



P2P e Gnutella

Rodrigo Rebouças de Almeida

rodrigor@dsc.ufcg.edu.br

<http://www.dsc.ufcg.edu.br/~rodrigor>

[Conteúdo]

- O que é P2P?
 - Arquiteturas P2P (Overview)
 - Gnutella
- Problemas com o Gnutella
 - Non-compliance
 - Connectivity
 - Lack of Search Results
 - Download Failure
- Conclusões

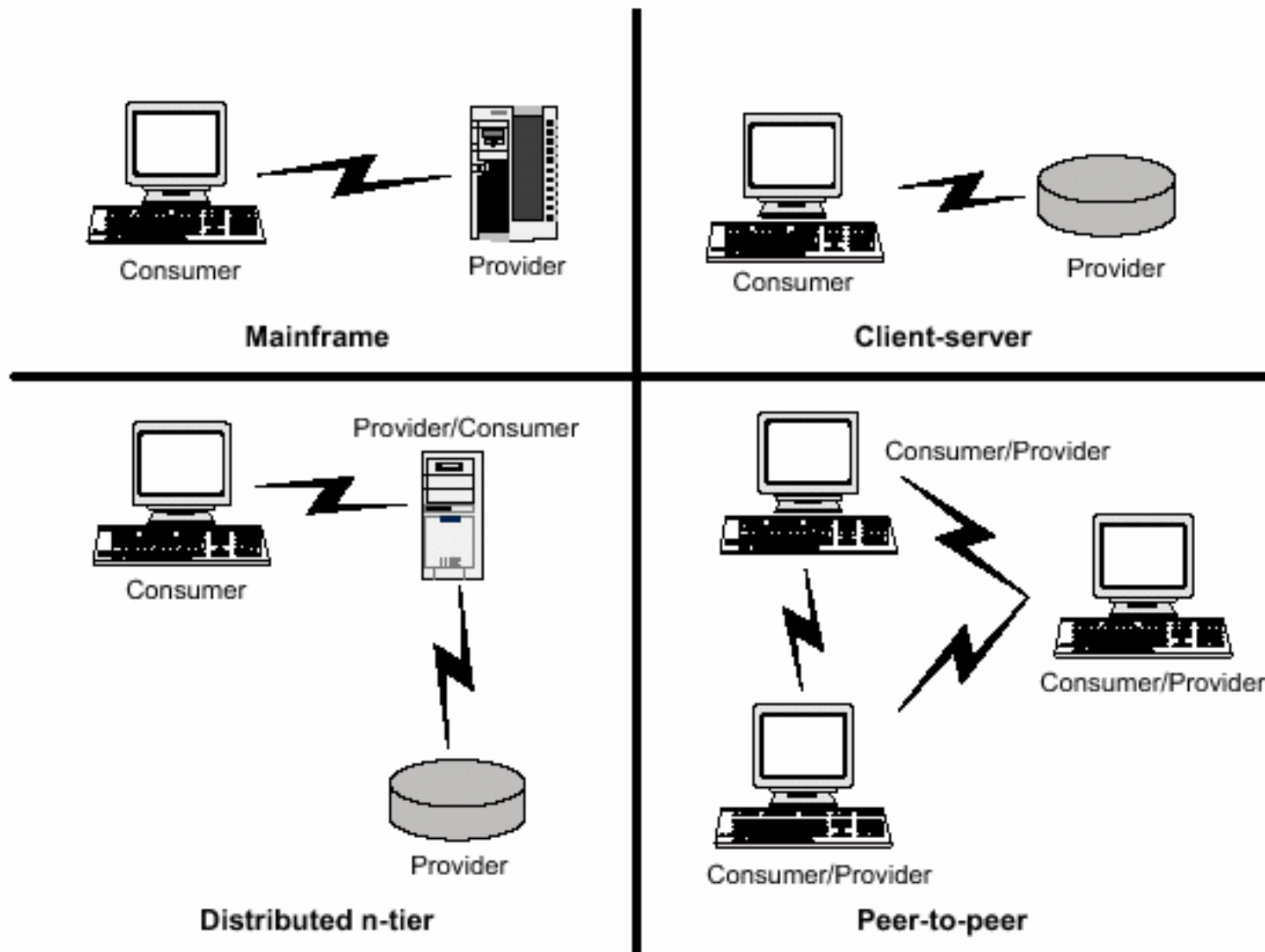
[O que é P2P?]

