Review of ”IP Core Identification in FPGA Configuration Files using Machine Learning Techniques”

John Doe (1234567)

February 15, 2024

Summary

The paper, ”IP Core Identification in FPGA Configuration Files using Machine Learning Techniques,” by Mahmood *et al.* evaluates the effectiveness of leveraging artificial neural networks (ANN) to identify IP cores within FPGA configuration files. It introduces a synthetic dataset comprised of 1884 partial bitstreams containing logic blocks that perform add, multiply, subtract, AND, OR, XOR, NAND, NOR, XNOR and NOT operations. The study evaluates two distinct architectures, Multiple Layer Perceptrons (MLP) and Convolutional Neural Networks (CNN), in their ability to detect whether or not a given logic block performs a chosen function. CNN architectures generally outperformed the MLP architectures across the selected metrics.

Reasons to accept the paper

1. Investigates using CNN architectures to identify logic blocks in FPGA configuration files, a first in the literature.
2. Responds to a highly relevant problem related to the cybersecurity of embedded systems, providing a solution that could help protect FPGA designs from supply chain compromise.
3. Follows a robust scientific process to analyzeanalyze and compare the effectiveness of the proposed approach, yielding insight into the effectiveness of CNN for IP core identification.

Reasons to reject the paper

1. No comparison of the proposed approach with the literature is provided, limiting the reader’s understanding of the paper’s contribution.
2. Lacks results or discussion on the applicability of the proposed approach for more complex designs, limiting prospects for real-world applications.
3. Provides limited details on the dataset’s composition and generation process, raising doubts as to the reproducibility and the validity of the conclusions
4. Features many grammatical errors and poor phrase structure, hindering the paper’s readability.

Comments

The manuscript balances an innovative contribution and a strong need for improvement. Given this, a recommendation of ”**Weak Accept**” is appropriate. This recommendation is supported by the timely and relevant application of CNN architectures to FPGA security. It is countered by the lack of comparison to existing work, the need for a detailed description of the dataset and its application to more complex designs, and substantial improvements in writing quality.

Major Comments

- Introduction

1. While the paper clearly states that the motivation behind the research is to secure FPGA devices, which are increasingly used in complex and critical applications, no problem statement was provided. A problem statement explaining how FPGA devices are prone to compromise via Trojan insertion would significantly enhance the quality of the paper.
2. At the start of the introduction, there is a link between FPGAs and the medical field. However, why this link is being made needs to be clarified; this is related to the lack of a problem statement in the introduction. A statement clearly outlining the impact of an FPGA system's failure or compromise would help drive the point being made. For instance, the point would be made more evident by saying: ”Furthermore, FPGA devices are often used in critical applications that require a high level of reliability. For instance, FPGA devices are used in medical systems such as ECG and blind source separation systems. Failures and weaknesses in these systems could have significant consequences...”
3. The main contribution, as stated in the introduction, needs to be more comprehensive and representative of what readers can realistically take out from the manuscript. A more pertinent main contribution would be as follows: ”The main contribution of this paper is to evaluate the effectiveness of Convolutional Neural Networks (CNN) and Multiple Layer Perceptrons (MLP) in identifying simple logic blocks in partial bitstreams.”
4. In the paper's first paragraph, the term ”another advantage” is used. However, the first advantage needs to be clarified. Furthermore, the advantages described could be better defined as applications. For instance, the following phrase would help better define these applications: ”Two applications of this research have been identified. First, hardware inspection of bitstream would allow for the detection of known protected patents being used without a license. Second, it would also allow for untrusted IP cores being used in deployed FPGAs.”
5. The reference to the previous work (reference [9]) could be more clearly stated to establish the link with the previously published paper. For instance: ”This work builds on our previous work using the same experimental setup. Unlike the previous experiments, in this paper, we present..”

- Related Works

6. The related works section in the manuscripts does not compare the strengths or weaknesses of the various approaches. The quality of the paper would be greatly improved if qualitative or quantitative comparisons were made to the literature. Furthermore, it would be much easier to perform this comparison if the section was placed towards the end of the paper, before the conclusion.
7. Reference [13] is not cited anywhere in the paper.

- Background

8. The background section does a great job of describing the hardware structure of the FPGA device. However, it would also benefit the readers if the bitstream generation process was explained.

- Inspection of Partial Bitstreams

9. In the second paragraph of Section 4, the hardware modules are listed, and five combinations are presented. As a reader, it is unclear why these particular combinations were selected. Is this a random choice, or is there a reason behind the selections?
10. Has the impact of constraining the design to a different bank within the FPGA been investigated? The location of the logic within the FPGA fabric could considerably impact the results depending on the features the model learns.
11. Could the need for zero-padding be bypassed by constraining the design to a consistently sized region? Such an approach would ensure a maximum of fidelity with the original bitstream.
12. Considering that the bitstreams used in the dataset are mainly composed of 0's, how would generating ”noisy” bitstreams, with lots of surrounding logic, affect the results?
13. The configuration bits for a given logic block are generally distributed over several frames. Thus, relevant bits are not necessarily adjacent to each other. Does compression remove redundant zeroes and ones and thus potentially affect some of the features learned by the model?
14. In Section 4, the use of optimizers to diversify the implementation of the logic is presented. Intuitively, with the dataset mainly composed of elementary circuits, the optimizers' impact on diversity needs to be further questioned. A metric to show how different the bitstreams used in training are from one another would be helpful.

