Beholder - Utilizando Redes Neurais MPL na Detecção de Intrusos

Fábio Bombonato
Universidade Católica de Brasília, Departamento de Ciência da Computação
Brasília, Brasil, 71966-700
bombonato@geleira.org

and

Flávia E. S. Coelho
Universidade Católica de Brasília, Departamento de Ciência da Computação
Servidores de Missão-Critica
Brasília-DF, Brasil, 71966-700
fcoelho@ucb.br

Abstract

Beholder is a system based on model that integrate SDIs and neural networks, aproaching a simple manner, but effective – to problems resolutions related to intrusion detection, moreover, obtain the advantages of use this approach, in comparing to tipical systems. The model proposed possibilities a implementation of a system anable to analyse and identify possible intrusions based on methods of anomaly detection, using a MPL neural network for detection of atacks manners – known as network scan at computer networks.
Keywords: Neural networks, IDS (Intrusion System Detection), Security.

Resumo

Beholder é um sistema baseado num modelo que integra os SDIs e as Redes Neurais, propondo uma forma simplificada – porém, efetiva – para a resolução de problemas relacionados à identificação de intrusão, além de obter as vantagens de se utilizar esta abordagem, se comparada aos sistemas usuais. O modelo proposto permitirá a implementação de um sistema capaz de analisar e identificar possíveis intrusões baseadas no método de detecção por anomalia, utilizando uma rede neural MLP para detecção de formas de ataque – conhecida como varredura em rede de computadores.
Palabras claves: Redes neurais, SDI (Sistema de Detecção de Intrusos), Segurança.

1. Introdução

É acompanhado constantemente em noticiários, informativos técnicos, *sites* especializados em segurança, o rápido crescimento dos ataques aos computadores, estes ligados ou não em rede. Este crescimento é perceptível a partir de estudos realizados por diversos órgãos como CERT [1], NBSO [2] e empresas como a Módulo [3].

Aumentar a segurança dos dados não é uma tarefa trivial. Para tal, se utilizam ferramentas com o propósito de melhorar a segurança, como *firewalls*, sistemas de filtragem de conteúdo (*proxy*), antivírus e os próprios Sistemas de Detecção de Intrusos (SDI) ou, *Intrusion Detection System* (IDS).

De acordo com Bace & Mell [4], detecção de intrusão é o processo de monitoramento de eventos que ocorrem em um sistema computacional ou em rede, analisando-os à procura de sinais de intrusão, definido como a tentativa de comprometer confiabilidade, integridade e disponibilidade ou por ultrapassar os mecanismos de segurança de um computador ou rede.

Os SDIs, normalmente, trabalham em três modos: 1) **SDI de *Host***, que se encontra enclausurado nas máquinas; 2) **SDI de Rede** (*Network-Based* IDS, ou NIDS) que analisa o que ocorre na rede e 3) **SDI híbrido** (*Hybrid IDS*) que seria a junção dos dois modos anteriores. No último modo, o monitoramento torna-se possível, através da análise e auditoria dos processos executados em cada máquina e pela captura e análise do tráfego de rede.

SDIs operam de duas formas: 1) via detecção de anomalias, baseados em **comportamento,** e 2) via identificação de assinaturas de ataques, estes baseados em **conhecimento** (também conhecido como **abuso**). A detecção por abuso refere-se à análise de algo conhecido como sendo "mau", sendo a técnica utilizada pela maioria dos sistemas de IDS. Já na detecção por anomalias, a análise é feita pela procura de padrões considerados anormais.

Conforme averiguado por [6], encontram-se algumas imperfeições, falta de eficiência ao tratar pacotes que serão analisados, alto número de falsos positivos (ocorrendo quando a ferramenta classifica uma ação como uma possível intrusão, quando na verdade trata-se de uma ação legítima) e flexibilidade limitada nos modelos vigentes de SDI. Muitos destes aspectos podem ser reduzidos ou até eliminados pelo uso das Redes Neurais Artificiais (RNA), força esta que vem alavancando diversas pesquisas, constituindo também um dos motivos para o desenvolvimento deste trabalho.

