

## UMA REVISÃO DE PROPOSTAS DE SOBREVIVÊNCIA PARA REDES DE SOBREPOSIÇÃO MULTICAST P2P

Autores	Filiação	Email
Manoel Camilo De Oliveira Penna Neto	PUCPR	penna@ppgia.pucpr.br
Rober Mayer	UTFPR	rober@utfpr.edu.br

Direitos de cópia - creative commons.

Recebido em:

27/8/2013

Aprovado em:

6/12/2013

Disponibilização no site

Páginas:

81-92

ID do artigo

1603

Editor Científico: Prof. Dr. Osni Hoss, Ph.D.

### RESUMO

O sucesso do *multicast* par-a-par ou P2P em redes de sobreposição para fluxos de dados de importância crítica depende de sua escalabilidade. Porém, como a funcionalidade *multicast*, neste caso, está associada a sistemas finais imprevisíveis, o alto grau de entrada e saída de nós da rede pode resultar em uma significativa perda de performance quando comparado a roteadores. A sobrevivência ou tolerância a falhas é um requisito chave para arquiteturas de distribuição de dados par-a-par. Neste artigo, apresenta-se uma revisão de propostas na literatura recente, que podem ser aplicadas em distribuição de dados em redes de sobreposição *multicasting* P2P visando manter a rede ativa e reduzindo redundância de dados.

**Palavras-chave:** Multicast, P2P, Sobreposição, Sobrevivência.

### A REVIEW OF PROPOSALS FOR SURVIVAL FOR OVERLAY MULTICAST NETWORKS P2P

#### ABSTRACT

The success of multicast peer-to-peer or P2P in overlay networks for data streams of critical importance depends on its scalability. However, as multicast functionality is pushed to autonomous, unpredictable end systems, significant performance loss can be resulted from their higher degree of churn when compared to routers. Survival or resilience is a key requirement for data distribution architectures of peer-to-peer. In this article, we present a review of proposals in recent literature, which can be applied to data distribution in overlay networks multicasting P2P, order to keep the network active and reducing data redundancy.

**Keywords:** Multicast, P2P, Overlay, Survival.

## 1 INTRODUÇÃO

A transmissão *multicast* é um mecanismo eficiente para suportar aplicações de comunicação em grupo, tais como conferência de áudio e vídeo e distribuição de conteúdo que elimina a maioria dos pacotes redundantes na rede (BIRRER e BUSTAMANTE, 2005). É particularmente útil para as sistemas de distribuição de fluxos de informação de uma origem para múltiplos destinos, por exemplo, a aplicação denominada vídeo *streaming* na qual um fluxo de vídeo é produzido em um nó fonte e distribuído para vários nós de destino. Um fluxo contínuo de informação que é transmitido de uma fonte para múltiplos destinos é denominado fluxo *multicast*.

Várias soluções baseadas em *multicasting* P2P têm recebido um ganho de popularidade nos últimos anos. Esse *multicasting* pode efetivamente dar suporte a aplicações de comunicação ao vivo com conteúdos variados, pela Internet. Neste mecanismo auto-adaptativo, abordando a camada de aplicação, os pares participantes configuram-se em uma topologia de sobreposição para entrega de dados. Nesta topologia cada enlace corresponde a um caminho *unicast* entre dois sistemas finais na Internet. Todas as funcionalidades relacionadas com *multicast* são implementadas nos sistemas finais, ao contrário dos roteadores, e o objetivo do protocolo *multicast* é construir e manter uma cobertura eficiente para transmissão de dados (BIRRER e BUSTAMANTE, 2006).

Os pares em um sistema P2P são mais vulneráveis a falhas por serem compostos por *hosts* finais, na maioria das vezes *desktops* ou *notebooks* de uso comum. Assim, a interrupção de um único *host* pode interferir no serviço oferecido por diversas conexões.

Neste contexto, a sobrevivência a falhas em redes P2P é quesito essencial no projeto e operação destas redes. A sobrevivência a falhas é a capacidade de uma rede não prejudicar, ou não interromper, as conexões de seus usuários quando ocorrer a falha de algum recurso da rede. Segundo Fei e Yang (2007), nas redes convencionais em malha, os mecanismos que oferecem sobrevivência a falhas, também denominados mecanismos de sobrevivência, são classificados, basicamente, em dois tipos: a proteção, que pré-computam e pré-locam os recursos de recuperação; e a restauração, que tratam dos recursos de recuperação de maneira reativa apenas quando ocorre a falha. Aplicando esses mecanismos de sobrevivência a falhas a uma rede de sobreposição P2P, enlaces e nós alternativos podem ser encontrados para que a rede continue a oferecer os serviços a que foi destinada, mesmo na presença de falhas

