

# Fluxo P2P

E. Leonardi e M. Mellia, do Politécnico de Torino (Itália)  
C. Kiraly e R. Lo Cigno, da Universidade de Trento (Itália)  
S. Niccolini e J. Seedorf, dos Laboratórios de Pesquisa da NEC (Alemanha)

Os fluxos P2P impõem demandas extremamente altas sobre a rede. O artigo propõe uma arquitetura geral para streaming de vídeo ao vivo, que suporta as necessidades dos usuários e provedores de conteúdo. As aplicações do protótipo que seguem a arquitetura já foram implementadas e estão rodando em instalações de peers e em demonstrações que comprovam a viabilidade e o desempenho da solução.

Nos últimos anos, a tecnologia P2P - *peer-to-peer* tornou-se cada vez mais popular para aplicações de *streaming* de vídeo, incluindo serviços de TV (P2P-TV), como o SopCast, TVAnts, PPLive, UUSee e TVUplayer. O interesse da comunidade de pesquisa, dos provedores de conteúdo e das operadoras de rede também está em contínuo crescimento, embora provavelmente devido a diferentes razões. Os provedores de conteúdo enxergam aí uma nova oportunidade de conquistar usuários. Porém, ao mesmo tempo, preocupam-se com as ameaças impostas aos seus modelos comerciais padrão. As operadoras de rede preocupam-se, principalmente, com o estresse imposto por uma aplicação com tamanha demanda de largura de banda e tão sensível a atrasos. A comunidade de pesquisa é estimulada pelas oportunidades oferecidas pela distribuição e radiodifusão de P2P ao vivo, buscando tanto novas soluções técnicas quanto modelos comerciais inovadores.

O NAPA-WINE - Network Aware Application over WIsE Networks) é um projeto de três anos desenvolvido dentro do 7<sup>th</sup> Research Framework da Comissão Europeia. O objetivo é encontrar soluções inovadoras para o *streaming*

P2P ao vivo, reduzindo as dificuldades enfrentadas pelos provedores de conteúdo, ao mesmo tempo em que elimina as preocupações das operadoras de rede. Já se propôs a cooperação entre provedores de rede e aplicações P2P e grandes consórcios, como o projeto P4P, também a levam em consideração. Porém, esses trabalhos tratam, principalmente, de aplicações P2P de compartilhamento de arquivos, enquanto, em geral, desprezam o P2P-TV ou o *streaming* ao vivo, provavelmente devido às dificuldades inerentes ao trabalho com um sistema com demandas de desempenho tão difíceis.

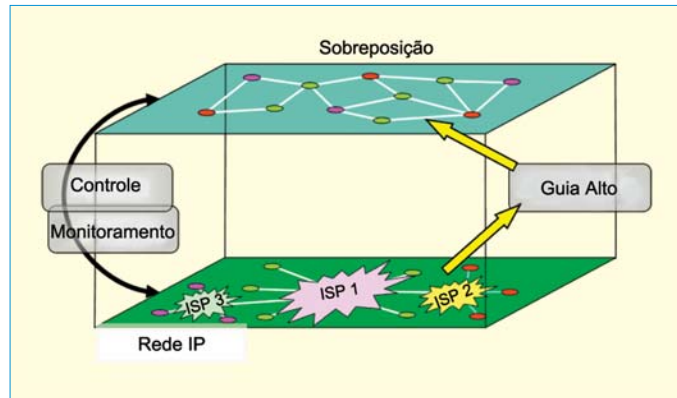
Em um sistema P2P-TV, uma fonte divide o fluxo de vídeo em blocos (*chunks*) de dados, que são trocados entre os nós, para distribuí-los a todos os *peers* participantes. Os *peers* formam uma topologia de sobreposição na camada de aplicação, onde os *peers* circunvizinhos se conectam e trocam blocos, com o uso da rede IP subjacente. A camada IP e de sobreposição são camadas de “rede”, pelo fato de ambas realizarem roteamento e encaminhamento de dados: pacotes na camada IP e blocos (normalmente uma sequência de pacotes) na camada de sobreposição.



