Aplicando a Teoria da Evidência na Detecção de Anomalias

Bruno F. O. Lins[1], Eduardo L. Feitosa[1,2], Djamel F. H. Sadok[1]

[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – Cidade Universitária – Recife – PE – Brasil

[2]Departamento de Ciência da Computação
Universidade Federal do Amazonas (UFAM)
CEP 69077-000 – Campus Universitário – Manaus – AM – Brasil

{bfol,elf,jamel}@cin.ufpe.br

Abstract. *In the current security situation of computer networks, traffic anomalies generated by many different reasons have affected an increasing portion of the Internet. This paper presents a tool for data fusion based on the theory of evidence to detect anomalous traffic. For validation, tests will be done in a controlled environment with the injection of anomalies together with normal traffic.*

Resumo. *Na atual situação de segurança das redes de computadores, anomalias de tráfego geradas pelos mais diversos motivos têm afetado uma porção cada vez maior da Internet. Este artigo apresenta uma ferramenta de fusão de dados baseada na teoria da evidência para detecção de tráfego anômalo. Para validação, serão realizados testes em ambiente controlado com a injeção de anomalias ao trafego normal.*

1. Introdução

No atual cenário de segurança das redes de computadores, em especial da Internet, problemas causados pelo aumento do volume de tráfego, má configuração de protocolos e serviços e ataques vêm crescendo consideravelmente. Tais problemas são, em geral, conhecidos como anomalias de tráfego [Sundaram 1996] e as atividades de detecção e prevenção vão além do tradicional gerenciamento de incidentes de segurança.

Por definição, uma anomalia é o desvio de uma condição típica ou normal. A detecção de anomalias parte da premissa de que qualquer coisa fora do comportamento normal é, então, anômala e constitui um ataque [Sundaram 1996]. No contexto do tráfego de rede, anomalias podem ser causadas pela proliferação de códigos maliciosos (vírus e worms), ataques de negação de serviço, spam, falhas de roteamento, entre outras. O resultado normalmente é o desperdício de recursos, causando prejuízos financeiros e degradando o desempenho e a confiabilidade das redes.

Parte desses prejuízos se deve a ineficiência das atuais soluções em identificar, reduzir e interromper anomalias de tráfego. A percepção tardia e a falta de cooperação entre soluções ou mesmo com a infra-estrutura de rede são alguns exemplos. Firewalls, regras de filtragem e sistemas antivírus são exemplos de soluções usadas no combate a determinados tipos de anomalias. Contudo, essas soluções têm baixo poder de contra-resposta e/ou apresentam resultados não muito eficazes na prevenção de anomalias. Sistemas especialistas como IDS (*Intrusion Detection Systems*) e IPS (*Intrusion*

Prevention Systems) sofrem com altas taxas de alarmes falsos positivos e negativos (tráfego normal classificado como anômalo e tráfego anômalo classificado como normal, respectivamente) e acabam tornando-se inadequadas para determinadas anomalias ou mesmo falham em questões de abrangência e escalabilidade. A questão é como resolver à incerteza de suas análises e conseqüentemente aumentar a precisão.

Neste contexto, técnicas de fusão de dados vêm sendo empregadas para aprimorar, e até mesmo agrupar, métodos de detecção de anomalias visando melhores e mais precisos resultados. A fusão de dados pode ser definida como o processo de combinar vários dados a fim de produzir informações mais valorosas para o usuário. Os dados podem provir de uma ou mais fontes. As fontes podem ser semelhantes ou desiguais, desde que o sistema tenha a capacidade de lidar com dados heterogêneos ou conflitantes. O uso de um eficiente esquema de fusão pode gerar vantagens significativas como a melhoria na confiança nas decisões, devido ao uso de informações extras, do desempenho de contramedidas e no desempenho em condições adversas.

Dentre as diversas técnicas e métodos de fusão de dados, a Teoria de Evidência de Dempster-Shafer (TDS) tem sido bastante utilizada na detecção de anomalias. Em comparação com outras técnicas como regra de Bayes [Pearl 1982] e teoria da possibilidade [Zadeh 1978], a TDS possui a habilidade de modelar o afunilamento do conjunto de hipóteses à medida que se acumulam evidências. Este procedimento reflete o processo que caracteriza o raciocínio usado em diagnóstico e o raciocínio especializado em geral [Gordon e Shortliffe 1984]. Maiores informações sobre as vantagens da TDS podem ser encontradas em [Uchoa 1997]

Sendo assim, este trabalho visa comprovar a eficácia da teoria da evidência como mecanismo de fusão de dados na construção de sistemas de detecção de anomalias. Para tanto, um protótipo, denominado ADS-Fusion, será construído para tratar incertezas e imprecisões de diferentes métodos de detecção de anomalias.

O restante desse artigo é organizado da seguinte forma: a seção 2 apresenta os conceitos básicos relativos à Teoria da Evidência de Dempster-Shafer; a seção 3 apresenta os trabalhos relacionados, discutindo vantagens e desvantagens; a seção 4 relata a implementação da ferramenta, detalhando os sensores utilizados e processo de fusão; a seção 5 trata da análise dos resultados obtidos durante os testes e simulações; por fim, a seção 6 apresenta as conclusões finais e perspectivas de trabalhos futuros.

2. Teoria da Evidência

A Teoria da Evidência de Dempster-Shafer, ou simplesmente TDS, é um dos modelos mais famosos para a representação da incerteza em sistemas baseados em conhecimento. Originada dos trabalhos de Arthur Pentland Dempster [Dempster 1967][Dempster 1967a], foi refinada e ampliada por Glenn Shafer [Shafer 1976].

