DIBSet: um Detector de Intrusão por Anomalias Baseado em Séries Temporais*

Roben C. Lunardi[1], Bruno L. Dalmazo[2], Erico M. H. do Amaral[2], Raul C. Nunes[2]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]Departamento de Eletrônica e Computação – Centro de Tecnologia – Universidade Federal de Santa Maria (UFSM) – Santa Maria - RS – Brazil

rclunardi@inf.ufrgs.br, {dalmazo,erico,ceretta}@inf.ufsm.br

Abstract. *Intrusion detection systems are used to detect attacks and malicious or inadequate use of the network. These systems work by analyzing signatures or anomalies in network traffic. This paper present the DIBSet, a time series based intrusion detector, that explores time series forecasting methods to detect anomalies and related attacks. From a concept proof, we show the DIBSet is able to detect SYN and SMURF attacks.*

Resumo. *Sistemas de detecção de intrusão são utilizados para detectar ataques e uso malicioso ou inadequado da rede, podendo ser baseadas em análise de assinaturas de ataque ou de anomalias no padrão de tráfego. Este trabalho apresenta o DIBSet, um Detector de Intrusões Baseado em Séries Temporais, que explora métodos de previsão por análise de séries temporais na detecção de anomalias/ataques. Os resultados demonstram que o DIBSet consegue detectar ataques do tipo SYN e SMURF, utilizados como prova de conceito.*

1. Introdução

Sistemas de detecção de intrusão (SDI) visam detectar ataques e uso malicioso ou inadequado da rede, sendo sua exploração chave para o desenvolvimento e uso da tecnologia da informação. Um SDI contém três componentes fundamentais [Dwyer 2003]: fonte de informações, análise e resposta; e costuma implementar coletores de informações e oferecer boas interfaces para visualização e análise (humana) do comportamento dos hosts e/ou redes, possibilitando uma melhor gerência de redes e proteção das informações. Entretanto, há uma carência por soluções eficientes para análise [Pohlmann e Proest 2006], sendo o desafio escolher um método eficiente que identifique uma intrusão, possibilitando a sinalização (alarme) e/ou a atuação (reconfiguração de dispositivos e/ou *software*) de maneira correta (verdadeiro positivo), sem gerar um número excessivo de falsas detecções (falsos positivos).

Este trabalho explora métodos de previsão por análise de séries temporais [Bowerman e O'Connel 1993] na detecção de anomalias/ataques e apresenta o DIBSet, um Detector de Intrusões Baseado em Séries Temporais desenvolvido e testado no

* Trabalho apoiado pelo Convênio UFSM/INPE e pela FAPERGS (Proc. 07503726).

trabalho de graduação de Lunardi (2008). Reutilizando bibliotecas de preditores previamente desenvolvidas no grupo de pesquisa, o DIBSet modela o padrão de comportamento do tráfego da rede como uma série temporal, quando deste só é conhecido o montante de pacotes trafegados por tipo, e realiza a análise assumindo que comportamentos de picos ou vales de utilização de determinados pacotes indicam uma anomalia de comportamento, o que é um indício de ataque [Dwyer 2003].

A grande maioria das vulnerabilidades dos computadores podem ser exploradas de diversas formas. Um ataque pode explorar uma única e específica vulnerabilidade, várias vulnerabilidades em simultâneo, um erro de configuração num componente de sistema, ou mesmo um *backdoor* criado por um ataque anterior. Por isto, para desenvolver este trabalho e realizar sua prova de conceito, utilizou-se alguns ataques conhecidos e com grande exploração em artigos: o SYN Attack [Peng, Leckie e Ramamohanarao 2007] e o SMURF Attack [Kumar 2007]. Como resultado dos testes foi demonstrado que o DIBSet é capaz de detectar anomalias geradas por estes dois tipos de ataques, o que demonstra a capacidade do uso de séries temporais na fase de análise de sistemas de detecção de intrusão.

O artigo está assim organizado: a seção 2 descreve os dois tipos de ataques considerados; a seção 3 apresenta os conceitos básicos sobre séries temporais; a seção 4 apresenta a arquitetura e testes do DIBSet; e a seção 5 apresenta as conclusões finais.

2. Ataques a Serem Analisados

O conhecimento detalhado do ataque possibilita saber quais contadores de pacotes deverem ser analisados e como deve ser o comportamento de cada um em casos de ataques. A seguir cada um dos tipos de ataques considerados neste trabalho é descrito detalhadamente.

