios fundamentais de comutação são os mesmos, independente da comunicação ser analógica ou digital.

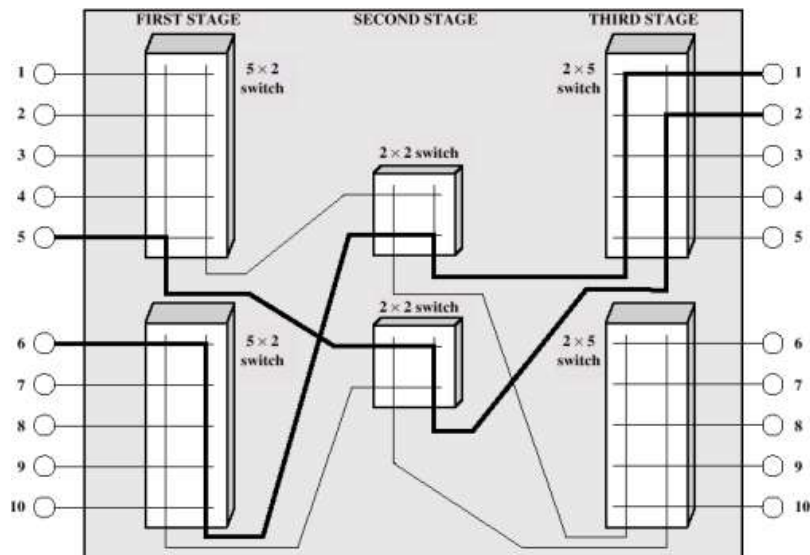
### 2.1.1 Comutadores de Divisão de Espaço

Como o próprio nome diz, um comutador de divisão de espaço explora os caminhos do sinal de cada uma das fontes. Esses caminhos cruzam de alguma forma com os caminhos de todas as outras fontes e o processo de comutação significa habilitar a ponte (*gate*) no cruzamento entre esses caminhos. Esse cruzamento é formado por um semicondutor ou um relé habilitado por algum mecanismo de controle dentro do comutador em função da demanda do assinante. Alguns exemplos podem ser dados para comutadores desse tipo. A figura abaixo ilustra um comutador de 10 entradas e 10 saídas *full-duplex*, onde a interconexão é possível pela habilitação do ponto de cruzamento entre os caminhos de origem e destino. Apesar da funcionalidade, esse tipo de *switch* possui algumas limitações:



- O número de pontos de cruzamento cresce exponencialmente com o número de linhas de entrada e saída;
- A perda de um ponto de cruzamento inviabiliza a conexão entre os dispositivos que usariam aquele cruzamento;
- Os pontos de cruzamento são usados de forma ineficiente. Mesmo quando todos os usuários estão conectados, apenas uma pequena fração dos cruzamentos estará sendo utilizada;
- É um comutador não bloqueante; basta a habilitação do ponto de cruzamento para viabilizar o caminho físico entre as partes.

Um outro exemplo de comutador de divisão de espaço é apresentado na figura a seguir:



Este tipo de arranjo é chamado de comutador de múltiplos estágios e possui vantagens em relação ao exemplo da matriz de cruzamentos apresentados anteriormente:

- O número de pontos de cruzamento é reduzido preservando as mesmas possibilidades de interligação das partes. Neste exemplo, o número de cruzamentos foi reduzido de 100 para 48;
- Existe mais de um caminho para conectar pontos finais, melhorando a confiabilidade;
- Exige um controle maior para a formação do caminho físico. Nesse caso é preciso determinar quem fará parte do caminho para depois habilitar os vários pontos de cruzamento;
- É considerado um comutador bloqueante, uma vez que determinadas conexões feitas podem atrapalhar a formação de outras. No exemplo, considerando as linhas mais escuras como caminhos já em uso, a linha de entrada 10 não poderá se conectar com as linhas 3, 4 ou 5, mesmo que essas estejam disponíveis.

OBS.: Um comutador de múltiplos estágios pode se tornar não bloqueante pelo incremento do número ou tamanho dos switches intermediários, mas isso, é claro, aumenta o custo.

### 2.1.2 Switches de Divisão de Tempo

A tecnologia de comutação de circuitos tem uma longa história, a maior parte dela coberta por uma fase onde a comutação analógica predominava. Com o advento das técnicas de digitalização da voz e das técnicas de multiplexação TDM, ambos, voz e dados, puderam ser transmitidos via sinais digitais. Essa característica provocou uma mudança enorme nos projetos de sistemas de comutação. Assim, os comutadores mais modernos contam com um controle inteligente de elementos de divisão de espaço e também de tempo.

Virtualmente, todos os comutadores modernos usam técnicas TDM para estabelecer e manter circuitos. A TDM permite agregar pequenas seqüências de dados de baixa velocidade em uma grande seqüência com taxas de velocidade maiores. Para entendimento desse conceito, é interessante analisar um tipo de comutador TDM simplificado que faz o compartilhamento do seu barramento principal.

Num primeiro momento, na fase de estabelecimento de conexão, a unidade de controle armazena as informações sobre qual porta de entrada deve estar ligada a qual porta de saída e qual a ordem de diálogo entre elas. Baseado nesse conceito, a unidade de controle permite o acesso ao barramento principal do comutador pela habilitação coordenada dos pares de portas comunicantes durante um certo limite de tempo. A varredura de todos os buffers das linhas de entrada durante um certo tempo permite a formação de um quadro. Cada slot de tempo deve ser suficiente para que o dado da linha de entrada seja transportado até a linha de saída, incluindo o atraso de propagação do sinal no barramento principal. Nesse contexto, se um comutador deseja atender 100 linhas de entrada simultaneamente, cada uma com taxa de 19.2Kbps, o barramento principal deve ter uma taxa superior a 1.92Mbps para funcionar adequadamente. Significa dizer que até 100 clientes o comutador é não bloqueante e a inserção de novos clientes torna-o bloqueante.

Os comutadores TDM simplificados podem acomodar linhas de entrada com variadas capacidades de transmissão, ou seja, os buffers de entrada (bit, byte ou bloco de bytes) não precisam ser de mesmo tamanho. Por exemplo, se uma linha de 9.6Kbps pegar um slot de tempo por quadro, uma linha de 19.2Kbps pode pegar dois slots de tempo por quadro. O importante aqui é entender que apenas linhas de mesma capacidade de transmissão podem se comunicar (a origem e o destino devem ter taxas iguais).

## 3. Roteamento em Redes de Comutação de Circuitos

Conforme já foi dito, as redes de comutação de circuito em geral são formadas por diversos comutadores interligados por troncos FDM ou TDM. Quando uma chamada é feita, se a linha de destino não estiver no mesmo comutador do chamador então será necessário lançar mão de técnicas de roteamento para identificar quais comutadores farão parte do caminho entre as partes. A decisão de roteamento entre comutadores passa por dois requisitos básicos:

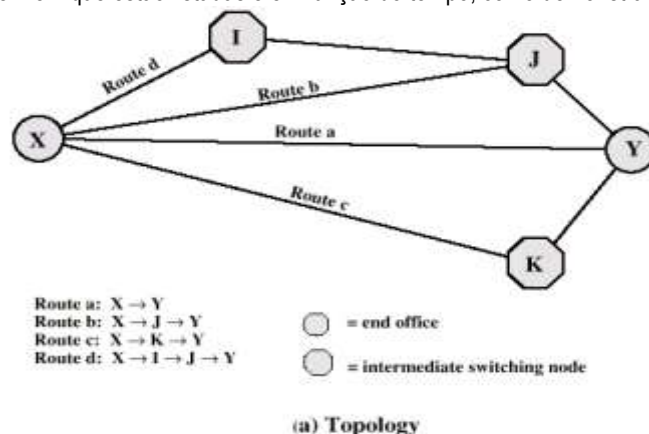
- **Eficiência** – É preciso minimizar a quantidade de equipamentos (nós e troncos) da rede para atender a demanda do assinante;
- **Resiliência** – É preciso garantir a demanda do assinante mesmo em momentos de picos de utilização ou em situações de falha de troncos ou nós de comutação intermediários.

