

IEEE Copyright Notice

IEEE copyright notice © 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted.

Accepted to be published in: 2021 IEEE 48th International Symposium on Circuits and Systems (IEEE ISCAS 2021), May 22-28, 2021

arXiv:2102.00724v2 [cs.CR] 29 Mar 2021

Side-Channel Trojan Insertion – a Practical Foundry-Side Attack via ECO

Tiago Perez, Malik Imran, Pablo Vaz and Samuel Pagliarini

Department of Computer Systems - Tallinn University of Technology, Tallinn, Estonia

Emails: {tiago.perez,malik.imran,pablo.vaz,samuel.pagliarini} @taltech.ee

Abstract—Design companies often outsource their integrated circuit (IC) fabrication to third parties where ICs are susceptible to malicious acts such as the insertion of a side-channel hardware trojan horse (SCT). In this paper, we present a framework for designing and inserting an SCT based on an engineering change order (ECO) flow, which makes it the first to disclose how effortlessly a trojan can be inserted into an IC. The trojan is designed with the goal of leaking multiple bits per power signature reading. Our findings and results show that a rogue element within a foundry has, today, all means necessary for performing a foundry-side attack via ECO.

Index Terms—hardware security, manufacturing-time attack, hardware trojan horse, side-channel attack, VLSI, ASIC.

I. INTRODUCTION

The ever-increasing cost to build high-end semiconductor manufacturing facilities – estimated to cost \$15-20B [1] – has made most design companies migrate to a fabless model. In practice, design houses can market integrated circuit (IC) solutions, but fabrication is outsourced to a third party. The practice of outsourcing can potentially affect the trustworthiness of an IC as a foundry (or a rogue element within the foundry) can manipulate the design for malicious purposes [2], [3].

Manufacturing-time attacks can tamper an otherwise trustworthy IC by inserting malicious logic or modifying specific aspects of the manufacturing process [4], [5]. These kinds of modifications are often referred to as hardware trojans (HTs). HTs are designed to leak confidential information, to disrupt a system’s specific functionality, or even to destroy the entire system. Various types of HTs have been recently studied [6]–[15], demonstrating the potential threat of this type of attack. Moreover, a class of HTs has emerged for assisting side-channel attacks (SCA). Lin *et al.* [6] were the first to propose an architecture for assisting a power SCA. This specific type of trojan is the centerpiece of our work and in the remainder of this text is referred to as a side-channel trojan (SCT).

An IC’s operating characteristics (e.g., timing, power consumption, electromagnetic radiation, etc.) can be used as a side-channel to indirectly reveal information that should be internal to the IC. For this reason, keys of crypto cores [16] are often targeted. However, to mount a successful SCA, it is necessary to acquire a large amount of data to perform correlation on. Using SCTs, on the other hand, the attack time and complexity is drastically reduced. The disadvantage of SCTs is that they require a circuit modification at fabrication time, which we later show is an **effortless exercise** for the attacker.

In [7], two lightweight SCT architectures are proposed, both with the intent to induce power consumption in order to leak a crypto key. The first architecture makes use of an adapted code-division multiple access (CDMA) scheme to distribute the leakage of bits over time. The modulated bits are forwarded to

a special “leakage circuit” that creates a CDMA channel over the power side-channel. The second architecture, in addition to the CDMA scheme, also implements intermediate states within the AES key schedule to facilitate a differential power analysis (DPA) attack. Both architectures are implemented in a field programmable gate-array (FGPA) platform.

A silicon validated HT is presented by the authors of [8], [9]. Their demonstration is a cryptographic IC composed of an AES core and an Ultra-WideBand transmitter that leaks the key together with the transmission of the 128 bits of ciphertext. To broaden the scope of SCTs from dedicated crypto hardware to general-purpose processors (GPPs), an interesting architecture is described in [12], where software models of crypto standards (AES and RSA) are executed on GPPs. A number of simple micro-architectural modifications has been described to induce information leakage via faulty computations or variations in the latency and power consumption of certain instructions.

