ABSTRACT
The globalized semiconductor supply chain is vulnerable to hardware attacks including: Trojans, piracy of intellectual properties (IPs) and/or overbuilding of integrated circuits (ICs), reverse engineering, side-channels, and counterfeiting. In this paper, we explain the threat models, the state-of-the-art defenses, and the metrics used to evaluate the defenses. The threat models outlined in this paper enable one to understand the attacks. Defenses and metrics can help defenders to build stronger countermeasures and evaluate them against other protection techniques using the metrics.

Keywords
Hardware Trojans, Reverse Engineering, IP/IC Piracy, Side-Channel Attacks, Counterfeiting, Camouflaging

1. INTRODUCTION
The semiconductor supply chain shown in Figure 1 is distributed worldwide [1, 2]. The figure shows SoC design flow and system design. Designing an SoC involves procuring intellectual property designs (IPs) from outside design houses, designing in-house components, combining them, and generating the layout through several synthesis and verification steps. The foundry manufactures the integrated circuits (ICs), which are then tested. Fault-free ICs are then packaged and sold.

This semiconductor supply chain is vulnerable to the following attacks. Rogue elements may insert malicious circuits (a.k.a., hardware Trojans) into the design [1]. An attacker may steal and claim ownership of the IP, resulting in IP piracy. An untrusted foundry may overbuild ICs and sell them illegally [3]. An attacker can reverse engineer the functionality of an IC/IP [4]. Furthermore, side-channels such as power and timing information can be used to compromise hardware implementation (for example, leaking secret keys of cryptographic algorithm implementations [5]). During system design, as shown in Figure 1 (shaded region on the right hand side), faulty, low-grade ICs can pollute the supply chain. In addition, ICs from outdated systems may be recycled and used into the target system. This is called counterfeiting.

This paper surveys these hardware attacks: Trojans (Section 2), IP piracy/IC overbuilding (Section 3), reverse engineering (Section 4), side-channels (Section 5), and counterfeiting (Section 6). For each attack, we explain the threat model, the state-of-the-art defenses, and the metrics used to evaluate the defenses. Section 7 concludes the paper.

2. HARDWARE TROJANS

Figure 2: Two hardware Trojan attack scenarios: (i) foundry and (ii) 3PIP vendor. The devil depicts an attacker, the shield represents a defender, and "?" indicates an untrustworthy entity.

A Hardware Trojan is a malicious modification to a circuit by an attacker. A Trojan can control, modify, disable, or monitor the contents and communications of the circuit [6, 7, 8].

2.1 Threat models
Figure 2 illustrates the two hardware Trojan attack scenarios. In the first scenario, an attacker in the foundry inserts a Trojan into the design [1]. An attacker may steal and claim ownership of the IP, resulting in IP piracy. An untrusted foundry may overbuild ICs and sell them illegally [3]. An attacker can reverse engineer the functionality of an IC/IP [4]. Furthermore, side-channels such as power and timing information can be used to compromise hardware implementation (for example, leaking secret keys of cryptographic algorithm implementations [5]). During system design, as shown in Figure 1 (shaded region on the right hand side), faulty, low-grade ICs can pollute the supply chain. In addition, ICs from outdated systems may be recycled and used into the target system. This is called counterfeiting.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICCAD 32, November 18 - November 21 2013, San Jose, CA, USA.
Copyright 2013 ACM 78-1-4799-1071-7/13/ $31.00.
2.3 Metrics

(i) Probability of detection: the ratio of the number of Trojans detected by the technique to the total number of Trojans in the design. (ii) Probability of false alarm: the ratio of the number of Trojan-free designs that are incorrectly classified as Trojan to the number of Trojan-free designs. (iii) Number of required clock cycles to detect a Trojan in a 3PIPs scenario.

3. IP PIRACY AND IC OVERBUILDING

An attacker with access to IP or an IC can steal and claim ownership of it and/or can overbuild and sell them illegally [3, 19].

3.1 Threat models

Figure 3 illustrates the threat model for piracy and overbuilding. In scenario 1, the attacker in the SoC integration house can pirate the 3PIP or use more than the licensed number of 3PIP instances. In scenario 2, the attacker in the foundry can pirate the 3PIP after extracting it from the layout of the design. In scenario 3, the attacker in the foundry can pirate the IC design and/or overbuild.

3.2 State-of-the-art defenses

Piracy and overbuilding can be prevented by obfuscation, watermarking, fingerprinting, and metering. In scenarios 1 and 2, the 3PIP vendor can protect his IP by obfuscating it, or by embedding his watermark or fingerprint it. In scenario 3, the SoC integrator can obfuscate or embed his watermark or fingerprinting the design before delivering it to the foundry.

