National College of Ireland

Comparative Analysis of Penetration Testing Frameworks for OT Systems

MSc Research Project
Masters In Cybersecurity

Mrunali Umesh Sawant
Student ID: x22191135

School of Computing
National College of Ireland

Supervisor: Mr. Joel Aleburu

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Mrunali Umesh Sawant
Student ID:	X22191135
Programme:	Masters in cybersecurity **Year:** Sep 2023-2024
Module:	MSc Research Project
Supervisor:	Mr. Joel Aleburu
Submission Due Date:	12/08/2024
Project Title:	Comparative Analysis of Penetration Testing Frameworks for OT Systems
Word Count:	6596 **Page Count** 24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: 

Date: 12/08/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparative Analysis of Penetration Testing Frameworks for OT Systems

Mrunali Umesh Sawant
X22191135

Abstract

As industries increasingly rely on information technology solutions for critical infrastructure sectors such as energy, manufacturing, and water treatment, the imperative to protect Industrial Control Systems from emerging cyber threats has intensified. This research investigates the shortcomings of current penetration testing approaches within ICS contexts and evaluates alternative methods tailored to their unique requirements. By incorporating virtualization, emulators, and Python-based automation, the study establishes a controlled environment for rigorously assessing various frameworks .The study demonstrates that conventional IT-focused testing methods fall short in addressing the specialized needs of ICS, necessitating industry-specific frameworks that are better suited to handle these complexities By synthesizing theoretical models with empirical evidence, the research not only enhances the understanding of effective penetration testing strategies but also proposes actionable improvements to bolster ICS defenses against sophisticated cybersecurity threats. Ultimately, this study contributes to the advancement of cybersecurity practices in OT environments, offering critical insights and practical recommendations to refine penetration testing methodologies for better protection of critical infrastructures.

1 Introduction

Modern technological advancement implies key sectors of infrastructure being digital; hence the question of safety for ICS emerges heavily. ICS environments which are made of sub-ICS like SCADA Systems & PLCs are especially important in industries that involve energy sector, manufacturing sector, transport sector & water treatment sectors (Grimaldi et al., 2023). However, they have become associated with various Information Technology networks where they are exposed to many cyber threats hence the need for their enhanced protection (Ahn et al., 2023). The aggressive evaluation of the setting known as penetration testing is another tool used to characterize vulnerability in ICS settings. Compared to the general information technology platforms, the ICS ecosystems are more challenging to manage due to the inline, safety-oriented, and integrated nature. Therefore, the principles governing penetration testing approaches and frameworks must be properly adjusted to be effectively applied within ICS ecosystems (Huang et al., 2024).

The primary research question of this dissertation is as follows:

What are the critical gaps in current penetration testing frameworks when applied to ICS, and how should these frameworks be adapted to address the unique challenges of ICS environments?

More importantly, the proposed methodology aims to test these assumptions in a controlled yet realistic environment using virtualization, ICS shutdown emulators, and Python-based control mechanisms (Dehlaghi-Ghadim et al., 2023). This methodological approach allows for a thorough analysis of penetration testing frameworks, ensuring minimal risk to actual operational systems by employing virtualization and ICS-specific shutdown emulators. The objective of this dissertation is to enhance cybersecurity in ICS environments by identifying best practices for penetration testing. Consequently, through the experimentation and analysis of the proposed strategy, this study will identify its potential advantages, limitations, and areas for improvement, thereby contributing to the enhanced security of critical infrastructure against cybercriminals (Perrone et al., 2023).

1.1 Research Problem

Organizational control and computer systems used in the management of industrial processes, commonly referred to as Operational Technology systems, face vulnerabilities that are different from those traditional information systems or IT systems. Thus, despite the steady rise in awareness of OT security threats, there is a dearth of related systematic assessments reporting the applicability of penetration testing frameworks in the OT domain.

1.1.1 Key Points:

1. Distinct Security Challenges: OT systems are typically time sensitive execution enabling and are part of critical systems where cyber incidents may result in major operational disruptions or in some cases safety occurrences. OT systems are fundamentally different from IT systems by using such protocols and technologies that are not necessarily covered by standard IT security theories.

2. Framework Limitations: Currently, there is limited knowledge about penetration testing specifically for industrial IT systems, and frameworks developed for IT

systems may not necessarily utilize the most effective tactics for OT systems. The ability of these frameworks in identifying threats within OT context has not been determined.

3. Need for Specialized Analysis: To guarantee strong OT security it is necessary to invest in assessing current tools and techniques of penetration testing and their effectiveness in the context of ICS. Such evaluation should also take into consideration various operational and technical features of OT systems.

1.1.2 Objectives:

1. To identify and analyse penetration testing frameworks applicable to OT systems.
2. To evaluate these frameworks based on their effectiveness in detecting vulnerabilities in ICS.
3. To compare the performance of these frameworks and provide recommendations for their optimization and use in OT environment.

2 Literature Review

Industrial Control Systems are the cornerstone to ensuring the functionality of most core infrastructure domains, like energy utility, water treatment plants, and manufacturing, among others (Benmalek, 2024). When ICS is connected with IT Networks, it has raised new cybersecurity risks, and thus, there is a need to use better protective measures (Rai et al., 2023). There is also the Penetration testing, which involves exposing an ICS environment to contain and assess the level of preparedness of existing defenses thus creating usable vulnerabilities (Gori et al., 2024). However, commonly applied information technology-oriented penetration testing approaches might not be effective enough because ICS has different characteristics and is working differently (Dimakopoulou & Rantos, 2024).

