Aplicação de um modelo estatístico em fluxo de dados para detecção de eventos em redes

André Proto, Leandro Arabi Alexandre, Adriano Mauro Cansian
Universidade Estadual Paulista "Júlio de Mesquita Filho" – R. Cristóvão Colombo, 2265, Jd. Nazareth, São José do Rio Preto - SP

***Resumo* — A segurança de computadores e redes se tornou essencial em qualquer ambiente que utilize tais tecnologias como fonte de trabalho. Muitas metodologias para defesa de perímetro em redes de computadores foram criadas buscando-se a detecção de eventos ilícitos. Porém com o crescimento das redes e a Internet, muitas dificuldades são encontradas na tentativa de identificar eventos em ambientes de larga escala. Este artigo propõe uma metodologia de detecção de eventos em redes de computadores de larga escala, cujo perímetro de defesa se estende a um ambiente de grande porte. A proposta aborda a detecção de eventos por anomalia utilizando o protocolo *NetFlow*, métodos estatísticos e o monitoramento do ambiente em tempo real.**

***Palavras-chaves* — Segurança, redes, *NetFlow*, detecção, defesa.**

I. INTRODUÇÃO

A. *Motivação e objetivos*

O desenvolvimento de aplicações de comunicação que utilizam a Internet ou redes locais é crescente nos dias atuais. Comunicadores instantâneos, aplicações de voz e vídeo, aplicações distribuídas entre outras, juntamente com o crescimento do número de usuários na Internet, contribuem para o aumento da quantidade de tráfego em redes de computadores e o número de incidentes de segurança em tais ambientes.

O grande desafio para administradores de redes é como monitorar o perímetro de uma rede de grande porte de forma escalável. Algumas metodologias e ferramentas foram criadas para a defesa de um computador conectado a uma rede, como por exemplo, antivírus, *firewalls* pessoais, *antispyware* e IDS (*Intrusion Detection System* – Sistema de Detecção de intrusão) baseados em assinaturas. Essas ferramentas defendem os usuários de certos tipos de ataques, como disseminação de *worms* e vírus, exploração de vulnerabilidades, entre outros. Porém, outra modalidade de ataques, como *DDoS* (*Distributed Denial of Service*) e ataques de força bruta a senhas de usuários [3], podem envolver não apenas um computador mas também um conjunto deles e não são detectáveis por esses tipos de ferramentas.

Este artigo propõe uma nova metodologia para detecção de ataques em redes de computadores de grande porte (perímetros com grande número de computadores) utilizando a detecção por anomalia de tráfego, o protocolo *NetFlow* (padrão IPFIX) [1][2] e técnicas estatísticas. O objetivo é detectar e alertar administradores de redes de anomalias no tráfego de um determinado serviço da rede (web, FTP, SSH, *telnet*, entre outros).

A. Proto, andreproto@acmesecurity.org, Tel +55-17-32212475, L. A. Alexandre, leandro@acmesecurity.org, Tel +55-17-32212475, A. M. Cansian, adriano@acmesecurity.org, Tel +55-17-32212201.

B. *Trabalhos relacionados*

Existem algumas metodologias para defesa de perímetro em redes de computadores. Uma ferramenta bastante difundida entre a comunidade é a *Snort* [4], que utiliza a metodologia de detecção de assinaturas para identificar eventos de rede. Essa ferramenta pode ser instalada em um dispositivo como *firewalls* ou *gateways*, identificando eventos em ambientes com algumas dezenas de máquinas conectadas. Porém em ambientes de maior porte a ferramenta pode apresentar problemas de desempenho, visto que sua metodologia é analisar cada pacote de dados trafegado por tais dispositivos. O grande número de máquinas em tais ambientes resulta em uma grande quantidade de pacotes de dados trafegados.