- Uma classe de aplicações capaz de:
 - Usar recursos presentes em qualquer máquina da rede;
 - Acessar livremente estes recursos de forma descentralizada;
 - Funcionar em um ambiente onde os IPs das máquinas são imprevisíveis e a conexão com as máquinas é instável;

[O que é P2P?]

- Mais uma arquitetura distribuída, como mainframes, cliente-servidor, n-camadas...
- P2P é uma rede colaborativa;
- Um ambiente onde qualquer máquina é capaz de prover e consumir informação;

Arquiteturas de sistemas, exemplos



[P2P ideal (Utopia)]

- Em uma rede P2P ideal:
 - Um host tem acesso aos outros hosts da rede;
 - A rede é vista como um recurso, onde qualquer host pode buscar e armazenar informações de forma irrestrita;
 - Ao acessar uma rede P2P, um host conhece automaticamente os outros hosts e pode fazer buscas por informações;
 - Todos os hosts podem prover e consumir informações da rede;

[Alguns problemas...]

- Firewalls
 - Como conectar-se a máquinas protegidas por firewalls?
- Endereçamento dinâmico
- Localização dos outros clientes
 - Ao entrar na rede, como um host descobre o endereço de outras máquinas da rede?

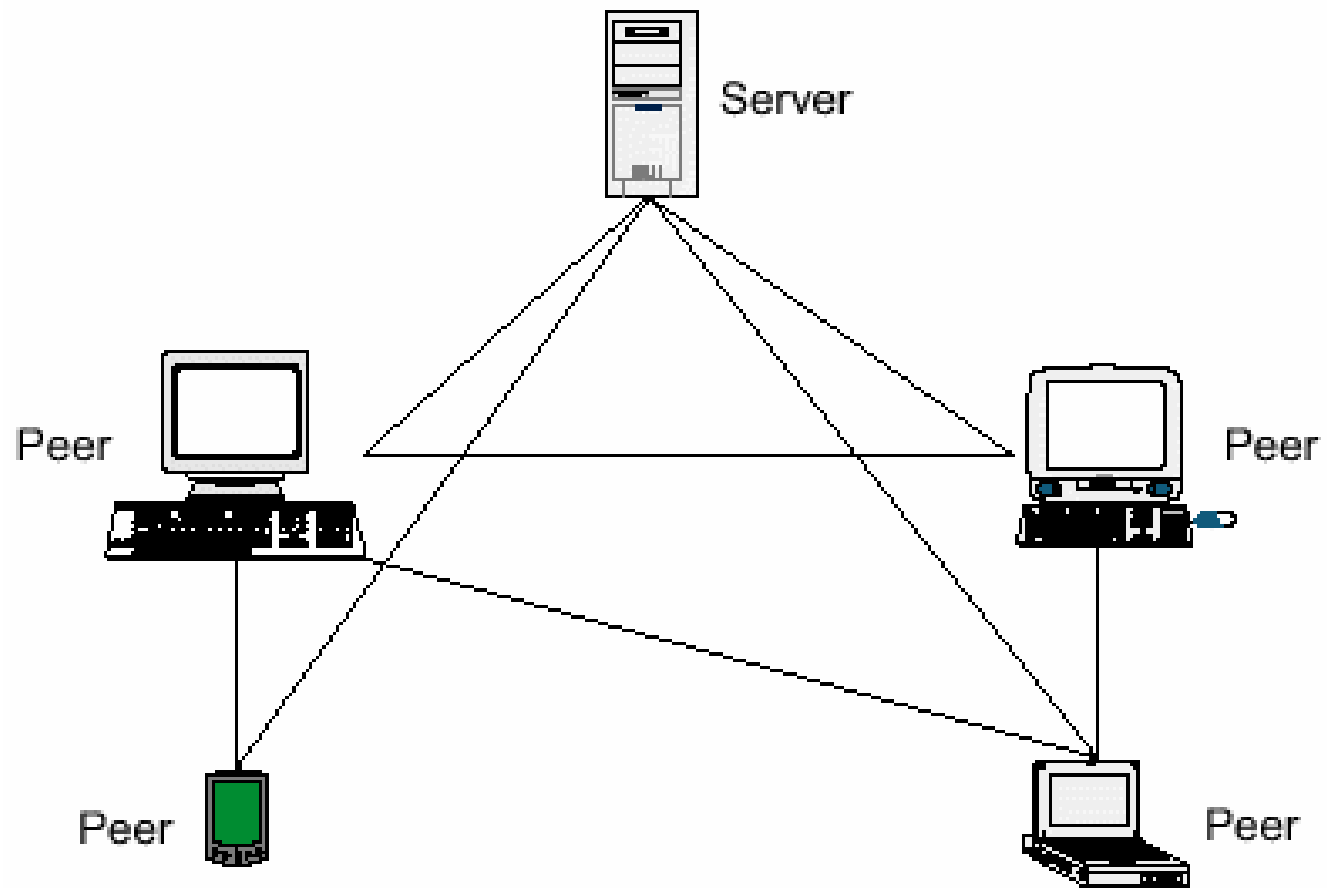
[Arquiteturas P2P]