Conforme pode ser percebido, esses requisitos são conflitantes entre si e a melhor estratégia de roteamento é aquela que consegue atender de forma satisfatória esses dois pontos.

Tradicionalmente, as operadoras de telefonia têm feito o roteamento de forma estática (utilizando sempre a mesma regra de roteamento) em cima de uma estrutura de sub-rede na qual os comutadores são organizados como uma árvore hierárquica. Dessa forma os dados do nó-folha de origem convergem para nós que compõem o tronco principal e depois são migrados para a ramificação que converge para o nó-folha de destino. Em alguns casos, onde a demanda é muito grande, a solução tem sido a inserção de troncos de atalho entre regiões (galhos) dessa árvore hierárquica para minimizar o surgimento de gargalos na sub-rede.

Atualmente, existe um outro enfoque para o roteamento em sistemas de comutação. A crescente demanda pelo uso da telefonia, tanto para voz como para dados, tem levado a uma decisão de se trabalhar com roteamento dinâmico ao invés de utilizar-se técnicas tradicionais de roteamento estático. Nesse contexto, a decisão de roteamento depende diretamente do tráfego atual da sub-rede de comutação. Nesse sentido, existem duas abordagens mais utilizadas.

O primeiro deles é o chamado Roteamento Alternado, onde cada comutador mantém uma seqüência de possíveis caminhos que podem ser adotados para se chegar a um determinado comutador de destino. Esses caminhos são previamente definidos e as rotas são listadas por ordem de preferência, de acordo com a configuração montada. Dessa forma, o comutador de origem seleciona uma rota apropriada em função da ordem em que estão listadas e em função do tempo, como demonstra o exemplo a seguir:



Time Period	First route	Second route	Third route	Fourth and final route
Morning	a	b	c	d
Afternoon	a	d	b	c
Evening	a	d	c	b
Weekend	a	c	b	d

(b) Routing table

Nessa figura, o roteador X possui quatro rotas para chegar a Y, e a escolha por uma delas depende, dentre outros fatores, do momento (manhã, tarde ou noite) em que será acionada.

Um segundo método, mais sofisticado, é denominado Roteamento Adaptativo. Ele pode ser utilizado para melhorar a escolha de rotas entre origem e destino. Nesse esquema, existe uma central de controle responsável por coletar, a cada tempo, informações sobre as condições de funcionamento de cada um dos comutadores da sub-rede. Com base nesses dados, o controlador determina qual é a melhor rota a ser seguida por cada comutador para chegar a um determinado destino. Significa dizer que toda definição de rota feita por um comutador deve ser autorizada pela central de controle. Como exemplo dessa abordagem, é possível citar o DTM (*Dynamic Traffic Management*), da Northern Telecom, cujo controlador central faz coleta de informações em intervalos de 10 segundos para descobrir a melhor alternativa de rotas para os comutadores.

#### 4. Controle de Sinalização

Em redes de comutação de circuitos, os sinais de controle são o mecanismo pelo qual a rede é gerenciada e pelo qual as conexões são estabelecidas, mantidas e encerradas. Tanto a gerência de uma chamada como a gerência da rede como um todo requer que as informações sejam trocadas entre o assinante e um comutador, entre comutadores e entre os comutadores e a central de gerência da rede.

O esquema de sinalização é razoavelmente complexo e sua aplicação se torna crítica, principalmente em redes grandes como os sistemas de telefonia atuais. Algumas funções providas por esses sinais podem ser relacionadas:

- Comunicação audível com o assinante, incluindo tom de linha, tom de discagem, sinal de ocupado etc.;
- Transmissão do número discado para os comutadores da rede que tentarão completar a conexão;
- Transmissão da informação entre comutadores indicando que uma chamada não pode ser completada ou que uma chamada terminou e que o caminho em uso pode ser liberado;
- Um sinal para fazer com que o telefone do usuário toque;
- Transmissão de informação usada para propósito de contabilização de uso (cobrança);
- Transmissão de informação sobre o estado de equipamentos ou troncos na rede. Essa informação pode ser usada para roteamento ou para manutenção;
- Transmissão de informação usada no diagnóstico e isolamento de falhas;
- Controle de equipamentos especiais tais como satélites de comunicação.

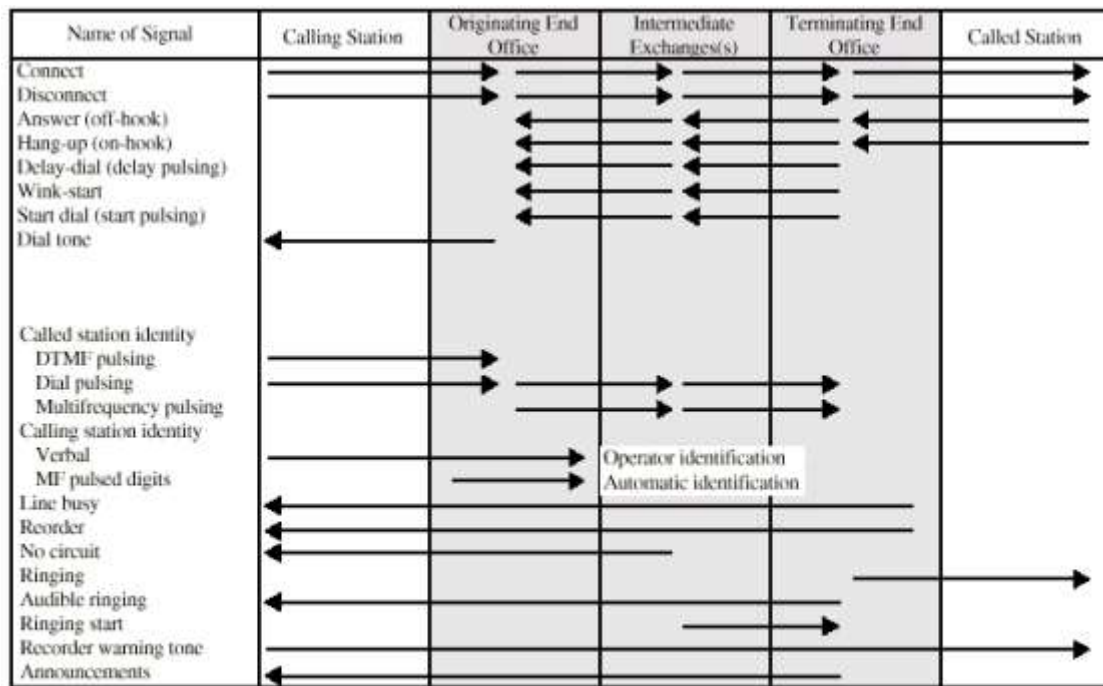
Como um exemplo do uso de sinais de controle, considere uma conexão típica de uma ligação entre dois usuários interligados por um mesmo comutador:

- i.* Em princípio, ambos os telefones não estão em uso (on-hook). A chamada começa quando um assinante tira o fone do gancho (off-hook); esta ação é automaticamente sinalizada para o comutador ao qual esse aparelho está ligado;
- ii.* O comutador responde com um sinal audível (dial tone), informando ao assinante que um número de destino pode ser discado;
- iii.* O chamador disca um número, o qual é comunicado como um endereço chamado para o comutador;
- iv.* Se o assinante chamado não estiver ocupado, o comutador alerta-o por enviar um sinal que provoca um sinal audível no telefone do chamado;
- v.* O retorno é provido ao assinante chamador pelo comutador dentro das seguintes condições:
  - a. Se o assinante chamado não está ocupado, o comutador retorna também um sinal audível para o chamador, durante o tempo em que o telefone do chamado está tocando,
  - b. Se o assinante chamado estiver ocupado, o comutador envia um outro tipo de sinal audível para o chamador,
  - c. Se a chamada não puder ser completada através do comutador, esse envia para o chamador uma mensagem audível solicitando nova tentativa de conexão;
- vi.* A parte que recebeu o chamado aceita a chamada por levantar o fone do gancho (off-hook), ação que é automaticamente identificada pelo comutador;
- vii.* O comutador encerra os sinais audíveis de toque nos dois lados e estabelece a conexão entre os dois assinantes;
- viii.* A conexão é liberada quando uma das partes encerra a conexão (normalmente a parte que iniciou a chamada).

