Uma Ferramenta para Análise e Controle de Aplicações P2P

Gustavo Beltrami Rossi
Laboratório Associado de
Computação e Matemática Aplicada
Instituto Nacional de Pesquisas Espaciais
rossi@lac.inpe.br

Antonio Montes Filho
Laboratório Associado de
Computação e Matemática Aplicada
Instituto Nacional de Pesquisas Espaciais
montes@lac.inpe.br

Resumo

Este artigo apresenta uma metodologia para o desenvolvimento de um software com capacidade de análise de tráfego de rede em camada de aplicação, permitindo a criação de regras de controle de banda de rede e emissão de relatórios estatísticos de protocolos ponto a ponto. Ajudando desta forma administradores de rede na implementação de políticas de segurança de uma maneira mais eficaz.

Abstract

This paper presents a method for the development of a software with application layer network traffic analysis capability, allowing the creation of bandwidth control rules and peer-to-peer protocols statistical reports emission. Thus helping network administrators in implementing efficiently security policies.

1. Introdução

Os software de comunicação ponto a ponto, referenciados neste documento simplesmente como P2P (*peer-to-peer*), elevam a idéia de que a Internet é uma rede de compartilhamento a novos patamares. Seus desenvolvedores pregam a livre distribuição de informação sem qualquer tipo de restrição.

Os software P2P permitem que computadores individuais troquem mensagens e compartilhem vários tipos de arquivos e informações através da Internet. Todos esses produtos, por definição, conectam computadores sem a utilização de um servidor central utilizando comunicação ponto a ponto real. A ausência de servidor central cria uma rede parecida com uma malha, onde todos os nós da rede estão interligados uns aos outros, e cada nó atua como cliente e servidor de informações. A Figura 1 exemplifica o modelo cliente/servidor e P2P.

1.1. Cenário atual

Recentemente, os software P2P estiveram em evidência nos noticiários devido às atuais e potenciais leis sobre direitos autorais. A Napster [8], companhia que criou um software para compartilhamento de arquivos do tipo MP3, foi condenada por infringir os direitos autorais de artistas e gravadoras ao permitir a troca de músicas entre os seus usuários.

Atualmente, estão sendo lançados software P2P que possuem o potencial de compartilhar todos os tipos de arquivos e devem atingir rapidamente uma popularidade ainda maior do que o Napster conseguiu. Este crescimento e popularidade podem ser facilmente observados verificando os software mais procurados em *sites* de *download*, como o c|net [9], por exemplo.

A Tabela 1 mostra uma lista dos software mais populares do site c|net da semana de 17 de março de 2002, onde só são indicados os software associados a aplicações P2P.

Dos 25 software mais procurados no *site*, 11 são aplicações P2P, sendo os três primeiros da lista em número de transferência.

2. Descrição e análise do problema

Mesmo com as restrições impostas pelos órgãos responsáveis pelos direitos autorais, como *Recording Industry de sua popularidade.Association of America (RIAA)* e *Motion Picture Association of America (MPAA)*, a utilização dos software P2P vem crescendo. Uma análise da companhia Webnoize estimou que 3.05 bilhões de arquivos foram transferidos utilizando a rede FastTrack [10] e Gnutella [11] durante agosto de 2001. Este valor é muito similar aos 2.79 bilhões de arquivos que foram transferidos utilizando a rede Napster em fevereiro de 2001 no pico de sua popularidade.

Porém, tão importante quanto a preocupação com os direitos autorais envolvidos nos arquivos compartilhados por esses software, está o potencial relacionado a problemas de segurança para corporações e

