Redes Neurais Utilizadas na Detecção de Intrusões em Redes de Computadores

Jessica Nancy Salas Poveda
Engenharia Elétrica – Universidade Presbiteriana Mackenzie
São Paulo – SP – 01302-907 – Brasil

e

Paulo Alves Garcia
Engenharia Elétrica – Universidade Presbiteriana Mackenzie
São Paulo – SP – 01302-907 – Brasil

Resumo — Este artigo investiga uma metodologia de apoio à detecção de ataques no tráfego de redes, baseada em redes neurais, testes de modelos de detecção e métodos para detecção de ataques no tráfego de rede baseados em redes neurais. O artigo apresenta também, estudos sobre as redes neurais que devem ser utilizadas em gerência de redes de computadores, analisando os resultados obtidos e dessa forma propondo o uso de um tipo de rede neural artificial para os sistemas de detecção de invasão.

Palavras chave — Internet, Detecção de Intrusos, Segurança da Informação, Redes Neurais, Inteligência Computacional.

1. INTRODUÇÃO

Atualmente, com o crescimento acelerado das redes de computadores, tem surgido uma constante preocupação com a privacidade e a segurança das informações que trafegam nessas redes. Empresas e organizações interconectadas através de redes de pacotes e da Internet são constantemente ameaçadas por ataques e intrusões em suas redes. Em função disso, vários mecanismos de segurança em redes de computadores tem sido projetados para impedir o acesso não autorizado a sistemas e a dados restritos. A fim de evitar situações inesperadas e indesejadas, e impedir a proliferação dos ataques continuamente lançados contra diferentes alvos na rede, foram implantados mecanismos de proteção, tais como *firewalls,* antivírus, sistemas de autenticação, mecanismos de criptografia e sistemas de detecção de intrusão nos ambientes de rede por todo o mundo.
Diversas técnicas para reconhecimento de intrusões têm sido propostas e disponibilizadas através de ferramentas de domínio público ou soluções comerciais. Apesar da existência de métodos modernos de segurança, sistemas denominados detectores de intrusos são muito utilizados na detecção de intrusões. Os sistemas de detecção de intrusão compõem uma parte essencial da infra-estrutura de segurança em camadas e têm por objetivo a análise de dados do tráfego
Redes Neurais Artificiais são sistemas computacionais com processamento altamente paralelo e distribuído, e que apresentam a capacidade de armazenar conhecimento experimental. Estes sistemas computacionais apresentam características similares e análogas às observadas no funcionamento do cérebro humano. Através da capacidade de generalização das redes neurais, espera-se que o sistema detecte novos ataques, mantendo assim uma alta taxa de acertos. A aprendizagem em redes neurais é caracterizada pela capacidade que as redes possuem de modificar o seu comportamento em resposta a eventos ou situações que ocorrem no ambiente externo e que fornecem um conjunto de entradas associadas a um conjunto de saídas.

2. REDES NEURAIS ARTIFICIAIS

As redes neurais artificiais são um método para solucionar problemas através da simulação do cérebro humano, inclusive em seu comportamento, ou seja, aprendendo, errando e fazendo descobertas. São técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

O desenvolvimento dos modelos de redes neurais artificiais surgiu como uma tentativa inicial de reproduzir o alto desempenho do cérebro humano em tarefas de alto grau de complexidade [1]. Uma rede neural artificial é uma estrutura composta de unidades processadoras simples, distribuídas e paralelas, que tem o propósito de armazenar conhecimento por meio de mecanismos empíricos, para assim torná-lo disponível para o uso.

O conhecimento da rede neural artificial se concentra nos pesos definidos para as conexões sinápticas da rede, formando uma representação compacta e distribuída desse conhecimento e proporcionando capacidades de generalização e adaptabilidade à rede neural. Porém, esta organização não baseada em regras, impossibilita às redes neurais a explicação de forma abrangente do processo computacional tradicional [2].

