# REGISTER-TRANSFER-LEVEL HARDWARE TROJAN CLASSIFICATION BOOSTED WITH GATE-LEVEL FEATURES

CHOO HAU SIM

UNIVERSITI TEKNOLOGI MALAYSIA

# REGISTER-TRANSFER-LEVEL HARDWARE TROJAN CLASSIFICATION BOOSTED WITH GATE-LEVEL FEATURES

CHOO HAU SIM

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

Malaysia-Japan International Institute of Technology
Universiti Teknologi Malaysia

DECEMBER 2022

## DEDICATION

This thesis is dedicated to my parents who provide me with relentless support and strength in every decision that I have made in my life.

**ACKNOWLEDGEMENT**

In preparing this thesis, I have received a lot of guidance and support. I would first like to thank my main supervisor, A.P. Dr. Ooi Chia Yee, for her guidance and insightful feedback through each stage of conducting the research. I would like to thank my co-supervisor, Mdm. Nordinah binti Ismail, for offering valuable advice and support with her professional experience. I would also like to thank Prof. Dr. Michiko Inoue for providing her advice and comments on my research and all the supports during my visit to Nara Institute of Science and Technology. In addition, I would like to express my gratitude to Prof. Dr. Koichiro Mashiko for sharing his useful opinion and experience in the area.

I am thankful to Malaysia-Japan International Institute of Technology for funding my study.

Lastly, I want to show my gratitude to my friends for all the support and fun while we are pursuing our postgraduate studies.

.

**ABSTRACT**

Hardware Trojan (HT) is an alarming hardware security threat which has gained increased awareness over the last decade. Due to the emerging threat of HT, ensuring trustworthiness in an integrated circuit (IC) has become an important aspect to be considered during manufacturing. Hence, the design process of ICs must be reviewed to avoid HT insertion by malicious third-party vendor. The purpose of this research is to develop a HT detection method in register-transfer-level (RTL) description with an improved HT coverage compared to the other previously proposed methods. The proposed method discovered HT branching statement in the RTL description by utilising a supervised machine learning classifier based on ten (10) proposed two-abstraction-level features. The proposed two-abstraction-level features relevant to HT characteristics included branching probability features extracted at RTL and net testability features extracted at gate-level (GL). The effectiveness of the proposed features in detecting HTs with 19 Trust-Hub benchmark circuits were demonstrated. The Minimum Redundancy Maximum Relevance (mRMR) feature selection algorithm was utilised to prove that the combination of the proposed features can achieve maximum accuracy (ACC) of 99.97% in detecting HTs during classifier training. To avoid overfitting issue, the trained classifiers were further evaluated with a classifier testing experiment on unseen circuit. The unseen circuit was completely independent of the training data, and it consisted of 24 HT circuits derived from a genuine Keccak encryption circuit. By using a set of proposed HT stealthiness assessment measures, the HT coverage of the classifiers was evaluated. The decision tree (DT) classifier with the two-abstraction-level features achieved the highest 87.5% HT coverage with 81.25% true positive rate (TPR), 88.44% true negative rate (TNR), and 88.24% ACC respectively. The results proved that the two-abstraction-level features outperformed single-abstraction-level features with higher HT detection coverage.

