

Capítulo 1

Introdução

A comunicação é uma das maiores necessidades da sociedade humana desde os primórdios de sua existência. Conforme as civilizações se espalhavam, ocupando áreas cada vez mais dispersas geograficamente, a comunicação a longa distância se tornava cada vez mais uma necessidade e um desafio. Formas de comunicação através de sinais de fumaça ou pombos-correio foram as maneiras encontradas por nossos ancestrais para tentar aproximar as comunidades distantes.

A invenção do telégrafo por Samuel F. B. Morse em 1838 inaugurou uma nova época nas comunicações. Nos primeiros telégrafos utilizados no século XIX, mensagens eram codificadas em cadeias de símbolos binários (código Morse) e então transmitidas manualmente por um operador através de um dispositivo gerador de pulsos elétricos. Desde então, a comunicação através de sinais elétricos atravessou uma grande evolução, dando origem à maior parte dos grandes sistemas de comunicação que temos hoje em dia, como o telefone, o rádio e a televisão [Keiser 89].

A evolução no tratamento de informações não aconteceu somente na área da comunicação. Equipamentos para processamento e armazenamento de informações também foram alvo de grandes invenções ao longo do nosso desenvolvimento. A introdução de sistemas de computadores na década de 1950 foi, provavelmente, o maior avanço do século nesse sentido.

A conjunção destas duas tecnologias — comunicação e processamento de informações — veio revolucionar o mundo em que vivemos, abrindo as

3

fronteiras com novas formas de comunicação, e permitindo maior eficácia dos sistemas computacionais. Redes de computadores são hoje uma realidade neste contexto. Para que possamos entendê-las, é necessário que observemos como se deu a evolução dos sistemas de computação até os dias de hoje, onde a distribuição do poder computacional é uma tendência indiscutível.

1.1 - Evolução dos Sistemas de Computação

Na década de 1950, computadores eram máquinas grandes e complexas, operadas por pessoas altamente especializadas. Usuários enfileiravam-se para submeter seus jobs (utilizando-se de leitores de cartões ou fitas magnéticas) que eram processados em lote (*batch*), como ilustrado na Figura 1.1. Não havia nenhuma forma de interação direta entre usuários e máquina. Longos períodos de espera eram comuns até que se pudesse obter algum resultado, dado que todo o processamento era feito job a job de acordo com a ordem em que eram submetidos.



Figura 1.1: Sistemas de processamento em batch.

Avanços na década de 1960 possibilitaram o desenvolvimento dos primeiros terminais interativos, permitindo aos usuários acesso ao computador central através de linhas de comunicação, como podemos observar na Figura 1.2. Usuários passavam a ter então um mecanismo que possibilitava a interação direta com o computador, ao mesmo tempo em que os avanços nas técnicas de processamento davam origem a sistemas de tempo compartilhado (*time-sharing*), permitindo que as várias tarefas dos diferentes usuários ocupassem simultaneamente o computador central, através de uma espécie de revezamento no tempo de ocupação do processador.

Mudanças na caracterização dos sistemas de computação ocorreram durante a década de 1970: de um sistema único centralizado e de grande porte, disponível para todos os usuários de uma determinada organização, partiu-se em direção à distribuição do poder computacional. O desenvolvi-

mento de minis e microcomputadores de bom desempenho, com requisitos menos rígidos de temperatura e umidade, permitiu a instalação de considerável poder computacional em várias localizações de uma organização, ao invés da anterior concentração deste poder em uma determinada área. Com o desenvolvimento tecnológico, a contínua redução do custo do hardware acompanhada do aumento da capacidade computacional levou também ao uso cada vez maior dos microcomputadores. Esses sistemas pequenos e分散式 eram mais acessíveis ao usuário, possuíam uma responsividade melhor e eram mais fáceis de utilizar do que os grandes sistemas centralizados com compartilhamento de tempo.

