UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS NO RECONHECIMENTO DE PADRÕES EM PACOTES DE DADOS TCP/IP

USING OF ARTIFICIAL NEURAL NETWORKS IN PATTERN RECOGNITION IN TCP/IP DATA PACKETS

PERES, Cassiano Ricardo de Oliveira[1], **FILHO,** Pedro Luiz de Paula[2], **MENEZES,** Paulo Lopes de[2]

[1] **cassianoperes@hushmail.com**

Resumo

O objetivo deste trabalho é apresentar resultados da utilização de inteligência artificial baseada em redes neurais artificiais (RNA), no reconhecimento de padrões em pacotes de dados com base no seu protocolo de redes de computadores. Também é apresentado uma comparação de desempenho no reconhecimento de padrões baseado no tipo de amostra utilizada para teste das redes neurais artificiais.

Palavras-chave: backpropagation, camadas, neurônio, teste, treinamento

Abstract

The objective of this paper is to present results of the use of artificial intelligence based on artificial neural networks (ANN), in recognizing patterns in data packets based on their protocol for cmputer networks. A comparison of recognition performance based on the type of sample used for testing artificial neural patterns is also presenteds.

Key-words: backpropagation, layers, neuron, testing, training

INTRODUÇÃO

Com o crescente uso de dispositivos com acesso à Internet, ocorreu um significativo aumento da quantidade de informações disponíveis na rede mundial de computadores. E com esse aumento, surgiram muitos desafios para a segurança da informação, tais como manter a privacidade dos dados, evitar roubos de informações, ataques de vírus de computadores, entre outros.

Segundo Silva (2005), apesar do aprimoramento diário das técnicas para a prevenção de ameaças, essas técnicas ainda possuem limitações que as impedem de estarem atualizadas contra novas ameaças, pois necessitam de conhecimento prévio das formas de ataque. Dessa maneira, cresce a exigência por alternativas na prevenção e tratamento de ameaças, buscando prevê-las e antecipar-se a elas. A utilização de redes neurais artificiais surge como uma alternativa no desenvolvimento de aplicações com esses objetivos.

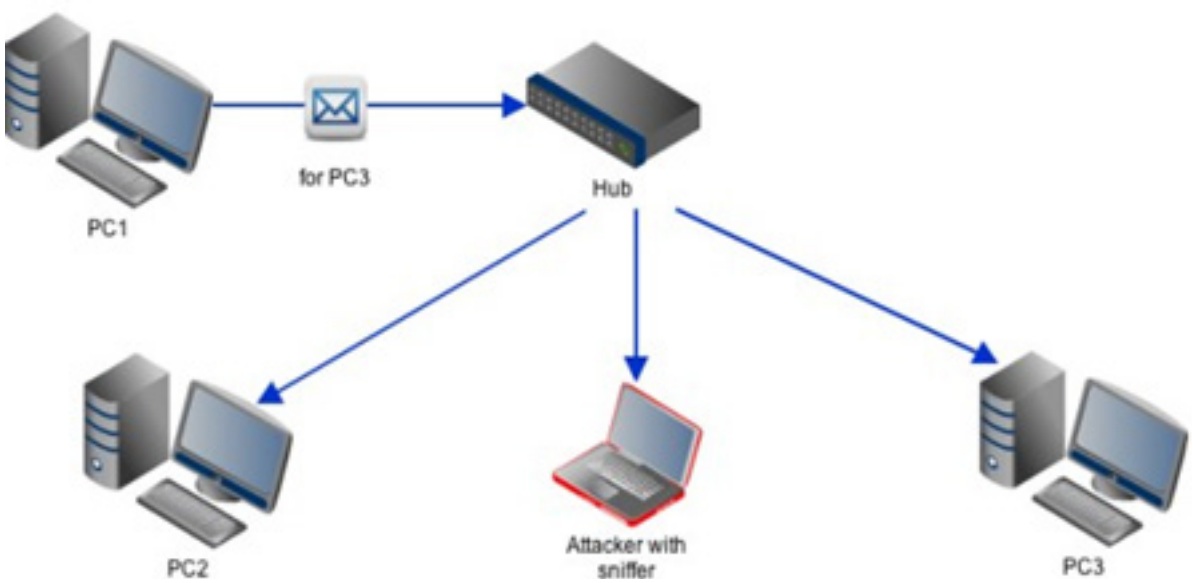
Há diversas propostas para a utilização de RNA em redes de computadores. Silva (2005) propôs a aplicação de RNA na detecção de intrusão em redes de computadores TCP/IP. Ahmad et al. (2009) aplicaram RNA na detecção de ataques de negação de serviço, Denial of Service (DOS), e Silva (2007) utilizou RNA na detecção de ataques no tráfego de dados em redes de computadores.

