Assignment 2 - Phase A

AI vs. Humans: Hardware Trojan Detection

Name: Daniel Horan

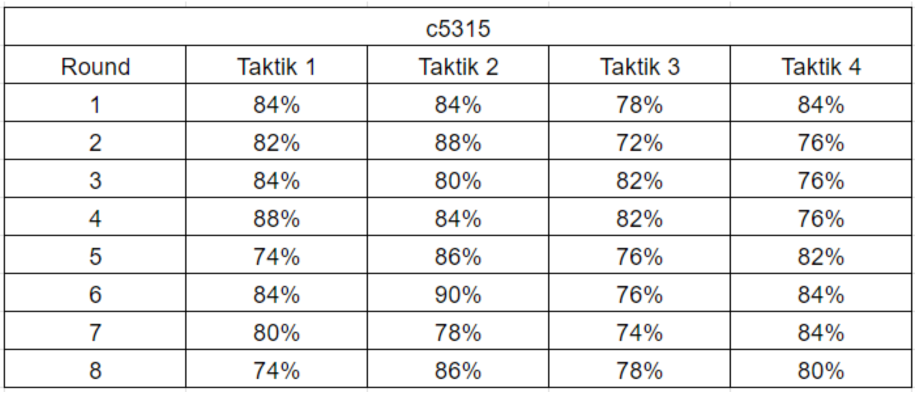
UIN: 527005307

Overview:

In the context of embedded system security, this assignment critically examines the insertion and detection of hardware Trojans (HT) within integrated circuit (IC) designs. Utilizing provided benchmark circuits, the objective is to methodically incorporate HTs and assess the success of various insertion strategies. Central to Phase A is the challenge of evading a non-AI detection tool, TARMAC. The outcomes from this phase will contribute to a deeper understanding of the vulnerabilities in IC design and the effectiveness of contemporary detection mechanisms.

Results Of Benchmarks:

For better data analysis, I have decided to do 8 benchmark runs for each strategy and circuit. Below are the success rates for each one:



c5315				
Round	Taktik 1	Taktik 2	Taktik 3	Taktik 4
1	84%	84%	78%	84%
2	82%	88%	72%	76%
3	84%	80%	82%	76%
4	88%	84%	82%	76%
5	74%	86%	76%	82%
6	84%	90%	76%	84%
7	80%	78%	74%	84%
8	74%	86%	78%	80%

Table 1.1: Benchmark results for c5315

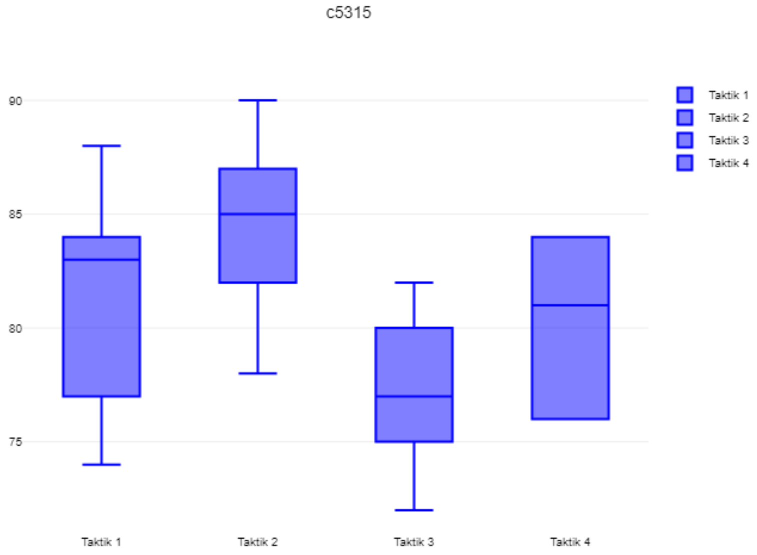
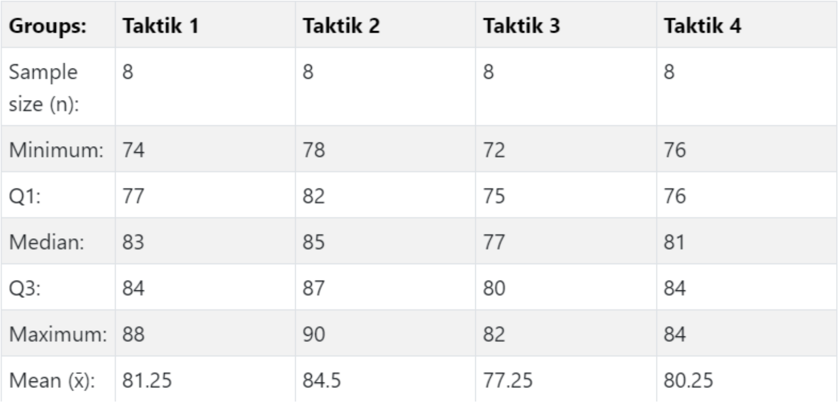
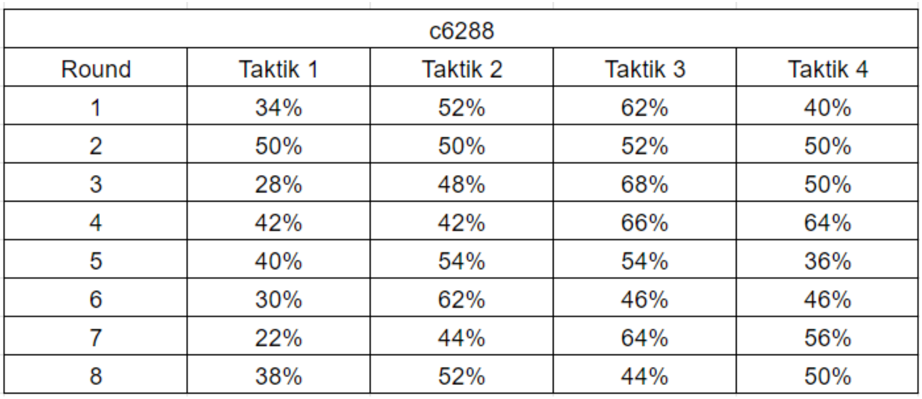