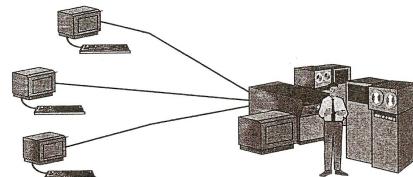


Figura 1.2: Sistema de time-sharing.

Embora o custo do hardware de processamento estivesse caindo, o preço dos equipamentos eletrônicos continuava alto. Mesmo no caso de dados que podiam ser associados a um único sistema de pequeno porte, a economia de escala exigia que grande parte dos dados estivessem associados a um sistema de grande capacidade centralizado. Pela mesma razão de custo, se justificava a utilização compartilhada de periféricos especializados tais como uma impressora rápida e de qualidade. Assim, a interconexão entre os vários sistemas para o uso compartilhado de dispositivos periféricos tornou-se importante (Figura 1.3).

A capacidade de troca de informações também foi uma razão importante para a interconexão. Usuários individuais de sistemas de computação não trabalham isolados e necessitam de alguns dos benefícios oferecidos por um sistema centralizado. Entre esses, encontram-se a capacidade de troca de mensagens entre os diversos usuários e a facilidade de acesso a dados e programas de várias fontes quando da preparação de um documento, ou mesmo da simples análise de dados ou implementação de um programa. Ambientes de trabalho cooperativo se tornaram uma realidade tanto nas empresas como nas universidades, tornando ainda mais necessária a interconexão dos equipamentos nessas organizações.



Figura 1.3: Compartilhamento de recursos.

A busca de soluções para os problemas de performance também impulsionou os pesquisadores a criar novas arquiteturas que propunham a distribuição e o paralelismo como forma de melhorar o desempenho, a confiabilidade e a modularidade dos sistemas computacionais.

1.2 - Evolução das Arquiteturas

A maioria dos computadores projetados até a década de 1980 teve sua concepção baseada no modelo original de *Von Neumann*. O casamento perfeito entre o modo como os programas são desenvolvidos e a maneira como são interpretados foi uma das razões para o grande sucesso de tal modelo. Ele oferece um mecanismo simples e bastante eficiente, desde que a computação seja puramente sequencial.

A revolução nos sistemas de computadores começou com os avanços da tecnologia de integração de circuitos, que reduziram em muito os custos das partes de tais sistemas. Várias arquiteturas foram então propostas, dentro das restrições de tecnologia de cada época, tentando contornar as limitações do modelo de Von Neumann no que diz respeito ao custo, confiabilidade e desempenho. Dentre as alternativas apresentadas, podemos citar os Sistemas de UCP Única com Múltiplas Unidades Funcionais, as Máquinas Pipeline e os Processadores de Matriz (Array Processors).

A idéia de seqüências múltiplas e independentes de instruções em um sistema composto por vários elementos de processamento compartilhando um espaço comum de memória aparece em uma outra arquitetura, contornando a restrição de controle centralizado do modelo Von Neumann, tendo sido citada na literatura como *Sistemas de Multiprocessadores Fortemente Acoplados*.

Acoplados. As principais características desses sistemas, de acordo com Enslow [Enslow 74], são as seguintes:

- Dois ou mais processadores de capacidades aproximadamente iguais.
- Todos os processadores dividem o acesso a uma memória comum.
- Todos os processadores compartilham os canais de entrada/saída, unidades de controle e dispositivos periféricos.
- O sistema total é controlado por um único sistema operacional.

Por último, surgiram os *Sistemas de Processamento Distribuído*, definidos por Eckhouse [Eckhouse 78] como "uma coleção de elementos de processamento interconectados, tanto lógica quanto fisicamente, para execução cooperativa de programas de aplicação, com controle geral dos recursos descentralizado". Tal definição exclui dessa classe os *Sistemas de Multiprocessadores Fortemente Acoplados*.