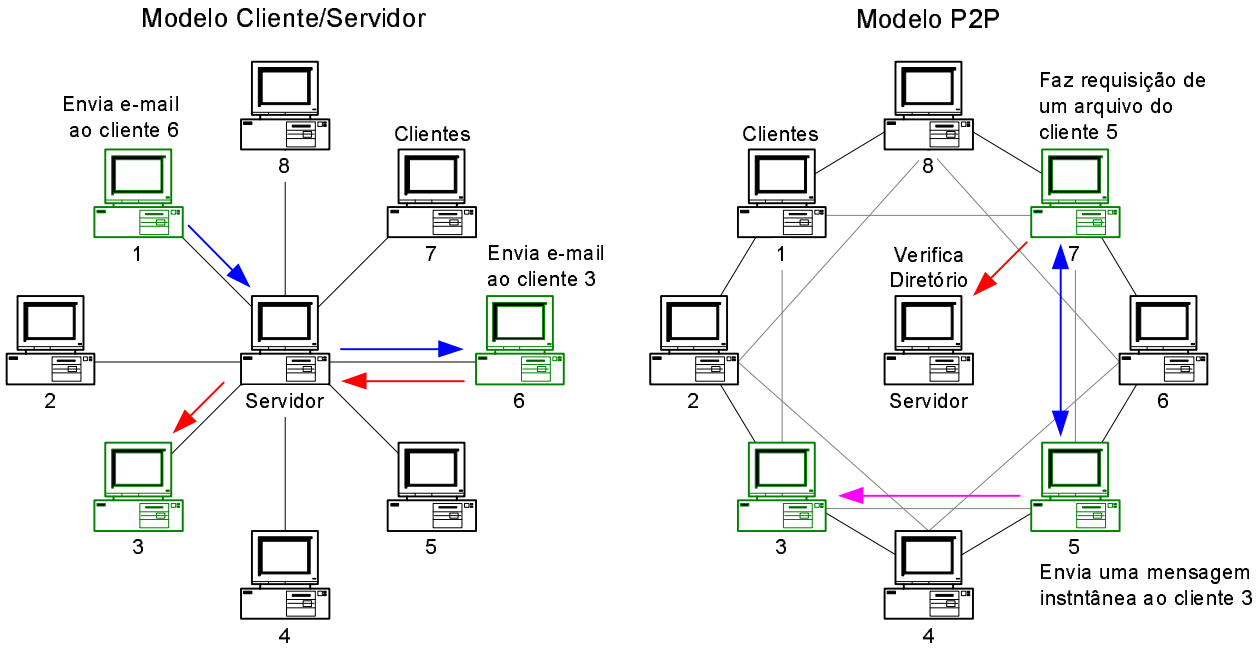

Figura 1: Modelo Cliente/Servidor e P2P [4]

This Week	**Most Popular titles in Windows** week ending March 17	**Last Week**	**Weeks on chart**	**Downloads this week**
1	Morpheus Preview Edition	1	46	3,322,409
2	KaZaA Media Desktop	2	6	3,079,149
3	ICQ 2001b	3	234	767,519
7	iMesh	7	99	267,662
9	LimeWire	10	53	190,699
12	BearShare	9	57	160,227
13	Grokster	11	23	155,059
15	Audiogalaxy Satellite	14	76	122,011
19	Trillian	18	13	83,776
23	MSN Messenger	27	31	54,474
24	WinMX	24	53	51,664

Tabela 1: Software mais populares do site c|net

usuários domésticos.

Recentemente, um furo de segurança descoberto em estudo do especialista Nicholas Weaver [12], da Universidade de Berkeley, mostra que o software Altnet [13] distribuído junto com o KaZaA [14], permite que atacantes utilizem o programa, oferecendo códigos maliciosos na forma de falsas atualizações.

Dessa forma, os software P2P ameaçam as corporações com:

- Consumo da largura de banda;
- Violação das responsabilidades e uso aceitável definidos pela corporação;
- Distribuição de vírus e cavalos de tróia;
- Revelação do endereço IP e MAC.

Para os usuários, os software P2P ameaçam:

- Revelação do endereço IP e MAC;
- Revelação da velocidade de conexão;
- Compartilhamento de arquivos;
- Distribuição de vírus e cavalos de tróia.

2.1. Preocupações para as corporações

Consumo de banda de rede

Para a maioria das companhias, consumo de banda de rede é uma grande preocupação. Internamente, cada usuário que estiver usando um software P2P está fazendo uma grande utilização de banda de rede.

Software como Gnutella, Morpheus [15] e WinMX [16] são geralmente utilizados para transferência de arquivos relativamente grandes, como MP3s, AVIs e MPGs. Se muitos usuários estão transferindo esses arquivos, o tráfego regular de rede associado aos negócios da companhia pode ser prejudicado. Universidades começaram a proibir tráfego de protocolos P2P mais conhecidos por essa mesma razão.

Violação das responsabilidades e uso aceitável

Outra preocupação para as corporações é a possibilidade de ser responsabilizada por violações de direitos autorais que seus usuários, utilizando software P2P, possam cometer. Além disso, a maioria das companhias tem uma política de uso aceitável para a Internet. É bem provável que perdas de tempo como transferências de arquivos MP3 sejam violações dessa política.