Nesse contexto, o projeto NAPA-WINE propõe uma arquitetura P2P em cooperação com a rede, que, explicitamente, tem como meta a otimização da qualidade percebida pelos usuários, ao mesmo tempo em que minimiza o impacto sobre a rede de transporte. O NAPA-WINE não impõe qualquer estrutura sobre a topologia de sobreposição, que pode ser qualquer tipo de malha genérica. A distribuição de vídeo baseia-se em blocos, porém a construção dos blocos é livre o bastante para acomodar qualquer coisa, desde um único quadro de vídeo (até menos, se necessário) até grandes trechos de um vídeo, no caso de aplicação sob demanda. O foco é o projeto de um sistema que capacite a futura TV de alta qualidade P2P, um sistema onde um *peer* fonte produz o fluxo de vídeo, corta em blocos e os injeta na sobreposição, onde os *peers* cooperam para distribuí-los, sem a necessidade de a fonte ter enormes recursos e largura de banda para suportar o serviço.

A arquitetura que antevemos está esquematicamente representada na figura 1. A sobreposição e a rede IP interagem por meio das capacidades de monitoramento e de controle, com o objetivo de garantir boa qualidade para os usuários e uso eficiente de recursos para operadoras de rede. Um elemento adicional, não obrigatório,

ALTO com a contribuição de parceiros do NAPA-WINE, o provedor de rede recebe a capacidade de dirigir a aplicação P2P, por exemplo, publicando, explicitamente, as informações sobre a condição de sua rede, como congestionamento de enlace ou preferências de roteamento AS.



**Fig. 1 - A visão do NAPA-WINE - roteadores IP e nós de sobreposição (peers) cooperam através das capacidades de monitoramento e de controle e da interação de ALTO para otimizar o desempenho e a utilização da rede**

mas definitivamente útil para a otimização do desempenho, é a presença de um guia ALTO - Otimização de Tráfego de Camada de Aplicação. Como objetivado pelo IETF, no grupo de trabalho

informações coletadas pela ferramenta de monitoração podem ser usadas para engatilhar a reconfiguração da sobreposição ou para orientar a esquematização do protocolo de distribuição de blocos.

A TRADIÇÃO DA MARCA COM A QUALIDADE E DISPONIBILIDADE DE MATERIAL E ATENDIMENTO.

**DISTRIBUIDOR AUTORIZADO:**

**COMMSCOPE®**  
AUTHORIZED DISTRIBUTOR

**SOLUÇÕES PARA:**

- SOLUÇÕES EM COBRE CAT5E E CAT6
- SOLUÇÕES EM FIBRA ÓPTICA OM3 E OM4 E ACESSÓRIOS
- RACKS, GABINETES E ORGANIZAÇÃO DE CABOS

**Conheça nossa empresa!**  
[www.laax.com.br](http://www.laax.com.br)  
[telecom@laax.com.br](mailto:telecom@laax.com.br)  
 11 2374-0300



## Visão geral da arquitetura

A arquitetura apresentada na figura 2 baseia-se em quatro blocos construtivos principais, mais o servidor externo ALTO, que pode suportar o gerenciamento da topologia, fornecendo informações que não se podem medir na camada de aplicação. A seguir, apresentamos um breve resumo do papel e dos recursos principais de cada bloco.

### Camada de usuário

A camada de usuário é principalmente responsável pela codificação do vídeo e seu empacotamento em blocos (na fonte) e pela decomposição dos blocos e decodificação, nos receptores. A camada de usuário pode usar as ferramentas padrão de codificação, como ffmpeg, cujo objetivo não é

pequenos, impondo altas demandas sobre o intercâmbio de blocos, a não ser que se construa uma sobreposição estruturada; e a segunda leva a uma severa degradação quando um bloco se perde, uma vez que as bordas do quadro são violadas e, normalmente, os participantes lidam mal com essa situação. O formador de blocos da camada de usuário permite técnicas inteligentes de formação de blocos, como por exemplo coletar quadros com diferentes recursos (por exemplo, quadros I, P e B de fluxos MPEG) em blocos separados, que também podem ser tratados com prioridades adequadas pelo negociador de blocos.