Obfuscation hides the functionality and implementation of a design by inserting additional gates into it. When a wrong value is applied to these gates, they modify the functionality of the design. In one type of obfuscation, additional (black) states are introduced in the finite state machine (FSM) [19, 20]. The FSM is modified in such a way that the design reaches a valid state only on applying the correct key. If the key is withdrawn, the design ends up in a black state. In another type obfuscation, XOR/XNOR gates [3, 21] and memory elements [22] are added. The obfuscated design will function correctly only on applying the correct value to these gates and memory elements.

Watermarking is done by embedding a designer’s signature in the design [23]. The designer can reveal the watermark and claim ownership of an IC/IP. Watermarks are constructed by adding black states to the FSM, and secret constraints during high-level synthesis, logical, and physical synthesis [24, 25] and during FPGA design [26].

Fingerprinting helps the defender to track down the source of piracy by embedding the buyer’s signature (for instance, his public key) along with the designer’s watermark [27]. When challenged, the designer can reveal the watermark to claim the ownership and the buyer’s signature to reveal the source of piracy. For example, the power, timing, or thermal fingerprint of an IC is revealed on applying a set of input vectors. Fingerprinting can be also be applied during high-level, logical, and physical synthesis [27]. Another possibility is to use fingerprints from an IC’s SRAM [28].

Some fingerprinting techniques use physical unclonable functions (PUFs). PUFs are circuit structures that extract a unique set of fingerprints from each IC [29].

Metering is a set of tools, methodologies, and protocols used to track the manufactured IC. In passive metering, part of the functionality is used for metering, even for the ICs manufactured from the same mask [30]. The identified ICs may be matched against their record in a database. This will reveal unregistered ICs or overbuilt ICs. In active metering, parts of the IC’s functionality can be only accessed, locked, or unlocked by the designer and/or IP rights owners [20].

3.3 Metrics

Obfuscation: Metrics for obfuscation include: (i) Number of brute force attempts required to unlock the FSM or to determine the secret key [20, 21]. (ii) Hamming distance between the outputs of an obfuscated netlist on applying an incorrect key (or configuration) and the original netlist [22, 31]. (iii) Number of input patterns that produce an incorrect output on applying an incorrect key to the design [19].

Watermarking [24]: Metrics for watermarking include: (i) Probability of a watermarking algorithm generating the same solution for different buyers’ signatures. (ii) Probability of an attacker changing one or more watermarking bits by modifying the design.

Fingerprinting: Metrics for fingerprinting include: (i) Average Hamming distance between the responses to the same challenge obtained from two different ICs. (ii) Average Hamming dis-
4. REVERSE ENGINEERING

Reverse engineering (RE) of an IC involves (i) identifying the device technology used [32], (ii) extracting its gate-level netlist [4], and/or (iii) inferring the implemented functionality [33]. Techniques and tools have been developed to reverse engineer ICs [34, 35]. RE can be misused to steal and/or pirate a design, to identify the device technology, and to illegally fabricate the target IC.

4.1 Threat models

Figure 4 illustrates the threat models for RE. In scenario 1, the attacker in the SoC integration house can reverse engineer the 3PIP. The functionality of 3PIP modules can be extracted by behavioral matching against a library of known components [36], or by performing Boolean satisfiability analysis against a library of components [37]. In scenario 2, the attacker in the foundry can extract the 3PIP from the layout of the IC. In scenario 3, the attacker in the foundry can reverse engineer the IC. He can extract the transistor-level netlist from the layout [35], and then the gate-level netlist from it [38]. In scenarios 4–8, the user performs reverse engineering. He may depackage the IC, delay it, image the layers, stitch those images, and extract the netlist.

4.2 State-of-the-art defenses

Obfuscation (see Section 3.2) and camouflaging are two defenses that have been proposed to thwart RE. In scenarios 1, 2, 4 and 7, a 3PIP vendor can obfuscate his IP. In scenarios 3, 5 and 6, an SoC integrator can obfuscate his design. A trusted foundry can camouflage the layout (scenarios 6–8), providing an additional layer of defense beyond obfuscation.

Camouflaging is a layout-level technique that hampers image processing-based extraction of a gate-level netlist from an IC. In one embodiment of camouflaging, the layouts of standard cells are designed to look alike, resulting in incorrect extraction of the netlist. IC camouflaging can leverage unused spaces in an IC by filling them with filler cells [39], can use programmable standard cells [40], or can use dummy contacts [41].

4.3 Metrics

Metrics for obfuscation are given in Section 3.3.

Reverse engineering: Metrics for reverse engineering include:
(i) Percentage of gates correctly extracted from a layout [4]. (ii) Percentage of gates whose functionality is correctly inferred [37]. (iii) Number of signals correctly matched between the signals in the component with known functionality and the signals in the target design [36].

Camouflaging: Metrics for camouflaging include: (i) Number of brute force attempts required to identify the functionality of camouflaged gates [42, 43]. (ii) Hamming distance between the outputs of the original netlist and the netlist in which the functionality of camouflaged gates are assigned by the attacker [42].

5. SIDE-CHANNEL ATTACKS

Side-channel attacks exploit the leakage of physical information when an application is being executed on a system [5]. Side-channel attacks are powerful and have broken all major cryptographic algorithms [44].

