

Combate a Poluição em Sistemas P2P de Mídia Contínua ao Vivo

Alex Borges, Jussara Almeida, Sérgio Campos

¹ DCC – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brasil

{borges, jussara, scampos}@dcc.ufmg.br

Abstract. *Peer-to-peer (P2P) live streaming media systems are becoming more popular each day. As in file sharing P2P system, they are susceptible to content pollution attack. In this attack, a peer alters the media content decreasing the perceived quality of the streaming. In this paper we evaluate the impact of content pollution attack in P2P live streaming and we present two reputation system to stop polluted content dissemination and isolate malicious peers. Our results show that a few number of polluters is capable to compromise all the application and, the two proposed reputation system can rapidly identify and isolate polluters and also be resistant to peers collusion.*

Resumo. *Sistemas de transmissão de mídia contínua ao vivo em arquiteturas par-a-par (P2P) se tornam cada vez mais populares e, de forma similar ao compartilhamento de arquivos em P2P, estão sujeitos a ataques. Um tipo de ataque comum é a disseminação de conteúdo poluído, onde um par altera o conteúdo da mídia degradando a qualidade percebida pelos demais pares. Neste artigo nós avaliamos o impacto de poluição em sistemas de mídia contínua ao vivo P2P. Apresentamos duas abordagens de reputação para impedir a disseminação de poluição e isolar os pares maliciosos do sistema. Os resultados indicam que um número pequeno de nodos maliciosos é capaz de comprometer toda a aplicação e, as abordagens propostas conseguem detectar e isolar pares maliciosos rapidamente e com resistência ao conluio dos atacantes.*

1. Introdução

Sistemas de distribuição de mídia contínua ao vivo na Internet são cada vez mais populares. Grandes canais de TV, como NBC¹, transmitem parte da programação diária através de seus sites na Internet. Além deste, vários outros sites, como FreeTV², oferecem uma grande diversidade de canais de mídia contínua ao vivo. Tal distribuição de conteúdo sofre principalmente em escalabilidade, por se basear no modelo tradicional cliente-servidor.

Transmissões de fluxo contínuo ao vivo com a utilização de arquitetura par a par (P2P), também conhecidos por *Peer-to-Peer (P2P) Live Streaming Systems*, estão sendo utilizadas como uma alternativa aos sistemas tradicionais de distribuição de conteúdo ao vivo de vídeo pela Internet. Em comparação ao modelo tradicional cliente-servidor, a abordagem P2P ultrapassa limitações como a banda de rede e a escalabilidade do sistema em relação ao número de clientes.

¹<http://www.nbc.com/>

²<http://www.freeetv.com/>

Nesses sistemas, há uma fonte sorvedora que codifica a mídia original e a divide em pequenos pedaços que serão transmitidos pela rede. Os pequenos pedaços, conhecidos como *chunks*, são distribuídos pelos vários clientes participantes da rede P2P. Os clientes, também conhecidos como pares ou nodos, reenviam conteúdo de vídeo recebido para outros clientes na rede, eliminando assim a necessidade de servidores poderosos ou uma imensa banda de rede em um único ponto. Tal serviço apresenta características muito peculiares como uma grande restrição no tempo de entrega esperado por um pedaço ou trecho da mídia, que está sendo transmitida pela rede. Neste ambiente, não seria aceitável que um trecho de um jogo de futebol chegue 30 segundos atrasado, principalmente se este for o momento do gol em uma final de campeonato.

Várias técnicas já foram propostas para organizar e estruturar os participantes do sistema P2P, a fim de alcançar um grande número de clientes e também um baixo atraso percebido na mídia. Existem diversas maneiras de interação entre os clientes e repasse do conteúdo de mídia ao vivo. A maioria dos sistemas populares, como PPLive, PPStream e GridMedia [PPLive 2007, PPstreaming 2007, Zhang et al. 2005, Hei et al. 2006], usam uma técnica de reenvio orientada por pedido de dados e estruturam seus participantes baseando-se em malha - *mesh-pull overlay network*. Estes sistemas funcionam de maneira similar ao sistema de compartilhamento de arquivos Bittorrent [Bittorrent 2007], onde o conteúdo do vídeo é quebrado em pequenas partes chamadas *chunks*, e, quando um novo cliente se junta à rede, este faz parceria com um subconjunto de clientes que já estão assistindo o conteúdo ao vivo. Os clientes trocam entre si mapas de *chunks* que anunciam os dados disponíveis e os desejados, e desta forma, eles realizam trocas de dados de acordo com suas necessidades e interesses.

De maneira geral, os sistemas populares P2P de vídeo assumem um comportamento altruísta e não malicioso de seus participantes e, pelo que sabemos, nenhum dos mais populares assumem que o conteúdo do vídeo pode ser alterado ou forjado durante sua transmissão [Dhungel et al. 2007]. Assim, um nodo mal intencionado pode alterar ou injetar mensagens falsas na mídia que é transmitido ao vivo, tornando parte da transmissão inútil. Parceiros ingênuos podem requisitar e retransmitir o conteúdo poluído, consumindo recursos do sistema com tráfego de dados indesejáveis.

Poluição de dados é comum em sistemas de compartilhamento de arquivo P2P, onde algumas entidades tentam parar a disseminação de arquivos, danificando-os. A mesma estratégia de conter disseminação de conteúdo pode ser tentada em sistemas de vídeo ao vivo em redes P2P. Mais ainda, o comportamento malicioso pode ser encorajado por concorrência entre provedores de serviço de vídeo, ou mesmo pela simples diversão de destruir um sistema popular.

