Em Busca de um Roteiro Experimental de Curta Duração para Avaliação de Sistemas de Detecção de Intrusão baseados em Rede

Leonardo Lemes Fagundes[1], **Luciano Paschoal Gaspary**[2]

Universidade do Vale do Rio dos Sinos (UNISINOS)
Centro de Ciências Exatas e Tecnológicas
[1]Instituto de Informática
[2]Programa Interdisciplinar de Pós-Graduação em Computação Aplicada (PIPCA)
Av. Unisinos 950 – 93022-000 – São Leopoldo, RS

{leonardo,paschoal}@exatas.unisinos.br

Abstract. *Intrusion Detection Systems (IDSs) have become an essential component to improve security in networked environments. The increasing set of available IDSs has stimulated research projects that investigate means to assess them to find out their strengths and limitations (in order to improve the IDSs themselves) and to assist the security manager in selecting the product that best suits specific requirements. Current approaches to do that (a) require the accomplishment of complex procedures that take too much time to be executed, (b) do not provide any systematic way of executing them, and (c) some require specific knowledge of IDSs internal structure to be applied. In this paper we address these limitations by proposing a script to evaluate network-based IDSs regarding their detection capability, scalability and false positive rate. Two Intrusion Detection Systems, Snort and Firestorm, have been assessed to validate our approach.*

Resumo. *Os sistemas de detecção de intrusão, Intrusion Detection Systems (IDSs), têm se tornado um componente essencial para aprimorar a segurança em ambientes interligados. O conjunto crescente de IDSs disponíveis tem estimulado projetos de pesquisa que investigam mecanismos para avaliá-los com o objetivo de identificar suas potencialidades e limitações (para proporcionar o aprimoramento dos próprios IDSs) e auxiliar gerentes de segurança na seleção do sistema que mais se adequa a requisitos específicos. As abordagens atualmente disponíveis para tal (a) requerem a realização de procedimentos complexos que consomem muito tempo para serem realizados, (b) não possuem nenhuma forma de sistematização, e (c) algumas requerem conhecimentos específicos da estrutura interna dos IDSs. Neste artigo nós tratamos essas limitações ao propor um roteiro para avaliar IDSs baseados em rede quanto a sua capacidade de detecção, escalabilidade e taxa de falsos positivos. Dois sistemas de detecção de intrusão, Snort e Firestorm, foram avaliados para validar a nossa abordagem.*

1 Introdução

Com o uso em grande escala da Internet observa-se um considerável aumento nos tipos e na quantidade de ataques realizados contra as mais diversas categorias de organizações. Através da exploração dos diferentes tipos de vulnerabilidades como falhas de configuração, falhas de implementação e uso indevido de recursos disponíveis, surge um universo de ataques possíveis. Exemplos desse universo de ataques incluem desde

varreduras de portas, negação de serviços, seqüestro de conexões até ataques mais sofisticados, tais como negação de serviço distribuído, inserção e evasão. Com o objetivo de minimizar as chances de um intruso obter sucesso em suas atividades, diversos mecanismos de proteção são utilizados. Entre esses mecanismos destacam-se a criptografia, a certificação digital, a infraestrutura de chaves públicas, os *firewalls*, os protocolos de autenticação e, ainda, os sistemas de detecção de intrusão.

Sistemas de detecção de intrusão representam uma importante técnica de monitoração, cuja principal função é detectar ações maliciosas, tais como tentativas de ataques e obtenção de informações. Atualmente, após aproximadamente mais de uma década de pesquisas, existem disponíveis no mercado diversos sistemas de detecção de intrusão. Entre esses IDSs destacam-se o Snort [Roesch, 1999], o Bro [Paxson, 1999], o NFR (Network Flight Recorder) [NFR, 2001], o Firestorm [Firestorm, 2001] e o RealSecure [ISS, 1999]. Diante desse conjunto crescente de sistemas, a identificação de suas potencialidades e limitações frente a um conjunto de métricas é essencial, quer com a finalidade de estimular nichos específicos de pesquisa na área (para aprimoramento dos próprios IDSs), quer com o objetivo de auxiliar gerentes de segurança na árdua tarefa de selecionar o sistema mais adequado para o ambiente sob sua administração.