- Afim de implementar uma rede P2P, existem várias arquiteturas que tentam solucionar o problema, veremos três delas:
 - Arquitetura Centralizada
 - Arquitetura Super P2P
 - Arquitetura P2P Pura

[P2P Centralizada]

- Os hosts conhecem os servidores;
- O “servidor” contém informações sobre os hosts ativos;
- Exemplo: Napster
 - Ao se conectar, o host manda informações sobre todos os arquivos que contém;
 - Os hosts mandam os critérios de busca para o servidor;
 - A conexão “peer-to-peer” ocorre apenas quando um usuário quer fazer download de um arquivo. Esta conexão é feita diretamente ao “peer”;

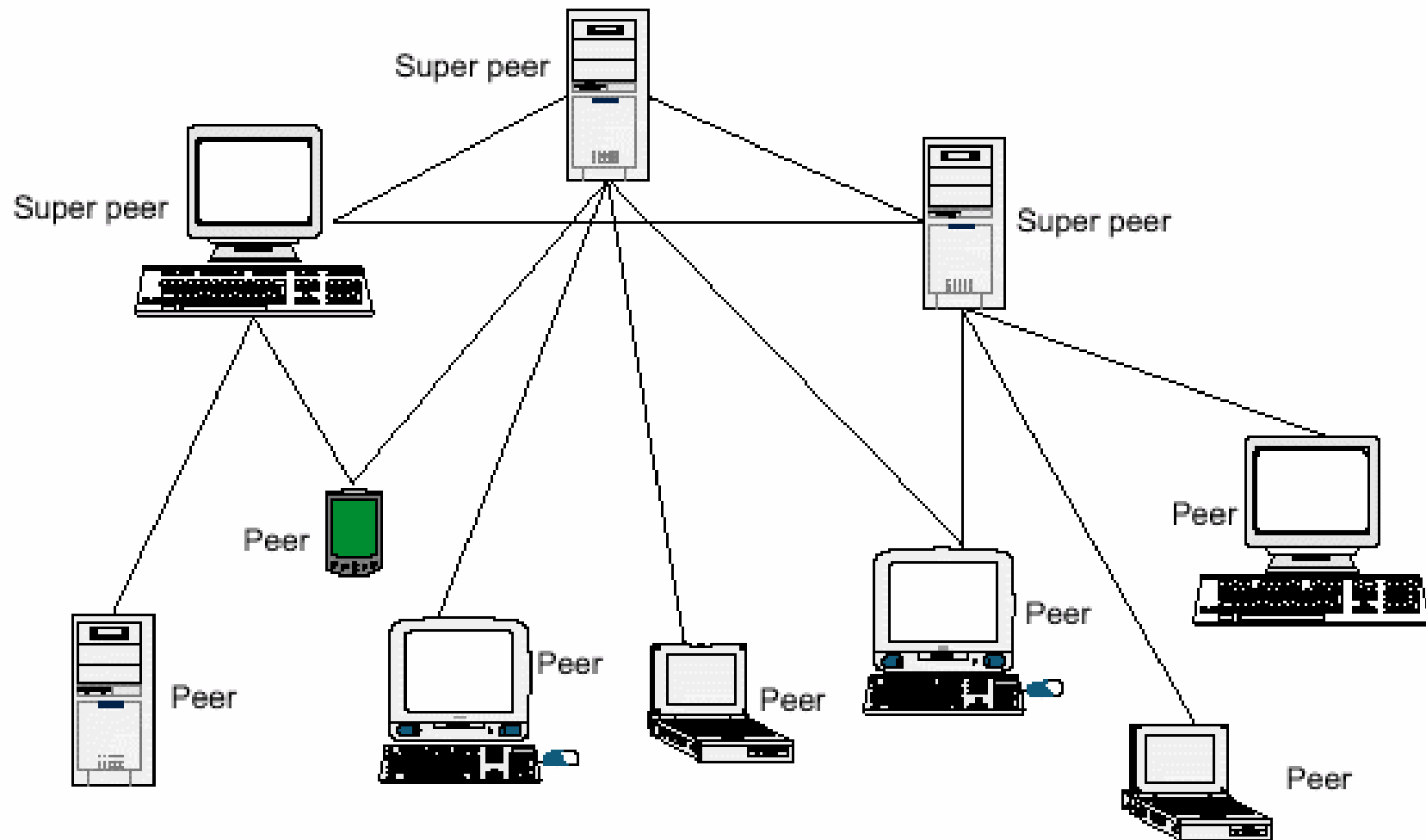
[P2P Centralizada]