5.1 Threat models

Information about secret keys can leak from an IC through its power consumption traces [5], timing traces [45], electromagnetic emanations [46], photonic emissions [47], scan chains [48], and faults injected in the designs [49]. Information leaked from side-channels do not completely overlap with each other. Hence, an adversary can combine the information leaked from several side-channels to increase the effectiveness of the attack [44].

5.2 State-of-the-art defenses

Leakage of information can be reduced by either decreasing the dependency between the side-channel trace and secret key, or by injecting noise into the side-channel [50]. The dependency between power consumption and secret key can be reduced by using dynamic and differential logic [51], asynchronous logic [52], and dual-rail with pre-charge logic [53].

Noise injected into the side-channels by adding dummy circuits that consume random amount of power for each transaction, or by performing random operation independent of the secret keys can reduce information leakage [54].

Though leakage reduction and noise injection do not provide theoretical security, they increase the effort of an attacker to extract the secret keys. For instance, decreasing the SNR of the side-channel information by a factor of $K$ linearly or quadratically increases the number of required input patterns for side-channel analysis [55].

Key update prevents the accumulation of side-channel information by regularly updating the secret key after a pre-determined number of input patterns [56].

Leakage-resilient cryptography entails designing cryptographic primitives that are intrinsically resilient to information leakage through side-channels [57, 58, 59].

5.3 Metrics

Metrics for defenses against side-channel attacks include: (i) Number of input patterns required by an attacker to retrieve the secret key. (ii) Maximum amount of information leaked per input pattern [55]. (iii) Correlation between the secret key and the side-channel trace per input pattern [60].
6. COUNTERFEITING

A counterfeit semiconductor component is an illegal forgery or imitation of the original component. Counterfeiting is often performed by one of the many entities in the semiconductor supply chain, including new product vendors or secondary (recycled) IC vendors. IC counterfeiters profit by selling a cheaper and low quality IC. Although the primary incentive for selling fake ICs is financial, the ease of inserting hardware Trojans or spyware in fake ICs makes them a real security threat for the system that contains them [61].

6.1 Threat models

<table>
<thead>
<tr>
<th>Design</th>
<th>Test</th>
<th>Re-packaging/Recycling</th>
<th>PCB Assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: Threat scenarios for counterfeiting.

Figure 5 illustrates the counterfeit IC threat models. In scenario 1, defective ICs, i.e., those which failed the manufacture-time testing and have been discarded, are used in consumer products [61]. An untrustworthy entity at the test facility can be the source of leaking defective ICs. In scenarios 2 and 3, a dishonest entity in the IC supply chain mislabels a product and sells it as another IC potentially through a vendor [61]. The functionality of the mislabeled IC is likely not the same as the intended IC specification. In addition, used and recycled ICs are repackaged as new [62, 63]. The attackers are the second-hand vendors who buy or collect old electronic systems and remove ICs from them. The extracted ICs are repackaged and sold as new, in particular for spare parts for older electronic systems which are out of production line.

6.2 State-of-the-art defenses

In scenario 1, the faulty and low-grade chips can be detected by re-testing them before deploying into a system. In scenario 2, proactive techniques like hardware metering, fingerprinting, watermarking (Section 3.2), and sensors to determine IC aging are the used at the design phase to enable counterfeit detection [64, 61, 65]. Mislabeled chips can be detected by visual inspection, depackaging, or X-ray photography of the packages [65]. In scenario 3, using non-invasive measurements such as power and timing, one can determine an IC’s age or reliability, thereby potentially detecting used/old/recycled ICs [64, 62, 63]. One can also implement aging sensors to report the amount of usage (or age) of the ICs which include them.

6.3 Metrics

The metrics for hardware metering, fingerprinting, watermarking are discussed in Section 3.3. Metrics for counterfeit detection using circuit-aging include: (i) Probability of detection is the ratio of the number of counterfeit ICs detected by the technique to the total number of counterfeit ICs [62, 63]. (ii) Probability of false alarm is the ratio of the number of genuine ICs that are incorrectly classified as counterfeit ICs to the number of genuine ICs [64, 62].

7. CONCLUSIONS

Threat models, the state-of-the-art countermeasures, and the metrics used to evaluate the defenses against Hardware Trojans, IC and IP piracy, reverse engineering, side-channels, and counterfeiting were introduced. Until now, most evaluations of defenses have been informal and anecdotal. The authors believe that the metrics are an important first step in formalizing the evaluation of the strengths of defenses. Similarly, a consistent classification of threat models was not available. By organizing the threat/defense scenarios, we hope the countermeasures can be compared against one another based on the target threat model and the corresponding metrics.

8. ACKNOWLEDGMENTS

This research was supported in parts by an Office of Naval Research grant (ONR R17460) and a NSF grants to Rice University (CNS-1059416) and NYU-Poly (CNS-1059328).

9. REFERENCES