Seu objetivo é solucionar problemas encontrados em modelar a incerteza quando se trabalha com métodos meramente probabilísticos. Neste aspecto, a TDS, diferentemente das teorias probabilísticas bayesianas, não necessita de um conhecimento prévio das distribuições de probabilidade dos elementos estudados, podendo, desta forma, atribuir valores de crença a subconjuntos das possibilidades e não só aos eventos simples. Outro diferencial é o fato de que a crença não atribuída a nenhum evento em particular, é atribuída ao ambiente e não ao restante das evidências.

A seguir serão apresentados alguns elementos essenciais para a TDS e as principais vantagens e desvantagens deste método.

2.1. Terminologia básica

Nesta seção serão apresentados os conceitos básicos da TDS.

2.1.1. Quadro de discernimento (Θ)

A TDS pressupõe um conjunto de hipóteses primitivas chamado de quadro de discernimento ou domínio do problema, denotado por Θ. Para que um conjunto de hipóteses seja considerado um quadro de discernimento é necessário que o mesmo apresente algumas características fundamentais:

- Θ deve ser exaustivo, ou seja, deve conter todas as possíveis hipóteses primitivas;
- Todas as hipóteses pertencentes ao Θ devem ser mutuamente exclusivas.

Um bom exemplo de quadro de discernimento é o seguinte:

Θ = {Normal {N}, Ataque TCP {T}, Ataque UDP {U}, Ataque ICMP {I}}

2.1.2. Função de massa (bpa)

Na TDS, a indicação de crença em determinada hipótese, dada um evidência, é associada a um valor no intervalo [0,1]. A relevância dada a cada um dos elementos do domínio do problema é representada por uma função denominada de atribuição de probabilidade básica (*bpa - basic probability assignment*) ou função de massa e notada por m. A função de massa representa a quantidade total de crença na evidência que remete um grupo de hipóteses.

2.1.3. Função de crença

Função de crença, denotada por $bel(\,)$, mede o total de crença atribuída a um determinado subconjunto de Θ. Na prática, $bel(Ev)$ é a soma das probabilidades básicas atribuídas a todos os subconjuntos de Ev de Θ. Para obter o total de crença atribuída Ev_1, deve-se adicionar a $m(Ev_1)$ os valores de $m(Ev_2)$ para todo subconjunto próprio Ev_2 de Ev_1.

2.1.4. Plausibilidade

A função de plausibilidade ou probabilidade superior, $pl(\,)$, determina a quantidade máxima de crença que pode ser atribuída a um determinado subconjunto de 2^{Θ}. Tem-se que $pl(Ev)$ representa o mesmo que $1 - bel(\overline{Ev})$. Desde que $bel(Ev) + bel(\overline{Ev}) \leq 1$, tem-se que $bel(Ev) \leq pl(Ev)\ \forall\ Ev\ \subseteq \Theta$.

2.1.5. Intervalo de Crença

Uma vez que a plausibilidade representa o quanto se pode acreditar em uma determinada hipótese e $bel(Ev\,)$ representa a crença atual em Ev, a TDS representa a crença em Ev como sendo o intervalo representado por $[bel(Ev), pl(Ev)]$. A esse intervalo dá-se o nome de intervalo de crença, representado por $\Im(Ev)$.

Tal intervalo exprime a faixa de probabilidade na qual se pode acreditar em uma determinada hipótese, sem correr o risco de graves erros com suposições.

3. Trabalhos Relacionados

Na literatura existem diversas abordagens e ferramentas de detecção de anomalias que utilizam mecanismos de fusão de dados para diminuir o número de falsos positivos e negativos e aumentar a eficiência com um todo. Dentre essas, as que utilizam a teoria da evidência destacam-se pela manipulação da incerteza em suas análises. Algumas dessas propostas serão discutidas a seguir.

A arquitetura proposta em [Siartelis 2003] descreve a utilização da TDS na elaboração de um sistema de detecção de ataques de negação de serviço distribuídos. É formada por um conjunto de diferentes sensores, espalhados em pontos distintos da rede e operando de forma autônoma, mas compartilhando suas crenças sobre o verdadeiro estado da rede, ou seja, se a rede está ou não sob ataque. Assumindo que a rede tem comportamento estocástico sem qualquer modelo funcional, os autores tentam inferir sob o estado do sistema sem conhecimento prévio, apenas usando os dados informados pelos sensores, que podem ter obtido suas "evidências" baseadas em critérios totalmente diferentes. O protótipo desenvolvido utiliza o SNORT [Snort 2008] e um coletor de dados SNMP como sensores.

O modelo proposto em [Tian 2005], denominado IDSDFM, emprega a TDS para fundir os diferentes alertas gerados por diferentes tipos de IDS. De modo simplificado, o IDSDFM correlaciona os alertas de acordo com seu grau de similaridade e de acordo com as regra de combinação de massa da TDS e define o *bpa* para cada uma dessas evidências.

A proposta apresentada em [Chen e Aickelin 2006] descreve um sistema capaz de "aprender" as características fundamentais do ambiente e assim determinar a crença em cada análise gerada por cada um dos vários sensores distribuídos pela rede, mesmo que utilizem diferentes metodologias de análises. Basicamente, emprega módulos específicos capazes de determinar a crença em cada uma das inferências e repassar esta informação para um módulo de combinação de dados, baseado na TDS, que tem o poder de tomar decisões. Tais módulos são sistemas baseados em aprendizagem que, após uma fase de treinamento, são capazes de estabelecer parâmetros que condizem com as características do ambiente e, desta forma, determinam o *bpa* de cada dado extraído dos sensores.

