Uma Solução Híbrida para Detecção de Anomalias em Redes

Lília de Sá Silva
DSS-LAC/INPE
lilia@dss.inpe.br

Antonio Montes
CENPRA-LAC/INPE
montes@lac.inpe.br

José Demisio Simões da Silva
LAC/INPE
demisio@lac.inpe.br

Resumo

Atualmente, observa-se um crescente interesse pelo desenvolvimento de sistemas de detecção de ataques por anomalia. Anomalias em redes podem ser descobertas pela observação de informações do host, tais como: acesso a arquivos, quantidade de system calls, dados de processos e de sistema, e também a partir de dados coletados do tráfego de rede, tais como endereços IP de origem e destino.Várias técnicas são utilizadas na construção destes sistemas, incluindo modelos estatísticos, regras de aprendizagem, abordagens de sistema imunológico artificial, redes neurais, e lógica fuzzy para identificar anomalias. São poucos os grupos que têm desenvolvido sistemas de detecção de intrusão por anomalia no tráfego de rede. Além disto, apesar de apresentarem resultados razoáveis, estes sistemas não são suficientemente bons para uso em ambiente de produção. Neste trabalho, é proposta uma metodologia híbrida, baseada na combinação de técnicas de redes neurais e regras fuzzy, que pretende-se utilizar no desenvolvimento de uma aplicação de detecção de anomalias no tráfego de rede.

Palavras-chave: *detecção de ataques, detecção por anomalia, ids, redes neurais, lógica fuzzy*

1. Introdução

Com o crescente volume de informações privilegiadas e sensíveis armazenadas e manipuladas em tempo real tanto na Internet como nas redes internas das organizações, manter a segurança dos sistemas e dados nestas redes é extremamente importante. Uma camada adicional de proteção aos recursos computacionais, relatando ou prevenindo contra ações maliciosas e acessos não autorizados, é composta pelas tecnologias de detecção de intrusão.

A abordagem de detecção de intrusão por anomalia tem sido intensivamente estudada desde a sua origem em 1987 e diferentes técnicas de aprendizagem têm sido aplicadas para solucionar problemas neste campo. A primeira técnica utilizada, segundo [13], foram os métodos estatísticos de detecção de *outliers* baseados em modelos probabilísticos. Além desta, muitas outras técnicas em data mining, consideradas o estado da arte, tais como as arvores de decisão, regras de associação, clusterização crisp, algoritmos de k-ésima vizinhança mais próxima e redes neurais [12] [17] [1], são encontradas na literatura. Técnicas de reconhecimento de padrões baseadas em aproximação estatística e *data mining* (mineração de dados), utilizam uma base de conhecimento previamente carregada e técnicas de inteligência artificial para explorá-la. Mas, em geral, a maioria destes métodos apresentam duas importantes questões [13]. A primeira, refere-se aos modelos de treinamento que devem ser construídos a partir dos dados rotulados. O segundo, é a capacidade de reconhecer apenas os atributos primitivos do ambiente monitorado. Atributos são dados que caracterizam o ambiente e podem ser classificados como primitivos e derivados. Atributos primitivos são dados básicos coletados, tais como endereço IP do host de origem, porta de destino ou *timestamp* (horário de captura do pacote), e atributos derivados são informações processadas a partir dos atributos primitivos, tal como o numero de conexões provenientes de um IP de origem para uma mesma porta de destino nos últimos N segundos.

Anomalias na rede podem ser detectadas a partir de informações do host, como acesso a arquivos; numero de *system calls*; dados de processos do kernel do sistema, compreendendo o número de processos filhos cancelados por processos-pai e quantidade do tempo gasto pelo processo nos modos usuário e sistema; dados do sistema, tais como uso de memória e carga de CPU; e dados coletados do tráfego de rede, envolvendo informações estatísticas relacionadas às camadas TCP, IP, ICMP, UDP da pilha de protocolos de rede, como, por exemplo, taxas de alteração do numero de pacotes recebidos, numero de pacotes recebidos com erro e numero de pacotes descartados.

Neste trabalho será dado enfoque à análise de anomalias no conjunto de dados coletado do tráfego de rede.