Quando o *peer* age como receptor apenas, o módulo de usuário remonta o fluxo de vídeo a partir de uma série de blocos recebidos, de forma que, após a decodificação, pode-se exibir o vídeo.

principais fatores para o funcionamento do sistema é a escolha dos *peers* com os quais se comunicar, ou seja, a seleção da vizinhança.

A figura 3 mostra a relação lógica entre as funções que compõem o gerenciamento de sobreposição e as que compõem o intercâmbio de blocos. As interfaces na direção de outras camadas e serviços externos, como ALTO, também são indicadas.

### Preenchendo o banco de dados de vizinhança

Preenche-se o banco de dados de vizinhança (em verde, com linhas grossas, na figura 3) tão logo um *peer* participa da distribuição de um canal de TV. Por uma questão de clareza, descrevemos a arquitetura como se houvesse apenas um canal, mas, como mostra a figura 2, pode haver múltiplas camadas de usuário e sobreposições para a gestão de múltiplos canais. Analisando a figura 3, os componentes em amarelo (linhas finas) estão relacionados à esquematização, transmissão e recepção de blocos, descritos mais tarde, enquanto os componentes em azul (linhas pontilhadas) referem-se à gestão da topologia e à sinalização em geral (intercâmbio de mapas de retenção, ou seja, a lista de blocos disponíveis a cada par, disponibilidade a blocos de serviços, etc.). As duas funcionalidades interagem por meio do banco de dados da vizinhança, assim como o *buffer* de bloco (ou seja, a estrutura onde os blocos são armazenados para intercâmbio e antes da transmissão) e os mapas de retenção correspondentes dos vizinhos.

A rede de sobreposição em sistemas P2P é o resultado de um algoritmo distribuído, que constrói e mantém a vizinhança de cada *peer*. Quando um *peer* junta-se ao sistema, seleciona seu conjunto inicial de vizinhos, então o conjunto de vizinhos de todo nó do sistema é dinamicamente otimizado ao longo do tempo.

A fonte ou um servidor da web pode ajudar a fase de *bootstrapping*, onde o usuário seleciona um canal, o

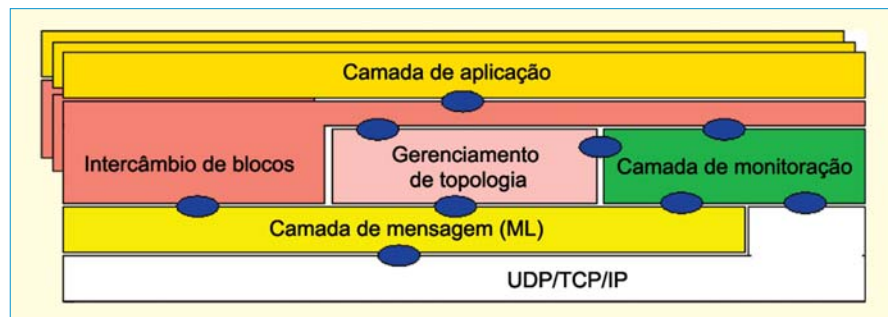


Fig. 2 - Arquitetura lógica de um peer NAPA-WINE

implementar um codificador de vídeo proprietário, mas suportar tantos formatos padrão quanto possível (MPEG1/2/4, H.264, etc.). Dependendo do tipo de fonte do vídeo, isso pode incluir conversão analógico/digital, codificação e qualquer outra manipulação de vídeo que o provedor de conteúdo queira fazer, como introdução de propaganda. Um recurso fundamental e diferencial da camada de usuário do NAPA-WINE é o suporte à flexibilidade da formação dos blocos. A maioria dos padrões de sistema P2P-TV tende a mapear um quadro em um bloco ou separar cegamente o fluxo em blocos de igual tamanho, alheia a quaisquer recursos de mídia. Ambas as soluções são sub-ótimas: a primeira pode levar a blocos muito

Várias instâncias da camada de usuário podem ser combinadas dentro da mesma aplicação, permitindo que se veja um canal e grave outro, ou que se tenham visualizações de diferentes canais na tela.

