REDES NEURAIS ARTIFICIAIS APLICADAS À DETECÇÃO DE INTRUSOS EM REDES TCP/IP

Renato Maia Silva
Centro de Estudos em Telecomunicações - CETUC
Pontifícia Universidade Católica do Rio de Janeiro
22453-900 Rio de Janeiro – RJ
renatomaia@uol.com.br

Marco Antonio Grivet M. Maia
Centro de Estudos em Telecomunicações – CETUC
Pontifícia Universidade Católica do Rio de Janeiro
22453-900 Rio de Janeiro – RJ
mgrivet@cetuc.puc-rio.br

RESUMO

Ataques e intrusões são uma ameaça constante para empresas interconectadas através de redes de pacotes e da Internet. Ferramentas tradicionais de detecção de ataques e intrusões dependem de conhecimento prévio sobre as técnicas de ataque não sendo capazes de detectar novas técnicas de ataques. Este artigo investiga a aplicação de redes neurais artificiais no auxílio a detecção de intrusão em redes de pacotes TCP/IP. Utilizando a capacidade de generalização das redes neurais, espera-se que o sistema detecte novos ataques mantendo uma alta taxa de acertos.

ABSTRACT

Computer attacks and intrusions poses significant threats to companies and organizations interconnected through packet networks and the Internet. Most current approaches to intrusion detection rely on previous knowledge of attack patterns and are not capable of detecting new intrusion techniques. This paper presents the application of artificial neural networks as a component of an intrusion detection system. Exploring neural networks generalization capabilities the system should be able to detect new attack patterns and sustain a high detection rate.

1 INTRODUÇÃO

Com o crescimento vertiginoso da Internet e sua ampla utilização por todos os setores da sociedade, a segurança de sistemas interconectados se tornou requisito obrigatório. Pesquisa realizada em 2003 pelo CSI – *Computer Security Institute* - em conjunto com FBI [1] mostrou que 56% das empresas participantes declaram ter sido vítimas de ataques nos últimos 12 meses. Esta situação é ainda mais crítica para empresas cuja operação e receita são dependentes completamente da utilização de serviços Internet. A necessidade de se proteger contra estes ataques despertou o interesse por ferramentas automatizadas para detectá-los.

Estas ferramentas, normalmente denominadas de Sistemas de Detecção de Intrusão, podem ser classificadas, quanto a sua área de atuação em :

- Sistema de Detecção de Intrusão para Host ("HIDS - *Host Intrusion Detection System*") são sistemas que monitoram um determinado computador ("*host*"). Normalmente estes sistemas monitoram variáveis como a taxa de utilização de processador, chamadas de funções de sistemas e modificações em arquivos críticos, procurando por atividades suspeitas.
- Sistema de Detecção de Intrusão para Redes ("NIDS - *Network Intrusion Detection System*") são sistemas que monitoram segmentos de redes de computadores. Normalmente estes sistemas capturam todos os pacotes de dados que trafegam em determinado segmento da rede procurando por atividades suspeitas.

Também podem ser classificadas em relação a estratégia adotada para detectar atividades suspeitas :

- Detecção de anomalias ("*Anomaly Detection*") consiste em técnicas que procuram determinar variações nas atividades em relação a um padrão de utilização do sistema/rede considerado normal.
- Detecção de abuso ("*Misuse Detection*") consiste em encontrar padrões de ataques conhecidos nos dados monitorados.

A maioria das soluções de detecção de intrusos empregada depende de conhecimento prévio dos padrões de ataque [2] [3]. Esta abordagem não é suficiente devido ao grande número de novas técnicas de ataque e intrusão que surgem a cada dia, aliados a rapidez com que elas passam a ser utilizadas.

O objetivo deste trabalho é propor um sistema de detecção de intrusão (NIDS) baseado em redes neurais artificiais. 4 configurações de redes neurais foram treinadas para detectar anomalias e posteriormente testadas com base que inclua padrões de ataques novos. Os resultados obtidos são apresentados.

2 DADOS PARA TREINAMENTO

Os dados utilizados para treinamento e testes das redes neurais correspondem a subconjuntos da base de dados disponibilizada na competição internacional de mineração de dados KDD Cup [4] – *Knowledge Discovery and Data Mining Competition* - em 1999 e gerados em projeto de análise de sistemas de detecção de intrusão realizado no MIT [5]. Esta base foi gerada através da captura, durante 9 semanas, de todos os pacotes TCP/IP em uma rede real.Um pré-processamento inicial destes pacotes foi efetuado, agrupando-os em conexões. Uma conexão consiste em troca de um ou mais pacotes entre dois sistemas, para um determinado serviço, durante um intervalo definido. A base de dados resultante possui aproximadamente 5.000.000 de registros de conexões. Cada conexão é caracterizada por 41 variáveis, classificadas em 3 categorias principais :

- Características básicas TCP/IP – são as informações essenciais e elementares obtidas ao se analisar um pacote da pilha de protocolos TCP/IP. Estão detalhadas na Tabela 1.
- Características sugeridas através de conhecimento da área ("domain knowledge") – são informações extraídas dos pacotes com auxílio de especialistas. Exemplo seria o registro de tentativas de login com usuário privilegiado em determinada conexão ou o número de acesso a arquivos realizados. Estão detalhadas na Tabela 2.
- Características obtidas através de uma janela de 2 segundos – são características temporais obtidas usando uma janela de 2 segundos. Informações importantes referentes a determinados ataques somente podem ser obtidas levando-se o tempo em consideração. Estão detalhadas na Tabela 3.

