WebMBO: Uma Ontologia para Comportamento de Malware Web

Alann Perini[1], Rodrigo Bonacin[1,2], André Grégio[3]

[1] Centro Universitário Campo Limpo Paulista (UNIFACCAMP)
Rua: Guatemala, nº 167, Campo Limpo Paulista - SP, 13231-220

[2] Centro de Tecnologia da Informação Renato Archer (CTI)
Rodovia Dom Pedro I, Km 143,6 Campinas, SP - CEP 13069-901

[3] Universidade Federal do Paraná (UFPR)
R. Evaristo F. Ferreira da Costa, 383-391, Curitiba - PR, 82590-300

alannkp@gmail.com, rodrigo.bonacin@cti.gov.br, gregio@inf.ufpr.br

Abstract. *Malicious software (malware) is a major threat to information security and it is largely associated with attacks to the Web. Knowledge about how malicious software behave is the basis to develop and maintain more secure information systems. However, current web malware has complex behavior, which cannot be represented by the traditional labeling system based on classes, which defines discrete types of malware. In this paper, we propose the use of ontologies to represent suspicious behavior on the web, since ontologies can provide formal and computer-interpretable models capable of representing complex behaviors. To this end, this paper presents the modeling process of the WebMBO, an ontology representing suspect behavior using OWL and SWRL.*

Resumo. *Programas maliciosos (malware) são uma grande ameaça à segurança de informação, sendo associados à maioria dos ataques presentes na Web. O conhecimento sobre o comportamento malicioso desses programas constitui a base para construir sistemas de informação mais seguros. Entretanto, malwares web atuais apresentam comportamentos complexos que não podem ser precisamente representados pela rotulação tradicional baseada em classes que definem tipos distintos. Neste artigo, é proposto o uso de ontologias para representar comportamentos suspeitos de malware Web, uma vez que elas podem prover um modelo formal e interpretável por computador capaz de representar comportamentos complexos. Para tanto, este artigo apresenta o processo de modelagem da WebMBO, uma ontologia que representa comportamentos suspeitos usando OWL e SWRL.*

1. Introdução

Desde seus primeiros anos, a Internet atraiu diversos ataques promovidos por entidades criminosas. Um *malware* é um programa que é executado em um sistema como qualquer outro, sendo que as intenções de um criminoso são traduzidas em instruções. Exemplares de *malware Web* são construídos para executar vários crimes virtuais, como tomar controle do sistema da vítima, roubar informações privadas, lançar ataques de negação de serviço e *spams*. Quando um *malware* não tem uma assinatura conhecida, se

torna mais difícil a sua detecção por ferramentas convencionais, sendo necessária a análise do comportamento de execução. Torna-se também necessária a formalização do conhecimento sobre o comportamento de *malware*, para fomentar a construção de sistemas de informação mais seguros. A análise de comportamento de *malware* é um passo importante para auxiliar na descoberta de padrões suspeitos, mas ao mesmo tempo uma tarefa difícil, pois depende do monitoramento do sistema da vítima durante a infecção [Grégio, Bonacin e Nabuco 2014].

O objetivo desse artigo é apresentar o processo de modelagem e descrição inicial de uma ontologia para representar os comportamentos de *malware Web*, visando a construção de sistemas mais seguros e mecanismos de detecção, bem como facilitar o entendimento comum de profissionais da área. Espera-se contribuir como ferramenta conceitual para tornar sistemas Web mais seguros; ao identificar em eventos coletados e representar de maneira formal comportamentos de *malware Web*.

Neste artigo, é apresentada a modelagem da WebMBO (*Web Malware Behavior Ontology*) por meio da análise de comportamento de *malware* coletados com o uso de um *honeypot*, estudo de relatórios e bibliografia da área e a consulta à especialistas. Por fim, é apresentada a análise prelimitar da representatividade da ontologia utilizando um *dataset* de comportamento de *malware Web*.