2. Detecção por Anomalias

Há um crescente interesse pela aplicação de algoritmos de aprendizagem de máquina e *data mining*, e anormal das atividades do sistema. A abordagem baseada em modelos de comportamento anormal é denominada *misuse detection* (detecção por abuso) e seu objetivo é comparar eventos do sistema em tempo real a cenários de ataques generalizados, ou seja, a modelos de detecção de abusos. A abordagem baseada em modelos de comportamento normal é denominada *anomaly detection* (detecção por anomalia). Diferente da detecção por abuso, a detecção por anomalia compara os eventos correntes do sistema ou do trafego de rede a perfis de atividades normais, ou seja, aos modelos de detecção por anomalia. Nesta abordagem, as anomalias são consideradas como possíveis intrusões [13].

A abordagem de detecção por anomalia apresenta como primeira grande vantagem a capacidade de detectar novos tipos de ataques, visto que quaisquer novos ataques diferem de comportamento normal.

Dentre as vantagens desta técnica destacam-se:

- Possibilidade de detectar ataques desconhecidos e complexos, não dependendo de uma base que armazene todos os ataques e vulnerabilidades possíveis, visto que a construção dos modelos de comportamento normal não requer que os padrões de ataque sejam rotulados. O que implica na possibilidade de uso de técnicas não supervisionadas na construção dos modelos.
- Produção de informações para a geração de bases de assinaturas (padrões de ataques)
- Esforço de manutenção reduzido, sem a necessidade de atualizações frequentes da base
- Dependência de plataforma reduzida
- Detecção de abusos de privilégios facilitada

A detecção por anomalias também apresenta problemas. A principal limitação é a alta taxa de alarmes falsos que ela produz. Isto ocorre porque nem toda atividade não usual é ilegítima ou representa um ataque. Deste modo, comportamentos legítimos do sistema podem ser equivocadamente reconhecidos como anomalias, e assinalados como possíveis intrusões.

Outras desvantagens deste técnica compreendem:

- Dificuldade de configuração e ajuste do limiar do objeto considerado anômalo ou não
os quais treinam sobre os dados históricos e permitem construir modelos do comportamento normal
- Desempenho reduzido, principalmente devido à grande quantidade de cálculos complexos exigidos
- Dificuldade no tratamento de mudanças do comportamento, principalmente se as alterações forem sutis, podendo comprometer a eficiência da técnica (atacantes podem desviar-se lentamente do comportamento padrao do sistema – se descoberto – até que seu ataque seja considerado normal)

3. Estado da Arte em Detecção de Anomalias no Tráfego de Rede

Há duas principais categorias de técnicas de detecção por anomalia: métodos supervisionados e não supervisionados [4].

A maioria das pesquisas em detecção de anomalias supervisionada tem por objetivo construir algum tipo de modelo com base nos dados normais e então verificar o quanto os novos dados se enquadram neste modelo. Algumas abordagens empregam modelos de predição obtidos através do treinamento de arvores de decisão sobre dados normais [10], enquanto outros usam redes neurais para obter o modelo [7] ou modelos não-estacionários [11] para detectar novos ataques. Bárbara [2] usou estimadores pseudo-Bayes para ressaltar detecção de novos ataques enquanto reduzir a taxa de alarme falso o máximo possível. Uma técnica desenvolvida pelo SRI no sistema EMERALD [8] usa registros históricos como dados de treinamento normal. Então compara as distribuições de novos dados com as distribuições obtidas destes registros históricos e as diferenças entre as distribuições indica uma intrusão. Trabalhos recentes, tais como [19] e [6] estimam parâmetros de um modelo probabilístico sobre os dados normais e calculam o quanto novos dados se enquadram neste modelo.

Zhang et al [21] desenvolveram o HIDE (Hierarchical Network Intrusion Detection System), um sistema de que utiliza processadores neurais que são treinados para analisar vários parâmetros do tráfego de rede simultaneamente. Este sistema é uma aplicação hierárquica distribuída, que consiste de várias camadas, cada uma contendo uma agente de detecção de intrusão (IDA – Intrusion Detection Agent) responsável pela monitoração das atividades de um host ou de uma subrede. Os IDAs de todos os níveis tem os componentes descritos a seguir:

- *Probe*: coleta o tráfego de rede de um host ou da rede, abstrai o tráfego em um conjunto de variáveis estatísticas para refletir o status da rede e periodicamente gera notificações ao *event preprocessor*.
- *Event Preprocessor*: recebe notificações de ambos *probe* e IDAs de níveis mais baixos, e converte a informação em um formato requerido pelo modelo estatístico
- *Statistical Processor*: mantém um modelo de referência das atividades típicas de rede, compara os relatórios do *event preprocessor* aos modelos de referência, e forma um vetor de estímulo para apresentar à rede neural classificadora.
- *Neural Network Classifier*: analisa o vetor de estímulo do modelo estatístico para decidir se o tráfego da rede é normal ou não.
- *Post Processor*: gera notificações aos agentes dos níveis mais altos. Ao mesmo tempo, pode mostrar os resultados através de uma interface com o usuário.