Uma breve análise das propostas apresentadas revela que a principal vantagem do uso da TDS é a não necessidade de qualquer conhecimento prévio do comportamento da rede. Além disso, a possibilidade de usar vários sensores ou filtros, baseados em diferentes mecanismos de classificação é extremamente interessante e tem se mostrado uma tendência na área de segurança. Por outro lado, questões como o profundo conhecimento do funcionamento dos sensores para que as crenças geradas nas hipóteses sejam condizentes com o real estado da rede e a dependência do desempenho dos sensores na geração de resultados se faz necessária. Além disso, a escolha dos sensores é muito importante. Sensores muito simples podem acabar por não contribuir para o estabelecimento do estado da rede. Já sensores baseados em assinaturas, como o SNORT, podem ter problemas com a detecção de novos ataques. Por fim, o processo de

relação de similaridade entre diferentes alertas pode ser um grande desafio para o estabelecimento da solução.

4. ADS-Fusion

A solução ADS-Fusion foi projetada com o objetivo de estudar técnicas de detecção de anomalia e desenvolver um sistema capaz de aumentar a eficiência da detecção através da fusão de dados usando TDS. Na prática representa um software capaz de fundir dados e produzir uma inferência com um grau de certeza maior que as certezas geradas individualmente pelos módulos de detecção.

O ADS-Fusion é composto por um módulo de coleta, sensores e um mecanismo de fusão de dados. O módulo de coleta é responsável pela monitoração do tráfego da rede e geração de arquivos em formato padronizado. Estes arquivos serão repassados para os diversos sensores para que os mesmos possam processá-los e inferir o estado da rede. Os sensores são os componentes responsáveis por analisar os dados gerados pelo módulo de coleta e detectar possíveis anomalias existentes no tráfego da rede. Outro papel fundamental destes sensores é definir, a partir de seus mecanismos de classificação, a crença em cada uma das inferências geradas. Por fim, o mecanismo de fusão é responsável pela tomada da decisão. Fazendo uso dos recursos das regras de combinação da TDS, correlaciona às diferentes análises dos vários sensores gerando inferências mais precisas e com um maior grau de exatidão.

4.1. Módulo de coleta

Antes de iniciar a explanação sobre o módulo de coleta é necessário informar que a atual implementação do protótipo funciona apenas em modo off-line, isto é, processa traces com o tráfego já capturado da rede em formato .pcap para efetuar todas as análises. Isto se deve ao fato de que as implementações dos sensores escolhidos somente operam de forma *off-line.*

O módulo de coleta foi desenvolvido usando a biblioteca libpcap [LIBPCAP 2008]. A função do módulo de coleta é ler os arquivos de tráfego da rede e gerar as saídas necessárias ao funcionamento dos diversos sensores. Para implementação do ADS-Fusion, foram escolhidos dois sensores (Profiling e TCPModel, explicados a seguir) que necessitam de entradas baseadas em fluxo e *socks*.

4.2. Sensores

Os sensores são componentes de análise responsáveis pela geração das hipóteses sobre o possível estado real da rede. Estes sensores podem variar de pequenos mecanismos de detecção de anomalias (IDS para host, por exemplo) a complexos sistemas especialistas, como os propostos em [Mirkovic 2002] e [Mirkovic 2003].

Uma vez que bons resultados na detecção de anomalias e caracterização do tráfego Internet têm sido obtidos com técnicas de análise comportamental ([Lakhina 2005] e [Karagiannis 2005], por exemplo), modelos matemáticos ([Gao 2006] e [Li e Lee 2005]) e estatísticos ([Abry 2007] e [Scherrer 2007]), foram escolhidas diferentes técnicas de análise para a implementação do protótipo, visando assim, uma maior abrangência no acompanhamento do estado da rede estudada. Detalhes sobre os mecanismos utilizados serão apresentados a seguir.

4.2.1. TCPModel

O TCPModel [Aschoff 2007] é um sistema, desenvolvido em Java, especialista na detecção de ataques DDoS. Baseado no comportamento de troca de mensagens do protocolo TCP entre dois hosts, é capaz de avaliar se existe algum comportamento anormal na rede através da razão entre a taxa de envio e recebimento de pacotes.

A premissa de detecção implementada no TCPModel foi proposta por [Mirkovic 2002]. Ela analisa a razão entre entrada e saída de pacotes TCP através de algoritmo de *Threshold Adaptativo* [Aschoff 2007]. Quando a razão ultrapassa um determinado limiar considerado normal, um alarme é disparado. Para analisar a entrada e saída de pacotes o TCPModel agrupa os pacotes em fluxos que possuem IP e porta de destino comuns, denominados *socks*

Agrupando os pacotes em *socks*, o TCPModel utiliza-se de um conjunto de métricas para definir o estado normal de uma comunicação de protocolo TCP e é capaz de reconhecer modificações neste padrão. Um fator importante neste método de análise é que por se tratar de um algoritmo adaptativo, é capaz de se moldar a possíveis modificações no comportamento da rede.

Para integrar o TCPModel ao modelo de fusão de dados proposto, foi necessário a criação de um mecanismo capaz de definir valores de crenças (*bpa*) a cada hipótese resultante do processo de análise. O mecanismo elaborado para a geração de *bpa* pelo TCPModel é bastante simples. Baseia-se na distância entre os valores obtidos na aplicação da função de indicação do *Threshold Adaptativo*. Sempre que o retorno da função for igual a 1, o sistema considera que a rede pode estar sobre ataque e, desta forma, calcula o *bpa* baseado na distância para o valor esperado para o intervalo anterior. Quanto maior for à distância, maior a crença no ataque.

