Método para Extração de Atributos a partir de Pacotes de Redes

Jean Douglas Gomes Valencio, Renato Bobsin Machado, Gustavo dos Santos Vieira
Laboratório de Pesquisa em Segurança Computacional (LAPSEC)
Universidade Estadual do Oeste do Paraná (UNIOESTE)
Foz do Iguaçu - Brasil

Abstract- The intense growth of computational incidents, denotes the fragility of the computational systems, mostly with the Internet. In this context, intrusion prevention and detection systems become increasingly important. These systems, for the most part, are based on the analysis of network packets, considering different attributes. The objective of this work is the development of an application, in real time, that captures network packets and extracts relevant attributes that will serve as input for computational intrusion detection systems.

Keywords - Network packet attributes; Computer network; Intrusion detection; Package Analysis.

***Resumo*— O forte crescimento de incidentes computacionais denota a fragilidade dos sistemas computacionais principalmente com a Internet. Neste contexto, os sistemas de prevenção e detecção de intrusão se tornam cada vez mais importantes. Tais sistemas, em sua maioria, baseiam-se na análise de pacotes de rede, considerando diferentes atributos. O objetivo deste trabalho é o desenvolvimento de uma aplicação, em tempo real, que capture pacotes de redes e extraia atributos relevantes, que servirão de entrada para sistemas computacionais de detecção de intrusão.**

Palavras-chaves—Atributos de pacotes de rede; Redes de computadores; Detecção de intrusão; Análise de Pacotes.

I. Introdução

A Internet é uma ferramenta indispensável nos dias de hoje. Os sistemas computacionais em rede podem facilitar nossas vidas, porém trazem riscos à segurança dos dados, incluindo ataques e invasões em dispositivos e sistemas [1].

Salienta-se que as redes são intrínsecas as mais variadas atividades em termos comerciais, sociais, culturais e pessoais atualmente. O rápido crescimento de conectividade e acessibilidade à Internet ocasionou o surgimento de diversos problemas de segurança [2]. Neste contexto, uma intrusão pode ser definida como uma ação (ou conjunto de ações) que tem como objetivo comprometer a integridade, a confidencialidade ou a disponibilidade de um recurso computacional. Dessa forma, acessos não autorizados, alterações realizadas de forma indevida e roubo de informações são exemplos de eventos de intrusão em ambiente computacional [3].

Com a finalidade de prevenir, identificar e reagir a tais práticas intrusivas, surgiram os chamados Sistemas de Detecção de Intrusão (SDI), visando determinar se os eventos ocorridos no ambiente correspondem à ações usuais do usuário ou eventos suspeitos, os quais podem sugerir algum tipo de atividade ilícita [4]. Diferentes tipos de métodos de detecção de intrusão vêm sendo desenvolvidos, buscando-se aumentar as taxas de acerto na classificação,e assim, tornando os sistemas cada vez mais eficazes. Dentre os métodos que compõem o estado da arte dessa linha de pesquisa, citam-se aqueles que utilizam-se de técnicas de aprendizado de máquina [10]. Tais métodos, em sua maioria, são modelos constituídos a partir de bases públicas contendo pacotes de eventos de rede representando diferentes tipos de ataque, e também tráfego normal.

Com o intuito de facilitar a aplicação destes modelos em tempo real, assim como para auxiliar pesquisadores a construírem novas bases de eventos computacionais direcionados a detecção de intrusão, no presente trabalho é proposto um método para extração de atributos, em tempo real, a partir de pacotes de rede. O intuito principal é auxiliar a tarefa de classificação de atividades intrusivas, por meio da análise do fluxo de pacotes de dados em redes de computadores, permitindo o uso de métodos de detecção distintos.

II. Base de Eventos

A base de eventos CICIDS2017 foi escolhida para a execução deste trabalho, tendo em vista que possui classes de ataques amplamente utilizados atualmente e com diversas abordagens de realizar tais ataques, aplicando ainda ferramentas e estratégias atuais [5].

A base CICIDS2017 possui dois formatos disponíveis, a base de dados de eventos brutos em formato .pcap e a base em formato atributo-valor em ".csv", já rotulados.

A base bruta de eventos, ".pcap", foi criada a partir da captura do tráfego de uma rede controlada, simulando atividades benignas e diversos ataques, entre os quais: DoS, DoS GoldenEye, Heartbleed port 444, Brute Force, DoS Hulk, DoS Slowhttptest, DoS slowloris, SSH-Patator, FTP-Patator, Web Attack Brute Force, Web Attack XSS, Web Attack Sql Injection, Infiltration Cool disk, Infiltration Dropbox download, Botnet ARES, PortScan, DDoS e DDoS LOIT.

