Supervisor Report

Student Name: Jianing

Supervisor Name: Hemant Ghayvat

Thesis Title: Multi-source Evidence Fusion Based Confidence Assessment Method for SQL Injection

Please find the below comments based on the review of your thesis report, though you have performed the research but the readability of the thesis needs to be improved:

Chapter 1: Introduction

Current status: Manages to achieve the false-positive rate (30-40) of standard tools such as SQLMap and OWASP ZAP. The gap is well established within it since the available tools do not offer continuous confidence values but binary ones.

Technical Refinement: In Section 1.3 (Problem Formulation), indicate clearly the mathematical character of the so-called knowledge gap. Although you say how to quantify, a sentence on the non-existence of a standardized probabilistic framework of vulnerability evidence would be beneficial to the research foundation.

Medical Context: You mention that the effects go to the patient's safety. To enhance that, Section 1.4 (Motivation) should be directly connected to HIPAA technical safeguards, in which the risk-based assessment is not possible, and only vulnerability scanning is needed.

Chapter 2: Method

Current State: You are using a Design Science Research (DSR) approach, which is very suitable to come up with new artifacts

Technical Improvement: You state in Section 2.3 (Reliability and Validity) a Cohen=1.86=1.86. This is a very powerful effect size. Nevertheless, you ought to comment on the threshold selection bias. Because you have selected the thresholds (0.50, 0.80), how did you come up with them- did you simply use them as a result of your 55 test cases or the result of a given statistical distribution?

Constraint Note: Be aware that with the use of DVWA and sqli-labs, constraint makes it appear more like a noise (such as network WAFs) than a real-world hospital system.

Chapter 3: Theoretical Background and Methodology

Current State: This will be the very heart of your technical contribution specifically the 8 feature categories and 4 heuristic rules.

Technical Improvement (Rule 2): Rule 2 (Divergence Penalty) was found to have no effect. Mathematically this implies that the divergence threshold of 0.30 may be too large. Propose a non-linear penalty which increases with the gap S{SQLMap} and S{ZAP} as opposed to a yes/no cutoff.

Rule 4 Logic: You refer to Rule 4 (Medical Field Bonus) as well, which did not play a role. This may be attributed to the fact that very few cases of test cases were used to activate the sensitive field detection (F8)). I strongly recommend switching to a multiplicative weight (instead of an additive bonus), which would be a factor of gamma=0.08 per any module categorized as either L1/L2.

Chapter 4: Implementation

Current State: Basic environment configuration using kali Linux and Ubuntu.

Technical Improvement: You classified 23 cases as "Secure Implementations" In order to enhance this part, elaborate the WAF (Web Application Firewall) configuration applied on TC051-TC052. When the WAF was in detection-only mode the tool response may vary greatly as compared to block mode which affects your confidence scores.

Chapter 5- 6: Results and Analysis.

Current Status: The Full Fusion + Heuristics method was the most effective of all with an 80.65% F1-score.

Technical Improvement: Section 6.3 has observed that L4 recall has dropped to 40 percent when threshold was increased to 0.60 as opposed to 80 percent. This is a negative important finding in usability. Suggest sensitivity analysis of the L4 threshold so as to establish a sweet spot (say 0.52) so as to minimize false positives without compromising the rule out half your detection rate.

Figures and Diagrams visualization and presentation: This needs to be improved for the readability and understanding of the readers.

Architecture and Methodology Figure 3.3 / 3.4:

Comment: The Three-Tier Architecture diagram is useful to put in context. Figure 3.4 (Methodology Design) in Appendix 1, however, is a bit cluttered.

Enhancement: Color-code the data inputs (URL, Scenario), processing layers (SQLMap/ZAP) and the logic layer (Heuristic Rules). Make sure that the "Rule Application Sequence" is shown in a visual form as a flow-chart and not placed in a box as text only.

Figure 5.1 (Confidence Boxplot)

Comment: This is your best visual evidence because there is a clear divide between True Positives and True Negatives.

Improvement: Mark the outliers (the grey points above 0.5 in the case of Secure Implementations). Identifying them by their TCID (e.g., TC037) would enable the reader to instantly cross-reference the reason why the system failed on the particular cases (e.g., delays due to load on the database).

Figure 5.3 (Adaptive Thresholds)

Comment: Successfully demonstrates the points of successful implementation of the method (L3) and the points of too conservative implementation (L4).

Improvement: Include the second Y-axis or a line graph over the bars to depict the Recall Rate. This would make the Recall 40% note of L4 far more meaningful, literally depicting the trade-off between reduction of False Positives and completeness of detection.

Please give the GitHub link of the research work with code and data.