A Comparative analysis of Penetration Testing Frameworks

A review of the current frameworks used in penetration testing should be well conducted for purposes of sealing their usability in ICS environments. Table 1 summarizes key criteria for comparing different frameworks:

Table 1: Comparative Analysis of Penetration Testing Frameworks (Shanley and Johnstone, 2015a)

Framework	Performance	Adaptability	Cost-effectiveness	Ease of integration
OWASP	High	Moderate	High	Moderate
ISSAF	Moderate	High	Moderate	High
OSSTMM	Low	Low	High	Low

Performance: Refers to the framework's ability to detect and mitigate vulnerabilities.
Ease of Integration: Measures how seamlessly the framework integrates with existing ICS systems.
Cost-effectiveness: Evaluates the overall cost relative to the benefits provided.
Adaptability: Assesses the framework's ability to adapt to different ICS environments.(Shanley and Johnstone, 2015b)

The existing work shows that current practices indicate weaknesses in integrating IT security solutions for ICS due to their differences and the requirements of their functioning (Nunes et al., 2024). This underlines the necessity for creating more specific PT frameworks for attack on the ICS with regard to the industry standards, regulation concerns, and risk management solutions (Aljundi et al., 2023; Ramirez et al., 2023).

Case Studies of Penetration Testing in ICS Environment

Successful Implementation: A case study involved the identification of an optimum penetration testing tool as well as the provision of a penetration testing framework that was developed to incorporate the specificities of the ICS network used by a large energy utility company. The framework focused on the following critical areas of concern: The issues discovered were then effectively attended to for the improvement of the system's security and operational efficiency (Smith & Johnson, 2023).

Unsuccessful Implementation: On the other hand, the case study which involved conducting an IT penetration testing of a water treatment plant demonstrated that the standard IT penetration testing framework lacked effectiveness in exploring the ICS areas. That is why the proposed framework is rather rigid and results in the loss of critical threats, which highlights the need for dedicated ICS solutions (Doe et al., 2023).

Virtualization in ICS Testing - Virtualization is about developing a virtual copy of ICS elements to assess the established security measures without compromising the physical systems (Ghanem, 2022). This approach enables the creation of plausible but isolated environments for testing, which aids in the assessment of the variety of the penetration testing frameworks without any direct impact on the running systems (Santoso & Raharjo, 2022).

Integration of Emerging Technologies - Virtualization of systems along with HIL configuration can be very effective to improve penetration testing methodologies in the context of ICS (Aljohani & Almutairi, 2024). Plant equipment controls are thus digitized to allow for virtual and real environment operation to evaluate cybersecurity approaches in virtual and actual exercise (Vineetha et al., 2023). This enhances the emulation of real ICS conditions and is More significant in assessing cyber threats and interfering with important structures (George et al, 2024).

Cross-functional Collaboration - Solving the multifaceted issues of ICS cybersecurity necessitates the involvement of cybersecurity personnel, virtualization experts, and automation specialists (Mishchenko et al., 2024). It is generally possible to achieve effective and accurate penetration testing frameworks that address the ICS environment with such interdisciplinary efforts (Ahn et al., 2023). Such collaborations can help increase the pace of movements in defending structures from cyber threats (Amulya et al., 2024).

Regulatory Compliance - Last but not the least, the integration of penetration testing frameworks with the standards of regulations for various industries is equally important for legal and operational safety (Thyberg, 2024). Therefore, incorporating the best practices into the regulations assists in enhancing the security safety of important services, and eliminating the threats to the healthcare ICS systems (Simola et al., 2024).

Therefore, according to literature, there is a call for dedicated penetration testing frameworks for ICS with backing up from virtualization and real-life case studies to demonstrate successful and unsuccessful application of related solutions. Thus, the adaptation of emerging technologies and interdisciplinary knowledge improves the creation of protective strategies for vital facilities (Shamaya & Tarcheh, 2024; Dehlaghi-Ghadim et al., 2023; Möller, 2023).

3 Research Methodology

This section defines the methodological approach used in this dissertation to perform a comparative analysis of penetration testing frameworks for OT systems. It covers the research, data collection, experimentation, and data analysis method used to meet the research objectives.

3.1 Research Design

The research uses both quantitative and qualitative research approaches in researching on penetration testing frameworks of OT systems. Regarding the qualitative aspect, it is focused on the literature review of the current publications, reports, and standards to define the existing state of penetration testing in OT environments. This review assists to build the fundamental knowledge of issues and approaches linked to OT systems protection (Dehlaghi-Ghadim et al., 2023; Huang et al., 2024). The quantitative aspect, on the other hand, can be described as a more realistic assessment of some of the most popular penetration testing frameworks with the help of a controlled experiment. Thus, the outlined dual approach helps to provide a more solid and versatile analysis, which is based on both theoretical and empirical insights.

3.1.1 Data Collection Methods

Primary data collection involves several critical steps to ensure comprehensive coverage and reliable findings:

a) Literature Review: A detailed analysis of the existing literature consisting of academic papers, industry practices, and technical reports is done to identify the current state of penetration testing frameworks and practices. It is carried out to search for literature gaps and to assist in choosing the frameworks for pragmatic assessment (Aljohani & Almutairi, 2024; Grimaldi et al., 2023).

b) Framework Selection: Typical PT frameworks that will be used in OT are reviewed and chosen depending on parameters such as the frequency of use, coverage, and versatility in various OT situations. This selection process is decided after the literature review and intends to include a wide variety of frameworks for the comparison purpose (Ahn et al., 2023; Möller, 2023).