### Amostragem de par e gerenciamento de topologia

Um cliente P2P-TV precisa se comunicar de forma muito eficiente com outros *peers*, para receber e redistribuir a imensa quantidade de informações embutidas em um fluxo de vídeo. As informações devem chegar quase em tempo real e com pequena variação de atraso.

O objetivo da aplicação é, então, entregar todas as informações do vídeo a todos os *peers* no sistema, no menor tempo possível. Um dos





qual pode se comportar como um rastreador do tipo torrent de bit. A manutenção da topologia baseia-se em um protocolo do tipo *gossip*,

como o Newscast ou o Cyclon (ambos suportados e implementados nos *peers* NAPA-WINE) que permitem a descoberta de *peers* na sobreposição.

Uma vez que se descobrem os *peers*, obtém-se a gestão ótima da topologia por meio de uma implementação *ad hoc* de Tman, que foi especificamente preparado e otimizado para *streaming* ao vivo e para interagir com serviços ALTO quando necessário.

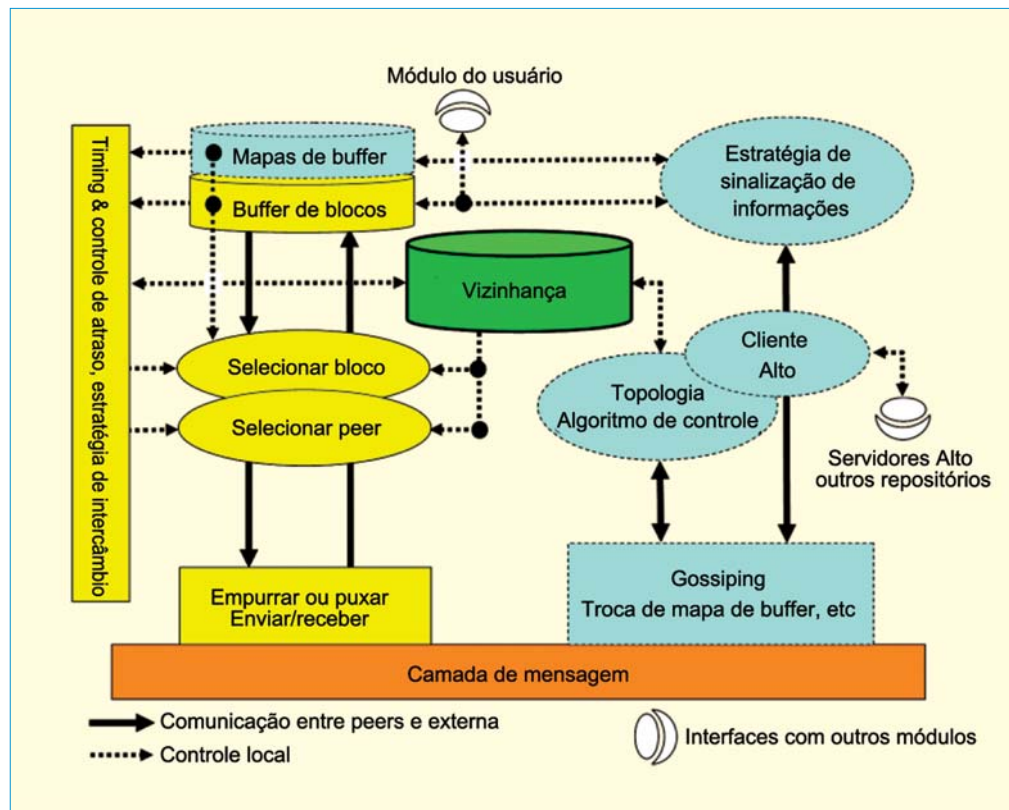


Fig. 3 - Arquitetura detalhada do gerenciamento de topologia e as camadas lógicas de intercâmbio de blocos