Quando o assinante chamado está ligado a um comutador diferente do chamante, mas existe uma conexão direta entre eles, as seguintes funções de sinalização ocorrem entre os comutadores:

- i.* O comutador ligado ao chamador apodera-se de um tronco disponível entre comutadores e envia uma mensagem off-hook para o comutador de destino avisando que um endereço será informado (telefone do destinatário);
- ii.* O comutador de destino envia um sinal off-hook seguido por um sinal on-hook indicando estar pronto;
- iii.* O comutador iniciador envia os dígitos de endereço para o comutador de destino onde ocorrerá o processo de identificação da linha associada a este número.

Este exemplo ilustra algumas das funções executadas entre as partes usando sinais de controle. A figura abaixo apresenta alguns sinais utilizados numa rede de comutação de circuitos apresentando a origem e o destino dos vários tipos de sinais:



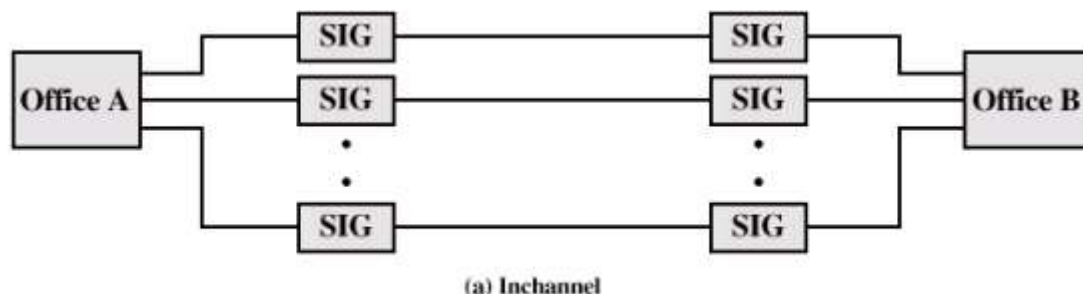
Note: A broken line indicates repetition of a signal at each office, whereas a solid line indicates direct transmittal through intermediate offices.

A sinalização utilizada em sistemas de comutação pode ser classificada da seguinte maneira:

- Sinais de supervisão – Funções de controle com caráter binário (verdadeiro ou falso), tais como requisição de um recurso, saber se um recurso está disponível ou não, alertas, respostas, etc.;
- Sinais de endereçamento – Identificam o assinante. Inicialmente um sinal de endereço é gerado pelo assinante chamador quando um número é discado. O endereço resultante pode ser propagado pela rede para suportar a função de roteamento e localizar e tocar no telefone do assinante chamado;
- Sinais de informação de chamada – Refere-se àqueles sinais audíveis que provêem informação para o assinante sobre o estado da chamada;
- Sinais de gerenciamento da rede – Enquanto que os três primeiros sinais são relacionados ao estabelecimento e encerramento de conexão, os sinais de gerenciamento são usados para manter a operação da rede como um todo. Tais sinais podem estar na forma de mensagens tais como uma lista de rotas planejadas e cobrem um largo escopo da rede aumentando a complexidade das redes de comutação.

Do ponto de vista de localização, as sinalizações podem se situar em dois contextos principais: sinalização entre o assinante e a rede e sinalização entre comutadores na rede. Tipicamente, a operação de sinalização é bem diferente nesses dois contextos.

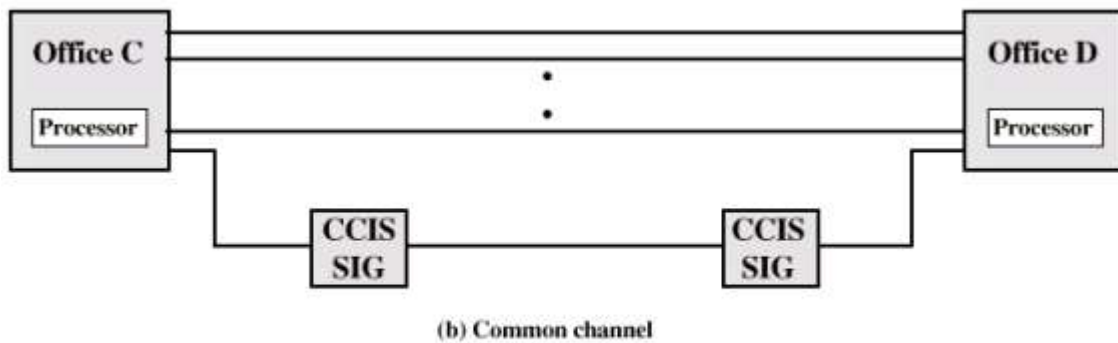
A figura abaixo apresenta um contexto onde a sinalização de controle da rede acontece em cada um dos circuitos estabelecidos, utilizando ou a frequência básica de 4KHz ou as frequências de guarda de cada canal de voz:



Na sinalização in band existem alguns fatores limitantes:

- A taxa de transferência de dados é limitada
- O atraso do call setup (estabelecimento da chamada) ainda é muito grande
- Ocorre sobreposição no uso dos canais de comunicação

Em função dessas limitações as pesquisas evoluíram no sentido de se criar um canal de sinalização comum a todos os circuitos conforme descreve a figura abaixo:



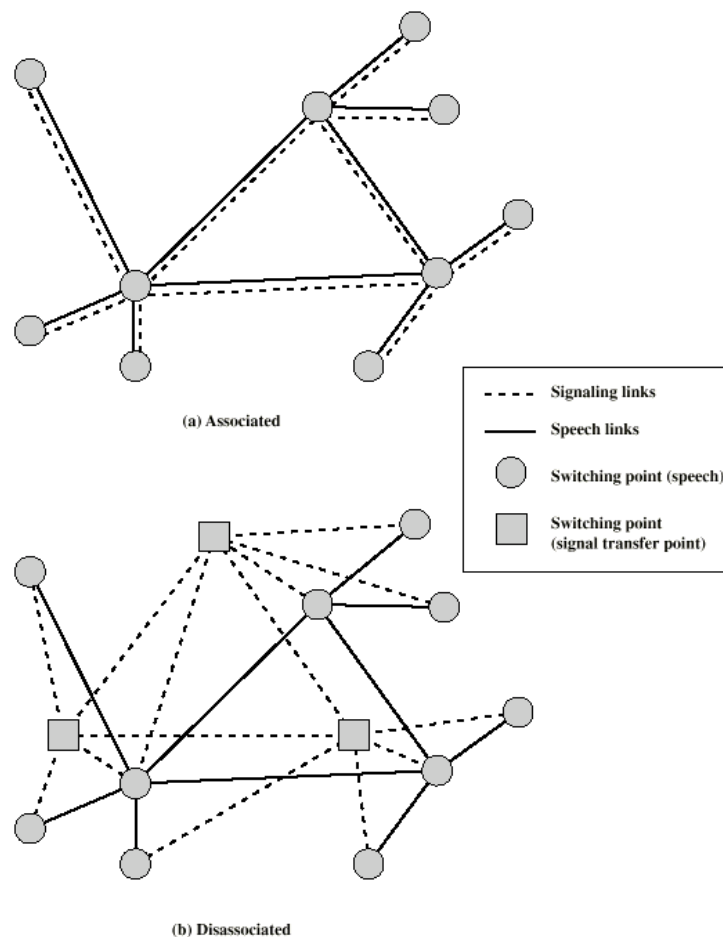
CCIS SIG: Common-channel interoffice signaling equipment  
SIG: Per-trunk signaling equipment

Algumas características da mecanismo de sinalização comum podem ser mencionados (CCS – Common Channel Signaling):

- Sinais de controle são carregados sobre caminhos independentes dos que são utilizados nos canais de voz (uma rede separada para sinalização)
- Um canal de controle pode carregar sinais de vários assinantes
- Taxa de transmissão de sinais é bem superior em relação à transmissão de sinais no esquema anterior.

