

GPRT/UFPE



GPRT/UFPE

Peer-to-Peer (P2P)

Computação Colaborativa na Internet
Minicurso SBRC - (alterado por Andrea Krob)

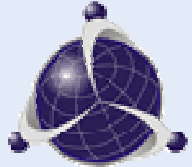


Autores Originais (UFPE)

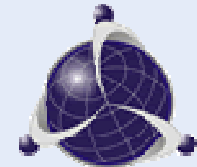
- João Rocha
- Marco Antônio Domingues
- Arthur Callado
- Eduardo Souto
- Guthemberg Silvestre
- Carlos Kamienski
- Djamel Sadok

Roteiro

- Introdução / Motivação
- Aplicações P2P
- Questões de Projeto
- Tecnologias e Soluções
- Considerações Finais



GPRT/UFPE



GPRT/UFPE

Introdução

Peer-to-Peer (P2P) - Computação Colaborativa na Internet

Cliente/Servidor

- Modelo mais usado atualmente na Internet
- Depende de servidores bem configurados e com informação acessível
- Não explora o potencial de computação distribuída proveniente da Rede
- A existência de um ou milhares de computadores é indiferente na **interação** de um usuário
- PCs clientes com capacidade razoável ficam "escondidos" dentro da rede

Definição P2P (1)

Quando temos o compartilhamento de recursos e serviços computacionais ocorrendo diretamente entre sistemas

Definição (2)

- É uma classe de aplicações que leva vantagem de recursos disponíveis nas **bordas** da Internet
- Quais recursos?
 - Armazenamento
 - Tempo de CPU
 - Conteúdo
 - etc.;

Características

- Sistemas distribuídos sem controle centralizado ou organização hierárquica
- Software executado em cada elemento (nó / peer) é equivalente em funcionalidade
- Têm recebido grande atenção:
 - Usuários
 - Empresas
 - Meios acadêmicos (pesquisas)

Características

- Cada participante age como cliente e servidor ao mesmo tempo (**servent**)
- Cada cliente "paga" a sua participação fornecendo acesso a (alguns de) seus recursos
 - Exceções são comuns: freeloaders ou vampiros

Características

- Sem coordenação totalmente centralizada
- Sem local único de falha ou gargalo
- Nenhum ponto tem visão global do sistema
- Comportamento global definido por interações locais
- Os dados e serviços são acessíveis de qualquer ponto
- Os pontos são autônomos
- As conexões são heterogêneas e não são confiáveis

Principais Vantagens

- Escalabilidade
 - Não há gargalo para crescimento
- Robustez
 - Não há ponto de falha único
- Flexibilidade
 - Auto-configuração / configuração dinâmica

Teste P2P

- O sistema aceita conectividade variável e endereços IP temporários?
- O sistema dá uma autonomia significativa aos computadores na borda da rede?
- Os nós podem trocar informações diretas entre si?

Requisitos dos Sistemas P2P

- Descoberta de recursos/serviços
- Roteamento
- Robustez e tolerância a falhas
- Armazenamento distribuído
- Escalabilidade
- Confiança nos pares (autenticação, etc.)
- Monitoramento de vizinhos

Modelos de Sistemas P2P

(Classificação 1/3: Índice)

- Modelo Centralizado
 - Índice global mantido por um autoridade central
 - Exemplo: Napster
- Modelo Descentralizado
 - Sem índice global (sem coordenação global)
 - Exemplos: Gnutella, Freenet
- Modelo Hierárquico
 - Introdução dos super-nós (super-nodes ou super-peers)
 - Mistura dos modelos centralizado e descentralizado
 - Exemplos: KaZaA, Morpheus

Modelos de Sistemas P2P

(Classificação 2/3: Busca)

- Centralized Service Location (CSL)
 - Busca centralizada
 - Exemplo: Napster
- Flooding-based Service Location (FSL)
 - Busca baseada em inundação
 - Exemplo: Gnutella
- Hash Table-based Service Location (DHT)
 - Busca baseada em tabela de hash distribuída
 - Exemplos: CAN, Pastry, Tapestry, Chord

Modelos de Sistemas P2P

(Classificação 3/3: Modelo)

- Modelo Centralizado
 - Napster, mensagens (ICQ, etc)
- Modelo Descentralizado e Estruturado
 - DHT
 - Chord, etc.
- Modelo Descentralizado e Não Estruturado
 - Super-Nós: KaZaA
 - Inundação: Gnutella

Aplicação: Troca de Mensagens

- IM (Instant Messaging)
- Aplicação popular na Internet, pela facilidade de enviar mensagens on-line
- Exemplos:
 - MSN Messenger (<http://messenger.msn.com>)
 - Yahoo! Messenger (<http://messenger.yahoo.com>)
 - Trillian Messenger – (<http://www.trillian.de>)

IM: Yahoo! Messenger e Trillian



Aplicação: Compartilhamento de Arquivos

- Aplicação de maior sucesso na Internet
- Permite usuários compartilharem diretamente seus arquivos, músicas, etc.
- Problema: pode apresentar problemas de violação de direitos autorais
- Características
 - Área de armazenamento
 - Disponibilidade de informações
 - Anonimato