Diversas metodologias para avaliação de sistemas de detecção de intrusão têm sido propostas [Puketza et al., 1997, Lippmann et al., 1999, Alessandri, 2000, Barber, 2001]. Essas metodologias, entretanto, não possuem uma forma sistematizada para a execução dos procedimentos previstos, são constituídas de uma série de atividades exaustivas que são realizadas durante semanas e exigem que os usuários da metodologia possuam conhecimentos específicos, tais como a estrutura interna dos IDSs (o que não é possível no caso de IDSs proprietários). Além disso, essas metodologias são pouco documentas e, em alguns casos, os resultados obtidos através de determinados testes são questionáveis devido à forma com que tais experimentos foram conduzidos.

Este artigo apresenta uma abordagem alternativa para avaliação de IDSs baseados em rede que constitui-se em um roteiro que possui um conjunto de procedimentos sistematizados, realizáveis em um curto espaço de tempo e que não exige o conhecimento prévio das ferramentas de detecção a serem avaliadas. Se por um lado a avaliação resultante não é tão detalhada no que se refere às informações oferecidas, por outro o roteiro pode ser facilmente executado (sem muitas exigências de recursos humanos e materiais). Os testes previstos nesse roteiro avaliam as seguintes características: capacidade de detecção, escalabilidade e taxas de falsos positivos dos sistemas avaliados. Esses critérios foram escolhidos por representarem aspectos determinantes no processo decisório de uma organização ao ter que optar por um IDS. O artigo está organizado da seguinte forma: a seção 2 descreve o roteiro proposto. Os resultados obtidos com um estudo de caso realizado com os IDSs Snort e Firestorm são apresentados na seção 3. A seção 4 finaliza o artigo com algumas considerações finais e apresenta perspectivas para trabalhos futuros.

2 Roteiro de Avaliação

O roteiro é composto por cinco etapas: seleção dos ataques, seleção de ferramentas, geração do tráfego dos cenários de avaliação, montagem do ambiente de avaliação e análise dos IDSs. A seguir é descrita cada uma das etapas citadas.

2.1 Seleção dos Ataques

Nesta etapa o objetivo é selecionar um conjunto de ataques que explore características técnicas únicas entre si. Ao invés de simplesmente reunir um conjunto de ataques, o

que se busca, ao finalizar esta etapa, é selecionar ataques cuja detecção seja possível a partir de diferentes mecanismos existentes em um IDS. Por exemplo, para que um IDS seja capaz de detectar um ataque de inserção, o *URL Encoding*, ele necessita mais do que simplesmente a capacidade de análise de um pacote HTTP, pois é necessário, ainda, um mecanismo de decodificação do conteúdo do cabeçalho desse pacote. Já o processo de detecção de um ataque de negação de serviço, como o *teardrop*, requer um mecanismo capaz de remontar pacotes IP fragmentados. Dessa forma, os ataques selecionados nessa etapa possuem um conjunto de características ímpar que permite avaliar as diferentes capacidades de detecção dos IDSs e não simplesmente a base de assinaturas dessas ferramentas.

A figura 1 ilustra um quadro onde nas colunas são listadas as características exploradas pelo conjunto de ataques inicialmente cogitados para serem usados na avaliação dos IDSs[1]. Nas linhas são listados todos os ataques. Por limitação de espaço a figura apresenta apenas as varreduras de portas, mas ataques de evasão, inserção e negação de serviços também foram considerados. Os ataques a serem utilizados na avaliação são aqueles que exploram combinações de características diferentes (em negrito na figura). Para amenizar o problema decorrente de o IDS não ter a assinatura do ataque selecionado e, com isso, concluir-se erroneamente que o IDS não é capaz de detectar os ataques com as características do mesmo, nós sempre selecionamos dois ataques que exploram as mesmas características: um antigo e outro recente. Como pode ser observado, essa forma de seleção permite reduzir o cenário inicial de ataques. Nós assumimos, por exemplo, que se um IDS é capaz de detectar o ataque *Ident Reverso TCP*, ele também deve ser capaz de detectar *TCPConnect* (que explora as mesmas características); basta que esteja configurado com assinaturas adequadas para tal.