Para melhorar o entendimento do leitor, um exemplo de geração do bpa é apresentado a seguir. Supondo que a saída de uma análise do TCPModel seja *[58.33.126.229:5576 -> 192.168.0.163:0; Pkt Send: 92 / PktRec: 0; threshold: 6.0],* onde uma conexão anômala entre os endereços IP 58.33.126.229 e 192.168.0.163, com taxa de 92/0 e com *threshold* calculado em 6, e supondo que nesta rede o *threshold* estabelecido como normal tem valor igual a 4, qualquer conexão que ultrapassar este limiar será considerada anômala. Desta forma, o exemplo acima ultrapassa esse valor e terá seu *bpa* calculado sobre o percentual da diferença. Fixando a crença do estado normal da rede sempre em 0,5 é possível determinar a crença do ataque como sendo a soma entre a crença normal e o percentual de aumento, ($6 \div 4 = 1{,}5$; o que representa um aumento percentual de 50%), obtendo, desta forma, um $bpa = 0{,}75$.

4.2.2. Profiling

A metodologia proposta em [Xu 2005] visa à identificação de anomalias de tráfego. O método faz uso de técnicas de mineração de dados e informação teórica (entropia) para automaticamente descobrir padrões de comportamento significantes no tráfego de dados. A metodologia (denominada de profiling) automaticamente descobre comportamentos do tráfego massivo e fornece meios plausíveis para entender e rapidamente reconhecer tráfego anômalo.

Basicamente, examina padrões de comunicação dos computadores (endereços e portas) que são responsáveis por um significativo número de fluxos em um determinado período de tempo. O processo do profiling inclui a extração de clusters significativos e a

classificação do comportamento deles baseado no relacionamento entre os clusters. Por exemplo, para um dado endereço IP de origem (srcIP) i, o processo do profiling inclui a extração dos fluxos com srcIP i dentro de um cluster (denominado de cluster srcIP) e a caracterização do padrões de comunicação (ou seja, comportamento) usando medições de teoria da informação (entropia) sobre as três dimensões de fluxos restantes, ou seja, endereço IP de destino (dstIP), porta de origem (srcPrt) e porta de destino (dstPrt).

A primeira etapa consiste em analisar um conjunto de fluxos baseado nas tuplas bem conhecidas para decidir sobre um cluster de interesse. O objetivo é extrair os clusters significativos de dimensões específicas, isto é, endereços IP de origem e destino e portas de origem e destino. Então, os clusters mais significativos são extraídos de uma dimensão fixa (por exemplo, endereço IP de origem) e o conceito de entropia é usado para medir a quantidade de incerteza relativa (RU - *Relative Uncertainly*) contida nos dados. A segunda etapa é responsável por descobrir relações entre os clusters, ou seja, encontrar padrões de comportamento comuns para o perfil do tráfego.

Visando alcançar este objetivo, a metodologia propõe uma classificação do comportamento baseado nos modelos de comunicação dos computadores de usuários finais e serviços. Desta forma, para cada cluster, uma RU é computada e usada como parâmetro para criar classes de comportamento (BC - *Behavior Classes*). Com essas classes é possível identificar qual delas representa tráfego anômalo ou indesejado.

Para integrar o Profiling ao modelo de fusão de dados proposto, foi estabelecida a freqüência de repetição entre as BCs e a quantidade de fluxos associados a elas como critério para geração de *bpa*. Supondo que usando *dstIP* como chave de grupo e analisando as interações (espaços predefinidos de tempo) percebe-se que na primeira (#1) os fluxos com IP de destino 10.108.40.X (450 fluxos) sejam classificados com o BC = 24 (o que represente um ataque DDoS para esta chave). Na interação seguinte (#2) o BC para este mesmo IP se manteve e o número de fluxos aumentou para 750. Então é possível aumentar a crença nessa inferência.

4.3. Mecanismo de Fusão de Dados

O mecanismo de fusão de dados é responsável pela tomada de decisões, definindo quando um determinado evento é classificado como anômalo ou normal. De modo geral é responsável por agrupar a "visão" da rede pela perspectiva de cada um dos sensores e inferir o comportamento geral baseado na crença individual de cada um. Para sua implementação foi utilizada a API EvidenZ [EvidenZ 2008], uma implementação gratuita em C++.

O funcionamento do módulo de fusão de dados pode ser resumido em quatro estágios (figura 1). O ponto de partida é a leitura das análises (arquivos que contém os elementos analisados e suas devidas classificações) feitas pelos sensores TCPModel e Profiling. Após este estágio é executado um processo de sincronização responsável por relacionar os elementos apontados pelos sensores de forma que os relatos de mesmos eventos possam ser combinados. Este processo se faz importante, pois é necessário garantir que cada uma das classificações a serem combinadas condiga ao mesmo período de tempo. O passo seguinte à sincronização é a combinação dos eventos, fazendo uso das regras de combinação de função de crença. A combinação é utilizada como mecanismo para a geração das inferências do estado real da rede. Para que seja possível combinar as saídas dos filtros, os atributos das saídas são transformados em

elementos da EvidenZ, que basicamente contém o estado identificado (NORMAL ou ANÔMALO) e a crença neste estado (bpa). Após combinados os eventos as inferências são geradas. É nessa fase que o ADS-Fusion determina se a rede está sob o efeito de uma anomalia ou não. Para tanto é necessária a definição do quadro de discernimento e a hipótese a ser avaliada.

