

Machine Learning Based Web Application Firewall

Batuhan İŞIKER
Computer Engineering Department
Gebze Technical University
Kocaeli, Turkey
bisiker@gtu.edu.tr
ORCID: 0000-0001-7737-2214

İbrahim SOĞUKPINAR
Computer Engineering Department
Gebze Technical University
Kocaeli, Turkey
ispinar@gtu.edu.tr
ORCID: 0000-0002-0408-0277

Abstract—Internet and computer systems are an indispensable part of daily life. The number of web applications have increased with the development of technology and digital transformation. Web applications have high risk for security since the applications is not developed securely, contains vulnerabilities and easily accessible by hackers. Web application firewall is used to protect web applications from attacks. Signature-based and anomaly-based methods are used in web application firewalls. In this research, anomaly-based web application firewall model was developed using natural language processing techniques and linear support vector machine learning algorithm. Word n-gram and character n gram natural language processing techniques were compared by performing separate models. Proposed model achieve higher detection performance with using the character n-gram compared to other studies. According to the results of the experiment proposed model is capable of detection web attacks effectively with the overall detection accuracy rate of 99.53%.

Keywords—*web application firewall, machine learning, natural language processing, n-gram, tf-idf*