Table 1.2: Box plots for c5315



Groups:	Taktik 1	Taktik 2	Taktik 3	Taktik 4
Sample size (n):	8	8	8	8
Minimum:	74	78	72	76
Q1:	77	82	75	76
Median:	83	85	77	81
Q3:	84	87	80	84
Maximum:	88	90	82	84
Mean (x̄):	81.25	84.5	77.25	80.25

Table 1.3: Additional data for c5315



c6288				
Round	Taktik 1	Taktik 2	Taktik 3	Taktik 4
1	34%	52%	62%	40%
2	50%	50%	52%	50%
3	28%	48%	68%	50%
4	42%	42%	66%	64%
5	40%	54%	54%	36%
6	30%	62%	46%	46%
7	22%	44%	64%	56%
8	38%	52%	44%	50%

Table 2.1: Benchmark results for c6288

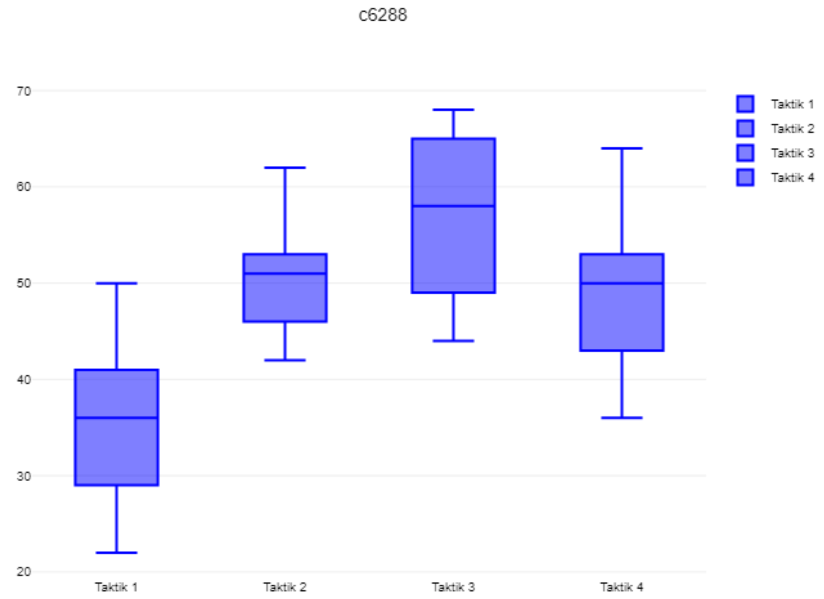
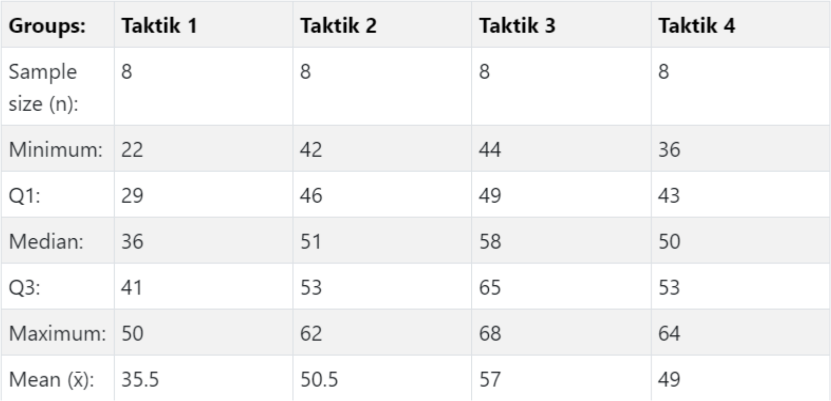
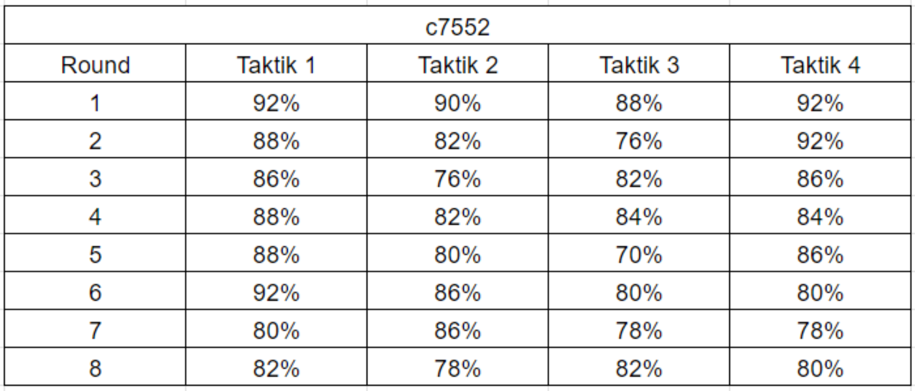