Este trabalho propõe a utilização de RNA no reconhecimento de padrões em pacotes de dados que trafegam em uma rede de computadores, afim de identificar os protocolos presentes nos mesmos. Estas técnicas podem ser utilizadas no desenvolvimento de aplicações voltadas à segurança da informação, tais como antivírus, aplicações para controle de acesso a certos tipos de conteúdo, detecção de comportamentos suspeitos de usuários de uma rede de computadores, previsão do consumo de banda de rede, entre outras aplicações.

METODOLOGIAS UTILIZADAS

Para este trabalho, foram obtidas amostras de pacotes de dados utilizados nos processos de treinamento, validação (obtenção da taxa de erro) e teste das RNA, a partir de duas máquinas virtuais enviando e recebendo pacotes de dados entre si.

Figura 1 – Representação esquemática de um sniffer de pacotes.



Para a captura dos pacotes foi desenvolvida uma aplicação do tipo sniffer. Segundo Qader et al. (2010), sniffer é um programa executado em um computador conectado a uma rede, que passivamente recebe cópias dos pacotes que passam em seu adaptador de rede, salvando-os em arquivos para análises posteriores.

A Figura 1 contém a representação de uma rede de computadores com uma das máquinas conectadas à rede possuindo um sniffer.

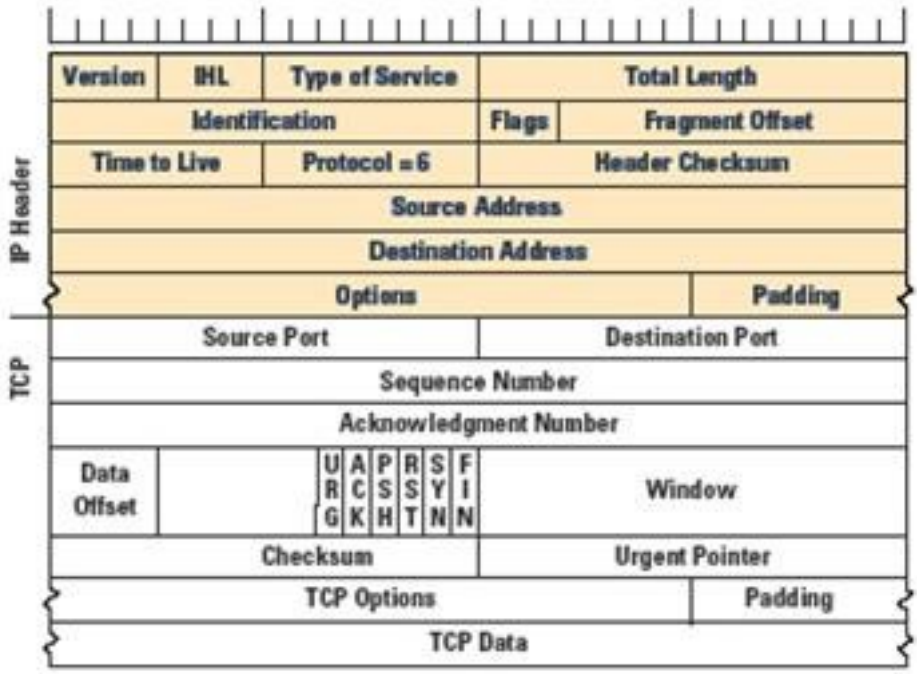
Para Torres (2009), pacotes de dados são formados por conjunto de bytes que carregam informações que o auxiliam a chegar a seu destino, tais como, entre outras características, os endereços IP de origem e destino, as portas de entrada e saída, tamanho do pacote e seu conteúdo A Figura 2 contém a estrutura de um pacote de dados TCP/IP com a descrição dos seus campos.

Os protocolos de rede analisados neste trabalho foram o Hiptertext Transfer Protocol (HTTP), File Transfer Protocol (FTP) e o Secure Shell (SSH), pertencentes à camada de aplicação do modelo TCP/IP.