Neste trabalho foi desenvolvido um algoritmo de similaridade que considera o número de ocorrências durante uma janela de tempo e um algoritmo para atualização do modelo de referência em tempo real, conforme descrito em [21].

Para classificar padrões e identificar possíveis intrusões, cinco redes neurais com aprendizagem supervisionada foram testadas: uma *perceptron*, uma PHB (*perceptron-backpropagation-hybrid*), que é a superposição de um perceptron e uma pequena rede backpropagation, backpropagation (BP), *radial-based function* (RBF) e Fuzzy-ARTMAP. Os resultados experimentais mostraram que as redes BP e PBH foram capazes de produzir classificação mais forte.

A detecção de anomalia não supervisionada, por outro lado, visa detectar comportamentos anômalos sem usar qualquer conhecimento dos dados de treinamento. Abordagens de detecção por anomalia não supervisionada são baseadas em técnicas estatísticas [16] [18] [20], *clustering* [9], esquemas de detecção de *outlier* [14], maquinas de estado, entre outros. Neste contexto, é muito usado o termo *outlier* [8] que equivale a um d*ata point* (ou ponto de dado) que é muito diferente dos demais dados quando analisado segundo uma métrica pré-definida.

Alguns algoritmos de detecção de *outliers* recentemente estudados pelo grupo de desenvolvimento do MINDS (Minnesota Intrusion Detection System) são baseados no processamento de distâncias dimensionais completas dos pontos (de um a outro) bem como no processamento das densidades das vizinhanças locais [5]. O modulo de detecção de anomalias do MINDS utiliza a técnica LOF (Local Outlier Factor) para identificar anomalias no trafego de rede. Em maio de 2003, o grupo MINDS realizou um estudo comparativo das seguintes técnicas de detecção de *outliers:* nearest neighbor (NN), abordagem baseada em distancia de Mahalanobis, esquema local outlier factor (LOF) e SVM (maquina de vetor de suporte) não supervisionada para detectar anomalias na rede de produção da Universidade de Minnesota. Com este trabalho, obtiveram resultados satisfatórios na detecção automática de várias intrusões novas que não poderiam ser identificadas pelo SNORT e constataram ser a técnica LOF mais robusta que as técnicas NN (Nearest Neighbor) e SVM (Support Vector Machine) não-supervisionada na detecção de ataques usando conexões únicas e múltiplas. A abordagem baseada em distância de *Mahalanobis* apresentou baixa performance (provavelmente devido à multimodalidade dos dados) e concluíram que LOF funciona bem para dados de rede reais.

A técnica de detecção de intrusão não-supervisionada foi utilizada também no trabalho de Silva et al [15] em 2003, cujo processamento é apresentado na figura 3.1.

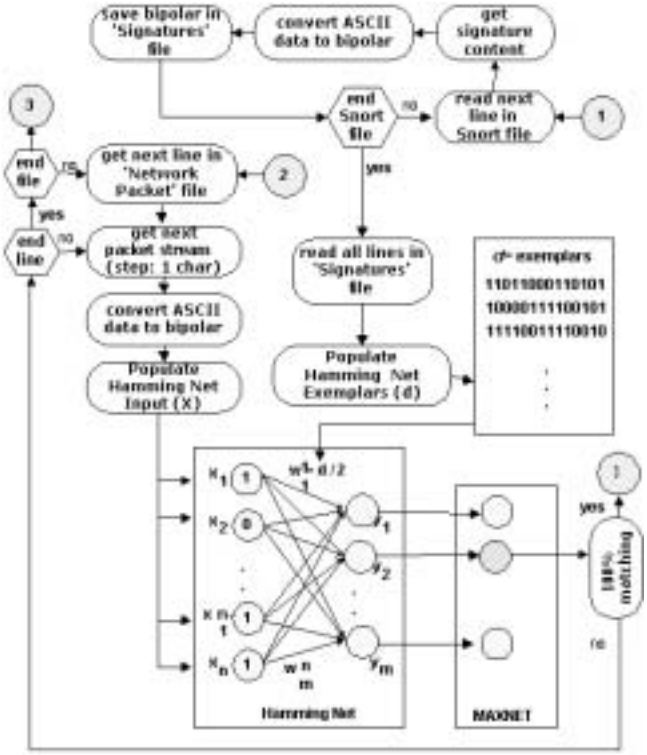

Figura 3.1 – Etapas de processamento da Hamming IDS

Embora esta aplicação não realize detecção de anomalias no tráfego de rede, e sim deteccao de assinaturas de ataque no conteúdo de pacotes de rede; é um exemplo de uso da rede neural *Hamming Net*, de

