DESENVOLVIMENTO EM JAVA DE UMA FERRAMENTA DE VISUALIZAÇÃO GRÁFICA DO TRÁFEGO DE REDE

Thiago Dias Mancilha[1], Lília de Sá Silva[2], Antônio Ésio Marcondes Salgado [3], Antônio Montes [4], Alderico Rodrigues de Paula Jr.[5]

[1] UNIVAP/FCC, Av. Shishima Hifumi, 2911, Urbanova, S.J.Campos – SP, thiago@dss.inpe.br
[2] INPE/DSS, Av. dos Astronautas, 1758, Jardim da Granja, S.J.Campos – SP, lilia@dss.inpe.br
[3] INPE/DSS, Av. dos Astronautas, 1758, Jardim da Granja, S.J.Campos – SP, tone@dss.inpe.br
[4] CenPRA, Rodovia SP-65, km 143,6, Campinas – SP, antonio.montes@cenpra.gov.br
[5] UNIVAP/FCC, Av. Shishima Hifumi, 2911, Urbanova, S.J.Campos – SP, alderico@univap.br

Resumo - O processo de detecção de intrusão em redes de computadores requer o uso de ferramentas que auxiliem o administrador de rede na identificação de ataques e operações maliciosas. Este artigo tem como objetivo implementar uma ferramenta que permita a representação gráfica do comportamento do tráfego de rede contido em um conjunto extremamente volumoso de dados, visando facilitar a identificação de eventos anormais neste conjunto. Com este trabalho, pretende-se que a representação gráfica desenvolvida exiba de forma mais clara o comportamento da rede, tomando por base os atributos obtidos das sessões de rede, de modo a facilitar a interpretação humana sobre o tráfego de rede pela diferenciação do comportamento do tráfego normal, do anômalo, apresentada na tela do computador.

Palavras-chave: Tráfego de Rede, Sessão de Rede, Anomalias no Tráfego de Rede, Gráfico de Coordenadas Paralelas, Visualização do Tráfego.
Área do Conhecimento: Ciências Exatas e da Terra

Introdução

Com o crescimento do volume de tráfego de rede malicioso, incluindo ataques de negação de serviço (DoS), *worms,* tentativas de invasão, intrusão, mau uso dos recursos computacionais compartilhados e outros tipos de ameaças, as ferramentas de análise e monitoração do tráfego desempenham um importante papel no contexto de gerenciamento de segurança da rede[1].

Ferramentas de análise visual do tráfego são importantes para que os administradores de rede possam detectar ataques e anomalias rapidamente e tomar uma ação apropriada para interromper os ataques antes que se propaguem pela rede e causem grandes prejuízos.

Para estudar e classificar o tráfego de rede baseado no uso e protocolos, várias ferramentas tais como FlowScan, Cisco's FlowAnalyzer e AutoFocus são usadas como analisadores do tráfego[1]. Algumas dessas ferramentas provêem recursos de relatórios em tempo real, mas em geral a análise é feita *off-line*.

Tendo em vista que uma representação gráfica para visualização do tráfego coletado facilita muito o trabalho dos analistas de rede na identificação de ataques e pode reduzir drasticamente o tempo de interpretação e análise dos dados[2][3][7][8], propõe-se, neste trabalho, o desenvolvimento de uma ferramenta para filtrar os dados relevantes do conjunto volumoso do tráfego de rede de modo *off-line* e a representação gráfica do comportamento do tráfego em diferentes intervalos de tempo.

Materiais e Métodos

Para a coleta dos dados que trafegam na rede é necessário um mecanismo que capture continuamente os pacotes de rede, denominado sniffer. Dois dos sniffers mais conhecidos são o tcpdump e o ethereal.

O sniffer utilizado neste trabalho (tcpdump) compõe um sensor posicionado no limite externo fora da proteção do firewall da rede monitorada[6], onde é possível ter acesso tanto aos pacotes que entram quanto aos que saem da rede, como ilustrado na Figura 1.

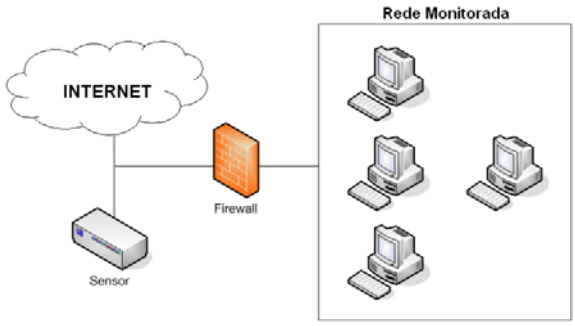

Figura 1 – Posicionamento do sensor de captura.