Para a construção da base atributo-valor, os eventos brutos foram processados após a realização da captura de todos os dados do período de tráfego controlado, aplicando a ferramenta CICFlowMeter [8].

III. Sistemas de Detecção de Intrusão

Um SDI busca identificar padrões intrusivos e emitir "alarmes de intrusão" para que os administradores de segurança da rede possam tomar as ações necessárias, tais como bloquear os intrusos ou isolar os pontos de ataque quando necessário [9].

Um SDI pode ser baseado no *host* ou baseado na rede, e com frequência, são adotados modelos híbridos contemplando as duas abordagens conjuntamente. Um sistema baseado somente no *host* irá monitorar a atividade em um único computador, evitando que usuários violem a política de segurança do sistema. Já um sistema baseado em rede, monitora a rede e compreende as ações que ocorrem nesta rede, com o intuito de detectar potenciais violações de segurança [6].

As principais classificações quanto ao método de detecção são:

- Detecção por anomalia: é procurado anormalidades no tráfego de rede, assim, sendo tomada a decisão, se o evento é normal ou suspeito. A construção deste detector começa por formar um perfil que é constituído como normal para o sujeito observado, podendo se um sistema ou um usuário, e depois decidir qual a frequência de atividades sinalizadas como anormais [6].
- Detecção por abuso: também chamada de detecção por assinatura. Consiste na classificação que tem padrões de ataque bem definidos, e os compara com as informações que estão sendo coletadas pelo SDI. Os padrões de ataques conhecidos quando coincidem com os padrões que estão sendo identificados pelo sistema pode sinalizar intrusão [7].
- Detecção híbrida: para obter uma melhor, e mais eficiente detecção, muitos sistemas utilizam a abordagem híbrida, que é a junção das duas abordagens anteriores, tendo por justificativa o fato de que cada abordagem é adequada para certos tipos de ataques [10].

IV. Atributos

A seleção de atributos tem uma grande importância para o presente trabalho, pois para cada evento um conjunto de atributos terá maior relevância para a detecção de uma possível intrusão ou de um tráfego normal. Os principais protocolos que serão explorados estão dentro do modelo TCP/IP, como: o protocolo IP (Internet Protocol) da camada de rede; o protocolo UDP (User Datagram Protocol) um protocolo de transporte não orientado à conexão, o protocolo TCP (Transmission Control Protocol) que é um protocolo orientado a conexão da camada de transporte, oferecendo ainda controle de transmissão, e um fluxo de bytes fim a fim confiável; e o protocolo ICMP (Internet Control Message Protocol), protocolo da camada de rede para controle de mensagens [9][11][12].

Segundo [5], para cada tipo de rótulo, é possível concentra em quatro atributos principais para a classificação do evento.

Por exemplo, um evento de DDoS tem como principais atributos: Packet Length Standard Deviation, Average Packet Size, Flow Duration, Flot Inter arrival Time Standard Deviation [5].

A base atributo-valor (no formato ".csv") CICIDS2017 possui 84 atributos, com sendo complexo determinar a relação entre os atributos. Sendo assim, um sistema é proposto para converter os dados brutos em dados que possuam somente os atributos necessários para identificar uma intrusão, e que o façam em tempo real, permitindo assim o uso dos modelos de detecção em situações reais de redes de computadores.

V. Metodologia

Neste trabalho propõe-se um sistema para realizar o processamento de dados brutos da rede (que são primeiramente gravados em formato ".pcap"), e transformá-los em uma base no formato atributo-valor. A partir desta atividade é possível a geração de novos atributos e filtragem dos atributos que sejam relevantes para a detecção de intrusão.

A arquitetura do método proposto é apresentada na Figura 1.

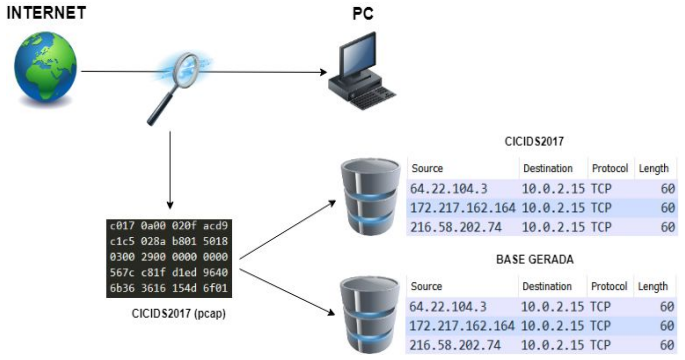

Fig. 1. Fluxograma do Funcionamento do Sistema

Para a extração dos atributos, o *script* irá gerar o *parsing* dos pacotes, buscando o relacionamento e agrupamento entre os mesmos, podendo assim criar atributos derivados que facilitem a futura rotulação dos eventos.