Neste artigo, com a motivação de contribuir para melhorar a robustez de aplicações críticas em redes P2P, revisa-se algumas propostas de mecanismos tolerantes a falhas e classifica-se os mesmos de acordo com a topologia de sobreposição e os métodos de controle de erros. Considera-se uma aplicação crítica aquelas que necessitam de grande confiabilidade, como sistemas de aviso de tempestades, atualizações de segurança e dados financeiros (WALKOWIAK e PRZEWOZNICZEK, 2011). Birrer e Bustamante (2007) realizaram uma comparação das propostas de tolerância a falhas baseadas em árvore e avaliaram sua efetividade contrastando o desempenho e o custo associado através de simulações e experimentos em áreas abrangentes. Em contraste a comparação apenas das propostas baseadas em árvore de Birrer e Bustamante (2007), a contribuição deste trabalho é uma revisão atualizada dos protocolos baseados em árvore e baseados em redes em malha propostos na literatura, e o fornecimento de detalhes a respeito dos

mecanismos que esses protocolos adotam para melhorar a tolerância a falhas e a robustez em ambientes altamente dinâmicos e críticos. Também contribui-se, com o acréscimo das propostas que visam a proteção de redes P2P estáticas, uma visão na qual o objetivo é a proteção contra falhas da rede e não contra a constante entrada e saída dos pares (*churn*) na rede.

O restante deste artigo está organizado como segue. A Seção 2 apresenta a revisão do *multicast* de sobreposição. De acordo com o projeto da topologia de dados é apresentada a classificação de várias propostas tolerante a falhas: baseadas em árvore, baseadas em redes em malhas e dirigidas por dados. Algumas propostas representativas em cada categoria são descritas na Seção 3. Uma nova linha de pesquisa é apresentada na Seção 4, a sobrevivência em redes de sobreposição *multicast* P2P estáticas. A Seção 5 conclui este trabalho.

## 2 VISÃO GLOBAL DO MULTICAST P2P

Sistemas P2P normalmente implementam redes de sobreposição, que são criadas na camada de aplicação, sobre as redes físicas. Essas redes são usadas para descobrir e identificar os pares e tornar a rede P2P independente da topologia da rede física. O conteúdo é compartilhado diretamente sobre a rede IP (*Internet Protocol*). Redes de sobreposição consistem de todos os pares participantes como nós da rede, portanto, são redes lógicas construídas sobre uma rede física já existente.

As redes de sobreposição são constituídas de enlaces lógicos criados entre os nós participantes da rede. Assim, existe a ideia de que a rede de sobreposição forma uma camada acima da rede física. Outro detalhe que precisa ser notado é que um enlace lógico da rede de sobreposição não precisa corresponder a um enlace físico da rede física (COUTINHO, 2006).

A primeira geração de sistemas *multicast* P2P, concentrou-se em permitir que o maior conjunto possível de pares se conectassem a sessões *multicast* em redes de sobreposição. Desde que Banerjee et al., (2003) explicitaram, primeiramente, um esquema tolerante a falhas usado para aumentar a distribuição de qualquer protocolo *multicast* enquanto mantém baixos limites de latência, várias técnicas têm sido propostas recentemente explorando o conceito de diversidade de caminhos, minimizando as dependências dos nós e a redundância de dados para prover rápida recuperação de dados quando um par falha e reduzir a taxa de perda de pacotes (CASTRO et al., 2003). Especialmente quando o *multicast* de sobreposição é combinado com transporte de dados de grande importância para um grande número de usuários, os sistemas *multicast* P2P entraram em uma nova geração, a qual considera, principalmente, a melhora da qualidade de serviço, aumento da robustez em relação ao erro, falha dos pares e redução do atraso fim-a-fim (WALKOWIAK e PRZEWOZNICZEK, 2011).

Em resposta às dificuldades de implantação da transmissão *multicast* nos protocolos da camada de rede IP (*Internet Protocol*), uma série de protocolos recentemente propostos adotam a alternativa *Peer-to-Peer* (P2P), com todas as funcionalidades da transmissão *multicast* implementadas exclusivamente nos sistemas finais (*hosts* de usuários) e não nos roteadores, e o objetivo do protocolo *multicast* é construir e manter uma cobertura eficiente para transmissão de dados (BIRNER e BUSTAMANTE, 2006). Neste mecanismo autoadaptativo, implementado na camada de aplicação, os pares participantes configuram-se em uma topologia de sobreposição (*overlay*), ou seja, uma rede lógica sobre uma rede física existente, para

entrega de dados. Na topologia de sobreposição cada enlace corresponde a um caminho *unicast* entre dois sistemas finais na Internet. Várias soluções para transmissão *multicast* em redes (P2P) têm recebido um ganho de popularidade nos últimos anos, demonstrando que podem efetivamente dar suporte a aplicações de comunicação ao vivo com conteúdos variados, pela Internet.

### 3 CLASSIFICAÇÃO DAS PROPOSTAS DE MULTICASTING P2P

As propostas de *multicasting* P2P podem ser organizadas dentro de três classes, de acordo com a construção da rede de sobreposição: baseadas em árvore (*tree-based*) (CASTRO et al., 2003), baseadas em redes em malha (*mesh-based*) (MAGHAREI e REJAIE, 2007) e dirigidas por dados (*data-driven*) (ZHANG et al., 2005).