[Super P2P]

- É um misto entre a arquitetura centralizada e a arquitetura pura;
- Alguns “peers” são eleitos “super peers”;
- Os “super peers” agem como os servidores na arquitetura centralizada, armazenando informações sobre os “peers” presentes na rede;

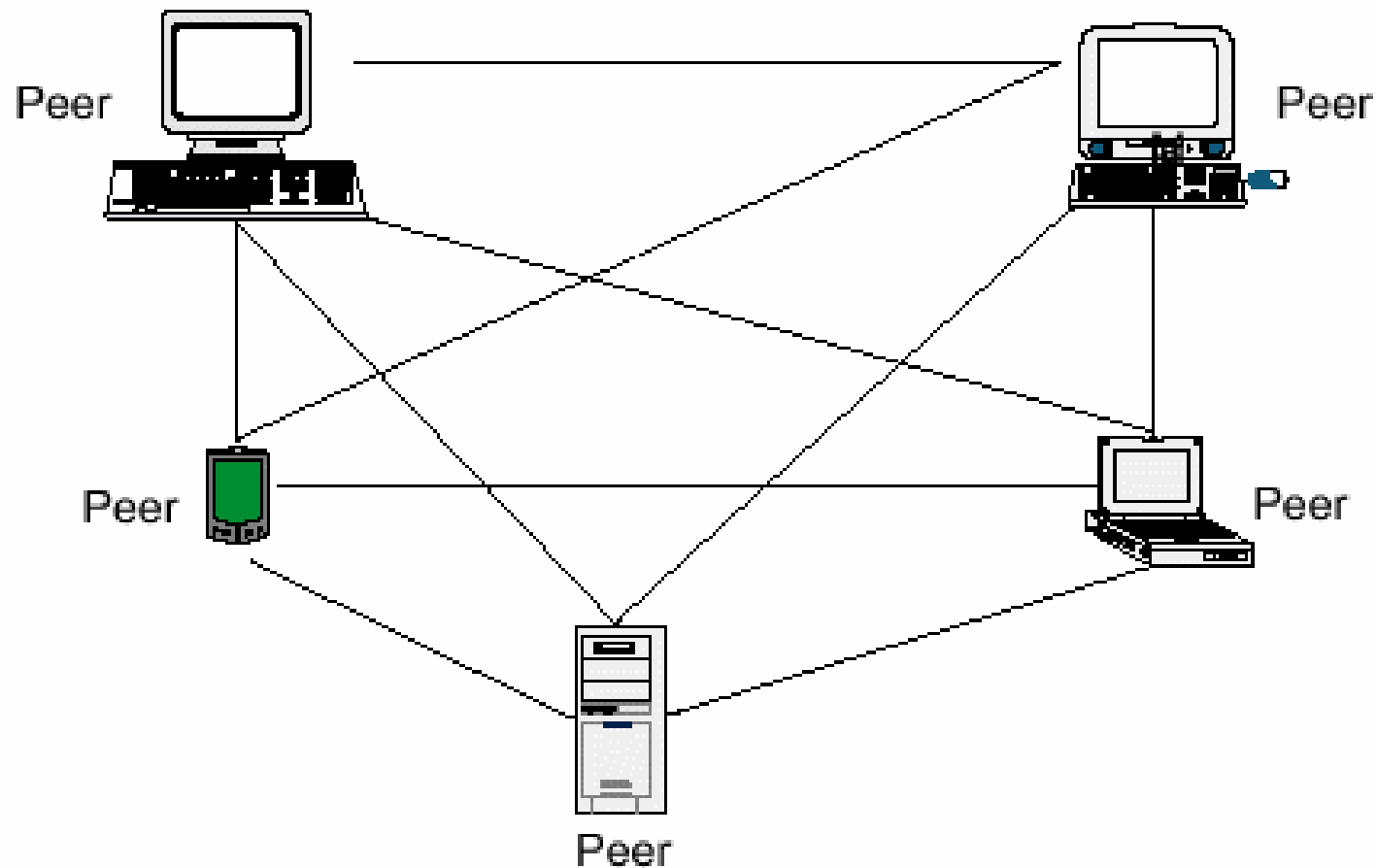
[Super P2P]



