Logic Obfuscation using Metasurface Holography

Mahabubul Alam, Yimin Ding, Xingjie Ni, Swaroop Ghosh
Department of Electrical Engineering
Pennsylvania State University
University Park, PA 16802 USA
{mxa890, yiminding, xingjie, szg212}@psu.edu

Abstract—Over the years, system design companies have developed sophisticated and accurate Reverse Engineering (RE) tools to debug the functionality of their own designs, ensure the legitimacy of their Intellectual Properties (IPs), and to some extent, gather information on their competitors. Technologies such as optical imaging, Scanning Electron Microscopy (SEM), Tunneling Electron Microscopy (TEM), X-ray Tomography, etc., have made it possible to acquire images of individual layers of a delayed Integrated Circuit (IC) with superfine resolution. State-of-the-art tools such as ICMWorks from Chipworks can perform automatic extraction of netlists from images of each layer. An adversary can unlock valuable IPs through such attacks and pirate the design or sell the cloned IP to other competitors for unjust motives. Logic and interconnect obfuscation have been proposed to make the RE effort behind such approaches prohibitively high for the adversaries. Several solutions have been proven to be vulnerable to Satisfiability (SAT) based attacks where an adversary can deobfuscate the circuit reasonably quickly. In this paper, we have presented a novel obfuscation technique based on the combination of cloaked nets and fake holographic gates using metasurface holography. The proposed approach can address optical and X-ray imaging based RE.

Index Terms—Metasurface Holography, IC Camouflaging, Reverse Engineering, Fake Gates, Cloaked Nets.

I. INTRODUCTION

Reverse Engineering (RE) of an Intellectual Property (IP) is a process of identifying its design from the fabricated chip. In the RE, the adversary de-packages the Integrated Circuit (IC), delays the IC, and takes pictures of each layer [1], [2]. The images of metal layers provide connectivity information whereas the image of the base layer is employed to identify the gate functionality. Finally, the information obtained from images are merged together to prepare a netlist unlocking the IP. RE has been originally used by industries with the mindset of gathering information on its competitors such as process parameters (e.g., channel length, the pitch of poly and metals), to confirm or debug the functionality of their own design, and to ensure the legitimacy of circuits from piracy. However, the advanced adversaries can exploit this technique with an ill intention to steal a design to illegally sell on the black market.

A. Motivation

Camouflaging or logic obfuscation has been proposed to hide the logic functionality of the gates and make the RE economically non-profitable or extremely difficult. Camouflaged gates can contain multiple functionalities such as, AND, OR, XOR, etc., that are selectable using electrical signals (e.g., Field Programmable Logic Array), vias [3]–[5], or transistor threshold voltage [6]–[8]. Since camouflaged gates are typically area, delay and power intensive, only a few gates from the design are chosen for camouflaging with the objective to increase adversarial RE effort while keeping the overheads minimal. Although the exact gate functionality is hidden, the adversary can still create a partial netlist with other known gates and go through the guess-and-validate process to RE the missing gate functionalities. The adversarial RE effort can be increased by fusing more functionalities in the camouflaged gate. Therefore, the RE effort translates to the area, delay and power overhead of the design. In such camouflaging techniques, the position of the camouflaged gates is known to the attacker, while the true functionalities of individual camouflaged gates are hidden. As the adversary can make a finite set of configurations of the camouflaged circuit (modeled by input vectors), query based attacks (i.e. SAT attacks) have been proven very successful in de-camouflaging these circuits [9], [10].

In this paper, we present a new approach using metasurface holography to protect the IP against optical and X-ray imaging based RE. The newly emerging metasurface — an ultrathin layer of engineered nanostructure that has the capability of locally tailoring the properties of light at the nanoscale — offers tremendous power for manipulating light and X-ray. The concept of the proposed obfuscation technique is illustrated in Figure 1(a)-(c). The metasurface layer (MS1) is fabricated in a shallow trench (~1-2um deep) with appropriately sized nano-antennas (can be as low as 30nm) to project the hologram of a gate at the same level as the real gates. Real interconnects will be designed to connect to the holographic gate’s drain, gate and source terminals. The output of the fake gate will be fed to the real design and a corresponding real interconnect will be cloaked by fabricating metasurface layer-2 (MS2) on top of it. The MS2 layer will project the dielectric and de-focused interconnect layer underneath for cloaking. During RE, the adversary will take the image of the interconnect layer (containing the cloaked net) before peeling the layer. Therefore, MS2 will successfully cloak the net. After the imaging, MS2 will get peeled along with the interconnect layer leaving no trace. At the base layer, the adversary will take an image of the layout containing the fake gate. Since MS1 is embedded deep in the substrate, an adversary will not be able to identify or peel it. Therefore, the adversary will obtain an incorrect netlist with fake gates and deleted nets as shown in Figure 1(c).