# ABSTRAK

Perkakasan Trojan (*HT*) adalah ancaman sekuriti litar bersepadu (*IC*) yang membimbangkan dan perhatian berkaitan dengannya telah meningkat selama sedekad kebelakangan ini. Disebabkan ancaman *HT* ini, pemastian kebolehpercayaan telah menjadi aspek penting untuk dipertimbangkan semasa pembuatan *IC*. Proses reka bentuk *IC* mesti disemak semula untuk mengelakkan *HT* disisip masuk oleh pihak ketiga yang tidak bertanggungjawab. Tujuan penyelidikan ini adalah untuk membangunkan kaedah pengesanan *HT* bagi deskripsi di peringkat pemindahan daftar (*RTL*) bagi liputan *HT* yang lebih baik berbanding dengan kaedah sebelumnya. Bagi mengesan pernyataan bercabang *HT* dalam deskripsi RTL, satu kaedah yang menggunakkan pengelas pembelajaran mesin tersedia dicadangkan di mana ia berasaskan 10 ciri abstraksi-dua peringkat. Ciri-ciri *HT* yang dicadangkan termasuk kebarangkalian percabangan yang diekstrak pada peringkat *RTL* dan ukuran testabiliti yang diekstrak pada peringkat get logik. Keberkesanan ciri-ciri tersebut telah dikenalpasti dengan menggunakan 19 litar tanda aras yang terdapat di dalam pangkalan data hab-kepercayaan. Algoritma pemilihan ciri Lebihan Minima Perkaitan Maksima (*mRMR*) digunakan untuk membuktikan bahawa gabungan ciri-ciri yang dicadangkan boleh mencapai ketepatan (*ACC*) maksimum sebanyak 99.97% semasa latihan pengelas. Untuk mengelakkan isu pemasangan limpahan, pengelas-pengelas terlatih diuji dengan data baru yang berbeza sepenuhnya daripada data latihan, dan ia terdiri daripada 24 litar *HT* yang diperolehi daripada litar penyulitan *Keccak* tulen. Nilai liputan *HT* kemudianya diukur dengan menggunakan langkah-langkah penilaian kesembunyian. Pengelas pokok keputusan (*DT*) dengan ciri-ciri abstraksi-dua peringkat telah mencapai liputan *HT* tertinggi iaitu 87.5% berbanding dengan kaedah dulu, dengan 81.25% kadar positif yang benar (*TPR*), 88.44% kadar negatif yang benar (*TNR*), dan 88.24% *ACC*. Hasil kajian ini menunjukkan bahawa penggunaan ciri-ciri abstraksi-dua peringkat mengatasi ciri-ciri abstraksi-satu peringkat dengan capaian liputan *HT* yang lebih tinggi.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| ACC | - | Accuracy |
| ADASYN | - | Adaptive Synthetic Sampling |
| ATPG | - | Automatic Test Pattern Generation |
| CC0 | - | Combinational 0-Controllability |
| CC1 | - | Combinational 1-Controllability |
| CO | - | Combinational Observability |
| COTD | - | Controllability And Observability Trojan Detection |
| DT | - | Decision Tree |
| EDA | - | Electronic Design Automation |
| ELM | - | Extreme Learning Machine |
| EMF | - | Electromagnetic Field |
| GL | - | Gate Level |
| HDL | - | Hardware Description Language |
| HT | - | Hardware Trojan |
| IC | - | Integrated Circuit |
| IP | - | Intellectual Property |
| k-NN | - | *k*-Nearest Neighbor |
| mRMR | - | Minimum Redundancy Maximum Relevance |
| RTL | - | Register-Transfer Level |
| SC0 | - | Sequential 0-Controllability |
| SC1 | - | Sequential 1-Controllability |
| SCOAP | - | Sandia Controllability/Observability Analysis Program |
| SO | - | Sequential Observability |
| SVM | - | Support Vector Machine |
| TNR | - | True Negative Rate |
| TPR | - | True Positive Rate |

## LIST OF SYMBOLS

| | | |
|---|---|---|
| $I$ | - | Mutual Information |
| $p$ | - | Probabilistic Density Function |

## LIST OF APPENDICES

# CHAPTER 1

# INTRODUCTION

## 1.1 Research Background

Trojan is a term that usually refers to a malicious module that conceals its content while deviously modifies the operation of a system. In the semiconductor industry, a term "hardware Trojan" (HT) was coined to refer to a specific Trojan that exists in a form of a circuit that is secretly inserted into an integrated circuit (IC) and attempts to attack the system at which the HT resides (Tehranipoor & Koushanfar, 2010). The HT can launch harmful attacks such as changing the circuit operation, leaking the critical information, degrading the circuit performance, and a Denial-of-Service attack (Salmani *et al.*, 2013). The HT effect could be activated upon receiving a specific signal, or upon system power up.