[Arquitetura P2P pura]

- Variação da arquitetura P2P ideal;
- Precisamos saber o endereço de pelo menos um “peer” para acessar a rede P2P;
- Gera um conjunto de redes isoladas, que se desconhecem;
- Exemplos: Gnutella e Freenet

[Arquitetura P2P pura]



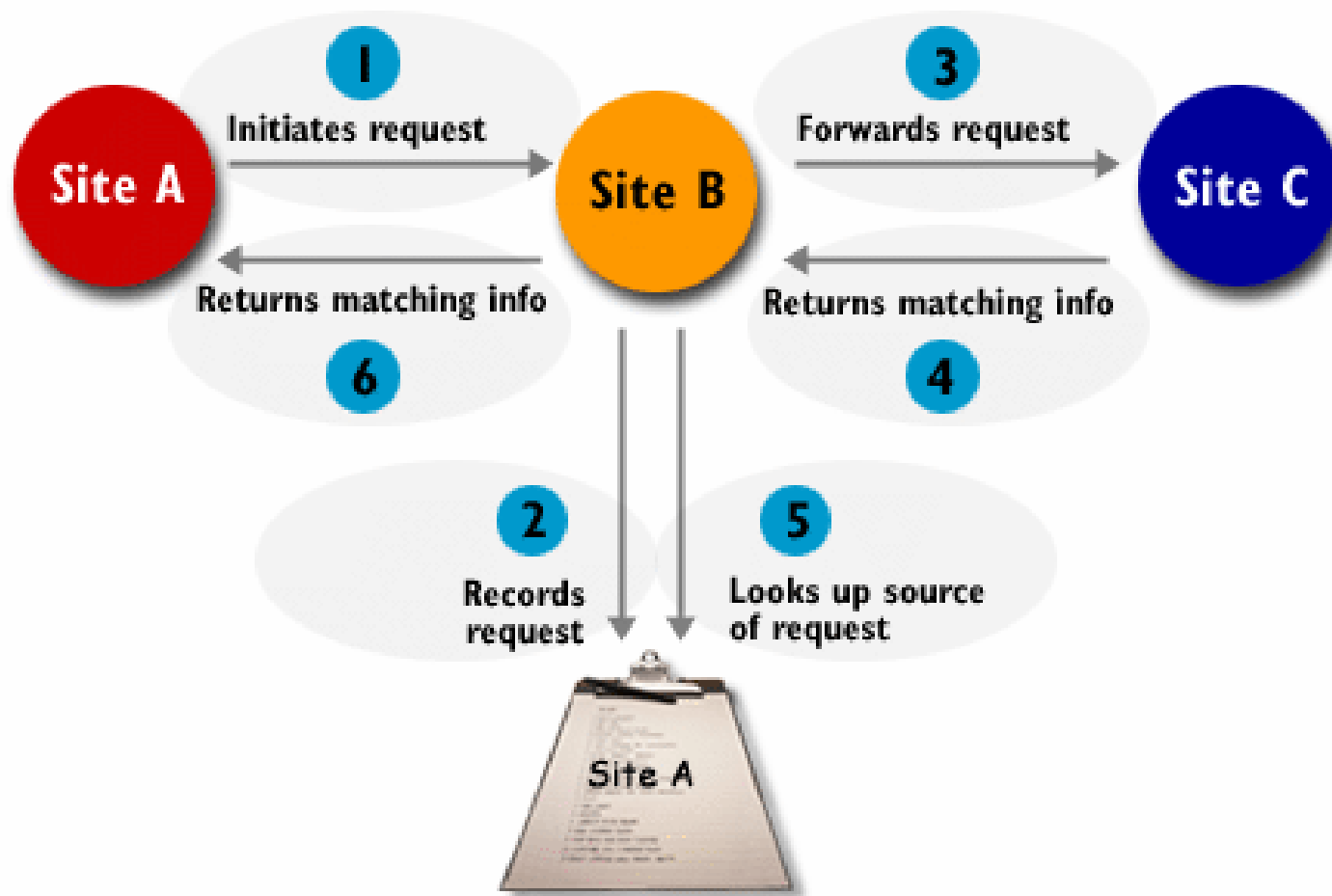
[Gnutella]

