Assignment 2 - Phase A
AI vs. Humans: Hardware Trojan Detection
Date: 10/18/2023 (Wednesday)

Assignment Description:

In this assignment, you will play the role of an attacker who inserts a hardware Trojan (HT) to disrupt the intended functionality of an integrated circuit (IC) design. You will be provided benchmark circuits, and you will be required to insert HTs following different strategies. This assignment is divided into two phases. For each phase, you will know the success rate of the strategies. You will then need to analyze the strategies and reason about which strategy works how well and why.

In the current phase, phase A, you are supposed to evade detection from a non-AI detection tool called TARMAC. All information in this manual after this paragraph is only about phase A. You will be provided a separate lab manual for phase B in due course of time.

For this assignment, you are expected to:

i. Insert HTs using four strategies for the benchmarks provided to you.
ii. Evaluate the HTs against a non-AI-based HT detection tool, TARMAC.
iii. Analyze the four strategies and the reason for their performance.

Resources required for this assignment:

i. Design files, supporting libraries, all necessary code, and state-of-the-art HT detection tools---these are included in the zipped folder "Assignment_2A".
ii. Access to the **Apollo server (apollo.ece.tamu.edu)** – you should already have access to this (if not, please email the TA).

Phase A: Insert and Evaluate Strategies for Inserting Hardware Trojans against a Non-AI-based Detection Tool

In this phase, you will insert HTs using the four strategies provided to you and evaluate the strategies against a non-AI-based HT detection tool, TARMAC. Your objective is to analyze the four strategies in terms of their performance and complexity and reason about them. All the files needed for this phase are in the "Assignment_2A" folder. The contents of this folder are as follows:

1. original_files - this folder contains the original Verilog netlists for the benchmark circuit - c5315, c6288, and c7552.

2. test_patterns - this folder contains the test patterns from TARMAC. You will use these patterns to evaluate your HT-inserted netlists and the four strategies.

3. src - this folder contains all the source codes and libraries needed for phase A.
    a. Trojan_generator_and_evaluator.py - this script is for inserting HTs and evaluating them using test patterns from TARMAC
    b. libsp_parser.py - supporting code for parsing a library file
    c. lib - folder containing some required library files

4. saved_simulations - this folder contains the simulation results for the original Verilog files. The data from the files in this folder is used in the codes to get the probabilities of the nets in the circuit.

5. TARMAC.pdf - this is the research paper detailing the workings of the TARMAC HT detection technique.

Before going over the details of the assignment and how to insert and evaluate HTs, you need to install some packages required for executing the provided codes. This setup process is explained below.

Setting up:

1. Log into your apollo.ece.tamu.edu (Apollo server) account using your NetID and your password. If you are not familiar with how to do that, you can log into your TAMU VOAL account, open MobaXterm, and follow the instructions mentioned here: link.

2. Once on the Apollo server, upload the "Assignment_2A.zip" file and unzip it with the following command on the terminal

```
unzip Assignment_2A.zip
```

3. Navigate to the "src" folder within the newly unzipped "Assignment_2A" folder and enter the following commands one by one on the terminal to install some required packages:

```
pip install --user numpy
pip install --user networkx==1.11
pip install --user pycosat
pip install --user tqdm
pip install --user pyqt5
```

Next, we outline the steps you need to take in order to insert and evaluate HTs in the original Verilog netlists.

To insert and evaluate HTs:

1. Navigate to the "src" folder and run the Trojan_generator.py script with the command

```
python3 Trojan_generator_and_evaluator.py
```