Em *Sistemas Distribuídos*, também chamados *Sistemas Fracamente Acoplados*, o estado do sistema é fragmentado em partes que residem em diferentes processadores e memórias, com a comunicação entre essas partes sujeita a retardos variáveis e desconhecidos. A diferença marcante entre sistemas fortemente acoplados e sistemas fracamente acoplados reside no fato de que, em sistemas fracamente acoplados a única forma de interação entre os módulos processadores se dá através da troca de mensagens, enquanto que em sistemas fortemente acoplados existe uma memória compartilhada entre os módulos. Em sistemas distribuídos é impossível forçar a simultaneidade de eventos. A mínima interferência na execução de tarefas paralelas vai permitir a obtenção de sistemas de grande desempenho. A não existência de qualquer elemento sem o qual o sistema pára totalmente lhe confere alta confiabilidade. A possibilidade de utilização em larga escala de um pequeno número de elementos básicos de hardware e software é responsável pelo elevado grau modularidade do sistema. Além disso, não existe nenhuma restrição inerente à estrutura que impeça o crescimento do sistema, o que lhe confere alta expansibilidade [Shimizu 80].

Várias são as razões para o uso de sistemas de múltiplos processadores (sejam eles sistemas fortemente ou fracamente acoplados):

- *Custo/desempenho:* a evolução da tecnologia de síntese de circuitos integrados tem conduzido os custos de microprocessadores e memórias a valores bem reduzidos. Um forte argumento para sistemas baseados em microprocessadores é seu alto potencial na relação entre o custo e o desempenho.

- *Responsividade:* um sistema de múltiplos processadores pode apresentar um grande potencial de processamento e responsividade, pois pode ser moldado à aplicação.
- *Modularidade:* existem várias razões para fazermos um sistema de computação modular. A primeira é uma relação custo/desempenho satisfatória para vários tipos de configurações. Por exemplo, um pequeno número de processadores para pequenos volumes de carga e um grande número para volumes elevados. Uma outra razão diz respeito ao crescimento incremental, ou expansibilidade. Um sistema bem projetado pode superar problemas de sobrecarga e/ou abrigar uma maior gama de aplicações pela simples inclusão de processadores. Uma terceira razão vem do fato de podermos utilizar em larga escala um conjunto de componentes básicos para a realização do sistema, o que simplifica não só o projeto, mas também sua futura manutenção.
- *Confiabilidade:* uma vez que a redundância é o ingrediente básico no projeto de um sistema confiável, uma arquitetura contendo um número elevado de componentes idênticos constitui-se em uma ótima estrutura onde a redundância pode ser incluída sem que o sistema seja duplicado como um todo. Além disso, o sistema pode possuir mecanismos de reconfiguração que o tornem tolerante a certas falhas, degradando apenas seu desempenho, podendo apresentar assim uma grande disponibilidade. Mais ainda: em sistemas centralizados as falhas não são confinadas, ou seja, a abrangência de uma falha simples é muito maior, resultando em uma operação de reconfiguração mais complexa e mais cara.
- *Concorrência:* máquinas destinadas a aplicações que requisitam alto desempenho exigem, em geral, a adoção de soluções que envolvam a utilização em larga escala de elementos concorrentes de processamento.

As desvantagens de um sistema de múltiplos processadores podem ou não mascarar as vantagens, de acordo com os requisitos particulares do sistema. Dentre elas podemos citar:

- O desenvolvimento de software aplicativo para tais sistemas pode ser mais complexo, e portanto mais caro, do que para sistemas centralizados, especialmente quando estão envolvidas máquinas de mais de um fabricante.
- A decomposição de tarefas é mais complexa, quer realizada automaticamente pelo software do sistema, ou explicitamente pelo programador.