Neste contexto, aplicou-se a biblioteca Scapy para a implementação, em Python, do *script*. Tal biblioteca é responsável pelas funções de manipulação de pacotes de redes, podendo gerar visualizações gráficas dos mesmos, ajudando a identificar campos com potencial valor para a detecção de intrusão.

A ferramenta WireShark, amplamente utilizado para *sniffing*, também serve como apoio, para a manipulação, geração e visualização dos arquivos ".pcap" para teste do script.

O hardware que está sendo utilizado para o desenvolvimento do trabalho é um notebook com processador i7-7500 GHz, 8 GB de memória, 1TB de armazenamento com o sistema operacional Windows 10 64 bits.

O script está sendo desenvolvido utilizando a linguagem python na versão 3.6.6, juntamente com a biblioteca Scapy na versão 2.4.0 e a ferramenta Wireshark na versão 2.4.5.

VI. Resultados Parciais

Como resultados parciais, um protótipo do *script* foi implementado e a partir de um tráfego de rede ".pcap" foi gerado pelo autor, com tráfego normal.

O *script* com a utilizando de um parser extraiu alguns campos, como endereço de origem, endereço de destino, protocolo, tamanho total, entre outros atributos que possuem relevância para o resultado, mostrado abaixo pela figura 2.

Ethernet	
dst	52:54:00:12:35:02
src	08:00:27:fc:24:44
type	0x800
IP	
version	4
ihl	5
tos	0x0
len	40
id	28961
flags	DF
frag	0
ttl	64
proto	tcp
chksum	0x1b1f
src	10.0.2.15
dst	216.58.202.70
options	[]
TCP	
sport	46146
dport	https
seq	3268251338
ack	35937172
dataofs	5
reserved	0
flags	A
window	65320
chksum	0xaeaa
urgptr	0
options	[]

Fig. 2. Parser de um pacote de rede

Como próximo passo será realizado o estudo, para tratar pacotes em janelas de tempo, assim agrupando os pacotes de forma correta, e gerando atributos derivados, que possam facilitar ainda mais a identificação de atributos com peso para a detecção do possível evento maligno.

Como exemplos de campos presentes na base temos endereço IP de origem, endereço IP de destino, protocolo, tamanho do pacote, duração da conexão, ociosidade, entre outros. A relação atributos podem representar um evento maligno.

O aprimoramento do *script* será realizado para contemplar essas mudanças no projeto, para identificar pacotes entre si, os relacionando.

A avaliação será realizada com a comparação da base gerada pelo script após processar a base ".pcap" da CICIDS2017 com a própria base CICIDS2017 ".csv". A comparação consistirá na quantidade de atributos que serão reduzidos da base original.

Referências

[1] NAKAMURA, E. T.; GEUS, P. L. de. Segurança de redes em ambientes cooperativos. [S.l.]: Novatec Editora, 2007.

[2] TSAI, C.-F. et al. Intrusion detection by machine learning: A review. Expert Systems with Applications, Elsevier, v. 36, n. 10, p. 11994–12000, 2009.

[3] MUKKAMALA, S.; JANOSKI, G.; SUNG, A. Intrusion detection using neural networks and support vector machines. In: IEEE. Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. [S.l.], 2002. v. 2, p. 1702–1707.

[4] ROWLAND, C. H. Intrusion detection system. [S.l.]: Google Patents, 2002. US Patent 6,405,318.

[5] SHARAFALDIN, I.; LASHKARI, A. H.; GHORBANI, A. A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: ICISSP. [S.l.: s.n.], 2018. p. 108–116.

[6] AXELSSON, S. Intrusion detection systems: A survey and taxonomy. [S.l.], 2000.

[7] CROSBIE, M.; SPAFFORD, E. H. Defending a computer system using autonomous agents. 1995.

[8] CICFlowMeter (2017). Canadian institute for cybersecurity (cic).

[9] GOODRICH, M. T.; TAMASSIA, R. Introdução à segurança de computadores. [S.l.]: Bookman, 2013.

[10] SOUZA, C. A. d. et al. Método híbrido de detecção de intrusão aplicando inteligência artificial. Universidade Estadual do Oeste do Paraná, 2018.

[11] TANENBAUM, A.; WETHERALL, D. Redes de computadores. [S.l.]: PRENTICE HALL BRASIL, 2011. ISBN 9788576059240.

[12] FOROUZAN, B. A. Comunicação de dados e redes de computadores. [S.l.]: AMGH. Editora, 2009.