Para Torres (2009), o HTTP é um protocolo utilizado amplamente para a transferência de arquivos de hipertexto entre máquinas remotas, onde as páginas de um site (arquivos de hipermídia) ficam armazenadas em servidores e são acessadas por máquinas clientes através de navegadores de Internet, denominado browsers.

Segundo Postel e Reynolds (1985), o protocolo de transferência de arquivos FTP, como o próprio nome sugere, é utilizado para a transferência de arquivos entre máquinas remotas.

Figura 2 – Estrutura de um pacote de dados TCP/IP



Fonte: Huston, 2004

O FTP possui três principais objetivos:

- Promover o compartilhamento de arquivos;
- Incentivar a utilização de máquinas remotas;
- Proteger os usuários das variações em sistemas de armazenamento de arquivos entre computadores, e;
- Fornecer uma transferência de dados confiável e eficiente.

Por fim, Lonvick e Ylonen (2006), definem o protocolo SSH como um protocolo de transporte seguro de baixo nível. Fornece criptografia forte, autenticação de host (computadores) em vez da autenticação de usuários e proteção da integridade na transmissão de pacotes.

O SSH foi projetado para ser simples e flexível para permitir que parâmetros sejam negociados, tais como métodos de trocas de chaves públicas e privadas, algoritmos de criptografia simétrica, autenticação de mensagens e algoritmos de hash.

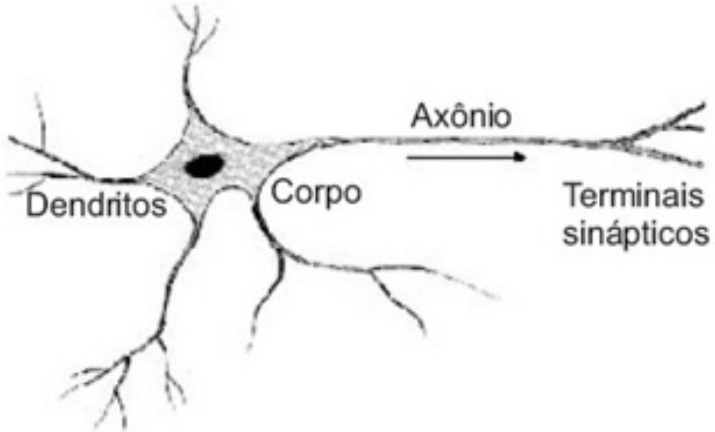
Após a realização da captura de uma determinada quantidade de pacotes, estes foram convertidos para a representação binária (0 e 1) e gravados em arquivos para a utilização nos processos de treinamento, validação e teste das RNA.

Para compor as amostras do primeiro cenário de testes, foi extraída a parte inicial dos pacotes, que contém o cabeçalho com as informações específicas do seu protocolo de rede.

Já no segundo cenário, foram obtidos bits de partes aleatórias do pacote, sempre do mesmo tamanho, mas podendo ter conteúdo variado, como partes do cabeçalho, conteúdo (informações) ou rodapé.

Redes neurais artificiais (RNA) são modelos matemáticos inspirados no cérebro humano, nas quais a unidade básica de processamento são os neurônios artificiais, baseados no neurônio humano. A Figura 3 contém o esquema básico da estrutura de um neurônio humano.

Figura 3 – Modelo simplificado de um neurônio humano



Fonte: Ferneda, 2006

Segundo Ferneda (2006), o neurônio humano é uma célula composta de três partes principais, sendo o dendrito, responsável por receber os estímulos externos, o corpo do neurônio responsável por processar esses estímulos e o axônio por transmiti-lo para outras células por meio de sinapses.

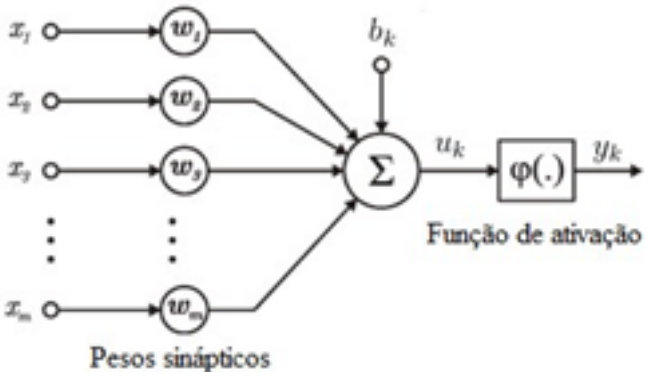
Esse processo pode ser repetido por várias camadas de neurônios e assim, após o estímulo ser processado, pode gerar reações físicas comandadas pelo cérebro.