				HTTP			IP					TCP										ICMP		UDP
	Múltiplos pacotes	Não estabelece conexão(*Half-Open*)	Estabelece conexão	Linha de Requisição	Codificação da requisição	Tamanho da requisição	Off-set de fragmento	Pacote IP com DF bit habilitado	Identificação do fragmento	Pacote IP raw	FLAGS de controle da fragmentação	Opções (NOP, MSS, Window, Timestamp)	Campo: Tipo de serviço	Número de sequência	TCP Initial Window	Pacote com *flag* SYN	Pacote com *flag* FIN	Pacote com *flag* ACK	Pacote com *flag* URG	Pacote com *flag* PUSH	Pacote com todos os *flag*s desativados	Pacote ICMP	Tamanho da mensagem ICMP de Erro	Pacote UDP de zero bytes
TCP Connect	X		X																					
Syn Scan	X	X														X								
Ack Scan	X	X																X						
Window Scan	X	X																X						
Fin Scan	X	X															X							
UDP Scan	X	X																						X
Null Scan	X	X																			X			
Xmas	X	X															X	X	X	X				
TCP Ping	X	X																X						
TCP Fragmentation	X	X					X		X		X					X	X	X	X	X				
Varredura IP	X	X								X														
Fingerpriting	X	X						X				X	X	X	X		X	X					X	
Ident Reverso TCP	X		X															X						

Figura 1: Descrição técnica do cenário inicial de ataques proposto

2.2 Seleção de Ferramentas

Esta etapa é dedicada à obtenção de ferramentas que permitam reproduzir os ataques selecionados na etapa anterior. Essa atividade pode ser realizada em um curto espaço de

[1]Essas características foram identificadas pelo nosso grupo de pesquisa após estudar um conjunto de aproximadamente trinta ataques.

tempo devido à facilidade com que atualmente se pode encontrar e utilizar tais ferramentas. Por exemplo, para reproduzir as varreduras de portas destacadas na figura 1 poderia ser utilizada a ferramenta Nmap 2.54 (GNU/Linux).

2.3 Geração do Tráfego do Cenário de Avaliação

O cenário de avaliação é formado pelos ataques selecionados e pelo tráfego de fundo (necessário para a análise de escalabilidade). A seguir são descritas as formas propostas nesse roteiro para (a) armazenar o tráfego de ataque e (b) gerar o tráfego de fundo.

2.3.1 Coleta do tráfego de ataque

Para que não seja necessária a manipulação de cada uma das ferramentas de ataque cada vez que os diversos testes (apresentados na seção 2.5) tiverem que ser realizados, sugere-se a coleta e o armazenamento prévio do tráfego dos ataques. Para tal é preciso montar um ambiente como o ilustrado na figura 2a. Na estação *Atacante* estão instaladas todas as ferramentas de ataque selecionadas na etapa de seleção de ferramentas. Já na estação *Vítima* estão instalados todos os serviços a serem atacados. A estação *Sniffer*, por fim, possui como atribuição coletar o tráfego (usando uma ferramenta como o `tcpdump`) gerado pelas ferramentas de ataque e pela estação atacada (quando, de alguma forma, reage a eles). Assim, para cada ataque a ser coletado e armazenado, sugere-se realizar a seguinte seqüência de passos: (a) iniciar o `tcpdump`, (b) executar o ataque, (c) parar o `tcpdump` e (d) armazenar o tráfego.

