e-Portfolio Activity: Research Proposal Review

Considering your thoughts on your chosen area of interest for your project:

- Which of the methods described in this week's reading would you think would suit your purpose?
- Which data collection methods would you consider using?

Note that you may find that you could be using a mixture of both the research methods and the data collection methods. These considerations should be included in your presentation of the Project Proposal in Unit 10.

As I intend to build an artifact that can perform automated API (passive) reconnaissance, I believe a mixed methods research approach, in which secondary sources would derive from previous studies involving IDS and vulnerability detection (Huang et al, 2021) and automation (Akshay et al., 2018), API vulnerability detection (Ball, 2022) and automation (Bogle et al., 2022), and parallel computing for web crawling (Akshay et al, 2018) along with primary research would best suit my research goals. The primary research would consist of an artifact that could produce a list of secondary and primary sources of testing targets' OSINT API documentation, which is a critical aspect of API vulnerability reconnaissance (Ball, 2022), an experiment with the following hypotheses:

H0 = Automated API reconnaissance does not significantly reduce equivalent manual reconnaissance testing times (status quo) while producing relevant search results

H1 = Automated API reconnaissance significantly reduces equivalent manual reconnaissance testing times (research question) while producing relevant search results.

The research itself would be to discover the best program design to prove H1 and disprove H0 (a baseline for manual testing times may best be found either through secondary research, or by survey, depending on available information). I must admit, I have some reservations about my ability to prove the validity of the experiment. While t-Tests and p-Values are good indicators of statistical validity, the p-Value itself “does not provide good enough evidence favoring any non-null hypothesis” (Dingledine, 2018). The concept of α and “the probability of the result based on prior knowledge” (Dingledine, 2018) are significant contributors to the actual quality of the results. So I would need to be very clear about my testing variables and how to evaluate them to prove or disprove the null hypothesis – something I can say that, at the moment, I am not.

The actual subject of the test would be the artifact algorithm, but the reconnaissance resources produced (data, .yaml files, API documents, etc.) and time management (time-to-complete) would be the test variables for validity. These variables would need to be numerical, with documents being web scraped from an algorithm of my design through parallel computing to decrease time-to-complete, and would have to depend on the definition of what ‘relevant search results’ indicate. This could be based on the relevance of a document title, the amount of matching keywords found in a document’s text, or another indicator of relevance not yet discovered. The sample size would need to be sufficiently large (at least n=30) to be able to have a t-test with $\alpha = 0.05$, as these parameters are less likely to result in a false rejection of H1 and a non-rejection of H0 (Berenson et al., 2015), and the statistical tests may need to take on more than two variables if the time management algorithm testing is to be robust. These are complicated aspects to consider while planning the project, and may require revision as initial research continues.

References

Akshay S et al. (2022) Automation of Recon Process for Ethical Hackers. In: *2022 International Conference for Advancement in Technology, Goa, India, 21 – 22 January, 2022.* IEEE: 1 – 6

Ball, C. J. (2022) Hacking APIs: Breaking Web Application Programming Interfaces. San Francisco, CA, USA. No Starch Press.

Berenson, L., Levine, D, & Szabat, K. (2015) *Basic Business Statistics: Concepts and Applications*. 13th Ed. Harlow, UK: Pearson.

Bogle, A., Mahmood, R, Pennington, J., Tran, T., & Tsang, D. (2022) A Framework for Automated API Fuzzing at Enterprise Scale. In: *IEEE Conference on Software Testing, Verification and Validation*. IEEE: 377 – 388

Dingledine, R. (2018) Why is It So Hard to Do Good Science? *eNeuro*, 5 (50. [Available Online] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6126587/

Huang, G., Li, J., Ren, J., Zhang, B. (2021) Efficiency and Effectiveness of Web Application Vulnerability Detection Approaches: A Review. *ACM Computer Survey*, 54 (9): 1 – 35