O primeiro modelo de RNA foi proposto por McCulloch & Pitts (1943). Segundo Haykin (1999),de forma semelhante ao cérebro humano, as RNA possuem a característica de armazenar conhecimento por meio de experiências e possuem conexões representadas por pesos sinápticos, cuja função é armazenar e processar o conhecimento adquirido.

No treinamento das RNA, foi utilizado o método de aprendizagem backpropagation, sendo este o mais popular para a aprendizagem supervisionada. Durante o período de treinamento, um vetor com os valores de entrada é apresentado à RNA em conjunto com seu valor de saída desejada.

Os pesos sinápticos da rede neural são inicializados com valores aleatórios e o aprendizado é realizado reajustando os pesos de forma iterativa. Os valores resultantes do processamento da entrada são comparados aos valores de saída desejados, resultando em um sinal de erro que é retropropagado através da rede, para permitir o reajuste dos pesos sinápticos. Esse processo é repetido até que sejam obtidos valores de saída satisfatoriamente próximos aos desejados (TISSOT, et al., 2012).

Figura 4 – Modelo matemático de um neurônio artificial



Fonte: Adaptado de Moreto, Rolim, 2010

Na Figura 4 é apresentado o modelo matemático de um neurônio artificial, e um neurônio pode ser matematicamente descrito através do par de equações (Equação 1.1 e Equação 2.2):

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (1.1)$$

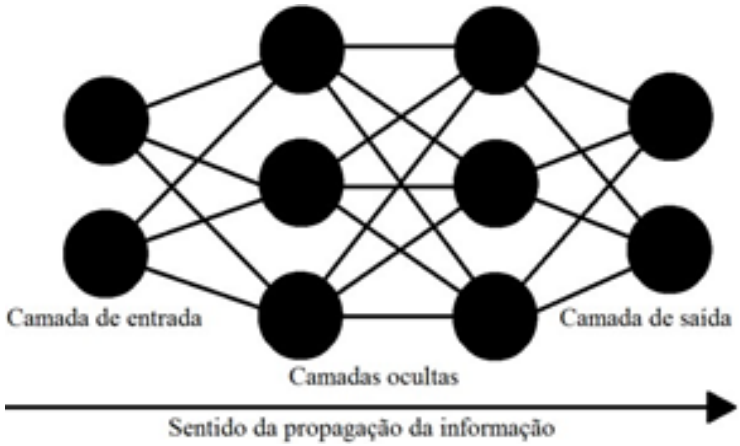
$$y_k = \varphi(u_k + b_k) \qquad (1.2)$$

em que $x_1, x_2, x_3, \ldots, x_m$ são os valores de entrada, são os pesos sinápticos do neurônio, é a saída do combinador linear dos valores de entrada, é o valor do bias, é a função de ativação, e por fim é o valor de saída do neurônio.

Segundo Haykin (1999) o uso do bias tem a função de aumentar o grau de liberdade da rede

a função de aumentar o grau de liberdade da rede neural, permitindo assim, que ela se adapte da melhor forma possível ao conhecimento que lhe é apresentado.

Figura 5 – Esquema de uma RNA multicamadas



Fonte: Adaptado de Haykin, 1999.

As RNA utilizadas neste trabalho foram do tipo MLP, Multilayer Perceptron, ou Perceptron de Múltiplas camadas, que possuem uma ou mais camadas ocultas entre as de entrada e saída.

Quanto à forma de conexão, esta foi do tipo feedforward na qual os valores apresentados à camada de entrada são propagados apenas no sentido das camadas seguintes, como apresentado na Figura 5. Também são conhecidas como RNA totalmente conectadas, pois todos os neurônios da camada anterior estão conectados com todos os da camada seguinte (HAYKIN, 1999).

RESULTADOS E DISCUSSÃO

Neste trabalho foram implementadas para cada um dos dois cenários, cinco RNA com topologias diferentes, nas quais a camada de entrada possui a quantidade de neurônios fixada em 350, na única camada oculta de cada uma das redes neurais foi de 100, 200, 300, 400 e 500 neurônios respectivamente, e na camada de saída, a quantidade foi fixada em 3 neurônios.