Neste artigo, foram utilizadas as redes neurais Perceptron Multicamadas, cuja modelagem da arquitetura envolve a escolha da quantidade de camadas e o número de unidades em cada camada.

O processamento de cada unidade é influenciado pelo processamento efetuado pelas unidades das camadas. Cada camada desempenha um papel específico, conforme representado na Figura 1 e de acordo com a descrição abaixo:

1) Camada de Entrada: Receptora de estímulos;
2) Primeira Camada Oculta: Cada unidade desta camada define uma reta no espaço de decisão, refletindo as características dos padrões apresentados;
3) Segunda Camada Oculta: Combina as retas definidas pela camada anterior, formando regiões convexas onde o

número de lados é definido pelo número de unidades da camada anterior conectado a unidade desta camada;

4) Camada de Saída: Combina as regiões formadas pela camada anterior, definindo o espaço de saída da rede.

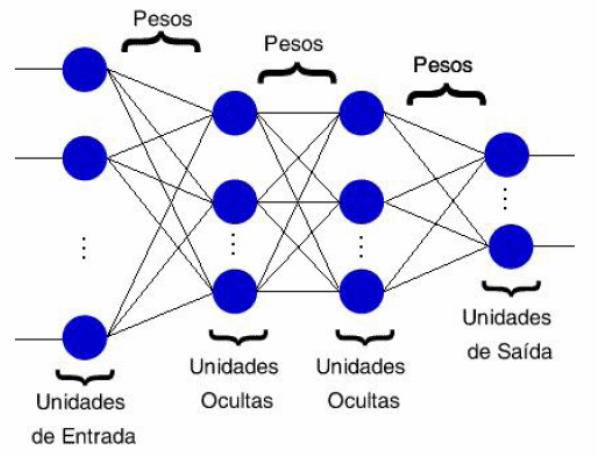

Figura 1 – Camadas de uma rede neural

3. DETECÇÃO DE INTRUSÕES

Os ataques dependem das habilidades do *hacker* e da sua facilidade de identificação com os processos. O ataque pode acontecer de forma indireta, com a utilização de ferramentas de intrusão. O *hacker* obtém dados de usuários que neste caso são vetores para alcançar informações de um objetivo maior, como uma empresa, organização ou até mesmo o governo.

Para a detecção de intrusão são utilizados sistemas especialistas e automatizados. Estes Sistemas de Detecção de Intrusão - SDI notificam o especialista humano sempre que detectarem alguma atividade considerada suspeita ou fora dos padrões normais.

Os Sistemas Detectores de Intrusão monitoram eventos que ocorrem em um sistema ou rede de computadores e os analisa em termos de possíveis incidentes, que correspondem a violações ou ameaças iminentes às políticas de segurança [3].

Detecção de intrusão em sistemas computacionais é uma tecnologia relativamente nova, uma vez que pesquisas nesta área tiveram início a partir de 1980. Os Sistemas de Detecção de Intrusão mais recentes possuem a capacidade de, opcionalmente, atuar automaticamente ao detectar uma anomalia.

Os sistemas de detecção de intrusão servem para indicar que alguma tentativa de intrusão foi feita no sistema. Existem dois tipos diferentes de detecção, os baseados na rede e os baseados na estação.

4. REDES NEURAIS APLICADAS À DETECÇÃO DE INTRUSÕES

O modelo proposto trata de um mecanismo de segurança capaz de detectar o comportamento intrusivo em determinadas seções de rede. Esta detecção é realizada através de um módulo que possui características desejáveis em sistemas de detecção de intrusão, como a utilização de redes neurais para o reconhecimento dos padrões de ataque de forma a incorporar à estrutura habilidades como adaptabilidade e generalização.

Optou-se neste artigo, pela utilização do *dataset* gerado pelo DARPA 98 [4].

O *dataset* é composto por padrões de invasões (ataques *Remote-to-local* e *User-to-Root*), também conhecidos como *exploits*, padrões de reconhecimento de vulnerabilidades (*Probes*) e ataques de negação de Serviço (DoS). Juntamente com padrões de tráfego normais.