Despite the encouraging results reported from the SCT studies mentioned so far, no study discusses how SCTs could be inserted **from the perspective of the attacker**. In this work, we present not only an SCT design methodology, but also a novel framework for SCT insertion. We assume that a rogue agent inside the foundry is the adversary and that he/she makes use of **readily available engineering change order (ECO)** capabilities of physical design tools.

II. THREAT MODEL AND ATTACKER CAPABILITIES

An attacker inside the foundry has the objective of inserting malicious logic in a finalized layout. Thus, he/she enjoys access to all technology and cell libraries utilized by the victim. This is particularly true for advanced nodes where only a handful of cell libraries per node exist. We assume the attacker is capable of identifying a crypto core in a layout, which is a reasonable assumption for well-known AES implementations that are often regular. We do not assume the adversary understands the entire victim’s design. Instead, we assume the adversary can recognize the layout/structure of a crypto core within a larger design. Our assumptions are in line with [7], [9]. Furthermore, we also assume the adversary: 1) is versed in IC design, 2) enjoys access to modern EDA tools. With the help of the inserted logic in the form of an SCT, the attacker will then attempt to leak confidential information via a power signature. For this reason, crypto cores are often the target in this type of attack [9], [10] – this is also the scenario in our work.

A typical physical implementation flow is described in the upper portion of Fig. 1. The attack takes place after the victim’s layout is sent for fabrication (see red portion of Fig. 1). Our threat model assumes that the attacker only has access to the layout (which is the norm when outsourcing IC fabrication)

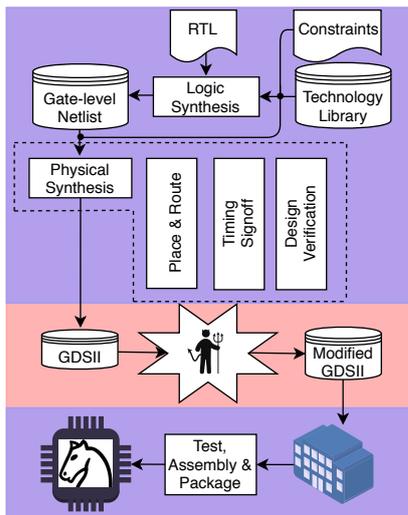


Fig. 1: A typical IC design flow. Highlighted in red is the untrusted fabrication stage where the attack takes place.

– he/she would not be able to insert the malicious logic by replicating the physical implementation flow since he/she does not have access to the RTL code, netlists, constraints, etc.

Nevertheless, EDA tools already have the capability to deal with finalized designs. This functionality is a feature referred to as ECO. Thus, an attacker holding only the layout could use an ECO to modify or insert additional logic in a finalized layout. An ECO flow requires four inputs: technology library, cell library, gate-level netlist, and a timing constraint. The adversary already possesses the first two, but must generate/estimate the others. A gate-level netlist can be effortlessly obtained from the victim’s layout through extraction [17]–[19], while the timing constraint can be estimated to a certain degree [20]–[22]. Our novel trojan insertion framework is shown in Fig. 2, where these two steps are considered.

III. SIDE-CHANNEL TROJAN DESIGN AND INSERTION

A. Side-Channel Trojan Design

Our proposed SCT is designed for creating an “artificial” yet controllable power consumption through which information is leaked. Since the majority of the power consumption in a circuit comes from the switching activity (dynamic power), a great candidate to be a power sink is a structure with a controllable frequency such as a dynamic ring-oscillator (RO). Our RO architecture implements delay stages broken into branches that are controlled by N_{leak} bits. Each RO branch has two active path options: a direct connection to the next branch or a series of delay cells. The power consumption created by paths is similar to a pulse-amplitude modulation with an order equal to $2^{N_{leak}}$. An example of SCT architecture for $N_{leak} = 2$ is illustrated in Fig. 2. The branch configuration is described in Table I, where the leaked bits are selectors $S0$ and $S1$.