#### 3.1 Propostas de Multicasting P2P Baseadas em Árvore

Os protocolos apresentados a seguir organizam os pares dentro de árvores lógicas simples ou múltiplas sobre as quais os dados *multicast* são distribuídos. As árvores são muito dependentes da confiabilidade de nós não folhas cujas falhas podem resultar em partições temporárias da árvore, podendo causar perda de pacotes de dados. Recentemente, para reduzir o impacto da dinâmica de entrada e saída dos nós no desempenho dos protocolos *multicast* das redes de sobreposição, diversas técnicas têm sido propostas na tentativa de melhorar a tolerância a falhas explorando a ideia de diversidade de caminhos e diminuindo a dependência em relação aos nós. As diversas propostas na classe com base em árvores podem ser classificadas em quatro categorias de acordo com o tipo de redundância: redundância por arestas cruzadas (*cross-edge*) (BANERJEE et al., 2002; BANERJEE et al., 2006) redundância interna da árvore de *multicast* (*multiple tree*) (BIRRER e BUSTAMANTE, 2005; TRAN et al., 2004), redundância por múltiplas árvores de multicast (WALKOWIAK e PRZEWOZNICZEK, 2011; CASTRO et al., 2003; PADMANABHAN et al., 2003; WALKOWIAK, 2009) e redundância baseada em rumores (*gossip-based tree*) (KACIMI et al., 2005).

**Redundância por arestas cruzadas:** O PRM (*Probabilistic Resilient Multicast*) proposto por Banerjee et al., (2006), foi o primeiro esquema proposto para multicast com tolerância a falhas. O PRM adota um componente pró-ativo para encaminhamento aleatório no qual cada nó da rede de sobreposição, com uma pequena probabilidade, escolhe um número constante de outros nós da rede de sobreposição e encaminha transmissões adicionais ao longo da árvore de distribuição e sobre algumas arestas cruzadas. O componente randomizado acrescenta pouca sobrecarga e pode garantir elevados índices de entrega de dados mesmo com alta quantidade de falhas de nós da rede de sobreposição. O PRM possui um segundo componente, este reativo, chamado *Triggerred NACKs* (*Negative acknowledgments*) para controlar a perda de dados devido a erros de enlace e congestão da rede, que foi originalmente implementado usando NICE (BANERJEE et al., 2002), que é um protocolo de aplicação, que tem como ideia principal a redução de complexidade e o ganho de desempenho utilizando uma estrutura hierárquica para sua rede, organizada em níveis.

**Redundância interna da árvore de multicast:** Nesta categoria os receptores de uma árvore são organizados dentro de uma hierarquia de *clusters* em que cada nível tem definido um líder e um líder associado. O Zigzag, idealizado por Tran et al., (2004),

é um protocolo *multicasting* P2P hierárquico para redes de sobreposição que se encaixa nessa descrição. Nele, a árvore é construída com os pares organizados em uma hierarquia multicamada recursiva de *clusters*. Com um líder e um líder associado em cada nível, o Zigzag separa o controle da manutenção da árvore da distribuição de dados, melhorando a tolerância a falhas. O líder é o responsável por monitorar os membros do *cluster* enquanto que o líder associado é o responsável por transmitir o conteúdo para os membros do *cluster*, que por sua vez trocam mensagens de controle para entrar e manter suas posições na árvore. A falha do líder não afeta a continuidade do serviço de outros membros, já que ele não é o responsável pela transmissão de dados. Quando um líder associado deixa a rede devido a uma falha ou por decisão individual, o líder define um novo líder associado. Devido ao uso de dois pares para compartilhar as responsabilidades na presença de falhas, a recuperação de falhas pode ser feita localmente no *cluster*, com impacto, no máximo em um número constante de receptores existentes e principalmente sem sobrecarregar o nó fonte com a manutenção da árvore e a distribuição de dados na árvore.

A proposta de Birrer e Bustamante (2005), é um protocolo *multicasting* P2P, para redes de grande escala, heterogêneas em ambientes altamente dinâmicos. Essa proposta define o protocolo Nemo (BIRRER e BUSTAMANTE, 2005), para *multicasting* P2P, organizando os pares dentro de uma hierarquia lógica baseada na proximidade da rede, sobre a qual as árvores de distribuição de dados estão implicitamente definidas por um conjunto de regras de encaminhamento. O Nemo atinge a tolerância a falhas através da inserção de co-líderes, líderes alternativos que compartilham a carga de encaminhamento com os líderes dos *clusters*, de maneira semelhante ao descrito no protocolo ZigZag. Quando um líder de um *cluster* recebe uma transmissão de um membro qualquer de seu *cluster*, ele deve encaminhar esta transmissão para os demais membros de seu *cluster* e para o líder do próximo *cluster*, com a ajuda do co-líder de seu *cluster*, alternando o receptor das mensagens entre eles. Como os co-líderes compartilham a carga das mensagens de encaminhamento com os líderes, eles também ajudam a reduzir a demanda de largura de banda dos líderes de *clusters*, melhorando a escalabilidade do sistema como um todo. O Nemo utiliza número de sequências e NACKs, para detectar perda de pacotes, através de um algoritmo heurístico para realizar a manutenção da rede de sobreposição. O Nemo ainda reduz o custo da manutenção da rede de sobreposição através da adoção de uma aproximação probabilística, na qual as operações são executadas com alguma prioridade, ou de forma alternativa, adiadas para o próximo intervalo.