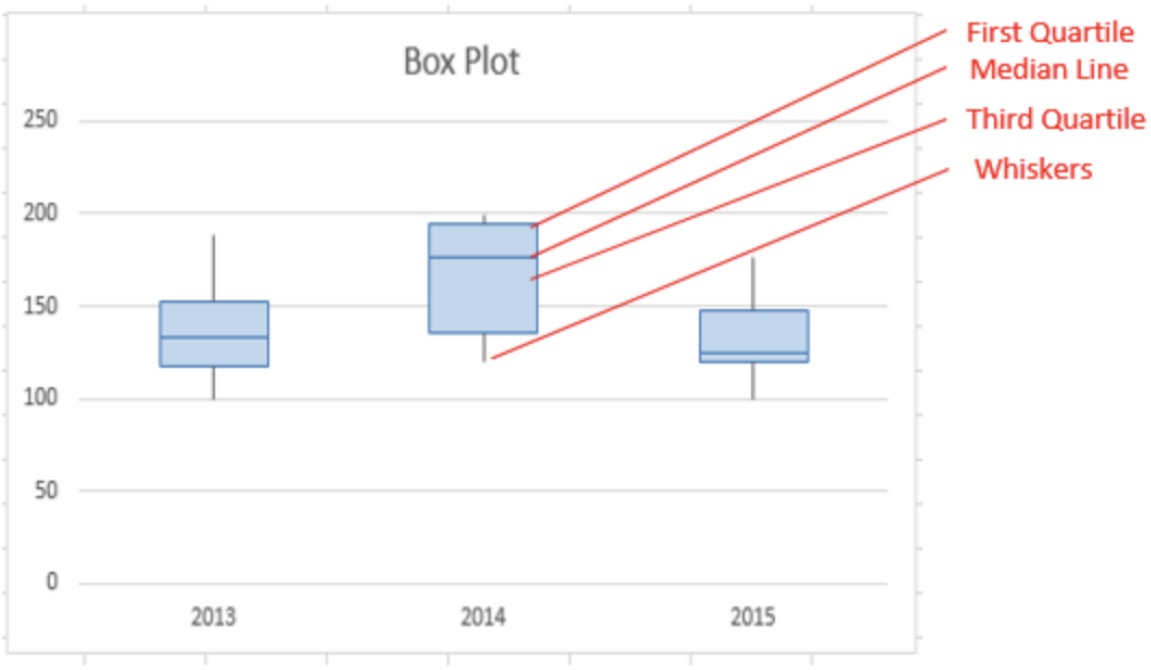
2. You will be prompted to enter the name of the benchmark you want to insert HT in.

3. Once you enter a valid benchmark name, you will be prompted to choose a strategy for generating the HTs. You can choose from four strategies: selecting rare nets that are very far from each other, selecting rare nets that are at an intermediate distance from each other, selecting rare nets that are very close to each other, or selecting rare nets randomly. You need to enter "1" for selecting the first strategy (selecting rare nets that are very far from each other), "2" for selecting the second strategy (selecting rare nets that are at an intermediate distance from each other), and so on.

4. Once you select the strategy, the script will automatically generate and evaluate 50 HTs according to your chosen strategy. This step can take a couple of minutes. Once all 50 HTs are generated and evaluated, the success rate, i.e., the percentage of HTs that evade detection, will be printed out.

5. Since there is randomness in each of the four strategies, you should run all four strategies for each benchmark multiple times and save the success rates you get for the different strategies.

6. Once you have enough trials for each strategy for each benchmark circuit, you need to plot the distribution of the success rates for each strategy for each benchmark circuit. You are free to use a plotting tool of your choice, but you should have a **box plot** that looks like this for each of the three benchmark circuits:



Here is a link that explains how to create a box plot in Excel: https://support.microsoft.com/en-us/office/create-a-box-plot-10204530-8cdf-40fe-a711-2eb9785e510f. However, you are free to use any software of your choice as long as you generate a box plot with the four strategies on the X-axis and the HT success rate on the Y-axis. Note that you need to create a separate box plot for each of the three circuits.

7. Next, you need to analyze the box plots and reason about the performance of the strategies for each of the three benchmark circuits. You should compare the three strategies in terms of their success rate as well as the complexity of implementing the strategies. Finally, you should also mention which strategy would you pick for inserting HTs and why. You need to put the box plots and your reasonings in the report document (more details about the deliverables below).

Due Date and Deliverables:

The due date to submit phase A of assignment 2 is **10/18/2023 (Wednesday) at 11:59 p.m. CT**.

You are required to submit one PDF report titled "Assignment_2A_<your_UIN>.pdf" containing the following details:

i. Box plots with four bars (one for each strategy) for each of the three benchmark circuits
ii. Detailed analysis of the performance of the four strategies for all three benchmark circuits
iii. Comparison of the four strategies in terms of their success rates and the complexity of implementing them
iv. Which strategy would you pick overall and why?

Grading Rubric:

The total number of points for Assignment 2 Phase A is 10. The points will be awarded based on the quality and the detail of the report, as well as your analyses of the results and reasoning.

If you have any questions, please feel free to contact me at gohil.vasudev@tamu.edu.