Violação da política de segurança

As companhias escrevem sua política de segurança para proteção e para estabelecer procedimentos que irão seguir para criar um ambiente de rede seguro. Comumente essa política e seus procedimentos definem a arquitetura que protege a rede, com ítens como *firewalls*, roteadores, servidores *proxy*, e controle de acesso a Internet que são neles habilitados. A maioria dos software P2P pode ser manipulada para contornar a arquitetura de segurança da rede. Isso torna a política de segurança menos efetiva.

Distribuição de vírus e cavalos de tróia

A publicação de arquivos com nomes falsos permite que usuários maliciosos possam distribuir livremente cavalos de tróia e vírus de forma oculta.

Se o usuário utiliza um protocolo P2P como o FreeNet [17], existe uma grande chance de ser atacado. O FreeNet não utiliza um servidor central e endereços IPs não são rastreados. De fato, um arquivo é copiado localmente de um cliente participante a outro cliente participante até ser recebido pelo cliente que fez a requisição. A possibilidade de se espalhar um cavalo de tróia ou vírus se torna mais fácil do que enviá-los por e-mail.

A rede Gnutella também é um terreno fértil para ataques de vírus. Em fevereiro de 2001, o Mandragore Worm [18] infectou a rede e foi rapidamente espalhado a computadores nela conectados.

2.2. Preocupações para os usuários domésticos

Os usuários domésticos estão em risco assim como as corporações.

Revelação do endereço IP e MAC

Mesmo com *firewalls* pessoais estando amplamente distribuídos e se tornando cada vez mais sofisticados, a grande maioria dos usuários domésticos não possui nada entre seus computadores e a Internet. Assim, quando um endereço é exposto, é o endereço direto de acesso ao computador. Uma simples varredura de portas contra o computador alvo pode dizer a um atacante o sistema operacional e as portas que estão abertas, facilitando a busca por ferramentas que exploraram uma vulnerabilidade.

Revelação da velocidade de conexão

Usuários de programas P2P como Napster, Morpheus e Gnutella têm acesso a informações referentes à velocidade de conexão. Como a maioria dos *links* de 56k e inferiores são conexões discadas, um atacante, provavelmente, não vai gastar esforços para tentar ganhar acesso. Porém conexões de 128k e acima podem indicar um DSL ou *cable modem*, que pode estar em uso por um usuário doméstico.

Compartilhamento de arquivos

A maioria dos usuários pode achar esses problemas alarmantes, mas normalmente perguntam: *"O que eu tenho no meu computador que um hacker pode querer?"* Porém, informações como números de cartão de credito, contas bancarias e outras informações privadas podem estar armazenas no computador.

Distribuição de vírus e cavalos de tróia

Como os usuários de corporações, os usuários domésticos estão expostos a cavalos de tróia e vírus. Porém com a agravante de que a maioria dos usuários não tem acesso a um *help desk* para ajudá-los em caso de emergência, não realizam cópia de segurança de seus dados, não possuem uma política de uso de um antivírus e como proceder para fazer a atualização das bases.

3. Metodologia

O objetivo desse artigo é descrever uma aplicação capaz de analisar o conteúdo do tráfego de rede e através da busca por padrões ou assinaturas, identificar o protocolo P2P utilizado. A partir dessa informação, tomar decisões sobre restrições ou bloqueio da aplicação utilizando uma configuração estabelecida.

Para o correto funcionamento da aplicação, é necessário seu posicionamento um ponto de passagem de tráfego, normalmente em um *gateway* da rede interna para saída para a Internet. A Figura 2 exemplifica o posicionamento da aplicação na topologia de rede.