Em redes P2P de compartilhamento de arquivo há o esforço para combater poluição e outros comportamentos maliciosos. Abordagens práticas foram implementadas e são apresentadas em trabalhos como [Costa et al. 2007, Walsh and Siner 2005, Damiani et al. 2002]. Entretanto, não há avaliações sobre a aplicação de tais abordagens ao contexto de mídia contínua ao vivo, e uma aplicação direta dessas abordagens pode não ser viável pelas diferenças entre as aplicações.

Apresentamos neste artigo, através de simulação, a extensão do dano em condições de ataque de poluição que um sistema de transmissão de mídia contínua ao

vivo em P2P pode sofrer. Também apresentamos alterações aos sistemas de reputação utilizados em compartilhamento de arquivo em P2P e avaliamos sua eficiência para banir os pares maliciosos e conter a disseminação do conteúdo indesejado.

Nós avaliamos, através de simulação, duas técnicas presentes em sistemas de compartilhamento de arquivos em P2P aplicados ao contexto de distribuição de mídia ao vivo. Os resultados mostram que uma abordagem centralizada de reputação com lista negra pode identificar e banir rapidamente um nodo malicioso. Porém, tal abordagem pode ser vulnerável a conluio de nodos maliciosos além de inibir a reabilitação de pares que foram mal identificados como maliciosos. A abordagem descentralizada utilizada apresenta uma rápida identificação dos nodos maliciosos e se mostrou imune ao conluio de pares maliciosos. Mais ainda, a carga adicional necessária para que tal sistema seja implementado é relativamente baixa, se comparada à retransmissão imposta quando o conteúdo poluído é barrado e tenta-se reaver o trecho de mídia perdido.

O restante deste artigo está organizado da seguinte maneira: a seção 2 apresenta os trabalhos relacionados na área. A seção 3 discute a poluição de conteúdo em sistemas de distribuição de mídia contínua ao vivo em P2P, evidenciando o quão danoso é este tipo de ataque. Apresenta ainda, os mecanismos de defesa adotados para banir nodos maliciosos e conter a disseminação do conteúdo poluído. O modelo simulado, a metodologia utilizada e os principais resultados, são apresentados nas seções 4 e 5. A seção 6 mostra as conclusões deste trabalho e possíveis trabalhos futuros.

2. Trabalhos Relacionados

Os trabalhos relacionados a comportamentos maliciosos em distribuição de mídia contínua em P2P concentram esforços no tratamento de participantes egoístas e negação de serviço. Trabalhos como [Jin et al. 2006, Conner et al. 2006, Zhong 2006] apresentam propostas para reputar pares em um sistema deste tipo, banindo pares que contribuem pouco com o sistema. Um sistema de reputação neste contexto, consiste em um sistema capaz de avaliar a atuação de um nodo na rede e, através de seu comportamento passado, atribuir-lhe uma nota. Pares com baixa relação *upload/download* tem menor prioridade em relação aos que possuem maior taxa e, dessa forma, pretende-se incentivar um comportamento altruísta dos participantes do sistema. Além disto, as três propostas, assim como grande parte dos sistemas de reputação em compartilhamento de arquivos, ponderam a nota enviada de um par em relação ao seu parceiro pela sua própria nota. Assim, pares confiáveis no sistema têm maior credibilidade no momento de reputar um parceiro.

Vários trabalhos que visam combater conteúdo poluído foram realizados em aplicações de compartilhamento de arquivos P2P. Sistemas como Credence [Walsh and Siner 2005], apresentam uma abordagem distribuída, na qual os participantes da rede assinalam reputação aos *objetos* descarregados em relação à sua autenticidade. Outros como Scruber [Costa et al. 2007], identificam e isolam pares maliciosos que ativamente disseminam conteúdo poluído.

Entre os primeiros trabalhos que tratam comportamento por poluição dos participantes da rede, além de negação de serviço, estão [Haridasan and van Renesse 2006, Maya Haridasan 2007]. Nesses trabalhos são apresentadas comparações entre quatro alternativas para verificar a integridade do dado distribuído no sistema de mídia contínua ao vivo em P2P. Em [Dhungel et al. 2007], é apresentado um experimento, no qual um

poluidor ativo é colocado em um sistema real e os resultados obtidos indicam que ataques de poluição podem destruir um sistema ao vivo de vídeo em P2P. Além disso, são sugeridas algumas técnicas possíveis para verificar a integridade do fluxo de mídia distribuído. Os três trabalhos mostram que há a possibilidade de marcar os dados do fluxo contínuo e verificar a sua integridade com um baixo custo adicional. Utilizando os esquemas propostos, há um custo de processamento em $O(n)$ para sinalizar os blocos de dados a serem transmitidos, onde n é a quantidade de *chunks* contida em cada bloco de verificação, e $O(1)$ para a verificação a ser realizada pelo cliente. O custo adicional de banda de rede é cerca de 5% da banda necessária para o envio do fluxo original. Apesar das propostas sugeridas, os trabalhos não avaliam o ataque de poluição com uma visão abrangente da estrutura de rede P2P, e também, não apresentam avaliações de sistemas de reputação com intenção de combater a disseminação de poluição.

3. Mecanismos de Defesa e Combate a Poluição

Esta seção descreve duas abordagens simuladas para combate e defesa contra poluição em sistemas de mídia contínua ao vivo em P2P. As duas propostas são inspiradas em outras, adotadas em sistemas de compartilhamento de arquivos em P2P. A primeira, é uma abordagem uma centralizadora de reputação com lista negra e a outra é distribuída.

Um aspecto importante a ser considerado, é a maneira como o conteúdo é classificado em poluído ou limpo. No contexto deste trabalho, qualquer dado corrompido é considerado poluído, mesmo que não tenha sido intencionalmente modificado. Consideramos que qualquer uma das técnicas de marcação e verificação de *chunks*, sugeridas em [Haridasan and van Renesse 2006, Maya Haridasan 2007, Dhungel et al. 2007], possa ser utilizada. Nós avaliamos apenas a sobrecarga relativa a troca de informações necessárias para o funcionamento dos sistemas de reputação em questão.