Os dados capturados pelo *tcpdump* são pacotes de rede baseados na arquitetura TCP/IP, largamente utilizada e difundida pela Internet.

Estes pacotes são remontados, ou seja, colocados em ordem para constituírem uma sessão de rede.

Uma sessão de rede TCP/IP pode ser definida como qualquer seqüência de pacotes, que caracterize a troca de informações entre dois endereços IP, e que tenha informação de início, meio e fim (mesmo que toda a comunicação esteja contida em um único pacote). O conjunto de sessões de um segmento de rede representa o tráfego daquela rede.

Neste trabalho, os pacotes de rede foram remontados através do uso do sistema RECON - *Sistema de Reconstrução de Sessões TCP/IP*[4].

Os dados do cabeçalho dos pacotes são considerados atributos primitivos do tráfego de rede e, isoladamente, carregam informação semanticamente fraca. Observa-se que, para a representação do tráfego, é conveniente o uso de atributos derivados que modelam as sessões do tráfego de rede, apresentando informações de maior relevância para o mapeamento do comportamento da rede.

Os atributos derivados utilizados no desenvolvimento da ferramenta aqui proposta são:

- TMPC - tamanho médio dos pacotes de rede em bytes recebidos pelo cliente;
- TMPS - tamanho médio dos pacotes de rede em bytes recebidos pelo servidor;
- NPC - número de pacotes recebidos pelo cliente;
- NPS - número de pacotes recebidos pelo servidor;
- PPP - porcentagem de pacotes pequenos ou que tenham menos de 130 bytes;
- DIR - direção do tráfego;
- TBC - total de dados (bytes) recebidos pelo cliente;
- TBS - total de dados (bytes) recebidos pelo servidor
- DUR - duração da sessão em milissegundos.

O sistema preliminarmente desenvolvido para *Representação Gráfica do Comportamento do Tráfego de Rede – RGCom* tem sua arquitetura ilustrada na Figura 2, a qual é composta de quatro módulos principais: módulo de captura de dados, módulo de reconstrução de sessões e armazenamento de atributos, módulo de seleção de atributos, módulo de impressão do gráfico.

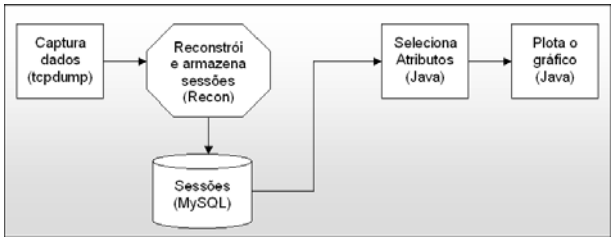

Figura 2 – Arquitetura do sistema.

Para a implementação de todos os módulos, foi utilizada a técnica de Programação Orientada a Objetos em ambiente de desenvolvimento JAVA, visando uma melhor interação entre os módulos, o reaproveitamento do código e a criação de uma ferramenta multiplataforma, para ser executada em qualquer sistema operacional.

O software *tcpdump* é utilizado no módulo de Captura de Dados para capturar todos os pacotes de um determinado segmento da rede, colocando, para isto, a interface de rede em modo promíscuo.

No módulo de reconstrução das sessões TCP/IP e armazenamento de atributos, o software *Recon* foi modificado para armazenar os atributos das sessões em uma base de dados *MySQL* instalada no servidor Linux.

Através do módulo de seleção de atributos o usuário pode escolher, entre os nove atributos derivados disponíveis anteriormente mencionados, aqueles desejados para análise do comportamento da rede.

Conforme a entrada de atributos especificada pelo usuário, estes são extraídos da base, normalizados e processados para a impressão do gráfico.

A figura 3 ilustra uma tela da aplicação, onde é possível selecionar os atributos desejados e o intervalo de sessões a que pertencem estes atributos. A seleção de atributos permite identificar quais atributos e associações destes são mais relevantes para a representação do comportamento do tráfego de rede.