A aplicação foi divida em três grandes módulos:

A relação entre os módulos ocorre de acordo com o modelo exposto na Figura 3. O módulode análise atua na interface de rede interna do *gateway* e aciona o módulo de controle para controlar o tráfego

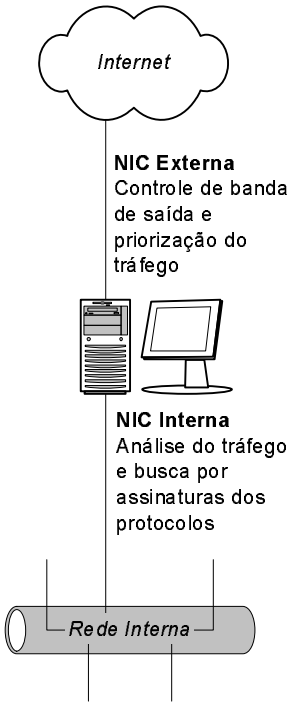

Figura 2: Posicionamento da aplicação na topologia de rede

Análise	Responsável pela captura e análise do tráfego de rede;
Controle	Responsável pelo controle de banda e qualidade de serviço de rede;
Relatório	Responsável pela criação de relatórios contendo a utilização de banda por aplicação.

Tabela 2: Descrição dos módulos da aplicação

P2P na interface de saída. Ambos módulos fornecem informações e dados para o módulo de relatório.

A seguir será melhor descrito o funcionamento de cada módulo.

3.1. Módulo de análise

O módulo de análise é responsável pela captura e análise do tráfego de rede, buscando padrões ou assinaturas que identifiquem a utilização de algum protocolo P2P configurado. Quando o padrão é encontrado, o módulo de controle é imediatamente ativado.

Para se configurar corretamente esses padrões, um profundo conhecimento do protocolo é necessário, e em muitos casos, o protocolo é proprietário e não documentado. Nesses casos são utilizados software de captura de pacotes com capacidade de remontagem de sessão em um ambiente controlado voltado para análise do tráfego de rede. Técnicas de engenharia reversa são aplicadas na sessão capturada com a finalidade de se encontrar padrões que identifiquem a

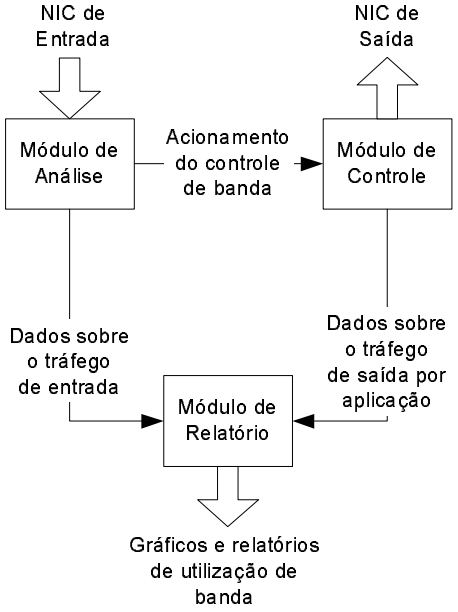

Figura 3: Integração entres os módulos principais

utilização do protocolo P2P. Essas técnicas não serão discutidas nesse documento que tem como foco principal descrever o funcionamento da aplicação de controle dos protocolos P2P e a interligação entre os seus módulos principais. Sendo assim, em toda a abordagem seguinte será utilizado o protocolo aberto Gnutella [19].

3.1.1. Protocolo gnutella

Um *servent*[1] se conecta na rede Gnutella estabelecendo uma conexão com outro *servent* já previamente conectado. A aquisição do endereço IP não faz parte da definição do protocolo, porém o serviço de *cache* de *hosts* é maneira mais comum de automatizar essa aquisição.

Conexão com a rede gnutella

Uma vez que o endereço de um *servent* é obtido, uma conexão TCP é criada, e a seguinte cadeia de caracteres de requisição de conexão é enviada:

```
GNUTELLA CONNECT/<protocol version>\n\n
```

Um *servent* desejando aceitar a requisição de conexão deve responder com a cadeia de caracteres:

```
GNUTELLA OK\n\n
```

Qualquer outra resposta indica que o *servent* não aceitou a conexão.

Uma vez que o *servent* esteja conectado, ele se comunica com os outros *servents* recebendo e enviando descritores.

[1] *Host* participante da rede Gnutella

Requisição de arquivo

O protocolo de transferência de arquivos utilizado na rede Gnutella é o HTTP/1.0. O *servent* que está inicializando a transferência envia uma requisição contendo a seguinte cadeia de caracteres ao servidor alvo:

```
GET /get/<Index>/<Name>/ HTTP/1.0\r\n
Connection: Keep-Alive\r\n
Range: bytes=0-\r\n
\r\n
```

onde `Index` é o índice do arquivo na busca previamente realizada, seguido pelo nome do arquivo `Name`. A entrada `Range` especifica ao servidor a posição de inicio do arquivo, para o caso de um cliente estar retomando uma transferência incompleta. Uma nova transferência utiliza posição 0. Cada linha é terminada utilizando '`\r\n`'.