3.1. Blacklist

Uma abordagem de lista negra é uma maneira simples de reputar parceiros e obter informações sobre os pares participantes da rede P2P. Nesse tipo de abordagem, os participantes monitoram o comportamento de seus parceiros e, periodicamente, reportam o comportamento observado à um servidor. A idéia principal é determinar quais são os nodos que originam o conteúdo poluído, e impedir que os repassem adiante. Isto é importante porque a identificação do conteúdo poluído não é suficiente para impedir a perda de qualidade experimentada pelos participantes da aplicação de mídia contínua.

Com *Blacklist* centralizado, os participantes do sistema consultam o servidor de reputação para obter informações sobre seus parceiros e obtêm como resposta, a reputação dos nodos requisitados. Com base na nota de reputação e uma classificação mínima aceitável, os participantes podem decidir realizar parcerias ou trocas de *chunks*.

Da mesma forma que é realizada em combate a nodos egoístas [Jin et al. 2006, Wang et al. 2006], a nota reportada por um nodo ao servidor de lista negra é ponderada proporcionalmente a sua própria reputação e assim, a reputação final de um nodo é a média ponderada de todas as reputações já enviadas, conforme mostrado na equação 1. A maneira como um nodo assinala a reputação de um parceiro pode ser realizada conforme a equação 2, que será detalhada adiante.

$$R_j = \left\{ \begin{array}{l} \sum_{k \in N} I_{k(j)} * R_k \\ \sum_{k \in N} R_k \end{array} \right. \quad (1)$$

Na equação 1, N é o conjunto de nodos que reputam o nodo j , R_j é a reputação de um nodo j do sistema de mídia ao vivo em P2P e, $I_{k(j)}$ é a reputação reportada por K , sobre o nodo j .

Feita a adoção de um servidor centralizado de lista negra, o sistema de reputação pode sofrer com escalabilidade, tolerância a falhas, problemas de segurança e, principalmente, com problemas relativos a ataques, que são facilitados pela existência de um ponto centralizador. Os poluidores também podem tentar enganar o sistema de reputação sendo menos agressivos na maneira que anunciam e enviam o conteúdo poluído, ou se associando a outros poluidores para obter notas de reputação melhores. Nossos resultados mostram que a centralização da reputação está sujeita a conluio dos participantes.

3.2. StRepS

StRepS, *Streaming Reputation System*, é um sistema de reputação distribuído, cujo objetivo é identificar e isolar os nodos maliciosos que disseminam conteúdo poluído no sistema de distribuição de mídia contínua em P2P. O sistema promove a reabilitação de um nodo, uma vez que há um incentivo para que estes nodos parem de repassar conteúdo poluído ou melhorem suas condições de rede, evitando assim, dados corrompidos. O desenvolvimento do StRepS é inspirado em um sistema de reputação proposto anteriormente para arquivos poluídos em sistemas de compartilhamento P2P [Costa et al. 2007], que foi estendido para capturar peculiaridades de poluição em distribuição de mídia contínua.

A reputação de um nodo é construída a partir de dois componentes: a *Experiência Individual* e os *Testemunhos dos Parceiros*. Cada nodo, baseado nesses dois componentes, atribui uma reputação à cada um de seus parceiros. A experiência individual é atualizada a cada momento em que um nodo recebe *chunks* dos seus vizinhos. Os testemunhos são obtidos através das trocas de informações entre parceiros.

A cada momento em que um nodo i recebe *chunks* de um parceiro j , ele atualiza a sua experiência individual de acordo com o comportamento de j . Assim, se j atua corretamente, o nodo i aumentará sua reputação, caso contrário, a diminuirá. A experiência individual de i do nodo j é calculada como segue:

$$I_{i(j)}^t = \left\{ \begin{array}{ll} \max(0, I_{i(j)}^{t-1} - \alpha_p * (1 + n/r)^y) & \text{se } n/r > \text{valor limite} \\ \min(1, I_{i(j)}^{t-1} + \alpha_g * n/r) & \text{caso contrário} \end{array} \right. \quad (2)$$

Na Equação 2, r é o total de *chunks* que i requisita a j , n é o total de conteúdo poluído que é provido, α_p e α_g são as recompensas e penalidades associadas interação de i com j , quando é classificada como limpa ou poluída, respectivamente. Um poluidor é identificado pela relação entre os dados poluídos e os limpos que ele transfere a um vizinho durante um intervalo de tempo. Para identificar rapidamente e penalizar os poluidores, o valor de α_p é inflacionado por $(1 + n/r)^y$, onde $1 \leq y \leq 2$. Assim, enquanto

um nodo é recompensado linearmente, este pode ser penalizado exponencialmente. Além disso, $\alpha_p > \alpha_g$, faz com que a experiência individual perca valor mais rapidamente.

O testemunho dos parceiros demonstra uma reputação geral sobre um dado nodo. Como dito anteriormente, cada nodo troca periodicamente com seus vizinhos, ou com um subconjunto deles, uma lista com suas experiências individuais. Esta informação é utilizada antes de cada nova interação e é atualizada como segue:

$$T_{i(j)} = \left\{ \begin{array}{l} \sum_{k \in N} I_{k(j)} * R_{i(k)} \\ \sum_{k \in N} R_{i(k)} \end{array} \right. \quad (3)$$

Na equação 3, N é o conjunto de vizinhos de i e $R_{i(k)}$, definida na equação 4, é a reputação corrente que o nodo i tem do nodo k . Se nenhum testemunho foi coletado em i sobre j , $T_{i(j)} = R_{init}$, onde R_{init} é um valor inicial para a reputação. Como a reputação de um nodo mostra a confiabilidade que i percebe do nodo k , o testemunho de cada nodo k é ponderado por sua própria reputação final. Conseqüentemente, as opiniões de nodos com maiores reputações têm maior impacto que as opiniões de nodos com baixa reputação.

