Detecção de ataques de negativa de serviço por meio de fluxos de dados e sistemas inteligentes

Adriano M. Cansian, Jorge L. Corrêa

UNESP – Universidade Estadual Paulista – Instituto de Biociências, Letras e Ciências Exatas (IBILCE) - Campus de São José do Rio Preto, SP - Brazil
ACME! Computer Security Research

Abstract. *This article presents a new model to anomalies and intrusion attempts detection based on the use of network flows (Netflow standard) and in the classification capacity of the artificial neural networks. The model is characterized by the behavior based detection of network environment together with the capacity of knowledge absorption of the intelligent systems. A new concept of signature is used, being tested several models along the evolution of the system. Several attacks like DoS, DDoS and worms activities are detected quickly, in a scalable and automated way for medium and big load environment, characterizing an effective monitor model for networks connected to the Internet.*

Resumo. *Este artigo apresenta um novo modelo de detecção de anomalias e tentativas de intrusão baseado na utilização de fluxos de dados (padrão Netflow) e na capacidade classificatória das redes neurais. O modelo caracteriza-se pela detecção baseada no comportamento do ambiente de rede juntamente com a capacidade de absorção de conhecimento dos sistemas inteligentes. Um novo conceito de assinatura é utilizado, sendo testados diversos modelos ao longo da evolução do sistema. Ataques como DoS, DDoS e atividades de worms são rapidamente detectados, de forma automatizada e escalável para ambientes de médio e grande porte, caracterizando um efetivo modelo de monitor para redes conectadas à Internet.*

1. Introdução

A Internet tem proporcionado a seus usuários diversas facilidades ao possibilitar a execução de tarefas no âmbito profissional, de entretenimento, comércio, entre outros. No entanto, juntamente com esta evolução, tornou-se constante o desenvolvimento de atividades ilícitas. Diversas tentativas de ataques ocorrem diariamente contra um ambiente computacional. Ainda, são constantes as tentativas de abusos por parte de usuários que utilizam serviços digitais sujeitos às políticas de segurança e utilização [Manoj 2007].

Acompanhando esta evolução, os monitores de redes são ferramentas importantes para a manutenção da ordem e segurança de um ambiente. Diversas ferramentas são utilizadas para este fim, tais como *firewall*, *proxy*, *access control list* [Allen 2001], filtro de conteúdo e sistemas detectores de intrusão. Considerando o paradigma atual de tráfego, composto por uma diversidade de serviços, altas taxas de

tráfegos e serviços multimídias, torna-se necessário um sistema capaz de monitorar e responder à eventos maliciosos de forma eficaz e escalável.

Este artigo trata de uma nova metodologia de monitoria de rede baseada na utilização de fluxos de dados como fonte de informação. Trata de uma arquitetura de monitoria cujas principais características são: adaptabilidade ao ambiente, de forma que o sistema possa ser empregado em redes distintas absorvendo as características comportamentais do ambiente, detecção de anomalias baseada no comportamento, escalabilidade e automatização do processo de detecção, determinando a ocorrência de diversas atividades maliciosas, como tentativas de prospecção, ataques de negativa de serviço e abusos.

Na seção seguinte serão mostradas algumas metodologias atualmente empregadas no monitoramento de um ambiente de rede. A seção 3 contém a descrição do modelo adaptativo de detecção de anomalias baseado na análise de fluxos de dados *Netflow*. Nas seções 4 e 5 são analisados os modelos de assinatura deste sistema. Na seção 6 são demonstrados os resultados obtidos e na seção 7 as conclusões.

2. Metodologias atuais

A monitoria de ambientes computacionais é uma tarefa realizada desde que os primeiros computadores passaram a se comunicar. Diversos sistemas são utilizados para aumentar a segurança de um ambiente e garantir a privacidade, confiabilidade e disponibilidade dos sistemas. Cada um deles possui seus limites de atuação e suas especialidades.

Nas últimas décadas, o número de redes conectadas à Internet tem crescido vertiginosamente. Aliado a este crescimento está a disseminação dos *worms* e ataques de negativa de serviços (*DoS* e *DDoS*), eventos que se caracterizam por interferir na eficiência de uma rede. Os *worms* possuem grande poder de auto-disseminação tornando-se a principal ameaça na Internet atual [Cert.br1 2006]. Os ataques *DoS* e *DDoS* ocorrem constantemente e visam afetar a disponibilidade no oferecimento de serviços.

Garantir a imunidade contra estes eventos é de extrema importância para os sistemas de rede atuais, devido a criticidade dos processos executados *online*, como transações bancárias e realização de pesquisas, por exemplo. Os monitores de rede atuam neste contexto, auxiliando administradores na detecção destas atividades maliciosas.

Dentre os monitores que atuam em uma rede podemos destacar os *Network Intrusion Detection Systems (NIDS)* [Mukherjee 1994]. Estes sistemas são capazes de analisar o tráfego de um ambiente em busca de informações que caracterizem atividades maliciosas. Normalmente baseiam-se em informações de *logs* (registros de auditoria) geradas por analisadores de pacotes. O método de detecção destas atividades é baseado na utilização de assinaturas de ataque. Estas assinaturas são registros que representam exatamente o evento o qual se deseja detectar. Por exemplo, uma assinatura pode ser baseada na análise de *strings* do campo de dados dos pacotes trafegados. Quando uma determinada *string* ocorrer em uma assinatura e for encontrada em um pacote, este representará uma possível atividade maliciosa. A principal desvantagem desta metodologia está na necessidade do desencapsulamento de todos os pacotes, o que pode

interferir na eficiência da rede, principalmente em atividades sensíveis ao atraso, como dados multimídia.

Outra abordagem utilizada por alguns monitores consiste na verificação dos cabeçalhos de pacotes em busca dos denominados pacotes mal-formados. Nestes pacotes, os cabeçalhos não estão de acordo com os padrões dos protocolos, definidos nos RFC's. Por exemplo, pacotes TCP podem conter uma combinação de *flags* não definidas no padrão, e serem utilizados para exploração de uma vulnerabilidade em aplicações que não estejam preparadas para receber esta combinação. A desvantagem desta abordagem é não possuir detalhes sobre o evento, baseando-se apenas na checagem de cabeçalhos. Ainda, possuem dificuldades na detecção de eventos como *DoS* e *DDoS*.