**Redundância por múltiplas árvores de multicast:** Para resolver o problema de gargalos no encaminhamento de dados em árvores isoladas, várias propostas surgiram com o objetivo de usar múltiplas árvores disjuntas sobre os quais os fluxos de dados são disseminados. Encaminhando diferentes fluxos sobre cada árvore e tornando cada par um nó interno em pelo menos uma árvore, a redundância é atingida através da distribuição da carga de dados de forma mais igualitária entre os pares participantes (WALKOWIAK e PRZEWOZNICZEK, 2011) e (WALKOWIAK, 2009).

Seguindo essa ideia, Castro et al., (2003) apresentaram o Splitstream, que introduz a difusão de conteúdo através de múltiplas árvores. A implementação do Splitstream é baseada no Scribe, um sistema de comunicação de grupo construído sobre Pastry que por sua vez é uma rede de sobreposição P2P estruturada. O Splitstream consegue atingir esse objetivo dividindo o conteúdo original em vários fluxos, e encaminhando cada fluxo usando uma árvore separada. Os pares juntam-se a tantas árvores quantos fluxos que eles desejam receber e especificam um limite

máximo de fluxos que eles estão dispostos a encaminhar. O desafio é construir uma floresta de árvores *multicast* em que um nó interno em uma árvore é um nó folha em todas as demais árvores e que as restrições de largura de banda especificada pelos nós sejam satisfeitas, ou seja, respeitando a capacidade de banda de cada par. Isso garante que a carga de encaminhamento possa ser espalhada através de todos os pares participantes. Dividindo o fluxo através de múltiplas árvores também melhora-se a tolerância a falhas para falha de nós. O Splitstream oferece robustez para falhas dos nós e súbitas partidas dos nós, explorando a diversidade de caminhos, garantindo que a maioria dos nós sejam nós internos em apenas uma árvore. Além disso, a falha de um simples nó causa a perda temporária de, no máximo, um fluxo.

O CoopeNet proposto por Padmanabhan et al., (2003), utiliza múltiplas árvores de distribuição e segue uma abordagem que mescla aspectos da arquitetura cliente-servidor com aspectos da arquitetura P2P. Assim como na arquitetura cliente-servidor, há um servidor central que provê todos os vídeos (pré-gravados ou ao vivo), além de servir como “porta de entrada” para as novas estações que queiram se conectar à rede. Assim como nas redes P2P, há conexões diretamente entre as estações que, aproveitando os recursos ociosos que possuem, auxiliam o servidor na distribuição do conteúdo, provendo uma grande escalabilidade ao sistema.

Trata-se de um sistema P2P de fluxo de mídia ao vivo que complementa, ao invés de substituir, a arquitetura cliente-servidor. Uma distribuição sincronizada de conteúdo aos clientes é realizada. Árvores de distribuição são formadas através de conexões *unicast*, estando o servidor em suas raízes. Cada cliente é incluído como nó nessas árvores, podendo ser admitido diretamente pelo servidor, caso este tenha recursos para isso, ou por outros nós que já estejam conectados no sistema, escolhidos aqueles que ainda tenham recursos excedentes. Cada cliente, então, repassa o fluxo recebido em uma determinada árvore de distribuição a todos os seus descendentes, até que o fluxo atinja os nós folhas daquela árvore. O grau de saída de cada nó do sistema é limitado pela banda de saída da respectiva entidade (servidor ou cliente). Em geral, o servidor apresenta uma banda de saída bem superior às disponibilizadas pelos clientes.

Para minimizar os problemas de rupturas em árvores decorrentes da constante entrada e saída de clientes, o sistema CoopNet utiliza, além de múltiplas árvores de distribuição, a codificação em múltiplos descritores (*Multiple Description Coding* ou MDC). A MDC é um método de codificação de áudio e/ou vídeo no qual um sinal é codificado em certo número de fluxos distintos, sendo cada fluxo chamado de descritor. Dessa forma, utilizando uma árvore de distribuição distinta para o transporte de cada descritor, o fluxo não é totalmente interrompido quando uma desconexão ocorre. Ao contrário, apenas um subconjunto dos descritores deixa de ser entregue temporariamente, ocasionando uma queda momentânea na qualidade de fluxo recebida.

O CoopNet emprega a sugestão do Splitstream, descrito anteriormente, que indica que cada cliente deve ser admitido como nó interior em uma única árvore, sendo nó folha nas demais. Com isso, consegue-se construir árvores o mais disjuntas possível, diversificando os caminhos de rede para aumentar a tolerância a falhas, bem como para melhor aproveitar as bandas de saída dos clientes (se um nó fosse folha em todas as  $n$  árvores, a sua banda de saída não seria aproveitada).