2. Trabalhos Relacionados

Com o objetivo de investigar o problema foram realizadas pesquisas nas seguintes bases: ieeexplore[1], acmdl[2], google scholar[3], e science direct[4], em agosto de 2016. Para tanto, foram utilizadas combinações das seguintes palavras-chave: *analysis, malware, ontology* e *"semantic web"*. De um total de 31 artigos relacionados, oito (8) foram selecionados para serem apresentados nesta seção, de acordo a contribuição e foco.

Karande *et al.* (2015) apresentam um sistema de segurança baseado em ontologia que prevê e classifica ataques em aplicações W*eb*, bem como dá sugestões para detectar e prevenir ataques. Já a proposta de Razzaq *et al.* (2014) inclui um método de detecção e classificação de ataques contra aplicações W*eb*. Nesse método, as ameaças são especificadas usando regras semânticas que estabelecem consequências de ataque comuns. Enquanto Moheeb *et al.* (2011) apresentam uma análise de dados recolhidos no período de quatro anos usando ferramentas para detecção de *malware Web*. Os resultados obtidos apontam limitações do uso de ferramentas para detecta-los.

Huang *et al.* (2010) propõem o desenvolvimento de um sistema inteligente para análise do comportamento de *malware*. Posteriormente, Huang *et al.* (2013) apresentam a *Taiwan Malware Análise Net* (*Twman*), um sistema baseado em ontologia para análise de *malware* por comportamento. Em Jasiul *et al.* (2014) é descrito um método de identificação de comportamento de *malware* chamado *PRONTO*, que usa ontologia e SWRL em seu mecanismo de detecção. Já Shoaib e Farooq (2015) propõem um sistema baseado em ontologias para classificar *spams*. O projeto visa diminuir a alta taxa de mensagens verdadeiras e legítimas enviadas indevidamente para caixas de lixo

[1] http://ieeexplore.ieee.org/Xplore/home.jsp
[2] http://dl.acm.org/
[3] https://scholar.google.com.br/
[4] http://www.sciencedirect.com/

eletrônico. Chang *et al.* (2014) apresentam um mecanismo baseado em representações por meio de ontologia para análise de comportamentos de *malware*. O trabalho busca expandir o modelo de análise ITFS (*Interval Type Fuzzy Logic System*) ao incluir a coleta de registros e um modelo para criar uma ontologia de comportamento.

Nosso artigo, está baseado em trabalhos anteriores de Grégio *et al.* (2016), que propõe ontologia para comportamento de *malware*. Entretanto, aquela ontologia não representa as especificidades do comportamento de *malware Web*, estando restrita ao comportamento de *malware* em MS-Windows.

Conforme apresentado, os trabalhos mais próximos a este artigo investigam comportamento de *malware* com o uso de Ontologias. Tais trabalhos apresentam soluções e clarificam desafios a serem abordados em cada caso. Outros artigos analisados na revisão bibliográfica também apontam a limitação da abordagem baseada em análise de código. Entretanto, não foram encontrados trabalhos que detalhem de maneira abrangente o comportamento de *malware Web* fazendo uso de ontologias e regras, bem como uma implementação que permita uma validação experimental.

3. Modelagem da WebMBO

A WebMBO é baseada nos aspectos conceituais e reuso da MBO (*Malware Behavior Ontology*) [Grégio *et al.* 2016]. Com esta estratégia, é possível partir de um modelo pré-existente e validado, para a sua extensão com foco em representar comportamento de *malware web*. O processo de modelagem da WebMBO também faz uso de comportamentos relatados em pesquisa na literatura e relatório do TOP10 OWASP [Owasp 2017]. Tais fontes visam dar fundamentação inicial ao modelo, podendo este ser expandido para considerar outras fontes de conhecimento. Além disso, a modelagem da ontologia também considera estudo com *logs* extraídos um *honeypot*. Por meio de seus históricos de ataques, é possível extrair padrões de comportamento que servem de base para a modelagem da ontologia. Conforme ilustra a Figura 1, a modelagem seguiu as seguintes etapas: (1) estudo e reuso da MBO, (2) análise de modelos e comportamentos de *malware* publicados na literatura, (3) instalação e uso de *honeypots* para extrair informações sobre comportamento de *malwares*, (4) modelagem da ontologia e (5) refinamento da modelagem com base em análises.