rápida classificação, que pode vir a ser utilizado na implementação da metodologia híbrida proposta neste paper. A aplicação desenvolvida mostrou ser uma técnica viável na identificação de padrões anormais em testes *off-line*. Dificuldades foram encontradas na modelagem adequada dos dados de interesse para entrada e processamento, segundo a arquitetura da rede neural, e na interpretação dos dados de carga útil dos pacotes de rede. Expandir o número de assinaturas no conjunto de exemplares da rede; refinar a codificação do programa; realizar testes com dados reais do pacote da rede e implementar a aplicação em tempo real são atividades a serem exploradas neste trabalho.

4. Solução Híbrida Proposta

No intuito de prover uma contribuição significativa no campo de detecção de intrusão, uma aplicação para detecção de anomalias no trafego de rede será desenvolvida, através da combinação de técnicas de redes neurais e regras *fuzzy*. A arquitetura da aplicação será composta basicamente dos seis módulos apresentados na figura 4.1 e descritos nas subseções seguintes.

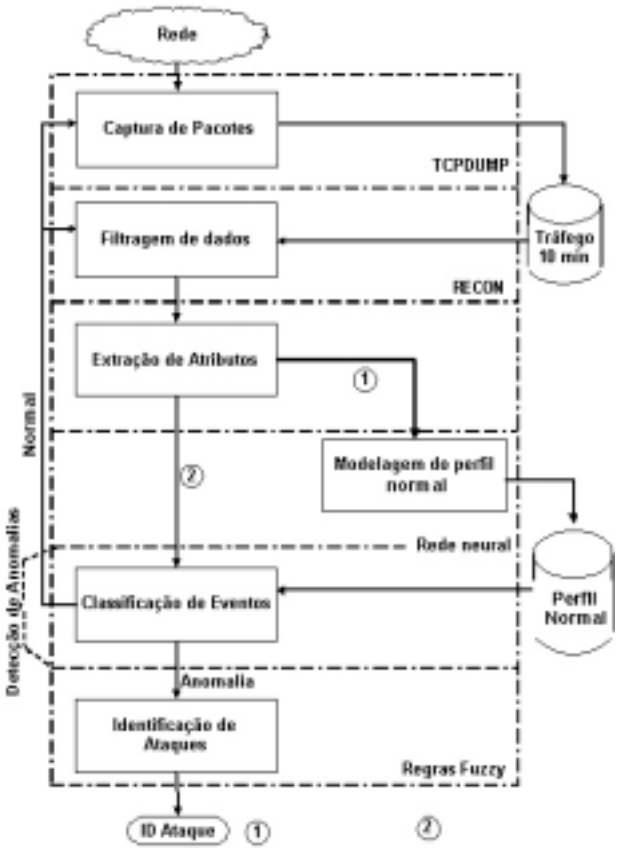

Figura 4.1 – Arquitetura da aplicação

O módulo de "Captura de Pacotes" tem por finalidade coletar os dados do tráfego de rede e reconstruir as sessões TCP/IP. Conjuntos de dados históricos de rede e dados de produção serão coletados para as etapas de processamento distintas: modelagem do perfil da rede (1) e classificação do tráfego de produção (2), como mostrado na figura 4.1.

Para a captura dos pacotes de rede, a reconstrução das sessões TCP-IP e a filtragem dos dados será utilizado o sistema RECON, desenvolvido por Chaves [3] em 2002. Esta aplicação utiliza a biblioteca *libpcap* para capturar dados do cabeçalho dos pacotes TCP, UDP e ICMP (em nível de Transporte), IP (em nível de Rede) e Ethernet (em nível de Enlace). O trafego de rede é representado pelo RECON em forma de sessões TCP-IP, representadas pela seqüência de pacotes caracterizando a troca de informações entre dois endereços IP e qute tenha inicio, meio e fim [3].

O módulo de "Filtragem de dados" tem a finalidade dividir o tráfego de uma janela de observação em sessões (TCP, UDP e ICMP) e em aplicacões (HTTP, FTP, SMTP, dentre outros).