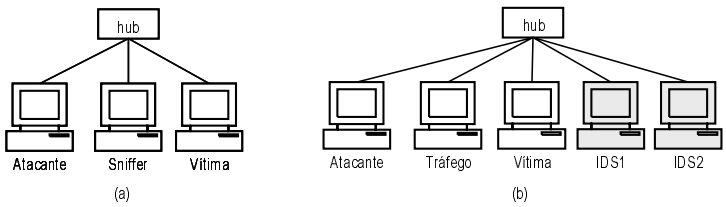

Figura 2: Ambiente de rede para (a) coleta e armazenamento do tráfego dos ataques e (b) para o cenário de avaliação dos IDSs

2.3.2 Geração do tráfego de fundo

O tráfego de fundo é necessário para realizar a análise de escalabilidade dos IDSs. Nosso grupo optou por utilizar tráfego de fundo artificial homogêneo nessa análise porque ao usar tráfego real, em função da oscilação da banda e da alternância das aplicações utilizadas, seria difícil identificar, por exemplo, a que taxa de transmissão o IDS começa a descartar pacotes. Além disso, o uso de tráfego real (a não ser que se conhecesse detalhadamente esse tráfego), poderia provocar alarmes não previstos (por exemplo, se ataques estivessem inseridos nesse tráfego), o que geraria ruído na avaliação em curso.

Tecidas essas considerações, sugere-se utilizar um tráfego de fundo composto somente por pacotes UDP de 256 bytes. Esse tráfego deve ser reproduzido em diferentes taxas de transmissão (ex: 4, 6, 8, 10 e 12 Mbps). A ferramenta indicada para tal é denominada `gerador_udp`, que foi desenvolvida no LAND/UFRJ. Como ela é de fácil manipulação, torna-se desnecessário armazenar esse tráfego para posterior reprodução.

2.4 Montagem do Ambiente de Avaliação

O ambiente de avaliação deve ser composto pelos IDSs a serem avaliados, pelas estações-alvo (*Vítimas*) que sofrerão os ataques, além de uma estação para reproduzir o tráfego de ataque (*Atacante*) e outra para reproduzir o tráfego de fundo (*Tráfego*) ao longo dos testes de escalabilidade. A figura 2b ilustra o ambiente utilizado na avaliação realizada pelo nosso grupo, cujos resultados são apresentados na seção 3.

A quantidade de estações vítimas e o sistema operacional instalado nessas estações podem variar conforme os ataques selecionados para compor o cenário de avaliação. Por exemplo, caso o cenário de avaliação seja constituído por ataques a estações Solaris e Windows 2000 Server, o ambiente de rede representado na figura acima deverá contar com mais duas estações vítimas, nas quais esses sistemas devem estar instalados e devidamente configurados. Além disso, a quantidade de IDSs avaliados também pode variar; por conseguinte, o número de estações para hospedar esses sistemas poderá ser maior.

2.5 Análise dos IDSs

Conforme já mencionado, o roteiro proposto se propõe a avaliar as seguintes características dos IDSs: capacidade de detecção, escalabilidade e taxa de falsos positivos gerados por esses sistemas. Capacidade de detecção é o teste a partir do qual é possível identificar as potencialidades de detecção dos IDSs, ou seja, quais os tipos de ataques que esse sistema está apto a detectar. O teste de escalabilidade permite identificar a partir de qual taxa os IDSs começam a descartar pacotes. Já a taxa de falsos positivos indica a tendência desses sistemas em gerar alarmes falsos, isto é, confundir um tráfego considerado normal com um ataque ou, ainda, quando submetido a um ataque, gerar alarmes referentes a outros ataques.

2.5.1 Capacidade de detecção

Para avaliar a capacidade de deteção dos IDSs, sugere-se realizar a seguinte seqüência de passos: (a) limpar os arquivos de *log* e iniciar o serviço de *log* dos IDSs, (b) reproduzir, um a um, o tráfego coletado dos ataques (vide seção 2.3.1), (c) parar o serviço de *log*, (d) salvar os arquivos gerados e (e) contabilizar e identificar os ataques detectados e os não detectados (a partir da análise dos *logs*). A reprodução dos ataques será realizada a partir da estação *Atacante*, a uma velocidade baixa (para que o IDS não descarte pacotes), usando-se alguma ferramenta como o `tcpreplay`.