Resposta do servidor

O servidor que está recebendo a requisição de transferência responde com `HTTP 200 OK` para o cliente, a menos que esteja sobrecarregado ou que o arquivo não esteja disponível. Nesses casos retorna o código de erro HTTP correspondente. Para manter a facilidade, o arquivo é sempre transferido em formato binário, seguido pelo tamanho total do arquivo em *bytes*. Depois de enviar a ultima seqüencia '`\r\n`' o arquivo é transmitido.

```
HTTP 200 OK\r\n
Server: Gnutella\r\n
Content-type: application/binary\r\n
Content-length: <Lenght>\r\n
\r\n
```

3.1.2. Análise do tráfego de rede e busca por padrões

Para realizar análise e busca por padrões dos protocolos P2P desejados, é utilizado a ferramenta Snort [20].

Snort é um sistema detector de intrusão, capaz de realizar análise de tráfego e auditoria de pacotes de redes IP em tempo real. Com ele é possível fazer análise de protocolo, busca e combinação de conteúdo e pode ser usado para detectar uma grande variedade de ataques e *probes*. O Snort utiliza uma linguagem flexível de regras para descrever o tráfego que deve ser armazenado ou permitido a passagem, assim como um motor de detecção que utiliza uma arquitetura modular de *plugins*. Ele tem a capacidade de realizar alerta em tempo real, assim como incorporar mecanismos de alerta externos.

O módulo de analise é constituído por série de regras contendo padrões ou assinaturas dos protocolos P2P. Para o protocolo Gnutella uma assinatura de requisição de arquivo pode ser configurada da seguinte maneira:

```
alert tcp any any <> $(HOME_NET) any
(content:"GET /get/"; content:"HTTP/1.0";
msg:"Gnutella GET statement";)
```

onde a cadeia de caracteres `'GET /get/ HTTP/1.0'` encontrada no conteúdo do pacote define a assinatura procurada.

O Snort é configurado utilizando os pré-processadores de defragmentação e remontagem da sessão do tráfego de rede, gerando um registro de alerta contendo os endereços IPs e portas envolvidos na conexão, quando uma assinatura é encontrada. O pré-processador `frag2` é utilizado para realizar a defragmentação dos pacotes IP. Já o pré-processador `stream4` e `stream4_reassemble` são utilizados para realizar a inspeção e remontagem da sessão TCP.

O registro gerado quando uma assinatura é encontrada segue o formato:

```
07/28-14:54:38.272453
[**] [1:0:0] Gnutella GET statement [**]
{TCP} x.x.x.x:a -> y.y.y.y:b
```

Um *script* Perl monitora o arquivo de alertas e extrai os IPs e portas do registro e os envia ao módulo de controle.

3.2. Módulo de controle

Com alocação de banda e modelagem de tráfego baseada em uma política estabelecida, o software permite que os administradores de rede atuem com inteligência na classificação, análise e outras funções de monitoração da rede. Assim aplicações críticas podem ser protegidas com garantia de banda.

3.2.1. QoS

Qualidade de Serviço [21] (QoS - *Quality of Service*) é um conjunto de capacidades que permite a entrega diferenciada de serviços para o tráfego de rede. Com QoS é possível garantir a largura de banda para o tráfego crítico, limitar o tráfego não crítico, e prover uma consistente resposta de rede, entre outras coisas. Isto permite que se utilizem as conexões de rede de uma maneira mais eficiente, e que se estabeleça acordo de nível de serviço (SLA - *Service Level Agreements*) com os usuários e consumidores da rede.

O objetivo primário da QoS é prover prioridade, incluindo banda de rede dedicada, controle de *jitter* e latência, e melhores características no controle de perda de pacotes. É importante salientar que prover prioridade para um ou mais fluxos de dados não causa parada nos outros fluxos.

Para realizar sua implementação é necessário definir políticas de controle de tráfego para os dispositivos de rede. Estas políticas podem diferenciar o tráfego baseado em categorias, como endereçamento, tipo de aplicação, conteúdo, etc. O software utilizado na implementação foi o ALTQ [22].