2.1. Syn Attack

O SYN Attack é um tipo de ataque de negação de serviço (DoS – *Deny-of-Service*) que explora vulnerabilidades do TCP e do protocolo IP [Peng, Leckie e Ramamohanarao 2007]. O ataque consiste na inundação de uma máquina por requisições TCP/SYN, fazendo com que ela não possa responder a outras requisições de conexão. O objetivo é manter a máquina alvo ocupada enquanto ocorre o ataque. A Figura 1 ilustra o *three-way handshake* utilizado na comunicação TCP explorado pelo SYN Attack e a Figura 2 ilustra o início do SYN Attack [Maselli, Deri e Suin 2003].

Para não sobrecarregar a máquina atacante, normalmente é realizada uma clonagem de IP com subseqüente alteração do endereço IP do emissor no cabeçalho do pacote enviado, para que outra máquina receba as respostas SYN/ACK. A Figura 3 ilustra o ataque usando redirecionamento de resposta, que faz com que tanto a máquina que recebe o SYN como a que recebe o SYN/ACK fiquem ocupadas e não possam responder a conexões de outras máquinas [Peng, Leckie e Ramamohanarao 2007].

Para detecção do ataque pode-se analisar somente picos de pacotes SYN ou analisar a relação de pacotes SYN com os outros tipos de pacotes utilizados em conexões TCP. Considerando somente pacotes SYN, pode-se modelar o padrão de tráfego como, por exemplo, uma série temporal, a fim de obter uma estimativa da quantidade de pacotes SYN por período de tempo. Se a quantidade de pacotes SYN

passar de certo limiar (*threshold*) em relação ao padrão de tráfego considerado, pode-se sinalizar um possível ataque (anomalia). Assume-se neste trabalho que toda anomalia detectada pode ser um possível ataque.

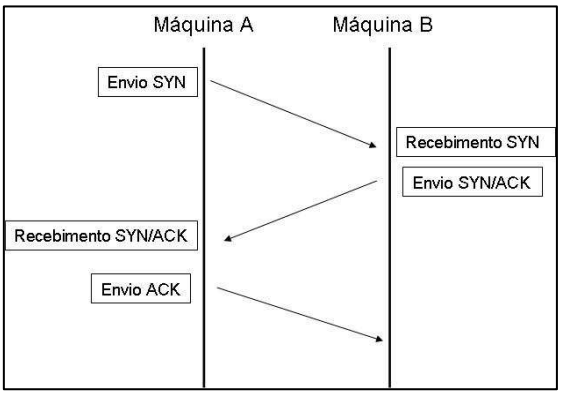

Figura 1. Início de comunicação TCP (*Three-way Handshake*) [Lunardi 2008].

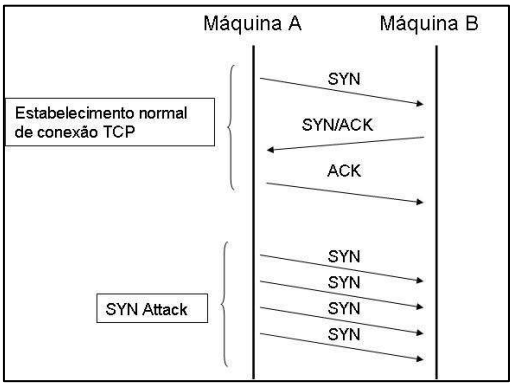

Figura 2. SYN ATTACK [Lunardi 2008].

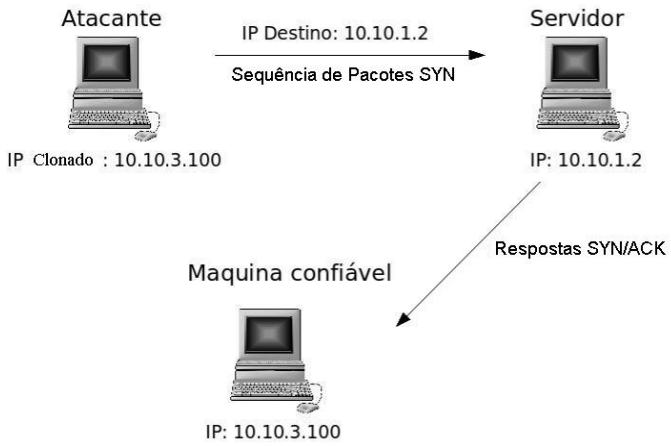

Figura 3. SYN Attack, com redirecionamento de resposta [Lunardi 2008].