2.5.2 Escalabilidade

Para que os resultados da análise de escalabilidade sejam os mais confiáveis possíveis, é fundamental que sejam reproduzidos apenas os ataques que cada um dos IDSs detectou na análise anterior. Por exemplo, se na avaliação da capacidade de detecção o IDS A identificou cinco ataques e o B três, então a análise de escalabilidade desses sistemas será realizada, respectivamente, com esses cinco e três ataques.

A figura 3 representa a relação entre o tráfego de ataque e o tráfego de fundo (homogêneo, gerado a uma taxa constante), reproduzidos concomitantemente nessa análise. Cada um dos tipos de ataques considerado na avaliação (no nosso caso, negação de serviços, evasão, inserção e varredura de portas) deve ser executado em uma bateria de testes distinta. Por exemplo, considerando que a figura 3 seja referente à análise de escalabilidade de um IDS frente aos ataques de negação de serviço, tem-se cinco ataques (Smurf, UDP Storm, Syn Flood, Teardrop e ICMP Fragmentation) sendo reproduzidos em paralelo ao tráfego de fundo (a). Os ataques são executados um após o outro separados por intervalos de dois segundos. O tráfego de fundo começa a ser gerado cinco segundos antes do primeiro ataque ser reproduzido e é interrompido somente após todos os ataques terem sido transmitidos. Tão logo essa seqüência tenha sido finalizada, (b) o sistema de *log* deve ser parado, (c) o número de alertas gerado pelo IDS deve ser registrado (para posterior contabilização), (d) o sistema de *log* deve ser reiniciado e, em seguida, (e) é preciso passar a reproduzir o próximo tipo de ataque. Quando, para cada um dos IDSs, todos os tipos de ataque tiverem sido reproduzidos, (f) deve-se aumentar a taxa de transmissão do tráfego de fundo e repetir o procedimento descrito (a).

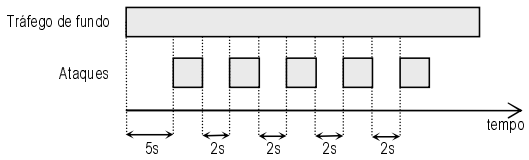

Figura 3: Seqüência de reprodução e geração dos tráfegos utilizados na análise de escalabilidade

2.5.3 Taxa de falsos positivos

Falsos positivos são todos os alarmes que indicam que um determinado ataque está sendo realizado, quando de fato o que está ocorrendo é outro tipo de atividade. Por exemplo, um usuário do suporte executa um comando `ping` para um servidor e o IDS registra este evento como um ataque. Outro exemplo de falsos positivos é quando a rede está sofrendo um determinado tipo de ataque (ex: *UDPstorm*), e o IDS gera alarmes tanto para este ataque, quanto para outros tipos de ataques (ex: *ICMP fragmentation*) que não estão ocorrendo e não possuem relação com o evento em questão.

A abordagem proposta pelo nosso grupo para a identificação de falsos positivos baseia-se na análise dos *logs* gerados nos testes de capacidade de detecção. Após executar um conjunto de ataques, o *log* do IDS registra os alarmes gerados. A separação dos alarmes que efetivamente identificam a detecção dos ataques e os demais alarmes (falsos positivos) pode ser facilmente realizada pelo gerente de segurança. A razão entre o número de alarmes adicionais sobre o total de alarmes gerados para um determinado conjunto de ataques representa um indicador importante de tendência do IDS em gerar falsos positivos. Para tal, os arquivos de *log* resultantes dos testes de capacidade de detecção devem ser novamente consultados.

3 Estudo de Caso

Nesta seção são apresentados os resultados alcançados pelos dois IDSs submetidos ao roteiro experimental descrito na seção anterior. Os sistemas de detecção de intrusão utilizados nesse estudo de caso foram o Snort 1.83 [Roesch, 1999] e o Firestorm 0.4.6 [Firestorm, 2001], ambos disponíveis sob licença GNU GPL versão 2.