A figura 1 ilustra o fluxo de eventos do módulo de fusão de dados.

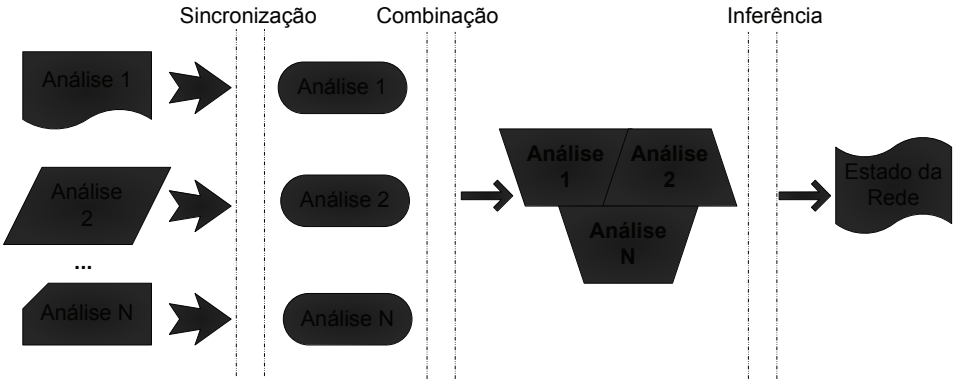

Figura 1. Fluxo de processamento do módulo de fusão

Visando simplificar a validação, o quadro de discernimento gerado contém apenas dois possíveis elementos que representam o estado da rede $\Theta = \{NORMAL, ANÔMALO\}$. Considerando que a rede encontra-se por maior tempo em estado NORMAL, a hipótese a ser questionada será sempre se o estado da rede é NORMAL. Desta forma, com o auxilio da EvidenZ, serão calculados a função de crença e a plausibilidade da hipótese, lembrando que esse cálculo leva como base a combinação das análises. De posse destes dois elementos é possível determinar o intervalo de crença que expressa à faixa de valores no qual é possível acreditar na hipótese H, ou seja, se é possível acreditar no estado normal da rede.

5. AVALIAÇÃO E RESULTADOS

Esta seção descreve o processo de avaliação da eficiência do ADS-Fusion. Para tanto serão descritos o ambiente de teste, o tráfego anômalo gerado e os resultados obtidos.

5.1. Ambiente de Teste

Para realização dos testes foram utilizadas as instalações do GPRT (Grupo de Pesquisas em Redes e Telecomunicações) da Universidade Federal de Pernambuco (UFPE), visando criar um ambiente controlado que se assemelha ao real e capaz de permitir anomalias de tráfego. Foram utilizados 52 PCs desktops, 2 switches com 24 portas 10/100/1000 Mbps e 2 servidores (processador Athlon XP 4200+ 64bits, com 2Gbytes de RAM e 160Gbytes de HDD), um para a execução do ADS-Fusion. Windows XP e Ubuntu Linux foram os sistemas operacionais utilizados.

5.2. Tráfego de Ataque

Na geração do tráfego anômalo foram utilizados 4 PCs executando um script de ataque que emprega a ferramenta Packit [Intrusense 2008] na criação de pacotes customizados e com endereços IP reais e/ou forjados. Foram escolhidos três tipos comuns de ataques:

TCP SYN de alta carga (*flood*), TCP SYN de baixa carga e SPAM. Os dois primeiros originados fora da rede do GPRT e destinados a um servidor dentro dela. O último foi originado dentro da rede GPRT e destinado a uma rede externa.

É importante ressaltar que a escolha por estes tipos de ataque se justifica pela presença de um sensor especializado em ataques DDoS baseados no protocolo TCP. Não foram incluídos ataques típicos dos protocolos UDP e ICMP.

5.3. Resultados

5.3.1. Ataque TCP SYN Flood

O ataque TCP SYN flood foi gerado com pacotes com endereço MAC, endereço IP de origem, porta de origem e porta de destino randômico. Cada pacote possui 1500 bytes de tamanho. A taxa de geração de pacotes é escolhida aleatoriamente, podendo variar entre 1000 e 8000 pacotes por segundo.

O ataque ocorreu no dia 20 de Novembro de 2008 entre as 13h00min e 14h00min, horário de grande número de acessos a rede GPRT. A figura 2 ilustra um cenário com dois ataques (o primeiro ocorrido às 13h55min com duração de 30 segundos e o segundo as 13h58min com duração de 20 segundos), onde é possível notar o aumento no número de fluxos (de aproximadamente 1450 fluxos por segundo para 8500 fluxos por segundo) durante os períodos de ataque.