Do ponto de vista de formação dessa rede paralela de sinalização existem duas possibilidades:

- ✓ Modo associado, onde a rede de sinalização segue a mesma topologia da rede de comutação de voz
- ✓ Modo não associado, onde a topologia é diferente da rede de comutação de circuitos tradicional. A figura abaixo apresenta essas duas configurações:



## COMUTAÇÃO DE PACOTES

Sistemas de comutação são formados por um conjunto de nós e enlaces de interconexão cuja estrutura pode ser utilizada para comunicação de longa distância (WAN) para a transmissão de voz, dados etc., conforme demonstra a figura a seguir:

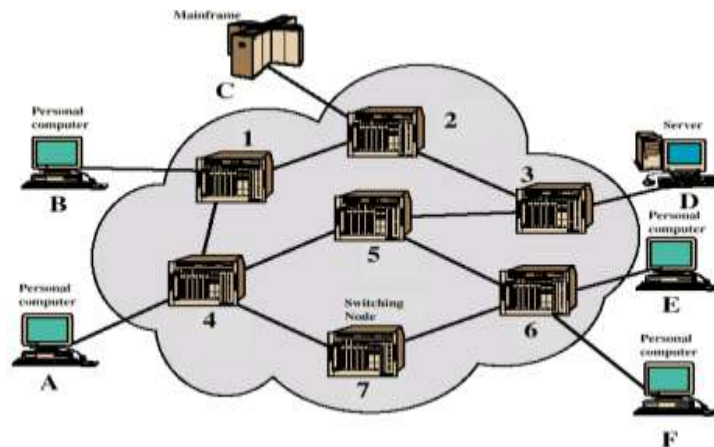


Fig.1

Conceitualmente os sistemas de comutação podem ser vistos como uma sub-rede de comunicação (representada pela nuvem da figura) com as seguintes características:

- ❑ Normalmente a rede não é completamente conectada (não existem enlaces cruzados entre todos os nós)
- ❑ Alguns nós possuem apenas a função de repassar adiante as informações que chegam, como é o caso dos nós 5 e 7. Os demais nós, além de repassar as informações adiante, devem atender as conexões com as estações usuárias finais
- ❑ Os troncos de interligação de nós normalmente usam algum esquema de multiplexação FDM ou TDM.

O sistema de transmissão de dados nas redes de comutação pode ser feito sob dois aspectos: o primeiro deles é a transmissão por comutação de circuitos, muito utilizados para telefonia; o segundo é a transmissão por comutação de pacotes, mais voltada para a transferência de dados. As seções que se seguem fazem uma abordagem sobre a comutação de pacotes:

### 1. Princípios da Comutação de Pacotes

Desde a invenção do telefone, a comutação de circuitos tem sido a tecnologia dominante nas operadoras de telecomunicações para transmissão de voz, apesar de, atualmente, existirem várias outras tendências tecnológicas em função das novas necessidades dos usuários. A comutação de pacotes surgiu para atender essas novas demandas. A transmissão de dados em redes de comutação de circuitos possui as seguintes características:

- ❑ A premissa da comutação de circuitos é que os recursos são dedicados para os usuários e não há mais de um usuário usufruindo um mesmo recurso (banda alocada, por exemplo).
- ❑ Numa transmissão entre computadores não existe a característica de transmissão contínua de dados. Normalmente o que ocorre são rajadas de dados seguidas por momentos de silêncio. Por isso os recursos da rede acabam ficando disponíveis na maior parte do tempo.
- ❑ A taxa de dados é fixa, ou seja, as entidades pares só podem se comunicar via rede de comutação de pacotes se as taxas de transmissão de ambas forem iguais.

Por sua vez, na comutação de pacotes os dados são transmitidos em pequenos blocos de dados de tamanho fixos, normalmente em torno de 1000 octetos. No caso de mensagens grandes, essas são quebradas em pedaços menores para se adequarem ao tamanho máximo do pacote. Além disso, cada pacote deve conter uma parte de dados do usuário e outra parte contendo informações de controle (dados sobre endereçamento de origem/destino etc.). Os pacotes, por sua vez, são recebidos, armazenados brevemente (bufferizados) e passados para o próximo nó. Essa técnica é conhecida como “store-and-forward”. A figura abaixo apresenta uma visão geral da utilização de pacotes para comunicação entre dois computadores:

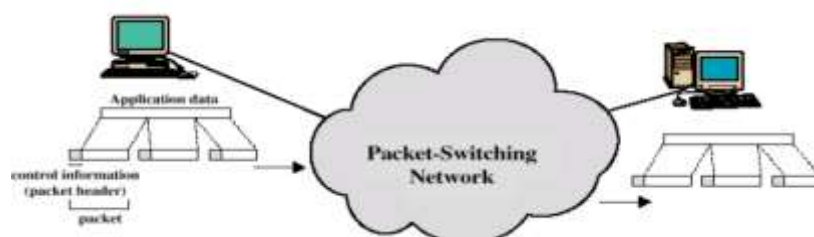


Fig.2

Algumas vantagens podem ser mencionadas sobre comutação de pacotes:



- ⇒ Melhoria no uso (mais eficiência) das linhas de transmissão. Isso ocorre por que um mesmo enlace ponto-a-ponto pode ser compartilhado por muitos pacotes no tempo e os pacotes são enfileirados e transmitidos tão rápido quanto possível, liberando o recurso para os demais.
- ⇒ Quanto à taxa de conversão de dados, cada estação pode se conectar ao nó local na velocidade que julgar mais apropriada. Por consequência, os comutadores devem possuir *buffers* de dados para compatibilizar essas velocidades.
- ⇒ Uma característica interessante é que os pacotes são aceitos mesmo quando a rede está ocupada. Nesse caso o que ocorre é que a liberação dos pacotes para o destino fica mais lenta.
- ⇒ Por último, é possível definir prioridades entre pacotes.

## 2. Técnica de Comutação por Pacotes

De um modo geral a comutação vê as mensagens do usuário como blocos de dados a serem despachados para o destino. Dessa forma as mensagens que chegam para a rede são segmentadas em pacotes de dados associados a um cabeçalho de controle. Do ponto de vista de comutação, os pacotes podem ser manuseados de duas maneiras:

a) Uma das formas de transporte de dados por comutação de pacotes é o sistema de datagrama. Nesse mecanismo os dados do usuário trafegam em pacotes independentes e há o risco de pacotes chegarem fora de ordem ou mesmo se perderem. Cabe ao receptor reordenar os pacotes e recuperar os perdidos, caso ocorram.

b) Uma outra maneira de transportar pacotes é na forma de circuito virtual. Nesse esquema pode-se mencionar as seguintes características:

- Uma rota deve ser definida antes que os pacotes sejam enviados
- Uma requisição de chamada e sua consequente aceitação (“call request” e “call accept”) estabelecem uma conexão
- Cada pacote contém um identificador de circuito virtual ao invés de um endereço de destino
- Nenhuma decisão de rota é requerida para cada pacote individualmente
- O encerramento da conexão (“clear request”) desfaz o circuito
- Não existe um caminho dedicado, ou seja, os recursos podem ser compartilhados.