Trabalhos semelhantes foram realizados nesta área como [Bonifacio e Cansian 1997]. O modelo proposto em [Bonifacio e Cansian 1997] apresenta dificuldades na detecção das classes de ataque de negativa de serviço. O modelo de detecção é baseado na análise do conteúdo dos pacotes (*payload*) e na geração de assinaturas a partir de padrões léxicos e semânticos do conteúdo dos pacotes que caracterizam um ataque. O modelo proposto neste artigo visa complementar as capacidades deste sistema, principalmente possibilitando a detecção de eventos de negativa de serviço. Uma vez que os fluxos *Netflow* são representações concisas das sessões em uma rede, o modelo baseia-se na geração de padrões de tráfego, levando em consideração a quantidade de fluxos gerados na rede. Os padrões são divididos em normais e anômalos e codificados em vetores de estímulo, utilizados para o treinamento da rede neural. A não análise do conteúdo dos pacotes acrescenta a característica de escalabilidade permitindo o monitoramento de redes com alta carga de tráfego.

Neste trabalho é apresentado um novo modelo de monitor de rede. O principal objetivo deste sistema é detectar atividades maliciosas em uma rede de grande porte sem afetar sua eficiência. Este modelo baseia-se em informações fornecidas pela tecnologia de fluxos de dados e pela utilização das características dos sistemas inteligentes de classificação de padrões.

3. Modelo adaptativo de detecção de anomalias

Um novo modelo de monitor de rede foi desenvolvido com o objetivo de monitorar ambientes de forma escalável, automatizada e baseada no comportamento [Geer 2006] do tráfego. Este modelo utiliza a tecnologia de fluxos de dados para geração de informações e um sistema classificador, capaz de indicar a ocorrência de uma anomalia no ambiente. Os fluxos de dados garantem a escalabilidade do processo de análise de pacotes. Esta tecnologia gera informações sobre o tráfego do ambiente baseada apenas nos cabeçalhos dos pacotes, não necessitando analisar seus conteúdos, método que gera latência na rede. Isto permite o monitoramento de um grande perímetro de rede. As redes neurais artificiais são responsáveis por conferir a capacidade de adaptabilidade e detecção ao sistema.

Este modelo baseia-se na utilização de assinaturas que descrevem um evento. No entanto, algumas peculiaridades são verificadas, principalmente pela utilização dos sistemas inteligentes de classificação. De modo geral, o estabelecimento do sistema segue as seguintes etapas:

- exportação dos fluxos em um ponto de observação (normalmente roteadores);
- coleta dos fluxos;
- geração de assinaturas contendo padrões normais e anômalos;
- treinamento do sistema neural;
- integração do sistema neural com o coletor de fluxos;
- monitoria em tempo real (geração e verificação de assinaturas).

3.1 Padrão IPFIX e *Netflow* de fluxo de dados

O conceito de fluxos de dados surgiu da necessidade de uma metodologia para análise de rede. Inicialmente, o IETF (*Internet Engineering Task Force – Internet Society*), órgão regulador de padrões para Internet, propôs a criação de um padrão cuja finalidade era estabelecer uma arquitetura para análise de tráfego na Internet [Quittek 2004]. Subseqüentemente, diversos padrões foram propostos, dentre eles o *Netflow* [Claise 2004], pela *Cisco Systems*, padrão utilizado neste projeto.

Um fluxo *Netflow* é definido como uma seqüência unidirecional de pacotes entre dois *hosts*, ou seja, é a representação de um tráfego com características comuns. Podemos entender um fluxo como uma tupla na qual as seguintes informações aparecem com o mesmo valor:

- endereço IP de origem e de destino;
- porta de origem e destino (referente ao protocolo da camada de transporte);
- valor do campo *Protocol* do datagrama IP;
- *byte Type of Service* do datagrama IP;
- interface lógica de entrada do datagrama no roteador ou *switch*.

Estes campos permitem que um fluxo represente concisamente o tráfego através de um ponto de observação (normalmente um roteador), no nível de cada campo de seus protocolos. Todos os pacotes com um mesmo valor nos campos desta tupla são contados, agrupados e empacotados em um registro de fluxo. Estes registros de fluxos são então exportados a um coletor.

Uma vez que os fluxos gerados nos equipamentos são exportados e armazenados, passam a constituir fonte valiosa de informações. É possível obter detalhes de cada conexão ou sessão estabelecida por qualquer máquina pertencente ao ambiente monitorado, fator este extremamente relevante na análise de segurança.

3.2 Redes neurais artificiais (RNAs)

Uma rede neural artificial (RNA) é um grupo de neurônios artificiais interconectados os quais utilizam um modelo matemático e computacional para o processamento de informações. São sistemas baseados na arquitetura do cérebro humano capazes de absorver conhecimento através de ajustes dos pesos nas conexões entre cada neurônio, chamadas sinapses. Este processo é conhecido como treinamento.

O modelo desenvolvido passa por algumas etapas até seu estabelecimento final. Uma destas etapas é o treinamento, na qual o sistema absorve conhecimento sobre os eventos as quais deve detectar. No modelo desenvolvido foi utilizada uma rede neural do tipo MLP (*Multi Layer Perceptron*) [Longstaff 1987]. O treinamento do sistema foi baseado no algoritmo *backpropagarion* [Wilkinson 1989], considerado extremamente eficiente para este tipo de RNA. A opção pelas redes MLP foi determinada tanto pelo

modelo de treinamento supervisionado, que possibilita a inserção de conhecimentos específicos de cada evento, quanto pela sua alta eficiência no reconhecimento de padrões, como já demonstrado em trabalhos da área como [Bonifacio e Cansian 1997].

A realização do treinamento é um processo que envolve a interação do administrador com o sistema. Uma vez que os fluxos do ambiente estejam armazenados, o administrador deve selecionar aqueles que considere relevantes para o ambiente. Este conjunto de padrões é utilizado para o treinamento do núcleo inteligente, conferindo conhecimento sobre os eventos detectáveis. Durante o desenvolvimento, foram utilizadas algumas ferramentas para esta análise inicial, como o *Flow-capture* [Fullmer 2006], um coletor de fluxos, e o *Flow-tools* [Fullmer 2006b], um conjunto de ferramentas para manipulação de fluxos.