The presence of HT causes the trustworthiness issue of ICs. Due to the rapid advancement of the semiconductor industry, the market demand for sophisticated IC is rising. To reduce the production's cost, outsourcing the production task is a common practice for semiconductor companies (Bhunia *et al.*, 2014). For instance, the companies would outsource part of their circuit design to a third-party intellectual property vendor. The companies focusing on circuit design business would outsource their circuit fabrication to other fabrication factory. This business model poses a security threat in which some untrusted third parties may have a chance to be involved in the supply chain and attempt to modify the circuit design. Furthermore, the increasing complexity of circuit allows the HT insertion with less effort because the HT can be stealthier in a large circuit and escape from detection more easily during conventional verification and testing (Cruz *et al*., 2018).

HT is gaining public awareness due to the emerging 5G and Internet-of Thing technology in which the digital devices around us are connected to each other and to

internet, in which our personal data could be stored. Without proving the trustworthiness of the IC, we cannot ensure our data privacy. Our devices could be vulnerable to HT attacks that may steal our personal data or kill the circuit operation to unexpectedly interrupt the system. Besides, HT also poses a threat to any system which is controlled by an IC, including military system, transportation system, healthcare system, and so on. In 2007, a critical failure of Syria's radar systems was reported to be the cause of incoming missiles detection failure (Adee, 2008). The experts suspected that the system failure was intentionally triggered by using a kill-switch which was secretly built into the microprocessors by the circuit manufacturer. Besides, there have been other rumors of secret HT insertion into ICs during the manufacturing process (Robertson & Riley, 2018). To resolve the HT threat, Toshiba Information Systems (Japan) Corporation (2020) have announced a new service of HT detection to their customers. These HT reports have drawn attention from academia, industry, and government over the past decade.

To assist in HT research, a few categorizations of HTs have been introduced. Xiaoxiao *et al.* (2008) introduced the first HT taxonomy which was later refined and expanded by other researchers (Karri *et al.*, 2010; Rajendran *et al.*, 2010; Tehranipoor and Koushanfar, 2010; Salmani *et al.*, 2013). The most widely referred HT categorization method is the one proposed by Salmani *et al.* (2013). The details of the HT taxonomy are illustrated in Figure 1.1. The authors have constructed a HT benchmark library based on the six attributes as in the proposed HT taxonomy: insertion phase, abstraction level, activation mechanism, effect, location, and physical characteristics.

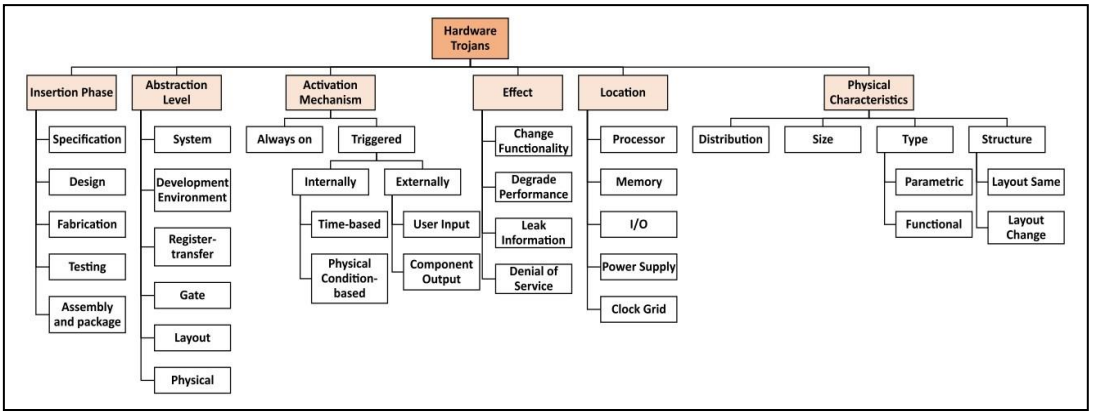


Figure 1.1 Taxonomy of Hardware Trojan (Source: Shakya *et al*., 2017).

