International Journal of Innovative
Computing, Information and Control
Volume **7**, Number **7(B)**, July **2011**

 ISSN 1349-4198
pp. 4347–4362

WEB MIMICRY ATTACK DETECTION USING HTTP TOKEN CAUSAL CORRELATION

Ching-Hao Mao[1], Hahn-Ming Lee[1,2], En-Si Liu[1] and Kuo-Ping Wu[1]

[1]Department of Information Science and Information Engineering
National Taiwan University of Science and Technology
No. 43, Sec. 4, Keelung Rd., Taipei 106, Taiwan
{ d9415004; hmlee; m9615046; wgb }@mail.ntust.edu.tw

[2]Research Center for Information Technology Innovation
Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

Received March 2010; revised November 2010

Abstract. *Web mimicry attacks evade anomaly-based web intrusion detection systems (WIDSs) by inserting non-functional characters. In this study, we propose a web mimicry anomaly detection method that uses HTTP token causal correlation. The proposed method extracts token sequences from web application HTTP requests and models the token correlation based on conditional random fields (CRFs) in order to identify web mimicry attacks. The CRF model is widely used for solving sequence labeling problems and suitable for capturing the dependency among different tokens in a token sequence. Since CRFs relax the strong independence assumptions that the other probabilistic sequence analysis methods (e.g., hidden Markov model) have, it can capture long term dependencies among the observed sequences of web tokens to improve the detection capability by observing significant attack patterns. The proposed method requires only HTTP request information and can be easily plugged into existing intrusion detection systems. Two datasets from "ECML/PKDD 2007's Analyzing Web Traffic challenge" and real world HTTP traffic data extracted from a private telecom company are used for an evaluation. The experiment results show that the proposed system performs well in the detection of both web mimicry attacks and general web application attacks even in heavily intersecting cases.*

Keywords: Web security, Mimicry attack, Conditional random field, Cross site script, SQL injection

1. **Introduction.** Since many economic activities such as shopping and cyber banking can be performed through web applications, web applications attract the attention of hackers who break into computer systems with the aim of obtaining valuable information [1,23]. The anomaly detection paradigm, which models normal behavior, is widely used in protecting web application services and particularly useful for detecting novel and unknown attacks. Attackers thus try to hide their malicious behavior by mimicking normal behavior to deceive an anomaly-based web intrusion system (WIDS). This type of attack behavior results in a high false negative detection performance for an anomaly based WIDS making it difficult to identify known attacks. The term "mimicry attack" was first introduced by Wagner et al. [21] to describe attacks that allow a sophisticated attacker to craft malicious actions to evade a detection mechanism. Mimicry attacks also exist at the network traffic application level, involving malicious web application behavior used to evade WIDSs [12,15]. In this study, we use the term web mimicry attacks to denote mimicry attacks involving the network traffic at a web application. Similar to host or network mimicry attacks, web mimicry attacks usually use evasion techniques to