Todos os padrões estão com etiquetas (*Label*) informando qual é o tipo de tráfego que ele pertence, facilitando o treinamento e também a análise de resultados do modelo.

Estas capacidades possibilitam a correta classificação de padrões semelhantes, mesmo que alguns destes padrões não façam parte do conjunto de treinamento. Esta característica colabora na composição de uma solução robusta e adaptável a mudanças no ambiente onde está inserida.

Exploits

Uma grande amostra de ataques reais de computador é necessária para testar um sistema de detecção de intrusão. Esses ataques devem cobrir as diferentes classes de tipos de ataque e conter *exploits* tanto para vulnerabilidades recentemente descobertas, como para as mais antigas. Cada novo *exploit* tem um período de tempo durante o qual é mais perigoso. Com o passar do tempo, mais pessoas tomam conhecimento da vulnerabilidade e aplicam os *patches* (remendos) aos seus sistemas para fazê-los resistentes aos *exploits*. Mesmo depois das notícias de uma vulnerabilidade se tornarem comuns, alguns sistemas podem não estar preparados. Alguns administradores de sistemas com menos experiência passam anos sem ter falhas de segurança populares detectadas e consertadas. Alguns desses ataques mais velhos foram incluídos no sistema de ataques usados para a avaliação DARPA 1998.

Pode-se utilizar uma taxonomia para classificar os ataques às redes de computadores. Uma boa taxonomia permite classificar ataques em grupos, com propriedades em comum. Uma vez que esses grupos foram identificados, o trabalho de testar apropriadamente um sistema de detenção de intrusão torna-se mais fácil, porque em vez de desenvolver cada possibilidade ataque, pode-se escolher um subconjunto representativo em cada grupo.

A taxonomia define uma aproximação para definir um ranking dos níveis de privilégio. Seguem a seguir as categorias de privilégio que são aplicadas neste artigo.

Ataques de negação de serviços: Um ataque de negação de serviço ocorre quando o atacante torna algum recurso do sistema ou uso da memória demasiadamente ocupado, impedindo o sistema de tratar pedidos legítimos; negando acessos legítimos aos usuários de uma máquina. Há muitas variedades de ataques de negação de serviço (ou DoS). Alguns ataques de DoS (como um *mailbomb*, Netuno, ou ataque de *smurf*) abusam de uma característica perfeitamente legítima. Outros (*teardrop*, *Ping of Death*) criam pacotes corrompidos que confundem a pilha TCP/IP da máquina que está tentando reconstruir o pacote.

Ataques *User-to-Root*: Exploits *User-to-Root* são uma classe de exploits na qual o atacante começa com acesso a uma conta de usuário normal no sistema (possivelmente através de um *sniffer*, um dicionário de ataque, ou engenharia social) e é capaz de explorar alguma vulnerabilidade para ganhar o acesso de *root* do sistema.

Ataques *Remote-to-Local*: Um ataque *Remote-to-User* ocorre quando um atacante tem a capacidade de enviar pacotes a uma máquina através da rede - mas não tem uma conta nesta máquina – e explora alguma vulnerabilidade para ganhar acesso local como usuário daquela máquina. Há muitos modos possíveis que um atacante pode obter acesso não autorizado a uma conta local em uma máquina.

Ataques *Probes*: Nos últimos anos, um número crescente de programas tem sido distribuído, os quais podem vasculhar automaticamente uma rede de computadores e reunir informação ou achar alguma vulnerabilidade conhecida. Esses *probes* de rede são bastante úteis a um atacante que está organizando um futuro ataque. Um atacante com um mapa de qual máquina e serviços estão disponíveis na rede, pode usar estas informações para procurar pontos fracos.