Uma das principais características das RNAs é seu poder de generalização. Quando uma RNA é treinada com um conjunto de padrões, ela torna-se apta a reconhecê-los. No entanto, a rede desenvolve capacidade de classificar padrões nunca antes vistos. Dessa forma, se um padrão *X* representa algum evento (um ataque, por exemplo), a rede classificará este padrão como o evento *Ataque A*. Se a rede for submetida a um padrão de entrada nunca antes visto, ela será capaz de classificá-lo como sendo um evento *Ataque A* caso possua alguma semelhança com o evento *X*, conhecido do treinamento. Esta classificação é realizada com base em uma probabilidade. Desta forma, o modelo responderá qual o evento detectado e com qual probabilidade. Isto deixa clara a eficiência da utilização de RNAs na detecção de anomalias em monitores de redes, tarefa na qual conhecemos alguns padrões e necessitamos de um sistema inteligente que tome decisões com um certa precisão para novas ocasiões.

A topologia desta rede neural depende do modelo de assinatura utilizado no ambiente. Foram desenvolvidos três modelos diferentes durante a evolução do sistema. Cada um destes modelos está diretamente relacionado com o conceito de assinatura. Veremos nas seções seguintes a definição do conceito de assinaturas para este modelo de monitor, bem como suas capacidades.

4. O conceito de assinatura

Uma assinatura de ataque é uma representação de aspectos, condições, disposição e inter-relação entre eventos que possam descrever um ataque [Cansian 2002]. Em um sistema detector de intrusos uma assinatura é utilizada como parâmetro de busca para os eventos detectáveis. Assim, as informações de auditoria, como registros de *log* e pacotes capturados, são analisadas e comparadas com as assinaturas. Caso seja encontrada uma correspondência, o evento em questão será detectado e considerado como uma possível ameaça.

Normalmente, a metodologia utilizada é conhecida como *fingerprint*. O princípio de funcionamento é encontrar uma cópia fiel da assinatura no sistema monitorado, ou seja, nos arquivos de *logs* ou pacotes analisados. Embora existam alguns algoritmos que melhorem a eficiência deste método, normalmente pequenas mudanças em algum evento implicam na geração de uma nova assinatura. Por exemplo, se um evento *A* é detectado por sua assinatura correspondente, um evento *A'* que seja uma

pequena modificação do evento original não será detectado utilizando a mesma assinatura.

Neste novo modelo de detecção, as assinaturas são representações de padrões de fluxos ocorridos no ambiente monitorado. Podem representar tanto eventos ilícitos quanto lícitos, de forma compatível com o modelo neural e, com a capacidade de detecção de eventos semelhantes a partir de uma mesma assinatura. Embora seja utilizado o conceito de assinatura, o modelo proposto pode ser classificado como um detector de eventos por anomalia, em contrapartida aos detectores por abuso que utilizam inerentemente assinaturas precisas dos eventos detectáveis.

4.1 Diferenças entre a metodologia de fingerprint e o novo conceito de assinatura

A utilização de uma rede neural como meio de detecção de anomalias implica diretamente na estrutura de uma assinatura. Embora o modelo do Perceptron de Rosenblatt [Rosenblatt 1958][Rosenblatt 1962] permita representar os valores de entrada de uma RNA como contínuos, as assinaturas deste modelo utilizam uma representação binária com zeros (0s) e uns (1s). Esta representação permite que uma assinatura seja visualizada de maneira semelhante a um gráfico cartesiano, de modo que possam ser escolhidas aquelas a compor o conjunto de padrões de treinamento. A geração de cada assinatura dependerá da escolha de um dos modelos apresentados na seção seguinte. Este processo é realizado por um módulo do sistema, capaz de quantificar os fluxos armazenados e gerar sua representação binária em uma linha de tempo para análise de um administrador. A partir desta representação, são selecionados padrões que representam tanto anomalias no tráfego quanto sua normalidade. Este mesmo módulo é utilizado então para geração do conjunto de treinamento (assinaturas) a partir dos padrões selecionados pelo administrador. De modo geral, os fluxos de dados são lidos do coletor e quantificados, gerando uma visualização binária na forma cartesiana (Figura 1). Esta quantificação permite a identificação de anomalias, pois eventos ilícitos caracterizam-se por causar distúrbios no número de fluxos, em relação ao comportamento normal da rede.

Embora uma seqüência de 0s e 1s não seja sugestiva, o modelo adaptativo de detecção tem como característica concentrar o conhecimento sobre os ataques dentro da RNA, um sistema inteligente capaz de absorver este conhecimento, e não apenas na assinatura. Uma vez que o conhecimento sobre um evento seja absorvido, uma assinatura será utilizada apenas para codificar um evento e ser testada pelo núcleo neural do sistema. Assim, qualquer evento que queiramos representar e analisar deve ser codificado seguindo este padrão.

O conceito de assinatura, até então utilizado como sendo uma representação concisa e exata de um ataque, apresenta uma modificação importante no que diz respeito a sua utilização. Nos monitores e IDSs tradicionais as assinaturas são utilizadas apenas na tarefa de detecção, ou seja, cada evento será comparado com uma assinatura durante a monitoria de um sistema ou ambiente. Neste novo modelo, além da utilização das assinaturas para verificação dos eventos, o papel fundamental está no treinamento da RNA. É durante este processo que todo conhecimento é passado ao sistema. O processo de classificação continua a existir. Adiciona-se a característica de detectar eventos nunca antes vistos que se assemelhem a algum evento visto no treinamento, e

não apenas os eventos que possuam uma assinatura que os representem exatamente, como na metodologia de *fingerprint*.