### Suporte de ALTO

O gerenciamento de topologia na arquitetura NAPA-WINE suporta totalmente a diretriz ALTO, por meio da integração de um cliente ALTO dentro do gerenciador de topologia (figura 3). A otimização da camada de tráfego da camada de aplicação é uma abordagem inovadora, que permite que as operadoras de rede tenham economia operacional, reduzindo a quantidade do tráfego de backbone da camada de aplicação. Essencialmente, o ALTO é um serviço dedicado, operado por ISPs, provedores de conteúdo ou provedores

## QUER SEGURANÇA, GRAVE NO VPON!



VP408G4/R VP416G4/R



VP4048BR



VPLITE 2



**BYCON**  
Soluções em tempo real

BYCON S/A  
(11) 5096-1900 - Conheça toda a linha em [www.bycon.com.br](http://www.bycon.com.br)  
[info@bycon.com.br](mailto:info@bycon.com.br)

**VISITE NOSSO STAND NA ISC - 24 A 26 DE ABRIL E EXPOSEC - 08 A 10 DE MAIO.**

independentes, que fornecem informações úteis sobre a camada de rede para clientes da camada de aplicação, sobre custos e recursos. O IETF formou um grupo de trabalho e, atualmente, está padronizando um protocolo ALTO. A figura 4 mostra a ideia geral subjacente a um serviço como o ALTO. Suponha-se que o cliente 2, na figura, queira conectar-se a um fluxo. O protocolo *gossip* oferece ao cliente 2 vários candidatos a *peer*, que podem oferecer o fluxo desejado. Na figura, o cliente 2 pode conectar-se ao cliente 1 ou ao cliente 3. O cliente 2 consulta um serviço ALTO para obter orientação sobre que cliente selecionar. O serviço ALTO pode responder a consultas com base nas informações fornecidas pelo ISP dos clientes ou pelo provedor de conteúdo. As consultas podem dizer respeito à topologia, estado de roteamento,

políticas ou custos operacionais. No exemplo apresentado, é provável que o serviço ALTO irá sugerir o cliente 3, porque este está fisicamente localizado na mesma rede que o cliente 2 e a sua escolha reduziria o tráfego entre AS.

#### Intercâmbio de blocos

O problema de intercâmbio de blocos está estritamente relacionado com a gestão de topologia, e é essa a razão pela qual os blocos lógicos foram apresentados juntos na figura 3. O objetivo do intercâmbio de blocos é receber o fluxo de forma suave (e com pequeno atraso) e cooperar no procedimento de distribuição. Assumimos que os *peers* são honestos e cooperativos, uma vez que o foco do NAPA-WINE é a otimização do desempenho e do sistema, porém, estamos todos bem cientes dos problemas relativos à segurança, privacidade e cooperação.

A transferência de blocos é a operação que mais afeta o desempenho e a conveniência para a rede. Inclui problemas de protocolo e de algoritmo. Primeiro de tudo, os *peers* precisam trocar informações sobre sua condição atual, para permitir a programação das decisões. As informações trocadas referem-se às condições do *peer*, com relação ao fluxo, ou seja, um mapa dos blocos necessários para que um *peer* transmita o fluxo de forma contínua. O bloco de estratégia de sinalização de informação (figura 3) realiza essa tarefa. Esse bloco está a cargo do envio de mapas de retenção para outros nós, com um *timing* adequado; da recepção desses mapas de outros nós e da reunião das informações no banco de dados do mapa de buffer; de negociar se essas e outras informações devem se disseminar por protocolos *gossip* ou não, e até que profundidade devem ser disseminadas na topologia.

## SOLUÇÕES DIAMOND

A Diamond do Brasil conecta pessoas, processos e sistemas de forma inovadora por meio de tecnologias ópticas de alta precisão. Nossa equipe antecipa tendências de mercado, oferecendo soluções inovadoras, confiáveis, robustas e com máxima performance para os segmentos de telecomunicações, indústria, data centers e redes locais, entre outros.

Visite → [www.diamond-brasil.com.br](http://www.diamond-brasil.com.br) e conecte-se com as tecnologias ópticas de última geração.