Desta forma, considerando pacotes SYN relacionados com pacotes ACK (recebimento de pacote) ou FIN (finalização da comunicação), tem-se uma relação conhecida, que só é perdida caso haja um problema na comunicação, ou um ataque [Levchenko, Paturi, e Varghese 2004]. Embora esta abordagem caracterize uma

assinatura, e não tenha sido alvo de estudos neste trabalho, técnicas híbridas de detecção de intrusão podem explorá-la.

2.2. Smurf Attack

O Smurf Attack [Kumar, 2007] também é um tipo de ataque de negação de serviço e seu objetivo é inundar uma máquina alvo com pacotes ICMP Echo Reply. Para tal e explora vulnerabilidades do protocolo IP, o atacante forja o endereço IP da máquina alvo, e envia pacotes ICMP Echo Requests via broadcast, inundando a maquina alvo com pacotes ICMP Echo Reply [Maselli, Deri e Suin 2003]. A Figura 4 ilustra a execução do SMURF Attack. O atacante altera seu endereço IP para um endereço conhecido ativo na rede (10.10.3.22), o qual será o alvo efetivo do ataque, e dispara uma seqüência de *pings* (verificação de tempo de demora para chegar a um determinado host, se este estiver alcançável) por broadcast. Como resultado todas as máquinas ativas na rede respondem o ping para a máquina alvo (máquina que o atacante copiou o IP), causando uma inundação de respostas nesta máquina.

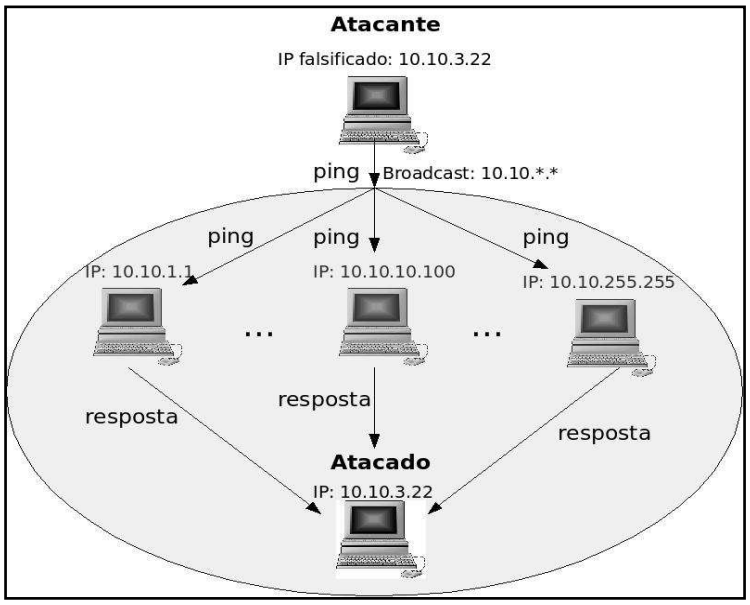

Figura 4. Smurf Attack [Lunardi 2008].

A detecção pode ser realizada analisando quando há um pico anômalo de pacotes ICMP Echo ou de mensagens broadcast. O ICMP Echo é um pacote que é utilizado pelos equipamentos de rede para verificar o estado dos links, logo são enviados periodicamente. A observação de pacotes ICMP Echo ou de mensagens de broadcast possibilita modelar seus comportamentos através de séries temporais. Desta forma, a análise pode indicar quando há extrapolação de limiares esperados, ou seja, uma anomalia resultante de um possível ataque.

3. Séries Temporais

Uma série temporal é um conjunto de observações de uma dada variável, ordenadas no tempo, geralmente em tempos eqüidistantes [Bowerman e O'Connel 1993]. Quando a variável é aleatória, como a contagem de tipos específicos de pacotes de rede, a série temporal pode ser definida como uma amostragem (com n observações) de um processo

estocástico, um fenômeno que varia de acordo com leis probabilísticas à medida que o tempo passa. Como a maioria das séries temporais não são estocásticas, ou seja, apresentam variações de nível e tendência, neste trabalho foi adotado o modelo de predição ARIMA (autoregressivo integrado e de médias móveis), descrito a seguir.

