

# ESTUDO DE REDES P2P PARA ARMAZENAR E RECUPERAR DADOS

**Klaus Korte (bolsista UEMS)**

Ciência da Computação, Universidade Estadual de Mato Grosso do Sul, UEMS, 021035@comp.uems.br,

**Nilton César de Paula (orientador)**

Ciência da Computação, Universidade Estadual de Mato Grosso do Sul, UEMS, nilton@comp.uems.br

Área CNPq: SISTEMAS DE COMPUTAÇÃO

## Resumo

Uma rede P2P (*peer-to-peer*) é uma tecnologia para disponibilizar dados através do conceito de compartilhamento sem se preocupar em definir papéis aos computadores envolvidos. Assim, os dados compartilhados estão armazenados em um computador dito “cliente”, tanto no modo de organização centralizado como descentralizado. A figura do “servidor” no modo centralizado é o realizador de autenticações e descoberta dos computadores da rede, enquanto no modo descentralizado ela não existe porque todos computadores realizam as mesmas funções, de “cliente” e “servidor”. Nesta proposta de Iniciação Científica o objetivo foi realizar um estudo de redes P2P para identificar arquiteturas e funcionalidades que podem ser utilizadas pelo sistema LIMA.

**Palavras-chave:** Gerenciamento de dados, sistemas distribuídos, redes P2P.

## 1. Introdução

Uma grade computacional tem sido uma alternativa para melhorar a execução de aplicações, por permitir o uso de vários recursos distribuídos como um único e poderoso ambiente de computação (FOSTER; KESSELMAN; TUECKE, 2001).

Para utilizar os recursos de uma grade computacional, tais como *clusters* e conjunto de computadores, precisa-se manter informações na grade sobre os recursos existentes (FOSTER; KESSELMAN, 1999). Para isto, vários sistemas de monitoramento de recursos foram propostos, tais como o Hawkeye (HAWKEYE, 2011), Ganglia (MASSIE; CHUN; CULLER, 2004), Monalisa (CIRSTOIU et al., 2007; LEGRAND et al., 2004) e, recentemente, o LIMA (*Light-weight Monitoring Architecture*) (DE PAULA, 2009).

No LIMA, os recursos são organizados em *pools* e cada *pool* é constituído de um *cluster* ou conjunto de computadores. As informações sobre os recursos são obtidas usando

módulos de captura instalados e configurados nos *pools* e elas são distribuídas entre os diversos *pools* da grade.

Com o acréscimo de novas funcionalidades no LIMA, como o uso de informações do PBS (*Portable Batch System*) (BERMAN; FOX; HEY, 2003) para ampliar a utilização de recursos de uma grade computacional, é interessante agregar outras tecnologias ao LIMA, como o uso de redes P2P (*peer-to-peer*) (KUROSE; ROSS, 2007). Com isto, o objetivo é adquirir experiência sobre esse tipo de rede para verificar a viabilidade de construção de ambientes autônomos de armazenamento e recuperação de dados para o LIMA.

## 2. Materiais e métodos

Esta proposta teve por objetivo estudar conceitos relacionados com redes P2P para apoiar uma análise de viabilidade futura em usar esse tipo de tecnologia no LIMA.

Os assuntos estudados relacionados a redes P2P foram: os tipos de arquiteturas (não estruturada, estruturada, híbrida e parcialmente centralizada), as principais funcionalidades dos *peers* (capacidade de auto-organizar, auto-tolerantes a falhas, mecanismos de roteamento), as tecnologias de desenvolvimento de sistemas P2P (JXTA e Groove Development Kit), os exemplos de aplicações (PeerDB, Edutella, Freenet, Gnutella e Napster).

Para o estudo, fontes bibliográficas e sites especializados foram pesquisados e, por fim, foi feita uma documentação dos estudos realizados.

## 3. Resultados

*Peer-to-peer* (tradução literal do inglês de "par-a-par" ou "entre pares"; tradução livre: ponto-a-ponto; sigla: P2P). A computação P2P pode ser determinada como um modelo genérico de arquitetura de *software*, classificada no mesmo nível da computação cliente/servidor.

Os sistemas P2P nada mais são que componentes de *software*, (SINGH, 2001) que são chamados de pontos/nós, *peer*, que compartilham e/ou utilizam os recursos, serviços e conteúdos, de outros pontos para a execução de tarefas de modo direto e descentralizado. Resumidamente, a idéia principal de sistemas P2P é a simplicidade em compartilhar arquivos, programas, e permitir a comunicação direta com outras pessoas pela internet.