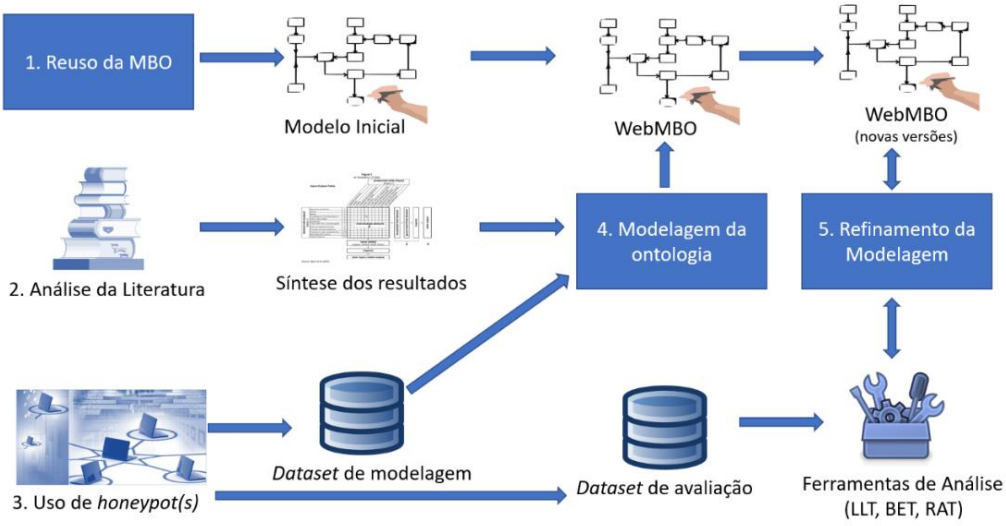

Figura 1 – Modelagem da WebMBO

3.1 Reuso da MBO

A Figura 2 apresenta as classes e relacionamentos de níveis hierárquicos mais altos da MBO, que representam respectivamente: *SuspiciousSoftware* que são execuções de programas suspeitos em um sistema hospedeiro (*System*). A classe *SuspiciousExecution* representa execuções suspeitas. *Action* são ações/eventos que estão relacionados a um objeto de origem (*SourceObject*) e outro que sofre essas ações (*TargetObject*). *ProcessAction* é uma ação de um processo, identificada por um nome e é associada à medida de tempo. *SuspiciousBehaviour* representa que, quando aplicável, um processo pode ser associado com uma execução suspeita.