3.1. Modelo ARIMA

O preditor ARIMA fornece previsões modelando a série como um processo auto-regressivo integrado de médias móveis (ARIMA - *Autoregressive Integrated Moving Average process*), o qual incorpora os modelos: puramente autoregressivo de ordem p (AR(p)); puramente médias móveis de ordem q (MA(q)); autoregressivo e de médias móveis de ordem p e q (ARMA(p,q)); e naturalmente o autoregressivo integrado e de médias móveis de ordem p, d e q (ARIMA(p,d,q)), onde d representa a ordem de integração, ou seja, o número de diferenças necessárias para transformar a série temporal não estacionária em estacionária. A previsão do contador de pacotes c segundo o ARIMA é computada pelo Sistema de Predição de Recursos (RPS[2]), e obedece a seguinte equação de predição

$$\hat{c}_t = \frac{\theta(B)}{\phi(B)\nabla^d} a_t$$

onde: B é o operador de deslocamento para trás (*backward shift operator*), definido por $B.c_t = c_{t-1}$; ∇ é o operador de diferença para trás (*backward difference operator*), definido por $\nabla. c_t = c_t - c_{t-1} = (1-B).c_t$; a_t é o ruído no instante t; d é a ordem de não estacionariedade; e $\theta(B)$ e $\phi(B)$ são polinomiais de B com coeficientes q e p, respectivamente, onde q é o número de termos de médias móveis e p é o número de termos autoregressivos [Bowerman e O'Connel 1993].

Enquanto os preditores tradicionais realizam previsões baseados em equações não polinomiais, para obter uma boa previsão o algoritmo geral de previsão adotado pelo ARIMA necessita reavaliar a estrutura polinomial e seus parâmetros periodicamente.

Com o uso das séries temporais, mais especificadamente com o modelo ARIMA, pretende-se estabelecer um padrão de comportamento do tráfego de rede e avaliar a cada nova amostra se a série está se comportando como esperado.

4. DIBSeT

Esta seção descreve o DIBSet, um detector de intrusões por anomalia baseado em séries temporais. Inicialmente é apresentada a arquitetura e funcionamento do DIBSet (seção 4.1) para então ser apresentado os testes realizados e as discussões (seção 4.2).

4.1. Arquitetura do DIBSeT

O DIBSet é um sistema de detecção de intrusão que avalia o comportamento de contadores de pacotes. Como entrada, ele utiliza dados capturados pelo NTOP [Maselli,

---

[2] RPS – An Toolkit for Resource Prediction in Distributed System. Disponível em http://rps.cs.northwestern.edu, último acesso em Julho de 2008.

Deri e Suin 2003], os quais são convertidos em uma série temporal. A série é então utilizada para realizar previsões de acordo com um preditor ARIMA, as quais são utilizadas para analisar o comportamento das amostras de dados vindas do NTOP. Caso haja um desvio anormal, um alarme é gerado. Finalmente, visando o ajuste do sistema e testes simulados com alimentação do DIBSeT a partir de dados antigos, o DIBSet gera logs dos contadores e dos níveis de alarme gerados. A Figura 5 ilustra a arquitetura do DIBSet, incluindo as bibliotecas necessárias para que o preditor ARIMA funcione.

Cada tipo de contador (referente a pacotes TCP/SYN, ICMP, etc.) gera uma série temporal que é passado para classes que implementam o modelo ARIMA, as quais foram herdadas de trabalhos prévios do grupo de pesquisa. A partir dos dados processados pelo gerador de séries temporais, é computado uma margem de segurança, que quando vinculado ao valor previsto estabelece um limiar superior e inferior para o comportamento considerado normal. É importante salientar que esta margem é computada em função do erro de predição e resulta em limiares dinâmicos (*adaptive thresholds*), o que possibilita uma adaptação do DIBSet ao histórico dos dados sendo coletados. A utilização de limiares dinâmicos contribui para reduzir o número de falsos positivos.