- O desenvolvimento do software de diagnóstico geralmente é mais difícil e, em consequência, mais caro.
- Um sistema distribuído é mais dependente da tecnologia de comunicação, em particular aqueles em que os processadores estão geograficamente dispersos e a demanda de tráfego de comunicação é alta.
- O tempo de serviço de um sistema com múltiplos processadores pode ultrapassar os limites máximos de tolerância se a estrutura de comunicação entre os processadores não suportar a taxa de transmissão de mensagens necessária.
- Uma falha na estrutura de comunicação pode fazer com que os sistemas de um defeito em um processador reflita em outros.
- Existe uma certa perda de controle em sistemas distribuídos. Neles é difícil gerenciar os recursos, forçar padronizações para o software e dados, e gerenciar informações disponíveis. A manutenção da integridade dos dados, da segurança e da privacidade é também uma tarefa mais complexa.

Embora difícil de caracterizar, a arquitetura de múltiplos processadores tem melhor aplicação em sistemas que exigem grande disponibilidade, grandes requisitos de vazão, tempos de resposta garantidos e baixos, alto grau de modularidade, e também onde as tarefas podem ser executadas de modo concorrente.

Tal como na Figura 1.3, um *Sistema Distribuído*, conforme a definição de Eckhouse, vai ser formado por um conjunto de módulos processadores interligados por um sistema de comunicação. Vemos assim, pelo que discutimos nesta seção e na seção anterior, que a interconexão de sistemas com poder computacional veio atender a duas necessidades distintas: (1) a construção de sistemas com maior desempenho e maior confiabilidade e (2) o compartilhamento de recursos. Alguns autores consideram como Sistema Distribuído apenas aqueles construídos para atender a primeira necessidade, classificando como *Redes de Computadores* os sistemas construídos com a finalidade de permitir o compartilhamento de recursos. Outros autores preferem classificar todos esses sistemas como Sistemas Distribuídos, e subclasseficá-los em Máquinas de Arquitetura Distribuída e Redes de Computadores.

Uma *Máquina de Arquitetura Distribuída* é composta por um número ilimitado mas finito de módulos autônomos de processamento interconectados para formar um único sistema, no qual o controle executivo global é implementado através da cooperação de elementos descentralizados. Não é suficiente que os processadores apareçam para o usuário como um sistema virtual único, é necessário que apareçam como um sistema real único em to-

dos os níveis de abstração. Conceitualmente, um único sistema operacional controla todos os recursos físicos e lógicos de maneira integrada, tendo, no entanto, seu núcleo e suas estruturas de dados distribuídos pelos vários processadores e memórias. Estas cópias do núcleo devem ser entidades individuais que executam concorrentemente, assincronamente e sem qualquer hierarquia ou relação mestre-escravo, de forma a constituir um organismo único [Jensen 78, Shimizu 80].

Uma *Rede de Computadores* também é formada por um número ilimitado mas finito de módulos autônomos de processamento interconectados, no entanto, a independência dos vários módulos de processamento é preservada na sua tarefa de compartilhamento de recursos e troca de informações. Não existe nesses sistemas a necessidade de um sistema operacional único, mas sim a cooperação entre os vários sistemas operacionais na realização das tarefas de compartilhamento de recursos e troca de informações.

Como veremos, ao estudarmos os níveis superiores de protocolos e os sistemas operacionais de redes, nos Capítulos 14 a 16, mesmo a subclassificação de Sistemas Distribuídos apresentada tende a se confundir, assumindo as Redes de Computadores também o papel de Máquinas de Arquitetura Distribuída. Abandonaremos, portanto, a preocupação e a polêmica em classificar tais sistemas, utilizando para Redes de Computadores a definição apresentada na próxima seção.

1.3 - Redes de Computadores

Uma *Rede de Computadores* é formada por um conjunto de módulos processadores (MPs)¹ capazes de trocar informações e compartilhar recursos, interligados por um sistema de comunicação, conforme ilustrado na Figura 1.4.

