Investigando o Desempenho do Classificador ARTMAP Fuzzy na Detecção de Intrusos

Nelcileno Araújo[1]
[1]Instituto de Computação
Universidade Federal de Mato Grosso
Cuiabá, MT, Brasil
nelcileno@yahoo.com.br

Ailton Akira Shinoda[2]
[2]Departamento de Engenharia Elétrica
Universidade Estadual Paulista Júlio de Mesquita Filho
Ilha Solteira, SP, Brasil
shinoda@dee.feis.unesp.br

Ruy de Oliveira[3]
Ed' Wilson Tavares Ferreira[4]
[3,4]Departamento de Informática
Instituto Federal de Educação. Ciência e Tecnologia de Mato Grosso
Cuiabá, MT, Brasil
ruy@cba.ifmt.edu.br
ed@cba.ifmt.edu.br

***Abstract* – A escolha de um classificador de padrões que seja rápido e preciso no treinamento de um IDS (Sistema Detector de Intrusão) tem sido alvo das pesquisas recentes na área de segurança em redes de comunicações. Neste artigo investigamos o uso de um classificador ARTMAP *Fuzzy* na detecção de intrusos em uma rede simulada. Nosso estudo mostra a eficácia do ARTMAP *Fuzzy* como elemento central no mecanismo de aprendizagem de um IDS. Os bons resultados obtidos em termos de classificação correta e duração de treinamento sugerem que o classificador avaliado é realmente viável para esse tipo de aplicação.**

I. INTRODUÇÃO

O acesso não autorizado em um sistema ou rede de computadores é uma das ameaças mais sérias a segurança de computadores. Por isso, a necessidade do desenvolvimento de uma aplicação que emita um alerta inicial de intrusão, para que uma determinada ação defensiva seja executada para impedir ou minimizar os danos. Esta ferramenta é comumente chamada de Sistema Detector de Intrusão.

O IDS pode reconhecer padrões de atividades associados com regras de intrusão previamente conhecidas (detecção por assinaturas), e/ou identificar padrões incomuns de atividades que não convergem com o comportamento normal esperado (detecção por anomalia)[1].

O módulo analisador de um IDS desempenha um papel muito importante pois é responsável, a partir dos dados coletados, em verificar se as atividades relatadas estão dentro do normal. Quando essa análise é feita apenas pela associação com regras de intrusão, limita-se a funcionalidade do IDS apenas aos ataques conhecidos. Por causa disso, há um crescente interesse em IDS's que empregam técnicas de aprendizagem de máquina na análise de dados. Contudo, existem algumas técnicas que são bastante criticadas pelo excessivo consumo de recursos computacionais, mesmo apresentando altas taxas de detecção[2].

A partir disso, investigamos neste artigo o uso da rede neural ARTMAP *fuzzy* no treinamento de um IDS. A escolha deste classificador deve-se ao fato dele oferecer uma única solução, que é conhecida como dilema da estabilidade-plasticidade, onde ela tem a capacidade de preservar o conhecimento anteriormente adquirido (estabilidade) e para adaptar-se a novos padrões de classificação (plasticidade) [3]. A investigação baseia-se nos requisitos de tempo gasto para treinamento do IDS, taxa de detecção global, taxa de precisão e taxa de falsos positivos sobre a base de treinamento do banco de dados de detecção de intrusão KDD99.

O restante do artigo está organizado como segue. Na segunda seção faremos o embasamento teórico sobre o classificador ARTMAP *fuzzy*, destacando o seu algoritmo. Na seção seguinte apresentamos o modelo proposto de treinamento para o IDS por meio do classificador ARTMAP *fuzzy* e analisamos o desempenho dele levando em conta alguns requisitos. Finalmente, na última seção realizamos as devidas conclusões obtidas com esta investigação e sugerimos os trabalhos futuros na continuação dessa pesquisa.

II. REDES ARTMAP FUZZY

Nos últimos anos, a ênfase aplicada no treinamento de um IDS tem voltado para um paradigma de aprendizagem por amostras. Neste sentido, as redes neurais tem sido amplamente utilizada na detecção de intrusão [2].

A rede neural ARTMAP *fuzzy* vem aliar esforços com mais um sistema de aprendizagem supervisionada, capaz de auto-organizar o reconhecimento estável de categorias em respostas aos padrões de entrada arbitrários [3].