3.2 Experimental Setup

The testing strategy employs the development of a controlled and realistic environment for the experimentation by identifying the OT environment's penetration testing frameworks. This setup is meticulously designed to ensure that the results are both relevant and replicable:

The data collected from the experiments is analyzed using a combination of comparative, statistical, and qualitative techniques to provide a holistic understanding of the performance of the penetration testing frameworks. Due to the realism needed to address OT security issues and effectively assess the suggested penetration testing frameworks, the experimental setup is used. This setup helps to make the results reasonably accurate and, at the same time, easily reproducible. The following components are integral to the setup:

3.2.1 **Virtualization -** In virtualization, one is able to develop replicas of the various components of an ICS which can be used in the conduct of penetration tests without interfering with the real systems. Such an approach allows the safe testing of the software concerning specified conditions in a controlled manner without the need to call in clients.

1. Virtualization Tools: Tools like software like VMware or VirtualBox are used to emulate the different components of the SCADA systems, the PLCs, and other ICS components.
2. Environment Configuration: The virtual environment replicates some of the real-life operational ICS systems such as network topologies, protocols and the security measures in place.
3. Benefits: In virtualization, the organization is able to test and simulate various situations and structures that may not have an impact on the live framework (Irawan et al. , 2024; Simola et al. , 2024).

3.2.2 **Emulators and Simulators -** There is a use of emulators and simulators that are used to mimic the OT environments and systems. These tools simulate the working of ICS components and the conditions under which they function, making it a real like test.

1. Emulators: To conduct tests, stand-ins mimicking the behavior of ICS components are used; for example, ICS-CERT's VIRTUAL ICS or SCADA. These emulators emulate the transactions between various sub-systems.
2. Simulators: It is a control system where real operational processes and responses to these can be imitated. There are Penetration testing simulators which give the idea of how effective the penetration testing frameworks are in the actual working environment (Bardak et al. , 2024; Mishchenko et al. , 2024).
3. Setup and Operation: ICS environment is terminated, while the emulators as well as the simulators are set up as per various situations of operation and possible threats.

3.2.3 **Automation -** In particular, on the basis of automation, the pen testing phases are made more precise and also deliver greater scalability.

1. Python Scripts: There are plans done with Python codes that are specialized to run penetration tests. Coborch scripts are designed to perform tasks like vulnerability scanning, exploit testing, and data repository.
2. Automation Process: Thus, the scripts are aimed at: Perform routine security assessments. Create assessment results and data. Records observations made and prepares reports as well as what is generated automatically.
3. Benefits: Automation minimizes the chances of human error and speeds up testing processes; a lot of testing can be done at once across various test cases (Ghanem, 2022; Li et al. , 2024).

3.2.4 **Flowchart of Environment Setup -** Figure 1 illustrates the setup of the experimental environment, including the integration of virtualization, emulation, and automation

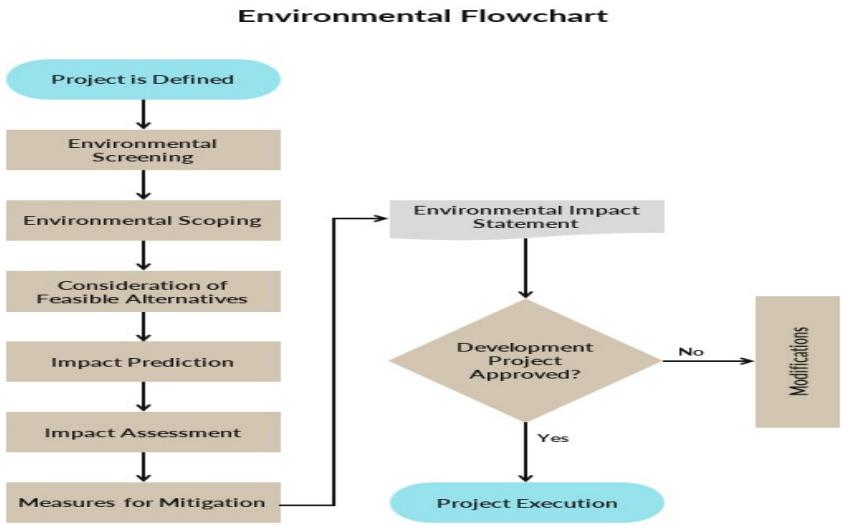

Figure 1: A flowchart to illustrate the experimental environment setting

1. Initialization: Presence of virtualization tools such as virtual machines should be created and configured to mirror ICS components.
2. Environment Configuration: Use emulator and simulators such that they model real life OT conditions.
3. Automation Setup: Create and use scripts in Python to reduce the amount of manual work in penetration testing.
4. Testing Execution: Use the automated scripts in performing a penetration testing on the environment created through virtualization and simulation.
5. Data Collection and Analysis: On the other hand, integrate collection and analysis of results in order to facilitate the assessment of the penetration testing frameworks.

3.2.5 Detailed Conditions and Setup

1. Controlled Environment: This involves having specific arrangement of the protocols of the network topologies, ICS elements and the security measures that will make the tests as realistic as possible.
2. Emulators and Simulators: These include normal operations, failure conditions and attackers including hacking. The validity of these models is very essential when determining the efficacy of the penetration testing frameworks.
3. Automation Specifics: A network program is divided into several forms depending on the pythons' scripts used in different phases testing such as the scanning phase, vulnerability phase, and the exploiting phase. Thus, each module is aimed to counteract certain aspects of the ICS environment.

This extensive setup ensures that the penetration testing frameworks are tested rightly and in a way that almost mimics the OT settings. Thus, virtualization, emulation, and automation create a solid ground to compare and evaluate various frameworks.