3.2.2. ALTQ

ALTQ é uma estrutura para sistemas operacionais Unix/BSD que implementa vários métodos alternativos de enfileiramento. Seu projeto básico é bastante simples; a interface de enfileiramento é projetada como um mecanismo de troca para um conjunto novo de disciplinas de enfileiramento.

O ALTQ suporta as principais disciplinas de enfileiramento, sendo elas: FIFO (*First-In-First-Out*), PQ (*Priority Queueing*), WFQ (*Weighted Fair Queueing*), SFQ (*Stochastic Fairless Queueing*), CBQ (*Class Based Queueing*) e RED (*Random Early Detection*).

A Figura 4 exemplifica o funcionamento dos dispositivos de classificação e priorização de tráfego e as disciplinas de enfileiramento do software.

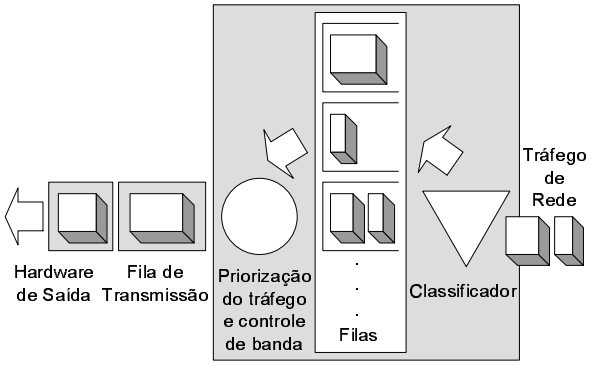

Figura 4: Funcionamento do software ALTQ

3.2.3. Implementando QoS para o controle do tráfego P2P

O módulo de controle recebe do módulo de análise os endereços IPs e portas dos *hosts* participantes de uma transferência de arquivo utilizando um protocolo P2P configurado. Essas informações são inseridas no arquivo de configuração de controle de banda do ALTQ.

Para se fazer uso das propriedades de controle de banda e prioridade de tráfego é utilizado a disciplina de enfileiramento CBQ. Assim as informações passadas pelo módulo de análise são inseridas da seguinte forma na configuração do software ALTQ:

```
class ROOT ne1 NULL pbandwidth 100
class cbq ne1 Gnutella ROOT \
  priority <pri> \
  exactbandwidth <bandwidth> \
  [borrow]
filter ne1 filter1 \
  <sip> <dip> <sport> <dport> 6
```

onde a prioridade é definida em `pri` podendo variar de 0 a 7. Números maiores tem maior prioridade. A configuração da largura de banda disponível para a classe é definida em `bandwidth`, onde o valor deve ser expresso em bps. O parâmetro `borrow` pode ser utilizado para permitir que banda sobressalente seja emprestada pela classe pai, no caso da classe configurada estar com o limite estourado. Nessa configuração a classe pai é definida pela classe `ROOT` e possui 100% da banda disponível definido na interface.

Os parâmetros `sip`, `dip`, `sport` e `dport`, correspondem as informações de endereço IP origem/destino e porta origem/destino respectivamente.

Uma vez a configuração atualizada, ela é relida e ativada pelo controle do ALTQ.

Outro fator importante tratado no módulo de controle é o momento em que a regra de filtragem deve ser removida da configuração. Para isso é utilizado um *script* Perl que monitora o arquivo de registro do pré-processador `strem4` do Snort. Nesse arquivo são inseridos os registros de todas as sessões realizadas, como no exemplo a seguir:

```
[*] Session stats:
Start Time: 07/28/02-19:31:54
End Time: 07/28/02-19:31:59
Server IP: x.x.x.x port: a pkts: 27 bytes: 9224
Client IP: y.y.y.y port: b pkts: 19 bytes: 3872
```

Assim, cruzando as informações sobre os *hosts* descritos no registro com a regra de filtragem, quando uma sessão é finalizada a regra de filtragem deve ser removida do arquivo de configuração.

3.3. Módulo de relatório

O módulo de relatório é baseado na geração de gráficos utilizando o software RRDtool [23] e exibidos em uma interface Web.

Dois tipos de gráficos são gerados:

- **Utilização de banda:** Para gerar o gráfico de utilização de banda os dados são extraídos diretamente das interfaces de rede de entrada e saída, utilizando o SNMP e um agente que envie as informações para o formato lido pelo RRDtool.
- **Utilização de banda por aplicação P2P:** Nesse gráfico os dados como horário inicio e fim da sessão, incluindo pacotes e quantidades de bytes transmitidos são extraídos do arquivo de registro do pré-processador `stream4` do Snort. De posse desses dados é possivel fazer a contabilização da quantidade de tráfego em função de uma fração de tempo.