B. Contributions

To this point, in this article, we (a) introduce the concept of holographic gates and cloaked nets for logic obfuscation; (b) present the techniques to utilize metasurface holography to project fake gates and cloak real nets so that during optical RE, the adversary gets manipulated images of different layers of the chip; (c) present a novel cloaking methodology based on holographic cloaked nets and fake gates.

The rest of the article is organized as follows: In Section II, we discuss metasurface holography and techniques to project fake gates and cloaked nets. In Section III, we discuss design obfuscation approaches using holographic fake gates and cloaked nets. In Section IV, we have analyzed the strength of the proposed technique under known attacks. Conclusions are drawn in Section VI.

II. METASURFACE HOLOGRAPHY

Traditional photography, which can modulate the amplitude distribution of scattered light, is only capable to construct 2D images. Holography is an imaging technique that can control both the phase and amplitude distribution of scattered electromagnetic waves and consequently form 3D images [11]. Unlike photography where images lie exactly on photos, the phase information and additional dimension can be used to construct a virtual object out of the hologram plane, i.e. an image can float above or beneath the hologram.

Metasurfaces, which are composed of artificially engineered sub-wavelength units (i.e. meta-atoms) on single-layer or few-layers structures, have emerged recently [12] [13] [14]. Meta-atoms give a different response to electromagnetic waves depending on their artificially controllable materials and structures. For example, analogous to an LC resonant circuit, a resonant meta-atom also gives a response to electromagnetic fields depending on their phase and amplitude modulation. By judicious design and arrangement of meta-atoms, the phase and amplitude distribution of scattered electromagnetic waves from metasurfaces can be controlled.
The unique ability of metasurfaces to locally control the amplitude and phase of electromagnetic waves gives us opportunities to construct wavefront control devices in ultra-thin structures including meta-lens [15], [16], invisibility cloaking [17]–[19]. In addition, the ultra-thin nature of metasurfaces makes them promising candidates for being integrated into nanoelectronics.

B. Cloaking Real Interconnects

We propose to cloak real nets by generating a holographic image floating beneath the metasurface. As shown in Figure 2(b), a metasurface on top of interconnecting nets can generate a holographic image that covers the real net and makes it transparent. The comparison between a real net (Figure 2(c)) and a cloaked nets image (Figure 2(f)) is shown. The hologram design methodology is same as above; however, the metasurface is fabricated on top of the net to be cloaked using the same material as the interconnect (Cu). Therefore, the distance between the holographic image and the metasurface is set to be 200nm.

III. OBFUSCATION METHODOLOGY

A. Interconnect Cloaking and Fake Gate Insertion Approaches

Following basic approaches can be used to insert a single fake gate into the design with no/single/multiple cloaked nets-

1. **Without cloaking interconnect:** If a fake gate (FG) is inserted into the design without cloaking any interconnect, in the recovered netlist, the adversary will find a gate input driven by multiple nets. One of the nets will be a fake gate fanout (FFO). For instance, in Figure 3(b), g8, and g3 gates are driving g4. The attacker will know one of these two gates (g8 and g3) must be a fake gate.

2. **Fake gate insertion at a single-fanout net with cloaked interconnect:** (d) Fake gate insertion at a multi-fanout net with cloaked interconnect.

A1) Without cloaking interconnect: If a fake gate (FG) is inserted into the design without cloaking any interconnect, in the recovered netlist, the adversary will find a gate input driven by multiple nets. One of the nets will be a fake gate fanout (FFO). For instance, in Figure 3(b), g8, and g3 gates are driving g4. The attacker will know one of these two gates (g8 and g3) must be a fake gate.

A2) Cloaking a single gate driving net: In this approach, the cloaked net will create two unconnected pins (one input and one output) in the recovered netlist. The inserted fake gate must feed the unconnected input. The fake gate itself must be fed by the unconnected output. In Figure 3(c), the interconnect (CN) between g1 (single-fanout) and g6 has been cloaked. An FG (g8) has been inserted which feeds g6 through FFO, and itself is fed by g1 fanout and another arbitrarily chosen fanout.

A3) Cloaking a segment of a multi-gate driving net: In this approach, the cloaked net will result in one unconnected input in the recovered netlist. The inserted fake gate has to feed this input. The gate itself can be fed by arbitrarily chosen inputs. In Figure 3(d), g3 (multi-fanout) is driving g4 and g5. Now, the connection between g3 and g4 (CN) has been cloaked and an FG has been inserted (g8) which is now feeding g4 through FFO.