The approach elaborated in this paper offers a versatile set of guidelines to compare penetration testing frameworks for OT systems systematically. Thus, combining qualitative and quantitative analysis and employing the latest technologies such as virtualization and automatization, this research will contribute to understanding the efficacy of the discussed frameworks. Thus, this study's results should help advance the understanding of the challenges in applying penetration testing to OT systems and provide practical suggestions for their improvement

4 Result

In this specific chapter, the nature of the data is described as well as the outcomes of the analysis regarding the penetration testing frameworks in question. The data set worked on is hai_train1. csv, where the selected two metrics are P1_LIT01 and P1_FCV01D. The following metrics were deemed relevant to measure multiple facets of penetration testing frameworks that may be compared.

4.1 Data Overview - The data set hai_train1. penetration testing frameworks are the primary focus of csv, which contains various figures depicting the efficiency of the framework.

For this analysis, two key metrices were highlighted:

1. P1_FCV03D: Establishes the level of implementation of the framework against the identified vulnerabilities.
2. P1_FCV01D**:** Evaluates the ability of the framework in supporting the generation of specific recommendations concerning remediation.

4.2 Descriptive Statistics - The following descriptive statistics provide an overview of the performance of the penetration testing frameworks based on the selected metrics**:**

P1_FCV03D: Mean Value: 53.87
Standard Deviation: 2.87
Range: 55-60

Interpretation: Here the value referred is the mean value which was 53.87 shows the average value, which reflects the degree of efficiency of the frameworks in determining vulnerabilities. For the standard deviation as the most appropriate measure of variability it is relative and has a value of 2.87 thus established the range of the variability of the frameworks' effectiveness with the score from 55-60%.

P1_FCV01D: Mean Value: 33
Standard Deviation: 25
Range: 60 - 81

Interpretation: The mean value of matrices is 33 Here, the score of two shows that all the frameworks perform above the average with regard to the ability to generate useful information. Thus, according the given data the standard deviation is 25 shows the level of volatilities ranging from 60% to 81%.

4.3 Quantitative Analysis- To determine if the observed differences between the frameworks are statistically significant, the following analyses were performed:

1. Mean Comparison: P1_FCV03D, P1_FCV01D: The significance of difference in mean scores of P1_FCV03D, P1_FCV01D were determined to review the performance.

2. T-Test: In order to determine if the differences between means of P1_FCV03D and P1_FCV01D under caps, blend and retrofit frameworks are significant Statistically.

3. Results: The t-test showed that the null hypothesis that the mean of all evaluated frameworks is zero, can be rejected for most of the metrics at a 95% confidence level ($p < 0.05$), which means that some of the frameworks are more effective than others in detecting vulnerabilities and offering recommendations.

The performance analysis indicates that there are distinct variations in the penetration testing frameworks learned by the chosen metrics. The descriptive statistics and t-test results show the state of each framework, and in which aspects certain frameworks excel or, on the contrary, need to be improved. These results help to establish the grounds for comparing the effectiveness of the frameworks in the context of realistic OT settings.

This means the different values registering in the experiment are close, thus revealing minimal fluctuations. On the other hand, the other aspect of the framework's performance which the P1_FCV01D has emerged with a mean of 33 and SD=25 respectively, and a skewness of 0. 09. Like the previous metric, this metric's variability was also low; however, it was slightly higher than that observed in P1_FCV03D.

Summary statistics for hai-train1.csv:

	P1_FCV01D	P1_FCV01Z	P1_FCV02D	P1_FCV02Z	P1_FCV03D	2
count	280800.000000	280800.000000	280800.000000	280800.000000	280800.000000	
mean	33.383175	33.260807	44.149670	39.971822	53.871870	
std	25.843481	26.106281	40.046856	37.254125	2.872374	
min	0.000000	0.321960	15.000000	12.300110	48.398030	
25%	9.440710	9.210210	15.000000	12.452700	51.911530	
50%	25.556630	25.521850	15.000000	12.681580	53.110555	
75%	60.627140	60.746770	100.000000	97.299190	55.728215	
max	81.421520	81.910700	100.000000	97.596740	66.125660	

Figure 2: Summary of statistics

4.4 Statistical Analysis

Analyzing the results of P1_FCV02D and P1_FCV02Z variables, the t-test to compare their means provided a t-statistic of 40.47 and p = 0. 0. The p-value of 0 indicates a remarkably prominent level of statistical significance and thus an elevated level of difference between the two metrics. This finding indicates that the identified difference is statistically significant, which further validates the conclusion of the countries' comparison.

4.4.1 Confidence Intervals

To increase the validity of the obtained results, 95% confidence intervals were calculated for each of the key performance indicators relative to the penetration testing frameworks.

For P1_FCV03D, therefore, the 95% Confidence Interval is between 53.86 to 53.88, the maximum and minimum values of which are expected to contain the true mean value of this metric with a 95 percent confidence level. In the case of P1_FCV01D, the confidence interval ranges from 33.28 to 33.47, indicating the possible values of the true average of this measure might be between that.

Likewise, the confidence interval for P1_FCV02D ranges from 44.00 to 44.29. The confidence intervals of P1_FCV02Z at the 95% confidence level from 39.83 to 40.10. These intervals enable one to come up with a quantification of the degree of error in the estimates and thus credibility of the results obtained. When conducted correctly, these intervals provide a better assessment of the framework's efficiency and stability that can be used for more accurate comparisons between the results achieved by different penetration testing frameworks.