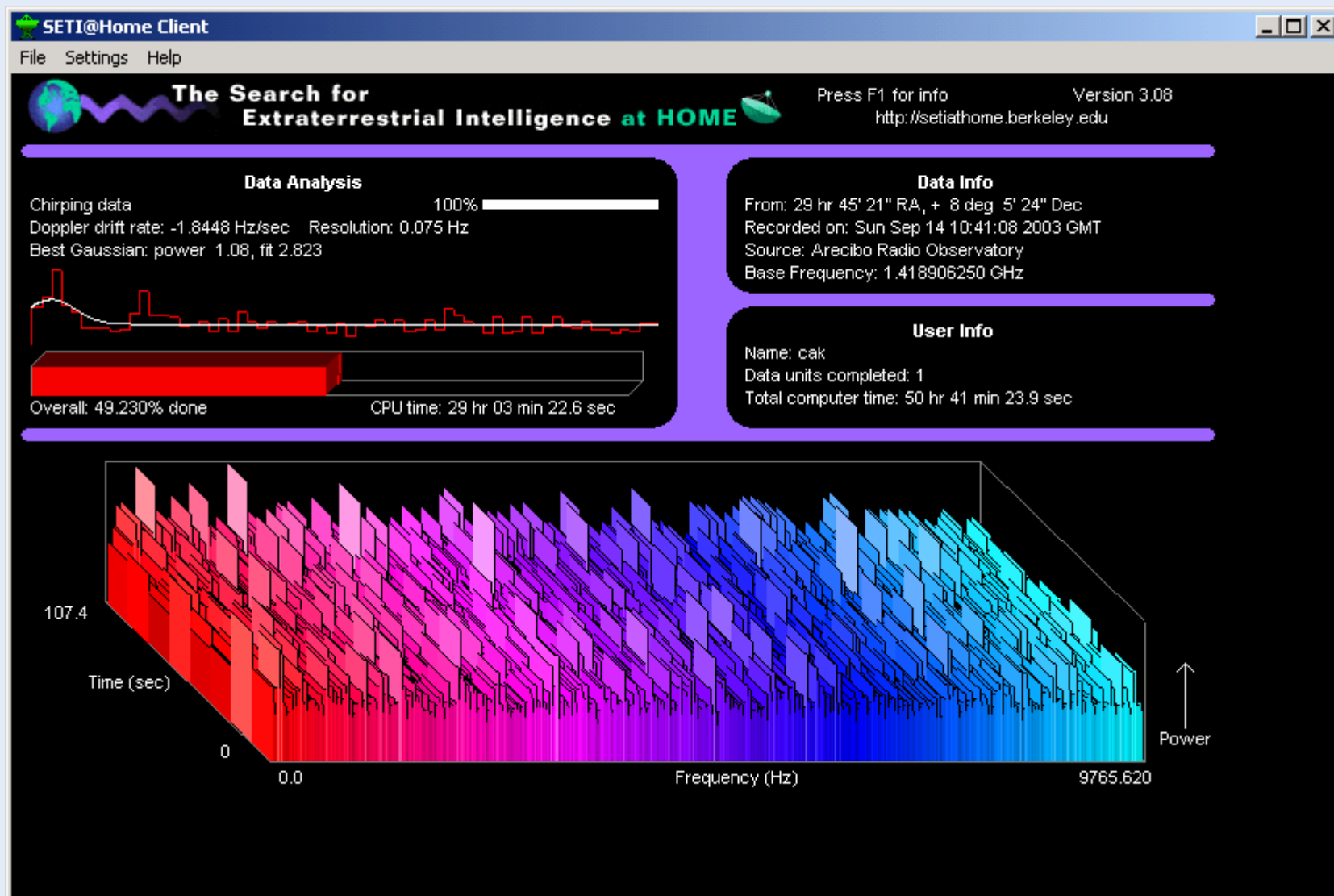
Aplicação: Compartilhamento de Arquivos

- Napster (<http://www.napster.com>)
- KaZaA (<http://www.kazaa.com>)
- Gnutella (<http://www.gnutella.com>)
- Freenet (<http://www.freenetproject.org>)
- Imesh (<http://www.imesh.com>)
- Morpheus (<http://www.morpheus.com>)
- Grokster (<http://www.grokster.com>)

Aplicações: Computação Distribuída

- A idéia de aproveitar recursos computacionais ociosos não é nova
- Grade Computacional (Grid)
 - Solução de computação distribuída baseada no compartilhamento de recursos em larga escala
 - Não é igual, mas semelhante ao P2P
- SETI@Home (<http://setiathome.ssl.berkeley.edu>)
 - The Search for Extraterrestrial Intelligence
 - Usuários executam partes da "busca"