A finalidade do módulo "Extração de Atributos" é selecionar os atributos primitivos de interesse relacionados ao tráfego analisado (endereços IPs de hosts e número de portas, por exemplo) e, a partir destes atributos, construir atributos derivados (por exemplo, numero de pacotes fluindo de tal origem para tal destino). Outra função deste modulo é converter os dados para um formato que possa ser reconhecido pela rede neural do modulo de "Modelagem de Perfil Normal".

Os atributos primitivos a serem extraídos do cabeçalho dos pacotes são: endereços IP de origem e de destino, portas de origem e de destino, tipo de protocolo ou serviço (IP, ICMP, TCP, UDP), duração da conexão, flag e número de bytes por pacote.
Os atributos derivados correspondem às características de conexões TCP individuais.

O módulo "Modelagem do Perfil Normal"será implementado como um processador inteligente (rede neural) capaz de aprender o comportamento normal do tráfego histórico da rede e armazenar este comportamento em uma base denominada "Perfil Normal" para uso posterior, quando o tráfego da rede de produção for analisado.

Para a construção deste módulo, serão analisados dados de acesso a aplicações de rede armazenados em arquivos de logs de servidores do INPE. Como a quantidade de dados disponível é muito grande, a estratégia é coletar amostras de registros de conexões normais. Os conjuntos de amostras de dados normais serão apresentados à rede e os pesos da rede reajustados até que a saída desejada seja obtida.

Dados de treinamento (dados históricos do tráfego em uma janela de tempo) serão coletados pelo módulo de "Captura de Pacotes" em intervalos regulares. Em seguida, o módulo de "Filtragem de Dados" dividirá o conjunto de dados de cada janela em sessões (TCP, UDP, ICMP) e em aplicações (HTTP, FTP, SMTP, etc) por sessão.

A cada tráfego capturado o analista humano classificará o tráfego como normal ou anômalo. O tráfego "normal" será apresentado à rede neural para treinamento. Enquanto o tráfego "anômalo" será descartado. Para cada conjunto de dados do tráfego de aplicação "normal" apresentado à rede, é atribuído um valor 0 como saída desejada.

O ciclo de treinamento se repetirá para as demais janelas de tráfego "normal" e, em cada etapa, os pesos serão reajustados até a rede convergir. Em seguida, a capacidade de resposta da rede será avaliada, através dos dados do tráfego de produção a serem utilizados para os testes. Se a rede retornar resposta similar para estes novos casos apresentados, constatar-se-á seu grau de capacidade de generalização.

Tendo sido a rede neural aprovada nos testes, o módulo de "Classificação de Eventos" será utilizado para classificar um tráfego de produção como "normal" ou "anômalo", com base no perfil da rede previamente armazenado.

Finalmente, regras *fuzzy* serão utilizadas para inferir sobre a natureza, ou seja, para identificar o tipo de tráfego classificado como "anômalo" pela rede neural. Esta etapa final será realizada pelo módulo denominado "Identificação de Ataques".

5. Conclusão

Os sistemas de detecção de intrusão por anomalia no tráfego de rede atualmente desenvolvidos têm apresentado resultados razoáveis, porém, não estão suficientemente preparados para uso em ambiente de produção.

Neste trabalho, pretende-se dar uma nova contribuição nesta linha de pesquisa, através do desenvolvimento de uma aplicação que combine as vantagens das técnicas de redes neurais e regras fuzzy para identificar anomalias no tráfego de rede.

Alguns dos desafios a enfrentar são:

- Mapear o comportamento normal para as aplicações de rede
- Selecionar os atributos do tráfego para análise
- Definir o tipo e arquitetura da rede neural
- Definir e modelar os parâmetros de entrada e saída desejada para a rede neural
- Construir as regras fuzzy para diferentes cenários de intrusões
- Reduzir a taxa de alarmes falsos gerada

Serão exploradas, por um lado, a capacidade de aprender e generalizar da rede neural, classificando um novo tráfego com base no comportamento aprendido e, por outro, a capacidade de suavizar a separação abrupta entre normalidade e anormalidade e a identificação mais aproximada de diferentes tipos de ataques que a lógica fuzzy pode proporcionar, visando reduzir a taxa de alarmes falsos e prover maior flexibilidade e desempenho ao sistema.

6. Referências

[1] Ambwani, T., "Multi Class Support Vector Machine Implementation to Intrusion Detection", Proceedings of The International Joint Conference on Neural Networks, 2003, 3, 2300-2305.