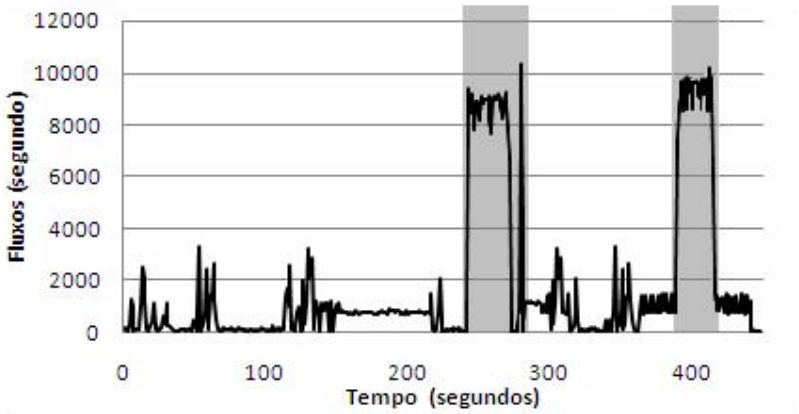

Figura 2. Cenário de ataque TCP SYN flood

O tráfego gerado foi capturado e transformado em fluxos para análise de ambos os sensores (TCPModel e Profiling). A figura 3 ilustra a detecção de anomalias efetuada por ambos os sensores durante o primeiro ataque às 13h55min, o que corresponde ao intervalo de tempo entre 200 e 300 segundos do trace capturado da figura 2.

É facilmente percebido pela figura 3 que ambos os métodos detectam o TCP SYN flood. Contudo, o sensor Profiling apresenta uma resposta mais rápida. No ponto 1 da figura, o TCPModel "demora" a atingir a carga máxima do ataque. No ponto 2, um resquício do ataque (provavelmente pacotes perdidos pela rede) é percebido somente pelo Profiling. Por outro lado, o TCPModel apresenta-se mais estável durante o período do ataque.

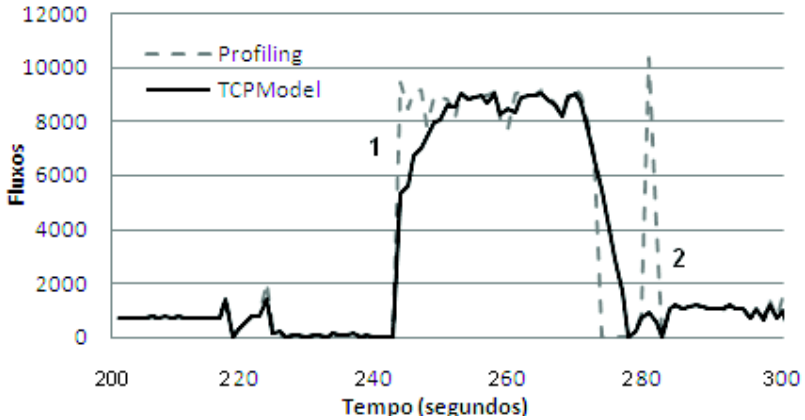

Figura 3. Detecção do Profiling e TCPModel

Uma vez que ambos os sensores foram capazes de detectar o ataque, o ADS-Fusion foi empregado para aumentar a precisão da detecção. Os resultados apresentados na tabela 1 exprimem tanto a crença individual dos sensores quanto do mecanismo de fusão sobre o estado NORMAL da rede e não sobre o estado de ataque.

Tabela 1. Resultados da fusão para o ataque TCP SYN Flood

Tempo	Profiling	TCPModel	ADS-Fusion
200-210	80%	83%	98,3%
210-220	78%	81%	97,9%
220-230	74%	72%	96,36%
230-240	82%	86%	98,74%
240-250	36%	58%	36,09%
250-260	16%	12%	8%
260-270	19%	13%	9,26%
270-280	53%	46%	87,31%
280-290	8%	78%	17,6%
290-300	81%	79%	98%

A primeira coluna representa a linha temporal da análise do primeiro ataques TCP SYN (100 segundos). As coluna do Profiling e do TCPModel exprimem a crença atribuída ao tráfego normal de acordo com as técnicas utilizadas para detecção. Percebe-se que entre os intervalos entre 240-250 e 260-270 (período de duração do ataque), ambos os métodos foram capazes de detectar o ataque e o resultado da fusão de ambas as crenças (coluna ADS-Fusion) apenas corroborou com as técnicas, aumentando a precisão do tráfego avaliado não ser normal.

Os intervalos entre 270-280 e 280-290, como explicado anteriormente, uma grande quantidade de fluxos originados pelo ataque foi detectada pelo sensor Profiling, onde a crença na atividade normal caiu de 53% para 8%, enquanto que o sensor TCPModel não registrou essa anomalia. Para o mecanismo de fusão essas crenças são conflitantes, mas mesmo assim podem ser combinadas. Para o intervalo entre 270-280,

a crença na normalidade do tráfego é alta, mas para intervalo seguinte, a crença, influenciada pelo Profling, aumentou a crença em que o tráfego não seja normal.

5.3.2. SPAM

Visando não somente avaliar ataques, foi realizada a injeção de tráfego SMTP visando à criação de "SPAM" para o servidor de email externo. Mais uma vez foi utilizada a ferramenta Packit para gerar tráfego forjado com duração de 60 segundos.

A figura 4 mostra que somente o sensor Profiling foi capaz de detectar a anomalia. Tal fato pode ser explicado da seguinte forma: para o TCPModel, o tráfego de e-mail é um tráfego legítimo tendo em vista que faz uso do mecanismo de *handshake* para estabelecer conexões. Por outro lado, o Profiling detecta a anomalia uma vez que o número de fluxos enviados ao mesmo destino cresce consideravelmente.