A Fig. 1 apresenta a arquitetura da rede ARTMAP *fuzzy*. Ela é composta por dois módulos ART_a nebuloso e ART_b nebuloso, que possuem a mesma estrutura da rede neural ART1 usando as operações envolvidas na lógica fuzzy [4]. Eles são interligados por um módulo conhecido como inter-ART que controla o treinamento de um mapa associativo de categorias de reconhecimento da ART_a para categorias de reconhecimento da ART_b. O inter-ART combinam os parâmetros de entrada com os parâmetros de saída através do *match tracking*, de forma a maximizar a generalização das categorias de reconhecimento e minimizar o erro da rede [3,5].

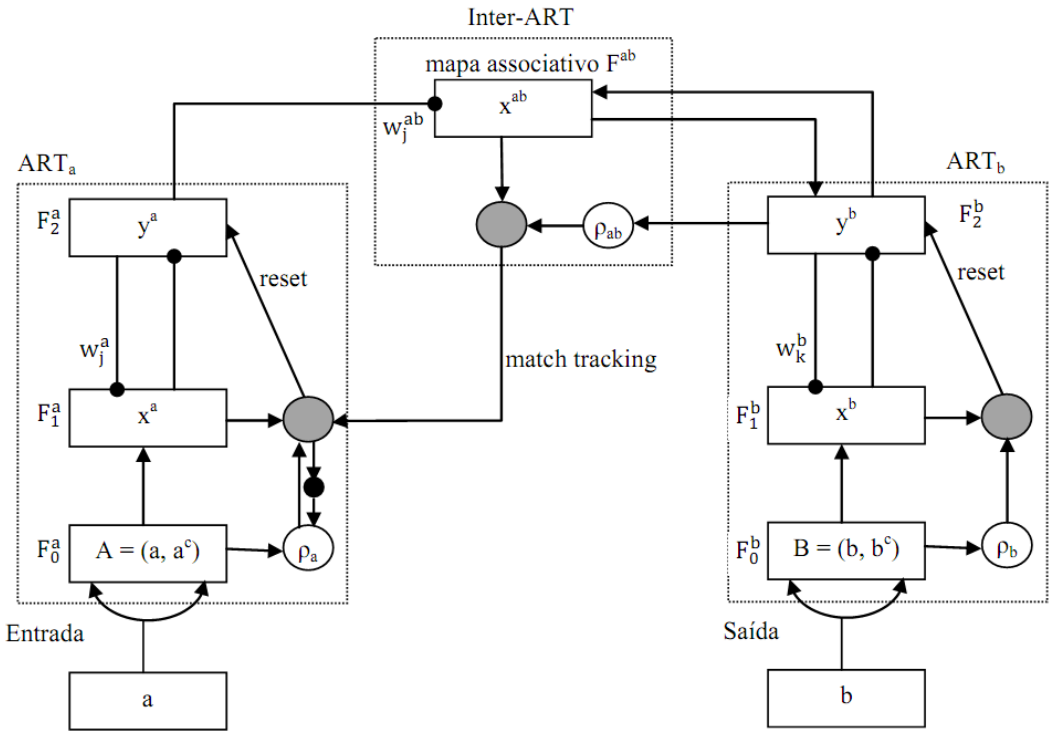

Fig. 1. Arquitetura da rede neural ARTMAP Fuzzy [4]

O algoritmo desta rede neural consiste nos seguintes passos [3,5]:

Passo 1: Normalizar os vetores de entrada ART_a e de saída ART_b, se necessário: Inicialmente, todos os valores dos neurônios devem ser normalizados se não estiverem entre 0 e 1.

Passo 2: Codificar os vetores dos módulos ART_a e ART_b: Um novo padrão de entrada deve sofrer uma codificação complementar preliminar para se preservar a amplitude da informação.

Passo 3: Iniciar os pesos e parâmetros dos módulos ART_a, ART_b e Inter-ART: deve-se iniciar os pesos (valor 1 = todas as categorias desativadas), taxa de treinamento (β entre 0 e 1), parâmetro de escolha ($\alpha > 0$) e parâmetro de vigilância (ρ_a, ρ_b e ρ_{ab} entre 0 e 1);

Passo 4: Escolha da categoria para os módulos ART_a e ART_b: Se mais de um neurônio está ativo, é escolhido aquele com o maior índice de ordenação (maior valor).