Dados de entrada do teste

A fim de se realizarem testes em redes neurais como classificadoras de SDIs, uma base de dados com padrões de ataques e padrões de tráfego normais foi necessária. Essa base de dados disponível tem sido utilizada há muito tempo e é oriunda do primeiro esforço para testes de sistemas de detecção de intrusão, que foi feito pelo DARPA em 1998 e 1999. Estas são as bases de dados prontas que podem ser utilizadas para os testes de avaliação de desempenho em SDIs [5].

A base de dados disponibilizada pelo DARPA está descrita na Tabela 1, onde se pode ver o número de dados ou linhas contidos em cada uma das bases, bem como a quantidade de ataques contidos em cada um dos conjuntos. O dataset foi separado em 4 conjuntos distintos, cada um representando uma classe de tráfego. Foi criado um arquivo para *Remote-to-Local,* outro para *User-to-Root*, outro para *Probe* e outro para DoS. Com isso, foi possível levantar a quantidade de padrões contidos em cada uma das classes para criar os conjuntos de treinamento, teste e validação, com o mesmo número de padrões de ataque e tráfego normal.

Cada linha do dataset é composta por 41 entradas descritas na tabela 1, na qual estão apresentados os dados discretos e contínuos, no entanto, para treinamento, usou-se dados contínuos e normalizados. Aos valores máximos e mínimos para os valores contínuos a serem normalizados e aos valores discretos, associou-se valores entre 0,0 e 1,0.

Treinamento das redes neurais

O treinamento usado para inserir conhecimento nas redes neurais foi executado utilizando o *dataset* de treinamento utilizado pelo DARPA 98. Para o treinamento também foram desenvolvidos dois programas, um para o treinamento da rede, e o outro para a validação do treinamento. Deste modo, treinou-se a rede utilizando para isso o programa de treinamento, e depois se verificaram-se os erros, utilizando-se o programa de validação. Para se iniciar o processo, foram criados os pesos aleatoriamente, definindo-se uma semente para a geração dos números aleatórios e definindo-se o valor mínimo e máximo para os pesos.

Nº	Campo	tipo	degraus
01	Duration	continuous	10
02	protocol_type	symbolic.	3
03	Service	symbolic	67 (binário)
04	flag:	symbolic	11
05	src_bytes:	continuous	20
06	dst_bytes	continuous	20
07	Land	symbolic	10
08	wrong_fragment	continuous	10
09	Urgent	continuous	10
10	Hot	continuous	10
11	num_failed_logins	continuous	10
12	logged_in	symbolic.	10
13	num_compromised	continuous	10
14	root_shell	continuous	10
15	su_attempted	continuous	10
16	num_root	continuous	10
17	num_file_creations	continuous	10
18	num_shells	continuous	10
19	num_access_files	continuous	10
20	num_outbound_cmds	continuous	10
21	is_host_login	symbolic	10
22	is_guest_login	symbolic	10
23	Count	continuous	10
24	srv_count	continuous	10
25	serror_rate:	continuous	10
26	srv_serror_rate	continuous	10
27	rerror_rate	continuous	10
28	srv_rerror_rate	continuous.	10
29	same_srv_rate	continuous	10
30	diff_srv_rate	continuous	10
31	srv_diff_host_rate	continuous	10
32	dst_host_count	continuous	10
33	dst_host_srv_count	continuous	10
34	dst_host_same_srv_rate	continuous	10
35	dst_host_diff_srv_rate	continuous	10
36	dst_host_same_src_port_rate	continuous	10
37	dst_host_srv_diff_host_rate	continuous	10
38	dst_host_serror_rate	continuous	10
39	dst_host_srv_serror_rate	continuous	10
40	dst_host_rerror_rate	continuous	10
41	dst_host_srv_rerror_rate	continuous	10

Tabela 1 – Descrição dos campos de entrada.