Independente da forma de transporte de dados adotada na comutação de pacotes, é possível perceber que há uma melhoria na utilização da rede, que passa a ser mais bem compartilhada entre os diversos usuários. Para ilustrar esse esquema, é interessante analisar a figura a seguir:

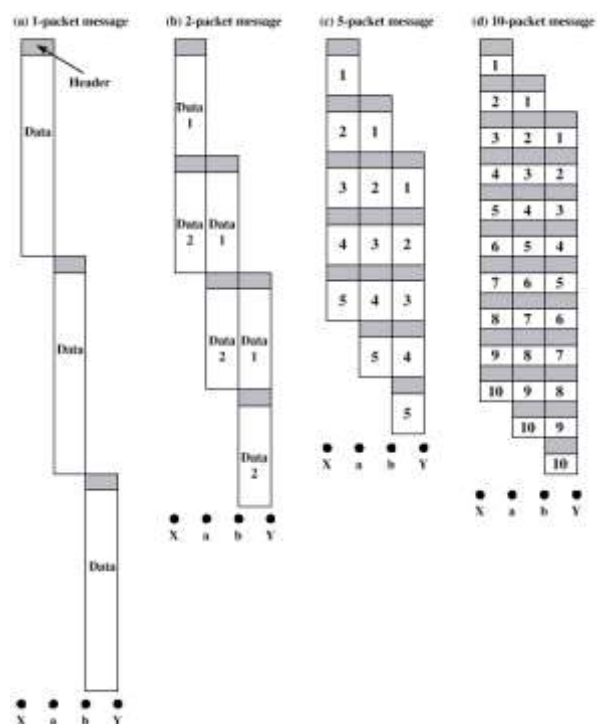


Fig.3

Neste exemplo, é assumido que existe um circuito virtual da estação X até a estação Y, passando pelos nós a e b. A mensagem a ser enviada possui 30 octetos e cada pacote contém 3 octetos de informação de controle (cabeçalho). Se considerarmos a transmissão serial dessa mensagem, ela será enviada como um único pacote de 33 octetos (30 de dados + 3 de cabeçalho). Dessa forma o pacote



é transmitido de X para a, de a para b e, finalmente, de b para Y, ocupando um tempo de transmissão equivalente a 99 octetos. Por outro lado, se essa mensagem for quebrada em dois pacotes de 15 octetos cada, ocorrerá o fenômeno de *pipelining*, onde os nós da rede trabalham em paralelo para atender uma determinada demanda. Nesse caso, a parte b da figura mostra que o tempo de transmissão será menor do que no primeiro caso (o equivalente a transmissão de 72 octetos). Seguindo esse raciocínio, se a mensagem for quebrada em 5 partes de 6 octetos cada, o tempo de transmissão diminui para 63 octetos. Por fim, a parte d da figura mostra que se houver uma fragmentação muito próxima ao tamanho do cabeçalho, o *pipelining* pode não ser tão compensatório (em outras palavras, existe um limite a ser considerado para o particionamento de mensagens). Ainda mais, o exemplo não apresenta o processamento e o enfileiramento das mensagens nos comutadores. Esses atrasos são tão maiores quanto mais pacotes são manuseados por uma única mensagem.

Obs.: Em contradição ao que foi dito, a tecnologia ATM trabalha com pacotes extremamente pequenos e, mesmo assim, produz uma performance bastante aceitável. Essa contradição ocorre por que existem outros aspectos do projeto ATM que devem ser considerados e que serão considerados num estudo mais adiante.

### 3. Comparação entre Comutação de Circuito e Comutação de Pacotes

Tendo visto o funcionamento interno de uma rede de comutação de pacotes, é possível compará-la com a comutação de circuitos com relação à performance. A figura abaixo ilustra um diálogo entre duas estações usando as tecnologias vistas. Nesse caso estão sendo considerados três tipos de atrasos:

- ❑ Atraso de propagação – É o tempo que o sinal leva para se propagar de um nó até o próximo. Este tempo normalmente é insignificante em função das capacidades de transmissão dos meios físicos atuais. Como exemplo, um sinal eletromagnético viaja por um cabo a uma velocidade em torno de  $2 \times 10^8$  metros/s.
- ❑ Tempo de transmissão – É o tempo que o transmissor leva para enviar um bloco de dados. Como exemplo, um transmissor em geral leva um segundo para transmitir um bloco de dados de 10 mil bits em uma linha de 10kbps.
- ❑ Atraso do nó – Esse é o tempo que um nó demora em realizar a comutação propriamente dita.

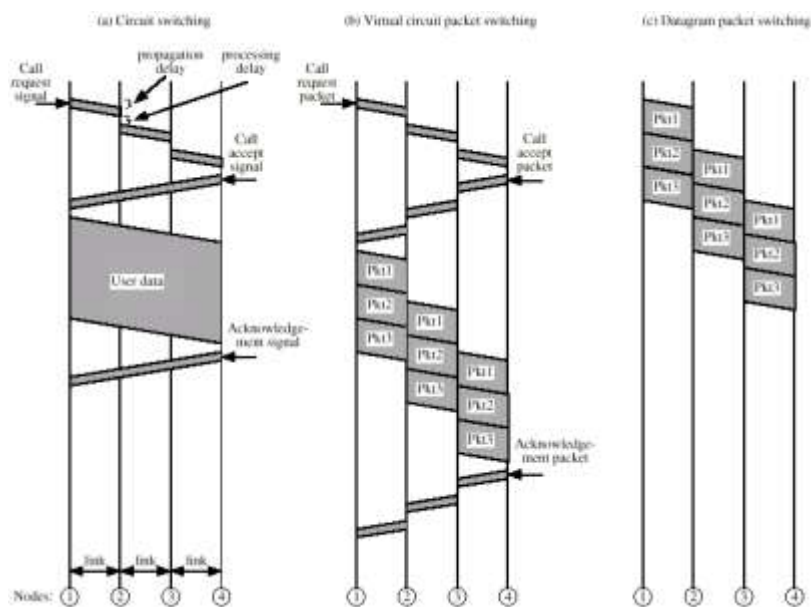


Fig.4

Algumas observações importantes:

- Na comutação de circuitos existe um atraso antes do envio da mensagem. Na solicitação de chamada ("call request") esse atraso é formado pelo atraso de propagação mais o atraso de processamento dos comutadores; na aceitação da chamada ("call accept") esse atraso é formado apenas pelo atraso de propagação uma vez que os recursos já foram alocados e registrados nos comutadores. Uma vez que o caminho está pronto, a mensagem pode ser enviada como um grande bloco e o atraso nos comutadores é insignificante.
- Na comutação de pacotes por comutação de circuitos também existe a necessidade de um estabelecimento de circuito, mas tanto a solicitação de circuito quanto a resposta da solicitação experimentam atrasos de propagação e de processamento dos nós. Além disso, como a transmissão de dados ocorre na forma de pacotes, os atrasos de processamento dos nós devem ser considerados (gerenciamento das filas de entrada e saída de pacotes). Deve estar claro para o leitor que a comutação de pacotes por circuito virtual não é necessariamente mais rápida do que a comutação de circuitos. Para concluir isso, basta perceber que na comutação de circuitos a taxa de transmissão é constante enquanto que na comutação de pacotes os atrasos são variáveis e podem aumentar com a carga da rede.

- Nas redes de pacotes por datagrama não existe fase de estabelecimento. Os dados são enviados diretamente e a performance, principalmente para mensagens curtas, é bem superior do que nas duas técnicas anteriores.
- É importante lembrar que apesar da figura acima apresentar uma caracterização de atraso, a performance de uma rede de comutação depende de uma série de fatores além dos que foram mencionados, tais como o tamanho e a topologia da rede, o padrão de carga etc.