O trabalho de [5] aborda uma nova metodologia para detecção de eventos em redes de computadores utilizando o protocolo *NetFlow* e o armazenamento de informações em Banco de Dados. Através de consultas SQL aos dados armazenados é possível identificar alguns eventos na rede incluindo ataques às redes de computadores. Porém o trabalho apenas propõe uma arquitetura para o armazenamento de informações, deixando a cargo de novos trabalhos a busca por metodologias mais robustas de detecção de intrusão aliadas a tal arquitetura.

Já o trabalho de [6] propõe um IDS que utiliza *NetFlow* para detectar ataques como *DDoS* e disseminação de *worms*. Este trabalho compara o formato de fluxos *NetFlow* buscando por semelhanças entre esses e fluxos qualificados como ataques. Ele também propõe técnicas para a reação a esses ataques (regras de bloqueios em roteadores ou *firewalls*), porém enfrenta dificuldades com falsos positivos na sua detecção.

II. CONCEITOS GERAIS

A. *Ataques e detecção por anomalia*

De acordo com [7], um ataque é qualquer ação que vise subverter pelo menos um dos eixos da segurança da informação: confidencialidade, autenticidade, integridade ou disponibilidade de um sistema computacional. Diversas técnicas para a subversão dos sistemas são difundidas atualmente pela Internet, sendo acessíveis por qualquer pessoa.

Qualquer perímetro de rede possui um determinado padrão de tráfego, em termos estatísticos, baseado no

comportamento dos usuários pertencentes a tal ambiente. Alguns tipos de ataques geram certa quantidade no tráfego de uma rede evidenciando uma anomalia perante um padrão normal de comportamento. Esta diferença no tráfego pode ser detectada por metodologias de detecção de intrusão por anomalias, que basicamente se baseiam em um padrão normal de tráfego e identificam discrepâncias no mesmo.
Alguns eventos relacionados à anomalia que serão utilizados neste trabalho são citados a seguir:

- **Prospecção de rede (*scan*):** Esta é normalmente a primeira fase de um ataque. A prospecção ou mapeamento de rede é utilizado para reconhecer serviços disponíveis em uma ou mais redes, identificando serviços e *hosts* vulneráveis a ataques;
- ***DDoS* (*Distributed Denial of Service*):** constitui um ataque de negação de serviço distribuído, ou seja, **um conjunto** de computadores é utilizado para tirar de operação um ou mais serviços ou computadores conectados à Internet [3];
- **Disseminação de *worms*:** *Worm* é um programa malicioso capaz de se propagar automaticamente através de redes, enviando cópias de si mesmo de máquina para máquina [3]. Normalmente ele explora vulnerabilidades em programas utilizados por usuários;
- **Ataque de dicionário:** Ataques desta categoria tentam, através da técnica de "tentativa e erro", adivinhar senhas de usuários em serviços como SSH (*Secure Shell Client*) ou *Web*. Normalmente este ataque tem sucesso, pois muitos usuários utilizam senhas fracas e de fácil dedução;
- **Envio de SPAMs:** SPAMs são e-mails indesejados enviados por pessoas cuja intenção é propagar programas maliciosos, propagandas de produtos ou conteúdo inadequado [3].

Os ataques citados anteriormente possuem como característica comum a comunicação com diversas outras máquinas ou servidores em um curto espaço de tempo, gerando um grande volume de tráfego. Esta característica peculiar será a base para a detecção desses tipos de ataques abordada neste artigo.