SETI@Home



Aplicações: Trabalho Colaborativo

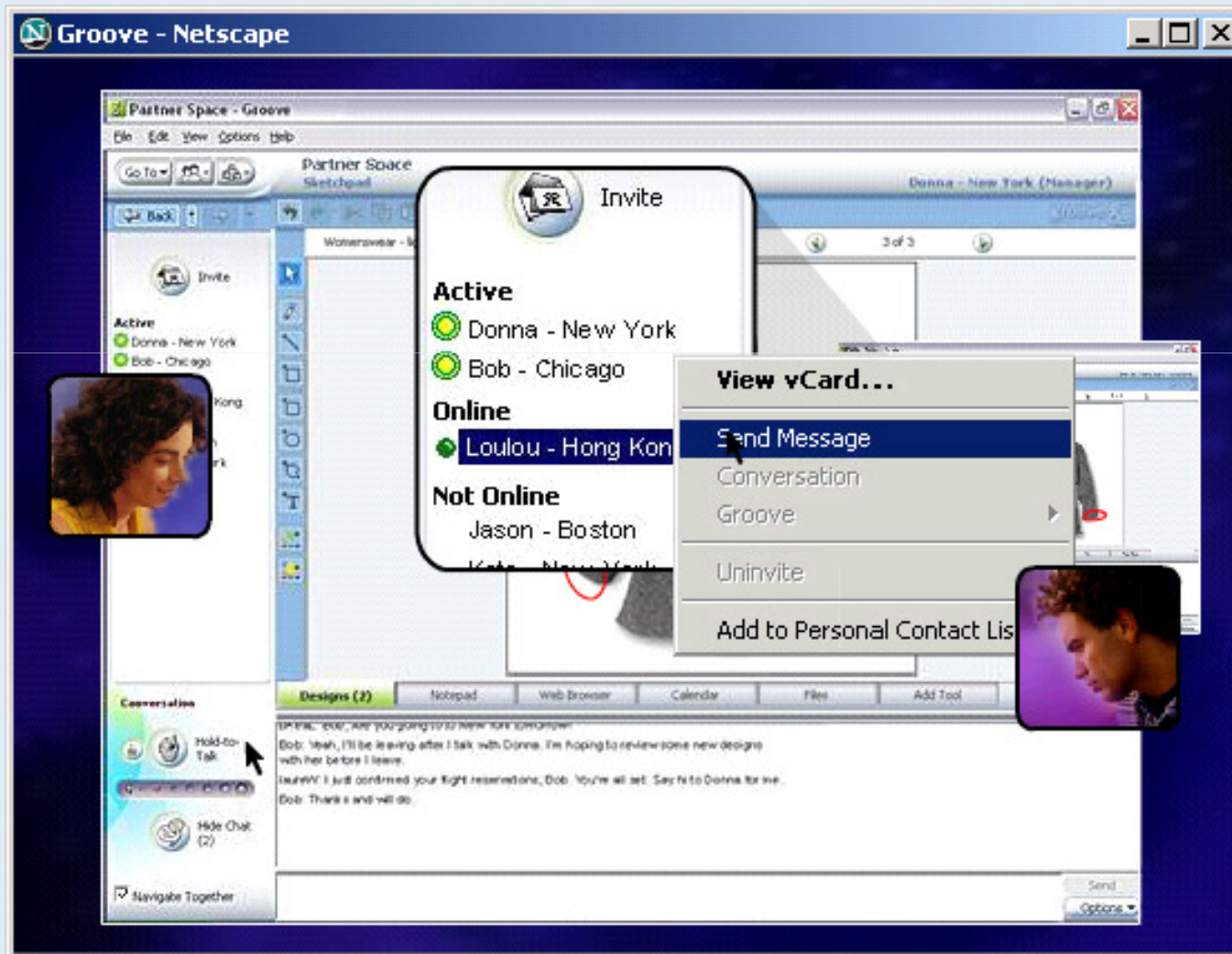
- Groupware

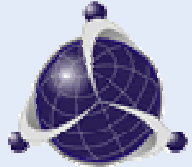
- Software que suporta colaboração, a comunicação e coordenação de vários usuários em uma rede
- Exemplo Cliente/servidor: Lotus Notes / Sharepoint

- Se combinarmos groupware com P2P teremos novas oportunidades, como o Groove (<http://www.groove.net>)

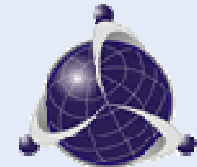
- Software que permite criar/compartilhar espaços de trabalho
- Compartilha arquivos e projetos entre várias equipes
- Permite o uso de servidores centralizados

Groupware: Groove





GPRT/UFPE



GPRT/UFPE

Questões de Projeto

Peer-to-Peer (P2P) - Computação Colaborativa na Internet

Questão 1: Endereçamento

- Comunicação P2P necessita de conexões diretas entre os *peers*, porém diversas barreiras de proteção podem impedir essa comunicação:
 - **Firewall**: bloqueia a comunicação de entrada/saída da rede, de acordo com critérios de segurança
 - **NAT**: permite apenas tráfego a partir de um IP externo se ele tiver sido gerado a partir da rede interna
 - **Proxy**: interpõem-se na comunicação fim a fim (http) para filtrar páginas indesejáveis (bloqueio)

Questão 2: Conectividade

- Heterogeneidade das conexões dos *peers*
 - Muitas variedades de conexões (dial-up, banda larga, redes acadêmicas, corporativas, etc.)
 - Assimetria nas bandas de upload e download
- Muitos *peers* com conexões de baixa capacidade e alta instabilidade
- Poucos *peers* com conexões de alta capacidade e baixa instabilidade

Questão 3: Escalabilidade

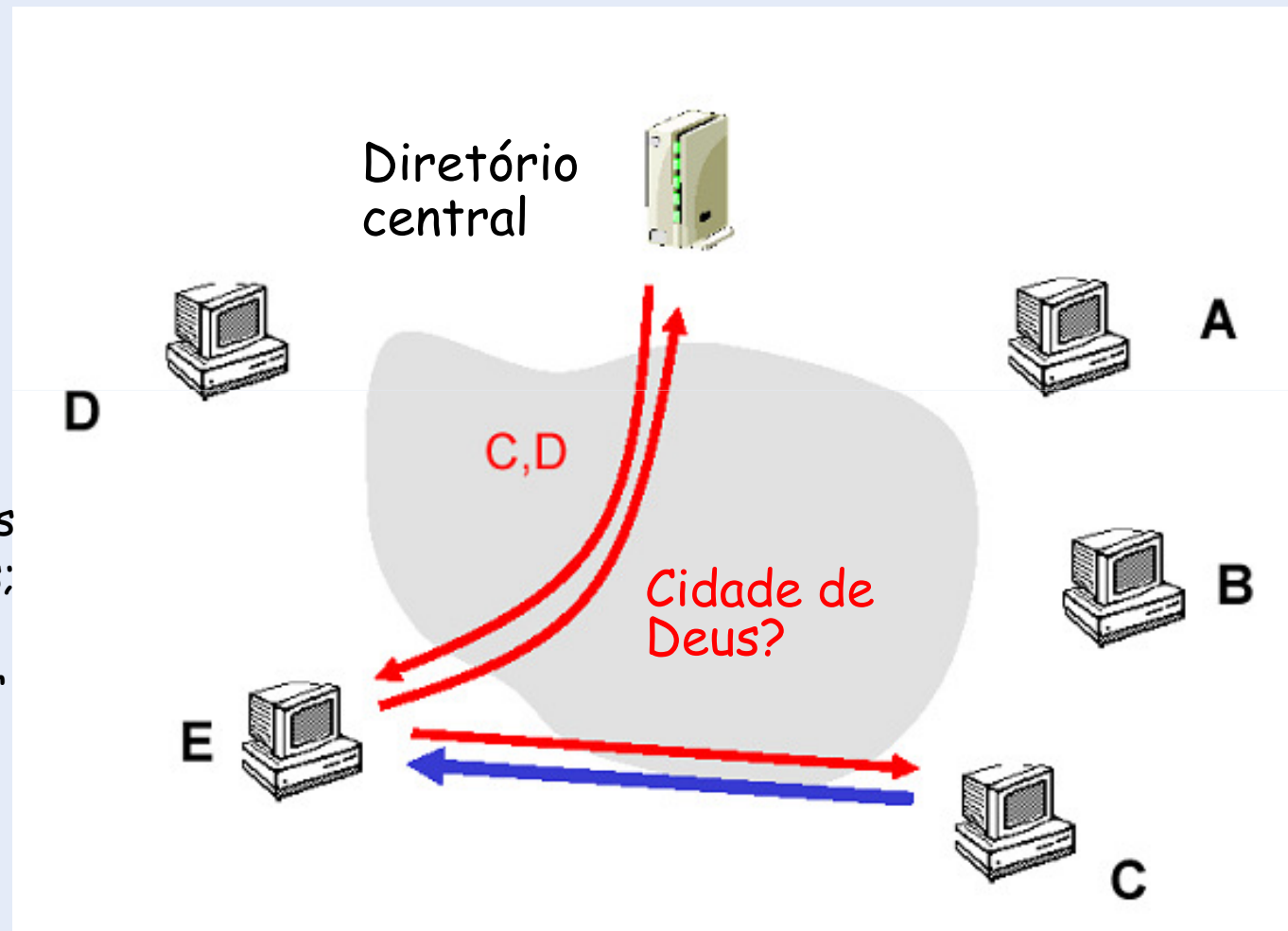
- Benefício imediato da descentralização
- P2P é mais escalável que cliente/servidor
 - Em um sistema cliente-servidor, os servidores são os únicos responsáveis por toda a carga do sistema
 - Em um sistema P2P, o número de servidores aumenta com o número de clientes
- Problemas de escalabilidade em P2P
 - Quantidade de operações centralizadas (ex.: Napster)