3.1 Capacidade de Detecção

A análise da capacidade de detecção foi realizada com base na seleção de ataques mencionada na seção 2.1. Essa análise foi realizada simultaneamente com os dois IDSs escolhidos. Os resultados obtidos tanto pelo Snort quanto pelo Firestorm são apresentados na figura 4. É importante ressaltar que, embora a seqüência de testes descrita na seção 2.5.1 tenha sido repetida dez vezes, os resultados obtidos foram sempre os mesmos (variância estatística foi zero).

Os resultados obtidos nessa análise demonstram que o Snort é uma ferramenta capaz de detectar ataques de inserção, evasão, varredura de portas e negação de serviço de forma bastante eficiente. Em relação ao Firestorm constatou-se que não há um mecanismo eficiente de decodificação de requisições HTTP. O desenvolvimento de melhorias em relação a este mecanismo consta como um dos objetivos para as futuras versões deste IDS.

3.2 Escalabilidade

Para avaliar sua escalabilidade, os IDSs foram submetidos ao procedimento descrito na seção 2.5.2, com tráfego de fundo a 4, 6 e 8 Mbps. Os testes foram repetidos 10 vezes

Ataques	Snort	Firestorm
Evasão		
Method Matching	X	X
Session Splicing	X	X
Inserção		
Long URLs	X	X
Self Reference	X	X
URL Enconding	X	
Varredura de Portas		
UDP Scan	X	
Xmas	X	X
TCP Fragmentation	X	X
Varredura do protocolo IP	X	
Fingerprinting	X	X
Ident Reverse TCP	X	X
Negação de serviço		
Smurf	X	X
UDP Storm	X	X
Syn Flood	X	X
Teardrop	X	X
ICMP Fragmentation	X	X
Obs: O caracter "X" representa que o ataque foi detectado pelo IDS em questão, ao passo que a ausência do caracter indica que o ataque não foi detectado.		

Figura 4: Resultados obtidos na análise da capacidade de detecção

e a variância estatística observada foi de 2,5%. Através dos experimentos realizados constatou-se que, a uma taxa de transmissão de 4 Mbps, os IDSs avaliados não apresentam descartes de pacotes. Sendo assim, o número de alertas (incluindo falsos positivos) gerados corresponde ao máximo possível para o conjunto de ataques em análise. A seguir, na figura 5, são apresentados os resultados obtidos pelos sistemas de detecção de intrusão avaliados. Conforme pode ser observado, o Snort apresentou uma pequena superioridade em relação ao Firestorm quanto à análise de escalabilidade frente a todos os tipos de ataques considerados. Destaca-se ainda que, mesmo a uma taxa de transmissão relativamente baixa (8 Mbps), é possivel que ambos os IDS já deixem de detectar alguns ataques.

	Evasão	Inserção	Varredura de portas	Negação de serviço
4 Mbps				
Snort	100%	100%	100%	100%
Firestorm	100%	100%	100%	100%
6 Mbps				
Snort	98,81%	97,86%	99,32%	99,83%
Firestorm	97,56%	95,18%	99,57%	99,36%
8 Mbps				
Snort	89,99%	86,10%	94,85%	94,41%
Firestorm	86,04%	83,42%	90,49%	92,24%
Obs: O valor em cada célula é obtido pela razão entre o número de alarmes armazenados no *log* ((incluindo falsos positivos) e o número máximo de alarmes esperados (quando o IDS não descarta pacotes). Este número é obtido no teste de capacidade de detecção.				

Figura 5: Resultados obtidos na análise de escalabilidade

3.3 Taxa de Falsos Positivos

A análise das taxas de falsos positivos, conforme descrita na seção 2.5.3, é realizada mediante consultas aos arquivos de *logs* criados pelos IDSs no momento da análise da capacidade de detecção. Embora a seqüência de testes descrita na seção 2.5.3 tenha sido repetida dez vezes, os resultados obtidos foram sempre os mesmos (variância estatística foi zero).

A figura 6 apresenta as taxas de falsos positivos geradas pelo Snort e pelo Firestorm. Os valores altos observados indicam que os IDSs possuem um conjunto de assinaturas pouco precisas (refinadas), o que provoca a geração equivocada de muitos alarmes. Como pode ser observado, essa questão é nitidamente mais crítica no Firestorm, embora os resultados apresentados para o Snort não sejam animadores.