O sistema de comunicação vai se constituir de um arranjo topológico interligando os vários módulos processadores através de enlaces físicos (*meios de transmissão*) e de um conjunto de regras com o fim de organizar a comunicação (*protocolos*). Redes de computadores são ditas *confiadas* quando as distâncias entre os módulos processadores são menores que alguns poucos metros. Redes Locais de Computadores são sistemas cujas distâncias entre os módulos processadores se enquadram na faixa de alguns poucos metros a alguns poucos quilômetros. Sistemas cuja dispersão é maior

¹ A definição de módulos processadores se refere a qualquer dispositivo capaz de se comunicar através do sistema de comunicação por troca de mensagens. Poderíamos citar, por exemplo, um microcomputador, uma máquina copiadora, um computador de grande porte, um terminal videotexto etc.

do que alguns quilômetros são chamados Redes Geograficamente Distribuídas.

Redes Locais (Local Area Networks — LANs) surgiram dos ambientes de institutos de pesquisa e universidades. Como vimos, as mudanças no enfoque dos sistemas de computação que ocorriam durante a década de 1970 levaram em direção à distribuição do poder computacional. O desenvolvimento de minis e microcomputadores de bom desempenho permitiu a instalação de considerável poder computacional em várias unidades de uma organização ao invés da anterior concentração em uma determinada área. Redes locais surgiram, assim, para viabilizar a troca e o compartilhamento de informações e dispositivos periféricos (recursos de hardware e software), preservando a independência das várias estações de processamento, e permitindo a integração em ambientes de trabalho cooperativo.



Figura 1.4: Rede de Computadores.

Pode-se caracterizar uma *rede local* como sendo uma rede que permite a interconexão de equipamentos de comunicação de dados numa pequena região. De fato, tal definição é bastante vaga principalmente no que diz respeito às distâncias envolvidas. Em geral, nos dias de hoje, costuma-se considerar "pequena região" distâncias entre 100 m e 25 Km, muito embora as limitações associadas às técnicas utilizadas em redes locais não imponham limites a essas distâncias. Outras características típicas encontradas e comumente associadas a redes locais são: altas taxas de transmissão (de 0,1 a 100 Mbps) e baixas taxas de erro (de 10^{-8} a 10^{-11}). É importante notar que os termos "pequena região", "altas taxas de transmissão" ou "baixas taxas de erro" são suscetíveis à evolução tecnológica; os valores que associamos a estes termos estão ligados à tecnologia atual e certamente não serão mais os mesmos dentro de poucos anos. Outra característica dessas redes é que elas são, em geral, de propriedade privada.

Quando a distância de ligação entre os vários módulos processadores começa a atingir distâncias metropolitanas, chamamos esses sistemas não mais de redes locais, mas de *Redes Metropolitanas (Metropolitan Area Networks — MANs)*. A definição do termo "rede metropolitana" surgiu com o aparecimento do padrão IEEE 802.6, conforme teremos oportunidade de discutir no Capítulo 9. Uma rede metropolitana apresenta características semelhantes às das redes locais, sendo que as MANs, em geral, cobrem distâncias maiores do que as LANs operando em velocidades maiores.

Redes Geograficamente Distribuídas (Wide Area Networks — WANs) surgiram da necessidade de se compartilhar recursos especializados por uma maior comunidade de usuários geograficamente dispersos. Por terem um custo de comunicação bastante elevado (circuitos para satélites e enlaces de microondas), tais redes são em geral públicas, isto é, o sistema de comunicação, chamado *sub-rede de comunicação*, é mantido, gerenciado e de propriedade de grandes operadoras (públicas ou privadas), e seu acesso é público. Face a várias considerações em relação ao custo, a interligação entre os diversos módulos processadores em uma tal rede determinará a utilização de um arranjo topológico específico e diferente daqueles utilizados em redes locais, como veremos no Capítulo 2. Ainda por problemas de custo, nos seus primórdios, as velocidades de transmissão empregadas eram baixas: da ordem de algumas dezenas de kilobits/segundo (embora alguns enlaces cheguem hoje a velocidades de megabits/segundo). Por questão de confiabilidade, caminhos alternativos devem ser oferecidos de forma a interligar os diversos módulos processadores.