Questão 4: Roteamento

- Localizar uma informação em redes grandes, volátil e distribuída não é simples
- A localização e busca de informações dependem do roteamento utilizado
- Abordagens comuns:
 - Modelo Centralizado
 - Modelo por Inundação
 - Modelo de Super-Nós
 - Modelo DHT

Modelo Centralizado

- Os peers se conectam a um diretório central e publicam o conteúdo a compartilhar;
- Ao receber uma requisição, o diretório escolhe os nós mais adequados;
- Exige um servidor central para gerenciamento;
- Ex: Napster

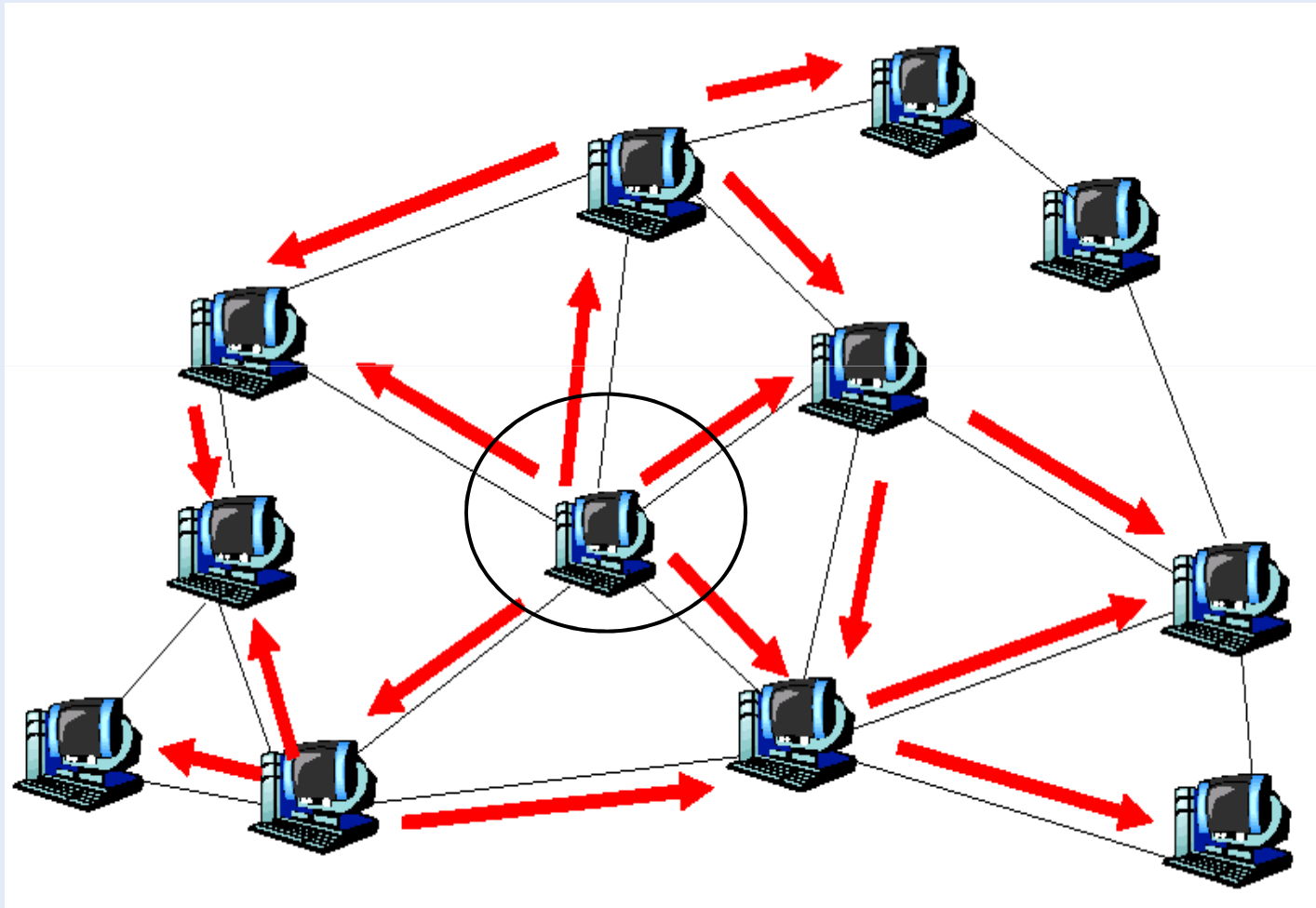


Modelo por Inundação

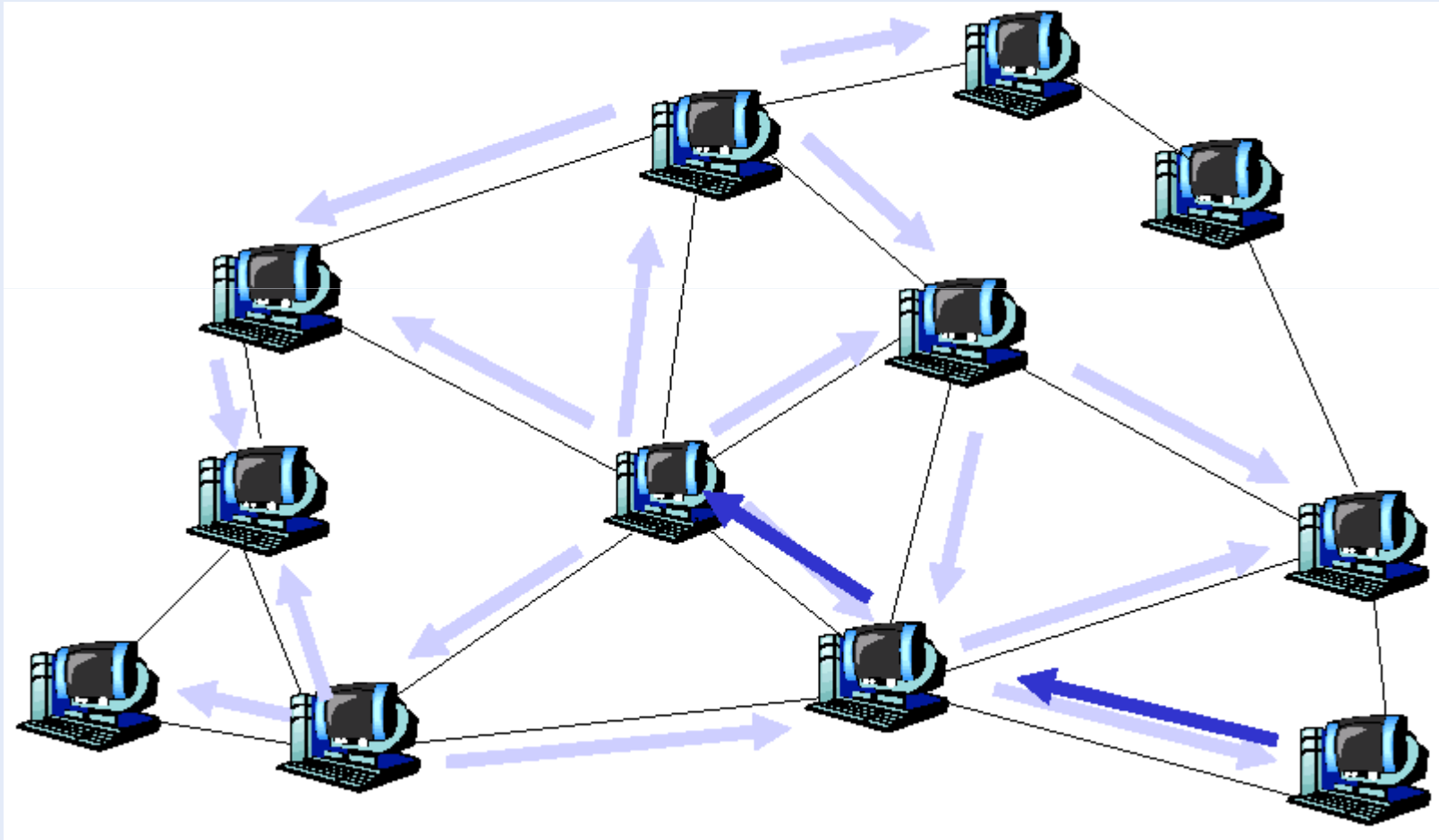
- Cada requisição é enviada para todos os peers diretamente conectados e assim por diante;