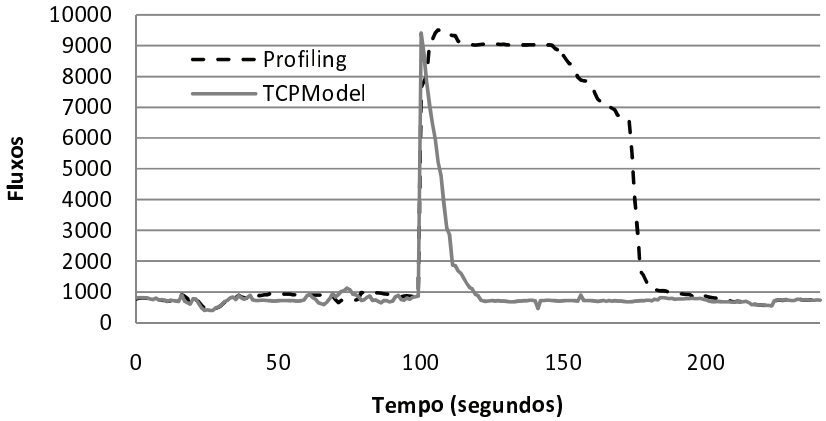

Figura 4: Detecção do tráfego de SPAM.

No processo de fusão dos dados, a não adequação do TCPModel influencia diretamente os resultados. A tabela 2 demonstra os resultados.

Tabela 2. Resultados da Fusão para o tráfego SPAM

Tempo	Profiling	TCPModel	ADS-Fusion
80-100	80%	81%	98,10%
100-120	19%	34%	13,27%
120-140	22%	81%	21,24%
140-160	37%	78%	23,07%
160-180	41%	82%	44,69%
180-200	78%	80%	97,8%

Os intervalos entre 100-120 a 140-160 representam o ataque e a detecção pelos sensores. Uma vez que o TCPModel não registrou essa anomalia, uma combinação padrão da TDS não seria capaz representar a anomalia. A solução encontrada foi trabalhar com crença especialista, ou seja, foi atribuído um peso maior as evidências geradas pelo Profiling. Na prática, o TCPModel não gera alertas e, então, sua inferência relacionada a anomalia não entra na combinação.

5.3.3. Ataque TCP SYN de baixa carga

O último ataque testado foi um TCP SYN de baixa carga, ou seja, foi empregada uma taxa de aproximadamente 48 pacotes por minuto. A figura 5 ilustra a detecção do ataque.

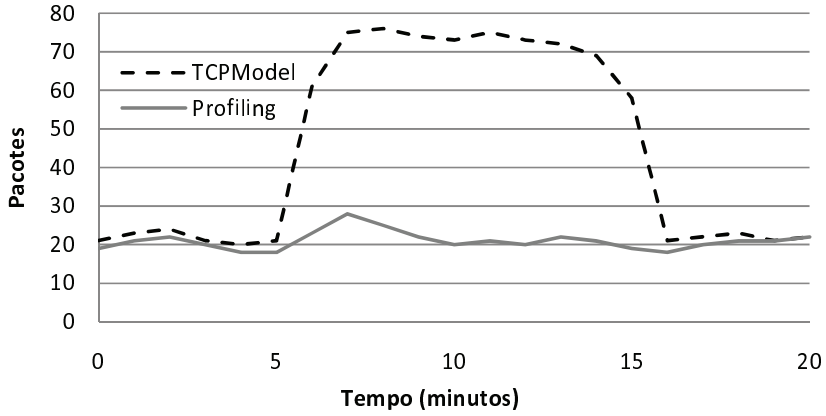

Figura 5. TCP SYN de baixa carga

Uma vez que este tipo de ataque é difícil de detectar, pois gera uma pequena quantidade de fluxos, o Profiling não é capaz de detectá-lo. Por outro lado, o TCPModel consegue perceber o número de pacotes TCP SYN destinados ao mesmo endereço e assim é capaz de detectar o ataque.

De modo similar ao SPAM, mais uma vez um dos sensores não foi capaz de detectar o ataque. Novamente, o uso de crença especialista foi utilizado. A tabela 3 exemplifica os resultados da combinação.

Tabela 3. Resultados da Fusão para o ataque de baixa carga

Tempo	Profiling	TCPModel	ADS-Fusion
0-5	78%	83%	98,13%
5-10	62%	21%	24,99%
10-15	74%	24%	20,12%
15-20	73%	67%	95,54%

6. Conclusão

Hoje em dia, anomalias de tráfego continuam causando inúmeros prejuízos a empresas e instituições. Ataques de negação de serviço, varreduras (*scans*), *worms*, vírus e outros tipos de males ainda geram problemas a milhares de administradores de rede e até mesmo a usuários comuns. Apesar do desenvolvimento e aperfeiçoamento das técnicas de detecção, o número de falsos positivos destes mecanismos ainda é preocupante. Desta forma, este estudo contribui com a discussão do uso da TDS aplicada como técnica de fusão de dados.

Outra importante contribuição deste trabalho é o uso da TDS para o estabelecimento de regras de avaliação de anomalias. Todos os resultados obtidos pela ADS-Fusion podem ser empregados para montar um histórico da(s) anomalia(s),

evitando assim processamento desnecessário. É importante ressaltar que a TDS requer um alto poder de processamento para tratar enumeras evidências (um grande *frame* de discernimento, por exemplo). Esta funcionalidade pode ser empregada em uma versão on-line.

Para resumir, este trabalho apresentou uma visão geral do mecanismo de fusão de dados e elementos da teoria da evidência, além de descrever estudos com fusão de dados visando à detecção de anomalias de rede. Também desenvolveu um protótipo capaz de agregar dados sobre anomalias baseado na inferência de sensores espalhados na rede e inferir o real estado da rede monitorada. A validação do protótipo foi realizada utilizando-se dados sintéticos contendo diversos ataques gerados na rede do GPRT-UFPE. Apesar dos resultados estarem em estado inicial e ainda existir um longo caminho até uma solução completa e definitiva, percebe-se que é possível e definitivamente positivo o uso da teoria da evidência como técnica de fusão de dados.