Passo 5: Teste de vigilância dos módulos ART_a e ART_b: A ressonância ocorre se é satisfeito o critério de vigilância. Caso contrário, se o critério de vigilância falhar para a categoria escolhida, ocorre o *reset* e um novo índice é escolhido (retornar ao passo 4). O processo de busca se repete até que o índice escolhido satisfaça o teste de vigilância.

Passo 6: *Match tracking* (teste de ressonância) entre os módulos ART_a e ART_b: Verificação se houve casamento da entrada com a saída. Se não houver casamento deve-se procurar outro índice que satisfaça o teste;

Passo 7: Adaptação dos pesos: O vetor (camada F_2) dos módulos ART_a, ART_b e inter-ART são atualizados com os novos pesos.

Passo 8: Repetir passos 4 à 7 para todos os pares a serem treinados.

Na próxima seção apresentamos como o classificador ARTMAP fuzzy pode ser aplicado no treinamento de um IDS.

III. INVESTIGANDO O DESEMPENHO DO CLASSIFICADOR ARTMAP FUZZY NA DETECÇÃO DE INTRUSOS

Na seção anterior vimos o funcionamento do classificador ARTMAP fuzzy. Agora, aplicamos esta técnica no treinamento de um IDS.

A. Treinamento de um IDS por meio do classificador ARTMAP Fuzzy

Primeiramente, definimos a estrutura dos vetores de entrada (F_0^a) e saída (F_0^b). Sendo, o vetor de entrada alimentado com os registros da base de treinamento da base de dados de detecção de intrusão KDD99. Contudo, aplicamos a base de treinamento otimizada proposta em [6], pois diminuimos o espaço vetorial apenas para os atributos mais significativos. Criando assim um vetor de entrada *a* com uma dimensão igual a 14.

A Tabela I descreve a formação do vetor de saída *b*, onde é empregado uma codificação binária para as grandezas de saída de 2 bits e são formalizados em duas classes, onde a classe S1 representa o tráfego com o comportamento esperado na rede simulada e a classe S2 representa o tráfego quando apresenta um padrão de atividade diferente do comportamento esperado.

TABELA I
VETORES DE SAÍDA (BINÁRIOS) CORRESPONDENTES ÁS CLASSES DE DETECÇÃO DE INTRUSOS

Tipo de detecção	Classe	Vetor de saída *b*
Normal	S1	01
Anomalia	S2	10

A partir disso, os dados contidos nos vetores são normalizados e codificados formando os vetores da primeira camada das redes ART_a (F_1^a) e ART_b (F_1^b).

A seguir, o classificador realiza o processamento das duas redes ART. Escolhendo a categoria ativa, por meio da função escolha $T_{j,k}$ definida em (1).

$$T_{J,K} = \frac{|I \wedge W_{J,K}|}{\alpha + |W_{J,K}|} \quad (1)$$

Sendo:
I – o vetor de entrada na primeira camada da rede ART (F_1)
W – a matriz de pesos da rede ART
α – o parâmetro de escolha

Logo que a ressonância seja confirmada em cada módulo ART pelo teste de vigilância [3] descrito em (2), têm-se a categoria ativa deles (J em ART_a e K em ART_b).

$$\frac{|I \wedge W_{J,K}|}{|I|} \geq \rho_a \quad (2)$$

Sendo:
I – o vetor de entrada na primeira camada da rede ART (F_1)
W – a matriz de pesos da rede ART
ρ_a – parâmetro de vigilância aplicado na rede ART

Através do processo de *match traking* verifica-se se a categoria ativa no ART_a corresponde ao vetor de saída desejada apresentado no ART_b. O critério de vigilância [3] é dado por:

$$\frac{|y^b \wedge W_{JK}^{ab}|}{|y^b|} \geq \rho_{ab} \quad (3)$$

Sendo:
y^b – vetor de saída do ART_b (padrão de atividade F_2^b)
W_{JK}^{ab} – matriz de pesos do módulo inter-ART
ρ_{ab} – critério de vigilância

Se (3) não for satisfeita, é feito um incremento mínimo no parâmetro de vigilância da rede ART_a, suficiente apenas para excluir a atual categoria e selecionar outra categoria, que se tornará ativa e entrará novamente no processo até que (3) seja satisfeito.