TABLE I: Ring oscillator active path configuration

$S0$	$S1$	Delay Cells	Inverter Cells	Freq.
0	0	N_{D1}	N_i	High
1	0	$N_{D1} + N_{D2}$	N_i	Mid-high
0	1	$N_{D1} + N_{D3}$	N_i	Mid-low
1	1	$N_{D1} + N_{D2} + N_{D3} + N_{D4}$	N_i	Low

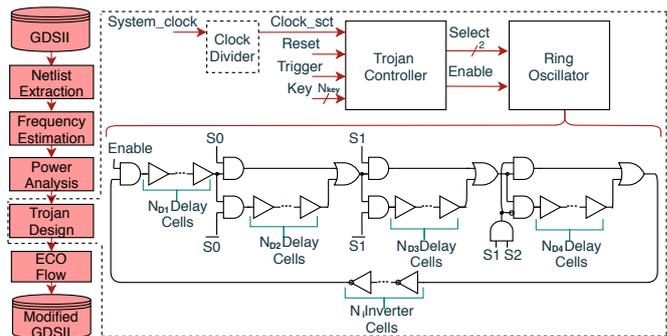


Fig. 2: Our SCT insertion methodology detailed.

A dual-sided constraint guides the attacker: he/she has to induce as much dynamic power as possible (i.e., to increase the effectiveness of the attack) while increasing as little leakage power as possible (i.e., to avoid detection). In this sense, not only the SCT has to be carefully planned, as well as when exactly will the trojan be triggered. Our approach is to not allow the trojan to compete with the dynamic power consumption of the crypto core. Therefore, when the core is actively working, the trojan is silent and the RO is not switching. When the crypto core is idle, the trojan kicks in. For this reason, our proposed SCT trojan has a *Trigger* signal that is connected to the *Done* signal coming from the crypto core, which marks the end of a cryptographic operation.

When triggered, the SCT connects a set of the leaking bits per clock cycle in the RO until all the N_{key} bits from the crypto key are leaked. Thus, our SCT requires a connection to the system clock and reset, a trigger signal, and the crypto key. Its architecture is illustrated in Fig. 2, consisting of three blocks: clock divider (DV), the trojan controller (TC), and the RO. The DV is required when the system clock is high and is responsible for dividing the frequency as the name suggests. Thus, the *Clock_sct* signal is either connect directly to the *System_clock* or to the DV. The TC is responsible for enabling the RO and for connecting the leaking bits in the RO. The RO starts running when the enable signal is asserted. The frequency is controlled by the select signals $S0$ and $S1$.

To reduce the detection probability and increase the attack’s feasibility, SCTs are tailored for each target circuit. Therefore, the SCT is designed with size and power constraints, i.e., we set thresholds for the SCT based on the target’s size and static power. The attacker has to acquire such information from the layout. According to Fig. 2, the layout is inspected as follows:

Netlist extraction: since the attacker only holds the layout, a gate-level netlist has to be extracted by a CAD tool [17]–[19].

Frequency estimation: the attacker needs to estimate the target circuit operating frequency by performing static timing analysis on the extracted gate-level netlist. The attacker can try different clock frequencies and, by observing the critical path(s), can increase/decrease the frequency as needed until the timing slack is positive but near zero. The caveat is that multi cycle and false paths are expected to violate STA, and for this reason we say the frequency of operation is *estimated*.

Power analysis: with the extracted gate-level netlist and the estimated operating frequency, the attacker can perform a

typical power analysis. For relatively large circuits, a near-accurate static power estimation can be achieved even without input vectors.

Therefore, after inspection, the attacker has estimated frequency and power consumption and is now ready to draw his SCT. The RO's dynamic power is tweaked by choosing an adequate number of delay cells in each individual branch as well as the number of inverter cells in the feedback path. The achieved amplitude steps have to be sufficiently different from one another for the attack to be successful.