5. Modelos de assinatura

Um dos objetivos primordiais do modelo adaptativo desenvolvido é permitir a detecção de eventos em tempo real. Os eventos de segurança têm como característica marcante estarem relacionados com o tempo. Considerando a análise temporal, juntamente com as necessidades de codificação para utilização de uma RNA, foram desenvolvidos três modelos de assinaturas. De forma geral, estes modelos de assinaturas são gráficos cartesianos relacionando *fluxos x tempo*, codificados de forma compatível com a entrada de uma rede neural. Em cada um dos modelos, cada linha representará uma média de fluxos recebidos, durante um período de tempo. Este período de tempo será representado por cada coluna, de forma discreta. Tanto o número médio de fluxos quanto o período são parâmetros escolhidos pelo administrador do sistema sendo, portanto, valores dinâmicos. Por exemplo, podemos considerar que cada coluna representa uma janela de tempo de 30s e cada linha uma média de 20 fluxos por segundo. A presença do algarismo 1 em uma linha representa que foram recebidos 600 fluxos por segundo (uma média de 20 fluxos por segundo vezes os 30 segundos de uma coluna: 20 x 30 fluxos por segundo). No entanto, não existe uma maneira determinística para definição destes valores, sendo estimada uma média a partir da análise de uma determinada quantidade de fluxos para o ambiente a ser monitorado, a fim de auxiliar a configuração final destes parâmetros.

5.1 Análise total de fluxos

Este modelo de assinatura não faz distinção entre tráfego de entrada e tráfego de saída. Para sua criação, o sistema utiliza uma função de quantificação considerando *todos* os fluxos, independente do tráfego sair ou entrar na rede. A figura 1 mostra a estrutura obtida.

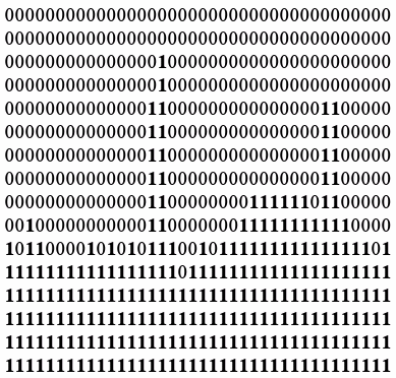

Figura 1: Modelo 1 – análise total de fluxos.

Trata-se de um modelo simples cuja estrutura é um gráfico cartesiano apenas com os valores positivos em seus eixos, representando todos os fluxos sem nenhuma diferenciação. Com relação à rede neural, este modelo requer uma quantidade menor de assinaturas durante o treinamento, devido sua grande abrangência em relação aos tipos de fluxos analisados. No entanto, torna-se mais difícil distinguir anomalias diferentes,

visto que a assinatura tem como característica principal a generalização dos fluxos do ambiente.

Embora simples, este modelo mostra uma grande capacidade no reconhecimento dos padrões que caracterizam eventos ilícitos, principalmente eventos que causam grandes distúrbios em uma rede. A perda de eficiência se dá não pela sua capacidade de reconhecimento, mas pela representatividade da assinatura que não absorve muitas informações sobre o ambiente. No modelo de análise total de fluxos, as assinaturas são simplesmente estruturas que representam uma quantificação, uma contagem média da ocorrência de fluxos, de forma generalizada para todo ambiente.

5.2 Análise de fluxos por destino

Assim como as assinaturas por análise total de fluxos, na análise por destino ainda podemos entendê-las como um gráfico cartesiano. No entanto, passamos a considerar o eixo das ordenadas de forma diferenciada. Neste modelo, os fluxos correspondentes aos tráfegos entrantes na rede são quantificados separadamente dos fluxos correspondentes aos tráfegos para redes externas. Desta forma, a assinatura fica dividida, com as ordenadas positivas representando os fluxos enviados e as ordenadas negativas representando os fluxos recebidos. A figura 2 mostra um exemplo desta estrutura.

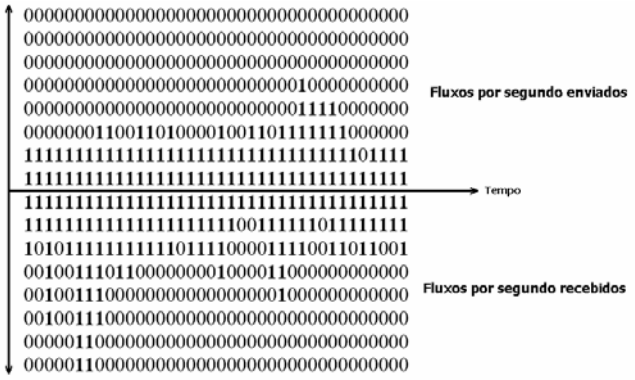

Figura 2: Modelo 2 – análise de fluxos por destino.

Este modelo proporciona ganhos quando comparado ao modelo de análise total. Por separar os fluxos em enviados e recebidos, a diferenciação entre um ataque recebido e um ataque gerado a partir da rede monitorada torna-se factível. Esta possibilidade facilita as tarefas de gerenciamento e monitoria. Considerar a possibilidade de ser um atacante, não somente uma vítima, faz parte do modelo atual de análise de segurança. Ameaças como os *worms* e situações de infecção interna são cada vez mais comuns. Estas ameaças caracterizam-se por serem automaticamente disparadas e disseminadas, sem o conhecimento do usuário. Desta forma, não é incomum a ocorrência de ataques partindo do próprio ambiente monitorado para redes externas. O modelo de análise por destino possibilita ao administrador reconhecer esta ocorrência de forma rápida, tomando providências para que os sistemas infectados envolvidos sejam corrigidos.

Esta mudança no modelo não altera a topologia da rede neural. No entanto, implica no processo de treinamento do sistema, tornando-o um pouco mais complicado. A quantidade necessária de padrões de treinamento aumenta e a geração das codificações torna-se mais detalhada.

A eficiência deste modelo é alta, sendo o reconhecimento de anomalias efetivo nos ambientes em que a diferenciação dos destinos é importante. A possibilidade de rápida contenção na disseminação de *malwares (malicious software)* é uma de suas grandes vantagens.

5.3 Análise de fluxos por serviços e visão geral da rede

O terceiro modelo de assinatura mostra uma maior complexidade, tanto em sua estrutura quando nas implicações de adaptação ao sistema. Este modelo foi criado com base no princípio de que os serviços apresentam comportamentos independentes e diferentes em cada ambiente sob monitoria. A utilização de um modelo que generaliza as características dos fluxos, como o modelo por análise total, leva a dificuldades na geração de alertas e informações mais apuradas sobre o andamento de uma rede. A filtragem de informações, principalmente de segurança, é importante para realização de perícias [Abad 2004]. Atividades periciais iniciadas com base na totalidade de informações geradas por um ambiente são dispendiosas e extremamente caras. Este modelo fornece meios para restringir o domínio de busca em atividades periciais, fornecendo uma maior especificação na análise de fluxos.