Diamond, há mais de 20 anos fornecendo soluções ópticas de alta performance para **Telecomunicações, Indústrias, Data Centers, Ambientes Comerciais e Residenciais.**  
Rio de Janeiro 21 3083 2000 • São Paulo 11 3284 6003 • Minas Gerais 31 2535 5779 • [diamond@diamond-brasil.com.br](mailto:diamond@diamond-brasil.com.br) • [www.facebook.com/diamonddobrasil](http://www.facebook.com/diamonddobrasil)



#### COMPONENTES ATIVOS



Com versões Fast Ethernet e Gigabit Ethernet, nossa linha de switches oferece uma solução altamente confiável para a conexão das mais modernas redes de comunicação, incluindo soluções FTTO.

#### COMPONENTES PASSIVOS

Conexões ópticas E-2000™. Proteção do ferrinho óptico, codificação mecânica, e por cores, no corpo do conector.





Além da troca do mapa de buffer, a sinalização inclui oferta/pedido/seleção de primitivos usados para o intercâmbio de blocos. Essas mensagens podem ser *piggybacked* em blocos, por uma questão de eficiência.

Outra decisão importante do protocolo diz respeito às informações de *pushing* (empurrar) ou *pulling* (puxar). Empurra-se um bloco quando o *peer* que possui o bloco decide oferecê-lo a algum outro *peer* e puxa-se um bloco quando um *peer* que precisa do bloco pede-o a outro *peer*. De um ponto de vista teórico, empurrar é mais eficaz.

Independentemente do protocolo e da estratégia de sinalização usados, o

de blocos e o gerenciador de sobreposição podem explorar as informações oportunas sobre a qualidade da conectividade entre *peers*, coletadas em tempo real pelos módulos de monitoração. Isso inclui, mas não se limita, a distância e largura de banda disponível entre dois *peers*, ou a presença de NAT - Tradução de Endereço de Rede. A camada de monitoração tem dois modos de operação: passivo e ativo. As medições passivas são realizadas por observação das mensagens trocadas entre *peers*. As medições ativas criam mensagens especiais de sondagem, que são enviadas para outros *peers*, a critério da camada de monitoração. O projeto e a realização do módulo

disponíveis aos outros *peers*, os quais podem delas se beneficiar. As informações adquiridas por cada *peer* são localmente armazenadas e compartilhadas no banco de dados da vizinhança. Também são resumidas e exportadas (publicadas) para outros *peers* ou para um repositório externo do qual podem ser recuperadas por outros *peers*, por exemplo, no *bootstrapping*.

Pode-se ver um serviço ou guia ALTO como um caso especial de repositório externo, que são, na realidade, uma generalização do conceito ALTO. Também podem ser vistos os repositórios locais como uma abstração da interação do gerenciamento de topologia e do

intercâmbio de bloco com o módulo de monitoração.

### Camada de mensagens

A camada de mensagens (a caixa inferior na figura 3) oferece primitivas a outros módulos para envio e recepção de dados de/para outros *peers*. Ela fornece uma interface abstrata para o transporte de protocolos (UDP/TCP) e dos pontos de acesso a serviços correspondentes, oferecidos pelo sistema operacional, por meio da