Walkowiak (2009), aborda o problema para transmissão *multicast* em redes P2P estáticas, formulando um problema de otimização para criação de árvores disjuntas em redes de sobreposição para proteger o sistema contra os seguintes tipos de falhas:

falhas dos nós raízes, falhas de enlaces de sobreposição, falhas de nós de *upload* e a falhas de enlaces de ISP (*Internet Service Provider*). Os objetivos são dois: minimizar o custo de transmissão e maximizar o *throughput* de todos os pares receptores da rede. Para resolver o problema, são usados um modelo de programação inteira mista (MIP) e simulações. No primeiro caso, os melhores resultados gerados pelo resolvidor CPLEX são apresentados e discutidos. O último caso representa um ambiente distribuído e um simulador de rede simples é aplicado para fazer os experimentos. Os resultados obtidos em ambos os métodos da pesquisa indicam que os requisitos adicionais necessários para a proteção não tem uma influência significativa sobre o desempenho do sistema P2P *multicasting*.

Walkowiak e Przewozniczek (2011) apresentam uma extensão do trabalho anterior, considerando, também, a métrica de atraso. O problema foi resolvido por modelos de programação inteira mista e por um método heurístico. No primeiro caso foram utilizadas duas formulações, a mesma formulação por fluxos anteriormente estudada e uma formulação por níveis que permite que restrições referentes ao número de níveis das árvores sejam implementados, reduzindo o tempo de execução do sistema. Novamente os testes foram executados sobre o CPLEX concluindo-se que os sistemas P2P *multicasting* podem ser melhorados com métodos de proteção adicionais, sem redução significativa do desempenho do sistema. Também é apresentando um algoritmo heurístico, chamado SuTeC (*Survivable Trees Constructor*) para resolver o problema de sobrevivência em P2P *multicasting*, com o objetivo de minimizar o custo dos fluxos, em redes de grande porte, para os quais os modelos de programação inteira mista não conseguem encontrar solução em tempo adequado. O SuTeC é um método heurístico estocástico que utiliza um algoritmo de propósito geral e que não utiliza mecanismo de restrições.

**Redundância baseada em rumores (gossip):** Em um algoritmo baseado em rumores, um nó envia uma mensagem nova para um conjunto aleatório de nós selecionados. Esses nós fazem o mesmo na próxima rodada, e assim sucessivamente até a mensagem chegar a todos os nós. A escolha aleatória dos alvos oferece resistência a falhas e permite operações descentralizadas. Porém, seu uso na transmissão de fluxos não é tão simples, pois como a escolha do destino da mensagem a ser enviada é aleatória, pode causar redundância significativa, o que é muito grave para aplicações de transmissão de fluxos de banda alta. Assim, foi desenvolvido um algoritmo inteligente de seleção de pares e um algoritmo de escalonamento de baixa sobrecarga que distribui dados de múltiplos parceiros por Kacimi et al., (2005), o que reduz significativamente a redundância.

O HON-P2P (*Hybrid Overlay Network*) proposto por Kacimi et al., (2005) combina o protocolo baseado em rumores com uma árvore de sobreposição não estruturada para prover fluxos de vídeo sob demanda em ambientes P2P. O HON-P2P é um mecanismo híbrido para compartilhamento de conteúdo multimídia, no qual os pares são agrupados em *clusters* de acordo com a similaridade do conteúdo multimídia. Entre os pares em potencial, cada nó realiza uma espécie de intercâmbio de informações com alvos aleatórios. As redes de sobreposição do HON-P2P, correspondem a hierarquias de classificação de domínio que definem conceitos compartilhados pelo par dos *clusters*. Similarmente, sobreposições baseadas em recursos são criadas para representar propriedades de dados de multimídia, como cor e textura. A arquitetura contém dois tipos de nós: os super pares, com alta capacidade de armazenamento e computação e que realizam consultas de roteamento, indexação e pesquisas de dados em nome de pares com menor poder de computação.

Os *clusters* das redes de sobreposição são construídos usando diferentes métodos de agrupamento. Um par pode juntar-se a diferentes redes de sobreposição. Para participar de um *cluster*, o par deve primeiro realizar uma análise de sobreposição para determinar sua semântica baseada na representação de recursos ou assinatura. Em seguida ele conecta-se a qualquer par na rede HON-P2P para obter a informação de sobreposição. Uma vez conectado, o par envia por meio da conexão inicial sua assinatura para super pares interessados que respondem com seus dados de assinatura de *cluster*. Baseado nas representações de *cluster* retornadas, o par decide juntar-se e cria enlaces para um ou mais *clusters*. Finalmente os super pares dos *clusters* unidos ao novo par recalculam seus valores de representação. Quando um par deixa a rede HON-P2P, ele notifica os vizinhos diretamente conectados a ele e os super pares dos *clusters*, utilizando-se do método de inundação.