Como a reputação de um nodo j construída por i apresenta dois componentes, devemos ponderar qual a importância de cada um deles. Assim, $(0 \leq \beta \leq 1)$ controla o peso dado para *Experiência Individual* e para os *Testemunhos dos Parceiros*. Baixos valores para β dão ênfase à experiência individual, enquanto altos valores enfatizam as opiniões dos parceiros. Logo, o nodo i calcula periodicamente a reputação de cada parceiro seu, da seguinte maneira:

$$R_{i(j)} = \beta * T_{i(j)} + (1 - \beta) * I_{i(j)} \quad (4)$$

Finalmente, seja $R_{min(i)}$, $(0 \leq R_{min(i)} \leq R_{init})$, a reputação mínima que um nodo deva ter para que seja considerado confiável por i . Por esse critério, o nodo i não envia nem recebe *chunks* de nenhum outro nodo que não se enquadre na faixa de valores de uma reputação considerada confiável.

4. Simulação de Poluição em Mídia Contínua em P2P ao Vivo

Esta seção apresenta a avaliação do BlackList e do StRepS, comparando-os a um sistema de mídia contínua ao vivo sem sistema de reputação. Conduzimos simulações para demonstrar os efeitos de ataques de poluição de conteúdo nesse tipo de aplicação, e verificamos o quão danoso estes podem ser, em função da quantidade de poluidores no sistema. Os demais experimentos evidenciam que sistemas de reputação devem ser bem elaborados, para que possam atingir os objetivos de isolar os parceiros maliciosos e evitar desperdício de recursos. Não basta que a aplicação verifique quais *chunks* estão poluídos, pois a qualidade percebida pelo usuário final pode ser afetada, devido a pedidos de retransmissão, atrasos e ocupação indevida da banda de rede/processamento.

4.1. Modelo de Simulação

O simulador de rede NS-2 [Mccanne et al.] foi utilizado para a realização da simulação em questão. Foi construído um conjunto de novos agentes, que simulam todas as en-

tidades participantes de um sistema de transmissão ao vivo em P2P. Esses novos agentes seguem um protocolo de reenvio orientado por dados e estruturam seus participantes baseando-se em malha (*mesh-pull overlay network*), da mesma forma que sistemas populares como PPLive, PPStream e GridMedia [PPLive 2007, PPstreaming 2007, Zhang et al. 2005, Hei et al. 2006]. A Seção 4.1 apresenta o modelo de simulação adotado, e os principais resultados são discutidos na Seção 4.2. As principais características da simulação são:

Modelo de Fluxo Contínuo ao Vivo P2P: modelamos um sistema de distribuição de mídia contínua ao vivo baseado em malha. Nesse tipo de sistema, existe um participante especial, o *Servidor*, de onde se origina a mídia a ser transmitida por toda a estrutura P2P. Um novo participante, ao se unir ao sistema, faz contato com um subconjunto de pares do sistema. Caso seja de interesse, o par contactado pelo novo cliente o adiciona a sua lista de pares e começa a interagir na troca de dados. Cada participante reconhece e troca dados apenas com parceiros aos quais estão conectados. O subconjunto inicial é obtido aleatoriamente entre todos os participantes do sistema, através de um mecanismo independente de inicialização (*bootstrap*).

A simulação não leva em consideração a codificação do fluxo contínuo, dessa forma, o cliente deve descarregar dados da rede sem erro, a uma taxa igual ou próxima a taxa de geração de dados pela fonte sorvedora. Para a aplicação, assumimos que os clientes possuem capacidade de armazenamento e transmissão suficiente para visualização da mídia, e que compartilhem um conteúdo recém capturado da rede por 2 minutos.

Modelo de Rede: nos experimentos foram utilizadas topologias de rede, criadas pelo gerador de topologias BRITE [Medina et al. 2001]. Cada topologia gerada apresenta mil nodos, as ligações entre os pares apresentam banda de rede suficiente para a aplicação de vídeo em questão, e a maneira como são realizadas as ligações e os tempos de atrasos entre elas é típica de uma estrutura de rede de grande abrangência.

Modelo do participante: cada participante do sistema é classificado como um *bom parceiro* ou como um *poluidor*. Os poluidores repassam apenas *chunks* poluídos e nunca abandonam o sistema. Além disso, os poluidores anunciam um mapa completo de *chunks*, ou seja, sempre têm disponível algum dado desejado. Assumimos também que, os participantes *bons* nunca abandonam o sistema e trocam mapas de *chunks* consistentes com seus dados disponíveis/desejados. Os pares são limitados por recursos como largura de banda e número máximo de conexões. Conseqüentemente, eles só realizam parcerias entre si caso tenham recursos disponíveis para atender e requisitar dados dessa nova parceria. Quando um par perde um vizinho ou deseja uma melhor condição do fluxo de mídia contínua do sistema, ele pode requisitar vizinhos adicionais, onde, a seleção dos novos parceiros é feita aleatoriamente. Em nossa simulação existe um nodo especial, o *servidor*, que é um participante *bom* e somente produz dados, sem a necessidade de consumir. Além disso, todos os participantes do sistema coletam informações sobre suas parcerias as tratam a cada período de 30 segundos.

Modelo de disseminação de poluição: em nossa simulação, o poluidor dissemina apenas *chunks* poluídos e sempre atende às requisições por dados feitas a ele. Pares *bons* podem entregar dados corrompidos com probabilidade de p_{error} , o que também é interpretado como poluição. Além disso, caso nenhum mecanismo de verificação de conteúdo



Figura 1. Sistema Simulado

tenha sido implementado, assumimos que os participantes repassam o conteúdo poluído ingenuamente, .