B. Side-Channel Trojan Insertion

After designing the SCT, the next step is its insertion. The attacker can utilize the ECO feature provided by commercial EDA tools for inserting the SCT. Typically, ECO is used to perform slight modifications in a finalized layout after its manufacturing (i.e., post-mask ECO). A special type of spare cell is utilized to enable ECOs. These cells do not add any functionality to the original design but, when needed, are instantiated by the ECO flow. By doing so, a new design can be generated with minimal changes in the fabrication mask set.

For the SCT insertion via ECO, since we previously established that the attacker can discern any gate in a layout, the attacker can replace both filler and spare cells by his malicious logic [23]. Contrarily to spare cells, every layout has filler cells. During placement, EDA tools have to spread the standard cells to assure routability, thus mandatorily leaving gaps between cells. For more details about the relationship between placement density and HT insertion, we direct the reader to [23].

According to Fig. 2, the ECO flow is the last step for the SCT insertion. In order to identify the filler/spare cells and remove them to create the gaps needed for the SCT, a single Cadence Innovus command is required. After the ECO, the attacker has to perform a timing sign-off to guarantee that the performance of the victim's design was kept. The SCT insertion is not likely to perturb the target's performance; it is only connected to a register (key storage) and some control signals, adding a small capacitance load. Besides, the coupling capacitance inserted by the additional routing wires is minimal due to the SCT's lightweight characteristic and the inherent goal of the ECO flow: not to disturb the existing logic. However, if the target circuit performance is perturbed, even if unlikely, it means that the size constraint used for designing the SCT was inappropriate - the adversary then proceeds to pick a different value and leak less bits per clock cycle. The attacker also has to check whether the SCT itself has timing violations. If so, the optional clock divider must be included. Every division by two requires one additional D-type flip-flop.

IV. EXPERIMENTS AND RESULTS

For our experimental investigation, we have utilized AES and Present (PST) crypto cores with $N_{key}=128$ and $N_{key}=80$, respectively. AES was chosen due to its standardized status while PST was chosen due to its lightweight characteristic [24]. To allow the analysis of changes in *frequency* and *density*, the combination of these variables is explored as low-frequency low-density (LFLD), low-frequency high-density (LFHD), high-frequency low-density (HFLD), and

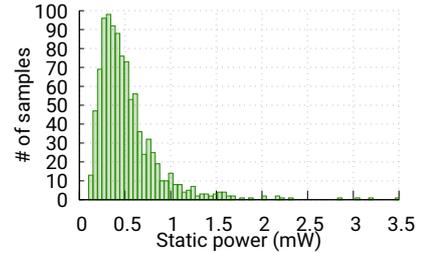


Fig. 3: PST_HFLD static power histogram, 10K MC samples.

high-frequency high-density (HFHD). Results from physical synthesis of the considered targets are presented in Table II. A 65nm CMOS technology was utilized to exercise very challenging placement densities (e.g., 75% for AES_LFHD) and frequencies (e.g., 0.95GHz for PST_HFLD). The values reported are for a typical corner.

Based on the pre-ECO results reported in Table II, different SCTs were designed for each target. We assume the attacker has no means to stop clock delivery to the whole circuit, so we included the clock tree power as it has to be accounted for in our SCT power constraint. Notice how the clock tree power is significant w.r.t. the leakage power of the targets, even for the LF variants. In the results that follow, we therefore set a power budget for our SCTs of 10% of the sum of leakage and clock tree power. Importantly, this is not a limitation of the methodology, an attacker can pick any other threshold.

Aiming to obtain a better representation of the static power of the cores, we performed a Monte Carlo (MC) simulation using Cadence Spectre. This simulation was performed for 1000 samples, varying only the process with temperature fixed to 25°C. The simulation results match the values reported in Table II for the typical corner. Fig. 3 depicts the static power distribution of PST_HFLD. As the SCT is implemented in the very same region of the IC as the target, we can also expect the same variation in its power.