A4) Cloaking multiple single-gate driving net: In this approach, the cloaked interconnects will create multiple unconnected inputs and outputs. The inserted fake gate has to drive all these unconnected inputs. The fake gate itself has to be driven by the unconnected outputs. In Figure 4(a), CN1 and CN2 are cloaked interconnects which has created 2 unconnected outputs (g1 and g5 outputs), and 2 unconnected inputs (g3 and g7 inputs). The inserted FG (g8) has to be connected to all these pins. The cloaked nets can be chosen in different fanin cones which will be useful to corrupt multiple primary outputs with a limited number of fake gates.

A5) Cloaking segments of multiple multi-gate driving net: In this approach, the cloaked segments will create multiple unconnected inputs and outputs. The inserted fake gate has to drive these inputs. In Figure 4(b), CN1 and CN2 segments are cloaked. A FG (g9) is feeding both g7 and g8 unconnected inputs through FFO. These net segments can be chosen in different fanin cones which will be useful to corrupt multiple primary outputs with a limited number of fake gates.
B. Relative Strengths and Weaknesses of the Approaches

Observation-1: Fake gate insertion without cloaked interconnects reveal the position of the possible fake gate to the attacker. Among the approaches discussed above, an attacker knows the position where a fake gate has been inserted only for approach A1 (shown in Figure 3(b)). Gate g4 is driven by both g8 and g5 in the recovered netlist. An attacker will be certain that either g8 or g5 is a fake gate.

Observation-2: Cloaking a segment of a multi-fanout net provides greater stealth to the camouflaging even if the fake gates are identified. Successful removal of the fake gate from the design will also reveal the cloaked interconnect in approach A2. In Figure 3(c), removing the fake gate from the design will create unconnected gate inputs. In Approach A3, an attacker can connect them together to construct the original netlist. For approach A4, an attacker will have an unconnected input after successfully removing the fake gate which can be connected to any other net of the netlist (g4 input in Figure 3(d)) will be unconnected after removing the FG from the netlist.

Observation-3: Using multi-fanout fake gates and multiple cloaked interconnects, we can minimize the number of fake gates for the desired level of output corruptibility. Both, approach A4 and A5 will provide more output corruptibility than approach A2 and A3 as the fake gate is driving multiple real gates in these approaches which cannot be connected to different fanout cones. Approach A5 will be stronger because successful removal of the fake gate does not give any clue about the cloaked interconnects. For approach A4, an attacker can select the missing interconnects from the fake gate inputs after removing the fake gate successfully (observation – 2).

Observation-4: While choosing between two interconnects for cloaking (assuming the same set of PO’s), the one that maximizes the number of selectable nets (details in next section) for the corresponding fake gate inputs also maximizes the attackers effort to reconstruct the corrupted netlist even if the fake gate identity is revealed. When an attacker identifies and removes the fake gate from the design, it results in a gate with undriven input (considering a segment of a multi-fanout net has been cloaked as in A3 and A5). Excluding the nets in the fanout path of the gate with the undriven inputs, any other nets can be logically connected to this undriven input to find all possible net connections for a missing input can be maximized during fake gate insertion, it will also maximize the attacker’s effort to reconstruct the circuit even if the fake gate is successfully identified.

IV. RE OF THE OBFUSCATED IC

To recover a netlist through RE of the obfuscated IC, an attacker has to delayer the chip, take optical images of different layers, recover the camouflaged netlist by stitching the images together, identify and remove the fake gates from the design, and find out the missing connections (cloaked interconnects) to reconstruct the original netlist. In our attack scenario, we have made the following assumptions:

Assumption – 1: The attacker is capable of producing a partially incorrect netlist of the obfuscated design using optical RE.

Assumption – 2: The obfuscation methodology and algorithms are known to the attacker.

Assumption – 3: Fake gates or the cloaked interconnects are not identifiable through the optical images, and hence, the positions of the fake gates and cloaked nets are unknown to the attacker.

Based on our Assumption – 3, the attacker has no clue about the positions or number of the fake gates and cloaked interconnects just from the recovered layout. For a novice attacker, each and every gate in the netlist will be suspected as possible fake gates (considering a fully camouflaged design where all the primary outputs are corrupted). However, in an obfuscated design where all the PO’s are not corrupted, an attacker can reduce the set of suspected fake gates by removing the gates that are driving the non-corrupted PO’s.
TABLE I