Feito isso, executou-se a primeira bateria de treinamento, que consistiu em treinar a rede para certa quantidade de períodos,

gerando-se novos pesos baseados nos pesos iniciais, como se pode ver na figura 2. Verificou-se o erro com o programa de validação para o conjunto de pesos gerados pelo programa de treinamento. Se o erro encontrado era menor que o anterior, executou-se outra bateria de treinamento, partindo-se com o conjunto de pesos validados como valores iniciais e assim reduzindo-se o erro. Se o erro aumentasse, era reduzida a quantidade de períodos de treinamento e verificando-se o resultado, caso fosse melhor, validava-se o conjunto de pesos e repetia-se o processo para o mesmo número de períodos. Caso o conjunto fosse pior, reduzia-se o número de períodos novamente repetindo-se o processo até o número de períodos chegar a zero
Portanto, através do programa de validação e da utilização de um *dataset* composto de padrões de ataques desconhecidos, verifica-se o erro global, e quais os padrões que não foram bem classificados e dessa forma foram obtidos os resultados.

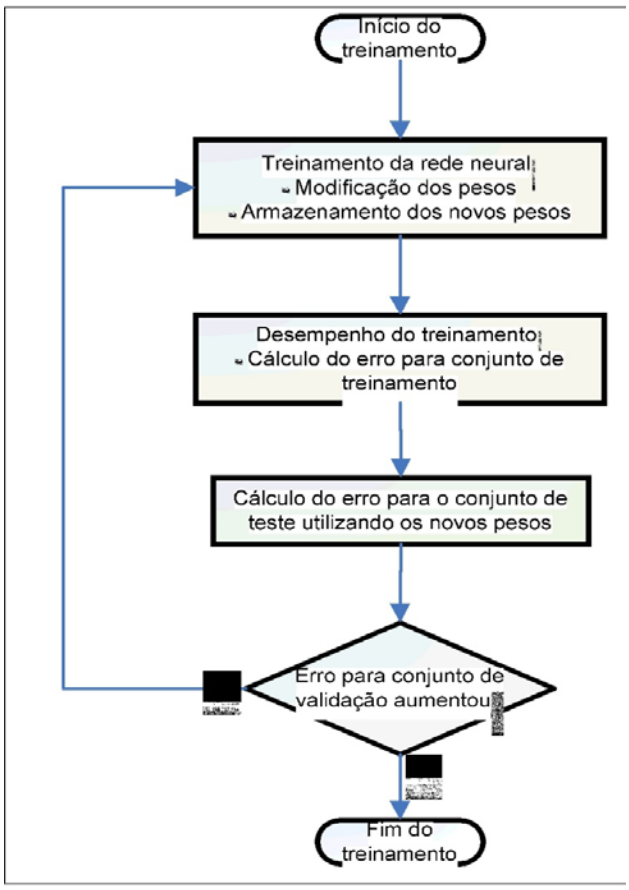

Figura 2 – Processo de Treinamento

5. ANÁLISE DOS RESULTADOS DOS TESTES

Com esses experimentos buscou-se obter um Sistema de Detecção de Intrusão, utilizando-se para isso, a capacidade de se detectar ataques novos ou desconhecidos. No entanto, para se obter um bom resultado, é necessário haver uma baixa taxa de falsos positivos e falsos negativos. Se um SDI tem uma ótima taxa de detecção de ataques, inclusive ataques novos, porém com uma taxa relativamente alta de falsos positivos, haverá problemas, pois os tráfegos normais estão sendo classificados como ataques. Sendo assim, o administrador terá alarmes para a maioria do tráfego que passa pela sua rede, sendo uma boa parte, falsos alarmes, provenientes do tráfego normal. Por outro lado, se um SDI tem baixa taxa de falso positivo e alta taxa de falso negativo, a rede monitorada pelo sistema estará vulnerável e o sistema não informará a maioria dos ataques sofridos.

Deste modo, para que a ferramenta seja confiável, espera-se que ela tenha a capacidade de detecção de novos ataques e que possua uma taxa baixa de falsos positivos. A partir dos treinamentos descritos anteriormente, obtiveram-se os resultados a seguir.