Modelo de conluio dos nodos maliciosos: cada nodo malicioso tem conhecimento de todos os outros atacantes no sistema. Assim, a cada interação com o servidor centralizado ou com os parceiros no sistema P2P de transmissão ao vivo, os poluidores reputam-se com uma nota normalmente distribuída acima da nota limite para uma boa reputação e abaixo da nota máxima permitida.

Configuração do experimento e cenário: A Figura 1 mostra o cenário da simulação de nossos experimentos. Adotamos uma configuração de rede com mil participantes, incluindo servidor e poluidores. O servidor produz vídeo a uma taxa de 6 *chunks* por segundo, o que é comum nesses tipos de aplicações [Dhungel et al. 2007]. Cada participante se conecta a um número máximo de vizinhos, normalmente distribuído entre 30 e 40 nodos. Inicialmente, um participante tenta se conectar diretamente à 60% do número máximo de vizinhos permitido e pode realizar novas parcerias durante a sua participação no sistema, hora por pedidos de outros pares, hora por necessidade de incrementar a quantidade de dados descarregados da rede.

4.2. Resultados

Nós realizamos experimentos com várias configurações e obtivemos resultados qualitativamente similares. Os resultados apresentados têm como parâmetros os valores encontrados na Tabela 1. Nos experimentos realizados, os nodos maliciosos continuamente tentam manter o seu número máximo parceiros e, permanecem no sistema do momento de sua entrada (entre os minutos 2 e 5) até o fim da realização dos experimentos (minuto 60). Os nodos se juntam a rede no início da experimentação, normalmente distribuído entre o tempo 5s e 300s e não abandonam o sistema. Os resultados apresentados são valores médios de 4 execuções, com coeficiente de variação abaixo de 2%.

Inicialmente discutimos os resultados relativos a presença de nodos poluidores em uma rede de distribuição de mídia contínua em P2P sem nenhum tipo de reputação ou verificação de dados poluídos. A Figura 2 apresenta os resultados de nossos experimentos com mil participantes e presença de poluidores no sistema. Pelas Figuras 2-a e 2-b, verificamos que os nodos maliciosos podem poluir totalmente um participante qualquer.

Tabela 1. Parâmetros da Simulação

Parâmetro	Valor
Número de participantes	1000
Taxa da mídia	300kbps = 6 chunks/s
Tempo de duração da seção	1 hora
Número de vizinhos	25 a 40
Intervalo de medições	30s

Pelas taxas de recepção e envio mostradas nestas figuras, percebemos que o participante analisado apresenta a quase totalidade de dados poluídos e se torna um poluidor passivo. A taxa de recepção não foi alterada e o nodo continua recebendo dados a uma razão próxima à taxa de codificação do vídeo. Estas duas figuras são visões extremistas do sistema e acontece quando um nodo é totalmente influenciado pelos poluidores.

A visão de um participante isolado no sistema pode levar a enganos de interpretação pois o nodo isolado pode ser influenciado totalmente por um poluidor ou simplesmente nem percebê-lo. Por este motivo, as Figuras 2-c e 2-d apresentam uma visão agrupada do sistema de mídia contínua em P2P. Estes dois gráficos mostram a probabilidade acumulada inversa da proporção de dados poluídos que um nodo qualquer recebe e envia durante um período de ataque. Nós mostramos a situação de ataque para 1, 10 e 100 poluidores na rede com mil participantes.

Com 1 poluidor no sistema, há cerca de 80% de probabilidade de se obter um fluxo com pelo menos 60% de dados poluídos. Se o sistema apresentar 100 poluidores, o que corresponde a 10% do total de participantes da rede, percebemos que um nodo recebe pelo menos 90% dos dados poluídos com os mesmos 80% de probabilidade.

O envio de chunks de um nodo para seus vizinhos segue o mesmo comportamento que a recepção, uma vez que os nodos repassam ingenuamente o conteúdo poluído aos seus pares. Durante o ataque de 100 poluidores, a probabilidade de se repassar um fluxo com pelo menos 60% dos dados poluídos é superior a 90%.

Podemos ver nas Figuras 3-a e 3-b os resultados de um ataque em relação à contaminação dos parceiros. Consideramos como um receptor/emissor contaminado, um vizinho de um nodo que tenha recebido/enviado pelo menos um *chunk* poluído em um intervalo de medição. Assim, um nodo no sistema com a participação de apenas um poluidor, contaminou mais de 80% de seus vizinhos em 98% dos casos. No mesmo sistema, com uma probabilidade superior a 93%, um nodo recebe conteúdo poluído de pelo menos 80% de seus vizinhos.

Caso seja realizado somente o pedido de retransmissão dos dados contaminados e não haja reputação dos nodos da rede, a taxa de envio e recepção de *chunks* nos nodos aumenta consideravelmente. A Figura 4 mostra a taxa de recepção de dados na rede P2P de transmissão de vídeo ao vivo. Quando temos 10 poluidores no sistema, o que corresponde a 1% da rede simulada, a taxa necessária de recepção situou-se cerca de 25% superior a taxa esperada. Entretanto, na presença de 100 poluidores (10% da rede), é necessário realizar uma recepção de dados a uma taxa cerca de 125% superior a esperada.