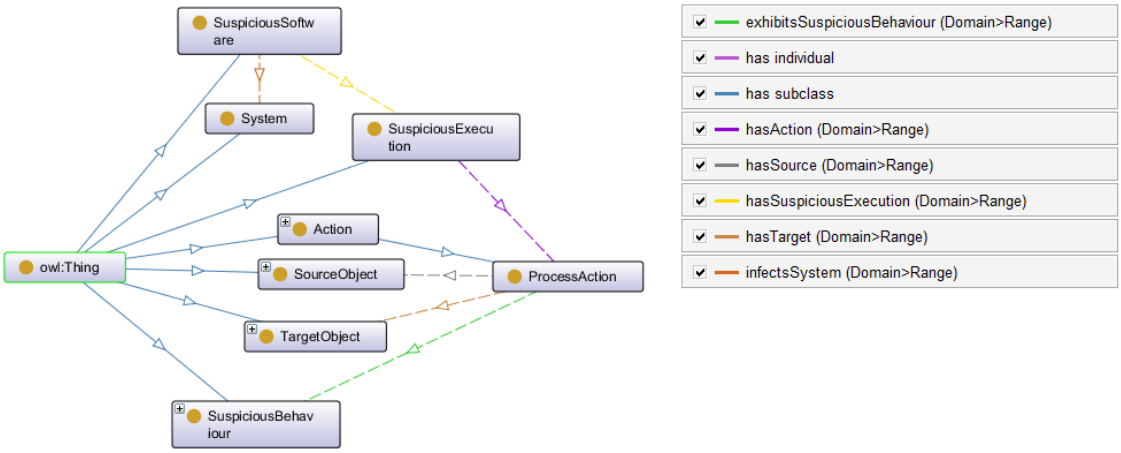

Figura 2 – Classes principais da MBO

3.2 Coleta com *Honeypot* e Validação

Foi instalado o *honeypot KFSensor Versão Professional 3.4.2* com o objetivo de coletar *logs* para análise do comportamento para a modelagem da ontologia, bem como para posterior validação da proposta. Foram objetos de monitoramento os serviços IIS HTTPS, POP3, SMTP, NNTP, ISSPROXY, IDS, entre outros. Ao analisar os arquivos de *log* é possível saber as portas, protocolos e dados usados pelo invasor, além de outras informações, tal como o seu IP, conteúdo enviado/recebido e as ações do *malware*.

No total *307.570* eventos foram coletados de 15 de fevereiro até 25 de julho de 2017. Para manipulação e análise, estes arquivos foram organizados em formato independente de *honeypot*. Estes arquivos possuem os campos propostos por Grégio *et al.* (2014), que inclui: tempo/data do evento, *SourceObject*, *ProcessActionName*, *TargetObject* e *TargetIdentifier*. Foi adicionado o campo *ReceivedData* que é utilizado para identificar e caracterizar tentativas de ataques *Web*. Em seguida foi realizada a análise dos *logs*, por exemplo, ao considerar a frequência, conteúdo dos dados, tipos de ações, entre outros fatores. O *logs* coletados também foram utilizados na etapa 5, para validação e aprimoramento da ontologia. Para tanto, foram utilizadas as ferramentas propostas em Grégio *et al.* (2016): a Ferramenta Carregadora de Logs (LLT - *Log Loader Tool*); a Ferramenta de Extração de Comportamento (BET - *Behavior Extraction Tool*); e a Ferramenta de Análise de Risco (RAT – *Risk Analysis Tool*).

4. Descrição da ontologia WebMBO

As principais classes da WebMBO (como extensão da MBO) incluem: *WebClient* representa situações em que o objeto de origem (*SouceObject*) é um cliente *Web;*

WebTargetObject contém objetos (*TargetObject*) que podem receber ações de *malware Web*; *WebEventAction* representa ações (*Action*) associadas a programas *Web* que podem resultar (ou não) em eventos suspeitos; *WebSuspiciousBehaviourEvent* representa classes de comportamentos suspeitos de ações provenientes da *Web*.

A Figura 3 ilustra a hierarquia de subclasses expandida da classe *WebSuspiciousBehaviourEvent*. Este conjunto pode ser atualizado conforme novos comportamentos são analisados ou surjam pela evolução de malware na Web. São exemplos de classes de eventos (para citar algumas): *DenialofServicesEvent* evento de ataque de negação de serviço, *SessionFixationEvent* evento de ataque que permite que um invasor sequestre uma sessão de usuário e *TrojanExploitEvent* evento de ataque por no qual o usuário frequentemente usa um *javascript* malicioso, que se hospeda em sites e após ser executado faz download de outros exemplares de *malware*. Já as subclasses de *WebTargetObject* contém portas que recebem ataques (*e.g.*, HTTP, SMTP, UDP, POP3), conforme registro de ataques pelo relatório TOP10 OWASP [Owasp 2017] e os registros de *logs* gravados pelo *Honeypot*.