One of the HT countermeasures is detection. We cannot apply only single approach to detect all HTs because of the diversity of HT types. Many HT detection methods have been introduced with each of them having different motivations. These HT detection approaches can be categorized based on their detection level. The detection level refers to the abstraction levels of the IC manufacturing process at which the detection approach is applied, as illustrated in Figure 1.2.

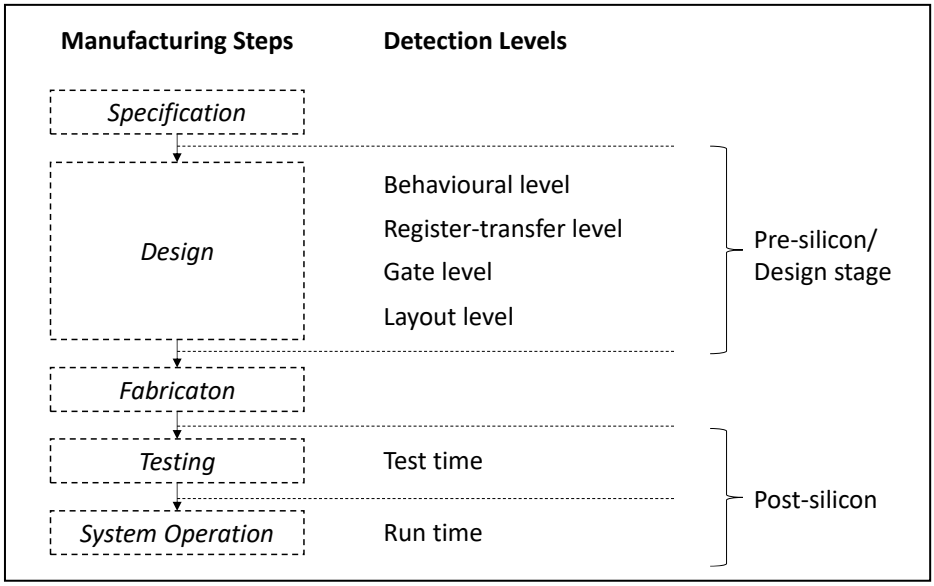


Figure 1.2 Detection levels corresponding to manufacturing process.

In typical IC design process, the abstraction levels involved are behavioral level, register-transfer level (RTL), gate level (GL) and layout level. The detection methods conducted at one of these levels are considered as pre-silicon detection. For instance, Bang *et al.* (2010) introduced a HT detection method by utilizing the

conventional verification techniques at GL. The method only requires the circuit netlist, which is simply a list of coding, instead of the physical circuit. Such method is known as the pre-silicon detection method which can identify HT at early manufacturing process. After the circuit is fabricated, post-silicon HT detection can be applied at test time and run time. For example, Jin *et al.* (2008) proposed a test-time circuit authentication method by analyzing the path delay parameter. The required information is extracted from the fabricated circuit. The method is considered as post-silicon detection which can conclusively authenticate the fabricated circuit.

Among the previously proposed detection methods, the machine-learning-based approach is relatively popular, especially for pre-silicon detection. Its self-learning ability helps in building an analytical model without an explicit programming. The model can be even expanded by fitting it with a larger database. Due to these reasons, machine learning can reduce the effort of HT detection. A test-time circuit authentication method utilizing a machine learning approach was suggested by Bao *et al.* (2014). The authors successfully developed a HT circuit classifier based on the scanning electron microscope image of the fabricated circuit. The first pre-silicon machine-learning-based HT detection method was introduced by Hasegawa *et al.* (2016). The authors suggested to classify each net in the GL netlist based on their structural measures during the design stages.

HT detection is preferable to be conducted during pre-silicon stage due to the difficulty of HT removal from a physically fabricated circuit. To prevent unnecessary investment in an infected circuit design, it is always better to identify the HT before the circuit design is mapped into later stages of design process. In addition, the information contained in the circuit design at every stage is different, it could be harder to trace the HT when the circuit design undergoes more mappings at lower abstraction levels. Due to these reasons, RTL is an early design stage which is suitable for conducting HT detection.