O desempenho de cada uma das RNA foi avaliado com base no erro SSE (Sum of Squared Errors), o qual determina a performance da RNA de acordo com a soma dos quadrados dos erros gerados pelo neurônio artificial no processamento dos valores de entrada (MATHWORKS, 2014). Foram utilizados 2250 pacotes, subdivididos em: 1470 pacotes para treinamento, 480 para validação e 300 para teste

CENÁRIO 1

No primeiro cenário as RNA foram treinadas, validadas e testadas utilizando bits da parte inicial do pacote, ou seja, parte do seu cabeçalho, com informações estruturais sobre o seu protocolo de redes de computadores.

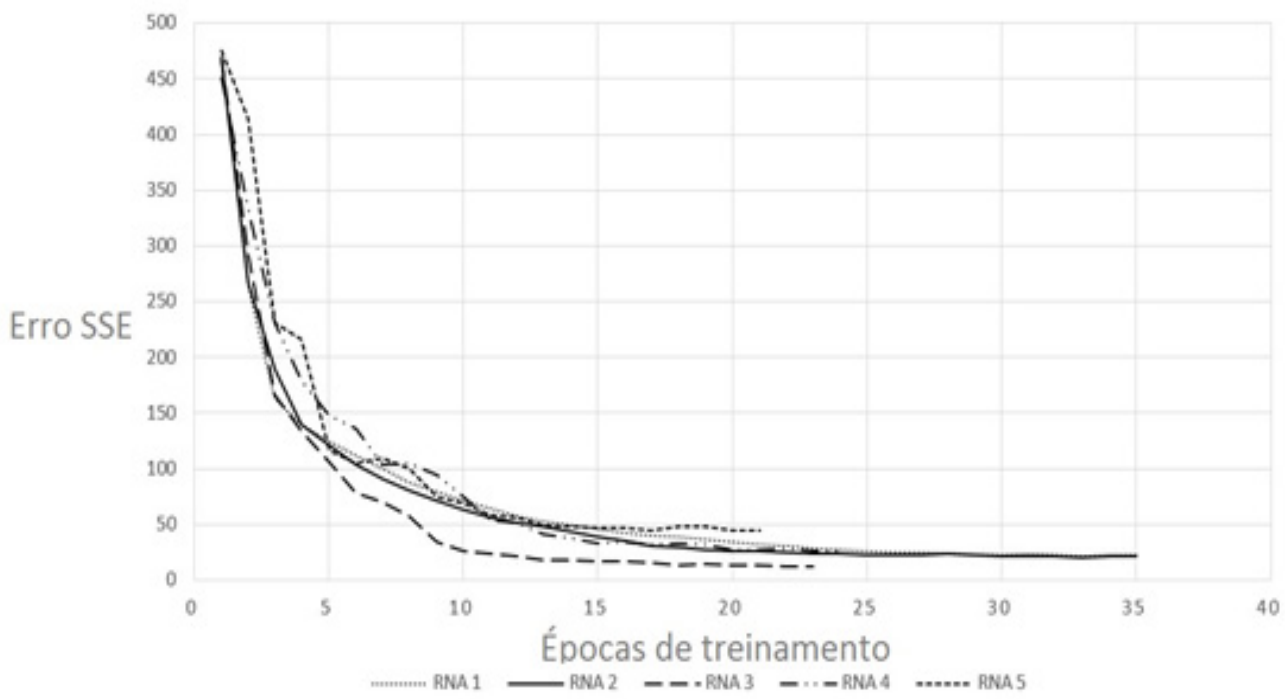
Com base nos resultados apresentados na Tabela 1, neste cenário, observou-se que a RNA com o menor erro foi a que possui 300 neurônios na camada oculta (RNA 3), permitindo assim, concluir que não necessariamente a RNA com a maior quantidade de neurônios apresentará o melhor desempenho

Tabela 1 – Desempenho das RNA no Cenário 1

RNA	Número de neurônios	Erro SSE
RNA 1	100	22.04625
RNA 2	200	20.96255
RNA 3	300	12.4371
RNA 4	400	26.46572
RNA 5	500	43.92732

Na Figura 6 são apresentadas as curvas de aprendizado de cada uma das RNA testadas, com suas respectivas épocas e erro SSE, o qual decresce no decorrer das épocas de treinamento.