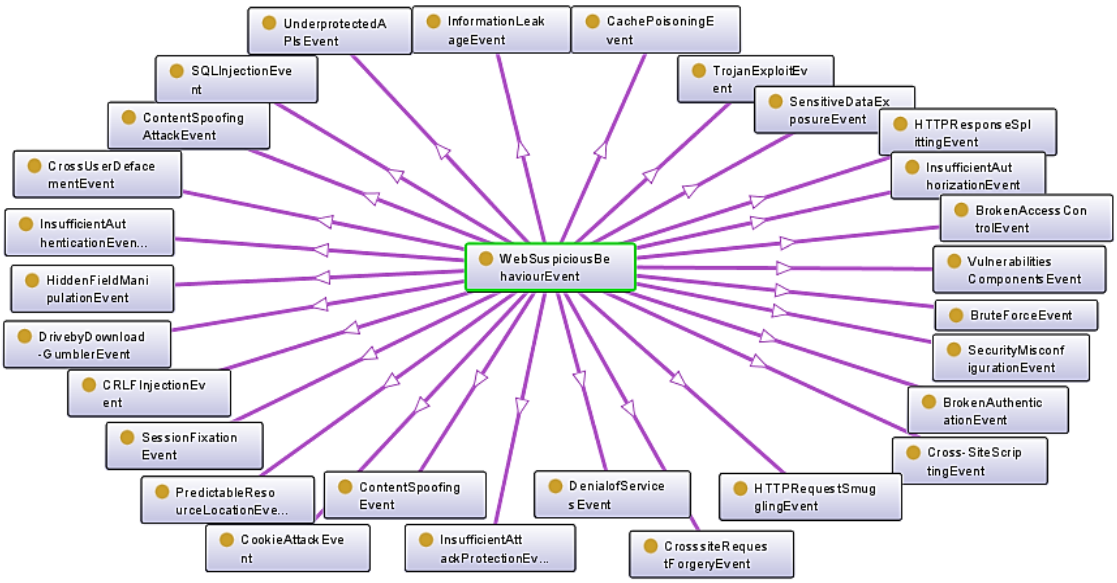

Figura 3 – Visão ampliada de eventos web suspeitos definidos, seus comportamentos associados e atividades suspeitas.

A ontologia possui um conjunto de regras associadas às classes de comportamentos suspeitos. No total, foram modeladas *30* regras. Para a modelagem e validação, foi criado um conjunto de tabelas com o formato apresentado na Tabela 1. Em uma análise preliminar utilizando *34.164* eventos, em um conjunto inicial de regras, mostra que as classes de eventos suspeitos mais frequentes foram *CookieAttackEvent* com *11,38%* e *SessionFixationEvent* com *10,05%*.

Esta análise inicial não é suficiente para determinar medidas de precisão e cobertura, em função do seu tamanho e abrangência. É importante ressaltar que não existe base padrão para tal finalidade, e criar tal base requer esforço em longo prazo. Entretanto, tal validação foi importante para verificar a pertinência das regras, erros conceituais e consistência.