Com a ressonância confirmada, os pesos dos módulos ART_a e ART_b, são atualizados em suas matrizes (W^a e W^b) por meio da equação descrita em (4).

$$W_J^{novo} = \beta\left(I \wedge W_J^{velho}\right) + (1 - \beta)W_J^{velho} \quad (4)$$

Sendo:
J – categoria ativa
W_J^{novo} – matriz de pesos atualizados
W_J^{velho} – matriz de pesos referente à atualização anterior
β – taxa de treinamento

A adaptação de pesos [3] para o módulo inter-ART é realizada da seguinte maneira:

$$W_{J,K}^{ab} = \begin{cases} 1, k = K \\ 0, k \neq K \end{cases} \quad (5)$$

Sendo:
k – a categoria do neurônio na matriz de pesos da rede ART_a
K – a categoria do neurônio na matriz de pesos da rede ART_b

Repetimos este processo para todos os registros da base de treinamento utilizada. A seguir, mostramos o desempenho do IDS treinado pelo classificador ARTMAP fuzzy segundo os requisitos avaliados.

B. Experimentos e Resultados

O banco de dados escolhido para o experimento é a base de dados KDD99. Ela está disponível em [7]. Esta base é amplamente empregada no treinamento e teste de IDS [8]. Utilizamos especificamente a base de treinamento (10% KDD99) para reduzir a quantidade de amostras. Além disso, empregamos o método de seleção de atributos usado em [6] para restringir nas amostras apenas os atributos mais significativos.

Para analisar o desempenho de um IDS treinado com um classificador ARTMAP *fuzzy* aplicamos a base de treinamento do KDD99 em três cenários descritos pela Tabela II. Esta base possui 125793 instâncias.

Os parâmetros utilizados no classificador ARTMAP *fuzzy* para o treinamento do IDS são os mesmos empregados em [5] que trata de uma rede neuro-*fuzzy-wavelet* para detecção e classificação de anomalias de tensão em sistemas elétricos de potência. A escolha desse conjunto de valores deve-se pela semelhança do objeto pesquisado no trabalho deles com o nosso, os dois procuram detectar distúrbios numa base de dados analisada. Dessa forma, aplicamos os seguintes valores apresentados na Tabela III.

O desempenho do IDS é investigado de acordo com os seguintes requisitos: tempo gasto para treinamento do IDS, taxa de detecção global, taxa de precisão e taxa de falsos positivos.

TABELA II
CONFIGURAÇÃO DOS CENÁRIOS SIMULADOS.

Cenário	Total de Conexões da Base de Treinamento KDD99 contida em cada fase (%)	
	Treinamento	Teste
1	33%	67%
2	50%	50%
3	66%	34%

TABELA III
PARÂMETROS DE CONFIGURAÇÃO USADOS NO CLASSIFICADOR ARTMAP FUZZY.

Parâmetros	Valor
Parâmetro de escolha (α)	0,001
Taxa de treinamento (β)	1
Parâmetro de vigilância da rede ART_a (ρ_a)	0,99
Parâmetro de vigilância da rede ART_b (ρ_b)	0,9
Parâmetro de vigilância do módulo inter-ART(ρ_{ab})	0,99

Todas as simulações foram realizadas por meio da ferramenta de programação WEKA [9] que já tinha sido utilizada em [6] e mostrou-se bastante eficiente na implementação de classificadores de padrões.

Na Tabela IV avaliamos os requisitos de desempenho tempo gasto para o treinamento do IDS e taxa de detecção global das instâncias analisadas. No primeiro parâmetro percebe-se que todos os cenários tiveram um duração de treinamento bem semelhante (por volta de 2 minutos) e um valor bem reduzido. Isso deve-se principalmente pela propriedade de estabilidade-plasticidade a qual faz com que o classificador empregue uma aprendizagem incremental (treinando somente os novos padrões de atividades) sem esquecer dos padrões de atividades anteriormente assimilados.

Sendo assim, nenhum novo treinamento será necessário se as novas amostras inseridas na base de treinamento não corresponderem a novos padrões de atividades. Por consequência, diminuição no período de treinamento do IDS.

Outro ponto importante a ser ressaltado é a pequena diminuição de tempo de treinamento no Cenário 2. Isso demonstra que não houve uma grande plasticidade na aquisição de conhecimento, pois quando novos padrões aparecem a tendência que ocorra uma certa elevação neste requisito, que é comprovado no Cenário 3 onde o valor registrado volta a crescer.