Este e outros fatos movimentam diversas pesquisas em Sistemas de Detecção de Intrusos (SDI), conforme verificado em [3] e [4], como por exemplo, o MIT, *Georgia University*, UBILAB *Laboratory*, *Research of* RST *Corporation*, além de iniciativas nacionais como o SADI (Sistema Adaptativo de Detecção de Intrusão) [6] e ACME! (*Advanced Counter-Measures Environment*) [8].

Esse artigo, portanto, irá apresentar um novo modelo de detecção de intrusos com diminuição do índice de falsos positivos, baseando-se em redes neurais.

2. IDS e Redes Neuras

Várias pesquisas na área de detecção de intrusos vêm sendo desenvolvidas, basicamente tentando diminuir a quantidade de falsos positivos e possibilitar um maior dinamismo no tratamento do conhecimento dos ataques. Estas pesquisas basicamente trabalham no desenvolvimento de novas técnicas de detecção de intrusos, sejam por abuso ou anomalia.

Por um lado, na detecção por abuso é possível encontrar algumas abordagens no desenvolvimento de IDS, segundo [7] e [9], dentre elas:

- Sistemas especialistas;
- Verificação de assinaturas;
- Redes de Petri (*Petri nets*);
- Diagramas de transição de estado.

Normalmente, a abordagem escolhida na detecção por abuso é, basicamente, a *verificação de assinaturas*, onde um sistema detecta os ataques que já são conhecidos, efetuando a comparação com uma assinatura invariável que é deixada pelos ataques.

Por outro lado, temos a detecção por anomalia, que segundo [7] e [9], incluem:

- Detecção por limiar (*threshold*);
- Estatísticas;
- Medidas baseadas em regras;
- *Data mining*;
- Algoritmos não-lineares ou classificadores (redes neurais, algoritmos genéticos, classificadores *Bayesianos)*.

Esta abordagem, muitas vezes, demonstra-se mais dinâmica às condições do ambiente. Contudo, a maior limitação para esta abordagem consiste em encontrar um limite correto sem prover alarmes falsos com muita frequência. Onde, tal fato, foi obtido pelo estudo em questão, uma vez que se desenvolveu um modelo que, através dos resultados, demonstrou ser capaz de reduzir, de forma considerável, os falsos positivos.

3. Descrição do Sistema Beholder

Dentre os diversos tipos de ataques por anomalia, foi escolhido trabalhar com os ataques de varredura em rede, por se tratar da fase inicial onde é feito um levantamento de informações relevantes que serão utilizadas no ataque propriamente dito.

O modelo apresentado trata, especificamente, da detecção de intrusos operando sobre redes TCP/IP, ou seja, será um IDS de Rede (NIDS). Procura-se detectar o comportamento de pacotes trafegando na rede, que sejam considerados anômalos. Para tal finalidade, será utilizado um módulo de redes neurais, para que este separe o tráfego considerado normal do suspeito, identificando o nível de periculosidade atribuído para o pacote analisado.

