In Confidence

Professor Josef Pieprzyk, PhD, DSc, IACR Fellow
Senior Principal Research Scientist
Data61, CSIRO
Distributed Systems Security
Cnr Vimiera and Pembroke Roads
Marsfiled, NSW 2122
PO Box 76, Epping NSW 1710, Australia
email: *josef.pieprzyk@csiro.au*

PhD Thesis Report

August 13, 2024

Thesis Title: **Provable Countermeasures against Hardware Trojans**
Author: *Małgorzata Gałązka*

Overview

The focus of this PhD study is on the security of algorithms implemented in hardware. Specifically, it addresses security countermeasures against malicious adversaries who tamper with hardware implementations. The security model examined in this work considers both intentional tampering by adversaries and random hardware faults. These two types of adversarial activities are collectively referred to as hardware Trojan horses (HTHs).

Hardware security is becoming an increasingly critical aspect of overall IT infrastructure protection. Its strategic importance is emphasized by the growing globalization of technology, where complex computing environments are assembled from hardware components supplied by various manufacturers with vastly different levels of trustworthiness. Even when hardware comes from a fully trusted source, it must still be tested to ensure it operates correctly and is free from production faults. Furthermore, when hardware operates in unprotected or adversarial environments, its security must be regularly verified to ensure that it continues to function as intended.

It's important to note that this vital area of security has lagged behind other developments, such as cryptography. As a result, there is a limited number of research publications addressing hardware security. This thesis makes a significant contribution to this crucial yet often overlooked area of research.

The thesis is organized as follows: Section 1 establishes the context and motivation for the study. Section 2 presents the security model and introduces key definitions. Section 3 outlines protection schemes against restricted HTHs. Finally, Section 4 describes simple compilers designed to verify hardware functionality against total hardware Trojan horses (THTHs).

Originality and Contribution to Knowledge of Discipline

Sections 3 and 4 present the author's original contributions. In particular, Section 3 focuses on the functionality verification of a hardware component (also known as a circuit) to protect it against

hardware Trojan horses (HTHs). The main idea is to modify the original circuit with functionality F by adding extra circuitry (gates) and additional control input/output wires. The resulting circuit, with the extended functionality $\hat{F}$, allows for the detection of HTH interference while preserving the basic functionality F when the additional input/output wires are set to zero.

The author has proposed two solutions, $\mathfrak{L}$ and $\mathfrak{R}$. In essence, solution $\mathfrak{L}$ is deterministic, while $\mathfrak{R}$ is probabilistic. Table 2 compares the two solutions and provides details about their parameters. Section 3 is based on two papers published at the ITCS 2023 and TCC 2023 conferences, referred to in the manuscript as [Bai+23a] and [Bai+23b], respectively. Both conferences are highly regarded and have stringent review processes. According to the Polish Ministry of Science, both conferences are valued at 140 points in the 2024 ranking. It is worth noting that these contributions are primarily theoretical, as acknowledged by the author. The main challenge in implementing these ideas is the lack of direct access to the internal structure of the original and protected circuits.

Section 4 explores a more practical approach to HTH security. The approach is similar to methods already used in critical flight hardware to detect faults. A device is duplicated, and the hardware is considered fault-free if all outputs generated by the duplicate devices are the same. The solution proposed in this section works in two phases: testing and deployment. In the testing phase, untrusted (Trojan) circuits are tested against a fully trusted and fault-free circuit. In the deployment phase (also called the wild phase), all untrusted circuits are tested using random inputs. They are considered Trojan-free if all outputs are identical for all inputs.

The author develops a necessary model to discuss the level of security achieved using a sequence of security games. This work was published at the TCC 2021 conference (see [Cha+21]), another highly regarded conference with rigorous quality evaluations of submitted work.

Details about the author's personal contributions can be found in Section 1.4, and I accept them as presented.