- Até que ocorra um número máx. (5 a 9) encaminhamentos ou a resposta seja recebida;

Ex: Gnutella



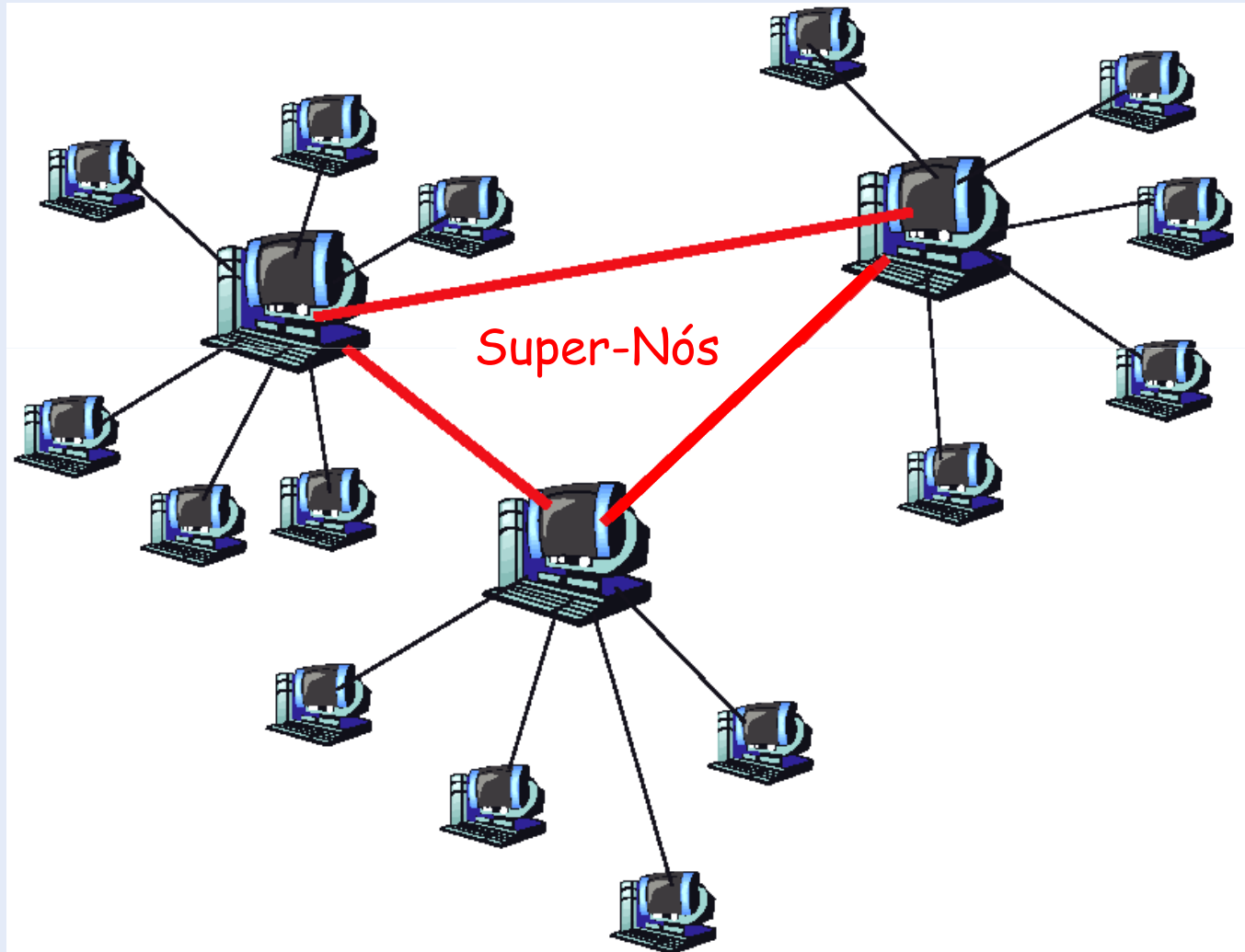
Modelo por Inundação



Modelo de Super-Nós

- É uma mistura dos anteriores (modelo hierárquico);
- Utiliza clientes "supernós" que concentram várias requisições;
- O armazenamento de pesquisas recentes (caching) é usado para melhorar a escalabilidade;

Ex: Kaaza



Modelo DHT

- É o modelo mais recente. Os nós usam uma tabela *hash* para separar o espaço de busca entre eles;
- **O que é Hash mesmo?**
É uma estrutura de dados que associa chaves de pesquisa a valores. Seu objetivo é, a partir de uma chave simples, realizar uma busca rápida e obter o valor desejado;
- Exemplos: Chord, CAN, Tapestry, Pastry, ...