Tabela 1 – Características básicas de conexões TCP/IP.

Nome	**Descrição**	**Tipo**
Duration	Duração em segundos da conexão	Contínua
Protocol_type	Tipo de protocolo usado na conexão, i.e. tcp, udp, etc.	Discreta
Service	Serviço de rede sendo utilizado, i.e. http, telnet, ftp etc.	Discreta
Src_bytes	Número de bytes enviados da fonte para o destino	Contínua
Dst_bytes	Número de bytes enviados do destino para a fonte	Contínua
Flag	Status da conexão (normal ou erro)	Discreta
Land	1 se conexão é de/para o mesmo host; 0 caso contrário.	Discreta
Wrong_fragment	Número de fragmentos errados.	Contínua
Urgent	Número de pacotes urgentes.	Contínua

Tabela 2 – Características de uma conexão sugeridas através de conhecimento da área.

Nome	Descrição	Tipo
Hot	Número de indicadores chaves (“*hot*”).	Contínua
Num_failed_logins	Tentativas de login sem sucesso	Contínua
Logged_in	1 se login efetuado com sucesso; 0 caso contrário.	Discreta
Num_compromised	Número de condições de “comprometimento”	Contínua
Root_shell	1 se shell root foi obtido; 0 caso contrário	Discreta
Su_attempted	1 se comando “su root” foi tentado; 0 caso contrário	Discreta
Num_root	Número de acessos como root.	Contínua
Num_file_creations	Número de operações de criação de arquivos.	Contínua
Num_shells	Número de “shells prompts” obtidos.	Contínua
Num_access_files	Número de operações em arquivos de controle de acesso.	Contínua
Num_outbound_cmds	Número de comandos externos em uma sessão ftp.	Contínua
Is_hot_login	1 se o login pertence a lista “hot”; 0 caso contrário.	Discreta
Is_guest_login	1 se o login usou a conta guest; 0 caso contrário.	Discreta

Tabela 3 - Características determinadas usando uma janela de 2 segundos.

Nome	Descrição	Tipo
Count	Número de conexões iguais a esta para este mesmo “*host*” nos últimos 2 segundos.	Contínua
Serror_rate	% de conexões que possuem erro SYN.	Contínua
Rerror_rate	% de conexões que possuem erro REJ.	Contínua
Same_srv_rate	% de conexões para um mesmo serviço.	Contínua
Diff_srv_rate	% de conexões para serviços diferentes.	Contínua
Srv_count	Número de conexões para o mesmo serviço que o usado nesta conexões nos últimos 2 segundos.	Contínua
Srv_serror_rate	% de conexões que possuem erro SYN para este serviço.	Contínua
Srv_rerror_rate	% de conexões que possuem erro REJ para este serviço.	Contínua
Srv_diff_host_rate	% de conexões deste mesmo serviço para hosts diferentes.	Contínua

Diversos tipos de ataques a redes TCP/IP [6] – 24 para treinamento e 14 para testes – foram inseridos e identificados na base de dados, juntamente com milhares de registros normais. Os ataques gerados podem ser agrupados em 4 categorias principais destacadas na Tabela 4.

Tabela 4 – Classes de ataques com alguns exemplos.

Classe de Ataques	**Exemplos**
Negação de Serviço – DOS ou *Denial of Service*	*Smurf* *Neptune* *Ping of Death* *Mailbomb* *UDP Storm*
Reconhecimento e Sondagem – *Probing*	*Ipsweep* *Nmap* *Satan* *MScan*
Remoto-Usuário	*Xlock* *ftp-write*
Remoto-Super-usuário	*Eject* *Ffbconfig*

As tabelas 5 e 6 representam os subconjuntos da base de dados KDDCup 1999 selecionados, respectivamente, para treinamento e testes das redes neurais. Os subconjuntos selecionados para testes possuem registros de conexões contendo tipos de ataques que não estavam presentes nos subconjuntos para treinamento.

Tabela 5 – Subconjuntos de treinamento

Subconjunto	**Tamanho (N° conexões)**	**Tipos de conexões**	**% Conexões normais**	**% Ataques**
TR1	1470	Normal, Smurf	76%	24%
TR2	18503	Normal, Smurf, Neptune	57%	43%
TR3	27005	Normal, Smurf, Neptune, Back	22%	78%

Tabela 6 – Subconjuntos de testes

Subconjunto	**Tamanho**	**Tipos de conexões**	**% Conexões normais**	**% Ataques**
TE1	836	Normal, Smurf, Neptune, Snmpgetattack	38%	62%
TE2	1025	Normal, Smurf, Neptune Satan	41%	59%
TE3	1226	Normal, Smurf, Neptune, Back Snmpgetattack Ipsweep Satan	28%	72%

As variáveis da base de dados do KDD Cup 1999 não estavam normalizadas e portanto, não podiam ser apresentadas como entrada de uma rede neural. Foi realizado um pré-processamento destes dados convertendo todas as variáveis para valores numéricos, normalizados dentro do intervalo [-1,1].