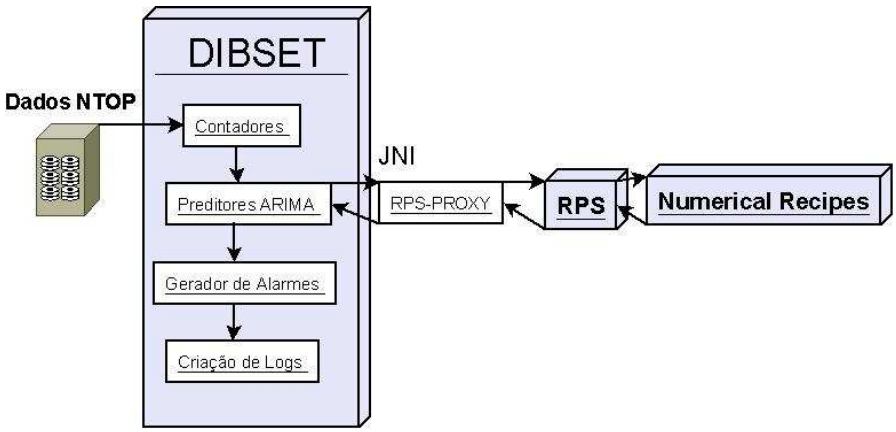

Figura 5 – Arquitetura do DIBSeT [Lunardi 2008].

A hipótese básica de funcionamento é que quando a quantidade de um dado tipo de pacote sob análise ultrapassar um limiar (*threshold*) estabelecido, um possível ataque pode estar ocorrendo (o DIBSet é um detector por anomalias). Assim, quando observado algum comportamento anômalo [Goodall 2006], o DIBSet indica o possível ataque através da classificação de alarmes em níveis. Os níveis de alarme geram dados mais detalhados para o software ou gerente que irá tomar a decisão. Os níveis são criados através de uma relação entre o contador atual (última amostra) e sua previsão (valor computado pelo preditor), considerando o limiar. Alarmes são gerados apenas quando o valor ultrapassa o limiar pré-estabelecido, isto é, só quando o valor do contador era maior que o valor previsto somado à margem de segurança. Quanto maior esta razão, maior o nível de alarme indicado. Neste trabalho foram considerados 5 níveis de alarme, sendo o valor zero correspondente a inexistência de anomalias. Deste modo pode-se ter uma idéia melhor da situação da rede, não apenas se houve uma anomalia, mas quão longe a anomalia se distanciou do limiar estabelecido (comportamento previsto).

Como as classes implementadas pelo grupo utilizavam funções e bibliotecas do programa RPS, foi necessária a instalação deste software. Para a instalação do programa RPS com as funções de Séries Temporais, é necessário também a instalação do programa de funções matemáticas Numerical Recipes[3]. Ambos programas foram de difícil instalação pois teve-se de utilizar versões antigas de bibliotecas e compiladores para linguagens como C++ e FORTRAN. Para que o RPS (programado em C++) e as classes de séries temporais (feitas em Java) pudessem trocar informações foi necessário a criação de um *proxy* utilizando JNI.

4.2. Teste e discussões

Os testes foram realizados em computadores PC e considerando uma amostragem do tráfego da Universidade Federal de Santa Maria durante 5 dias. Abaixo estão gráficos de uma parte das amostras coletadas. Na Figura 6 pode ser observado algumas variação do contador ICMP_echo e na Figura 7 os níveis de alarme resultantes. Pode ser visto que as anomalias geradas apresentam níveis de alarmes correspondentes aos picos de variação.

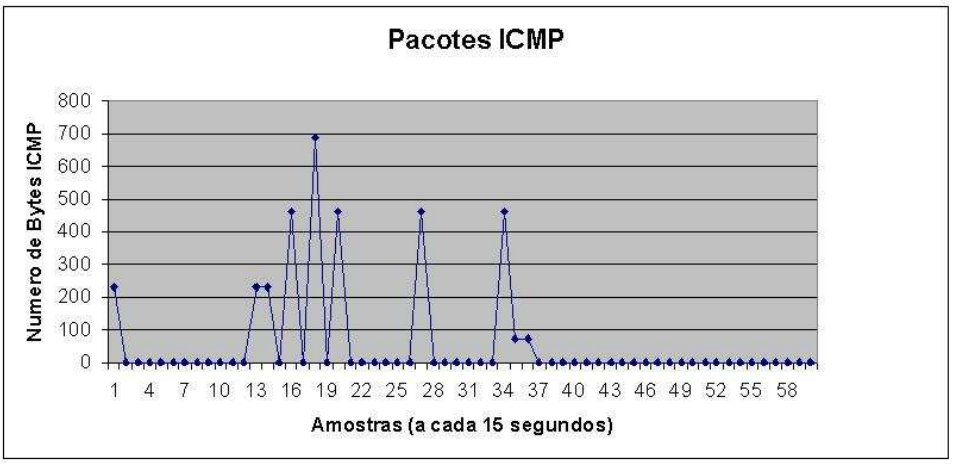

Figura 6 – Gráfico de Bytes ICMP.