B. Fluxo de redes

A *Cisco Systems* define um fluxo de rede como uma seqüência unidirecional de pacotes entre hosts de origem e destino. Pode-se dizer em resumo que o *NetFlow* provê a sumarização de informações sobre o tráfego de um roteador ou *switch*. Fluxos de rede são altamente granulares; eles são identificados pelos endereços IP de origem e destino, bem como pelo número das portas da camada de transporte. O *NetFlow* também utiliza, para identificar unicamente um fluxo, os campos "*Protocol type*" e "*Type of Service*" (*ToS*) do cabeçalho IP e a interface lógica de entrada do roteador ou *switch*. Os fluxos mantidos no *cache* do roteador/*switch* são exportados para um coletor nas seguintes situações: permanece ocioso por mais de 15 segundos; sua duração excede 30 minutos; uma conexão TCP é encerrada com a *flag* FIN ou RST; a tabela de fluxos está cheia ou o usuário redefine as configurações de fluxo. É importante notar que o tempo máximo que um fluxo permanece no dispositivo antes de ser exportado é de 30 minutos.
A Fig. 1 ilustra os campos do protocolo *NetFlow v5*, bem como o seu cabeçalho. Os campos que realmente interessam neste trabalho estão descritos em *"Flow Record Format"*. Eles são responsáveis por representar as informações sumarizadas de uma conexão/sessão entre dois *hosts*, descrevendo endereços de origem e destino, portas de origem e destino, interfaces de entrada e saída do roteador ou *switch*, número de pacotes e octetos envolvidos, *timestamp* de criação do fluxo e *timestamp* de sua última atualização (campos *first* e *last*), *flags* do TCP, entre outros.

Flow Header Format (Netflow v.5)

0	1	2	3	4	5	6	7
version		count		sys_uptime			
unix_secs				unix_nsecs			
flow_sequence				engine_ type	engine_ id	sampling_interval	

Flow Record Format (Netflow v.5)

0	1	2	3	4	5	6	7
srcaddr				dstaddr			
nexthop				input		output	
dPkts				dOctets			
fist				last			
srcport		dstport		pad1	tcp_flags	prot	tos
src_as		dst_as		src_mask	dst_mask	pad2	

Fig. 1. Formato do datagrama *NetFlow*.

C. Conceitos de estatística

Para este trabalho alguns conceitos de estatística descritiva [8] foram utilizados buscando-se a identificação de anomalias no tráfego de uma rede de computadores. A estatística descritiva é usada para descrever e sumarizar um conjunto de dados. Os principais elementos a serem utilizados neste trabalho são:

- **Mediana:** Em um conjunto de dados ordenados, o elemento que separa ao meio tal conjunto, ou seja, possui a mesma quantidade de elementos tanto abaixo quanto acima do mesmo, é considerado mediana. Em alguns casos em que o conjunto possui números pares (não há um único elemento central), dois elementos centrais são somados e divididos por 2 para resultarem na devida mediana;
- **Quartil:** É um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais. Cada parte representa exatamente um quarto da amostra de dados. O 2º quartil, por exemplo, representa a mediana de uma amostra. Neste trabalho serão usados o 1º e 3º quartis;
- **Pontos discrepantes:** Também conhecidos como *outliers*, são elementos de uma amostra nos quais se mostram distantes do resto dos elementos da mesma.

Estes conceitos serão fundamentais para a proposta deste trabalho. A próxima seção descreve a metodologia deste projeto e os pontos principais para a execução dos objetivos.

III. METODOLOGIA

Como descrito anteriormente, o objetivo deste trabalho é identificar eventos de segurança de uma rede de computadores utilizando protocolo *NetFlow* e metodologias de detecção por anomalia baseadas em técnicas estatísticas. Para realização dos objetivos, os seguintes itens devem ser definidos:

- Coleta e armazenamento dos dados providos pelo *NetFlow*;

- Modelo para a definição do padrão de tráfego na rede;
- Modelo para discrepantes de pontos discrepantes.

A. Arquitetura de armazenamento dos dados

A coleta e armazenamento dos dados do *NetFlow* será de fundamental importância para o fornecimento de informações sobre a rede a ser defendida. Utilizou-se então a arquitetura de armazenamento proposta em [5], provendo robustez e versatilidade no armazenamento e consulta dos dados providos pelo *NetFlow*. A arquitetura permite o armazenamento dos fluxos em um banco de dados relacional. Uma tabela em especial armazena uma janela dos últimos trinta minutos de fluxos gerados pelo ambiente. Esta tabela é de essencial importância para o monitoramento em tempo real do tráfego da rede. Maiores detalhes da arquitetura são descritos em [5].