	Evasão	Inserção	Varredura de portas	Negação de serviço
Snort	36,73%	29,97%	4,71%	4,63%
Firestorm	41,32%	42,97%	12,93%	7,47%
Obs: O valor em cada célula é obtido pela razão entre o entre o número de alarmes adicionais (falsos positivos) sobre o total de alarmes gerados.				

Figura 6: Resultados obtidos na análise das taxas de falsos positivos

4 Conclusões e Trabalhos Futuros

As principais metodologias voltadas à avaliação de sistemas de detecção de intrusão partem da premissa de que quanto maior a quantidade de ataques reproduzidos, mais detalhado é o processo de avaliação. No entanto, o que de fato se pode observar é que não existem critérios pré-estabelecidos para a seleção desses ataques. Sendo assim, muitos deles exploram as mesmas características e, portanto, não possibilitam uma avaliação ampla e detalhada das potencialidades dos IDSs. Além disso, este tipo de seleção de ataques torna os experimentos previstos extremamente exaustivos dada a quantidade de ataques a serem realizados. O roteiro proposto neste artigo descreve um método de seleção de ataques, a partir do qual o cenário utilizado no processo de análise dos IDSs é composto apenas por ataques que apresentam características únicas entre si. Através desta proposta de seleção, descrita na seção 2.1, reduziu-se o cenário inicial de ataques em aproximadamente 50%.

Com relação à avaliação de escalabilidade dos IDSs acreditamos que, embora o roteiro seja baseado em tráfego homogêneo transmitido a uma taxa constante, é possível ter uma noção clara da capacidade sustentada do sistema sendo avaliado. Pelos resultados obtidos é possível observar que mesmo quando submetidos a um tráfego relativamente baixo, os IDSs começam a descartar pacotes (comprometendo o processo de detecção). Um trabalho ainda a ser realizado é estender a avaliação da escalabilidade desses IDSs para taxas de transmissão superiores a 10 Mbps (ex: até 100 Mbps).

A avaliação das taxas de falsos positivos geradas, da forma como foi proposta nesse roteiro, é influenciada pelo poder de expressão das linguagens para descrição de assinaturas oferecidas e pela precisão do gerente de rede ao especificá-las. Um mecanismo adicional para esta análise, que leve em consideração o tráfego de fundo típico encontrado na organização em que o IDS vai ser utilizado, ainda precisa ser investigado.

Outros trabalhos futuros previstos incluem (a) a ampliação do cenário de avaliação, através da seleção de outros tipos de ataques, (b) a investigação de procedimentos para avaliar outros critérios (ex: capacidade de manipular ataques concorrentes) e (c) o desenvolvimento de uma ferramenta para auxiliar a execução do roteiro proposto.

Referências

Alessandri, D. (2000). Using rule-based activity descriptions to evaluate intrusion-detection systems. In *Third International Workshop on Recent Advances in Intrusion Detection (RAID)*, pages 183–196.

Barber, R. (2001). The evolution of intrusion detection systems - the next step. *Computer & Security*, 20(2):132–145.

Firestorm (2001). *Firestorm network intrusion detection system Homepage*. http://www.scaramanga.co.uk/.

ISS (1999). *Real Secure Systems Inc. Homepage.* http://iss.net.

Lippmann, R., Haines, D., Fried, D. J., Das, K. J., e Korba, J. (1999). Evaluating intrusion detection systems: the 1999 darpa off-line intrusion detection evaluation. *Computer Networks*, 34(4):579–595.

NFR (2001). *Network Flight Recorder, Inc. Homepage*. http://www.nfr.com/.

Paxson, V. (1999). Bro: a system for detecting network intruders in real-time. *Computer Networks*, 31(23-24):2435–2463.

Puketza, N., Chung, M., Olsson, R. A., e Mukherjee, B. (1997). A software plataform for testing intrusion detection systems. *IEEE Software*, 14(5):43–51.

Roesch, M. (1999). Snort - lightweight intrusion detection for networks. In *USENIX LISA Conference*.