Cada serviço executado no ambiente é monitorado separadamente. Embora seja possível analisar todos os serviços, este trabalho baseou-se na monitoria dos mais utilizados no ambiente de testes: FTP, SSH, SMTP, DNS e HTTP. Ainda, buscando ressaltar a característica do comportamento do ambiente, este modelo possibilita uma visão geral de todos os serviços. Existe uma redundância de informações gerada a partir da estrutura da assinatura e da topologia da rede neural.

Comparado com os modelos anteriores, a análise por serviços é uma sobreposição de assinaturas. Os gráficos cartesianos com a representação temporal continuam a existir, agora representando cada serviço do ambiente. Estes gráficos são sobrepostos formando uma estrutura tridimensional. A figura 3 mostra a estrutura deste modelo de assinatura.

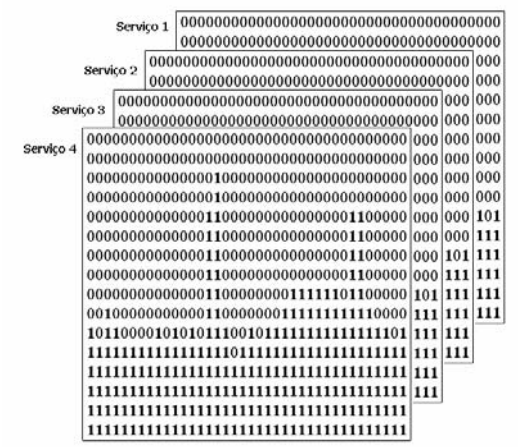

Figura 3: Análise de fluxos por serviços considerando o comportamento normal do ambiente.

Cada serviço será monitorado de forma independente. O processo de treinamento definirá qual é o comportamento normal de cada um deles e, conseqüentemente, o comportamento normal de todos os serviços no ambiente. Esta

visão geral dos serviços é obtida pela extração de conhecimento da própria estrutura do modelo. A figura 4 mostra um esquema da topologia da rede neural para este modelo.

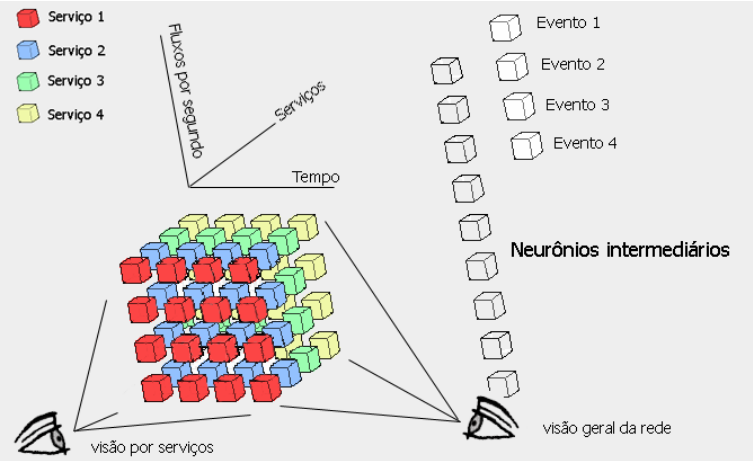

Figura 4: Topologia da rede neural para assinaturas do modelo de análise por serviços.

Todos os serviços possuem suas próprias representações, baseadas nos tempos em que ocorreram os fluxos. A última coluna de cada serviço representa os fluxos mais recentes gerados no ambiente. A extração do conhecimento para compor a visão geral dos serviços é realizada por meio de uma máscara aplicada à estrutura da assinatura. Esta máscara seleciona exatamente os neurônios que representam os fluxos mais recentes de cada serviço, de forma compor um plano que é ligado a uma parte intermediária da rede neural, dedicada a analisar o andamento geral do ambiente.

Devido à complexidade da estruturação deste modelo, o processo de treinamento requer uma quantidade considerável de padrões, que devem ser minuciosamente determinados. A geração deste conjunto é mais complexa devido à necessidade de integração de cada serviço, para gerar o modelo tridimensional de entrada da RNA. No entanto, a eficiência adquirida pelo sistema o torna um monitor extremamente eficaz em uma rede. Qualquer discrepância nos fluxos, em relação ao comportamento normal, poderá ser detectada tanto da parte dos serviços quanto da parte de análise geral do comportamento da rede.

6. Resultados

Os resultados obtidos estão baseados na monitoria de um ambiente com aproximadamente 460 *hosts*, baseado na exportação de fluxos de dados no padrão *Netflow* versão 5. Os fluxos foram gerados no roteador responsável por todo o tráfego do ambiente com a Internet. Ainda, para fins de enriquecimento de dados na geração de padrões de treinamento, foi utilizada uma *Honeynet*, gerando fluxos referentes aos ataques sofridos.

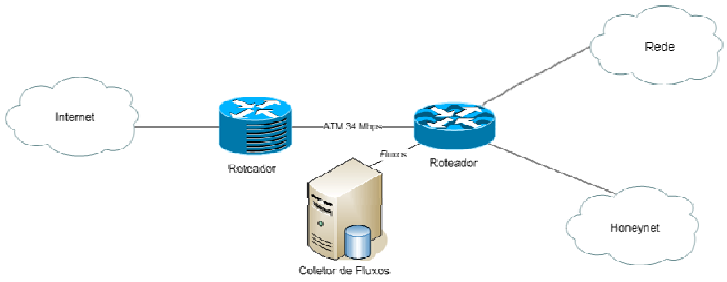

Figura 5: Ambiente monitorado – gateway responsável por todo tráfego com a Internet.

Para monitoria do ambiente, o sistema opera em modo *daemon*. Desta forma, é capaz de receber os fluxos recém exportados, gerar as assinaturas e classificá-las utilizando a rede neural já treinada. Este processo ocorre em tempo real, assim que os fluxos são obtidos do ambiente. Diversos eventos puderam ser observados com este modelo de monitor, utilizando os três modelos de assinatura.