4. Trabalhos futuros

A comunicação entre o módulo de análise e o módulo de controle é realizada utilizando um *script* Perl que

verifica o arquivo de registro de alertas gerado pelo Snort. O principal problema dessa abordagem está relacionado com a perda de performance em redes com diversos nós e com grande capacidade de *link* de acesso. Assim, faz-se necessário criar um módulo externo de registro específico para o Snort, fazendo com que os alertas sejam enviados diretamente ao módulo de controle.

Outro fator importante a ressaltar no módulo de análise está relacionado com a ampliação da base de protocolos P2P identificados, incluindo, além do Gnutella, protocolos como: Grokster [24], KaZaA, eDonkey [25], iMesh [26] e WinMX.

Já no módulo de relatório, ainda é preciso desenvolver um *script* para o gerar o gráfico de utilização de banda por aplicação, onde as informações sejam retiradas do arquivo de registro do pré-processador `stream4` e a contabilização da quantidade de pacotes e *bytes* trafegados em uma determinada fração de tempo seja processada.

Referências

[1] Stevens, R. W. (1994). *TCP/IP Illustrated, Volume I: The Protocols*. Addison-Wesley.

[2] Comer, D. E. (2000). *Internetworking With TCP/IP Volume I: Principles, Protocols and Architecture*. 4. ed; Prentice Hall.

[3] Tanenbaum, A. S. (1996) *Computer Networks* 3. ed; Prentice Hall.

[4] McKean, C. (2001) *Peer-to-Peer Security and Intel's Peer-to-Peer Trusted Library*. http://rr.sans.org/threats/peer.php

[5] Petruzzi, M., et al. (2000) *Security Concerns for Peer-to-Peer Software*. Key Technologies and Security, Inc. http://downloads.securityfocus.com/library/Security_Concerns_Peer-to-Peer_KTSI.pdf

[6] Chappell, L. (2001) *Security Alert: Just Say Gno!*. Novell Connection, pp. 33-35. http://www.nwconnection.com/2001_09/gnutel91/

[7] Chappell, L. (2001) *Security Alert: Capturing Peer-to-Peer Applications*. Novell Connection, pp. 36-40. http://www.nwconnection.com/2001_12/securityd1/

[8] Napster, Inc. http://www.napster.com/

[9] c|net Download. http://download.com.com/

[10] FastTrack Protocol. http://www.fasttrack.nu/

[11] The Gnutella Network. http://www.gnutella.com/

[12] Weaver, Nicholas. *Reflections on Brilliant Digital. Single Points of Internet 0wnership*. http://www.cs.berkeley.edu/~nweaver/0wn2.html

[13] About Altnet. http://www.kazaa.com/en/aboutaltnet.htm

[14] KaZaA Media Desktop. http://www.kazaa.com/en/index.htm

[15] Streamcast Networks. Morpheus OS. http://www.morpheus-os.com

[16] WinMX. http://www.winmx.com/

[17] The Free Network Project. http://freenetproject.org/

[18] Kaspersky Anti-Virus AVP. *Virus Alerts & Advisories: Gnutella-Worm.Mandragore* http://www.avp.ch/avpve/worms/net/gnutella.stm

[19] Clip2. *The Gnutella Protocol Specification v0.4*. Document Revision 1.2. http://www.clip2.com/GnutellaProtocol04.pdf

[20] Snort. *The Open Source Network IDS*. http://www.snort.org/

[21] Cisco Systems. *Internetwoking Technologies Handbook*. Quality of Service Networking, Capítulo 49, pp 49.1-49.32. http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.pdf

[22] Cho, K. (2001) *The Design and Implementation of the ALTQ Traffic Management System* Keio University, Japão. ftp://ftp.csl.sony.co.jp/pub/kjc/papers/dissertation.ps.gz

[23] Oetiker, T. *About RRDTool*. http://people.ee.ethz.ch/~oetiker/webtools/rrdtool/index.html

[24] Grokster for P2P file sharing. http://www.grokster.com/

[25] eDonkey 2000. http://www.edonkey2000.com/

[26] iMesh. http://www.imesh.com/