### 3.2 Propostas de Multicasting P2P Baseadas em Malha

Em sistemas de distribuição de fluxos P2P em redes em malha (*mesh*), os pares formam uma rede de sobreposição em forma de malha incorporando a entrega de conteúdo. Os pares requisitam independentemente porções de vídeo sobre um conjunto de vizinhos e fornecem dados disponíveis para os seus vizinhos (MAGHAREI e REJAIE, 2007). A tolerância a falhas é atingida de acordo com a atualização dos pares. As propostas desta categoria adotam um método dinâmico que permitem as mudanças de membros e condições de rede facilmente, o método *pull-push*, que trata as mudanças de membros e condições de rede através da descrição de como os conteúdos devem ser requisitados ao servidor. No modelo *push* o cliente sinaliza o servidor para que a transmissão inicie. E no modelo *pull* o cliente solicita trechos de dados conforme vai necessitando, repetindo as solicitações até possuir todos os dados desejados.

O primeiro mecanismo baseado em rede em malha para fluxos P2P que permite a um par, dentro de uma rede P2P, fazer o *download* de um mesmo arquivo de vários outros pares simultaneamente é o PRIME (MAGHAREI e REJAIE, 2007). Nele, os pares participantes formam uma rede em malha aleatória conectada e direcionada. Cada par participante na rede de sobreposição tem múltiplos pais e múltiplos pares filhos. Cada par periodicamente envia seus novos pacotes disponíveis para os seus filhos e solicita pacotes específicos de seus pais. Todas as conexões na rede de sobreposição possuem controle de congestão e são sempre iniciadas pelo par filho correspondente. Cada par tenta manter um número suficiente de pais que podem coletivamente atender aos seus requisitos de recursos. Quando um par precisa de um (ou mais) novo(s) pai(s), ele contata um nó de inicialização para aprender sobre um subconjunto aleatório de outros pares no sistema e então solicita a esses pares para servir como seu pai. As redes de sobreposição baseadas em malha são de fácil manutenção e são muito resistentes a saídas voluntárias dos pares. As conexões de entrada e saída de cada par são mais propensas a ter diversos caminhos que por sua vez reduzem a probabilidade de gargalo compartilhado entre eles.

### 3.3 Propostas De Multicasting P2P Dirigidas Por Dados

Nas redes de sobreposição dirigidas por dados, é a disponibilidade dos dados que definem as direções do fluxo, e não uma estrutura de sobreposição específica. Um nó sempre encaminha os dados para outros que estejam aguardando por este dado, sem utilização de regras prescritas do tipo pai/filho, interno/externo e

*downstreaming/upstreaming*. Esse mecanismo atinge a tolerância a falhas, adaptando-se a rápida mudança dos pares e conteúdos entre múltiplos fornecedores.

O DONet sugerido por Zhang et al., (2005), se encaixa neste tipo. Sua operação é bem simples: cada nó periodicamente troca informações sobre seus dados disponíveis com seus parceiros, recebe dados novos de um ou mais parceiros, ou fornece dados a eles. As três vantagens principais desse modelo são a facilidade de implementação, pois não é necessário montar e manter uma estrutura global complexa; a eficiência, já que o encaminhamento de dados é determinado dinamicamente de acordo com a disponibilidade, sem restrições a direções específicas; e a robustez e tolerância a falhas, onde os pares permitem uma troca rápida entre os diversos fornecedores.

### **3.4 Similaridades e Diferenças Das Propostas Baseadas em Árvore e Baseadas em Malha**

As propostas mencionadas adotam a diversidade de caminhos para resolver o problema da constante entrada e saída dos pares da rede de sobreposição *multicast*. A redundância por arestas cruzadas usa encaminhamento probabilístico enquanto a redundância interna da árvore adota a ideia de adicionar um caminho em uma única árvore *multicast*, que permite fácil otimização para a latência ou outras métricas e é fácil de implementar.

Enquanto a redundância em uma única árvore é frágil, a redundância por múltiplas árvores, baseada em redes em malha e baseada em rumores são resistentes a falhas de nós/enlaces. Mas, é difícil garantir baixa latência em sistemas usando a redundância baseada em rumores ou otimizar múltiplas árvores ao mesmo tempo e a coordenação entre as múltiplas redes de sobreposição é difícil.

Para o balanceamento de carga, uma única árvore não é tão boa quanto as outras abordagens. As propostas de redundância por múltiplas árvores, baseada em rumores e baseada em redes em malha podem alcançar uma alta taxa de transmissão, mas a largura de banda de uma única árvore da rede de sobreposição diminui ao longo da árvore, desde a raiz até os nós folhas.

As abordagens baseadas em árvore ou malha possuem uma grande quantidade de semelhanças. Primeiro, mesmo que essas abordagens utilizem algoritmos diferentes para construção de redes de sobreposição, a forma e seu resultado final são bem parecidos. Especificamente a visão de sobreposição de diversas árvores é na verdade a mesma visão de uma malha aleatória. Segundo, o fornecimento de conteúdo em ambas as abordagens permite aos pares receberem diferentes partes do conteúdo. No nível do par, cada par recebe conteúdo de vários pais e envia conteúdo para diversos filhos. No nível do sistema, o conjunto de arestas utilizadas para fornecer um único pacote da fonte para todos os pares participantes formam uma árvore que tem como raiz essa fonte.