Com relação a taxa detecção global pode ser visto que a partir do Cenário 2 ocorre uma convergência para um valor em torno de 88%. Essa convergência é o resultado da diminuição de ocorrências de novos padrões de atividades. Dessa forma, aumenta-se as amostras na base de treinamento mas novas categorias de reconhecimento não são apresentadas.

Os resultados apresentados na Fig. 2 demonstram que quando alimentamos o IDS com uma base de dados com bastante diversidade de dados conseguimos uma taxa de precisão maior. Por isso, a partir do Cenário 2 quando temos a base de treinamento envolvendo 50% das conexões atingi-se um precisão em torno de 90% que é bastante alta levando em conta a baixa duração de treinamento.

TABELA IV
RESULTADOS OBTIDOS NOS CENÁRIOS DE SIMULAÇÃO.

Cenário	Requisitos de desempenho	
	Tempo gasto para treinamento do IDS (seg)	Taxa de detecção global (%)
1	122,97	72,85
2	118,81	87,20
3	121,54	88,91

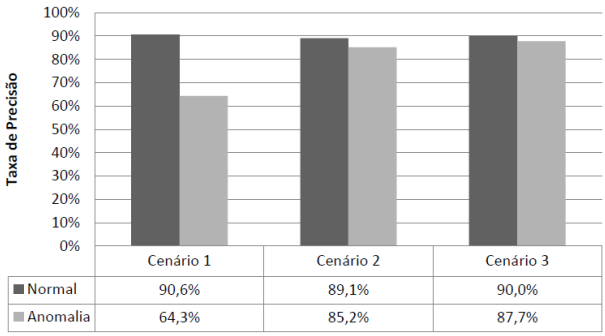

Fig. 2. Resultados obtidos pela taxa de precisão nos cenários simulados

Um outro requisito muito importante é a taxa de falsos positivos porque em alguns casos o classificador pode ter uma elevada taxa de detecção das categorias de reconhecimento mas com uma alta taxa de falsos positivos, identificando uma baixa precisão do classificador no reconhecimento de padrões [6].

Na Fig. 3 observa-se que o classificador neuro-*fuzzy* quando possui uma base de informações limitada (33% das conexões utilizadas para treinamento) apresenta uma dificuldade na detecção da categoria de reconhecimento *anomalia*, rotulando quase a metade das conexões normais como um comportamento não esperado. Contudo, com o crescimento da base de treinamento utilizada no IDS ocorre um decréscimo acentuado na taxa de falsos positivos na categoria de reconhecimento *anomalia*. Como também, uma convergência no valor apresentado (em torno de 10%) neste requisito para as duas categorias de reconhecimento investigadas.

Os resultados apresentados nesta investigação demonstram a viabilidade do classificador ARTMAP *fuzzy* no treinamento de um IDS. A principal vantagem diagnosticada é exatamente o tempo gasto para o treinamento do IDS. Mesmo quando submetido a uma grande quantidade de amostras (em torno de 85000), o tempo gasto foi pequeno devido a sua característica de armazenar os conhecimentos adquiridos e retreinar o IDS apenas quando surgir novos padrões de atividade.

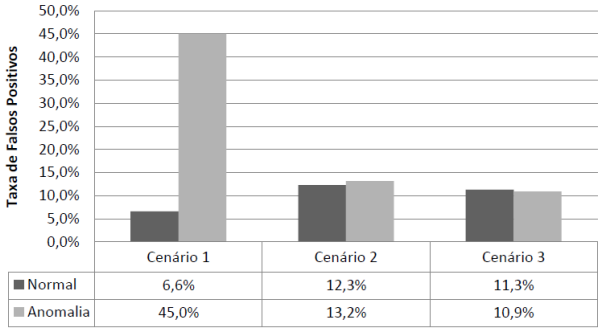

Fig. 3. Resultados obtidos pela taxa de falsos positivos nos cenários simulados

Esta propriedade é muito importante pois a maioria dos classificadores necessitam retreinar toda base de dados sempre que novas amostras são inseridas. Quando os dados estão sendo retreinados, ás vezes mudanças significativas ocorrem no domínio da classificação. Além disso, quando um novo padrão

é assimilado, não há nenhuma garantia que a topologia de rede e o parâmetro de aprendizagem anterior ainda ofereceram uma boa solução. Dessa forma, a duração de treinamento aumenta porque novas regiões de decisão são necessárias, a rede terá mais camadas escondidas e o tamanho da matriz de entrada da rede neural será maior [10].