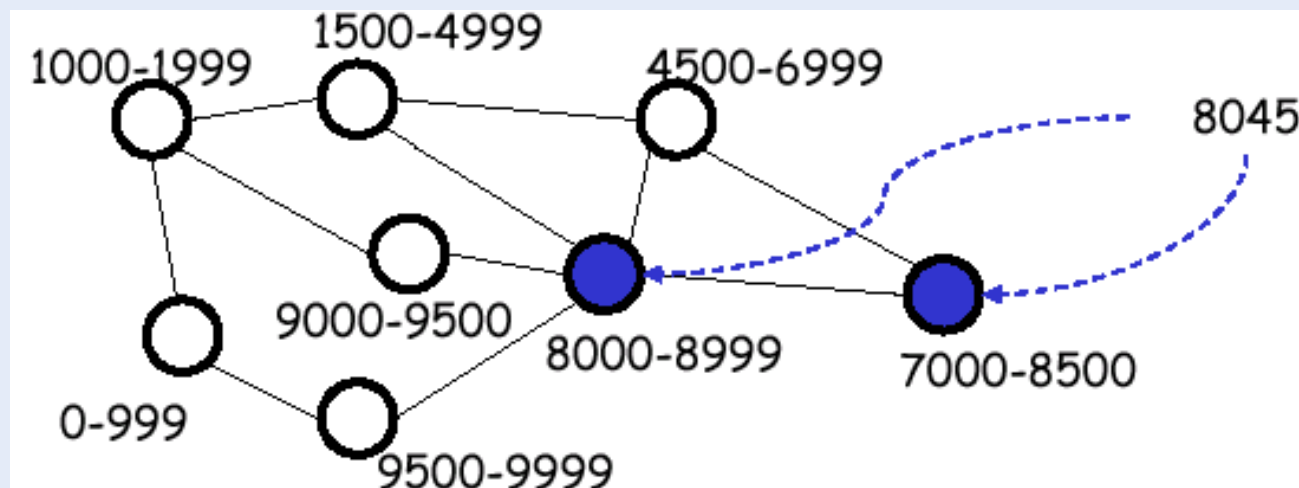
DHT: Funcionamento Básico

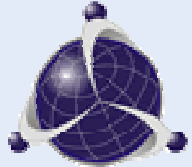
- Cada nó possui um ID associado a ele;
- Quando um documento é compartilhado, um ID único também é associado a ele (*hash* conteúdo + nome);
- O documento é encaminhado até o nó com ID mais próximo ao ID do documento;
- Quando houver uma requisição, ela irá direto para o nó com ID mais semelhante ao documento, descobrindo a localização dele.

DHT: Funcionamento

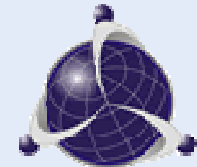
- Na prática, os nós possuem uma faixa de *hash* e conhecem os objetos que ali estão inseridos;
 - Os nós podem armazenar tanto **objetos** quanto **apontadores** para objetos na sua faixa de *hash*

Ex publicação de arquivo : $h(\text{"Aquarela do Brasil"}) \rightarrow 8045$





GPRT/UFPE



GPRT/UFPE

Tecnologias e Soluções

Peer-to-Peer (P2P) - Computação Colaborativa na Internet

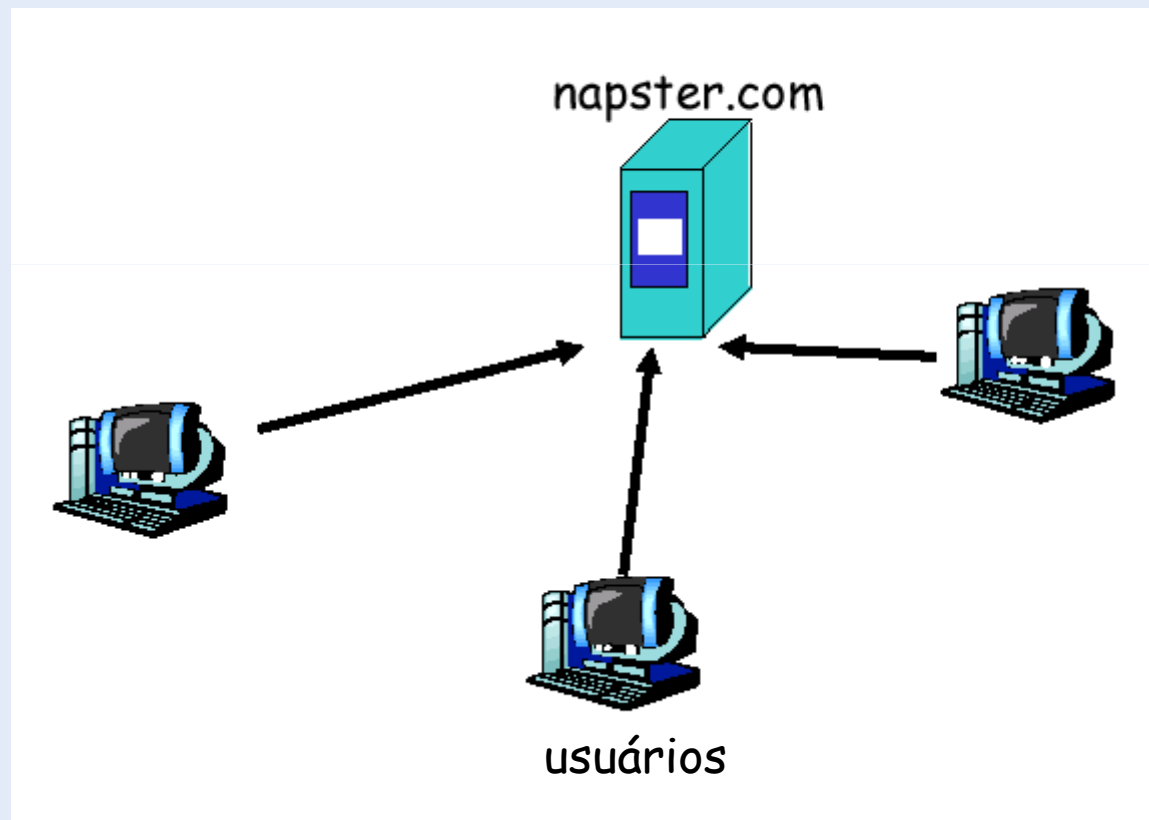
Redes Públicas de Compartilhamento

- Existem algumas redes públicas e vários aplicativos que as acessam
- Vários aplicativos acessam a mesma rede e várias redes podem ser acessadas por um aplicativo
- Algumas Redes P2P:
 - Napster (extinta)
 - Gnutella (BearShare, LimeWire, WinMX, etc.)
 - Freenet
 - Edonkey
 - Fastrack