6.1 Eventos detectados

A indicação sobre qual evento foi detectado ocorre pela camada de saída da rede neural. A MLP utilizada possui uma camada de entrada, para o padrão de fluxos, uma camada intermediária e uma camada de saída, na qual cada neurônio representa a ocorrência de um evento. Assim sendo, a camada de entrada possuirá quantos neurônios forem necessários para representação do modelo de assinatura utilizado. A camada intermediária possuirá uma quantidade de neurônios maior, conforme aumentamos a complexidade do modelo de assinatura. Por fim, a camada de saída possuirá um número de neurônios igual ao número de eventos a qual se deseja detectar, eventos estes bem definidos no conjunto de treinamento. A diferenciação entre cada evento se dá pela diferença entre as assinaturas presentes no conjunto de treinamento, absorvidas durante este processo. Portanto, os eventos a serem detectados devem ser escolhidos de maneira criteriosa para geração dos conjuntos de treinamento. Diversos eventos puderam ser detectados durante a fase de teste do sistema.

Tráfego lícito: tráfego que represente o padrão considerado normal no ambiente.

Prospecções de hosts e serviços: anomalia causada por ferramentas maliciosas como *worms* que varrem diversos *hosts* em busca de um serviço específico. Sua representação é um distúrbio na quantidade de fluxos, caracterizando um pico. Os três modelos de assinatura testados mostraram-se capazes de detectar este tipo de anomalia. No entanto, a análise por serviços permite a detecção efetiva destes eventos, indicando exatamente qual serviço encontra-se sob ataque.

DoS e DDoS: estes eventos foram facilmente detectados com todos os modelos de assinaturas, pois sua característica é um distúrbio contínuo na quantidade de fluxos, de forma a causar um pico que permanece por certo período de tempo. Isto pode ser explicado pela abertura de diversas conexões com o *host* alvo, a fim de esgotar seus recursos e fazer com que passe a negar serviços. Uma vez que os fluxos representam todas as conexões de um ambiente, para cada uma delas novos fluxos serão gerados, causando uma discrepância com o comportamento normal.

Eventos FlashCrowd [Jung 2002]: estes eventos são bastante semelhantes aos eventos de negativa de serviço, pois são representados por distúrbios contínuos nos fluxos do ambiente. No entanto, este é um evento legítimo causado por uso acentuado de determinados serviços na rede. São diferenciados dos ataques *DoS* e *DDoS* durante a mudança de estado da rede: enquanto ataques de negativa de serviços causam picos repentinos de fluxos, os eventos *flashcrowd* aumentam a taxa de fluxos gradualmente. Este evento pode ser detectado normalmente, desde que o processo de treinamento seja realizado com padrões bem definidos, diferenciando-os dos ataques de negativa de serviço, principalmente quanto ao período de transição do padrão de fluxos na rede.

Ataques de dicionário [Goyal 2005]: visam serviços com necessidade de autenticação, onde diversas tentativas de *login* são efetuadas com várias combinações de usuários e senhas. O distúrbio causado nos fluxos é semelhante aos ataques de prospecção. Desta forma, os modelos de análise total e por destino apenas detectam uma anomalia, enquanto o modelo de análise por serviços identifica precisamente o serviço atacado.

Uso indevido: este evento é detectado quando o sistema é treinado para alertar sobre uso indevido de recursos, caracterizando violação de políticas de uso e segurança, como por exemplo, aplicações de *file sharing*.

Problemas na conectividade: são eventos legítimos, de cunho não malicioso, mas que interferem no funcionamento de uma rede. São causados por erros de roteamento, falhas de enlaces e perda de conectividade. A diminuição abrupta de fluxos caracteriza um padrão de perda de conectividade, podendo emitir alertas mais específicos, no caso de modelo de serviços.

Tráfego ilícito em canais mitigados (*covert channels*): os canais mitigados caracterizam-se pela ocultação de informações dentro de mensagens legítimas, como a utilização de um servidor de nomes (DNS) especial capaz de estabelecer um túnel IP. Existem várias ferramentas para esta finalidade, como o NSTX [Gil 2006]. Assim, um usuário pode ter um processo servidor de nomes modificado e utilizá-lo como *proxy*. Uma vez que os serviços de autenticação normalmente permitem tráfego DNS (como nos serviços *wireless* em aeroportos e cafés), este indivíduo poderá utilizar um túnel IP para seu servidor e obter acesso à Internet, via encapsulamento de dados utilizando os segmentos UDP do DNS. Este tipo de utilização viola as políticas de uso tanto do ambiente onde o usuário se encontra, quanto do ambiente onde seu processo servidor do túnel executa. Esta atividade é detectada pelo modelo de serviços, pois a utilização de um túnel IP via serviço de resolução de nomes causa distúrbios no padrão normal de utilização do DNS. Apesar de não constar nas codificações das assinaturas, o número de octetos gerados com este evento facilitará sua perícia, uma vez que o túnel trafegará grandes quantidades de informações. Detectada uma anomalia de comportamento, facilita-se a identificação deste evento pelas informações sobre os octetos armazenadas nos fluxos.

6.2 Erros e eficiência

Podemos afirmar que a eficiência do sistema de detecção de anomalias está intimamente relacionada aos erros cometidos pela classificação neural. Se o processo de treinamento do sistema apresentar resultados que indicam uma convergência para a minimização dos

erros de classificação, a eficiência na detecção das anomalias será obtida como conseqüência.

No processo de treinamento foi utilizada a ferramenta SNNS [Stuttgart 2006], gerando um tempo médio de 1 minuto para cada modelo de assinatura, considerando os conjuntos de treinamento descritos na tabela 1. Os gráficos a seguir demonstram os erros quadráticos médios em função dos ciclos de treinamento, para cada modelo de assinatura. A linha cheia refere-se à curva de aprendizado enquanto a linha pontilhada à curva de validação. A curva de validação é extremamente importante, pois demonstra os resultados da classificação para um conjunto de padrões nunca visto (os padrões de validação não ocorrem no conjunto de treinamento e podem ser entendidos como eventos novos as quais o sistema deve classificar baseado em seu conhecimento). A aproximação entre estas curvas representa uma baixa taxa de erros na classificação de padrões nunca antes analisados, em relação ao conhecimento absorvido dos padrões de treinamento. Uma não convergência demonstraria que, mesmo após o treinamento, a classificação de padrões novos ocorreria sob taxas de erros não aceitáveis.