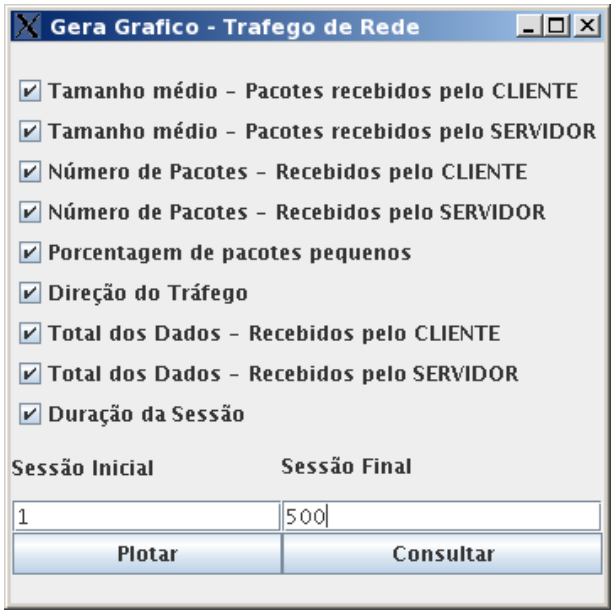

Figura 3 – Tela de seleção dos atributos e intervalo de sessões

Quando informado o intervalo de sessões é feita uma busca na base de dados dos valores de máximo e mínimo (limiares) dos atributos

pertencentes ao intervalo selecionado que participarão dos cálculos para posicionamento dos pontos no gráfico.

Os atributos são impressos na tela em diferentes coordenadas no gráfico, cada uma referente a um determinado atributo, denominadas 'Coordenadas Paralelas'. Esta metodologia tem como objetivo a visualização geométrica das informações em n dimensões. A visualização é obtida pela representação das dimensões em eixos paralelos igualmente espaçados. Em cada um destes eixos, que reproduzem uma dimensão, podemos representar coordenadas por seus valores. Traçando-se uma linha interligando os eixos pode-se representar pontos e planos em n dimensões[5].

Os valores dos atributos, que se diferenciam em escala e unidade de medida, são normalizados através do Teorema de Tales de modo que exista uma proporcionalidade entre os pontos de atributos na tela, sendo apresentados dentro de um mesmo espaço amostral.

Os pontos de atributos selecionados são interligados através de linhas e representam o comportamento de uma sessão do tráfego de rede. Ao conjunto de sessões impressas na tela, pode-se observar o comportamento do tráfego como um todo.

Atualmente algumas ferramentas já possuem os recursos de impressão gráfica que permite a utilização de coordenadas paralelas para representação de dados, como o *VisEd*[9], o *XLSTAT*[9], o *Orca*[10] e o *Mondrian*[11], porém, a ferramenta RGCom em desenvolvimento possibilita a interação entre diferentes módulos necessários para a modelagem visual do trafego de rede.

Resultados

Como resultado deste trabalho foi construído um gráfico de linhas verticais, contendo até nove coordenadas paralelas, sendo que em cada paralela é apresentado o valor de um atributo de sessão selecionado pelo usuário. O gráfico é atualizado de forma estática, ou seja, a cada lote de dados lido, são impressos na tela os conjuntos de pontos (atributos), cada conjunto deste carregando dados de uma sessão de rede específica obtida pelo módulo de Captura.

A figura 4 representa graficamente um tráfego simulado em condições controladas, sem ataques, tendo sido os pacotes capturados na operação de acesso a páginas Web, contidos em 500 sessões normais do tráfego. Portanto, este gráfico representa o comportamento normal do tráfego de rede que descreve uma atividade de acesso de paginas Web no ambiente de testes configurado.

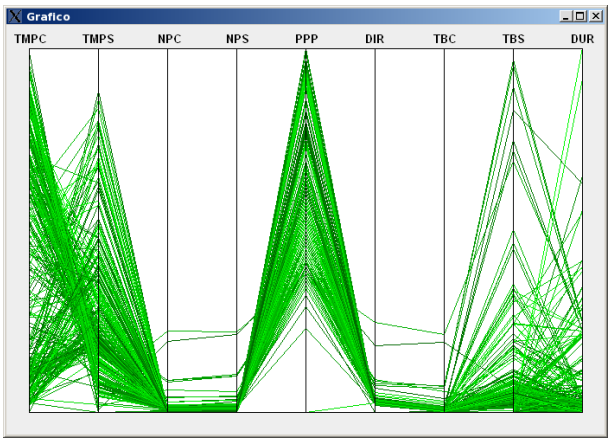

Figura 4 - Representação do tráfego normal.