Além das questões de performance existem outros pontos de comparação que podem ser mencionados. A tabela abaixo relaciona essas três tecnologias apresentando outras características de cada uma delas:

Comutação de Circuitos	Com. Pacotes – Datagrama	Com. Pacotes - Circuito Virtual
Caminho de transmissão dedicado	Não há caminho dedicado	Não há caminho dedicado
Transmissão contínua de dados	Transmissão de pacotes	Transmissão de pacotes
A unidade de informação é a <u>mensagem</u> e essas não são armazenadas nos comutadores	A unidade de informação é o pacote e esses podem ser armazenados e depois liberados.	A unidade de informação é o pacote e esses são recebidos e liberados imediatamente.
O caminho é estabelecido para o diálogo inteiro	Uma rota é estabelecida para cada pacote	Rota estabelecida para a conversa inteira
Atraso de "call setup"; atraso de transmissão insignificante	Atraso de transmissão do pacote	Atraso de "call setup" e atraso de transmissão dos pacotes.
Sinal de ocupado se destinatário está ocupado	Chamador pode ser avisado se o pacote não for entregue.	Chamador será notificado sobre conexão não permitida.
Sobrecarga pode bloquear novos "call setup's"; Carga grande não gera atraso nas chamadas já ativas.	Sobrecarga aumenta atraso na entrega dos pacotes	Sobrecarga pode bloquear "call setup's"; aumenta o atraso na entrega de pacotes
Nós de comutação são eletromecânicos ou computadorizados	Pequenos nós de comutação (computadores)	Pequenos nós de comutação (computadores)
Usuário é o responsável por proteção quanto à perda de mensagens	A rede pode ser responsável por pacotes individuais	A rede pode ser responsável por sequências de pacotes
Normalmente não há conversão de códigos ou de velocidades de transmissão.	Há conversão de códigos e taxas de transmissão	Há conversão de códigos e taxas de transmissão
Transmissão em largura de banda fixa	Usa largura de banda variável	Usa largura de banda variável
Não há bits de "overhead" após o "call setup"	Há bits de "overhead" em cada pacote	Há bits de "overhead" em cada mensagem

#### 4. Operação Interna e Externa das Redes de Comutação

Neste ponto é possível diferenciar entre o que é oferecido para o usuário (visão externa da rede) e o modo como a rede operacionaliza o serviço proposto. A figura abaixo mostra uma visão externa de uma rede de comutação de pacotes:

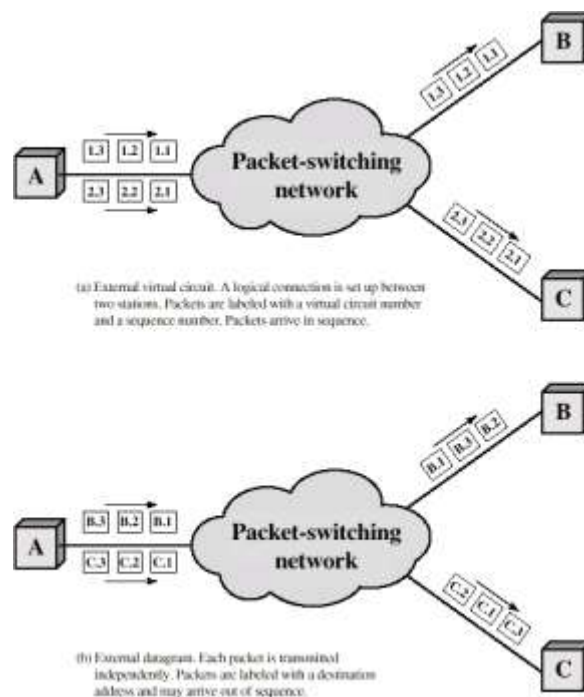


Fig.5

Na parte (a) da figura, uma conexão lógica é estabelecida entre as duas estações finais, garantindo a entrega ordenada de pacotes. Cada pacote é rotulado com um identificador de circuito virtual e um número de sequência para prover garantias ao usuário. Como exemplo, é possível citar o protocolo TCP da camada de transporte do modelo TCP/IP. Na parte (b) da figura, não existe uma conexão lógica entre as partes comunicantes e os pacotes podem chegar fora de ordem no destino. Um exemplo de implementação desse mecanismo é o protocolo UDP do modelo TCP/IP.

Do ponto de vista interno a rede pode também funcionar como um circuito virtual ou como um datagrama, independente do que foi proposto ao usuário como serviço, conforme demonstra a figura a seguir:

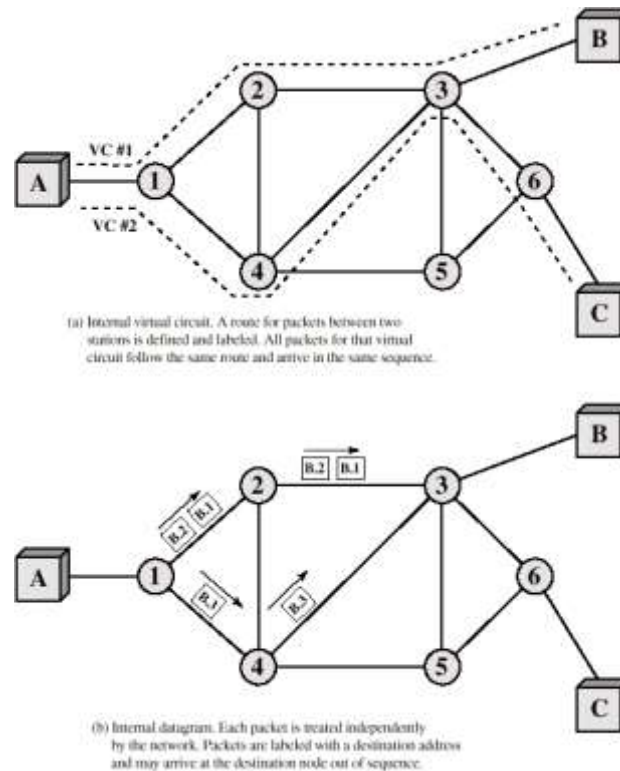


Fig.6

Na parte (a) da figura, uma rota entre duas estações é formada e rotulada. Todos os pacotes para um determinado circuito virtual seguem a mesma rota e chegam em sequência. Na parte (b) não há formação de circuito virtual e os pacotes são roteados independentemente, podendo até mesmo chegar fora de ordem. Como exemplo do primeiro caso, pode-se mencionar a tecnologia Frame Relay (vista mais adiante). O protocolo IP do modelo TCP/IP é um caso prático de rede por datagrama.

A questão que surge é qual a melhor escolha a ser feita tanto para a visão externa como para a visão interna da rede. Essa resposta depende dos objetivos de comunicação da rede e os fatores de custo que devem prevalecer. Algumas observações podem ser lembradas quanto aos dois mecanismos:

- O serviço de datagrama interno e externo permite uso eficiente da rede; não há fase de “call setup” e não há necessidade de pendurar pacotes enquanto uma retransmissão por erro é feita. Essa última característica é interessante para aplicações de tempo real.
- O serviço de circuito virtual pode prover um sequenciamento fim-a-fim e controle de erro. Este serviço é atrativo para suporte de aplicações orientadas à conexão tais como transferência de arquivo e terminal de acesso remoto.

Na prática o serviço de circuito virtual é muito mais comum do que o serviço de datagrama, em função da confiabilidade e conveniência associadas (90% das aplicações Internet trabalham com TCP, 10% usam UDP).

## 5. Roteamento em Redes de Comutação de Pacotes

Um dos aspectos mais complexos e cruciais no projeto de uma rede é o roteamento. Uma estratégia de roteamento em geral envolve os seguintes requisitos: correteza, simplicidade, robustez, estabilidade, justiça, otimalidade e eficiência.

Correteza e simplicidade são auto explicativos. Robustez é a habilidade da rede em liberar pacotes via alguma rota alternativa para fugir de eventuais falhas e/ou sobrecargas em partes da rede. Idealmente a rede deve reagir a tais contingências sem perder pacotes ou quebrar circuitos virtuais. O projetista que procura robustez talvez não consiga atender o requisito de estabilidade, uma vez que esses são conflitantes. As técnicas que reagem às condições de mudança têm uma tendência infeliz de, ou reagir muito lentamente

aos eventos ou experimentar movimentos instáveis de um extremo ao outro. Por exemplo, uma rede pode reagir a um congestionamento em uma área por transportar a maior parte da carga para uma segunda área. Agora a segunda área está sobrecarregada e a primeira está subutilizada, causando um segundo movimento de carga. Durante essa movimentação de carga, os pacotes podem viajar através da rede por um bom tempo sem chegar a lugar algum.