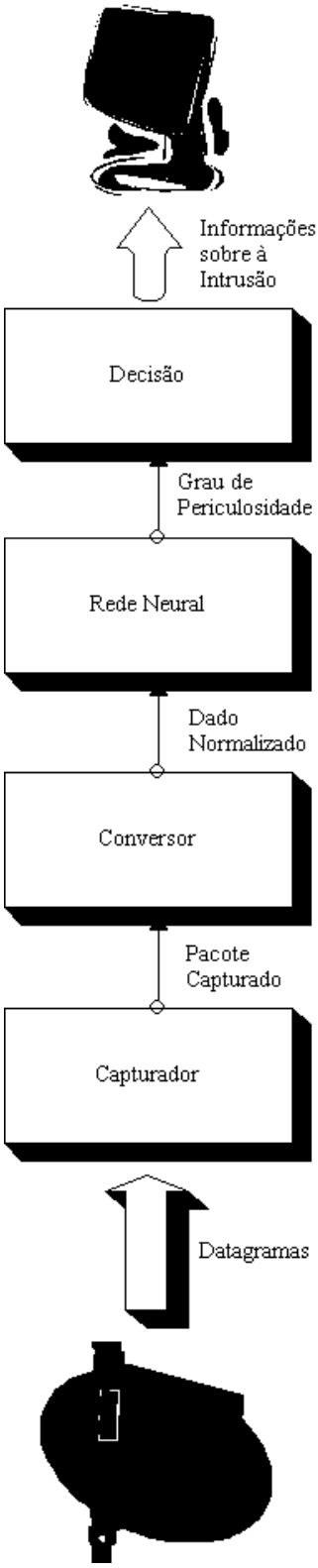

Figura 1 - Estutura geral do *Beholder*

A proposta inovadora para este sistema é a capacidade de efetuar a detecção de comportamentos anômalos utilizando redes neurais artificiais, cujo modelo aqui utilizado é, de certa forma, mais natural. Esta naturalidade é obtida, pois os dados são capturados da forma que estão sendo trafegados na rede, sendo então submetidos para a análise da RNA, evitando assim, um trabalho extra de pré-classificação e separação de fluxos.
Basicamente, o modelo geral de detecção de intrusão proposto (Figura 1), é composto de um capturador que obtém os dados da rede, um conversor que efetua a normalização destes dados recebidos da rede para então serem submetidos à rede neural artificial, ao qual, retorna um valor, indicando o grau de periculosidade referente a cada pacote analisado. Estes dados, por sua vez, são repassados ao decisor que, então, formata e apresenta ao operador conforme as duas categorias possíveis: ataque ou situação normal.

4. Fase de Aprendizagem

Para a concepção da Rede Neural utilizada, foram efetuadas algumas simulações no EasyNN e SNNS, permitindo moldar um modelo de RNA adequado para ser utilizado pela *engine Beholder*.
Os dados utilizados na simulação foram obtidos da captura de pacotes (datagramas) obtidos no barramento Ethernet, obtendo padrões normais de uso em rede e padrões considerados como ataque (obtidos de forma isolada), conforme exemplo na Tabela 1.

Protocolo	6
IP Origem	192.168.0.254
IP Destino	192.168.0.3
Porta Origem	279
Porta Destino	43309
Flags	0x14
Tamanho Header	20
Tamanho Dados	6
Tipo ICMP	0

Tabela 1 – Dados obtidos pelo Capturador

Após a coleta e seleção dos dados que serão utilizados, estes são repassados para o módulo conversor (sub-módulo simulador) que efetuará a normalização destes dados de entrada sendo convertidos para seus valores correspondentes no formato *long,* utilizado na linguagem C, permitindo assim que todos sejam formatados considerando um padrão. Feita a normalização, para o exemplo anterior, os dados seriam então apresentados no formato da Tabela 2.

Protocolo	6.0
IP O1	192.0
IP O2	168.0
IP O3	0.0
IP O4	254.0
IP D1	192.0
IP D2	168.0
IP D3	0.0
IP D4	3.0
Porta Origem	279.0
Porta Destino	43309.0
Flags	20.0
Tamanho Header	20.0
Tamanho Dados	6.0
Tipo ICMP	0.0

Tabela 2 – Dados normalizados

Com os dados normalizados foi possível efetuar a simulação que utilizou as seguintes quantidades de padrões, conforme Tabela 3.