Um ponto a ser melhorada no classificador ARTMAP *fuzzy* é aumentar a taxa de precisão em consonância com a redução da taxa de falsos positivos, mas mesmo assim o valores apresentados não são desanimadores, levando-se em conta a simplicidade do algoritmo apresentado. Com certeza, podemos melhorar a precisão do IDS e consequentemente reduzir a taxa de falsos positivos se aplicarmos o classificador neuro-*fuzzy* em conjunto com outras técnicas de baixo impacto computacional (wavelet, *rough sets*, algoritmos genéticos).

IV. Conclusões e trabalhos futuros

Os resultados apresentados (duração de treinamento ≅ 120s, taxa de detecção ≅ 90%, taxa de precisão ≅ 90%, taxa de falsos positivos ≅ 10%) demonstram a viabilidade do classificador ARTMAP *fuzzy* na detecção de intrusos. Principalmente, pela capacidade de diminuir a duração de treinamento do IDS sem acarretar uma degradação acentuada nos valores de classificação correta das amostras analisadas.

Como continuação desse trabalho, pretendemos investigar o uso do classificador ARTMAP *fuzzy* em arquiteturas híbridas de treinamento, ou seja, combiná-lo com técnicas de classificação computacionalmente leves. Além disso, procurar técnicas de seleção de atributos que reduzam ainda mais o espaço vetorial das amostras empregadas na aprendizagem do IDS. Contudo, o nosso objetivo deve ser norteado na limitação do consumo dos recursos computacionais e aumento da capacidade de detecção e classificação de intrusos.

Agradecimentos

Este material é baseado num projeto de pesquisa financiado pela Fundação de Amparo à Pesquisa de Mato Grosso (FAPEMAT) sobre a supervisão do Grupo de Pesquisa em Redes e Segurança (GPRS). O GPRS é gerenciado pelo Instituto Federal de Educação, Ciência e Tecnologia do Estado de Mato Grosso (IFMT) em conjunto com a Universidade Federal de Mato Grosso (UFMT), a Universidade Estadual Paulista (UNESP) e a Universidade Federal de Uberlândia (UFU). Os autores agradecem as instalações e equipamentos oferecidos pelo IFMT para o desenvolvimento deste trabalho.

Referências

[1] P. Souza, *Estudo sobre sistemas de detecção de intrusão por anomalias: uma abordagem utilizando redes neurais*, Dissertação de Mestrado, Universidade Salvador/Salvador, 2008.

[2] C. B. Vilakazi & T. Marwala, "Application of feature selection and fuzzy ARTMAP to intrusion detection," in *Proceedings of 2006 IEEE International Conference on Systems, Man and Cybernetics*, 2006, pp. 4880-4885.

[3] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynold & D. B Rosen, "Fuzzy ARTMAP: A neural network for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Network*, vol. 3, n. 5, pp. 689-713, 1992.

[4] G. A. Carpenter, S. Grossberg & D. B Rosen, "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, n. 1, pp. 759-771, 1991.

[5] F. C. V. Malange, *Rede Neuro-Fuzzy-Wavelet para detecção e classificação de anomalias de tensão em sistemas elétricos de potência*, Tese de Doutorado, Universidade Estadual Paulista/Ilha Solteira, 2010.

[6] N. V. de S. Araújo, A. A. Shinoda, R. de Oliveira & E. T. Ferreira, "Identificando características importantes na base de dados de detecção de intrusão KDD99 por meio da seleção de atributos com abordagem híbrida," in *Proceedings of the 8th International Information and Telecommunication Technologies Symposium*, 2009.

[7] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba & K. Das, "The 1999 DARPA off-line intrusion detection evaluation," *Computer Networks*, vol.34, n.4, pp. 579-595, 2000.

[8] I. Ahmad, M. A. Ansari & S. Mohsin, "Performance comparison between backpropagation algorithms applied to intrusion detection in computer networks systems," *Recent Advances in Systems, Communications and Computers*, pp. 47-52, 2008.

[9] R. R. Bouckaert et al., WEKA manual for version 3-7-0. http://www.cs.waikato.ac.nz/ml/weka/, Acesso: 20 de agosto de 2009.

[10] R.P Lippmann, "Pattern classification using Neural Networks," *IEEE Communication Magazine*, vol. 27, n. 11, pp. 47-62, 1989.