

## Estrutura dos Sistemas Distribuídos

### Referência:

Silberschatz, Abraham. **Sistemas Operacionais com Java**. 7 ed. Rio de Janeiro: Elsevier, 2008.

Cluster desenvolvido na UNIP campus Indianópolis

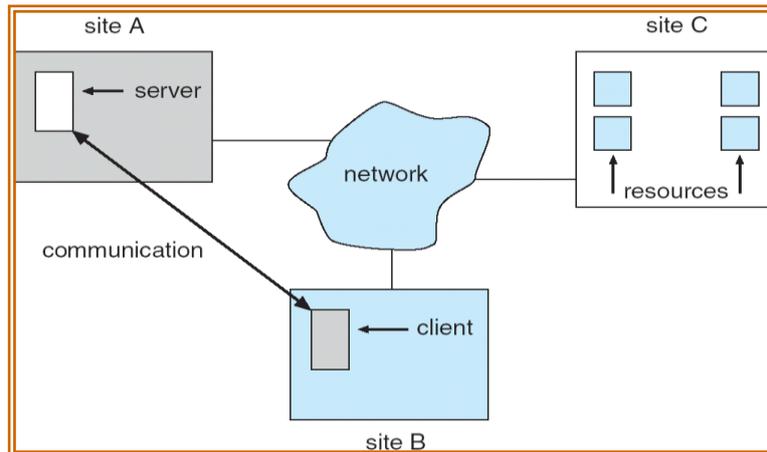
[http://cassiopea.ipt.br/teses/2005\\_EC\\_Fabio\\_Vieira\\_Amaral.pdf](http://cassiopea.ipt.br/teses/2005_EC_Fabio_Vieira_Amaral.pdf)

- Fundamentos
- Tipos de sistemas operacionais em rede
- Estrutura de rede
- Topologia de rede
- Estrutura de comunicação
- Protocolos de comunicação
- Robustez
- Aspectos de projeto

### Fundamentos

- Sistema distribuído é a coleção de processadores pouco acoplados, interconectados por uma rede de comunicações.
- Os processadores são chamados de vários nomes, como *nós*, *computadores*, *máquinas*, *hosts*.
  - Site é o local do processador.
- Motivos para sistemas distribuídos
  - Compartilhamento de recursos.
    - compartilhamento e impressão de arquivos em sites remotos.
    - processamento de informações em um banco de dados distribuído.
    - uso de dispositivos de hardware especializados.
  - Aumento da velocidade de computação – compartilhamento de carga.

- **Confiabilidade** – detecta e recupera de falha do site, transferência de função, reintegra site que falhou.
- **Comunicação** – passagem de mensagem.



**Sistema Distribuído**

### **Tipos de sistemas operacionais distribuídos**

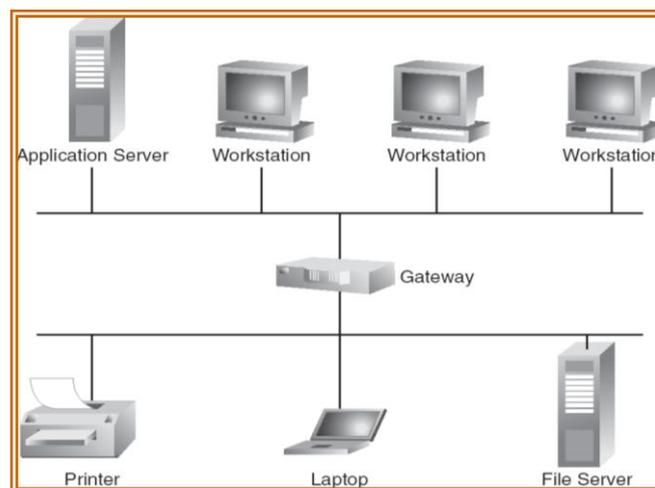
- **Sistemas operacionais de rede**
  - Os usuários estão cientes da multiplicidade de máquinas. Acesso aos recursos de várias máquinas é feito explicitamente por:
    - Logging remoto na máquina remota apropriada (telnet, ssh)
    - Desktop remoto (Microsoft Windows)
    - Transferência de dados de máquinas remotas para máquinas locais, via mecanismo de File Transfer Protocol (FTP)
- **Sistemas operacionais distribuídos**
  - Usuários não cientes da multiplicidade de máquinas
  - Acesso a recursos remotos semelhante ao acesso a recursos locais
  - Migração de dados – transferência de dados transferindo arquivo inteiro, ou transferindo apenas as partes do arquivo necessárias para a tarefa imediata
  - Migração da computação – transferência da computação (ao invés dos dados) pelo sistema