Padrão	Quantidade
Treinamento	5577
Validação	200
Teste	18

Tabela 3 – Quantidade de padrões utilizados

O treinamento utilizado no EasyNN foi o *backpropagation* com *momentum*. Para tal, foram utilizados os seguintes parâmetros para o treinamento da rede neural utilizada:

- •Taxa de Aprendizagem: **0.30**
- •Momento (*Momentum*): **0.80**
- •Parar quando todos os erros estejam abaixo de **0.0500**

5. Rede Neural

Dadas as configurações utilizadas na fase de aprendizagem, foi possível definir a topologia da RNA para a MLP (*Multilayer Perceptron*), tendo os neurônios de entrada e saída sendo fixos, 15 padrões de entrada (conforme apresentado anteriormente) e 1 de saída (correspondendo a um valor que tende à situação de ataque ou normal). Já a camada intermediária (oculta) foi fixado para uma única camada contendo 4 neurônios.
Para esta topologia e treinamento efetuado, foram executados 44.059 ciclos de treinamento, onde a taxa de validação alcançada chegou a **99.58%** dos 200 exemplos de validação. Para os erros, foram obtidos os resultados apresentados na Tabela 4.

Tipo de Erro	Erro Obtido
Erro mínimo	0.00000
Erro médio	0.001768
Erro máximo	0.670616

Tabela 4 – Taxas de Erros

Vale ressaltar que, apesar da taxa de erro máximo estar alta, o erro médio está muito próximo do erro mínimo, o que demonstra que a rede consegue reconhecer uma grande quantidade de situações normais e de ataque. O erro máximo encontra-se elevado devido a alguns poucos padrões que não identificou de forma correta devido à similaridade com os padrões normais, conforme será visto nos resultados.

6. Resultados Obtidos

Os resultados obtidos para o ambiente proposto, podem ser resumidos abaixo:

- •Acesso Normal: Foram reconhecidos **100%** do trafego como normal. Não gerou nenhum falso positivo;
- •Ataques por Varredura: TCP SYN, *Stealth* FIN, *Xmas Tree*, *Null*, RPC: **99.5%** detectado. TCP *connect()* não foi detectado.

Os resultados apresentados foram satisfatórios, demonstrando que a utilização das redes neurais possibilita lidar com as tentativas de ataques de forma efetiva e adaptativa, visto que estas possuem uma característica interessante, conhecida como generalização. Além disso, foi constatada a grande diminuição dos falsos positivos, se comparado aos modelos usuais.
Também é importante salientar que não foi possível detectar a técnica de varredura conhecida como TCP *connect()*, tal fato já era esperado à medida que o projeto vinha sendo desenvolvido e pelo trabalho sobre os padrões de treinamento. Isto ocorre devido a grande semelhança entre os padrões relacionados ao ataque TCP *connect()* e uma situação normal, sendo praticamente idênticos.

7. Conclusões

As redes neurais artificiais demonstraram ser capazes de efetuar a detecção de intrusão baseada em anomalia, mas precisamente, para as técnicas de varredura em rede de computadores utilizando a rede neural MLP. A utilização das RNA no contexto de SDIs é considerada uma área relativamente nova, sendo necessários mais experimentos e melhoramentos para que esta abordagem se torne mais efetiva e possa ser utilizado em grande escala, ressaltando que este trabalho mostrou que esta é uma possibilidade real.

Referências

[1] Bace, R., Mell, P. NIST Special Publication on Intrusion Detection Systems. (2001)

[2] Martino, S. A Mobile Agent Approach to Intusion Detection. *Joint Research Center Institute for Systems*, Informatics and Safety, Italy. (1999).

[3] Cannady, J. Artificial Neural Networks for Misuse Detection, *School of Computer and Information Sciences*, Nova Southeastern University, (1998)

[4] Philippe Jean, Application of Neural Networks to Intrusion Detection, *Information Security Reading Room*, SANS Institute, (2001).

[5] Tavares, D. M., Castejon, E. F., Rossi, G. B. ACME! (Advanced Counter-Measures Environment), *Monografia de Projeto Final apresentada ao Departamento de Ciências da Computação*, Instituto de Biociências Letras e Ciências Exatas – UNESP, São José do Rio Preto, (1999).

[6] Oliveira, C. B. Reconhecimento de Padrões com Auxílio à Detecção de Intrusão em Redes de Computadores, *Tese de Mestrado*, Universidade Católica de Brasília, Brasília, 2001.