Figura 6 – Curvas de aprendizado das redes neurais treinadas no Cenário 1.



A matriz de confusão da Tabela 2 apresenta o resultado percentual de acerto das RNA no reconhecimento dos padrões, onde os valores em destaque são as taxas de acerto, e os demais valores são as taxas de erro (falsos-positivos), do protocolo associado à coluna.

Tabela 2 – Matriz de confusão da rede neural selecionada no Cenário 1

	HTTP	FTP	SSH
HTTP	95.558	0.972	0.124
FTP	4.409	86.011	5.183
SSH	0.033	13.017	94.693

CENÁRIO 2

Neste segundo cenário as RNA implementadas foram treinadas, validadas e testadas com amostras extraídas de partes aleatórias do pacote de dados. Dessa forma, uma amostra pode conter partes tanto do cabeçalho, do conteúdo ou do rodapé do pacote.

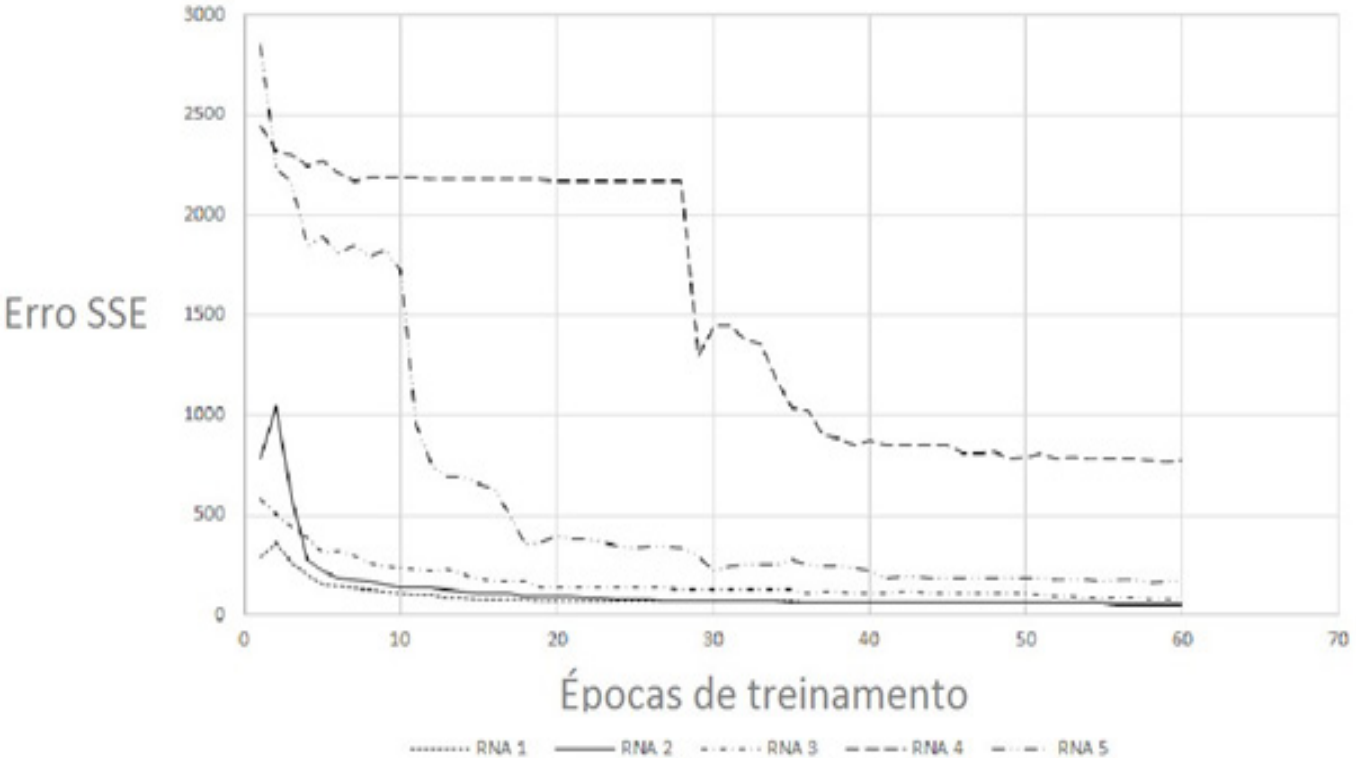
A Tabela 3 contém os resultados das taxas de erro SSE geradas pelas RNA treinadas no segundo cenário.