- É um protocolo para busca de informações;
- Retorna a informação sem informar a origem;
- Foi produzido pela Nullsoft, uma divisão da AOL, mas o projeto foi cancelado por causa do seu potencial para infringir leis de copyright;
- Brian Mayland fez engenharia reversa do protocolo e iniciou um projeto para produzir clientes Gnutella;
- Hoje, Gnutella é um projeto de código aberto, com os clientes registrados sob licença GNU;

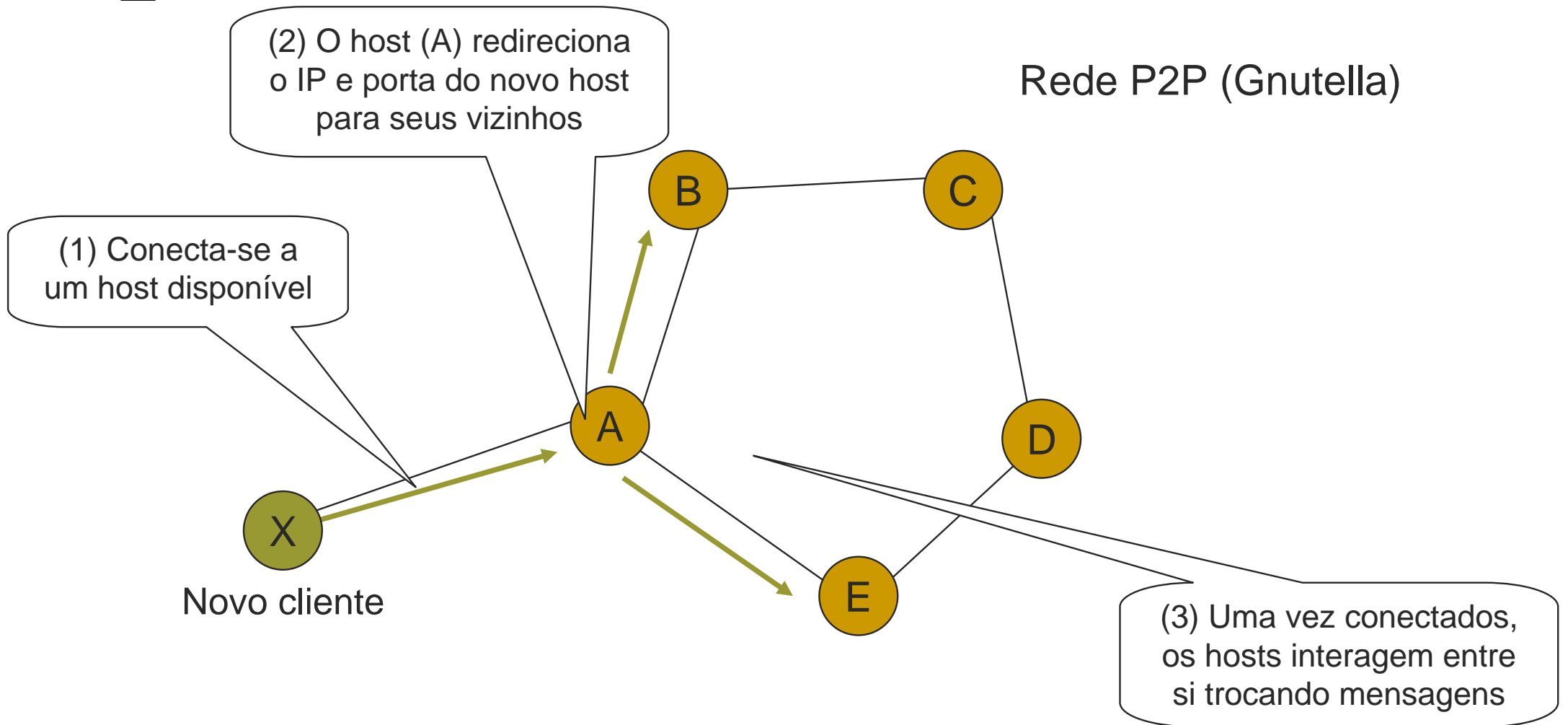
[Gnutella: Protocolo]

- O protocolo define apenas como uma string de busca deve ser encaminhada para outros hosts;
- Não há especificação de como um host deve tratar esta string;

[Gnutella: Protocolo]



Gnutella: Overview



[Gnutella: Mensagens]

- Mensagens do protocolo:
 - PING
 - PONG
 - IP, porta, numero e tamanho dos arquivos disponíveis;
 - Query Messages
 - Possuem um ID, fonte desconhecida, percorre toda a rede (teoricamente)
 - Query Response Messages
 - IP, porta, etc. informação relevante para fazer download da informação;
 - Get / Push Messages
 - Requisita por um arquivo em um host. Push é usado caso o host esteja sob um Firewall;

[Gnutella: Mensagens]

- As mensagens são propagadas através dos hosts;
- Para evitar uma propagação infinita, as mensagens possuem um flag (Time to live);
- Esta abordagem cria uma fronteira de busca à rede P2P