As coordenadas paralelas possibilitaram o mapeamento de valores com unidades diferentes de forma proporcional no gráfico, exibindo uma representação aproximada em escala normalizada do tráfego da rede. Nesta ilustração pode-se observar as regiões de maior concentração de dados de sessões, indicando os traços de comportamento normal da rede. Ou seja, em algumas regiões do gráfico o fluxo das linhas é mais intenso, demonstrando que o comportamento das sessões é semelhante nestes pontos.

Conclusão

Para que um grande volume de dados do tráfego de rede possa ser avaliado por interpretação humana é necessário que se desenvolva uma técnica visual que modele e apresente as informações relevantes do conjunto de modo simplificado, sugestivo e preciso. A ferramenta de visualização do tráfego deve ser capaz de facilitar o entendimento do comportamento do tráfego, reduzir o tempo de observação e ampliar a precisão dos resultados da análise.

A ferramenta gráfica desenvolvida apresentou, até o momento, um resultado preliminar satisfatório na representação do tráfego, através da utilização de coordenadas paralelas e dos cálculos de normalização, porém deverá ser atualizada em vários pontos.

Dentre os próximos passos deste trabalho serão simulados ataques e operações maliciosas no ambiente de rede controlado para avaliar a ferramenta RGCom quanto a precisão na identificação de eventos anômalos e suspeitos. Espera-se que condições anormais sejam observadas na área fora das regiões de concentração de linhas de cada eixo, consideradas regiões de normalidade. Para tanto, deverão ser avaliadas as associações relevantes de atributos, ou seja, as combinações de atributos que melhor representem o tráfego da rede.

Além disto, a ferramenta deverá ser melhorada para permitir a visualização do tráfego em diferentes intervalos de tempo e utilizando algum método que facilite a comparação de sessões do tráfego antigas e recentes.

Enfim, com a evolução do trabalho e como metas futuras, estuda-se a forma de tornar dinâmica a impressão dos dados na tela à medida em que as sessões vão sendo lidas nos intervalos de tempo pré-definidos e permitir a seleção de serviço de rede específico a analisar.

Referências

[1] KIM, S.S.; REDDY A. L. N. **A Study of Analyzing Network traffic as Images in Real-Time**. Department of Electrical Engineering – Texas A&M University, 2005.

[2] KIM, S.S.; REDDY A. L. N.; VANNUCCI M. **Detecting Traffic Anomalies through aggregate analysis of packet header data**. Proc. of Networking 2004, LNCS 3042, pp 1047-1059, Athens, Greece, mai 2004.

[3] KIM, S.S.; REDDY A. L. N. **Modeling Network traffic as Images**, Proceedings of IEEE International Conference on Communications, Seoul Korea, mai 2005.

[4] CHAVES, M. H. P. **Análise de Estado de Tráfego de Redes TCP/IP para Aplicação em Detecção de Intrusão**. Dissertação de Mestrado em Computação Aplicada - INPE, set 2002.

[5] MÜLLER N. C. **Representação Visual de Bases de Documentos**. Dissertação de Mestrado em Ciência da Computação – PUC. Porto Alegre, ago 2002.

[6] NORTHCUTT S. **Desvendando Segurança em Redes : o guia definitivo para fortificação de perímetros de rede usando Firewalls, VPNs, roteadores e sistemas de detecção de invasores**. Rio de Janeiro, Campus, 2002.

[7] PLONKA D. FlowScan: **A Network Traffic Flow Reporting and Visualization Tool**. Proc. of the USENIX 14th System Administration Conference, New Orleans, LA, dez 2000.

[8] BARFORD P.; KLINE J.; PLONKA D.; RON A. **A Signal Analysis of Network Traffic Anomalies**. Proc. of ACM SIGCOMM Internet Measurement Workshop, Marseille, France, nov 2002.

[9] NASCIMENTO H. A. D.; FERREIRA C. B. R. **Visualização de Informações – Uma Abordagem Prática**. XXV Congresso da Sociedade Brasileira de Computação, 2005.

[10] MARTIN T. **Interactive Data Visualization using Mondrian**. Journal of Computational and Graphical Statistics, Augsburg, Germany, 2003.

[11] LUMLEY T. **Orca [R [RJava]]**. Proc. of the 2nd International Workshop on Distributed Statistical Computing, Vienna, Austria, 2001.