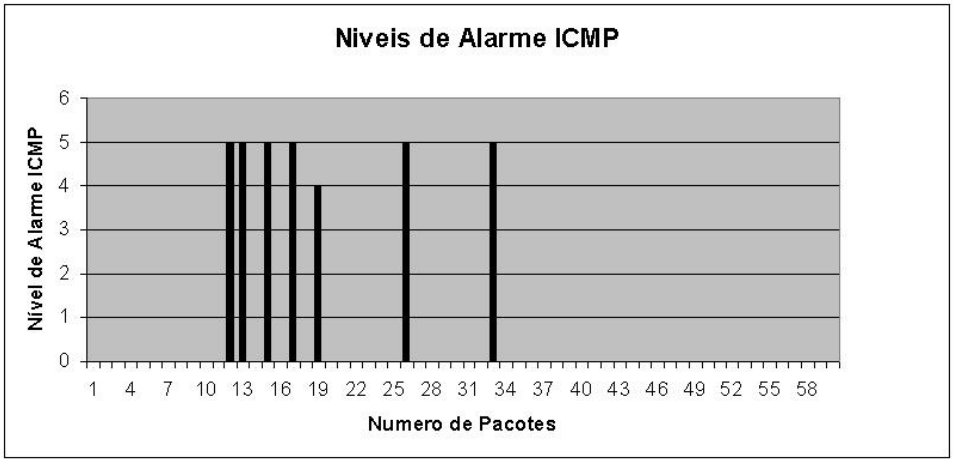

Figura 7 – Gráfico de Níveis de Alarme ICMP.

[3] Numerical Recipes. Disponível em http://www.nr.com, último acesso em Julho de 2008.

Na Figura 7, também pode ser observado que os níveis de alarmes muitas vezes são altos com quedas bruscas para zero. Este comportamento se deve ao fato da indisponibilidade, gerada pelo ataque, no servidor onde está instalada a sonda de captura de dados.

Com o objetivo de gerar dados mais robustos para o gerente ou software que tenha que tomar alguma decisão a partir dos níveis gerados, foram analisados também os dados que ficaram abaixo da faixa inferior, ou seja, além do contador ter de ficar abaixo do limiar superior, também precisa ficar acima do limiar inferior. Adotaram-se níveis negativos para identificar as amostras que ficaram abaixo do limiar inferior. O monitoramento de violações de limiares inferiores é importante para poder correlacionar duas variáveis sob observação. Num ataque enquanto alguns contadores podem ter picos positivos outros podem apresentar picos de redução (ficar abaixo do esperado), como por exemplo, pela indisponibilidade do sistema depois do ataque a algum contador específico ou pela troca da assinatura de ataque. Desta forma, todo contador que fica abaixo do esperado, pode ser resultado de um ataque e um alarme deve ser indicado para o administrador do sistema.

Nas Figuras 8 e 9 podemos verificar que após ser gerado um Smurf Attack, entre as amostras 177 e 188, os níveis de alarme invertem de sinal, isto é ficam abaixo do esperado pois o servidor que estava sendo analisado caiu.

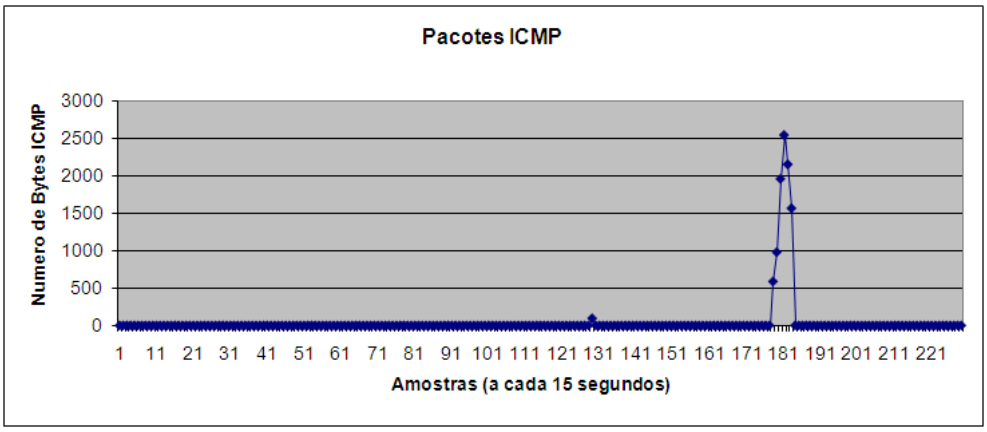

Figura 8 – Gráfico de pacotes ICMP.