 - A medida que acrescentamos novos hosts, com novas informações, e os colocamos fora da “fronteira”, seu conteúdo se torna inalcançável;

[Problems: Non-compliance]

- O protocolo não define como os hosts devem ser implementados;
- Um bug em um host pode afetar toda a rede;
- Um host pode corromper mensagens, alterar requests, etc.
- Como o protocolo não segue uma especificação, um sistema é dito compatível com o protocolo se este é inserido na rede e funciona. :-)

[Problems: Connectivity]

- Os clientes Gnutella precisam de pelo menos um host para se inserirem na rede;
- Não existe uma forma de descobrir sozinho o endereço de algum host da rede, o usuário precisa buscar outras fontes;
- Para tentar sanar este problema, algumas iniciativas criaram caches e listas de hosts ativos (gnoCache);

Problems: Lack of Search Results

- Com o aumento da rede, vários hosts com conexão dial-up foram alvo de uma grande quantidade de requisições, tornaram-se os gargalos da rede;
- Algumas ações foram tomadas para resolver o problema, como:
 - A introdução de “Reflectors”, um servidor Gnutella que funciona como um proxy para usuários com links de baixa capacidade;
 - Acrescentar algumas regras de preferência de conexão, que levaram os clientes com baixa velocidade a serem menos requisitados.