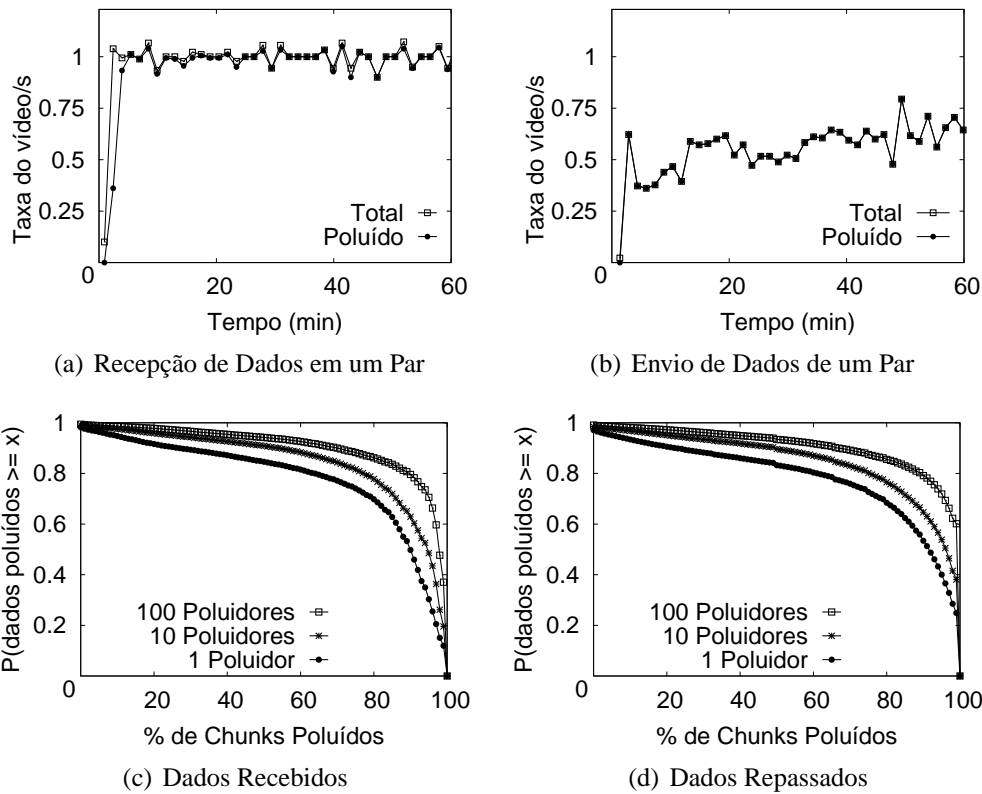


Figura 2. Características do Sistema com 1000 nodos e Presença de Poluidores

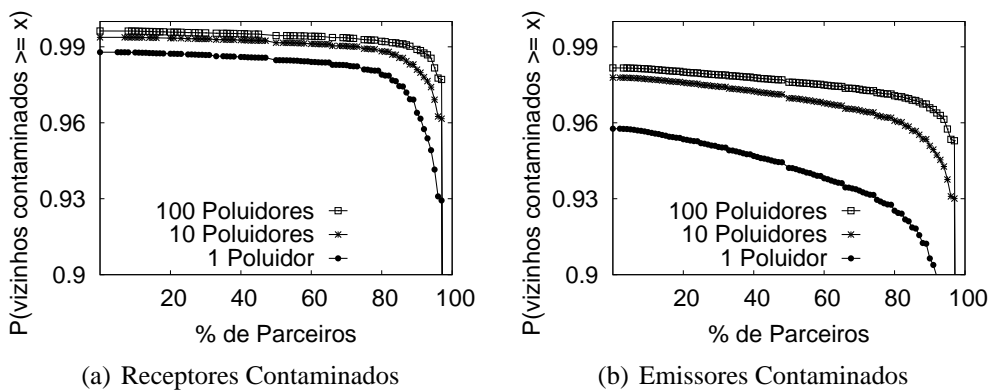
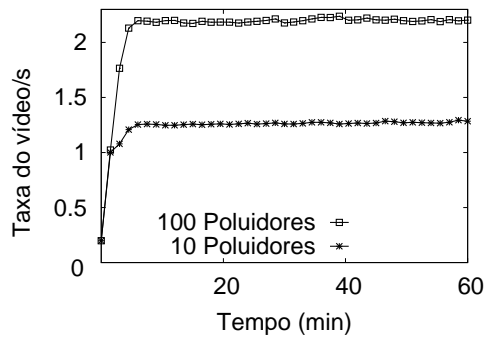


Figura 3. Características do Participantes com Presença de Poluidores

Tal diferença na banda necessária para retransmissão de dados em um sistema com 10 e 100 poluidores pode ser explicado pelo comportamento dos nodos não maliciosos. Ao se realizar um pedido de retransmissão de dados, um nodo não o faz para o mesmo parceiro anterior. Porém, com o aumento do número de poluidores, aumenta-se a probabilidade de um nodo fazer o pedido de retransmissão a um outro poluidor, o que justifica uma taxa de transmissão extremamente alta para a rede com 10% de poluidores. Este tipo de situação mostra a necessidade de banir rapidamente um poluidor da rede.

A Figura 5 apresenta os resultados do sistema utilizando o *BlackList* como mecanismo de reputação. Neste sistema há 100 poluidores que não realizam conluio. Como



(a) Bloqueio de Chunks Poluídos

Figura 4. Taxa de Transmissão de Dados na Rede

esta abordagem é centralizada, há uma alta taxa de recepção de testemunhos dos participantes em um único ponto, e desta forma, o sistema rapidamente identifica os nodos maliciosos. Assim, uma vez que um participante é marcado como malicioso, os demais nodos do sistema evitam fazer parceria com ele. Como podemos verificar pela Figura 5-a, esta abordagem apresentou uma rápida atuação e demorou cerca de 3 minutos para isolar todos os 100 poluidores do sistema.