Ataques *User-To-Root*

Em relação ao primeiro experimento, cujo objetivo foi detectar ataques do tipo *User-To-Root*, obteve-se 100% de detecção e 0% de falso alarme, ou seja, todos os padrões foram classificados corretamente.
Portanto, para este tipo de ataque, conseguiu-se o ideal, ou seja, nenhum falso negativo e nenhum falso positivo, o melhor resultado de detecção de intrusão em relação a todos os outros tipos de ataque dos experimentos realizados. A rede neural foi capaz de identificar os novos ataques, bem como diferenciá-los do tráfego normal, atingindo um excelente resultado.

	User-To-Root	Tráfego Normal
User-To-Root	228	0
Tráfego Normal	0	228
%	100	100

Tabela 2 – Matriz Confusão *User-to-Root*

Ataques *remote-to-local*

No segundo experimento, voltado para ataques *Remote-To-Local,* foram obtidos 11 falsos negativos e 23 falsos positivos, ou seja, 11 ataques foram classificados como tráfego normal e 23 padrões de tráfego normal foram classificados como ataques.
Para esse tipo de ataque, a rede neural teve um bom desempenho, com uma taxa de detecção acima de 95% e obtendo-se apenas 6,15% de falso positivo. Deste modo, a rede poderia ser utilizada na detecção dessa intrusão, pois apresentou uma baixa incidência de falsos positivos e não seria significativa na triagem de alarmes do administrador da rede.

	Remote-To-Local	Tráfego Normal
Remote-To-Local	325	23
Tráfego Normal	11	313
%	96,73	93,15

Tabela 3 – Matriz Confusão *Remote-to-Local*

Ataques *Probes*
O terceiro experimento apresentou 100% na detecção de ataques, porém, 61 padrões normais foram classificados como ataques ou falsos positivos. A rede teve um bom desempenho na detecção dos ataques, apesar apresentar a maior taxa de falsos positivos entre todos os ataques avaliados, ou seja, 7,75%. Mesmo assim, a rede neural pode ser utilizada para essa detecção, já que de todos os ataques detectados, apenas alguns padrões de tráfego normal foram classificados como ataque.

	Probes	Tráfego Normal
Probes	1294	61
Tráfego Normal	0	1223
%	100	92,95

Tabela 4 – Matriz Confusão Probe

Ataques negação de serviço (*Denial of Service* – DOS)
No último experimento, voltado para ataques de DOS, obteve-se 17 falsos negativos e 26 falsos positivos. A rede teve uma boa taxa de detecção, acima de 97%, e uma taxa de falso positivo aceitável, 5,56%. Logo, também seria possível a utilização da rede na detecção desse tipo de intrusão, o que auxiliaria no combate aos ataques de negação de serviço. Apesar de apresentar falsos positivos, estes são de pequena grandeza, e não seriam significativos na verificação da veracidade dos alarmes gerados.

	DoS	Tráfego Normal
DoS	1294	61
Tráfego Normal	0	1223
%	100	92,95

Tabela 5 – Matriz Confusão DOS

Comparação dos resultados
A fim de avaliar o desempenho de detecções de intrusões com as redes neurais Multi-layer Perceptron, realizou-se um comparativo entre outros tipos de implementações voltados a detecção de ataques, como se pode ver na figura 3. Foram realizaram estudos de outros tipos de tecnologia em relação aos mesmos ataques analisados neste trabalho, utilizando para isso a mesma base de dados do DARPA. Liu e Chen utilizaram o Supporting Vector Machines (SVM) [6], Laskov e Dussel o Regression Tree (R.T.) [7], Zhang e Jiang avaliaram o Self Organized Maps (SOM) [8], e Cabrera e Mehra realizaram análises com o Multivariate Adaptative Regression Splines (MARS) [9]. Pode-se constatar que o tipo de rede neural utilizado neste estudo (MLP), teve um desempenho muito próximo comparado aos outros tipos de implementações testados, e no ataque casos User-to-Root teve o melhor desempenho entre todos, atingindo 100%.