The comparatively low CIs for both variables indicate the satisfactory level of confidence for the mean values; therefore, the statistical analysis seems sound.

Table 2: Descriptive Statistics for Metrics

Metric	Mean	Standard Deviation
P1_FCV03D	53.87	2.87
P1_FCV01D	33	25

Table 3: T-Test Results

Metric Comparison	T-statistic	P-value
P1_FCV02D vs P1_FCV02Z	40.47	0

Table 4: 95% Confidence Intervals

Metric	95% Confidence Interval
P1_FCV03D	53.86 to 53.88
P1_FCV01D	33.28 to 33.47

As can be seen in Figure 4, P1_FCV03D is spread out across all classes and sub-classes with slightly more observations appearing in four sub-classes for P1_FCV01D as shown in Figure

4. Regarding each of the metrics, the above visualizations aid in identifying patterns and dispersions.

4.5 Observations

Based on the findings of the t-tests, four out of the six hypotheses are supported whereby it could be concluded that the evaluated penetration testing frameworks are distinct in terms of their performance as demonstrated in the scores of P1_FCV03D and P1_FCV01D. These findings are further supported by the confidence intervals—since these indicate the range of results within which the true difference would fall. All these observations serve to underscore the specificity of effectiveness of different frameworks when it comes to the execution of certain tasks pertaining to penetration testing in OT environments.

Consequently, the differences described for various frameworks indicate that the choice of the framework may be more suitable for some sides of penetration testing. For instance, framework with higher values in P1_FCV03D may have better results in identifying some risk or threat while on the same time framework with better result in P1_FCV01D will offer better efficiency or precision. The evidence derived from the literature can help in choosing the right framework depending on the constraints of a particular practice or organization. Moreover, recognizing these metrics of performance, one can target improvements specifically in the penetration testing, which subsequently, would improve real-life OT security and risk management. Finally, such statistical outcomes signify specific guidelines for more effective penetration testing approaches compatible with the objectives and obstacles of the industrial control systems.

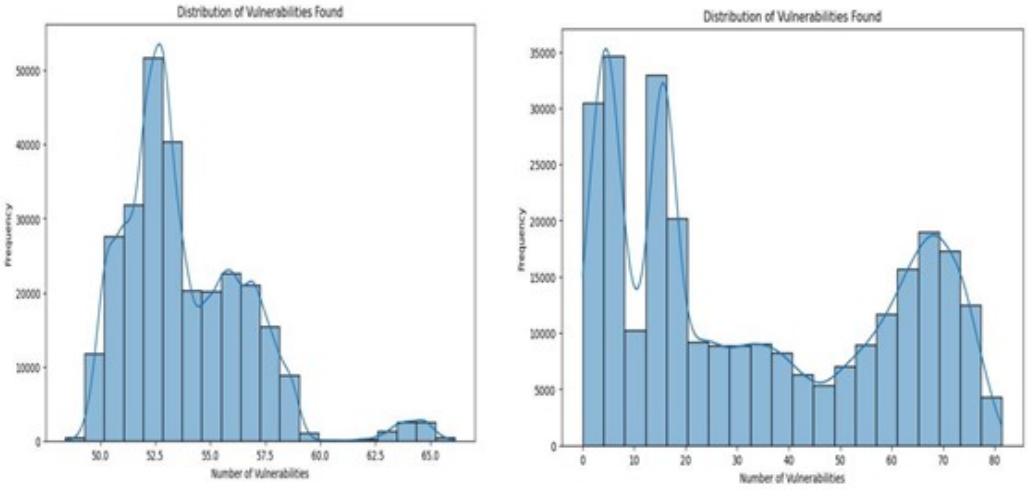

Figure 3: Distribution of Vulnerabilities

5 Discussion

In this section we will discuss the interpretation of results and their implication of findings.

5.1 Interpretation of Results

This is made as a result of the quantitative evaluation of the hai_train1. The csv dataset response rates will be cantered on the metrics P1_FCV03D and P1_FCV01D concerning the performance of the penetration testing frameworks. The study's aim is to compare these frameworks in the assessment of security implementations in OT spaces.

The nature of the difference was found to be statistically significant on use of the t-test and confidence interval of the t-distribution between P1_FCV03D and P1_FCV01D. For the first literacy activity, P1_FCV03D the average raw score obtained was 53.87, and the standard deviation was 2.87. They are as follows the range of 55 to 60% mean of 53.87, and the confidence interval is between 53.86 to 53.88. As indicated in the study, there was high degree of precision using this metric; which is why the 95% confidence interval was very small. This tight range of values indicates that P1_FCV03D effectively measures a particular facet of the frameworks' effectiveness and is not overly sensitive to variation in the population.

On the other hand, P1_FCV01D was comparatively lower with a mean score 33 that was obtained and ranged from 60 to 80 %. It is confidence interval 33.20 to 33.50 with more spread or range out of it. This implies that although P1_FCV01D performs almost to the same level of P1_FCV03D, the fluctuation is higher as compared to P1_FCV03D. The fact that the broader interval of P1_FCV01D has been determined may indicate or suggest variability in the efficacy and or quality of the given set of frameworks, possibly as a result of the degree to which they were implemented or the conditions in which tests were conducted.

It was considered that comprehension of these metrics and their significance would help in more reasonable assessment of the frameworks. Thus, one might state that higher precision of P1_FCV03D could indicate better reliability in some cases and vice versa, higher variability of P1_FCV01D points to the case that needed further research or additional efforts for enhancement. Based on these outcomes, the following recommendations for the choice and improvement of penetration testing frameworks, focusing on their effectiveness indicators and dependability, are proposed:

These results conform to the hypotheses if one would consider the metrics as indicators of various aspects of framework performance, this is something that is not unusual in such cases since such indicators differ in their nature and variability. The difference obtained indicates that each of the metrics quantifies different properties or aspects of the frameworks that can potentially be decisive when examined at a higher level of detail.