B. Definindo o padrão de tráfego de uma rede

Toda detecção por anomalia necessita da definição do padrão do tráfego de uma rede de computadores. Esta definição deve se basear no comportamento de cada ambiente em determinados horários, isto porque o tráfego de uma rede em horário comercial não possui o mesmo comportamento do tráfego da mesma rede no período noturno. A Fig. 2 ilustra o comportamento do tráfego de uma rede ao longo do dia.

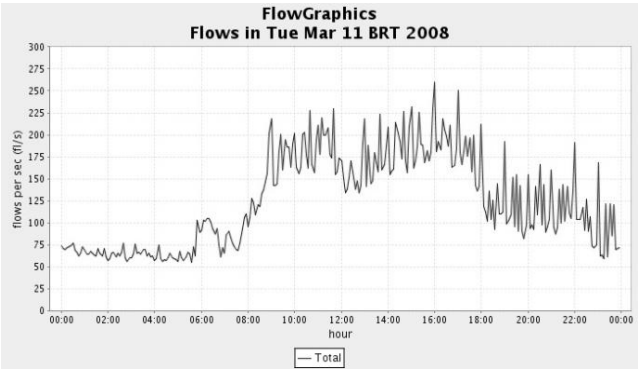

Fig. 2. Exemplo de comportamento do tráfego de uma rede.

Neste trabalho em especial, é necessário não apenas definir o padrão de tráfego da rede como um todo, mas também de cada serviço utilizado nela. Para realizar essa tarefa, foram coletados dados do tráfego em uma janela de três meses. Para cada dia deste intervalo, foram selecionadas janelas de tempo de cinco minutos de fluxos; isso significa que a cada cinco minutos de dados, são computadas as médias de fluxos por segundo relacionadas a cada serviço de rede no ambiente. Para isso foi utilizada uma consulta em linguagem SQL [9] ao banco de dados relacional, com o formato descrito a seguir:

SELECT date_sub(subtime(first,second(first)),interval mod(minute(first),5) minute) as Time, dstport as Service, input, count(*) as Flows, sum(dPkts) as Pkts, sum(dOctets) as Bytes FROM TableDay WHERE dstport < 1024 GROUP BY month(first),day(first),hour(first),(minute(first) div 5),dstport,input order by Time,dstport,input;

O resultado desta consulta pode ser vista na Fig. 3. Nela são obtidas as médias de fluxos para serviços que rodam em portas TCP/UDP menores que 1024. Esses dados foram armazenados em uma nova tabela, a fim de serem novamente trabalhados em uma segunda etapa.

Time	Service	input	Flows	Pkts	Bytes
2009-01-01 00:00:00	0	28	24	484	188311
2009-01-01 00:00:00	21	28	2	3	144
2009-01-01 00:00:00	22	28	2	6	304
2009-01-01 00:00:00	25	28	35	285	103863
2009-01-01 00:00:00	53	28	633	665	47678
2009-01-01 00:00:00	80	28	48	688	50636
2009-01-01 00:00:00	123	28	65	170	12920
2009-01-01 00:00:00	137	28	75	75	16575
2009-01-01 00:00:00	143	28	4	22	1364
...	...	...	...	...	...
2009-01-01 23:55:00	445	28	10	14	560
2009-01-01 23:55:00	496	28	95	100	4800
2009-01-01 23:55:00	500	28	2	2	1352
2009-01-01 23:55:00	769	28	117	128	9928
2009-01-01 23:55:00	771	28	72	141	12288

5282 rows in set (3.62 sec)

Fig. 3. Resultado da consulta SQL.