- **Migração de processo** – executa um processo inteiro, ou partes dele, em locais diferentes
- **Balanceamento de carga** – processos distribuídos pela rede ou pela carga de trabalho
- **Ganho de velocidade de computação** – sub-processos podem executar simultaneamente em diferentes locais
- **Preferências de hardware** – execução do processo pode exigir processador especializado
- **Preferência de software** – software exigido pode estar disponível apenas em um local em particular
- **Acesso a dados** – executa processo remotamente, ao invés de transferir todos os dados localmente

### Estrutura de rede

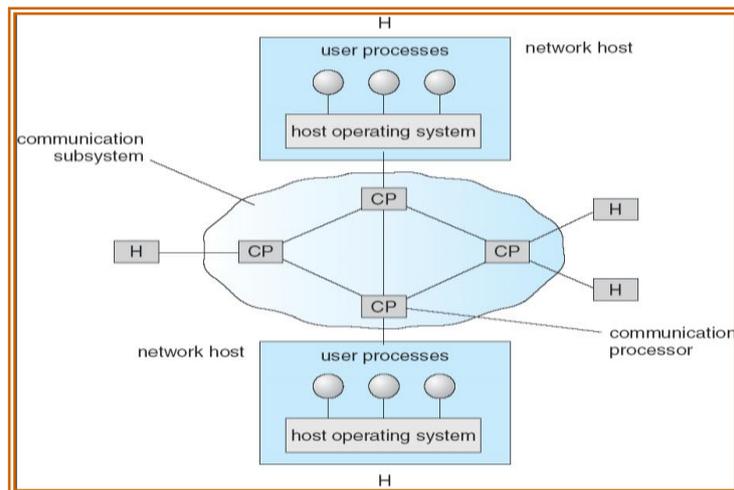
- **Rede local (LAN)** – criada para abranger pequena área geográfica.
  - Rede de barramento multiacesso, anel ou estrela
  - Velocidade  $\approx 10 - 100$  megabits/segundo
  - Broadcast é rápido e barato
  - Nós:
    - normalmente, estações de trabalho e/ou computadores pessoais
    - alguns (normalmente um ou dois) mainframes

### Representação de uma LAN típica



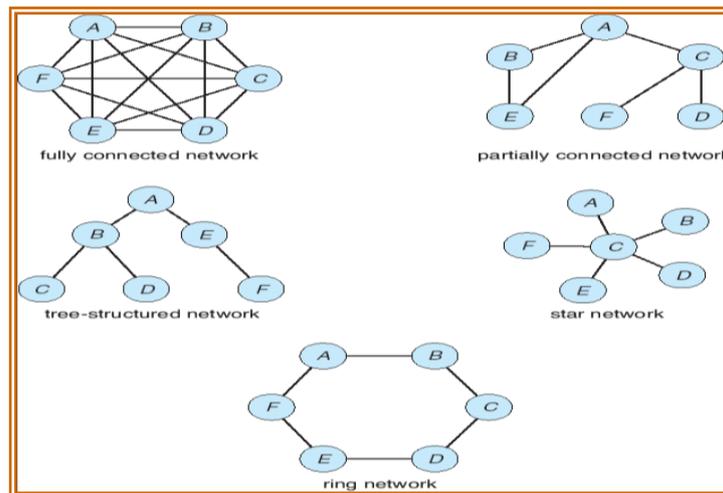
- **Rede remota (WAN) – liga sites separados geograficamente**
  - **Conexões ponto-a-ponto por linhas de longa distância (normalmente, alugadas de uma companhia telefônica)**
  - **Velocidade  $\approx 1,544 - 45$  megabits/segundo**
  - **Broadcast normalmente requer múltiplas mensagens**
  - **Nós:**
    - **normalmente, uma alta porcentagem de mainframes**