   15. The manuscript states that a small learning rate will result in more iterations before a minima is reached, but is that the only impact? Also, an explanation needs to be provided regarding the selection of the batch size, learning rate, number of layers, or the number of neurons used.

- Evaluation

   16. In Section 5, an image analogy is introduced to justify using partial bitstreams instead of complete bitstreams. A reference to this analogy would be appreciated. Similarly, in a couple of sentences below, object detection is discussed. A reference would be helpful here as well.

- Conclusion

   17. In the conclusion, no clear qualitative or quantitative terms are used to describe the conclusions drawn in the manuscript. For instance, based on the results, a batch size of 100 yields the best results, and this could be an excellent metric to use in the conclusions. Also, a large portion of the interpretation is left to the reader. For instance, what can be concluded from this batch size?
   18. Towards the end of your conclusion, you mention factors affecting the performance of a machine learning algorithm; however, it is unclear what these factors are. A definition of these factors would be insightful.

Minor Comments

- Introduction
   - (Para. 1, line 2): Change ”which allow” to ”which allows”
   - (Para. 1, line 8): Change ”certain” to ”specific”
   - (Para. 1, line 22-23): Change ”to real-life image” to ”to a real-life image”
   - (Para. 2, line 2): Use ”prove” instead of ”proof”
   - (Para. 2, line 17): Change ”have been” to ”were”
- Related Works
   - (Para. 1, line 1): Change ”that conducts analysis of” to ”to analyze”
   - (Para. 1, line 6): Change ”performs analysis on” to analyzes”
   - (Para. 1, line 18): Change ”correspond” to ”corresponds”
   - (Para. 2, line 1): Change ”presents” to ”present”
   - (Para. 2, line 2): Change ”on Bil tool” to ”on the Bil tool”
   - (Para. 2, line 4): Change ”by means of Bil” to ”using Bil”
   - (Para. 2, line 8): Change ”proof” to ”prove”
   - (Para. 2, line 11): Change ”to data that” or ”to detect data that”
   - (Para. 3, line 2): Change ”and its contents” to ”and their contents”
   - (Para. 4, line 10): Change ”analyzes” to ”analyze”
- Background
   - (Para. 1, line 3): Change ”correspond” to ”corresponds”
   - (Para. 1, line 8): Change ”classsification” to ”classification”
   - (Para. 2, line 6): Change ”to Sigmoid function” to ”to the Sigmoid function”
   - (Para. 3, line 8): Change ”, hidden layer” to ”, a hidden layer”
   - (Para. 3, line 14): Change ”flatten” to flattened”
   - (Para. 6, line 11): Remove space after ”FPGA”
- Inspection of Partial Bitstreams
   - (Para. 2, line 6): Change ”data set” to ” test datasets”

  - (Para. 3, line 2): Change ”previous” to ”previously”
  - (Para. 3, line 11): Change ”arrangement” to ”arrangements”
  - (Para. 3, line 16): Change ”loosing” to ”losing”
  - (Para. 6, line 3): Change ”represents” to ”represent”
  - (Para. 12, line 1): Add space between ”N55:” and ”The”
  - (Para. 13, line 2): Change ”that involves” to ”, which involves”
  - (Para. 13, line 4): Repeated Phrase
  - (Para. 15, line 5): Open sentence
  - (Para. 16, line 2): Change ”Small” to ”A small”
  - (Para. 18, line 1): Change ”activation” to ”an activation”

- Evaluation
  - (Para. 1, line 3-19): Figure 3 overlap with the text on the left column of the manuscript.
  - (Para. 1, line 3): Change ”The used metrics” to ”The metrics used”
  - (Para. 1, line 17): Change ”the precision” to ”precision”
  - (Para. 1, line 18): Change ”Equation 4 the definition” to ”Equation 4 shows the definition”
  - (Para. 1, line 19): CapitalizeCapitalize ”python” and ”tensorflow”
  - (Para. 2, line 4): Add space between ”IV” and ”shows”
  - (Para. 2, line 4-5): Change ”both training” to ”both the training”
  - (Para. 4, line 4): Change ”the better” to ”better”
  - (Para. 4, line 5): Change ”the tables” to ”tables”
  - (Para. 6, line 8): Change ”case” to ”the case”
  - (Para. 6, line 13): Change ”were having” to ”had”
  - (Para. 6, line 16): Change ”show” to ”shows”
  - (Para. 7, line 2): Change ”as classification model” to ”as a classification model”
  - (Para. 7, line 10): CapitalizeCapitalize ”trojans”

- Conclusion
  - (Para. 1, line 2): Change ”that identification” to ”that the identification”
  - (Para. 1, line 4): Replace the phrase ”fairly possible” with a quantitative or qualitative term.
  - (Para. 1, line 7): Change ”Convolutional Neural Network” to ”Convolutional Neural Networks”
  - (Para. 1, line 17): Change ”Use of dynamic” to ”The use of dynamic”
  - (Para. 1, line 21-22): Change ”this work is empirical” to ”this work are empirical”
  - (Para. 1, line 22): Change ”be showing” to ”show”
  - (Para. 2, line 5): Change ”between size” to ”between the size”