Tabela 3 – Desempenho das RNA no Cenário 2

RNA	Número de neurônios	Erro SSE
RNA 1	100	58.416916
RNA 2	200	53.460453
RNA 3	300	77.244644
RNA 4	400	767.769348
RNA 5	500	171.043472

Na Figura 7 é apresentado o gráfico com as curvas de aprendizado geradas pelas RNA durante o período de treinamento.

Figura 7 – Curvas de aprendizado das redes neurais treinadas no Cenário 2.



É apresentado na Tabela 4 a matriz de confusão com o percentual de acertos das RNA que obteve a menor taxa de erro SSE (RNA 2) no reconhecimento dos protocolos.

Tabela 4 – Matriz de confusão da rede neural selecionada no Cenário 2

	HTTP	FTP	SSH
HTTP	1,233	1,957	1,060
FTP	4,213	5,314	5,048
SSH	94,553	92,729	93,890

CONCLUSÃO

Com base nos resultados das matrizes de confusão, observou-se no primeiro cenário, que a RNA selecionada (RNA 3) teve um bom percentual de acerto, sendo 95.558% no reconhecimento de pacotes HTTP, 86.011% em pacotes FTP e 94.693% em pacotes SSH. Conclui-se também, que a RNA foi bastante eficiente na identificação de pacotes FTP e SSH, resultado atribuído em partes à utilização da parte inicial do pacote que contém informações especificas do protocolo, e também à capacidade de aprendizado do algoritmo backpropagation.

Já no segundo cenário, os resultados da RNA selecionada (RNA 2) não atingiram uma alta taxa de acerto no reconhecimento dos protocolos HTTP (1,233%) e FTP (5,314%), e no protocolo SSH obteve uma alta taxa de acerto (93,890%).

Esse resultado pode ser atribuído a diversos fatores, tais como a complexidade da estrutura dos dois primeiros protocolos, ou como a estrutura do protocolo SSH é mais simples comparada à estrutura destes.

Também pode ser considerado o fato das amostras utilizadas para treinamento, teste e validação serem extraídas de partes aleatórias do pa-

a tarefa de assimilação e reconhecimento dos padrões e resultando em menores taxas de acerto.

REFERÊNCIAS

AHMAD, I.; ABDULLAH, A. B.; ALGHAMDI, A. S. Application of Artifical Neural Network in Detection of DOS Attacks. 2nd International Conference on Security of Information and Networks, 2009.

HAYKIN S. Neural Networks. v2. Symon & Schuster, 1999.

HUSTON, G. TCP – How it works, The ISP Column, 2004.

JOHN, V. Countering Packet Sniffers Using Encrypted FTP, Jscape, 2012.

LONVICK,C; YLONEN T. The Secure Shell (SSH) Transport Layer Protocol, Network Working Group, 2006

MORETO, M.; ROLIM. J. Análise Automática de Oscilografias em Sistemas Elétricos de Potência, Sba Controle & Automação [online], 2010.

MATHWORKS, Documentation Center, sse, disponível em < http://www.mathworks.com/help/nnet/ref/sse.html>, aceso em 29/04/2014.

MCCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, vol. 5, pp. 115-133. 1943.

POSTEL, J.; Reynolds J. File Transfer Protocol, Network Working Group, 1985.

QADEER,M.; et al.. Network Traffic Analysis and Intrusion Detection using Packet Sniffer, 2010 Second International Conference on Communication Software and Networks, 2010.

SILVA, L. S. Uma metodologia pra detecção de ataques no trafego de redes baseada em redes Neurais, INPE, 2007

SILVA, R. M. Redes Neurais Aplicadas à Detecção de Intrusão em Redes TCP/IP, Pontifícia Universidade Católica do Rio de Janeiro, 2005.

TISSOT H. C.; CAMARGO L. C.; POZO A. T. R. Treinamento de Redes Neurais Feedforward: comparativo dos algoritmos Backpropagation e Differential Evolution, Brazilian Conference on Intelligent Systems 2012, 2012.

TORRES. G. Redes de Computadores, Curso Completo, Axel Books do Brasil, 2001

Artigo submetido em: 31/08/2014

Artigo aceito para publicação em: 22/12/2014