## 1.2 Problem Statement

Most of the previous RTL HT detection research focused on detecting the HT by examining the circuit connectivity and circuit operation. These approaches may require additional structural analysis or simulation. Moreover, HT detection at RTL is less explored compared to the other abstraction levels such as GL and layout level, especially the machine-learning-based approaches. At RTL, the circuit functionality is described. This information is difficult to be converted into numerical data which is required to be used as the input features for machine learning classifier.

The purpose of pre-silicon HT detection is to identify any suspicious component or circuit part, and to stop the infected circuit from undergoing further mapping. Therefore, the sensitivity of the classifier towards HTs is always emphasized to increase the possibility of successful detection of HT. We always want to use a detection approach with high HT coverage to cover as many types of HTs as possible. However, this is not a trivial task due to the diversity of HT types. Forcing a machine learning classifier to learn the features of all HTs is not a good move because the machine learning classifier will probably result in overfitting whereby the classifier is biased towards the training data and not robust against unseen data, especially when the input information or features are not enough. Although we could just use different detection approaches to detect different types of HTs, extending an existing detection approach to increase its detection performance would be a better option that possibly gives a better overall result with less overhead.

Since machine-learning-based approaches may be subject to overfitting issues, classifier testing is important to determine the classifier performance against unseen data. However, most of the previous works did not conduct a proper classifier testing which used a truly unseen dataset. Besides, by just looking at the classification accuracy, we are unable to precisely tell the classifier performance against specific HTs, and thus the HT coverage is unknown. Currently, there is no suitable performance metric that could assist in HT research to evaluate HT detection methods in terms of HT coverage.

## 1.3 Research Objectives

To resolve the problems as stated, the objectives for this research are established as follows:

(a) To propose new features of branching statement in RTL description based on branching probability and net testability, to develop a machine-learning classifier with high accuracy in classifying HTs.

(b) To engineer the HT-relevant features at different abstraction levels, specifically RTL and GL, in order to achieve higher HT detection coverage.

(c) To introduce a set of HT assessment measures to describe the HT stealthiness to assist in HT design and classifier testing in terms of HT coverage and detection accuracy against unseen data.

## 1.4 Research Scope

This research proposes a HT detection method utilizing supervised machine learning and two-abstraction-level features. The proposed features are extracted at RTL and GL. The proposed method can classify suspicious HT branching statements in the RTL description. Our proposed detection method is only effective to the HTs that fulfil our threat model assumption. For the analysis of circuit signal, we consider logic values ‘0’ and ‘1’.

Since our main purpose here is to demonstrate the effectiveness of the features with supervised machine learning, the choices of the algorithms are not emphasized. Three well-known supervised machine learning algorithms for binary classification are used in this research, which are decision tree (DT), k-nearest neighbor (k-NN) and support-vector machine (SVM).

We train the proposed machine learning classifier by utilizing the HT benchmark circuits provided by an open-source library, Trust-Hub (Salmani *et al.*, 2013). We collected the RTL descriptions, which are written in Verilog hardware description language (HDL), of 19 HT circuits from the library. Since the number of training data is reasonably large (>10,000), we apply *k*-fold cross validation with $k = 10$ to evaluate the trained classifiers. A set of self-designed circuits based on Keccak (SHA-1) encryption circuit is used for classifier testing to determine the detection performance against unseen circuit. The performance metrics analyzed in this research are accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and HT detection coverage.

## 1.5 Significance of the Study

This research proposes a HT detection method which can detect HT hiding in RTL description and help to increase the trustworthiness of RTL digital circuit design. The proposed method is a pre-silicon detection method which can identify HT at a very early design stage and before the circuit is fabricated, and thus prevents unnecessary investment wasted in an infected circuit. The proposed method is based on a supervised machine learning approach which has a self-learning ability and is expendable with a larger HT database. This can ease the HT detection effort to tackle known HTs in the future. Besides, this research also introduces a set of HT stealthiness assessment measures. It can assist in HT research by systematically categorizing HTs.