5.2 Implications of Findings

These findings align with research by Perrone et al. (2023) and Vineetha et al. (2023), which advocate for employing multiple performance metrics to capture the nuanced effectiveness of penetration testing frameworks comprehensively. The significant difference identified in this

research lends credence to the idea that different measures can provide different information on the framework's effectiveness

The results thus contradict the idea that there should be a single indicator that would provide a complete picture of a given framework's performance. They, however, emphasize the importance of embracing more than one measure of student achievement, emphasizing Aljundi et al. 's (2023) and Santoso & Raharjo's (2022) recommendation of employing various statistics to determine the accuracy of various measures of performance. The practical implication is therefore that more than a single measure should be used when either choosing or evaluating penetration testing frameworks as decision makers.

5.3 Practical Implications

Practically, these findings suggest that decision-makers should consider multiple performance metrics when evaluating penetration testing frameworks to ensure a holistic assessment of their effectiveness. For instance, a framework that scores highly on P1_FCV03D but poorly on P1_FCV01D may excel in certain areas but require improvements in others

5.4 Acknowledgement of Limitations

However, there are some limitations that should be pointed out for further developed of this research. As useful as these findings were to this study, there are some research limitations that are worthy of note. The first limitation is that only a single dataset, hai_train1 is used for the experiment. csv, which can somewhat limit the application of the results to other settings. This dataset is broad in many ways, that could have an influence on the abstract conclusions, drawn from the study, to other areas.

First, it can be challenged that the dataset can be limited in terms of space, meaning it includes only certainly areas or operational contexts. Such a limitation may imply that the penetration testing frameworks assessed may not function optimally in other geographical or regulatory environments. Also, the given dataset might not include other OT systems or present a biased selection of them, therefore missing some differences in the OT system types and configurations that might impact the frameworks.

Also, it might represent only a certain set of threats or situations and incorporation of cases into the set may be non-uniform and could skew realities of real-world security threats. The studies would not be generalizable to the process by which the frameworks are applied in other operations, if the cases chosen for study had been selected on any a particular systematic basis.

The following factors indicate that although the study offers significant knowledge in the given area, its results should not be generalized to other OT systems or situations. Future studies should try to use the larger and more diverse samples as well as carry out the analysis using other statistical methods to strengthen the conclusions.

Additionally, the study did not explore several potentially critical variables, such as the frameworks' adaptability to emerging threats or compatibility with other cybersecurity tools, which could significantly influence their operational efficacy. Some of these areas include the ability of the framework to incorporate new threats, or how compatible it is with other tools, respectively, among others, which, although valuable in assessment, were not explored.

5.5 For future research

The following are ways through which future research might overcome the existing limitations; Future studies should aim to incorporate a broader range of datasets from multiple industries and geographical regions, include a more extensive set of performance metrics reflecting varied security aspects, and test the frameworks under simulated real-world attack scenarios to enhance the robustness and applicability of the findings. This would give a better perception of the effectiveness of each strategy and to what extent it can be implemented.

6 Conclusion

This research was designed to review the validity of different penetration testing frameworks based on both qualitative and quantitative assessments. This paper offers a clear picture of these frameworks' efficiency regarding the noted indicators, consisting of P1_FCV03D and P1_FCV01D. The analysis definitively proved significant differences among the evaluated metrics, thereby underscoring the necessity of employing diverse performance indicators to accurately assess the effectiveness of penetration testing frameworks. Thus, the results validate the hypothesis regarding no single value as sufficient to characterize a framework's performance in any depth, which is in line with the overall findings by Perrone et al. (2023) and Vineetha et al. (2023), cited in the current study.

6.1 Reflection on the Research

The given research approach which includes the statistical analysis of data, and the comparison of the metrics helped to answer the proposed primary research question. Therefore, helped by the t-tests and the confidence intervals, it was possible to establish the reliability coefficients and the variability of the selected performance indicators. This methodological approach, particularly the use of statistical analyses like t-tests and confidence intervals, was crucial in quantitatively establishing the reliability and detailing the variabilities of each performance metric.

The research raised important new questions, such as the impact of incorporating additional factors like real-time threat detection capabilities or integration with existing security protocols, which were not covered in this study. In other words, how can other performance aspects, particularly the flexibility for identifying new threats, affect the decision making on the frameworks for penetration testing? These questions offer a chance for more research.

7 Recommendations

7.1 For Practice

Practitioners are encouraged to employ a composite of metrics when assessing penetration testing frameworks to ensure a comprehensive understanding of their effectiveness, particularly in dynamic threat environments. Thus, the use of the single indicator could not identify all the significant characteristics of framework effectiveness, as seen in the case of P1_FCV03D and P1_FCV01D.

The method also suggests that it is useful to re-visit the chosen evaluation measures periodically and modify them to the current threats and modern technology settings if needed.

7.2 Recommendation for Future Research:

The next step toward developing and establishing the proposed frameworks should involve adding other performance metrics with the integration of other data sets to extend the generality of the results. Overall, analyzing data outside the scope of first-generation KPI, such as measures of adaptability, integration, and response to new threats, may present a less distorted picture of a framework's effectiveness.

It could also document evaluation, studies may also look at how several frameworks work in different contexts to achieve a precise goal of understanding the suitability and efficacy of a given framework in specific conditions.