Corrections and Suggestions

Overall, the manuscript reads well, though there is significant room for improvement. To enhance readability and presentation, the following aspects of the thesis need to be addressed:

- English: The use of articles (the/a/an) can be tricky, as can punctuation. To refine the English, I suggest seeking help from a native English speaker. Alternatively, AI tools (such as ChatGPT, for example) can be used to correct the text and provide helpful hints. However, a word of caution – AI tools may use terms that are not consistent with the vocabulary used in the field or discipline.
- Structure: The content of the sections is unbalanced. For instance, Section 2 is very short (4 pages). Parts of Section 3 that address the tampering model and notation could be moved to Section 2. Please consider moving Section 3.1.4 (Related Works) to Section 2, where it should discuss all works relevant to both Sections 3 and 4.
- Discussion of Previous Works: Instead of using the past simple tense, it is recommended to use either the present tense or the present perfect tense. For example, "some cases were investigated in [Ish+06]" could be revised to "some cases have been investigated in [Ish+06]" or "some cases are investigated in [Ish+06]." To see why this makes sense, imagine writing an overview with all the works you reference laid out in front of you. Additionally, if you refer to your results already proven in previous sections, you should use the present perfect tense.

- Theorems/Lemmas/Conclusions: Please follow the standard syntax, which consists of two components: (1) Assume/Given/Let – this is where all assumptions are stated; and (2) Then—this is where the statement that needs to be proven is presented.
- The Closing Section: The closing section on page 103 requires attention. It should (1) summarize the author's original contributions; (2) discuss the impact of the results achieved and highlight their potential in solving real-life security problems; and (3) provide prospects for future research directions.
- Capitalization of Names: Capitalization is overused in the text of the thesis. While there is a trend in some disciplines to capitalize certain names and notions, formally, only proper nouns should begin with capital letters. For example, there is no need to capitalize "hardware Trojan horse" as it is not a proper noun. To illustrate, you might encounter many different hardware Trojan horses. Typically, a proper noun is singular.

Please note that minor errors, typos, and suggestions are provided in Appendix at the end of this report.

Summary

The thesis addresses an intriguing and critical topic in the field of hardware security, specifically focusing on protection against side-channel attacks. This area of computer security has been somewhat overlooked and has not received the level of attention it truly deserves. The thesis makes significant contributions to both the theory and practice of hardware security, effectively filling several research gaps. These contributions have been recognized through publication in three papers accepted at top-tier, highly-ranked conferences.

It is worth noting that while the papers are co-authored by up to six individuals, this should not be seen as a detriment to the thesis author's work. On the contrary, it highlights the author's ability to collaborate effectively within a large research team. Such collaboration offers numerous advantages, including rapid feedback on individual progress, the exchange of new ideas, and the development of essential communication skills, among others.

Overall, the thesis presents a substantial amount of original work that undoubtedly qualifies for the awarding of a PhD. Therefore, I recommend that the Scientific Council of the Disciplines of Mathematics and Computer Science at the University of Warsaw accept the thesis and admit Małgorzata Gałązka to the subsequent stages of the doctoral program.

Josef Pieprzyk

Appendix – Minor Corrections

The list given below should provide a more detailed (and by far not complete) list of errors and mistakes.