Tabela 1. Exemplo de regra modelada

Log: <event sensorid="Admin" id="237301" type="Connection" action="SimStdServer" name="Blaster, Trojan" simname="Command console" protocol="HTTP" severity=" High"> <start>2017-06-19 17:26:35:469</start> <end>2017-06-19 17:26:35:749</end> <client domain="ip-195-182-138-178.clients.cmk.ru" ip="195.182.138.178" port="80" /> <host ip="172.16.1.250" bindip="" port="4444" /> <connection closedby="Server" /> <recBytes>43</recBytes> <received size="43" coding="kf"><![CDATA[%03%00%00+&%E0%00%00%00%00%00Cookie: mstshash=hello%0D%0A Command: QUERY "select shell("C:\\ProgramData\\Microsoft\\Windows\\Start Menu\\Programs\\Startup**yizai4.exe**")"%0D%0A %01%00%08%00%03%00%00%00]]> </received>
Regra: Service_Port(?y), WebEventAction(?x), hasTarget(?x, ?y), processActionName(?x, ?n), contains(?n, "CONNECTION"), receivedData(?x, ?w), contains(?w, ".exe"), targetName(?y, ?z), contains(?z, "4444") -> TrojanExploitEvent(?x), riskLevel(?x, "3")
Explicação: *(Service_Port(?y))* é uma porta de serviço **E** é *WebEventAction(?x)* uma ação correspondente a um evento Web **E** *?x* tem como alvo a porta de serviço *?y* (*hasTarget(?x, ?y)*) **E** nome desta ação contem "CONNECTION" (*processActionName(?x, ?n), contains(?n, "CONNECTION")*) **E** os dados recebidos contém ".exe" (*receivedData(?x, ?w), contains(?w, ".exe")*) **E** o nome do alvo contém (*targetName(?y, ?z), contains(?z, "4444")*) **-> Então** este será um evento que pode estar associada a exploração de Trojans que possui o nível de risco 4 (*TrojanExploitEvent(?x), riskLevel(?x, "3")*)

5. Considerações Finais e Trabalhos Futuros

Programas maliciosos tornaram mais complexos e capazes de explorar vulnerabilidades, resultando na necessidade de entender e representar o conhecimento sobre *malware web* de maneira explícita e formal. Para tanto, este artigo propôs a WebMBO, uma ontologia para representar comportamentos suspeitos exibidos por *malware web*. Por meio de um método de modelagem baseado em dados empíricos, análise de publicações e o reuso da MBO, foi defino um modelo OWL que compõem eventos suspeitos e seus comportamentos associados; bem como, foi descrito um conjunto de regras SWRL para definir o comportamento. Análise preliminares com *logs* mostram a consistência do modelo e a incidência dos comportamentos modelados.

Como próximos passos desta pesquisa, é proposta a análise extensiva do modelo e expansão da ontologia para um conjunto maior de comportamentos de *malware web*. Além disso, é proposto estender a ontologia para representar *malware* em sistemas móveis.

Referências

Grégio, A., Bonacin, R., Marchi, A.C., Nabuco, O.F., e Geus, P.L. (2016) An ontology of suspicious software behavior. *Applied Ontology.* 11 (1), 29 - 49.

Grégio, A., Bonacin, R., e Nabuco, O.F. (2014) "Ontology for Malware Behavior: A Core Model Proposal." IEEE 23rd International. WETICE, 453-458.

Huang H., Chuang T., Tsai Y. e Lee C., "Ontology-based intelligent system for malware behavioral analysis," International Conference on Fuzzy Systems, Barcelona, 2010, pp. 1-6.

Huang, H.D., et al. Fuzzy markup language for malware behavioral analysis. *On the Power of Fuzzy Markup Language.* Springer Berlin Heidelberg, 2013. 113 - 132.

Jasiul, B., Sliwa, J., Gleba, K., e Szpyrka, M. (2014) “Identification of malware activities with rules.” IEEE, WARSAW: Computer Science and Information Systems, 101-110.

Karande, A., Harshal, S., Gupta, S., Gupta, D., e Kulkarni, A.P. (2015) Security against Web Application Attacks Using Ontology Based Intrusion Detection System. *International Research Journal of Engineering and Technology.* 89-92.

Owasp. (2017) Top10 OWASP - The Ten Most Critical Web Aplication Security Risck. The Open Web Aplication Security.

Razzaq A., Latif, K., Ahmad, H. F., Hur, A., Anwar, Z., e Bloodsworth, P.C. (2014) Semantic security against web application attacks. *Information Sciences* , 254, 19-38.

Shoaib, M, e Farooq, M. (2015) “A User Centric Ontology Driven Spam Detection.” 48th Hawaii Inter. Conference on IEEE. Hawaii: System Sciences, 3661-3669.