Justiça e otimalidade também são critérios contraditórios. Por exemplo, uma situação ótima e que resulta numa melhoria de performance da rede como um todo, é dar maior prioridade para troca de informações entre estações mais próximas entre si, como é o caso das estações A/A', B/B' e C/C' na figura abaixo. O problema é que essa otimização acaba produzindo uma injustiça com as estações mais distantes, como é o caso das estações X/X' apresentadas na figura.

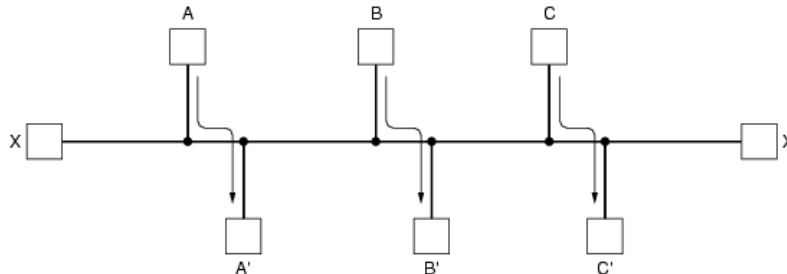


Fig.7

Finalmente, qualquer técnica de roteamento envolve algum "overhead" de processamento em cada nó e também nas linhas de transmissão, impactando na eficiência da rede. Dependendo das características das métricas e parâmetros utilizados para a decisão de roteamento, esse impacto pode ser maior ou menor.

Existem critérios Inerentes aos algoritmos de roteamento que vale a pena mencionar. Alguns deles estão colocados nos itens a seguir:

- Critério de Performance** – A seleção de uma rota é geralmente baseada no critério de melhor performance. O critério mais simples é a escolha de uma rota que passe pela menor quantidade de nós ("hops"). No caso da figura abaixo, o melhor rota entre o nó 1 e o nó 6 é 1-3-6. No entanto, se consideramos como métrica o caminho cujo custo das linhas seja o menor, então o melhor caminho entre 1 e 6 seria 1-4-5-6 (custo = 4, menor do que o custo = 10 de 1-3-6).

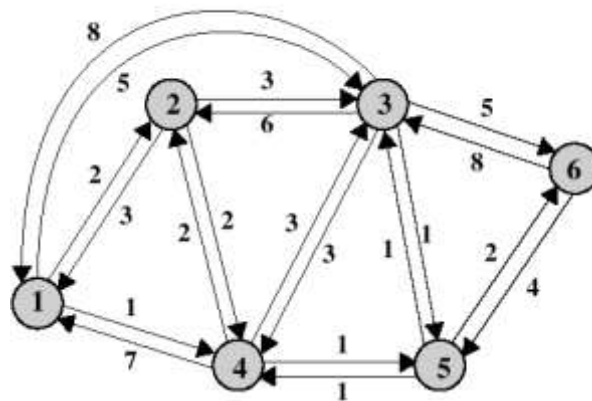


Fig.8

**Obs.:** O custo apresentado na figura pode ser oriundo de uma série de fatores, tais como: tamanho das filas nos comutadores, tráfego nas linhas, taxa de transmissão das linhas etc.

- Em que momento a decisão de roteamento irá ocorrer** – Para redes tipo circuito virtual, a decisão de roteamento na fase de estabelecimento da conexão lógica, enquanto que nas redes datagrama a decisão ocorre no momento em que os pacotes chegam no comutador
- Quem vai tomar a decisão de roteamento** – Essa decisão pode ser: centralizada num único roteador (que mantém informações sobre a rede como um todo), pode ser distribuída entre os diversos comutadores ou pode ser concentrada na estação de origem da informação ("source routing").
- Qual é a fonte de informação de rede para o roteamento** – A maioria das estratégias de roteamento determinam que decisões sejam tomadas com base no conhecimento sobre a topologia da rede, sobre a carga na rede e sobre os custos dos enlaces. Nesse caso as informações normalmente vêm dos comutadores vizinhos ou mesmo de todos os comutadores da rede (como é o caso do roteamento centralizado). Existem estratégias que se baseiam apenas em informações locais como por exemplo, o tamanho das filas internas das portas de saída dos comutadores. Por fim, existem estratégias que não se baseiam em nada, como é o caso do "flooding" (visto mais adiante) que não analisa métrica nenhuma para realizar roteamento.

- e) Qual é o momento de atualizar as informações da rede – Se nenhuma informação de rede é necessária, então essa questão de tempo não se aplica. No entanto, para estratégias que coletam dados de outros nós da rede (nós da rede inteira ou dos nós adjacentes), é preciso definir de quanto em quanto tempo essas informações são trocadas. As alternativas de roteamento que são fixas não precisam sofrer atualizações, mas as técnicas que procuram adaptar o roteamento em função das mudanças na rede precisam ter dados coletados a intervalos regulares.

Quanto à estratégia de roteamento os algoritmos podem ser agrupados em algoritmos adaptativos e não adaptativos. As sub-seções que se seguem abordam cada um deles.

### 5.1 Algoritmos Não Adaptativos

Esses algoritmos não se baseiam nas medidas ou estimativas de tráfego corrente nem em mudanças na topologia da rede. A escolha de rota entre um par de nós qualquer é feita antecipadamente quando a rede é inicializada. Este procedimento é algumas vezes chamado de roteamento estático. Algumas técnicas podem ser descritas:

- ⇒ Menor Caminho (“Shortest Path”) – Existem várias estratégias para implementar esse tipo de algoritmo, dentre elas a mais famosa é uma solução dada por Dijkstra, na qual a rede é vista como um grafo de nós passíveis de fazerem parte de uma rota. Inicialmente todos os nós são marcados como nós-tentativa. Na medida em que a rota vai sendo formada, os nós-tentativa vão sendo transformados em nós-permanentes. A figura a seguir mostra o processo de descoberta do menor caminho entre A e D.

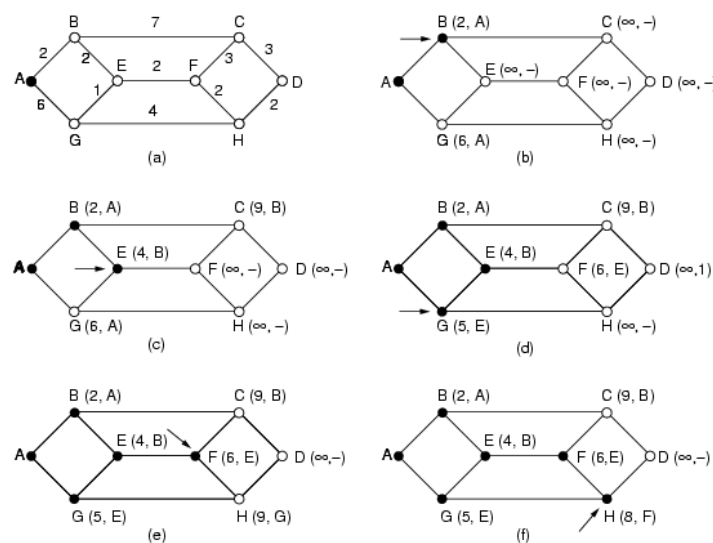


Fig.9

- ⇒ Inundação (“flooding”) – Essa técnica não requer informação da rede para funcionar. Um pacote é enviado da origem para cada um de seus vizinhos que, por sua vez, retransmitem os pacotes de entrada pelas portas de saída, conforme demonstra a figura abaixo:

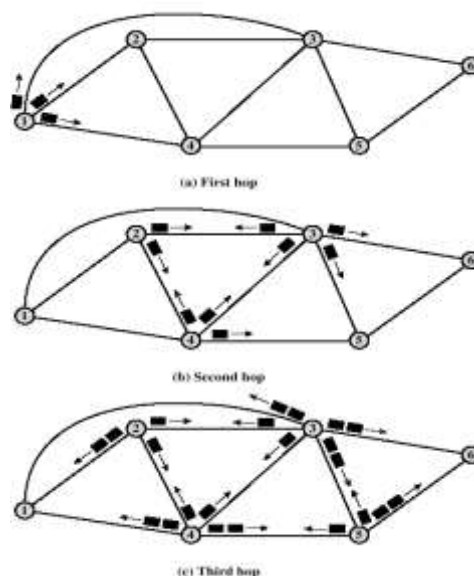


Fig.10

- ⇒ Roteamento Randômico – O roteamento randômico é similar ao “flooding”, só que os comutadores liberam os pacotes de entrada pelas portas de saída cuja probabilidade de alcançar o destino seja maior. O cálculo das probabilidades de cada porta não é assunto dessa apostila.