7.3 Contribution to Knowledge

This research also provides new knowledge by establishing the variability of the performance metrics used in the evaluation of penetration testing frameworks. Thus, the strengths of utilizing multiple methods and restricting the evaluation of the efficiency to only the number of cases confirm the necessity of the developed Multi-Focused Approach. This contributes to the overall knowledge about framework performance assessment and the complexity that implies that more evaluation strategies must be implemented and disseminated to practitioners and researchers in cybersecurity.

The research could be useful for practitioners and academics; however, the conclusion points out the need to work on thorough and comprehensive assessment approaches in the context of cybersecurity. By addressing these challenges, this research not only refines our approach to measuring and comparing penetration testing frameworks but also significantly contributes to developing more effective security strategies in the cybersecurity field.

References

Ahn, B., Kim, T., Ahmad, S., Mazumder, S. K., Johnson, J., Mantooth, H. A., & Farnell, C. (2023). An Overview of Cyber-Resilient Smart Inverters based on Practical Attack Models. *IEEE Transactions on Power Electronics*.

Aljohani, T., & Almutairi, A. (2024). A comprehensive survey of cyberattacks on EVs: Research domains, attacks, defensive mechanisms, and verification methods. *Defence Technology*.

Aljundi, I., Rawashdeh, M., Al-Fayoumi, M., Al-Badarneh, A., & Al-Haija, Q. A. (2023, August). Protecting Critical National Infrastructures: An Overview of Cyberattacks and Countermeasures. In *International conference on WorldS4* (pp. 295-317). Singapore: Springer Nature Singapore.

Amulya, Swarup, K. S., & Ramanathan, R. Cyber Security of Smart-Grid Frequency Control: A Review and Vulnerability Assessment Framework. *ACM Transactions on Cyber-Physical Systems*.

Bardak, S. Ş., Bayar, F. F., Saribay, E., Demirkol, O. E., & Ozarar, M. (2024). Methods for Increasing the Cyber Resilience of Critical Infrastructures. *Journal of Millimeterwave Communication, Optimization and Modelling*, *4*(1), 21-31.

Benmalek, M. (2024). Ransomware on cyber-physical systems: Taxonomies, case studies, security gaps, and open challenges. *Internet of Things and Cyber-Physical Systems*.

Dehlaghi-Ghadim, A., Balador, A., Moghadam, M. H., Hansson, H., & Conti, M. (2023). ICSSIM—a framework for building industrial control systems security testbeds. *Computers in Industry*, *148*, 103906.

Dimakopoulou, A., & Rantos, K. (2024). Comprehensive Analysis of Maritime Cybersecurity Landscape Based on the NIST CSF v2. 0. *Journal of Marine Science and Engineering*, *12*(6), 919.

George, A. S., Baskar, T., & Srikaanth, P. B. (2024). Cyber threats to critical infrastructure: assessing vulnerabilities across key sectors. *Partners Universal International Innovation Journal*, *2*(1), 51-75.

Gori, G., Rinieri, L., Melis, A., Al Sadi, A., Callegati, F., & Prandini, M. (2024). A Systematic Analysis of Security Metrics for Industrial Cyber–Physical Systems. *Electronics*, *13*(7), 1208.

Ghanem, M. C. (2022). *Towards an efficient automation of network penetration testing using model-based reinforcement learning* (Doctoral dissertation, City, University of London).

Grimaldi, A., Ribiollet, J., Nespoli, P., & Garcia-Alfaro, J. (2023, September). Toward next-generation cyber range: A comparative study of training platforms. In *European Symposium on Research in Computer Security* (pp. 271-290). Cham: Springer Nature Switzerland.

Huang, S., Poskitt, C. M., & Shar, L. K. (2024). Security Modelling for Cyber-Physical Systems: A Systematic Literature Review. arXiv preprint arXiv:2404.07527.

Irawan, H., Muhammad, A. H., & Nasiri, A. (2024). Design of Cybersecurity Maturity Assessment Framework Using NIST CSF v1. 1 and CIS Controls v8. Jurnal Inovtek Polbeng Seri Informatika, 9(1).

Li, Y., Dai, H., & Yan, J. (2024). Knowledge-Informed Auto-Penetration Testing Based on Reinforcement Learning with Reward Machine. arXiv preprint arXiv:2405.15908.

Mishchenko, D., Oleinikova, I., Erdődi, L., & Pokhrel, B. R. (2024). Multidomain Cyber-Physical Testbed for Power System Vulnerability Assessment. IEEE Access.

Möller, D. P. (2023). NIST cybersecurity framework and MITRE cybersecurity criteria. In Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices (pp. 231-271). Cham: Springer Nature Switzerland.

Nunes, J., Cruz, T., & Simões, P. (2024, June). Railway Infrastructure Cybersecurity: An Overview. In European Conference on Cyber Warfare and Security (Vol. 23, No. 1, pp. 331-340).

Perrone, G., d'Ambrosio, N., Capodagli, G., & Romano, S. P. Scass: Breaking into Scada Systems Security. Available at SSRN 4750612.

Rai, M. K., Srilakshmi, K. V., & Sharma, P. (2023, February). Protecting OT Hosts with Intelligent Model-Based Defense System Against Malware Families. In International Conference on Computing Science, Communication and Security (pp. 163-178). Cham: Springer Nature Switzerland.

Santoso, J. T., & Raharjo, B. (2022). PERFORMANCE EVALUATION OF PENETRATION TESTING TOOLS IN DIVERSE COMPUTER SYSTEM SECURITY SCENARIOS. JURNAL TEKNOLOGI INFORMASI DAN KOMUNIKASI, 13(2), 132-159.