Como trabalho futuro pretende-se: (i) modificar o ADS-Fusion para ser capaz de operar com coleta e análise de dados em modo on-line, gerando assim alertas em tempo real; (ii) implementar outros tipos de regras de conflito para o módulo de fusão TDS como, por exemplo, os propostos em [Campos 2005][Sentz e Ferson 2002], visando um melhor tratamento na resolução de conflitos de evidências; (iii) desenvolver ou adaptar outros sensores de forma a aumentar a abrangência de anomalias que possam ser detectadas.

References

Abry, P., Borgnat, P. e Dewaele, G. (2007) "Sketch based anomaly detection, identification and performance evaluation", In *IEEE/IPSJ SAINT Measurement Workshop*, páginas 80-84.

Aschoff, R. R. (2007) "ChkModel: Um Mecanismo de Defesa Contra Ataques DDoS", Trabalho Final de Graduação. Centro de Informática. Universidade Federal de Pernambuco.

Campos, F. F (2005) "Uma extensão a matemática da evidência", Tese de Doutorado. Centro de Informática. Universidade Federal de Pernambuco.

Chen, Q. e Aickelin, U. (2006) "Anomaly Detection Using the Dempster-Shafer Method", In *International Conference on Data Mining, DMIN 2006*, Nevada, USA.

Dempster, A. P. (1967) "Upper and Lower Probabilities Induced by a Multivalued Mapping", In *Annals Mathematics Statistics*, 38, páginas 325-339.

Dempster, A. P. (1967) "Upper and Lower Probability Inferences Based on a Sample from a Finite Univariate Population", *Biometrika*, 54, páginas 515-528.

EvidenZ. (2008), http://www.lrde.epita.fr/cgi-bin/twiki/view/Projects/Evidenz

Gao, J., Hu, G., Yao, X., and Chang, R. (2006) "Anomaly Detection of Network Traffic Based on Wavelet Packet", In *Asia-Pacific Conference on Communications (APPC´06)*.

Gordon, J. e Shortliffe, E. H. (1984) "The Dempster-Shafer Theory of Evidence", *Rule-based expert systems*. New York, Addison-Wesley, páginas. 272-292.

Intrusense. (2008) "Packit", http://www.intrusense.com/software/packit/.

Karagiannis, T., Papagiannaki, K. e Faloutsos, M. (2005) "BLINC: Multilevel traffic classification in the dark", *ACM SIGCOMM Computer Communication Review*, 35(4), páginas 229-240. ACM Press.

Lakhina, A., Crovella, M. e Diot, C. (2005) "Mining anomalies using traffic feature distributions", *ACM SIGCOMM Computer Communication Review*, 35(4), páginas 217-228. ACM Press.

Li, L. e Lee, G. (2005) "DDoS attack detection and wavelets", In *Telecommunication Systems*, Vol. 28, No. 3-4, páginas 435-451.

Mirkovic, J., Prier, G. e Reiher, P. (2002) "Attacking DDoS at the Source", In *Proceedings of 10th IEEE International Conference on Network Protocols*, páginas 312-321, Paris, França, Novembro.

Mirkovic, J., Robinson, M., Reiher, P. e Kuenning, G. (2003) "Forming Alliance for DDoS Defenses", In *Proceeding of the New Security Paradigms Workshop (NSPW 2003)*, ACM Press, páginas 11-18, Agosto.

LIBPCAP. (2008) TCPDUMP/LIBPCAP Public Repository, http://www.tcpdump.org/

Pearl, J. (1982) "Reverend Bayes on inference engines: a distributed hierarchical approach", In *Proceedings of the Second National Conference on Artificial Intelligence*, Pittsburgh, PA, páginas. 133-136.

Scherrer, A., Larrieu, N., Owezarski, P., Borgnat, P. e Abry, P. (2007) "Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies", *IEEE Transactions on Dependable and Secure Computing*, volume 4, páginas 56-70.

Sentz, K. e Ferson, S. (2002) "Combination of Evidence in Dempster-Shafer Theory", Relatório Técnico. http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf

Siaterlis, C., Maglaris, B. e Roris, P. (2003) "A novel approach for a Distributed Denial of Service Detection Engine", Relatório Técnico, Network Management and Optimal Design (NETMODE) Lab; National Technical University of Athens.

Shafer, G. (1976) "A mathemathical theory of evidence". Princeton, Princeton University Press.

Snort. (2008) "Snort: The open source network intrusion detection system". http://www.snort.org. Outubro.

Sundaram, A. (1996) "An Introduction to Intrusion Detection", ACM Crossroads 2.4.

Tian, J., Zhao, W, Du, R. e Zhang, Z. (2005) "D-S Evidence Theory and its Data Fusion Application in Intrusion Detection", In *Sixth International Conference on Parallel and Distributed Computing Applications and Technologies*, páginas 115 – 119

Uchôa, J.Q.; Panontim, S.M.; Nicoletti, M.C. (1997) "", Relatório Técnico RT-DC 007/97, UFSCar/DC, São Carlos, páginas 1-29.

Xu, K., Zhang, Z-L. e Bhattacharrya, S. (2005) "Profiling internet backbone traffic: Behavior models and applications", *ACM SIGCOMM Computer Communication Review*, 35(4), páginas 169-180. ACM Press.

Zadeh, L. A. (1978) "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, 1, páginas 3-28.