Scholar Name	Nancy Agarwal
Supervisor Name	Dr. Syed Zeeshan Hussain
Department	Department of Computer Science (Faculty of Natural Science)
Thesis Title	Intrusion Detection for Web-based Applications

Abstract

A great deal of efforts has been spent by the research community to design an effective IDS for web applications. But yet, designing a suitable IDS for websites is still a major concern as attackers do find their ways to bypass this protection shield. There is a need of an intelligent IDS that is exclusively designed to monitor web application traffic, capable of learning the business logic and policies of a web application and uses multitudes of detection methodologies to block all the classes of web-based attacks.

As the first contribution, the thesis proposes the *taxonomy* of web IDS. Being naïve in the domain of web application security is one of the reasons for not having efficient detecting systems for protecting websites. IDS have been primarily designed to observe and detect intrusive activities on the network. The proposed taxonomy will assist in determining the set of essential features required to create a robust and ideal architecture of the webIDS. To the best of our knowledge, a systematic classification of webIDS does not exist so far.

In the next contribution, an experimental study is conducted on a well-known signatures-based IDS, PHPIDS to investigate the potential causes that yield a weak signature set. Signature-based detection is one of the detection methodologies that mostly leverages regular expressions to store the pattern of a known attack vector in order to discriminate out malicious web requests. A weak signature set may severely reduce the performance of the signature-based IDS (SIDS) by steadily increasing the number of false alarms. The experimental case study helped in identifying

various potential reasons behind the design of poor signatures. Based on these reasons, the weak signatures are divided into six categories, namely *incomplete, irrelevant, semi-relevant, susceptible, redundant* and *inconsistent*. To the best of our knowledge, the presented work identified novel signature design issues. Moreover, these design issues are defined mathematically.

In the third contribution, a novel machine learning-based mutation testing framework is proposed and implemented that assists in identifying particularly incomplete signatures in the IDS. The proposed framework consists of five components, namely, *attack-vector preprocessor, cluster program, mutation program, attack bot* and *incomplete signature identifier (ISI) system*. DBSCAN algorithm with Levenshtein distance function has been utilized for grouping similar attack vectors in order to reduce the mutant dataset.

The mutation-based testing model yields 28 signatures of PHPIDS as incomplete. A new version of the quality signature set is also proposed in this contribution, named as *ePHPIDS* by rectifying the identified incomplete rules and removing the extremely generic signatures of PHPIDS. It is shown that *ePHPIDS* outperformed the PHPIDS signatures. The proposed model will assist in enforcing the quality signatures in SIDS.

In the end, the last and fourth contribution of the thesis presents an integrated hybrid framework of webIDS. The presented framework follows modular architecture where the whole system is divided into five components, namely *Preprocessor, Detector, Defender, Logger* and *Response Controller*. The features of the proposed framework have been compared with 5 well-known detection systems, namely *AppSensor, PHPIDS, ModSecurity, Shadow Daemon* and *AQTRONIX WebKnight*.