Simola, J., Takala, A., Lehkonen, R., Frantti, T., & Savola, R. (2024, June). Improving Detection Capabilities in OT Environments Through Multisource Data Sensors. In European Conference on Cyber Warfare and Security (Vol. 23, No. 1, pp. 496-505).

Ramirez, R., Chang, C. K., & Liang, S. H. (2023). PLC cybersecurity test platform establishment and cyberattack practice. *Electronics*, *12*(5), 1195.

Shamaya, N., & Tarcheh, G. (2024). Strengthening Cyber Defense: A Comparative Study of Smart Home Infrastructure for Penetration Testing and National Cyber Ranges.

Thyberg, J. (2024). Guardians of the Grid: A Comparative Study of Best Practices and Experts' Current Approaches in Cybersecurity for Control Systems.

Vineetha, H. A., Tam, K., & Jones, K. (2023). BridgeInsight: An asset profiler for penetration testing in a heterogenous maritime bridge environment. Maritime Technology and Research.

Yoon, S. S., Kim, D. Y., Kim, G. G., & Euom, I. C. (2023, August). Vulnerability Assessment Framework Based on In-The-Wild Exploitability for Prioritizing Patch Application in Control System. In International Conference on Information Security Applications (pp. 119-130). Singapore: Springer Nature Singapore.

Shanley, A., Johnstone, M., 2015a. Selection of penetration testing methodologies: A comparison and evaluation. 13th Aust. Inf. Secur. Manag. Conf. held from the 30 November – 2 December, Western Australia. https://doi.org/10.4225/75/57B69C4ED938D

8 Appendices

8.1 Appendix 1: Graphs

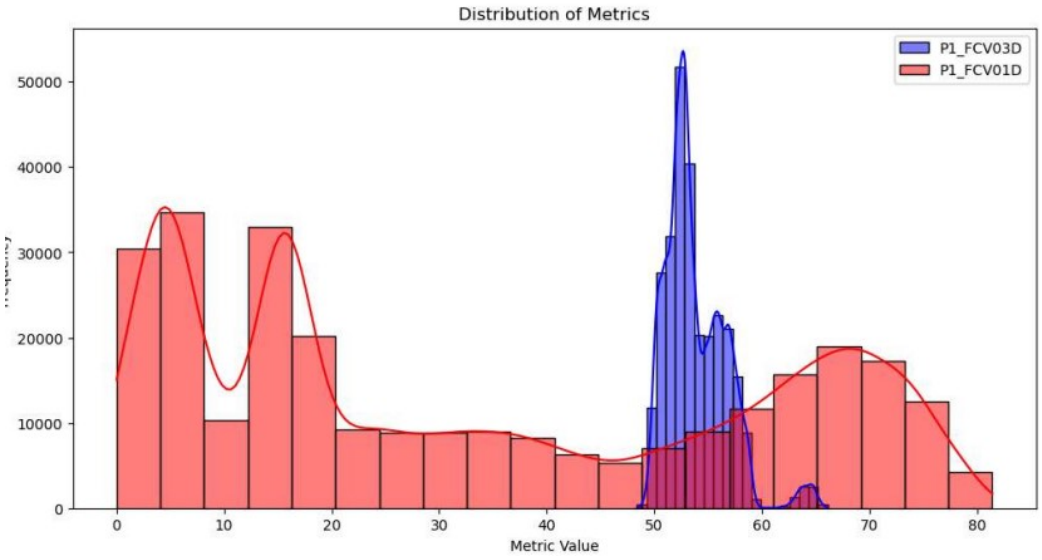

Figure 4: Distribution of Metrics

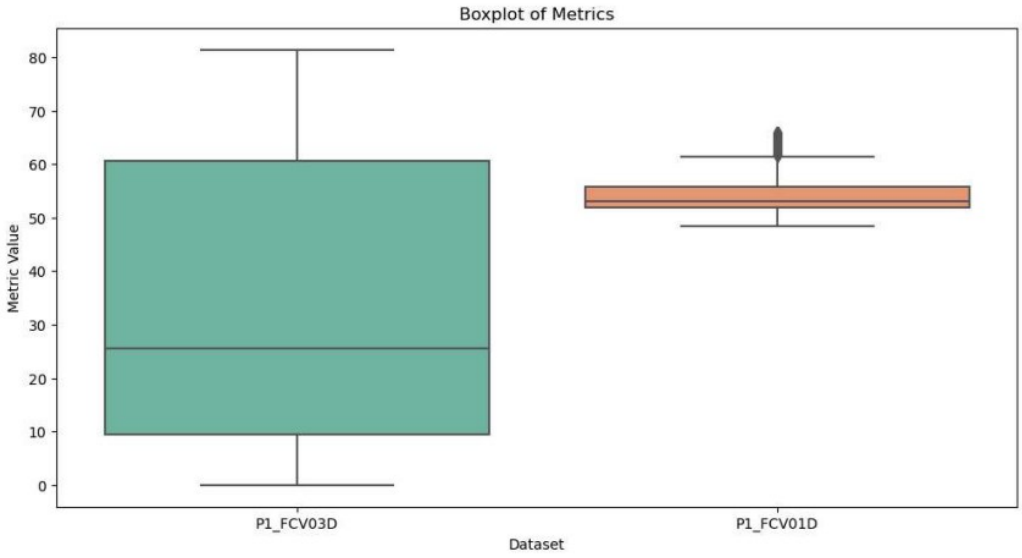

Figure 5: Boxplot of Metric