### Processadores de comunicação em uma rede remota



### Topologia de rede

- **Os sites no sistema podem ser conectados fisicamente de várias maneiras; eles são comparados com relação aos seguintes critérios:**
  - **Custo básico - Qual o custo de ligar vários sites no sistema?**
  - **Custo de comunicação – Quanto tempo é necessário para levar uma mensagem do site *A* ao site *B*?**
  - **Confiabilidade – Se um enlace ou um site no sistema falhar, os sites restantes ainda podem se comunicar um com o outro?**
- **As diversas topologias são representadas como grafos cujos nós correspondem aos sites**
  - **Uma aresta do nó *A* ao nó *B* corresponde a uma conexão direta entre os dois sites**



## Estrutura de comunicação

O projeto de uma rede de *comunicação* deve focar quatro *questões* básicas:

- **Nomeação e tradução de nomes – Como dois processos localizam um ao outro para se comunicar?**
  - **Sistemas de nomes na rede**
  - **Endereça mensagens com o process-id**
  - **Identifica processos nos sistemas remotos com par <nome-host, identificador>**
  - **Serviço de nome de domínio (DNS) – especifica a estrutura de nomeação dos hosts, além de tradução de nome para endereço (Internet)**
- **Estratégias de roteamento – Como as mensagens são enviadas pela rede?**
  - **Roteamento fixo – Um caminho de A para B é especificado em avanço; o caminho só muda se uma falha de hardware o desativar**
  - **Como o caminho mais curto normalmente é escolhido, os custos de comunicação são minimizados**
  - **Roteamento fixo não pode se adaptar a mudanças de carga**
  - **Garante que as mensagens serão entregues na ordem em que foram enviadas**

- **Circuito virtual** - Um caminho de *A* para *B* é fixado pela duração de uma sessão. Diferentes sessões envolvendo mensagens de *A* para *B* podem ter caminhos diferentes.
- **Remédio parcial para adaptar a mudanças de carga**
- **Garante que as mensagens serão entregues na ordem em que foram enviadas**
- **Roteamento dinâmico** – O caminho usado para enviar uma mensagem do site *A* para o site *B* é escolhido apenas quando uma mensagem é enviada
- **Normalmente, um site envia uma mensagem para outro site no enlace menos usado nesse momento em particular**
- **Adapta-se a mudanças de carga evitando mensagens de roteamento no caminho muito usado**
- **Mensagens podem chegar fora de ordem**
- **Esse problema pode ser remediado acrescentando-se um número de seqüência a cada mensagem**
- **Estratégias de conexão** – Como dois processos enviam uma seqüência de mensagens?
  - **Comutação de circuitos** – Um enlace físico permanente é estabelecido pela duração da comunicação (por exemplo, sistema telefônico)
  - **Comutação de mensagem** – Um enlace temporário é estabelecido pela duração de uma transferência de mensagem (por exemplo, sistema de correios)
  - **Comutação de pacotes** – Mensagens de tamanho variável são dividida em pacotes de tamanho fixo, que são enviados ao destino
  - **Cada pacote pode tomar um caminho diferente na rede**
  - **Os pacotes devem ser remontados em mensagens à medida que chegam**
  - **Comutação de circuitos requer tempo de preparação, mas gera menos overhead para entrega de cada mensagem, e pode desperdiçar largura de banda de rede**

- **Comutação de mensagem e pacote requer menos tempo de preparação, mas gera mais overhead por mensagem**
- **Contenção – A rede é um recurso compartilhado, logo, como resolvemos demandas em conflito pelo seu uso?**
  - **Vários sites podem querer transmitir informações por um enlace ao mesmo tempo. Algumas técnicas para evitar colisões repetidas são:**
    - **CSMA/CD - Carrier sense with multiple access (CSMA); collision detection (CD)**
    - **Um site determina se outra mensagem está sendo atualmente transmitida por esse enlace. Se dois ou mais sites começarem a transmitir exatamente ao mesmo tempo, então eles registrarão uma CD e pararão de transmitir**
    - **Quando o sistema estiver muito ocupado, muitas colisões podem ocorrer, e o desempenho pode ser degradado**
    - **CSMA/CD é usado com sucesso no sistema Ethernet, o sistema de rede mais comum**
    - **Passagem de tokens – Um tipo de mensagem exclusivo, conhecido como token, circula continuamente no sistema (normalmente, uma estrutura de anel)**
    - **Um site que deseja transmitir informações deve esperar até que o token chegue**
    - **Quando o site termina sua rodada de passagem de mensagens, ele retransmite o token**
    - **Um esquema de passagem de tokens é usado por alguns sistemas IBM e HP/Apollo**
    - **Slots de mensagens – Uma série de slots de mensagens de tamanho fixo circula no sistema (normalmente, uma estrutura de anel)**