Table 2.2: Box plots for c6288



Groups:	Taktik 1	Taktik 2	Taktik 3	Taktik 4
Sample size (n):	8	8	8	8
Minimum:	22	42	44	36
Q1:	29	46	49	43
Median:	36	51	58	50
Q3:	41	53	65	53
Maximum:	50	62	68	64
Mean (x̄):	35.5	50.5	57	49

Table 2.3: Additional data for c6288



c7552				
Round	Taktik 1	Taktik 2	Taktik 3	Taktik 4
1	92%	90%	88%	92%
2	88%	82%	76%	92%
3	86%	76%	82%	86%
4	88%	82%	84%	84%
5	88%	80%	70%	86%
6	92%	86%	80%	80%
7	80%	86%	78%	78%
8	82%	78%	82%	80%

Table 3.1: Benchmark results for c7552

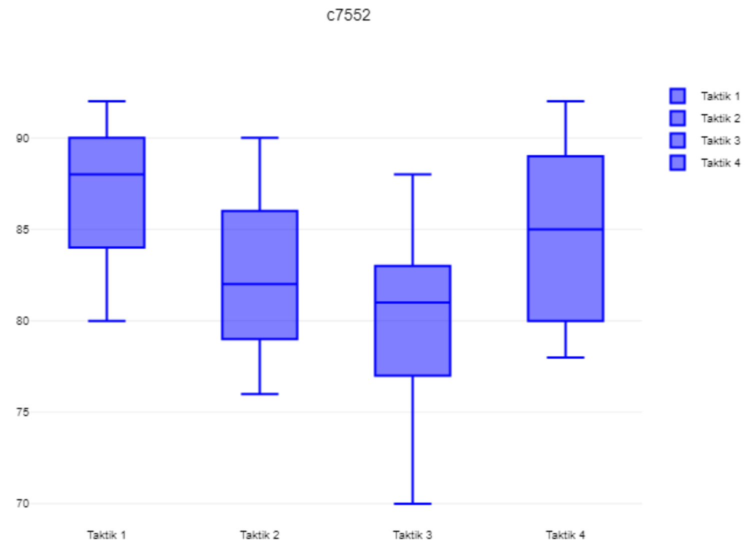
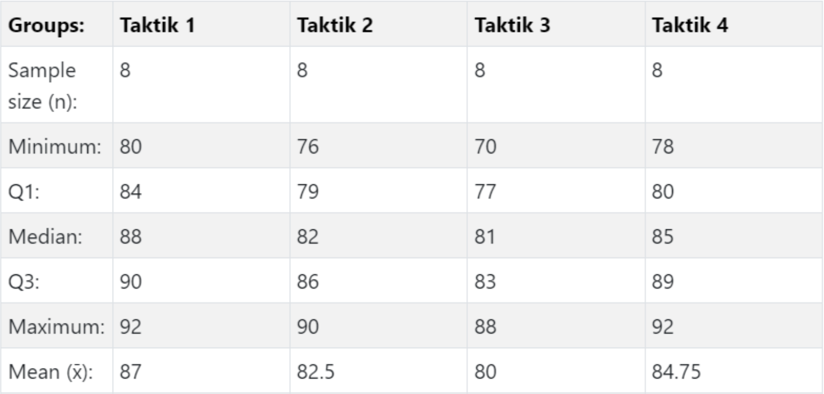


Table 3.2: Box plots for c7552



Groups:	Taktik 1	Taktik 2	Taktik 3	Taktik 4
Sample size (n):	8	8	8	8
Minimum:	80	76	70	78
Q1:	84	79	77	80
Median:	88	82	81	85
Q3:	90	86	83	89
Maximum:	92	90	88	92
Mean (x̄):	87	82.5	80	84.75

Table 3.3: Additional data for c7552

Analysis Of The Results:

Circuit: c5315

Strategy 1:
Range: 74% - 88%
Mean = 81.25, SD = 4.5
Observation: Demonstrates consistent performance across all rounds.