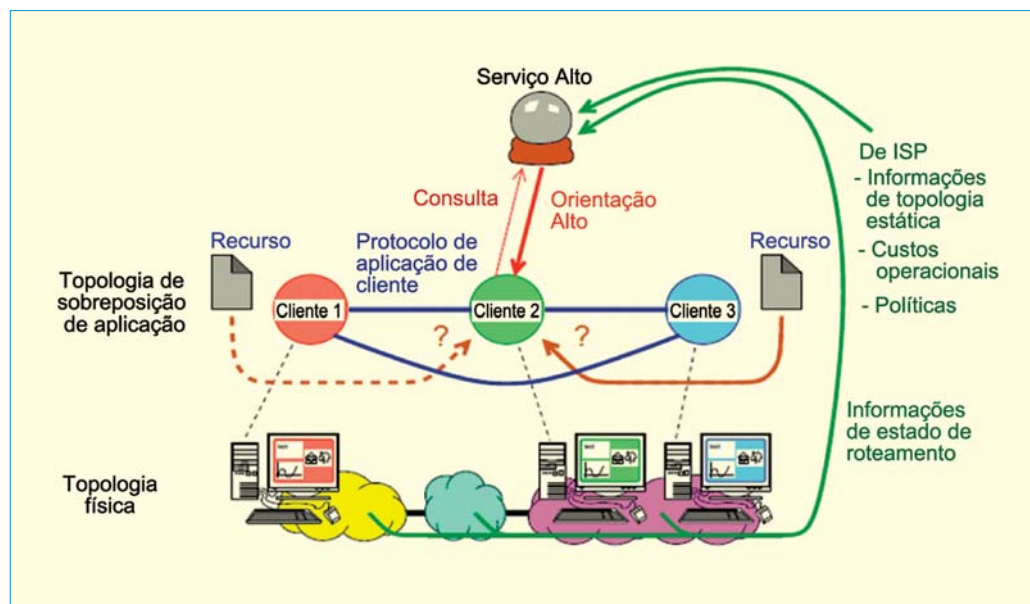


Fig. 4 - Seleção de peer habilitada pelo ALTO

núcleo de um agendador de tarefas é o algoritmo para escolher os blocos a serem trocados e os *peers* com os quais se comunicar. Foram estudadas muitas estratégias diferentes, incluindo ambos os algoritmos fundamentais e sua adaptação aos cenários heterogêneos, múltiplos subfluxos, etc. Os algoritmos resultantes estão disponíveis no protótipo com o qual podem ser testados.

### Camada de monitoração

Além das informações fornecidas pelo servidor ALTO, o programador

de monitoração são soluções inovadoras do NAPA-WINE, pois muitas das suas medidas, como a largura de banda disponível em sistemas de grande escala, estão longe de ser triviais.

### Repositórios

“Repositórios” são os bancos de dados de nível de aplicação, onde as informações sobre cada *peer* são compartilhadas por todos os outros *peers*. Sem dúvida, as informações geradas pelo módulo de monitoração não são apenas exploradas pelo *peer* local, mas também se fazem

extensão de suas capacidades e do endereçamento de nível de aplicação, ou seja, atribuindo um único identificador para cada *peer*. Por exemplo, fornece a capacidade de enviar um bloco para outro *peer*, que precisa ser segmentado e depois remontado para caber nos datagramas UDP. A camada de mensagens também fornece uma interface para o módulo de monitoração, invocada para medições passivas: sempre que uma mensagem é enviada ou recebida, será dada uma indicação ao módulo de monitoração, que pode então atualizar as suas estatísticas. O



último recurso importante da camada de mensagens são os mecanismos para a passagem de NAT. O NAT permite anexar vários computadores à Internet, usando apenas um endereço IP globalmente único. Portanto, é muito popular com clientes privados, que também são a principal audiência para P2P TV. Entretanto, a presença de um dispositivo NAT pode evitar que os *peers* estabeleçam conexões com outros *peers*. Portanto, as funções de passagem especial de NAT são oferecidas pela camada de mensagens.

## Situação de implementação

O projeto NAPA-WINE está implementando todos os recursos e blocos lógicos descritos anteriormente. Implementam-se o

algoritmos e arquiteturas, promovendo o desenvolvimento de novas ideias e aplicações.

Outras partes, como a camada de monitoração, são liberadas como bibliotecas isoladas, que podem ser publicadas também para aplicações P2P de *não-streaming*.

Vários *streamers* já foram implementados, explorando diferentes algoritmos e técnicas.

Estritamente relacionado à avaliação de QoS, um conjunto de ferramentas denominado PSNR para a avaliação da qualidade objetiva de vídeo, em termos de PSNR (relação pico de sinal e ruído) e outras medidas similares de reconstrução do fluxo de vídeo, começando a partir de um vestígio dos blocos recebidos e comparando-o ao vídeo original não codificado. Essas ferramentas podem ser usadas tanto em conjunto com as

arquitetura geral ainda se inclina para aplicações de compartilhamento de arquivo e opera sem qualquer coordenação com a rede IP, o que quase sempre resulta numa sofrível utilização de recursos e até mesmo em desperdício. Isso impedirá que suportem a TV de alta qualidade do futuro, ou que façam a transição para a alta definição, o que requisitará vários Mbit/s por *peer*.