Once the power constraint has been established, the attacker can proceed to estimate the multiple operating frequencies of the RO (and the associated power values that effectively leak the key). Moreover, as previously alluded, we have to take into account the placed and routed version of the SCTs. For this goal, we have taken each of our SCTs and performed a custom simulation using Cadence Spectre. The oscillation frequency and power consumption of the ROs are reported in Table III, where each RO has been termed with a "DXIY" suffix. X and Y represent the number of delay and inverter cells, respectively. Notice how we do not differentiate density in the results reported in Table III: either the trojan fits or it does not. The SCT design is nearly agnostic to placement density.

A visual representation of how the SCT performs is given in Fig 4. The set of leaked keys in the image is {00-10-01-11} and the target circuit is AES_LFHD. We also highlight an extreme case in the RO_{D6I4} which targets the PST_LF core. Here, the SCT alone represents about 10% of the size of the PST core. Since area and leakage have a linear dependency, the SCT's leakage already is about 10% of the target's leakage. Hence, the power constraint is violated. This extreme example assumes the entire IC consists of a single PST core. For a large system-on-chip containing multiple cores, the power budget for

TABLE II: Physical synthesis results for our considered targets, before and after trojan insertion.

Core	Frequency (MHz)	Before SCT insertion				After SCT insertion			
		Density (%)	Leakage (μW)	Clock Tree Power (μW)	Total Power (μW)	Density (%)	Leakage (μW)	Clock Tree Power (μW)	Total Power (μW)
AES_LFLD	100	61	77.4	115.2	1670	63.45	80	115.8	1720
AES_LFHD	100	75	75.8	116.7	1660	78.20	79	117.6	1720
AES_HFLD	1000	58	1048	1228	22800	59.37	1052	1238	23015
AES_HFHD	1000	72	1036	1241	22610	73.02	1040	1252	22830
PST_LFLD	95	53	14.13	32.05	371.3	67.33	20.71	34.75	483.4
PST_LFHD	95	70	14.09	31.89	371.2	82.05	17.72	32.85	428.5
PST_HFLD	950	52	34.02	325.30	3744	60.89	36.85	338.1	4022
PST_HFHD	950	69	34.13	329.10	3785	80.26	36.96	341.5	4015

TABLE III: RO operating frequency and power consumption

Target core	RO	Power & Frequency (μW & MHz)			
		S=00	S=01	S=10	S=11
AES_LF	RO_{D6I10}	19.52@65	16.89@45	14.94@34	12.96@20
AES_HF	RO_{D10I10}	198.4@551	182.5@483	160.7@390	139.8@300
PST_LF	RO_{D6I4}	15.95@112	11.55@58	10.22@39	8.7@20
PST_HF	RO_{D8I10}	42.02@79	35.56@61	30.88@46	25.66@31

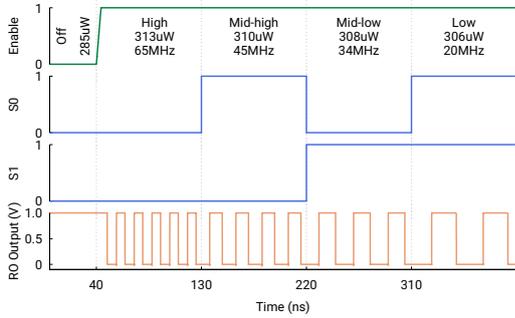


Fig. 4: Side channel trojan functionality example for the AES_LFHD, where the SCT utilizes the RO_{D6I10} .

designing the SCT would be much more forgiving.

Alongside the custom-simulated ROs, the SCTs are synthesized for each N_{key} and at the same clock frequency of the target. Exclusively for the HF targets, we added the CD block to ensure the SCT does not violate timing. For AES_HF, the system clock was divided by 8 while for PST_HF it was divided by 16. Area and cell count for the SCTs are given in Fig. 5.

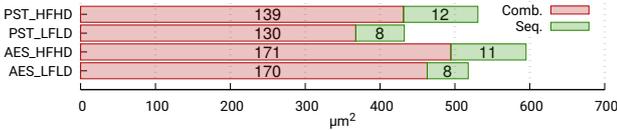


Fig. 5: Comparison of area and number of cells between SCTs.