Strategy 2:
Range: 78% - 90%
Mean = 84.5, SD = 3.5
Observation: Emerges as the most effective tactic for this circuit.

Strategy 3:
Range: 72% - 82%
Mean = 77.25, SD = 3
Observation: Shows moderate consistency with a slightly diminished success rate compared to Strategies 1 and 2.

Strategy 4:
Range: 76% - 84%
Mean = 80.25, SD = 3
Observation: Offers consistent results, albeit not surpassing the effectiveness of Strategy 2.

Final Observation: Strategy 2 possesses the highest mean success rate and a relatively low standard deviation, making it the most consistent and successful for c5315.

Circuit: c6288

Strategy 1:
Range: 22% - 50%
Mean = 35.5, SD = 9
Observation: Exhibits a broad range of success rates, indicating variability in performance.

Strategy 2:
Range: 42% - 62%

Mean = 50.5, SD = 7
Observation: Stands out as the most effective tactic for this circuit.

Strategy 3:
Range: 44% - 68%
Mean = 57, SD = 8
Observation: Achieves commendable success, slightly surpassing the average of Strategy 2.

Strategy 4:
Range: 36% - 64%
Mean = 49, SD = 9
Observation: Maintains a consistent performance, though not reaching the peak effectiveness of Strategy 3.

Final Observation: While Strategy 3 has the highest mean, its standard deviation suggests slightly more variability than Strategy 2. This variability might hint at potential sensitivity to different conditions or varying performance in diverse scenarios.

Circuit: c7552
Strategy1:
Range: 80% - 92%
Mean = 87, SD = 4
Observation: Exhibits exemplary consistency and effectiveness for this circuit.

Strategy 2:
Range: 76% - 90%
Mean = 82.5, SD = 4.5
Observation: Delivers consistent results, albeit slightly trailing behind Strategy 1.

Strategy 3:
Range: 70% - 88%
Mean = 80, SD = 6
Observation: Offers moderate consistency and effectiveness, not surpassing Strategy 1 or 2.

Strategy 4:
Range: 78% - 92%
Mean = 84.75, SD = 4.75
Observation: Achieves consistent and impressive results, closely aligned with the performance of Strategy 1.

Final Observation: Strategy 1, with the highest mean and a low standard deviation, showcases robustness and effectiveness for c7552.

Choice Of Strategy:

Circuit c5315:
Strategy 2 stands out with an average success rate of 84.5%. Its consistent performance across rounds suggests a reliable and robust approach tailored for this specific circuit design.

Circuit c6288:

Strategy 3 achieves a 57% average success rate, indicating its adaptability and effectiveness for this circuit. The variance in success rates across rounds suggests that this strategy might be leveraging a more dynamic approach.

Circuit c7552:

Strategy 1 registers an 87% average success rate. Given the multifunctional nature of this circuit, the high success rate of Strategy 1 underscores its comprehensive and holistic approach.

Comparative Analysis of Strategies:

Success Rates: While each strategy has its strengths, Strategy 2 and Strategy 1 consistently emerge as frontrunners for the c5315 and c7552 circuits, respectively. Strategy 3, although marginally superior to c6288, demonstrates its adaptability.

Complexity: The TARMAC documentation emphasizes its prowess in handling large designs with rare trigger conditions. This suggests that strategies aligned with TARMAC, such as Strategy 3, might possess inherent complexities but offer unparalleled efficiency in intricate designs.

Overall Strategy Selection:

Considering the individual circuit performances and the comparative analysis across circuits, *Strategy 2* emerges as the most consistent in terms of success rate. Especially when considering diverse IC designs, this consistency could be vital. While Strategy 1 demonstrated remarkable performance for specific circuits like c7552, Strategy 2's average success rate across the board makes it a suitable choice for a broader range of applications. However, the final strategy choice should be contingent upon the specific requirements and intricacies of the IC design in question.

Conclusion:

In the context of embedded system security, this study evaluated HT insertion strategies within IC designs. Benchmarking results revealed the effectiveness of these strategies against the TARMAC detection tool. Notably, Strategy 2 was optimal for circuit c5315, Strategy 3 for circuit c6288, and Strategy 1 for circuit c7552. These findings underscore the importance of strategic decision-making in IC design security and highlight areas for future research and refinement.