[Problems: Download Failure]

- “busy signals”: apesar do resultado de uma busca atender a expectativa do usuário, este não é muito útil se os arquivos não estão disponíveis para download;
- Os hosts podem ser inalcançáveis, graças a firewalls;
- As aplicações podem ter bugs, serem incompletas;

[Problems: Download Failure]

- Os hosts pode se desconectar, alterar o conteúdo compartilhado, etc.
- “freeloading”, usuários fazendo mais download que upload de arquivos;

[Resultado Estudo da Xerox]

- 70% dos usuários não compartilham arquivos;
- 50% de todas as respostas vêm de 1% dos hosts;
- A rede tende a assumir uma característica cliente-servidor;

[Conclusões]

- Gnutella tem muitos problemas que podem comprometer seu futuro;
- The tragedy of Digital Commons
 - Uso > contribuição ao sistema;
 - Limitação do número de conexões e da banda de conexão;
- Falso senso de segurança;
 - Pequeno risco do sistema ser desligado?
 - Poucos hosts compartilham muitos arquivos, atuando como servidores centralizados;

[Conclusões]

- What Next?
 - Muitos problemas para resolver;
 - Não existe uma comunidade de desenvolvedores preocupada com a padronização e evolução do protocolo;
 - Alguns clientes alteram o protocolo para o bem de seus usuário, não pensam coletivamente; Criam sub protocolos incompatíveis entre si;

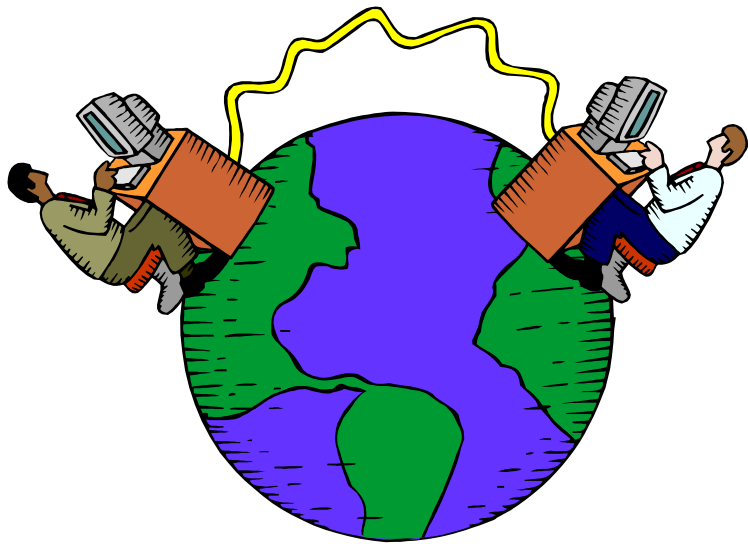
[Referências]

- Free Riding on Gnutella
http://www.firstmonday.dk/issues/issue5_10/adar/
- In Praise of Freeloaders
http://www.openp2p.com/pub/a/p2p/2000/12/01/shirky_freeloading.html
- The Value of Gnutella and Freenet
http://www.webreview.com/pi/2000/05_12_00.shtml
- Gnutella: Alive, Well, and Changing Fast
<http://www.openp2p.com/pub/a/p2p/2001/01/25/truelove0101.html>
- What is P2P and What Isn't
<http://www.openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html>
- Gnutella and Freenet Represent True Technological Innovation
<http://www.oreillynet.com/pub/a/network/2000/05/12/magazine/gnutella.html>
- Peer-to-peer computing in E-Business
Cutter Consortium Vol. 4, NO. 3

[A Synapse Tech]

- Empresa de treinamento e consultoria em Tecnologia da Informação
 - Cursos com garantia de satisfação
 - Mais de 15 instrutores altamente capacitados
 - Aulas teóricas e práticas
 - Cursos abertos e fechados para empresas
 - Atuamos em todo o Brasil
- Contato:
 - <http://www.synapsetech.com.br>
 - contato@synapsetech.com.br

[Obrigado!]



Rodrigo Rebouças de Almeida

rodrigor@dsc.ufcg.edu.br

<http://www.dsc.ufcg.edu.br/~rodrigor>