A diferença chave entre as abordagens baseadas em árvore e malha é como a árvore de entrega de um determinado pacote é formada. Na abordagem baseada em árvore, a árvore de distribuição de cada pacote está, de fato, presa ao mecanismo de construção da árvore por causa do mapeamento estático de descrições (*descriptions*) de árvores. Em contraste, na abordagem baseada em malha a árvore de distribuição de pacotes individuais é dinamicamente formada de acordo com a travessia do pacote através da rede de sobreposição. A formação dinâmica da árvore de distribuição permite as propostas baseadas em malha utilizarem eficazmente os recursos disponíveis.

#### 4 SOBREVIVÊNCIA EM REDES MULTICAST P2P ESTÁTICAS

A tolerância a falhas (*resilience*) em uma rede de *multicasting* P2P deve ser tratada em dois casos, quando um nó falhar ou quando ele deixar a sessão *multicast* voluntariamente (FEI e YANG, 2007). A questão chave é como reconstruir a árvore de sobreposição após esses eventos dentro de um tempo adequado (FEI e YANG, 2007; BANERJEE et al., 2003; PADMANABHAN et al., 2003). Na transmissão *multicast* em redes IP, os nós intermediários na árvore de distribuição são roteadores, que são relativamente estáveis e não deixam a árvore *multicast*. Já em uma rede *multicasting* P2P, um nó corresponde a um processo executando em um *host* final, que é potencialmente mais suscetível a falhas do que um roteador, principalmente devido à possibilidade de o *host* final poder deixar a sessão *multicast* unilateralmente. Esses eventos causam a desconexão de nós que dependem do nó que falhou (ou que deixou espontaneamente a rede) até a reconstrução da árvore de *multicast*. Perdas devido a falhas nos nós da rede de sobreposição são mais significativos que perda de pacotes regulares e podem causar interrupção dos fluxos de dados por períodos na ordem de dezenas de segundos (BANERJEE et al., 2006).

A recuperação da árvore de *multicast* na rede de sobreposição pode ser realizada de duas maneiras: reativa e pró-ativa. No primeiro caso a restauração da árvore, que acontece somente após a ocorrência do evento, costuma levar um tempo maior para reparar a árvore de *multicast*. A dificuldade é encontrar um novo pai para cada nó desconectado, devendo contatar vários nós na árvore antes de encontrar uma locação apropriada (FEI e YANG, 2007). No segundo caso, a ideia básica para a recuperação é que cada nó não folha na árvore pré-calcula um plano de resgate antes das falhas ou partidas dos nós (FEI e YANG, 2007).

Ainda no caso de recuperação de forma ativa, é possível trabalhar com um método *off-line*, que executa um algoritmo para pré-calcular (antes da operação) um plano de restauração (WALKOWIAK e PRZEWOZNICZEK, 2011) e (WALKOWIAK, 2009). Assim, ao se iniciar o sistema, todos os pares de uma rede *multicasting* P2P, já conhecem todas as possíveis falhas e seus respectivos planos de restauração. Isso é uma vantagem, pois libera os pares de trabalhos extra no momento de uma falha, ou no momento em que estão transmitindo dados, ganhando em desempenho e até mesmo exigindo menor uso de recursos computacionais.

Desconsiderando a dinamicidade nas redes P2P, Walkowiak (2009) e Walkowiak e Przewozniczek (2011), assumem em seus trabalhos que o cenário de *multicasting* P2P é relativamente estático. Isto é, que os nós participantes têm interesse no sucesso da comunicação e que deixam a rede voluntariamente. Consequentemente, seus trabalhos focam nas falhas de rede, não considerando questões relacionadas ao dinamismo de sistemas P2P, como entradas e saídas de pares. Com base nessas premissas os autores argumentam que mecanismos de sobrevivência semelhantes aos estabelecidos na transmissão *unicast* podem ser usados em transmissões *multicast*. Duas estratégias são primordialmente empregadas na proteção de redes *unicast*: proteção e restauração (WALKOWIAK & PRZEWOZNICZEK, 2011) e (WALKOWIAK, 2009). Essas técnicas já foram amplamente discutidas na sobrevivência das redes convencionais e a distinção entre as duas consiste na escala de tempo diferente nas quais elas operam. Enquanto a proteção aloca recurso para a sobrevivência da rede antecipadamente, a restauração usa dinamicamente os recursos de sobrevivência. A principal vantagem da proteção é a rápida reação a falhas, garantindo tempo de restauração pequeno.

## 5 CONCLUSÕES

Neste trabalho, apresentou-se uma revisão de propostas de sobrevivência para redes *multicasting* P2P, de acordo com o *churn* causado pela falha do par/enlace, para topologias de redes de sobreposição baseados em árvores e baseados em malha. A tabela 1 compara as várias propostas.

A inspiração para este trabalho são os métodos de sobrevivência propostos para redes com pares participantes estáveis. Essa visão diferenciada de uma rede P2P permite que trabalhos avancem na construção de novos métodos de proteção para falhas na rede, não se preocupando com as constantes entradas e saídas no sistema dos pares, amplamente estudada em uma rede P2P dinâmica, como o caso das falhas dos nós intermediários, ou nós fonte.