## 1.6 Thesis Outline

This thesis consists of five chapters which are organized as follows. In Chapter 2, previous studies related to this research will be reviewed and discussed. In Chapter 3, we describe our proposed supervised machine learning classifier for HT detection. The detail of the classifier development will be clearly explained. In Chater 4, the experimental results will be analyzed and discussed to show the performance of our proposed solution. The comparison between our proposed method and previous studies

will also be discussed to demonstrate the improvement achieved by our method. Lastly, in Chapter 5, we summarize this research and discuss the potential extensions of this research.

## REFERENCES


Adee, S. (2008). The Hunt For The Kill Switch. *IEEE Spectrum*, *45*(5), 34–39.

Banga, M., & Hsiao, M. S. (2010). Trusted RTL: Trojan detection methodology in pre-silicon designs. *Proceedings of the 2010 IEEE International Symposium on Hardware-Oriented Security and Trust, HOST 2010*, 56–59.

Bao, C., Forte, D., & Srivastava, A. (2014). On application of one-class SVM to reverse engineering-based hardware Trojan detection. *Proceedings - International Symposium on Quality Electronic Design, ISQED*, 47–54.

Bao, C., Xie, Y., Liu, Y., & Srivastava, A. (2016). On Reverse Engineering-Based Hardware Trojan Detection. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *35*(1), 49–57.

Bazzazi, A., Manzuri Shalmani, M. T., & Hemmatyar, A. M. A. (2017). Hardware Trojan Detection Based on Logical Testing. *Journal of Electronic Testing: Theory and Applications (JETTA)*, *33*(4), 381–395.

Bhunia, S., Hsiao, M. S., Banga, M., & Narasimhan, S. (2014). Hardware trojan attacks: Threat analysis and countermeasures. *Proceedings of the IEEE*, *102*(8), 1229–1247.

Chen, X., Liu, Q., Yao, S., Wang, J., Xu, Q., Wang, Y., Liu, Y., & Yang, H. (2018). Hardware Trojan Detection in Third-Party Digital Intellectual Property Cores by Multilevel Feature Analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *37*(7), 1370–1383.

Cruz, J., Farahmandi, F., Ahmed, A., & Mishra, P. (2018). Hardware trojan detection using ATPG and model checking. *Proceedings of the IEEE International Conference on VLSI Design*, 91–96.

Deepthi, S., Ramesh, S. R., & Nirmala Devi, M. (2021). Hardware Trojan Detection using Ring Oscillator. *Proceedings of the 6th International Conference on Communication and Electronics Systems, ICCES 2021*, 362–368.

Goldstein, L. H., & Thigpen, E. L. (1980). SCOAP: Sandia Controllability/Observability Analysis Program. *17th Design Automation Conference*, 192–196.

Han, T., Wang, Y., & Liu, P. (2019). Hardware trojans detection at register transfer level based on machine learning. *Proceedings - IEEE International Symposium on Circuits and Systems*, 19–23.

Hasegawa, K., Oya, M., Yanagisawa, M., & Togawa, N. (2016). Hardware Trojans classification for gate-level netlists based on machine learning. *2016 IEEE 22nd International Symposium on On-Line Testing and Robust System Design, IOLTS 2016*, 203–206.

Hasegawa, K., Yanagisawa, M., & Togawa, N. (2017a). Hardware Trojans classification for gate-level netlists using multi-layer neural networks. *2017 IEEE 23rd International Symposium on On-Line Testing and Robust System Design, IOLTS 2017*, 227–232.

Hasegawa, K., Yanagisawa, M., & Togawa, N. (2017b). Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier. *Proceedings - IEEE International Symposium on Circuits and Systems*.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *Proceedings of the International Joint Conference on Neural Networks*, 1322–1328.

Inoue, T., Hasegawa, K., Kobayashi, Y., Yanagisawa, M., & Togawa, N. (2018). Designing subspecies of hardware trojans and their detection using neural network approach. *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, 1–4.