Com os dados coletados, é possível realizar o cálculo do padrão de tráfego da rede. Este padrão representará, a cada conjunto de cinco minutos de tráfego, uma amostra contendo a média de fluxos envolvidos em determinado serviço. Forma-se então uma série de amostra de dados, sendo cada amostra a representação do tráfego de um serviço de determinado horário (considerando o conjunto de cinco minutos). A Fig. 4 apresenta um exemplo de amostras de dados coletados referentes a um serviço qualquer de rede.

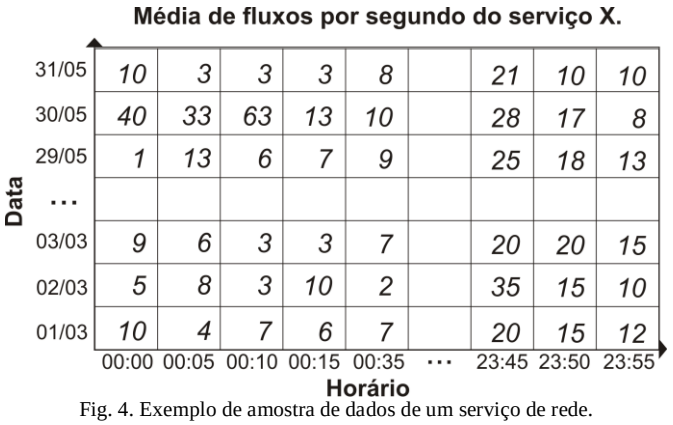

Fig. 4. Exemplo de amostra de dados de um serviço de rede.

Uma questão a ser levantada refere-se a como remover possíveis tráfegos anômalos dentro dos três meses coletados para criação do padrão. Isso é resolvido utilizando-se a mesma fórmula de identificação de pontos discrepantes, que será discutida na próxima subseção.

C. Modelo para identificação de pontos discrepantes

Segundo [8], uma amostra de dados pode ser dividida em cinco pontos de sumarização: o mínimo, o máximo, a amostra de mediana e a 25ª e 75ª porcentagem empírica dos dados. A 25ª porcentagem empírica é chamada de 1º quartil (Q_1); já a 75ª porcentagem empírica é chamada de 3º quartil (Q_3). A distância entre o Q_3 e Q_1 é chamada de *interquartile range (IQR)* e pode ser vista em (1). O ponto mínimo de uma amostra é aquele que está a 1.5*IQR de distância de Q_1 (2); o ponto máximo está a 1.5*IQR de distância de Q_3 (3). Qualquer elemento da amostra que estiver fora dos limites de mínimo e máximo é considerado um ponto discrepante [8].

$$IQR = Q_3 - Q_1 \quad (1)$$

$$Min = Q_1 - 1.5\ IQR \quad (2)$$

$$Max = Q_3 + 1.5\ IQR \quad (3)$$

Para este trabalho apenas o ponto máximo de uma amostra é utilizado. Sejam então as amostras de dados descritas na subseção anterior, aqueles elementos de uma determinada amostra cuja quantidade de fluxos ultrapasse o valor máximo da mesma são considerados anômalos.

A metodologia citada no parágrafo anterior será usada como base tanto para identificar anomalias no tráfego quanto na remoção de pontos discrepantes das amostras coletadas em três meses de fluxos. Para este último, o seguinte algoritmo de remoção de pontos discrepantes foi utilizado:

```
Faça{
        existe_outlier = 0;
        Calcule Q1 e Q3 da amostra X;
        Faça MAX = Q3 + 1.5*(Q3-Q1);
        Percorra todos os elementos da amostra X {
                Se elemento > MAX
                        Remova elemento;
                        Faça  existe_outlier = 1;
        }
} Enquanto existe_outlier = 1;
```

O algoritmo citado calcula o ponto máximo e remove os pontos discrepantes até que esses não mais existam. Ele é executado para cada amostra coletada de cada serviço da rede. Terminado o algoritmo, o valor de MAX será utilizado como limiar para definir a anomalia de um tráfego. Assim um sistema monitora em tempo real a média de fluxos dos últimos cinco minutos da data atual. Para cada valor coletado, o sistema compara com o valor MAX de determinada amostra relacionada ao serviço e hora/minuto correspondente. Como exemplo, supondo que dados do serviço SSH (porta 22) foram coletados no intervalo entre 10h00min e 10h05min. A média de fluxos por segundo deste intervalo será comparado com o valor MAX da amostra correspondente ao mesmo intervalo. Caso seja superior, então concluir-se-a que há uma anomalia no tráfego SSH neste intervalo de tempo.