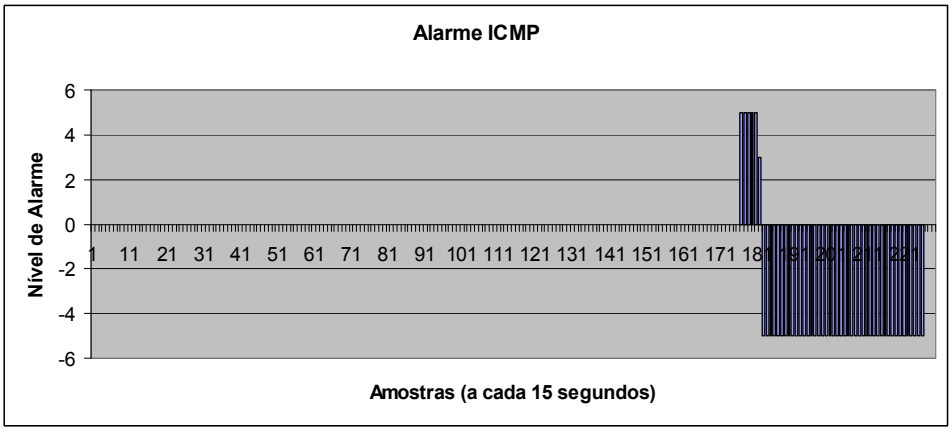

Figura 9 – Níveis de Alarme ICMP.

Nas Figuras 10 e 11 podemos verificar que após ser gerado um SYN Attack, entre as amostras 113 e 127, os níveis de alarme também invertem de sinal, isto é ficam abaixo do esperado, pois o servidor que estava sendo analisado caiu novamente.

Do ponto de vista estabilização do sistema uma observação importante é que nas primeiras horas de análise, a série temporal gera muitos resultados fora do esperado. Isto se deve ao fato de ser necessária a alimentação da série temporal para poder gerar um histórico de comportamento, logo o período de aprendizado deve ser considerado. Adicionalmente, os parâmetros de níveis de alarmes precisaram ser modificados de acordo com os parâmetros de tráfego da rede alvo, pois somente depois de diversos testes de ataques foi possível verificar valores condizentes com o esperado.

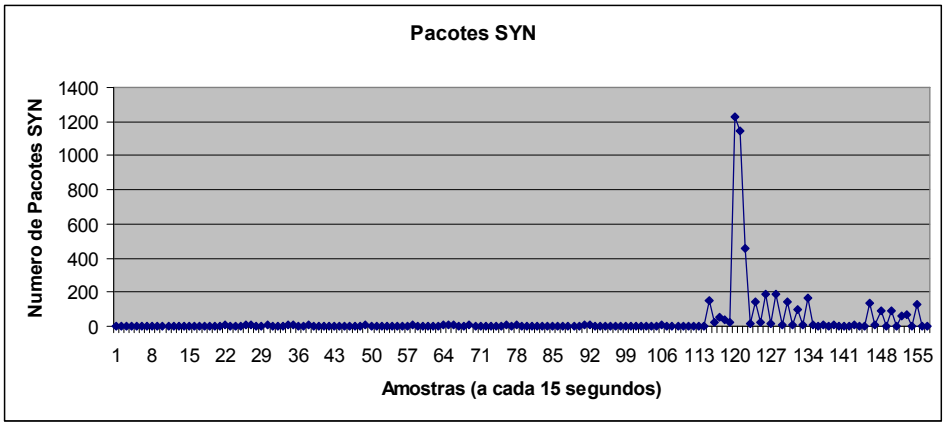

Figura 10 – Gráfico de pacotes SYN.