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>( p = 0.1 )</th>
<th>( p = 0.3 )</th>
<th>( p = 0.6 )</th>
<th>( p = 1 )</th>
<th>Fully Gate Camouflaged</th>
</tr>
</thead>
<tbody>
<tr>
<td>c432</td>
<td>243</td>
<td>731</td>
<td>1462</td>
<td>2437</td>
<td>320</td>
</tr>
<tr>
<td>c499</td>
<td>317</td>
<td>953</td>
<td>1908</td>
<td>3180</td>
<td>404</td>
</tr>
<tr>
<td>c880</td>
<td>664</td>
<td>1994</td>
<td>3987</td>
<td>6646</td>
<td>766</td>
</tr>
<tr>
<td>c2670</td>
<td>2678</td>
<td>8034</td>
<td>16070</td>
<td>26783</td>
<td>2538</td>
</tr>
<tr>
<td>c3540</td>
<td>3587</td>
<td>10762</td>
<td>21525</td>
<td>35875</td>
<td>3338</td>
</tr>
<tr>
<td>c5315</td>
<td>5204</td>
<td>15612</td>
<td>31255</td>
<td>52041</td>
<td>4614</td>
</tr>
<tr>
<td>c7552</td>
<td>8333</td>
<td>25001</td>
<td>50002</td>
<td>83336</td>
<td>7026</td>
</tr>
</tbody>
</table>

If the number of suspected gates is high, the number of input vectors will be too large to be solved by any SAT solver for any reasonable circuit. Let \( N \) be the number of the gates in a combinational circuit, \( p \) is the ratio between the number of suspected gates and \( N \). Table I shows the number of input vectors required for different benchmark circuits (ISCAS85) for different values of \( p \), and compared with the value of a fully gate camouflaged design where each gate can have one of four possible functionalities. We considered \( x = 2 \) for all the circuits. It is evident from the data in the table that for a number of suspected gates greater than 10% of the number of total gates in the design, the number of input vectors can be significantly large.

Fig. 7. System CPU time taken by incremental SAT solver [9] to resolve the obfuscated design varying the possible number of circuit configurations modeled by input vectors.

To demonstrate the difficulty for a SAT solver to come up with a solution with a vast number of circuit configurations, we have simulated (ISCAS85) benchmark circuits varying the number of possible configuration models modeled by input vectors (by varying the number of possible fake gates as illustrated in Figure 6). We have used the incremental SAT solver demonstrated in [9]. This simulation was executed on an Intel Core i7-6700 3.4GHz Quad Core processor with 16Gb of RAM running Ubuntu 16.04 LTS x86_64 Operating System. The RE effort is calculated in terms of seconds (System CPU time) and the results that are hovering above \( 10^6 \) seconds were deemed to be unsolvable. The results for c499, c1908, c5315, and c7552 are shown in Figure 7.

We have found that for reasonably large circuits (starting from c2670), when the number of input vectors exceeds 1000 bits, the circuit becomes unsolvable. The number of possible circuit configurations exceeds \( 2^{1000} \) when the input vectors exceed 1000 which is too large a search space for a SAT solver to solve in any acceptable time-frame for reasonably large circuits. If only 10% of the gates in the design are suspected (\( p = 0.1 \)), the number of input vectors to model the obfuscated circuit exceeds well over 1000 for large circuits.

B. VLSI Test-based Attacks on Obfuscated IC

In our proposed obfuscation methodology, the inserted fake gates will corrupt all the POs. For the golden chip, a fake gate output will not propagate to the PO. Yet, in the attackers recovered netlist, it will always propagate to some POs. However, the attacker is unaware of the number of the fake gates and their positions in the design. Hence, he can not ensure that the suspected fake gate output has not interfered with other fake gates in the design during propagation. Additionally, all the gates in the fanout/fanin cone of the fake gate will propagate corrupted output to the POs. Therefore, any approach to isolate and validate a fake gate through fault excitation and sensitization will not be effective in deobfuscating these circuits.

V. DESIGN OVERHEAD

The holographic gates do not consume any power. The metasurface used to cloak the interconnect segments can be made of conductive material (Cu) which will reduce the interconnect resistance by providing a parallel conducting path. The floating interconnects used for routing the fake gate inputs and outputs will result in additional capacitive loads to the connected gates. Nevertheless, only for a few fake gates, this capacitive load will be negligible. Fabrication of the metasurfaces will require a few more masks and additional process steps which can be easily integrated to the current CMOS processing steps [12], [14].

VI. CONCLUSION

We have presented a novel IC obfuscation technique based on metasurface holography which can be used to protect semiconductor IPs. We showed that an attacker will recover a partially incorrect netlist using optical or X-ray based RE on such circuits. Recovery of the original netlist from this obfuscated netlist will require to identify the fake gates and cloaked interconnects which have been deliberately introduced in the design. Since the attacker cannot identify the locations of the fake gates from layout inspection, the RE effort is considerably higher than the conventional camouflaging methodologies. Only a few fake gates and cloaked nets in the design can provide more security against optical RE than a fully camouflaged design using existing methodologies.

ACKNOWLEDGEMENTS: This work is supported by SRC (2847.001), NSF (CNS- 1814710, CCF-1725577, CCF-1718474, DGE- 1723687 and DGE-1821766) and DARPA Young Faculty Award (D15AP00089). X. Ni and Y. Ding knowledge support from the Gordon and Betty Moore Foundation.

REFERENCES