IV. Resultados

Esta seção descreve testes e resultados obtidos com o novo modelo de detecção de anomalias proposto por este trabalho. A subseção A descreve o ambiente utilizado para os testes. A subseção B descreve os serviços monitorados, seus limitantes obtidos após o cálculo do tráfego padrão e a remoção dos pontos discrepantes. Por fim a subseção C descreve os resultados obtidos.

A. *Ambiente de testes*

O ambiente de testes foi montado em uma universidade que conta com mais de mil dispositivos de rede incluindo computadores, roteadores e dispositivos móveis. O ambiente possui um roteador CISCO 7200 VXR que exporta fluxos *NetFlow* versão 5. A máquina coletora é um PC x86 Pentium D 3.4GHz, 2GB RAM e HD SATA 200GB, dedicado a coleta, armazenamento e análise dos fluxos no banco.

As amostras foram coletadas no período de três meses de fluxos (março a maio de 2009). O tempo médio, por serviço, para a coleta das amostras esteve entre 5 e 6 minutos. Deve-se ressaltar que esta coleta ocorre apenas uma vez; os resultados são armazenados no banco de dados. A consulta responsável por coletar a quantidade de fluxos em tempo real é executada em aproximadamente 0.52s por serviço. O tempo de processamento, por serviço, na identificação de anomalias em um determinado intervalo de tempo é de aproximadamente 0.1s.

B. *Serviços monitorados*

Para este artigo, quatro serviços em particular foram monitorados pelo sistema proposto: FTP (protocolo de transferência de arquivos), SSH (protocolo de acesso remoto), SMTP (protocolo para serviço de e-mails) e HTTP (protocolo para serviços web). Os pontos máximos calculados na fase de definição do padrão de tráfego são mostrados na Fig. 5 (FTP e HTTP) e Fig. 6 (SMTP e SSH). Estes gráficos já mostram a representação dos limitantes de tráfego dos serviços excluídos os pontos discrepantes das amostras.

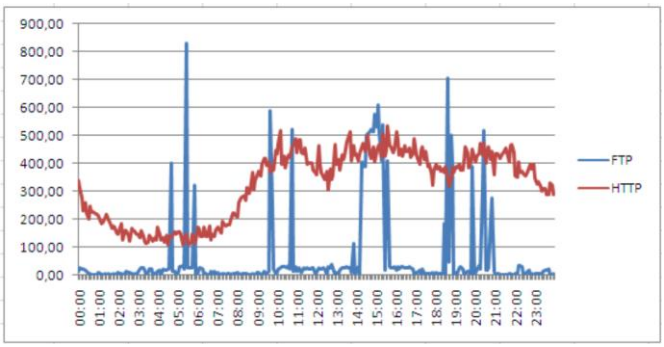

Fig. 5. Limitantes superiores dos serviços FTP e HTTP por tempo.

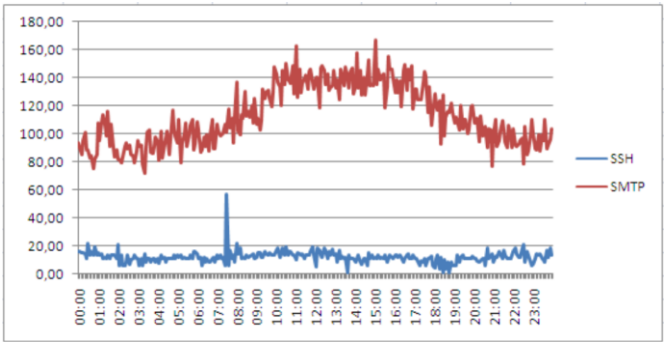

Fig. 6. Limitantes superiores dos serviços SSH e SMTP por tempo.