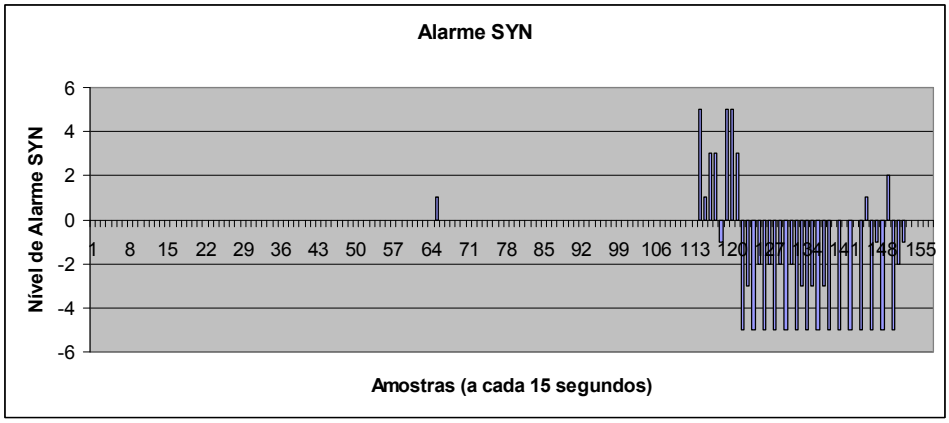

Figura 11 – Níveis de Alarme SYN.

5. Conclusões

Este trabalho explorou a previsão baseada em séries temporais na fase de análise de sistemas de detecção de intrusão por anomalias e apresentou o DIBSet, um sistema de detecção de intrusão baseado em séries temporais. Abordou-se no trabalho a caracterização de dois ataques de negação de serviço bem conhecidos, para que fosse possível testar a partir deles se o DIBSet realiza detecções corretamente analisando apenas contadores de pacotes. Como a escolha do modelo de previsão ARIMA,

parâmetros *p*, *d* e *q*, é muito importante para poder se chegar a resultados satisfatórios, o DIBSet baseou-se em preditores já desenvolvidos.

Os resultados indicam sucesso na detecção de ataques do tipo SYN e SMURF, e apontam para a necessidade de geração de alarmes para extrapolação de limiares superiores e inferiores. De maneira similar, observa-se que a detecção considerando limites possibilita a geração de alarmes em níveis, o que contribui para a tomada de decisão final. A adaptatividade do limiar ao comportamento do tráfego pode reduzir o número de falsos positivos. Em síntese, os testes demonstraram que pode-se aplicar séries temporais para detectar ataques baseados em inundação, e que para tal a abordagem é identificar anomalias no comportamento do tráfego.

Como continuação deste trabalho pretende-se utilizar os dados do IAS [Pohlmann e Proest 2006] para a obtenção dos contadores de pacotes de redes, dado a recente cooperação estabelecida entre a FHGe/Alemanha e a UFSM/Brasil. Após o refinamento do código de geração de alarmes, com o objetivo de gerar menos falsos positivos, pretende-se também expandir o número de contadores analisados, isto é, a capacidade de detectar um maior número de ataques; além de incluir a analise de dados de contadores de diferentes máquinas para obter um comparativo entre máquinas servidoras e máquinas de usuários.

Referências

Bowerman, Bruce L. and O'Connel, Richard T. (1993), Forecasting and Time Series: an Applied Approach. Belmont: Duxbury Press.

Dwyer, D. (2003) “Network Intrusion Detection.” 3rd Edition, New Riders Publishing.

Goodall, J. (2006) “Visualizing Network Traffic For Intrusion Detection” In: ACM Symposium on Designing Interactive Systems, pages 363-364.

Kumar, S. (2007) “Smurf-based Distributed Denial of Service (DDoS) Attack Amplification in Internet”. Second International Conference on Internet Monitoring and Protection (ICIMP IEEE 2007).

Levchenko, K., Paturi, R. e Varghese, G. (2004) “On the Difficulty of Scalably Detecting Network Attacks”. CCS-ACM.

Lunardi, R. (2008) “Um analisador de intrusões baseado em Séries Temporais”. Trabalho de Graduação n°255, Curso de Ciência da Computação, UFSM.

Maselli, G., Deri, L. e Suin, S. (2003) “Design and Implementation of an Anomaly Detection System: an Empirical Approach”. Proceedings of Terena Networking Conference (TNC 03), Zagreb, Croatia.

Peng, T., Leckie, C. e Ramamohanarao, K. (2007) “Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems”. In: ACM Computing Surveys, Vol. 39, No 1, Article 3.

Pohlmann, N. and Proest M. (2006) “Internet Early Warning System: The Global View”. In: Vieweg, Securing Eletronic Business Process, pages 377 – 386.