1.4 - Parâmetros de Comparação

A escolha de um tipo particular de rede para suporte a um dado conjunto de aplicações é uma tarefa difícil. Cada arquitetura possui certas características que afetam sua adequação a uma aplicação em particular. Nenhuma solução pode chamar para si a classificação de ótima quando analisada em contexto geral, e até mesmo em particular. Muitos atributos entram em jogo, o que torna qualquer comparação bastante complexa. Esses atributos dizem respeito ao custo, à confiabilidade, ao tempo de resposta, à velocidade, ao desempenho, à facilidade de desenvolvimento, à modularidade, à capacidade de reconfiguração, à complexidade lógica, à facilidade de uso, à disponibilidade, à facilidade de manutenção, à dispersão geográfica e a outros fatores não técnicos ou quase técnicos.

1.4.1 - Custo

O custo de uma rede é dividido entre o custo das estações de processamento (microcomputadores, minicomputadores etc.), o custo das interfaces com o meio de comunicação e o custo do próprio meio de comunicação. Uma vez que o desenvolvimento tecnológico continuará reduzindo cada vez mais o custo das estações, é necessário que o custo das conexões (interfaces) seja minimizado.

O custo das conexões dependerá muito do desempenho que se espera da rede. Redes de baixo a médio desempenho usualmente empregam poucas estações com uma demanda de taxas de dados e volume de tráfego pequeno. Isso vai permitir o desenvolvimento de interfaces de baixo custo, a despeito de suas limitações para outras aplicações.

Redes de alto desempenho já requerem interfaces de custos mais elevados, devido em grande parte ao protocolo de comunicação utilizado e ao meio de comunicação.

1.4.2 - Retardo de Transferência

Antes de definir o que é retardo de transferência faz-se necessário discutir o que se entende por retardo de acesso e retardo de transmissão.

Chamamos *retardo de acesso* o intervalo de tempo decorrido desde que uma mensagem a transmitir é gerada pela estação até o momento em que a estação consiga obter para ela e somente para ela o direito de transmitir, sem que haja colisão de mensagens no meio. Em outras palavras, retardo de acesso é o tempo que uma estação espera, a partir do momento em que uma mensagem está pronta para ser transmitida, até o momento em que ela segue transmitir essa mensagem com sucesso (sem que outras estações na rede a perturben).

Chamaremos de *retardo de transmissão* o intervalo de tempo decorrido desde o início da transmissão de uma mensagem por uma estação de origem até o momento em que a mensagem chega à estação de destino.

Podemos agora definir o *retardo de transferência* como a soma dos retardos de acesso e de transmissão. Assim, o retardo de transferência inclui todo o tempo de entrega de uma mensagem, desde o momento em que se deseja transmiti-la, até o momento em que ela chega para ser recebida pelo destinatário.

O retardo de transferência é, na grande maioria dos casos, uma variável aleatória, como veremos mais adiante. No entanto, em algumas redes o maior valor que o retardo de transferência pode assumir é limitado

(costuma-se dizer que o retardo de transferência é *determinístico*, embora a palavra, como vemos, seja mal empregada).

A rede deve poder ser moldada ao tipo particular de aplicação de modo a assegurar um retardo de transferência baixo. O sistema de comunicação entre os módulos deve ser de alta velocidade e de baixa taxa de erro, de forma a não provocar saturação no tráfego de mensagens. Em algumas aplicações (em particular as de controle em tempo real) a necessidade de retardo de transferência máxima limitado é de vital importância. Voltaremos a esse assunto mais adiante.

1.4.3 - Desempenho

Várias são as medidas que caracterizam o desempenho de um sistema, entre elas o retardo de transferência anteriormente mencionado, vazão etc. Vamos definir desempenho de uma rede, quando não especificado de outra forma, como a capacidade efetiva de transmissão da rede. Conforme veremos nos próximos capítulos, a utilização efetiva do sistema de comunicação é apenas uma porcentagem da capacidade total que ele oferece, devido a vários fatores que serão vistos.