3 IMPLEMENTAÇÃO

A "*toolbox*" de redes neurais do software Matlab R11 foi utilizada para testar o desempenho de algumas configurações de Redes Neurais artificais aplicadas a detecção de intrusão, utilizando-se os subconjuntos para treinamento e testes já apresentados. Após alguns testes preliminares definiu-se pela utilização de uma rede neural MLP [7] – *MultiLayer Perceptron* – composta por duas camadas escondidas com 10 elementos processadores cada, função de ativação tangente hiperbólica (faixa de valores no intervalo [-1,1]). Em apenas uma das redes utilizamos as 41 entradas, correspondendo a todas as 41 características dos dados das conexões. Nas demais foram utilizadas apenas 9 entradas referentes às categorias básicas do TCP/IP. Uma das redes foi projetada para classificar os registros de conexões em normal, "*smurf*", "*neptune*" utilizando apenas 1 saída. Uma segunda rede foi projetada para, usando 4 saídas, determinar se o registro da entrada corresponde a uma conexão normal ou aos ataques "*neptune*", "*smurf*" e "*back*". As demais tinham apenas uma saída identificando o registro de conexão entre normal e ataque. As configurações de redes neurais utilizadas estão detalhadas na tabela 7.

Tabela 7 – Configurações das redes utilizadas

Rede	Entradas	Camada Escondida 1	Camada Escondida 2	Saídas	Valores Saída
Rede1	9	10	10	1	(-1) – Normal (1) – Ataque
Rede2	41	10	10	1	(-1) – Normal (1) – Ataque
Rede3	9	10	10	1	(-1) – Normal (0) – Neptune (1) – Smurf
Rede4	9	10	10	4	(-1 1 1 1) – Normal (1 -1 1 1) – Neptune (1 1 -1 1) – Back (1 1 1 -1) – Smurf

4 RESULTADOS

As configurações de redes definidas foram treinadas e testadas utilizando os subconjuntos dos dados do KDD Cup 1999 definidos anteriormente. O treinamento utilizou a função de treinamento padrão e os resultados obtidos estão apresentados na tabela 8.

Tabela 8 - Resultados obtidos

Rede	**Dados Treinamento**	**Dados Testes**	**Número de épocas**	**Erro final (MSE)**	**Taxa Acerto**	**Falso positivo**
Rede1	TR1	TE1	200	0,1260	**0,940**	0,0598
Rede1	TR1	TE1	2	0,0009	**0,960**	0,0390
Rede2	TR2	TE2	2	0,0061	**0,950**	0,0144
Rede3	TR2	TE1	18	0,0054	**0,975**	0,0000
Rede3	TR2	TE2	18	0,0054	**0,960**	0,0020
Rede4	TR3	TE3	8	0,0085	**0,920**	0,0000

5 CONCLUSÃO

Apresentamos resultados obtidos com a aplicação de redes neurais MLP ao problema de detecção de intrusos em redes TCP/IP. Taxas de acerto acima de 90% foram obtidas para todas as configurações testadas. Em nenhuma das configurações testadas a rede gerou um falso-negativo, ou seja, identificou um ataque novo presente nos dados de teste como conexão normal. Para as configurações com saídas

que identificavam o ataque específico um ataque novo foi sempre categorizado como um dos ataques conhecidos pela rede neural. Para os cenários avaliados foi possível comprovar que a capacidade de generalização das redes neurais permite a identificação de padrões de ataques desconhecidos. Como parte integrante deste projeto, cenários adicionais envolvendo todos os 38 ataques presentes na base de dados do KDD Cup 1999 serão analisados.

Finalmente, a utilização de um comitê de redes neurais, com várias redes, cada uma especializada em detectar uma classe de ataques ou ,até mesmo especializada em categorias específicas das variáveis de entrada, pode aumentar a precisão da resposta do sistema, simultaneamente diminuindo a quantidade de falsos-positivo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] RICHARDSON, Robert; 2003 CSI/FBI Computer Crime and Security Survey. Disponível em: <http://www.gocsi.com>.

[2] CANNADY, James; HARREL, Jay. A Comparative Analysis of Current Intrusion Detection Technologies. In : Proceedings of the Fourth Technology for Information Security Conference' 96 (TISC'96). Houston, 1996.

[3] DEBAR, Hervé et al. A Revised Taxonomy for Intrusion Detection Systems. IBM Research Report. Zurique, 1999.

[4] KDD Cup 1999 Dataset, UCI KDD repository, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[5] LIPPMAN, Richard et al. The 1999 DARPA Off-line Intrusion Detection Evaluation. In: Computer Networks, In Press, 2000.

[6] KENDALL, Kris. A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems. Boston : MIT, 1998. Tese de Mestrado.

[7] HAYKIN, S., Neural Networks: A Comprehensive Foundation, 2nd Edition. Prentice-Hall, 1999.