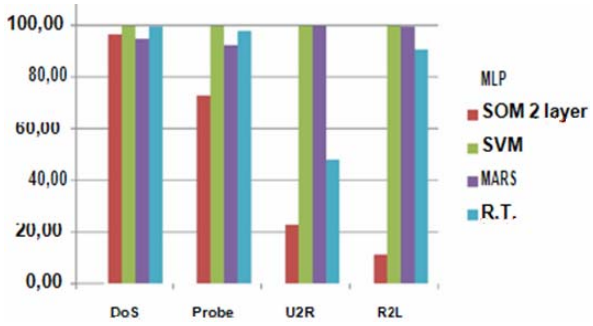

Figura 3 – Gráfico Comparativo de Desempenho

6. CONCLUSÃO

Sistemas de detecção de intrusões baseados em regras, exigem a atualização de suas bases de dados, a fim de que possam detectar qualquer tentativa de invasão na rede. Caso sua base de dados não esteja atualizada, o sistema não será capaz de detectar ataques e sua segurança estará comprometida
Com o intuito de não depender ostensivamente das atualizações em sua base de dados, utilizam-se redes neurais artificiais na detecção de possíveis ataques, valendo-se para isso da capacidade de generalização e aprendizado dessas redes em relação ao reconhecimento de padrões. Essas redes reconhecem ataques desconhecidos baseando-se no reconhecimento de padrões de ataques semelhantes.
Através da análise dos experimentos realizados, verificou-se que as redes neurais MLP apresentaram excelentes resultados na detecção de intrusão, pois identificaram tanto ataques conhecidos quanto desconhecidos. A análise de dados apresentada neste artigo, permitiu observar um baixo índice de falsos negativos e falsos positivos, eliminando assim desperdício de tempo e de produtividade analisando alarmes falsos nos sistemas. Também observou-se que na detecção do ataque do tipo *User-to-Root,* a rede teve o melhor desempenho, pois detectou todos os ataques e tráfegos normais corretamente. Outra rede que apresentou 100% de acertos foi a que detectou o tipo de ataque *Probe*, embora tenha apresentado uma taxa de 92,25% de acerto para tráfego normal. As análises permitiram constatar ótimos resultados com a utilização desse tipo de rede neural artificial.

7. REFERÊNCIAS BIBLIOGRÁFICAS

[1] Fausett, L. *Fundamentals of neural networks: architectures, algorithms, and applications*. Primeira edição, Editora Prentice-Hall, New Jersey, 1994.
[2] Haykin, S. *Redes neurais princípios e práticas.* Segunda Edição, Editora Bookman, Porto Alegre, 2001.
[3] Russel, S; Norving, P. *Inteligência artificial*. Editora Campus/Elsevier, Rio de Janeiro, 2004.

[4] Darpa. *Defence Advanced Research Projects Agency*. Disponível em: <http://www.darpa.mil/> Data de acesso: 18 de Mai. 2010.

[5] Netto, R. *Detecção de Intrusão Utilizando Redes Neurais Artificiais no Reconhecimento de Padrões de Ataque*. Tese de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Itajubá, Itajubá, Minas Gerais, 2006.

[6] Liu, F; Chen, Z. *Intrusion Detection Based on Multi-Layer Minimax Probability Machine Classifier.* Tese de mestrado, Shangai Institute of Technology, 2004.

[7] Laskov, P; DusseL, P; Schafer, C; Rieck, K. *Learning Intrusion Detection: supervised or unsupervised.* Fraunhofer, 2008.

[8] Zhang, C; Jiang, J; Kamel, M. *Comparison of BPL and RBF Network in Intrusion Detection System*. Tese de Doutorado, Waterloo, Canadá, 2007.

[9] Cabrera, J. B. D; Mehra, R.K. *Control and Estimation Methods in Information Assurance.* Conference on Decision and Control, Nevada, 2002.