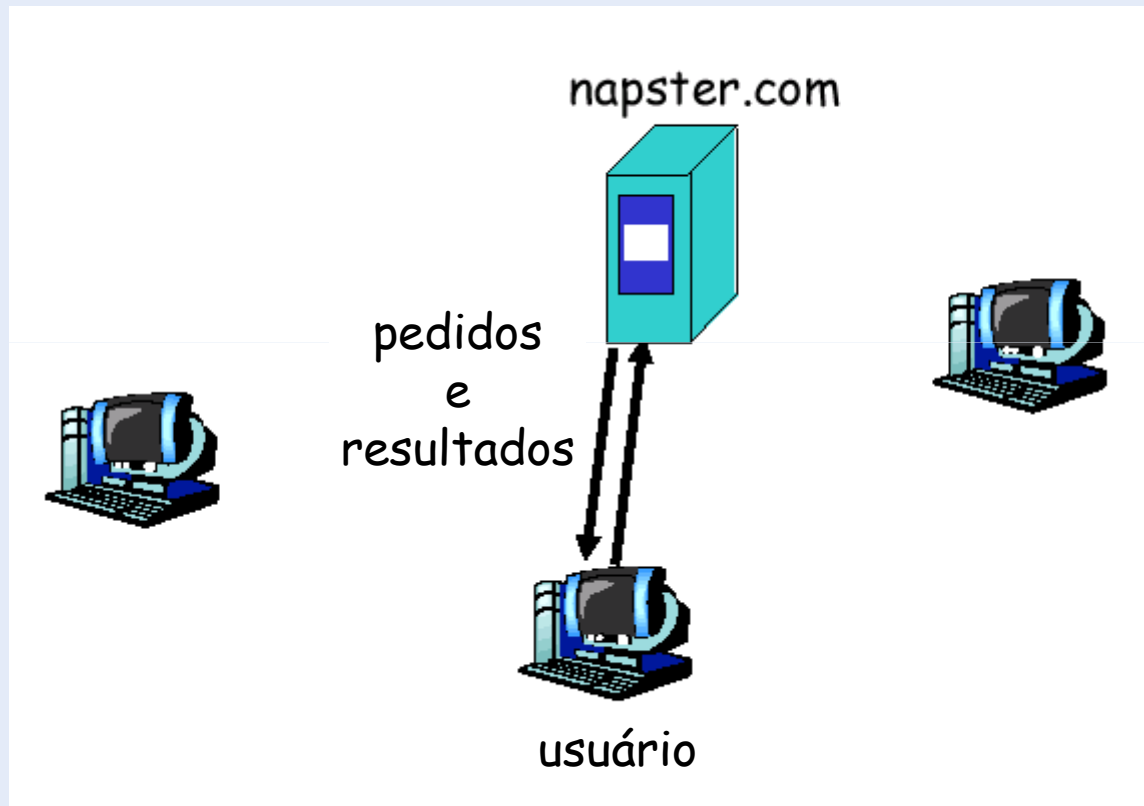
Napster: funcionamento

1. Um cliente se conecta com o servidor e envia a sua lista de arquivos compartilhados
2. O cliente interessado envia palavras-chave para fazer busca na lista completa
3. O cliente testa a taxa de transmissão dos pares que têm o arquivo solicitado (ping)
4. O arquivo é transferido entre os pares

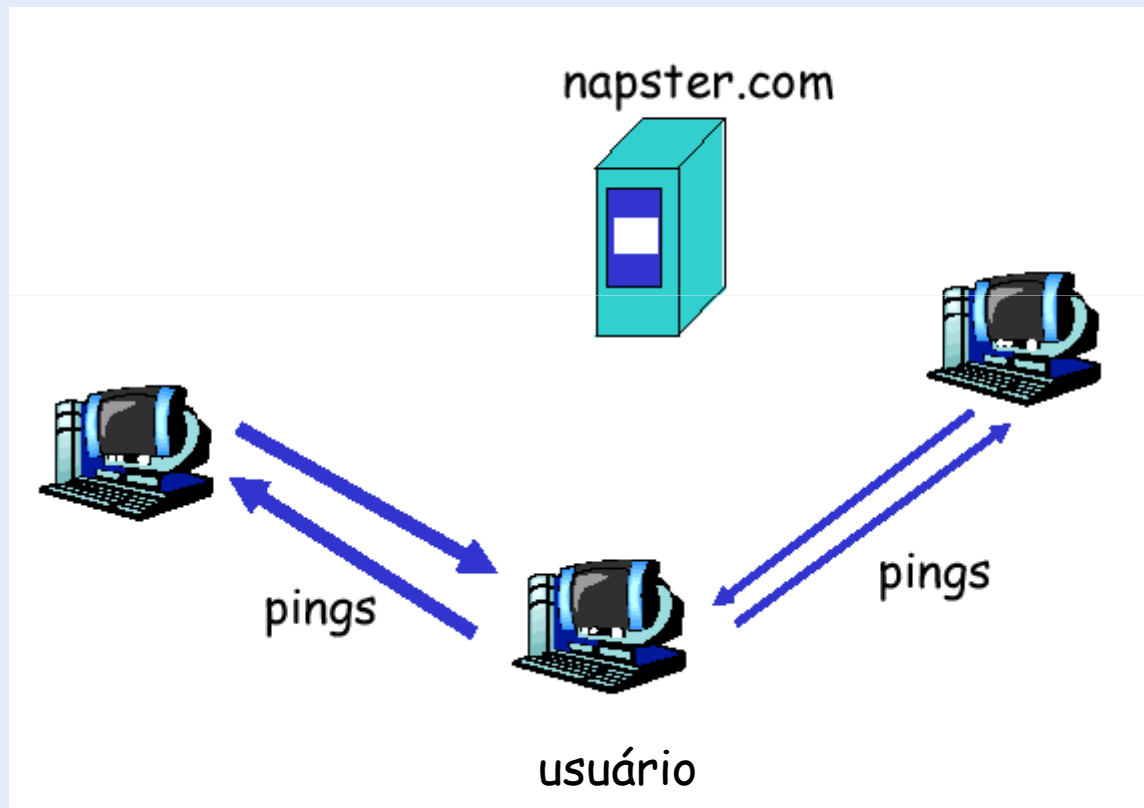
Napster: funcionamento (1)