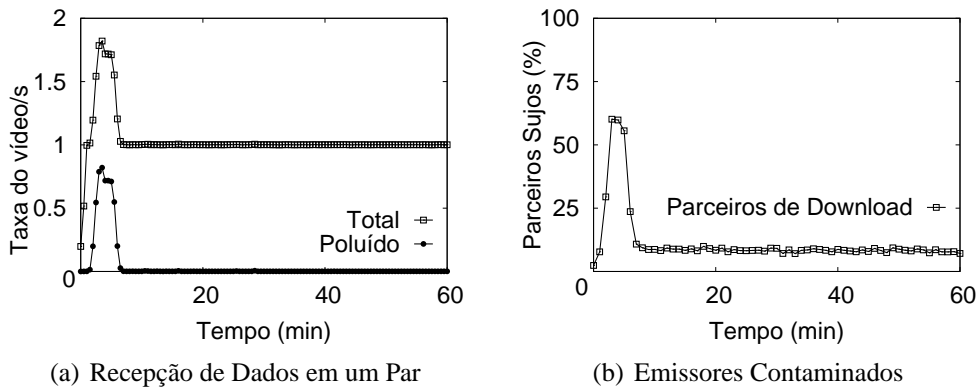


Figura 5. BlackList Centralizado com 100 Poluidores Sem Conluio

A Figura 6 apresenta o resultado do *BlackList* sob efeito de ataque de poluição e conluio dos nodos maliciosos. Quando os atacantes se combinam e se reputam bem, uma abordagem centralizada não atinge o seu objetivo. Os atacantes continuam influenciando fortemente a rede P2P conforme podemos verificar pela alta taxa de transferência presente na rede como um todo. Esta alta taxa é necessária para realizar as retransmissões e como um nodo malicioso não é identificado como tal, a retransmissão pode ser dirigida a ele novamente. Como os nodos maliciosos sempre se reputam bem, eventualmente sempre existirá um número suficiente de poluidores bem reputados que poderá afetar a aplicação de mídia contínua.

As Figuras 7 e 8 apresentam os resultados obtidos utilizando o StRepS mediante conluio dos poluidores. Comparando a Figura 5-a com a 7-a, nota-se que o StRepS necessita de um período maior para isolar os poluidores. Porém, a taxa de rede adicional necessária para suportar retransmissões é baixa, com pico de de 30% acima da original.

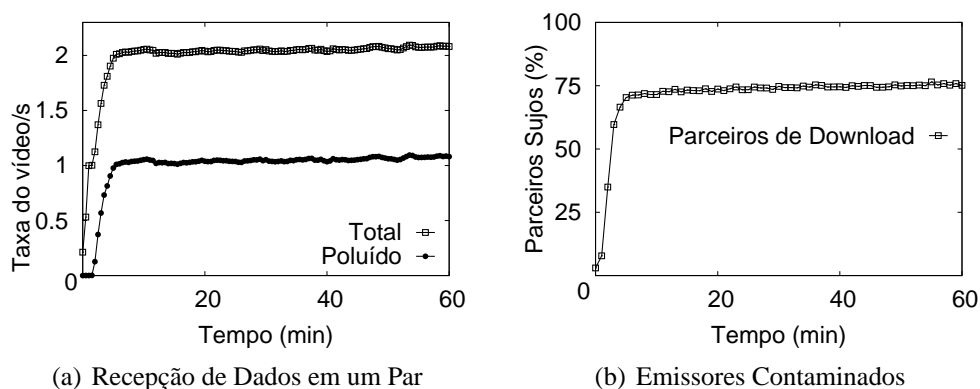


Figura 6. BlackList Centralizado com 100 Poluidores e Conluio dos Poluidores

Percebemos pela Figura 7-b, que o número de parceiros que enviou pelo menos 1 *chunk* poluído a um determinado nodo é ligeiramente maior que o ocorrido no BlackList sem conluio. Isto acontece porque os nodos maliciosos estão realizando um conluio, e eles eventualmente conseguem voltar a realizar parcerias e enviar dados poluídos. Porém, eles são identificados e banidos novamente.

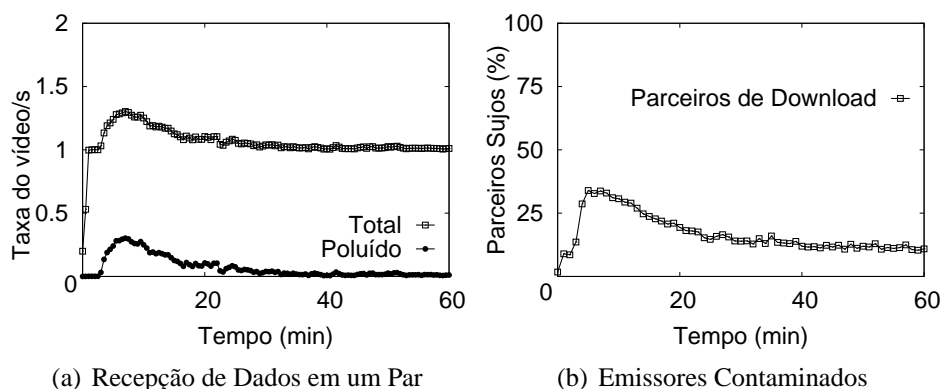


Figura 7. StRepS com 100 Poluidores Sem Conluio dos Poluidores

Comparando a Figura 5-a com a Figura 8-a, ambas com 100 poluidores presentes, o StRepS novamente apresenta um tempo maior para atingir o seu melhor desempenho e uma banda de rede adicional similar ao BlackList. Porém, o StRepS nesta figura, esta sujeito ao conluio dos nodos maliciosos além do ataque de poluição, o que justifica o sistema como um todo ficar sob influência dos poluidores durante um período de tempo maior. Para o sistema com conluio dos poluidores, o StReps necessita de um pico de 90% de banda de rede adicional.