After designing the RO and synthesizing the remainder of the SCT logic, the attacker is ready to perform the insertion via ECO. Insertion results are described on Table II ('After SCT insertion'). For all considered scenarios, the ECO flow was capable of placing and routing the SCT successfully, even for dense layouts. Considering that high density implies less routing resources, we verified that the ECO flow purposefully utilizes the least congested metal layers. We also provide a visual comparison of the density increase for the PST_HFHD SCT in the left side of Fig. 6. Note that the placement of the target was kept identical and only filler cells were removed

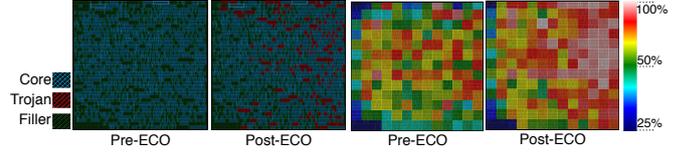


Fig. 6: Placement view (left) and density map (right) of the PST_HFHD core, before and after SCT insertion via ECO.

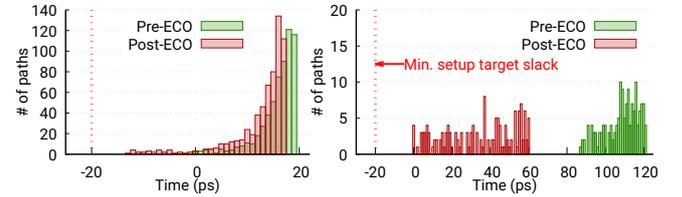


Fig. 7: Pre- and post-ECO setup timing slack comparison of AES_HFHD (right) and PST_HFHD (left).

during ECO. This is the key finding of this paper: **an adversary can effortlessly insert an SCT into a finalized layout.**

Besides enabling the SCT insertion, the ECO flow also has to preserve the performance of the target circuit. The impact on the performance of AES_HFHD and PST_HFHD cores is illustrated in Fig. 7. The difference in pre- and post-ECO timing slack is attributed to additional load and coupling capacitances. One can appreciate how the red bars in Fig. 7 are shifted to the left (w.r.t. the green bars). However, this shift was not sufficient to degrade the performance of any core. The PST_HFHD implementation is affected slightly (which is explained by the increase in density reported in Table II) but does not violate our safety margin of 20ps applied to all paths. Furthermore, we argue that our proposed methodology is not only capable of inserting an SCT in a high density layout, but also of keeping the target's performance regardless of its (challenging) frequency. Finally, there are very few techniques that would assuredly counter the ECO-enabled trojan insertion [3], [25].

V. CONCLUSIONS

In this work, we proposed an SCT design methodology as well as a novel framework for SCT insertion via ECO. The SCT insertion was detailed step by step, showing that a rogue element inside a foundry can replicate it effortlessly. Furthermore, our results show how efficient an otherwise benign ECO flow can be when used for malicious reasons. Our future work includes a silicon demonstration of the inserted HT. A tapeout was completed during the writing of this paper and the fabricated ICs are expected to arrive by Jan/2021.

ACKNOWLEDGMENT

This work has been partially conducted in the project “ICT programme” which was supported by the European Union through the European Social Fund.