Os trabalhos futuros apresentarão resultados para redes com pares participantes estáveis, por meio de um mecanismo *off-line* para proteção para sobrevivência de fluxos *multicast* em redes de sobreposição P2P.

Propostas	Baseada em Árvore			Baseado em Rumores	Baseada em Malha	Dirigida por Dados
	Conteúdos	Arestas Cruzadas	Interna na Árvore			
Permite Otimização	Sim	Sim	Não	Não	Não	Não
Balanceamento de Carga	Fraco	Fraco	Bom	Bom	Bom	Médio
Potencial de Taxa de Transmissão	Médio	Médio	Alto	Alto	Alto	Alto
Controle de Sobrecarga	Baixo	Baixo	Alto	Baixo	Alto	Médio
Implementação	Fácil	Fácil	Difícil	Fácil	Difícil	Fácil
Casos Representativos	Nice, PRM	Zigzag, Nemo	Splitstream, CoopNet, SuTeC	HON	PRIME	DONet

TABELA 1 – Comparação entre as várias propostas

## REFERÊNCIAS

- BANERJEE, S.; Lee, S.; Bhattacharjee, B.; Srinivasan, A. **Scalable Application Layer Multicast**. SIGCOMM'02, Aug. 19-23, Pittsburgh, Pennsylvania, USA. Copyright 2002 ACM 1-58113-570-X/02/0008.
- BANERJEE, S.; Lee, S.; Bhattacharjee, B.; Srinivasan, A. **Resilient multicast using overlays**. Proc. of ACM SIGMETRICS, Jun. 2003.
- BANERJEE, S.; Lee, S.; Bhattacharjee, B.; Srinivasan, A. **Resilient Multicast Using Overlays**. IEEE/ACM Transactions on Networking, Vol. 14, nº. 2, Apr. 2006, 063-6692, 2006.
- BIRNER, S.; Bustamante, F. E. **Resilient Peer-to-Peer Multicast without the Cost**. Proc. of MMCN, Jan. 2005.
- BIRNER, S.; Bustamante, F. E. **Resilience in Overlay Multicast Protocols**. Proc. of the 14th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, (MASCOTS '06) 0-7695-2573-3/06, 2006.
- BIRNER, S.; Bustamante, F. E. **A Comparison of Resilient Overlay Multicast Approaches**. IEEE Journal on selected areas in communication, vol.25, nº.9, Dez. 2007.
- CASTRO, M.; Druschel, P.; Rowstron, A.; Kermarrec, A.; Singh, A.; Nandi, A. **SplitStream: High-**

**Bandwidth Multicast in Cooperative Environments.** OSP'03, Oct. 19–22, 2003, Bolton Landing, New York, USA. ACM 1-58113-757-5/03/0010.

COUTINHO, G. L.. **Delay Tolerant Networks (DTN).** Rio de Janeiro, 2006. Universidade Federal do Rio de Janeiro – UFRJ - Escola Politécnica. Disponível em: [HTTP://WWW.GTA.UFRJ.BR/GRAD/06\\_2/GUSTAVO/INICIAL.HTM](http://www.gta.ufrj.br/grad/06_2/gustavo/inicial.htm). Acesso em 03 nov. 2011.

FEI, Z.; Yang, M. **A Proactive Tree Recovery Mechanism for Resilient Overlay Multicast.** IEEE/ACM Transactions on Networking, Vol. 15, nº. 1, Feb. 2007, 1063-6692.

KACIMI, M.; Yétongnon, K.; MA, Y.; Chbeir, R. **HON-P2P: A Cluster-based Hybrid Overlay Network for Multimedia Object Management.** Proc. of the 2005 11th International Conference on Parallel and Distributed Systems, (ICPADS'05) 0-7695-2281-5/05.

KUROSE, J. F.; Ross, K. W. **Redes de Computadores e a Internet – Uma abordagem Top-Down.** 5ª Edição. São Paulo: Pearson, 2010.

MAGHAREI, N.; Rejaie, R. **PRIME: Peer-to-Peer Receiver-driven MESH-based Streaming.** INFOCOM, 2007.

PANDMANABHAN, V. N.; Wang, H. J.; Chou, P. A. **Resilient Peer-to-Peer Streaming.** Proceedings of the 11th IEEE International Conference on Network Protocols, (ICNP'03)1092-1648/03, 2003.

TRAN, D. A.; Hua, K. A.; Do, T. T. **A Peer-to-Peer Architecture for Media Streaming.** IEEE Journal On Selected Areas In Communications, Vol. 22, nº. 1, Jan. 2004, 0733-8716/04.

WALKOWIAK, K. **Survivability of P2P Multicasting.** IEEE 7th International Workshop on the Design of Reliable Communication Networks, 978-1-4244-5048-0/09, 2009.

WALKOWIAK, K.; Przewozniczek, M. **Modeling and Optimization of Survivable P2P Multicasting.** Computer Communications, 34 (2011), p.1410–1424.

ZHANG, X.; Liu, J.; Li, B.; Yum, T.-S. P. **DONet/CoolStreaming: A Data-driven Overlay Network for Live Media Streaming.** Proc. IEEE Infocom, Miami, USA, Fev. 2005.