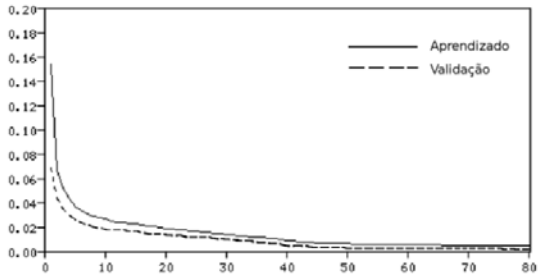

Figura 6: Percentual de erros quadráticos médios em função do número de ciclos de treinamento para o modelo de análise total de fluxos.

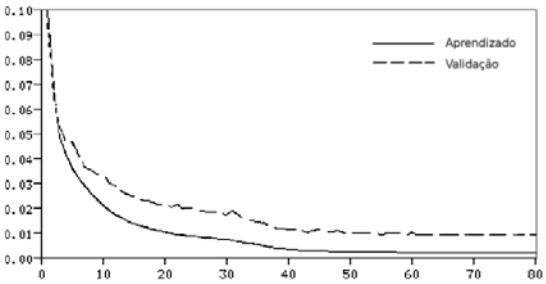

Figura 7: Percentual de erros quadráticos médios em função do número de ciclos de treinamento para o modelo de análise por destino.

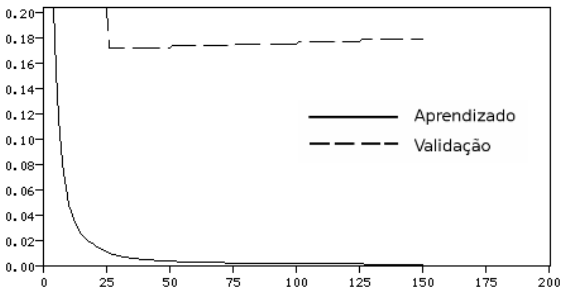

Figura 8: Percentual de erros quadráticos médios em função do número de ciclos de treinamento para o modelo de análise por serviços.

Os erros cometidos em cada um dos modelos de assinatura mostraram resultados promissores. Pelos gráficos anteriores podemos verificar que a taxa de erros média está sempre abaixo dos 0,2%. Com base nos dados obtidos durante o treinamento do sistema, para cada modelo, podemos dispor os resultados como mostra a Tabela 1.

Tabela 1: Dados utilizados e resultados obtidos no processo de treinamento do sistema.

	Análise de fluxos por		
	total	destino	serviços
Conjunto de treinamento	1051	1390	1570
Conjunto de validação	651	469	385
Ciclos de treinamento	100	100	150
Percentual de erros	0,15%	0,13%	0,18%

Uma vez que o processo de treinamento se mostrou promissor, com baixas taxas erros, fica comprovada a eficiência do sistema. É comum em sistemas desta natureza medidas de falsos positivos ou falsos negativos. No entanto, o sistema absorve conhecimento sobre os padrões de treinamento selecionados. Como ataques de negativa de serviço e *worms* modificam o comportamento do ambiente em relação ao comportamento normal dos fluxos, basta que padrões de ataque estejam entre os padrões de treinamento para que o sistema os reconheça. A possibilidade de acontecer um falso positivo se dá unicamente quando um padrão de tráfego lícito se parece com um padrão anômalo. Embora seja difícil na análise de comportamento, esta ocasião pode ocorrer. Uma possibilidade são eventos *FlashCrowd* que muito se assemelham aos ataques de *DoS*. A diferenciação ocorrerá se o treinamento apresentar padrões criteriosos que diferenciem tais eventos. De qualquer forma, uma anomalia será detectada, sendo o erro restrito ao tipo de evento acusado pelo sistema. O caso de um falso negativo é menos provável, uma vez que os padrões anômalos são bastante diferentes dos padrões normais. Assim, as medidas quantitativas de falsos positivos e falsos negativos estão fortemente relacionadas às taxas de erros cometidas pela classificação neural, dispostas nos gráficos e na tabela.

6.3 Uma prova de conceito

Durante os testes do modelo adaptativo de monitoria, o sistema operou em modo *daemon* por um determinado período de tempo. Dentro deste período um *host* do ambiente monitorado foi infectado por um *malware* denominado Kuang [Kuang 2006]. Trata-se de um *Trojan Horse* [Cert.br2 2006] que instala diversos componentes no *host* infectado, dentre eles um *backdoor*. Embora o sistema não conhecesse este artefato, alguma assinatura vista durante o treinamento representava um padrão semelhante ao causado pelo Kuang. O conhecimento absorvido pelo sistema neural possibilitou a detecção das atividades do *malware* de forma imediata.

Todos os modelos de assinaturas puderam detectar a atividade maliciosa no ambiente. As inúmeras tentativas de conexão no *backdoor* geraram distúrbios no comportamento dos fluxos do ambiente, desviando-o do considerado normal.

7. Conclusão

A tarefa de detectar intrusos em ambientes computacionais é desempenhada há tempos sendo uma das grandes áreas da segurança de computadores. Diversos modelos de sistemas de detecção e monitores de redes surgiram com a evolução computacional. O modelo adaptativo de detecção de anomalias e tentativas de intrusão demonstrado neste artigo tem se mostrado bastante eficiente perante o atual modelo de rede, caracterizado pela diversidade de serviços e altas taxas de tráfego.

A utilização de fluxos de dados para monitoria de redes torna o modelo escalável e aplicável em sistemas de grande porte, possibilitando a monitoria de grandes perímetros. O núcleo inteligente garante adaptabilidade ao ambiente e automatização do processo. Uma vez que o sistema esteja treinado, torna-se capaz de monitorar um ambiente de rede e emitir alertas na ocorrência de eventos relevantes.