[2] Barbara, D.; Wu, N.; Jajodia, S., "Detecting Novel Network Intrusions using Bayes Estimators", First SIAM Conference on Data Mining, Chicago, Il, 2001.

[3] Chaves, M.H.P.C., "Análise de Estado de Tráfego de Redes TCP/IP para Aplicação em Detecção de Intrusão", São Jose Dos Campos, INPE, 2002.

[4] Dokas, P.; Ertoz, L.; Kumar, V.; Lazarevic, A.; Srivastava, J.; Tan, P. "Data Mining for Network Intrusion Detection", Tutorial at the Pacific-Asia Conference on Knowledge Discovery in Databases, Seoul, 2003.

[5] Ertoz, L.; Eilertson, E.; Lazarevic, A.; Tan, P.; Dokas, P.; Srivastava, J.; Kumar, V., "Detection and Summarization of Novel Network Attacks Using Data Mining", disponível em: http://www.cs.umn.edu/ research/minds/papers/raid03.pdf, acesso em: jul 2004.

[6] Eskin, E.; Lee, W; S. J. Stolfo, "Modeling System Calls for Intrusion Detection with Dynamic Window Sizes", Proceedings of Darpa Information Survivability Conference and Exposition I1 (DISCEX I1), Anaheim, Califórnia, 2001.

[7] Ghosh, A.; Schwartzbard, A., "A Study in Using Neural Networks for Anomaly and Misuse Detection", Proceedings of the 8th Usenix Security Symposium, 1999.

[8] Javitz, H.S.; Valdes, A. "The Nides Statistical Component: Description and Justification", Technical Report, Computer Science Laboratory, SRI International, 1993.

[9] Lazarevic, A.; Ertoz, L.; Ozgur, A; Srivastava, J.; Kumar, V., "A Comparative Study Of Anomaly Detection Schemes In Network Intrusion Detection", Proceedings of Third SIAM Conference on Data Mining, San Francisco, May 2003.

[10] Lee, S.; Stolfo, J., "Data Mining Approaches for Intrusion Detection", Proceedings of the 1998 Usenix Security Symposium, 1998.

[11] Mahoney, M.; Chan, P., "Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks", Proceeding of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002.

[12] Mukkamala, S.; Janowski, G.; Sung, A. H., "Intrusion Detection Using Neural Networks and Support Vector Machines", Proceedings of the International Joint Conference on Neural Networks, 2002, 2, 1702-1707.

[13] Petrovskiy, M. A., "Fuzzy Kernel-Based Method for Real-Time Network Intrusion Detection, Innovative Internet Community Systems", Lecture Notes in Computer Science, 2003, 2877, 189-200.

[14] Ramaswamy, S.; Rastogi, R.; Shim, K., "Efficient Algorithms For Mining Outliers From Large Data Sets", Proceedings of the ACM Sigmod Conference, 2000.

[15] Silva, L. S.; Santos, A. C. F.; Silva, J. D. S; Montes A., "A Neural Network Application for Attack Detection in Computer Networks", International Joint Conference on Neural Networks, Budapeste , Hungria, 2004.

[16] Staniford, S.; Hoagland, J. ; Mcalerney, J. , "Practical Automated Detection of Stealthy Portscans", Journal of Computer Security, 2002, 10, 1-2, 105-136.

[17] Sung, A.H.; Mukkamala, S. "A Comparative Study of Techniques for Intrusion Detection", Proceedings of the International Conference on Tools with Artificial Intelligence, 2003, 570-577.

[18] Yamanishi, K.; Takeuchi, J.; Williams, G.; Milne, P., "On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms", Proceedings Of The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, Ma, 2000, 320- 324.

[19] Ye, N. "A Markov Chain Model of Temporal Behavior for Anomaly Detection", Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, 2000.

[20] Ye, N. ; Chen, Q., "An Anomaly Detection Technique Based On A Chi-Square Statistic For Detecting Intrusions Into Information Systems, Quality And Reliability Engineering International", 2001, 17(2), 105 - 112.

[21] Zhang, Z.; Li, J.; Manikopoulos, C.N.; Jorgenson, J.; Ucles, J. "Hide: A Hierarchical Network Intrusion Detection System Using Statistical Preprocessing ad Neural Network Classification", Proceedings of the 2001 IEEE, Workshop on Information Assurance and Security United States Military Academy, West Point, NY, 2001, 5-6 June.