5. Conclusões e Trabalhos Futuros

Neste artigo, nós tratamos o ataque de poluição a sistemas de transmissão de mídia contínua ao vivo em arquiteturas P2P. Primeiro, mostramos que tal ataque tem um alto impacto no sistema como um todo e, mesmo um número pequeno de poluidores consegue contaminar todos os participantes do sistema. Nós apresentamos, através de simulação, a extensão do dano sofrido pelo sistema em condições de ataque de poluição. Segundo,

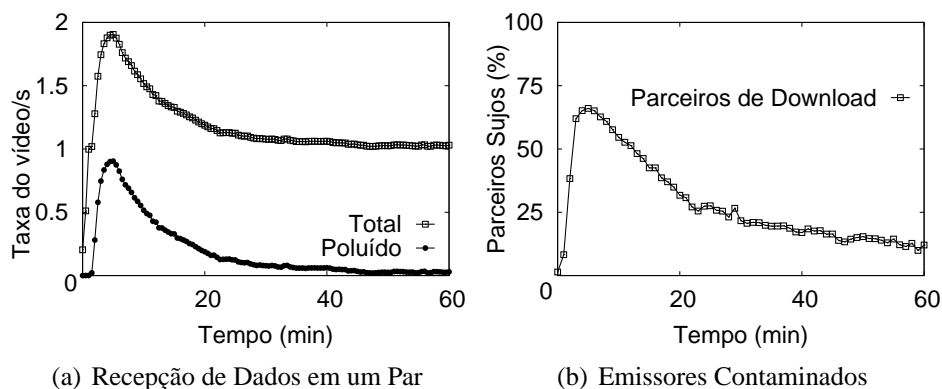


Figura 8. StRepS com 100 Poluidores e Conluio dos Poluidores

apresentamos alterações a sistemas de reputação, utilizados em compartilhamento de arquivos em P2P, e avaliamos a sua eficiência para banir os pares maliciosos do sistema de distribuição de mídia contínua ao vivo, além de conter a disseminação do conteúdo indesejado. Os resultados mostram que, tanto uma abordagem centralizada de reputação com lista negra, quanto a abordagem distribuída, identificam e banem rapidamente os nodos maliciosos. Porém, os sistemas centralizados estão vulneráveis ao conluio dos atacantes, enquanto o sistema distribuído apresentado, o StRepS, se mostra resistente a este tipo de comportamento.

Trabalhos futuros incluem uma avaliação extensa de sistemas de reputação nesse contexto, especialmente sob ataques combinados. Incluem também a prototipagem do sistema de reputação distribuído, o StRepS, e a inclusão do mesmo em um sistema real de distribuição de mídia contínua em P2P.

Referências

- Bittorrent (2007). The bittorrent website. <http://www.bittorrent.com/>.
- Conner, W., Nahrstedt, K., and Gupta, I. (2006). Preventing dos attacks in peer-to-peer media streaming systems. In *Thirteenth Annual Multimedia Computing and Networking Conference (MMCN'06)*, San Jose, CA.
- Costa, C., Soares, V., Almeida, J., and Almeida, V. (2007). Fighting pollution dissemination in peer-to-peer networks. In *SAC '07: Proceedings of the 2007 ACM symposium on Applied computing*, pages 1586–1590, New York, NY, USA. ACM.
- Damiani, E., di Vimercati, D. C., Paraboschi, S., Samarati, P., and Violante, F. (2002). A reputation-based approach for choosing reliable resources in peer-to-peer networks. In *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, pages 207–216, New York, NY, USA. ACM Press.
- Dhungel, P., Hei, X., Ross, K., and Saxena, N. (2007). The pollution attack in p2p live video streaming: Measurement results and defenses. In *Proc. SIGCOMM Peer-to-Peer Streaming and IP-TV Workshop*.
- Haridasan, M. and van Renesse, R. (2006). Defense against intrusion in a live streaming multicast system. In Montresor, A., Wierzbicki, A., and Shahmehri, N., editors, *Peer-to-Peer Computing*, pages 185–192. IEEE Computer Society.

- Hei, X., Liang, C., Liang, J., Liu, Y., and Ross, K. W. (2006). Insights into pplive: A measurement study of a large-scale p2p iptv system. In *In Proc. of IPTV Workshop, International World Wide Web Conference*.
- Jin, X., Chan, S.-H., Yiu, W.-P., Xiong, Y., and Zhang, Q. (2006). Detecting malicious hosts in the presence of lying hosts in peer-to-peer streaming. In *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE.
- Maya Haridasan, R. v. R. (2007). Securestream: An intrusion-tolerant protocol for live-streaming dissemination. In *Journal of Computer Communications. Special issue on Foundation of Peer-to-Peer Computing*. Elsevier.
- Mccanne, S., Floyd, S., and Fall, K. ns2 (network simulator 2).
<http://www-nrg.ee.lbl.gov/ns/>.
- Medina, A., Lakhina, A., Matta, I., and Byers, J. (2001). BRITE: Universal topology generation from a user's perspective. Technical Report 2001-003.
- PPlive (2007). The pplive website. <http://www.pplive.com/en/>.
- PPstreaming (2007). The ppstreaming website. <http://www.ppstreaming.com>.
- Walsh, K. and Sirer, E. G. (2005). Fighting peer-to-peer spam and decoys with object reputation. In *P2PECON '05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 138–143, New York, NY, USA. ACM.
- Wang, W., Xiong, Y., Zhang, Q., and Jamin, S. (2006). Ripple-stream: Safeguarding p2p streaming against dos attacks. In *ICME*, pages 1417–1420. IEEE.
- Zhang, M., Luo, J.-G., Zhao, L., and Yang, S.-Q. (2005). A peer-to-peer network for live media streaming using a push-pull approach. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 287–290, New York, NY, USA. ACM Press.
- Zhong, Y. T. L. S. M. Z. S. Y. Y. (2006). A novel distributed and practical incentive mechanism for peer to peer live video streaming. In *ICME 2006, IEEE International Conference on Multimedia & Expo, July 9-16, 2006, Toronto, Canada*.