Quanto ao desempenho, o sistema mostrou-se extremamente eficiente. As detecções dos eventos ocorreram em um intervalo entre 30 segundos e 3 minutos. Embora o sistema ainda não implemente ações reativas, a rapidez na detecção de atividades maliciosas auxiliam na administração de um ambiente. Exemplo disto foi a detecção do Kuang, possibilitando a rápida correção do *host* infectado.

O novo conceito de assinatura busca retratar uma detecção analisando o que o evento causa no ambiente, e não o que ele é (por exemplo, um código malicioso descrito em uma assinatura convencional). A análise deixa de ter caráter comparativo e passa a ter caráter comportamental. Os modelos analisados mostram eficiência dentro de seus limites. A utilização de mais de um modelo simultaneamente é totalmente viável e pode contribuir para aumento da eficiência do sistema.

A utilização de monitores e sistema de segurança em redes é uma área de intensa pesquisa, principalmente pela grande disseminação da Internet e processos críticos nela executados. Assim, a definição e teste deste modelo visam contribuir para o aprimoramento das técnicas de monitoria de redes atualmente existentes. A integração de novas metodologias e sistemas devem sempre contribuir para a obtenção de sistemas cada vez mais eficazes.

Referências

[Abad 2004] Abad, C.; LI, Y.; Lakkaraju, K.; Yin, X.; Yurcik, W. Correlation Between NetFlow System and Network Views for Intrusion Detection, In Workshop on Link Analysis, Counter-terrorism, and Privacy held in conjunction with the SIAM International Conference on Data Mining (ICDM), 2004.

[Allen 2001] Allen, Julia H. The CERT Guide to System and Network Security Practices. The SEI Series in Software Engineering. Addison Wesley Professional, 2001, ISBN-10: 0-201-73723-X; ISBN-13: 978-0-201-73723-3.

[Bonifacio e Cansian 1998] Bonifacio Jr., J. M., Moreira, E. S., Cansian, Adriano Mauro e Carvalho, A. C. P. L. F. An Adaptive Intrusion Detection System Using Neural Networks. In: Proceedings of the 14th Int. Information Security Conference (IFIP/Sec'98, part of the 15th IFIP World Computer Congress), 31 Aug - 4 Sep, 1998, Vienna, Budapest e Austria, Hungary (joint conference), 1998. IFIP, Austrian Computer Society.

[Cansian 2002] Cansian, A. M.; Silva, A. R. A. da; Souza, M. de. An attack signature model to computer security intrusion detection. Milcom 2002: Proceedings, v. 2, p.1368-1373, 07-10 out. 2002.

[Cert.br1 2006] CERT.br. Incidentes Reportados ao Cert.br - Janeiro a Dezembro de 2006. Disponível em: http://www.cert.br/stats/incidentes/2006-jan-dec/tipos-ataque.html.

[Cert.br2 2006] CERT.br - Cartilha de Segurança para Internet, versão 3.1. São Paulo: Comitê Gestor da Internet no Brasil, 2006.

[Claise 2004] Claise, B. RFC 3954: Cisco Systems NetFlow Services Export Version 9. Published by Internet Engineering Task Force (IETF). Internet Society (ISOC) RFC Editor. USA. oct. 2004. Disponível em: http://www.ietf.org/rfc/rfc3954.txt. Acessado em: 11 dez. 2006.

[Fullmer 2006] Fullmer, M. Flow-caputre: Manage storage of flow file archives by expiring old data. Disponível em: http://www.splintered.net/sw/flow-tools/docs/flow-capture.html. Acessado em: 11 dez. 2006.

[Fullmer 2006b] Fullmer, M. Flow-tools Description. Disponível em: http://www.splintered.net/sw/flow-tools/docs/flow-tools.html. Acessado em: 17 ago. 2006.

[Geer 2006] Geer, D. Behavior-based network security goes mainstream. In: IEEE Computer, v.39, n.3, pp. 14-17, 2006.

[Gil 2006] Gil, T. M. NSTX: IP-over-DNS. Disponível em: http://thomer.com/howtos/nstx.html. Acessado em: 11 dez. 2006.

[Goyal 2005] Goyal, V. et al. CompChall: Addressing Password Guessing Attacks. International Conference on Information Technology: Coding and Computing (ITCC), 2005. v1, n.1, p.739-744, april 2005.

[Jung 2002] Jung, J.; Krishnamurthy, B.; Rabinovich, M. Flash Crowds and Denial of Service Attacks. Proceedings of WWW-2002, Hawaii, v.1, p.293-304, May 2002.

[Kuang 2006] Virus Profile: W95/Kuang.gen Trojan. Disponível em: http://us.mcafee.com/virusInfo/default.asp?id=description&virus_k=10213. Acessado em: 11 dez. 2006.

[Longstaff 1987] Longstaff, I. D., Cross, J. F. A pattern recognition approach to understanding the multi-layer perceptron. Pattern Recogn. Lett. 5, 5 (May. 1987), 315-319.

[Manoj 2007] Manoj Parameswaran, Xia Zhao, Andrew B. Whinston, Fang Fang, "Reengineering the Internet for Better Security," Computer, vol. 40, no. 1, pp. 40-44, Jan., 2007.

[Mukherjee 1994] Mukherjee, B., Heberlein, L. and Levitt, K. Network Intrusion Detection, IEEE Network, vol. 8, pp. 26-41, May/June 1994.

[Quittek 2004] Quittek, J. et al. RFC 3917: Requirements for IP Flow Information Export: IPFIX. 2004. Published by Internet Engineering Task Force (IETF). Internet Society (ISOC) RFC Editor, USA, oct. 2004.

[Rosenblatt 1958] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev., 65:386-408, 1958.

[Rosenblatt 1962] Rosenblatt, F. Principles of Neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books, New York, 1962.

[Stuttgart 2006] University of Stuttgart. SNNS - Stuttgart Neural Network Simulator. Disponível em http://www-ra.informatik.uni-tuebingen.de/SNNS/. Acessado em: 23 jul. 2007.

[Wilkinson 1989] Wilkinson, T. S.; Mighell, D. A.; Goodman, J. W. Backpropagation and its application to handwritten signature verification. In Book: Advances in neural information processing systems 1, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1989.