Inoue, T., Hasegawa, K., Yanagisawa, M., & Togawa, N. (2017). Designing hardware trojans and their detection based on a SVM-based approach. *Proceedings of International Conference on ASIC*, 811–814.

Jap, D., He, W., & Bhasin, S. (2016). Supervised and unsupervised machine learning for side-channel based Trojan detection. *Proceedings of the International Conference on Application-Specific Systems, Architectures and Processors*, 17–24.

Jin, Y., & Makris, Y. (2008). Hardware Trojan detection using path delay fingerprint. *2008 IEEE International Workshop on Hardware-Oriented Security and Trust, HOST*, 51–57.

Jin, Y., Maliuk, D., & Makris, Y. (2012). Post-deployment trust evaluation in wireless cryptographic ICs. *Proceedings -Design, Automation and Test in Europe, DATE*, 965–970.

Kok, C. H., Ooi, C. Y., Inoue, M., Moghbel, M., Baskara Dass, S., Choo, H. S., Ismail, N., & Hussin, F. A. (2019). Net Classification Based on Testability and Netlist Structural Features for Hardware Trojan Detection. *Proceedings of the Asian Test Symposium*, *2019-Decem*, 105–110.

Kumar, P., & Srinivasan, R. (2014). Detection of hardware Trojan in SEA using path delay. *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2014*.

Kurihara, T., & Togawa, N. (2021). Hardware-trojan classification based on the structure of trigger circuits utilizing random forests. *Proceedings - 2021 IEEE 27th International Symposium on On-Line Testing and Robust System Design, IOLTS 2021*, 31–34.

Li, J., Ni, L., Chen, J., & Zhou, E. (2016). A novel hardware Trojan detection based on BP neural network. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 2790–2794.

Lodhi, F. K., Hasan, S. R., Hasan, O., & Awwadl, F. (2017). Power profiling of microcontroller's instruction set for runtime hardware Trojans detection without golden circuit models. *Proceedings of the 2017 Design, Automation and Test in Europe, DATE 2017*, 294–297.

Lu, R., Shen, H., Feng, Z., Li, H., Zhao, W., & Li, X. (2021). HTDet: A clustering method using information entropy for hardware Trojan detection. *Tsinghua Science and Technology*, *26*(1), 48–61.

Ngo, X. T., Danger, J. L., Guilley, S., Najm, Z., & Emery, O. (2015). Hardware property checker for run-time Hardware Trojan detection. *2015 European Conference on Circuit Theory and Design, ECCTD 2015*, 1–4.

Nguyen, L. N., Yilmaz, B. B., Prvulovic, M., & Zajic, A. (2020). A Novel Golden-Chip-Free Clustering Technique Using Backscattering Side Channel for Hardware Trojan Detection. *Proceedings of the IEEE International Symposium on Hardware Oriented Security and Trust, HOST 2020*, 111–121.

Peng, H., Long, F., & Ding, C. (2005). Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-

Redundancy. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *27*(8), 1226–1238.

Rajendran, S., & R, M. L. (2021). A Novel Algorithm for Hardware Trojan Detection through Reverse Engineering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *0070*(c).

Robertson, J., & Riley, M. (2018). The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies. Bloomberg Businessweek. Retrieved January 2, 2022, from https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies

Salmani, H. (2017). COTD: Reference-Free Hardware Trojan Detection and Recovery Based on Controllability and Observability in Gate-Level Netlist. *IEEE Transactions on Information Forensics and Security*, *12*(2), 338–350.

Salmani, H., Tehranipoor, M., & Karri, R. (2013). On Design Vulnerability Analysis and Trust Benchmarks Development. *2013 IEEE 31st International Conference on Computer Design (ICCD)*, 471–474.

Tehranipoor, M., & Koushanfar, F. (2010). A survey of hardware trojan taxonomy and detection. *IEEE Design and Test of Computers*, *27*(1), 10–25.