Este artigo discutiu a arquitetura NAPA-WINE para um sistema P2P-TV, que está em desenvolvimento no projeto, com o objetivo de eficiência e cooperação entre a aplicação e as operadoras de rede e provedores de conteúdo. Protótipos de *peers* e do sistema já estão rodando na Internet e foram demonstrados em locais importantes. Os recursos importantes são a presença de monitoramento contínuo e controle de sobreposição, obtidos também com o suporte de ALTO, cujas diretrizes permitem a troca de informações entre a rede e a aplicação.

O software e os protótipos implementados no projeto NAPA-WINE são oferecidos como bibliotecas de software sob licença GPL ou LGPL: estão evoluindo e podem ser baixados gratuitamente a partir do projeto ou de web sites de parceiros.

O gerenciamento de topologia de sobreposição e a programação de blocos de informações foram identificadas como importantes recursos para que a aplicação faça uso conveniente da rede. A primeira função permite a construção de topologias de sobreposição eficientes e racionais, que são corretamente mapeadas sobre a estrutura da rede de transporte (por exemplo, considerando o número mínimo de saltos entre vizinhos, os sistemas autônomos com relação à localidade, etc.). A segunda função garante que a capacidade da rede seja explorada sem desperdício (por exemplo, com a minimização das retransmissões e buscando uma distribuição eficiente de blocos, etc.).

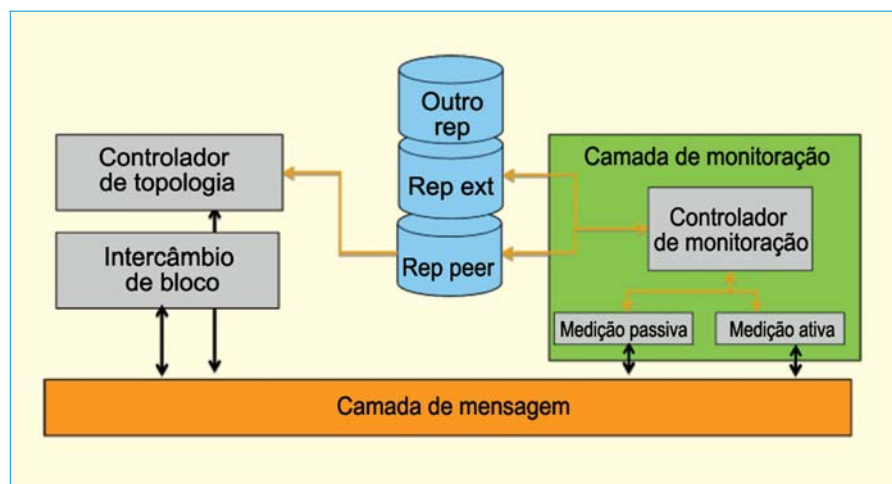


Fig. 5 - Arquitetura detalhada dos módulos de monitoração

gerenciamento de topologia e o intercâmbio de blocos oferecendo diferentes instâncias e algoritmos para fins experimentais. Isso se obtém por meio de projeto de um kit de ferramentas de desenvolvimento e um conjunto de bibliotecas, chamadas de GRAPES - Generic Resource Aware P2P Environment for Streaming. A GRAPES é inteiramente escrita em C, de modo que a reutilização e a ligação com qualquer linguagem são fáceis. Fornece um conjunto de blocos de construção que os pesquisadores podem usar, combinar e modificar para testar e comparar

aplicações reais (dado que o fluxo de vídeo original esteja de alguma forma disponível, por exemplo, por meio do envio do vestígio dos blocos recebidos/faltosos na fonte), como também acopladas com simuladores desenvolvidos para algoritmos preliminares para teste no NAPA-WINE.

## Conclusão

As aplicações de *streaming* de vídeo que exploram P2P como um paradigma de comunicação são uma realidade comercial, mas a sua