Antigamente, quando se desejava desenvolver um sistema P2P, (MILOJICIC; KALOGERAKI; LUKOSE; NAGARAJA; PRUNYNE; RICHARD; ROLLINS, 2002) era preciso antes, programar um protocolo próprio de comunicação para garantir a comunicação

dos *peers* nas redes. Mas com isso ficou impossibilitado para que um sistema P2P não se comunicasse com outro sistema de P2P, pois seus protocolos não eram iguais, mas com a popularização dessas redes teve a necessidade de se desenvolver um protocolo que todos se comunicassem. Dentre as plataformas destaca-se JXTA (GONG, 2002) e Groove Development Kit.

### **A Plataforma JXTA**

Diversas plataformas para prover comunicação *peer-to-peer* foram desenvolvidas. Cada uma com seu próprio foco, mas JXTA é considerada a mais madura atualmente. JXTA é uma plataforma livre que foi desenvolvida pela Sun Microsystems em 2001, para o desenvolvimento de sistemas P2P. A aplicação torna acessível um conjunto de protocolos que são usados para a comunicação de celulares, PDA, servidores, etc. O objetivo de JXTA é a independência da plataforma, isto é, não é limitado a um código, sistema operacional e uma rede. Além disto, a facilidade de qualquer dispositivo digital de se conectar usando JXTA proporciona transparência entre os sistemas de computação.

### **Groove Development Kit (GDK)**

O GDK é uma plataforma para desenvolver aplicações P2P gratuitas, mas os produtos que foram construídos baseados nela necessitam de uma licença. No desenvolvimento do Groove foi utilizada Microsoft Component Object Model (COM), portanto para desenvolver uma aplicação utilizando este modelo é necessário somente o Kit Visual Basic, Delphi, dentre outros. O Groove utiliza uma arquitetura híbrida, o que permite a utilização de uma solução centralizada ou descentralizada.

A seguir são apresentados os estudos sobre os exemplos de aplicações P2P.

### **PeerDB**

O PeerDB (OOI; SHU; TAN, 2003) é um sistema de gerenciamento de dados que foi implementado na plataforma Best Peer. O Best Peer teve início em 2000 na Universidade Singapura, seu objetivo principal foi estudar como tecnologias P2P poderiam ser usadas no gerenciamento de dados distribuídos. PeerDB visa incorporar dados relacionais e possui uma forte dependência de interatividade com os usuários na subordinação de consultas.

### **Edutella**

Edutella (NEJDL; SIBERSKI; SINTEK, 2003) conecta pares altamente heterogêneos (heterogênea em seu tempo de atividade, desempenho, tamanho de armazenamento, funcionalidade, número de usuários, etc.). No entanto, cada ponto pode fornecer informações de meta dados disponível como um conjunto de declarações RDF. O objetivo é fazer com que a natureza distribuída dos pares individuais RDF conectados à rede Edutella completamente transparente, especificando e implementando um conjunto de serviços Edutella. Cada ponto é caracterizado pelo conjunto de serviços que oferece .

### **Freenet**

Uma exposição sobre P2P não estaria completa sem mencionar a Freenet (FREENET, 2012). A prioridade da Freenet é a liberdade de expressão e o anonimato. Para obter isso, a comunicação entre os nós é cifrada, assim como a informação armazenada em cada nó. O usuário contribui com um pedaço do seu HD, mas ele não sabe o que exatamente está armazenado lá.

### **Gnutella**

O Gnutella (GNUTELLA, 2012) é uma rede de compartilhamento de músicas, filmes e *software*, é um sistema P2P. O *software* cliente Gnutella se conecta na rede e a transferência é passada de um nó para o outro circularmente.

O real benefício de se ter a Gnutella tão descentralizada é fazer com que seja muito difícil da rede cair. Diferente do Napster, onde a rede inteira depende de um servidor central, a rede Gnutella não cai ao se desligar qualquer um dos nós.

### **Napster**

O Napster (NAPSTER, 2012) foi o pioneiro no conceito de rede de compartilhamento de arquivo *peer-to-peer*. A versão antiga do Napster, as pessoas armazenavam os arquivos que queriam armazenar no disco rígido e dividiam diretamente com as outras pessoas, tendo um *software* que possibilitava esse compartilhamento, sendo assim que a máquina do usuário tornava um mini-servidor.