REFERENCES

- [1] M. Lapedus, “Big trouble at 3nm,” [Online]. Available at: <https://semiengineering.com/big-trouble-at-3nm/>.
- [2] U. Guin *et al.*, “Counterfeit integrated circuits: A rising threat in the global semiconductor supply chain,” *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1207–1228, 2014.
- [3] S. Pagliarini, J. Sweeney, K. Mai, S. Blanton, S. Mitra, and L. Pileggi, “Split-chip design to prevent ip reverse engineering,” *IEEE Design & Test*, 2020.
- [4] M. Tehranipoor and F. Koushanfar, “A survey of hardware trojan taxonomy and detection,” *IEEE Design Test of Computers*, vol. 27, no. 1, pp. 10–25, 2010.
- [5] M. Rostami, F. Koushanfar, and R. Karri, “A primer on hardware security: Models, methods, and metrics,” *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1283–1295, 2014.
- [6] L. Lin, W. Burlison, and C. Paar, “Moles: Malicious off-chip leakage enabled by side-channels,” in *2009 IEEE/ACM International Conference on Computer-Aided Design*, pp. 117–122, 2009.
- [7] L. Lin *et al.*, “Trojan side-channels: Lightweight hardware trojans through side-channel engineering,” in *Cryptographic Hardware and Embedded Systems - CHES 2009*, pp. 382–395, 2009.
- [8] Y. Jin and Y. Makris, “Hardware trojans in wireless cryptographic ics,” *IEEE Design Test of Computers*, vol. 27, no. 1, pp. 26–35, 2010.
- [9] Y. Liu, Y. Jin, and Y. Makris, “Hardware trojans in wireless cryptographic ics: Silicon demonstration & detection method evaluation,” in *Int. Conf. on Computer-Aided Design (ICCAD)*, pp. 399–404, 2013.
- [10] R. Kumar, P. Jovanovic, W. Burlison, and I. Polian, “Parametric trojans for fault-injection attacks on cryptographic hardware,” in *2014 Workshop on Fault Diagnosis and Tolerance in Cryptography*, pp. 18–28, 2014.
- [11] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, “A2: Analog malicious hardware,” in *2016 IEEE Symposium on Security and Privacy (SP)*, pp. 18–37, 2016.
- [12] J.-F. Gallais *et al.*, “Hardware trojans for inducing or amplifying side-channel leakage of cryptographic software,” in *Trusted Systems*, pp. 253–270, 2011.
- [13] K. Hasegawa, M. Yanagisawa, and N. Togawa, “Trojan-feature extraction at gate-level netlists and its application to hardware-trojan detection using random forest classifier,” in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, 2017.
- [14] S. Bhasin and F. Regazzoni, “A survey on hardware trojan detection techniques,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2021–2024, 2015.
- [15] S. Yu, C. Gu, W. Liu, and M. O’Neill, “A novel feature extraction strategy for hardware trojan detection,” in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2020.
- [16] P. Kocher, J. Jaffe, and B. Jun, “Differential power analysis,” in *Advances in Cryptology — CRYPTO’ 99* (M. Wiener, ed.), pp. 388–397, 1999.
- [17] Cadence Design Systems, “Virtuoso Layout Suite,” [Online]. Available at: https://www.cadence.com/en_US/home/tools/custom-ic-analog-rf-design/layout-design/virtuoso-layout-suite.html.
- [18] Mentor, “Calibre @,” [Online]. Available at: https://www.mentor.com/products/ic_nanometer_design/.
- [19] Synopsys, “Synopsys Custom Design Platform,” [Online]. Available at: <https://www.synopsys.com/implementation-and-signoff/custom-design-platform.html>.
- [20] R. Torrance and D. James, “The state-of-the-art in semiconductor reverse engineering,” *Design Automation Conference*, pp. 333–338, 2011.
- [21] N. Albartus, M. Hoffmann, S. Temme, L. Azriel, and C. Paar, “Dana - universal dataflow analysis for gate-level netlist reverse engineering,” 2020. <https://eprint.iacr.org/2020/751>.
- [22] G. L. Zhang, B. Li, B. Yu, D. Z. Pan, and U. Schlichtmann, “Timingcamouflage: Improving circuit security against counterfeiting by unconventional timing,” in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 91–96, 2018.
- [23] T. Trippel *et al.*, “ICAS: An Extensible Framework for Estimating the Susceptibility of IC Layouts to Additive Trojans,” *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 1078–1095, 2020.
- [24] S. Ghandali *et al.*, “Side-channel hardware trojan for provably-secure sca-protected implementations,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 6, pp. 1435–1448, 2020.
- [25] T. D. Perez and S. Pagliarini, “A survey on split manufacturing: Attacks, defenses, and challenges,” *IEEE Access*, vol. 8, pp. 184013–184035, 2020.