## 5.2 Algoritmos Adaptativos

Os algoritmos adaptativos são aqueles que possuem métricas associadas ao comportamento atual da rede, tais como tráfego na rede, tamanho das filas, falhas num comutador, queda de um enlace, congestionamento etc. Para o roteamento adaptativo ser possível, informações sobre o estado da rede devem ser trocadas entre os nós de comutação. Nesse caso é preciso existir um balanceamento entre a qualidade da informação trocada e o “overhead” gerado. Ou seja, quanto mais informações são trocadas e quanto menor o intervalo entre essas trocas, melhor a qualidade do roteamento. Por outro lado, quanto mais informações de controle na rede, maior o “overhead”, causando degradação de performance na rede. Existem vários pontos negativos associados com roteamento adaptativo, dentre eles:

- ⇒ A decisão de roteamento é mais complexa e, portanto, a responsabilidade de processamento nos comutadores cresce bastante;
- ⇒ Na maioria dos casos, as estratégias adaptativas dependem de informação de estado (“status”) que é coletada em um local e processada em outro, o que também gera aumento de tráfego na rede;
- ⇒ Uma estratégia adaptativa pode reagir muito rapidamente, causando oscilação na produção de congestionamento. Por outro lado, se a reação é lenta demais, o resultado não será significativo.

Independente desses perigos reais, as estratégias de roteamento adaptativo são largamente utilizadas por duas razões:

- ⇒ Uma estratégia de roteamento adaptativo pode melhorar a performance da rede, como pode ser percebido pelo usuário da rede;
- ⇒ Uma estratégia de roteamento adaptativo pode ajudar no controle de congestionamento da rede, conforme pode ser visto mais adiante.

Esses benefícios podem ou não ser alcançados, dependendo da clareza do projeto da rede e a natureza da carga. Em redes largas e complexas esses problemas são bastante difíceis de gerenciar apropriadamente. Uma demonstração disso é o fato da rede ARPANET (e outras redes produzidas pela IBM e DEC) que é o pilar da Internet já ter lançado várias versões do mesmo protocolo tentando resolver os problemas percebidos na versão anterior.

Para ilustrar roteamento adaptativo vamos considerar na rede da figura 8, que existe uma necessidade de roteamento do nó 1 até o nó 6. Nesse caso os comutadores tomam as decisões de roteamento baseado apenas nas informações internas do comutador e esses, por sua vez, liberam os pacotes nas portas cujo tamanho de fila seja o menor possível. Aqui existe um outro parâmetro, denominado BIAS que é o custo ou probabilidade dessa porta de saída ser apropriada para alcançar o destino.

Imagine agora que o nó 4 acabou de receber um pacote e precisa repassá-lo adiante em direção ao nó 6. A figura abaixo apresenta a fila Q e o BIAS B de cada uma das portas de saída. Nesse caso a porta escolhida será a que se conecta ao nó 3 por que é a porta que tem a menor soma entre esses dois parâmetros.

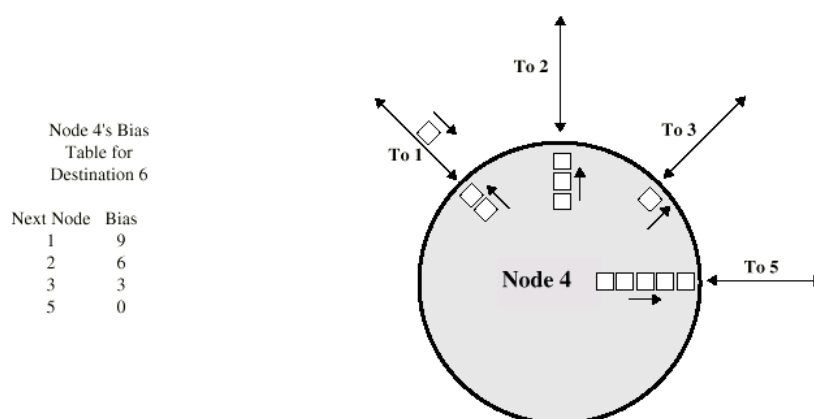


Fig.11

Esquemas baseados apenas em informação local são raramente usados por que eles não exploram facilmente a informação disponível. Ou seja, é difícil chegar a uma conclusão próxima do ideal. As estratégias baseadas em informações de nós adjacentes e de todos os nós são mais comumente encontradas. Dois algoritmos mais famosos que seguem essa linha de raciocínio são o “distance vector” usado no protocolo RIP (‘Routing Information Protocol’) e o “link state” usado no protocolo OSPF, bastante popular atualmente.

No “distance vector” por exemplo, cada roteador mantém uma tabela de rotas indexada por uma entrada para roteador da sub-rede. Esta entrada contém duas partes: uma linha de saída preferencial para usar como destino e uma estimativa de tempo e distância para aquele destino. A métrica usada pode ser o número de saltos (“hops”), o atraso em milissegundos, o número total de pacotes enfileirado

ao longo do caminho ou algo similar. Ao roteador é assumido conhecer a “distância” de cada um de seus vizinhos, de acordo com a métrica decidida. Como um exemplo assume-se que o atraso é usado como métrica e que cada roteador troca informações com seus vizinhos contendo os atrasos estimados para cada destino.

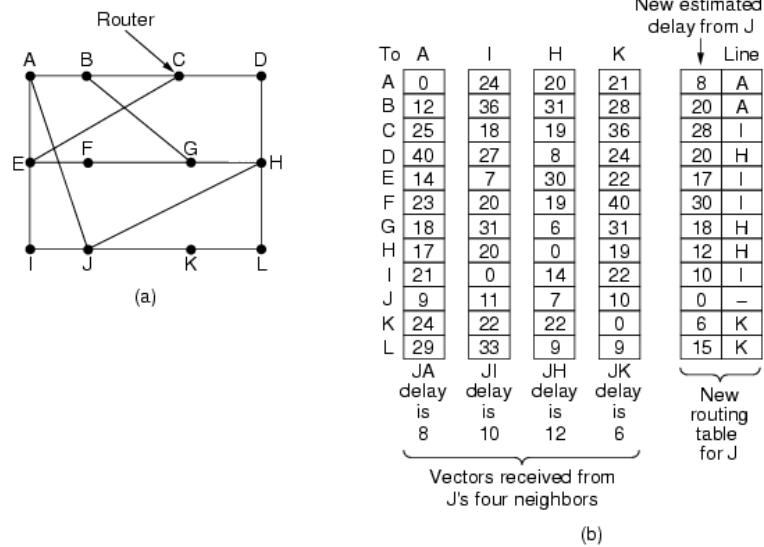


Fig.12

Na figura acima a parte (a) representa a sub-rede como um todo e a parte (b) apresenta os vetores de distância recebidos por J de seus vizinhos. Como J sabe o seu atraso em relação aos vizinhos, será possível recalcular um novo vetor de melhor caminho, considerando as linhas cujo atraso acumulado seja menor em relação a um determinado destino. A parte direita da figura representa a nova tabela de roteamento calculada em função das informações recebidas.