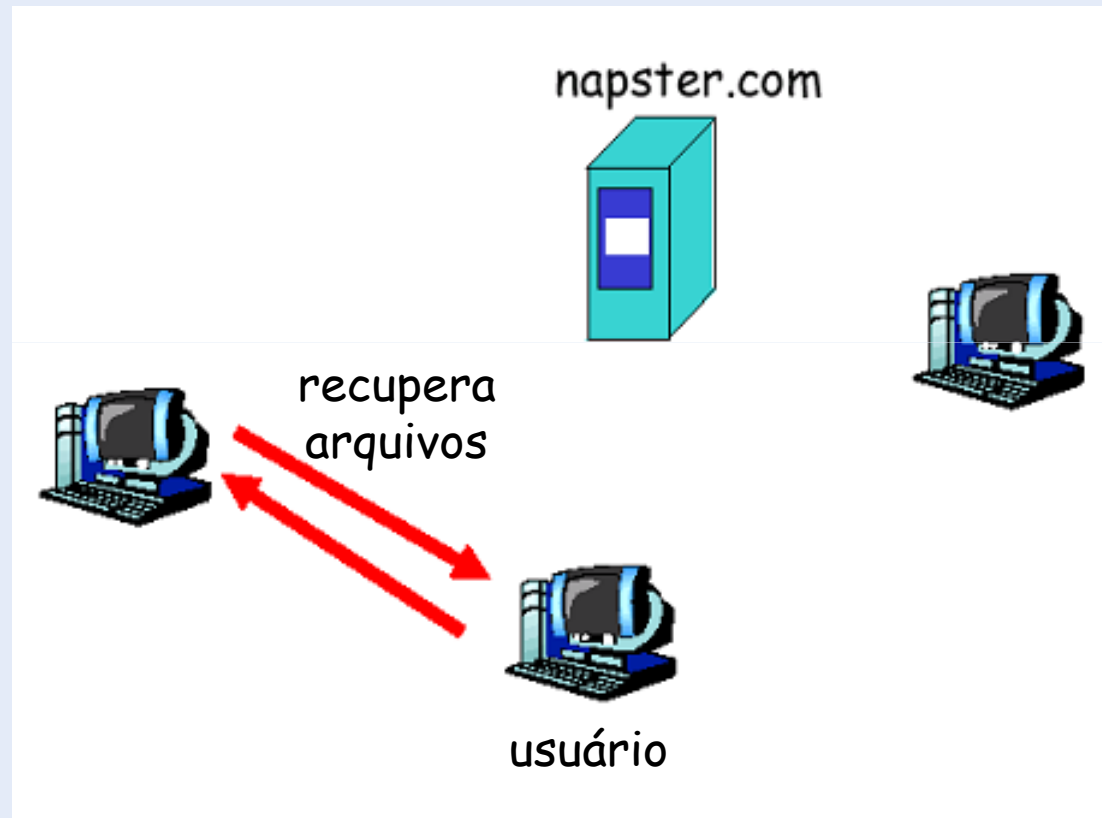
Napster: funcionamento (2)



Napster: funcionamento (3)



Napster: funcionamento (4)



Napster - Comentários

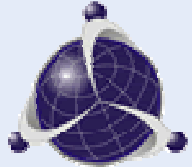
- Servidor centralizado
 - Ponto único de falhas
 - Pode usar o DNS para balancear carga entre servidores
 - Sujeito a congestionamentos
- Nenhuma segurança
 - Senhas enviadas sem criptografia
 - Sem autenticação
 - Sem anonimidade (identidade revelada)

Gnutella

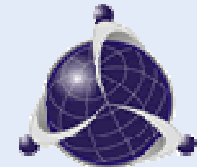
- Sistema de busca totalmente distribuído
- Busca baseada em **inundação**
- História:
 - 14/03/2000: Disponibilizado sob licença pública GNU no servidor web da Nullsoft (pertencente à AOL)
 - Retirado apenas algumas horas depois (decisão judicial)
 - Tarde demais: muitos usuários fizeram download
 - O protocolo Gnutella foi "descoberto" através de engenharia reversa
 - Outros clientes foram disponibilizados e Gnutella começou a se popularizar

Gnutella

- A sua principal característica é a Busca Distribuída
- Problema
 - Tráfego gerado: Escalabilidade
- Despertou grande interesse na comunidade acadêmica
 - Não depende de servidor central



GPRT/UFPE



GPRT/UFPE

Considerações Finais

Peer-to-Peer (P2P) - Computação Colaborativa na Internet

Questões Jurídicas

- Facilidade de compartilhar arquivos ilegalmente, violando direitos autorais
- Napster foi interrompido por decisão judicial
 - Voltou a funcionar, com conteúdo pago
- Anonimato: bom ou ruim?
 - Bom para o usuário (liberdade de expressão)
 - Encobre atitudes ilegais
 - Redes obrigadas a revelar identidade de usuários

Problemas

- Faltam simulações de grande porte
- Sistemas P2P requerem comportamento colaborativo dos usuários
 - *Compensações para usuários que colaboram*
- Diversas questões como:
 - *Disponibilidade das informações*
 - *Informações críticas em nós não confiáveis*

Dúvida

- As redes P2P tem sucesso devido à tecnologia ou ao conteúdo de graça?
- Quais serão as aplicações emergentes para o futuro? O P2P continuará sendo utilizado?

Trabalho Prático

1. Formar duplas e escolher um tópico abaixo de estudo. Preparar uma apresentação para o dia 18/10;
 - Algoritmos de Roteamento (CAN, CHORD, TAPESTRY)
 - Trabalho Colaborativo (Groove, Kimbo, Consilient)
 - Redes Públicas (Gnutella, Freenet, Edonkey, Fasttrack)
 - Mensagens Instantâneas (AIM, MSN, ICQ)
2. Explicar características e o funcionamento.
3. Não é necessário artigo, apenas PPT.



Peer-to-Peer (P2P)

Computação Colaborativa na Internet



Obrigado!