Toshiba Information Systems (Japan) Corporation. (2020). *ハードウェアトロイ検出ツール 「Htfinder]*. Retrieved January 2, 2022, from https://www.tjsys.co.jp/english/lsi/index.htm

Wang, C., Cai, Y., & Zhou, Q. (2017). Automatic security property generation for detecting information-leaking hardware trojans. *Proceedings - 35th IEEE International Conference on Computer Design, ICCD 2017*, 321–328.

Wang, C., Li, J., Yu, M., & Wang, J. (2013). An intelligent classification method for Trojan detection based on side-channel analysis. *IEICE Electronics Express*, *10*(17), 1–6.

Wang, S., Dong, X., Sun, K., Cui, Q., Li, D., & He, C. (2016). Hardware Trojan detection based on ELM neural network. *2016 1st IEEE International Conference on Computer Communication and the Internet, ICCCI 2016*, 400–403.

Wang, X., Tehranipoor, M., & Plusquellic, J. (2008). Detecting malicious inclusions in secure hardware: Challenges and solutions. *2008 IEEE International Workshop on Hardware-Oriented Security and Trust, HOST*, 15–19.

Xiao, K., Forte, D., Jin, Y., Karri, R., Bhunia, S., & Tehranipoor, M. (2016). Hardware trojans: Lessons learned after one decade of research. *ACM Transactions on Design Automation of Electronic Systems*, *22*(1).

Xie, X., Sun, Y., Chen, H., & Ding, Y. (2017). Hardware trojans classification based on controllability and observability in gate-level netlist. *IEICE Electronics Express*, *14*(18), 159–172.

Xue, M., Wang, J., & Hux, A. (2017). An enhanced classification-based golden chips-free hardware Trojan detection technique. *Proceedings of the 2016 IEEE Asian Hardware Oriented Security and Trust Symposium, AsianHOST 2016*.

Yao, S., Chen, X., Zhang, J., Liu, Q., Wang, J., Xu, Q., Wang, Y., & Yang, H. (2015). FASTrust: Feature analysis for third-party IP trust verification. *Proceedings - International Test Conference*, 1–10.

Yasaei, R., Yu, S. Y., & Al Faruque, M. A. (2021). GNN4TJ: Graph Neural Networks for Hardware Trojan Detection at Register Transfer Level. *Proceedings -2021 Design, Automation and Test in Europe, DATE 2021*, 1504–1509.

Zhang, X., & Tehranipoor, M. (2011a). Case study: Detecting hardware Trojans in third-party digital IP cores. *2011 IEEE International Symposium on Hardware-Oriented Security and Trust, HOST 2011*, 67–70.

Zhang, X., & Tehranipoor, M. (2011b). RON: An on-chip ring oscillator network for hardware Trojan detection. *Proceedings -Design, Automation and Test in Europe, DATE*.

## LIST OF PUBLICATIONS

### Indexed Journal

1. **Choo, H. S.,** Ooi, C. Y., Inoue, M., Ismail, N., & Kok, C. H. (2020). 'A Review of Hardware Trojan Detection: An Overview of Different Pre-Silicon Techniques', *Defense S&T Technical Bulletin*, *13*(1), 1-21. **(Indexed by SCOPUS)**

2. **Choo, H. S.,** Ooi, C. Y., Inoue, M., Ismail, N., Moghbel, M., & Kok, C. H. (2020). 'Register-Transfer-Level Features for Machine-Learning-Based Hardware Trojan Detection', *IEICE Trans. Fundamentals*, *E103-A*(2), 502-509. **(Indexed by Web of Science)**

### Indexed Conference Proceedings

1. **Choo, H. S.,** Ooi, C. Y., Inoue, M., Ismail, N., Moghbel, M., Dass, S. B. & Kok, C. H. (2019). 'Machine-Learning-Based Multiple Abstraction-Level Detection of Hardware Trojan Inserted at Register-Transfer Level', *Defense S&T Technical Bulletin*, *12*(1), 61-78. **(Indexed by Web of Science)**