O requisito baixo custo leva frequentemente ao sacrifício do desempenho. No entanto, uma rede deve proporcionar capacidade suficiente para viabilizar as aplicações a que é destinada.

Encontramos às vezes na literatura a distinção entre redes locais (Local Area Networks – LANs) e redes locais de alta velocidade (High-Speed Local Networks – HSLNs). Redes locais de alta velocidade são projetadas de forma a fornecer um alto desempenho na comunicação entre os dispositivos. Na maioria dos casos tais redes têm um custo de conexão mais elevado. Em todo o livro só faremos distinção entre esses dois tipos de redes quando necessário, uma vez que quase tudo será válido para ambas. Na realidade, muitos não gostam desta distinção, pelo menos sob o ponto de vista didático. Na prática, o conceito de "alta velocidade" em redes locais tem se tornado bastante relativo, dados os avanços tecnológicos na área de transmissão de dados.

Os termos velocidade, desempenho e retardo de transferência estão intimamente relacionados. A escolha adequada da arquitetura, incluindo a estrutura de conexão, o protocolo de comunicação e o meio de transmissão vão influenciar em muito no desempenho, velocidade e retardo de transferência de uma rede.

Em resumo, como veremos, a topologia, o meio de interconexão, o protocolo de comunicação, bem como a velocidade de transmissão influenciam

em muito na adequação de uma rede a uma aplicação particular. A seleção de um mecanismo de interconexão orientado para a natureza da aplicação é essencial para o bom desempenho de uma rede local.

1.4.4 - Confiabilidade

Confiabilidade pode ser avaliada em termos de tempo médio entre falhas (Medium Time Between Failures — MTBF), tolerância a falhas, degradação amena (graceful degradation), tempo de reconfiguração após falhas e tempo médio de reparo (MTTR — Medium Time to Repair).

O *tempo médio entre falhas* é geralmente medido em horas, estando relacionado com a confiabilidade de componentes e nível de redundância. *Degradação amena* é geralmente dependente da aplicação. Ela mede a capacidade da rede continuar operando em presença de falhas, embora com um desempenho menor. *Reconfiguração após falhas* requer que caminhos redundantes sejam acionados tão logo ocorra uma falha ou esta seja detectada. A rede deve ser tolerante a falhas transitórias causadas por hardware e/ou software, de forma que tais falhas causem apenas uma confusão momentânea, que será resolvida em algum nível de reiniciação. Obviamente, falhas de alguns componentes críticos ou destruição de programas não podem ser resolvidas sem recursos de redundância, mas essas não são de modo algum as únicas falhas possíveis. O *tempo médio de reparo* pode ser diminuído com o auxílio de redundância, mecanismos de autoteste e diagnóstico e manutenção eficiente. Várias redes têm incluídos, em suas interfaces, mecanismos de autoteste e diagnóstico para auxílio na manutenção e na realização de medidas de desempenho. Algumas possuem até estações especiais para esses fins.

1.4.5 - Modularidade

Modularidade pode ser caracterizada como o grau de alteração de desempenho e funcionalidade que um sistema (rede) pode sofrer sem mudar seu projeto original. Os três maiores benefícios de uma arquitetura modular são a facilidade para modificação, a facilidade para crescimento e a facilidade para o uso de um conjunto de componentes básicos.

No sentido de facilidade de modificação, modularidade diz respeito à simplicidade com que funções lógicas ou elementos de hardware podem ser substituídos, a despeito da relação íntima com outros elementos. No sentido de facilidade para crescimento, modularidade diz respeito a configurações de baixo custo (por exemplo, uma rede com um pequeno número de módulos para pequenos volumes de carga, e um grande número para volumes ele-