- ❑ Como um slot só pode conter mensagens de tamanho fixo, uma única mensagem lógica pode ter que ser desmembrada em diversos pacotes menores, cada um enviado em um slot separado
- ❑ Esse esquema tem sido adotado no Cambridge Digital Communication Ring experimental

### Protocolo de comunicação

A rede de comunicação é dividida nas seguintes camadas múltiplas:

- **Camada física** – trata dos detalhes mecânicos e elétricos da transmissão física de um stream de bits
- **Camada de enlace de dados** – trata dos *frames*, ou partes de tamanho fixo dos pacotes, incluindo qualquer detecção e recuperação de erro que ocorreu na camada física
- **Camada de rede** – fornece conexões e roteia pacotes na rede de comunicação, incluindo o tratamento do endereço dos pacotes de saída, decodificação do endereço dos pacotes que chegam e manutenção de informações de roteamento para a resposta apropriada em níveis de carga variáveis
- **Camada de transporte** – responsável pelo acesso em baixo nível para a rede e pela transferência de mensagens entre os clientes, incluindo o desmembramento de mensagens em pacotes, manutenção da ordem dos pacotes, controle de fluxo e geração de endereços físicos
- **Camada de sessão** – implementa sessões, ou protocolos de comunicação de processo a processo
- **Camada de apresentação** – resolve diferenças nos formatos entre as diversas instalações na rede, incluindo conversões de caracteres e modos half-duplex e full-duplex (eco de caracteres)
- **Camada de aplicação** – interage diretamente com os usuários, trata de transferência de arquivo, protocolos de login remoto e correio eletrônico, bem como esquemas para bancos de dados distribuídos

## Robustez

- **Detecção de falha**
  - A detecção de falha do hardware é difícil
  - Para detectar uma falha de enlace, pode-se usar um protocolo de handshaking
  - Suponha que o site A e o site B tenham estabelecido um enlace
  - Em intervalos fixos, cada site trocará uma mensagem *estou-ativo*, indicando que estão ativos e funcionando
  - Se o site A não receber uma mensagem dentro do intervalo fixo, ele considera que (a) o outro site não está ativo ou (b) a mensagem se perdeu
  - O site A pode agora enviar uma mensagem *você está ativo?* ao site B
  - Se o site A não receber uma resposta, ele pode repetir a mensagem ou tentar uma rota alternativa ao site B
  - Se o site A por fim não receber uma resposta do site B, ele conclui que houve algum tipo de falha
  - Tipos de falhas: Site B parado
    - O enlace direto entre A e B está interrompido
    - O enlace alternativo de A para B está interrompido
    - A mensagem foi perdida
  - Porém, o site A não pode determinar exatamente por que a falha ocorreu

## Reconfiguração

- Quando o site A determina que uma falha ocorreu, ele deve reconfigurar o sistema:
- Se o enlace de A para B tiver falhado, isso precisa ser transmitido por broadcast a cada site no sistema
- Se um site tiver falhado, cada outro site também deve ser notificado, indicando que os serviços oferecidos pelo site que falhou não estão mais disponíveis

- Quando o enlace ou o site se tornarem disponíveis novamente, essa informação deve novamente ser transmitida por broadcast a todos os outros sites

#### **Aspectos de projeto**

- **Transparência** – o sistema distribuído deve aparecer como um sistema convencional, centralizado ao usuário
- **Tolerância a falhas** – o sistema distribuído deve continuar a funcionar em caso de falha
- **Escalabilidade** – quando demandas aumentam, o sistema deve aceitar facilmente o acréscimo de novos recursos para acomodar a maior demanda
- **Clusters** – uma coleção de máquinas semi-autônomas que atuam como um único sistema.

[http://cassiopea.ipt.br/teses/2005\\_EC\\_Fabio\\_Vieira\\_Amaral.pdf](http://cassiopea.ipt.br/teses/2005_EC_Fabio_Vieira_Amaral.pdf)