## **4. Conclusões**

Podemos tirar como base dos estudos realizados que as redes P2P podem ser uma forma boa de compartilhar arquivos, sendo que esteja tudo legal conforme as leis, podemos verificar que não é apenas uma simples conexão entre computadores que atrás de tudo isso estão programas que se preocupam em soluções viáveis para uma melhor conexão.

Também podemos verificar que para um programa de compartilhamento ser bom, existem vários fatores responsáveis para que isso ocorra, uma característica que ficou bem evidente foi a forma de organização dos computadores conectados a rede.

Com os estudos, verificou-se que para aplicar P2P no LIMA há necessidades de um maior aprofundamento no JXTA, uma plataforma muito utilizada para construção de sistemas P2P e muito complexa.

## 5. Agradecimentos

Agradeço pelo apoio financeiro concedido pela Universidade Estadual de Mato Grosso do Sul e também ao orientador prof. Nilton César de Paula que conduziu os passos do estudo, bem como os que contribuíram diretamente ou indiretamente para a realização desse trabalho.

## 6. Referência Bibliográficas

BERMAN, F.; FOX, G. C.; HEY, A. J. G. **Grid Computing: Making the Global Infrastructure a Reality**. John Wiley & Sons, Inc, 2003.

CIRSTOIU, C. et al. **Monitoring, Accounting and Automated Decision Support for the ALICE Experiment Based on the MonALISA Framework**. In Proceedings of the 2007 Workshop on Grid Monitoring, p.39-44, 2007.

DE PAULA, N.; C. **Um Ambiente de Monitoramento de Recursos e Escalonamento Cooperativo de Aplicações Paralelas em Grades Computacionais**. Tese (Doutorado), Universidade de São Paulo, 2009.

FOSTER, I.; KESSELMAN, C. **The Grid: Blueprint for a New Computing Infrastructure**. Morgan Kaufmann Publishers, 1999.

FOSTER, I.; KESSELMAN, C.; TUECKE, S. **The Anatomy of the Grid: Enabling Scalable Virtual Organizations**. International Journal of High Performance Computing Applications, v.15, p.200-222, 2001.

FREENET. Disponível em: <[http:// http://pt.wikipedia.org/wiki/Freenet](http://pt.wikipedia.org/wiki/Freenet)>. Acesso em: 19 jun. 2012.

GNUTELLA. Disponível em: <[http://www.gta.ufrj.br/grad/06\\_1/p2p/gnutella.html](http://www.gta.ufrj.br/grad/06_1/p2p/gnutella.html)>. Acesso em: 19 jun. 2012.

GONG, L. Project jxta: a technology overview. Technical report, Sun Microsystems, Inc. Palo Alto, CA, USA, 2002.

HAWKEYE. Disponível em: <<http://www.cs.wisc.edu/condor/hawkeye/>>. Acesso em: 30 out. 2011.

KUROSE, J. F.; ROSS, K. W. **Redes de computadores e a Internet**: uma abordagem top-down, 3. ed. São Paulo, Pearson Addison Wesley, 2007.

LEGRAND, I. C.; NEWMANN, H. B.; VOICU, R.; CIRSTOIU, C.; GRIGORAS, C.; TOARTA, M.; DOBRE, C. **MonALISA**: An Agent Based, Dynamic Service System to Monitor, Control and Optimize Grid Based Application. In CHEP, Interlaken, 2004.

MASSIE, M. L.; CHUN, B. N.; CULLER, D. E. **The Ganglia Distributed Monitoring System**: Design, Implementation, and Experience. In Proceedings of Parallel Computing, v.30, issue 7, 2004.

MILOJICIC, D.; KALOGERAKI, V.; LUKOSE, R.; NAGARAJA, K.; PRUNYNE, J.; RICHARD, B.; ROLLINS, S.; XU, Z. **Peer-to-peer computing**. Technical report, HPL-2002-57, HP Labs, 2002.

NAPSTER. Disponível em: <<http://informatica.hsw.uol.com.br/gnutella1.htm>>. Acesso em: 12 mai. 2012.

NEJDL, W.; SIBERSKI, W.; SINTEK, M. **Design Issues and Challenges for RDF- and Schema-Based Peer-to-Peer Systems**, ACM, 2003.

OOI, B. C.; SHU, Y.; TAN, K. L. **Relational data sharing in peer-based data management systems**. ACM SIGMOD Record, 35(3): 59-64, 2003.

SINGH, M. **Peering at peer-to-peer computing**. IEEE Internet Computing, 5, 2001.