Na Fig. 5 nota-se no serviço HTTP uma maior quantidade de tráfego em horário comercial (8h às 18h), bem diferente o apresentado em horários noturnos (0h às 6h). Porém o que mais chama a atenção é a representação do tráfego FTP, que possui pontos no tempo cujo valor é excessivamente alto se comparado a outros valores. Isso pode ser explicado, pois dada a amostra coletada a quantidade de ataques ao serviço FTP nesses pontos apresentados foi tão freqüente que estes não foram considerados pontos discrepantes e conseqüentemente não foram removidos. Na Fig. 6 é possível notar uma maior normalidade nos limitantes, sendo que o SMTP apresenta característica parecida ao comportamento do serviço HTTP e o SSH possuindo apenas um dos limitantes

com valor excessivo se comparado aos outros, devido aos mesmos motivos apresentados pelo FTP.

C. Resultados de detecção

O sistema de detecção monitorou o ambiente proposto por um período de seis dias, o que equivale a analisar 1728 períodos de cinco minutos cada. Na Tabela I é apresentado o número de eventos detectados nos serviços propostos e a quantidade de falso-positivos referentes ao mesmo. Para o cálculo de falso-positivos foi realizada a analise de *logs* nos servidores sob ataque ou a utilização de técnicas descritas em [5] para a comprovação ou não de cada evento detectado.

TABELA I EVENTOS ANÔMALOS DETECTADOS.

Serviço	Eventos detectados	Falso-positivos	Porcentagem de acertos
FTP	10	0	100%
SSH	615	2	99,67%
SMTP	153	4	97,38%
HTTP	33	14	57,57%

Para o serviço FTP, dez eventos foram detectados e dentre estes, nenhum falso-positivo foi encontrado. Todos estes ataques se caracterizavam como *scan* de serviço, no qual o atacante deseja saber se em uma determinada rede existe um servidor FTP sendo executado.

Já na detecção de eventos no serviço SSH observa-se um grande número de eventos detectados, sendo apenas dois falso-positivos encontrados. Os ataques em geral se caracterizam por "ataques de dicionário" ou *scans* no serviço. O grande número de eventos pode ser explicado pelo maior número de atacantes utilizando tal técnica de ataque e pelo fato de que o "ataque de dicionário" é executado por várias horas seguidas. Dentre os falso-positivos encontrados, um deles refere-se há um dia em específico no qual pesquisadores utilizaram por diversas vezes o serviço em um horário não convencional (entre 1h e 3h da manhã).

Quanto à detecção de eventos no serviço SMTP (e-mail), 153 eventos foram detectados, sendo apenas quatro eventos falso-positivos. A modalidade de ataques neste serviço pode ser dividida entre *scans* e envio de SPAMs. Para identificar SPAMs, uma das principais técnicas identifica se o computador que enviou tal mensagem tem permissão para enviar e-mails pelo domínio a qual ele pertence e se ele possui um serviço SMTP sendo executado. Dentre os 149 eventos detectados e comprovados, 5 são *scans*, 139 são SPAMs e 5 são *scans* e SPAMs detectados no mesmo evento.

Por fim a detecção de eventos no serviço HTTP foi o que apresentou maior número de falso-positivos. Dos 33 eventos detectados, 14 foram falso-positivos, totalizando aproximadamente 57% de acertos. A modalidade de ataques detectada neste serviço ficou restrita a *scans* apenas. Dois fatores explicam tal resultado: o comportamento do serviço HTTP varia com grande freqüência tanto entre os horários de um dia quanto em períodos do ano; a utilização do protocolo HTTP v1.0 [10] implica na geração de uma conexão para cada objeto de uma página web requisitado, resultando em um grande número de fluxos. De fato, a quantidade de tráfego do serviço HTTP medido há alguns meses é bem diferente do mês atual, devido à grande popularidade de tal serviço e o crescimento de usuários utilizando-o em tal ambiente.

V. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou uma proposta para a detecção de eventos em redes de computadores utilizando metodologias estatísticas e análise de fluxos de dados *NetFlow*. O objetivo é utilizar tal proposta para defender o perímetro de um ambiente de redes de computadores em tempo real, detectando ataques através de anomalias no tráfego e alertando administradores quando necessário. O trabalho busca modelar o padrão do tráfego de um ou mais serviços de rede através de amostras do tráfego baseadas em intervalos de tempo. Dada as amostras, são removidos os pontos discrepantes e, pela mesma técnica, são definidos os valores máximos para separação entre tráfego normal e anômalo.

Testes foram realizados com o monitoramento de quatro serviços bastante utilizados por usuários da Internet: FTP, SSH, SMTP e HTTP. Dentre eles, os serviços FTP, SSH e SMTP apresentaram resultados bastante positivos no que diz respeito à detecção de eventos e o número de falso-positivos. Porém o serviço HTTP apresentou o maior número de falso-positivos, explicado pelas características do serviço e pela sua escalabilidade no ambiente. Apesar dos testes apresentados referirem-se apenas a esses quatro tipos de serviços, este modelo pode ser aplicado a qualquer outro serviço que o usuário desejar.

A grande dificuldade para este trabalho foi contabilizar o número de falso-negativos nos testes realizados. Isso porque para realizar tal quantização é necessário executar outras metodologias ou ferramentas no ambiente e comparar os resultados de detecção; porém as restrições administrativas do ambiente não permitiram a realização de tais testes.

Por fim como trabalhos futuros a coleta de amostras do tráfego deverá ser aperfeiçoado de modo que se atualize ao longo do tempo, tratando os problemas citados como no serviço HTTP. Também está prevista a utilização de novas técnicas para a remoção de pontos discrepantes a fim de comparar tais remoções com a técnica já utilizada neste trabalho.

REFERÊNCIAS

[1] J. Quittek, T. Zseby, B. Claise, S. Zender, "RFC 3917: Requirements for IP Flow Information Export: IPFIX." 2004. Disponível em: http://www.ietf.org/rfc/rfc3917.txt.

[2] B. Claise, "RFC 3954: Cisco Systems NetFlow Services Export Version 9". 2004. Disponível em: http://www.ietf.org/rfc/rfc3954.txt .

[3] Cert.BR. "Cartilha de Segurança para Internet". 2006. Disponível em: http://cartilha.cert.br/ .

[4] Sourcefire, "Snort.org", 2008, Disponível em: http://www.snort.org .

[5] J. L. Corrêa, A. Proto, A. M. Cansian, "Modelo de armazenamento de fluxos de rede para análises de tráfego e de segurança". *VIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg)*, 2008, Gramado – RS.

[6] W. Zhenqi, W. Xinyu, "NetFlow Based Intrusion Detection System", *International Conference on Multimedia and Information Technology*, 2008, Phuket, Thailand.

[7] D. Gollmann, *Computer Security*, 1ª Ed., John Wiley & Sons, New York, USA, 1999, ISBN 0471978442.

[8] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, L. E. Meester, "A modern introduction to probability and statistics", Springer, 2005, ISBN 1852338962, pp. 234-244.

[9] R. E. Elmasri, S. Navathe, *Sistemas de Banco de Dados*. 4ª Ed., Addison-Wesley, 2005, ISBN 8588639173.

[10] J. F. Kurose, K. W. Ross, *Redes de Computadores e a Internet*, 1ª Ed., Addison-Wesley, 2003, ISBN 8588639106, pp 65-80.