- page 14, line 5 from the top – "private key trustworthy" – wrong order of words;
- page 15, lines 3 and 4 from the top – awkward sentence;
- page 15, lines 9 to 11 from the bottom – awkward sentence;
- page 16, line 10 from the top – in Cryptography, "physical attack" is better known as "side-channel attack";
- page 17, lines 15 to 18 from the bottom – please correct the sentences;
- page 18, line 16 from the top – should f be replaced by F?
- page 19, line 4 from the top – expand to "by a cheat code or a time-bomb or a combination of the two";
- page 19, line 10 from the bottom – "Very Simple Compilers" – why capital letters? In general, only proper nouns are capitalised. A very simple compiler is not a proper noun;
- page 20, line 15 from the bottom – "never end up with a working device!" – be precise, a device can work incorrectly;
- page 20, after line 15 from the bottom – you may add a short discussion about faults during identification. They are probabilities of false acceptance (FA) and false rejection (FR). The main challenge is to minimise both FA and FR;
- page 24, line 6 and 7 from the bottom – "must be protectable against HTHs." – replace by "must be protected against HTHs."
- page 25, line 4 and 5 from the bottom – correct to "Figures 2 and 3";
- page 26, line 10 from the top – " and which (possibly) not" – correct to "and which are (possibly) not";
- page 27, last line – correct to "Figures 2 and 3";
- page 28, line 5 from the bottom – "Scope of modif existication" – modification?
- page 30, Figure 3 – I do not know what is going on here – clearly both circuits have the same truth table as they are identical. The only difference is that the top XOR gets a negated input. Please rephrase the sentence;
- page 31, the first sentence of Section 3.1.1 – please elaborate why the tampering model reflects the reality;
- page 32, line 7 from the top – "Negation of values transmitted by wires sometimes happens temporarily" correct to "Negation faults are typically intermittent";
- page 32, line 16 from the top – are you sure about 16 gates? Do you count the constants 0 and 1?
- page 33, caption of Figure 4 – replace "The truth table of both functions is identical." by "It is easy to see that the two functions are identical.";
- page 33, line 1 from the top – replace "The centre" by "The focus";
- page 33, line 6 from the bottom – remove "for formality";
- page 33, Theorem 1 - rephrase it to the correct syntax "Given . . .. Then . . ..";
- page 34, line 10 from the top – correct to "a testable circuit with the probability that approaches 1 when the number of tests grows.";
- page 34, line 13 from the top – replace "quality" by "advantage";
- page 34, line 16 from the top – correct "require using some dedicated tools." to "require the use of some dedicated tools.";
- page 35, line 6 and 7 from the bottom – correct to "The idea of an adversary who modifies circuit computation by applying the zero, one and neg operations has been investigated in [Ish+06]";
- page 38, line 9 from the top – correct to "Equation (1)";
- page 38, last line – too many parentheses;

- page 39, last line – do you mean "values" instead of "valuations"?
- page 40, line 1 from the top – do you mean "values" instead of "valuations"?
- page 41, Figure 6 – too sketchy, please add more details;
- page 41, please explain what you mean by topological order of wires and gates. Such order is well defined for input and output wires of a device. The situation is different for gates. They can be positioned in 3D space;
- page 42, Algorithm 1 – please insert more in-text comments;
- page 51, line 8 from the top – correct to "At the first sight";
- page 53, Definition 11 – use b-index instead of index$_b$;
- page 53, paragraph below Definition 12 – if $b \in \{0,1\}$, then $1-b$ is equivalent to $1+b$, the addition/subtraction operations are the same;
- page 53, Theorem 6 – put "For $b \in \{0,1\}$ and $j \in [k]$" and write it in correct form;
- page 55, proof of Theorem 7 – equation $1-b$ is equal to $1+b$;
- page 58, Algorithm 5 – please do not use "TestableCircuitCompiler". The standard "Testable Circuit Compiler" is much better;
- page 58, Theorem 8 – syntax;
- page 60, proof of Theorem 9 – please move Definition 13 out of the proof (either at the from or at the back);
- page 60, Theorem 9 – syntax;
- page 63, Proposition 1 – it should follow the theorem syntax;
- page 68, Lemma 2 – syntax;
- page 76, the first statement of Section 3.6 – correct to "We have described...";
- page 79, line 2 and 3 from the bottom – put "If so, it forwards the received output otherwise it blocks";
- page 80, the first paragraph – you describe the organisation of Section 4. There is nothing about organisation of other sections. Please remove the paragraph or alternatively, introduce similar paragraphs in other sections;
- page 81, Lemma 4 and Theorem 12 – syntax;
- page 82, 2 last lines – remove the last sentence. Maybe you can write a separate section of future research directions;
- page 84, Lemma 5 – syntax;
- page 86, Corollary 6 and Lemma 6 – syntax;
- page 89, Claim 1 – syntax;
- page 90, Theorem 13 and Lemma 7 – syntax;
- page 92, line 20 from the top – "sequence of hybrids" or sequence of games? Please be consistent and use either a hybrid or a game throughout the text;
- page 103 – the work conclusions need to be re-written. The first sentence makes no sense to me;