A Survey on Integrated Circuit Trojans

Halit Türksönmez* Mehmet Hilal Özcanhan
Department of Computer Engineering, Dokuz Eylül University, PO box 35390, İzmir, Turkey

Abstract
Traditionally, computer security has been associated with the software security, or the information-data security. Surprisingly, the hardware on which the software executes or the information stored-processed-transmitted has been assumed to be a trusted base of security. The main building blocks of any electronic device are Integrated circuits (ICs) which form the fabric of a computer system. Lately, the use of ICs has expanded from handheld calculators and personal computers (PCs) to smartphones, servers, and Internet-of-Things (IoT) devices. However, this significant growth in the IC market created intense competition among IC vendors, leading to new trends in IC manufacturing. System-on-chip (SoC) design based on intellectual property (IP), a globally spread supply chain of production and distribution of ICs are the foremost of these trends. The emerging trends have resulted in many security and trust weaknesses and vulnerabilities, in computer systems. This includes Hardware Trojans attacks, side-channel attacks, Reverse-engineering, IP piracy, IC counterfeiting, micro probing, physical tampering, and acquisition of private or valuable assets by debugging and testing. IC security and trust vulnerabilities may cause loss of private information, modified/altered functions, which may cause a great economical hazard and big damage to society. Thus, it is crucial to examine the security and trust threats existing in the IC lifecycle and build defense mechanisms against IC Trojan threats. In this article, we examine the IC supply chain and define the possible IC Trojan threats for the parties involved. Then we survey the latest progress of research in the area of countermeasures against the IC Trojan attacks and discuss the challenges and expectations in this area.

Keywords: IC supply chain, IC security, IP privacy, hardware trojans, IC trojans
DOI: 10.7176/CEIS/12-2-01
Publication date: April 30th 2021

1. Introduction
For a long time, hardware has been assumed as root-of-trust for the entire computer system and used as a virtual layer that runs the code sent from the software layer. Meanwhile, computer system security has been associated with software security or information security. Consequently, the studies on hardware security are mostly associated with the performance improvement of crypto-related algorithms embedded in hardware, such as crypto ICs (Preneel & Takagi 2011; Jin 2015). Hardware copyright protection is also considered in the hardware security domain (Rad et al. 2008). For many years, computer systems security researchers assumed that adversaries could not compromise ICs easily, or profit by compromising the ICs. The assumption was so extensive that the security of the IC supply chain was not even considered. The alarm was raised when illegal IC duplicates started to appear in the market.

ICs are the main building blocks of any electronic device that forms the fabric of a computer system. Their usage has been increased over the years, from handheld calculators and desktop computers to servers, smartphones, and Internet-of-Things (IoT) devices. In Figure 1, the income from the global semiconductor market for ICs between 2009 and 2021. In 2021, the income from IC sales is predicted to reach US $383.84 billion by Statista (Alsop 2020).

However, this significant growth in the semiconductor market created intense competition among IC vendors such as Intel, Samsung, Broadcom, and Qualcomm, leading to new trends in IC manufacturing (Salmani 2018). SoC design based on IP, and a globally spread supply chain for production and delivery of ICs are the foremost of these trends.
A typical example of globally spread IC production process, spanning multiple countries, is shown in Figure 2 (SIA 2016). These emerging trends are followed by the reduction of IC manufacturers’ control over the design and production stages. Consequently, many security and trust weaknesses and vulnerabilities arise (such as IC Trojans attacks, IP piracy and reverse-engineering, IC counterfeiting, etc.) (Salmani 2018; Bhunia & Tehraniipoor 2018; Bhunia & Tehraniipoor 2019; Farahmandi et al. 2020; Behnam 2018; Rostami et al. 2013; Rostami et al. 2014; Bhunia et al. 2014; Belous & Saladukha 2020; Qu & Yuan 2014; Li et al. 2016; Karri et al. 2010; Özcanhan & Türksönmez 2020).

In the rest of this article, we examine the IC supply chain and describe the potential IC Trojan threats faced by the parties included. Then we explain and compare IC security and IC trust. Finally, we survey the advances in recent studies about the countermeasures against the Trojan attacks and discuss the challenges and expectations in this area.

2. Differences between IC Security and IC Trust
IC security problems emerge from its built-in vulnerability to attacks, such as scan-based attacks, side-channel attacks, and probing attacks. However, IC trust problems emerge from the participation of untrusted parties in the lifecycle of an IC, such as:
- untrusted intellectual property (IP) or electronic design automation (EDA) / computer-aided design (CAD) tool vendors,
- untrusted design,
- untrusted fabrication,
- weak testing,
- insecure distribution facilities.

The parties included in the above IC production activities are liable to violate the trustworthiness of consumers towards an IC. Potentially, they may cause deflections from expected trustworthiness, functionality, reliability, or performance. Trust problems usually escalate to security problems. For instance, an untrusted IP vendor could insert malware entities in a design, which may cause information theft, or denial-of-service (DoS) when the IC goes on to the field. Moreover, trust issues may also cause other problems, such as low energy-
efficiency or performance, or reduced safety, or reliability problems. The horizontal structure of the semiconductor trade model and the growing nature of the globally spread IC supply chain are causing the IC trust problems even more important. Thus, it drives novel research and development studies on IC design for trust assurance and trust verification (Bhunia & Tehranipoor 2019; Farahmandi et al. 2020; Behnam 2018).

3. Vulnerable IC Supply Chain

Figure 2 shows a typical IC production process spanning over multiple countries, while Figure 3 shows detailed IC supply chain phases. IC supply chain shown in Figure 3 spreads globally, which causes new IC security and trust issues to emerge from the global trends in IC design, fabrication, and distribution. IC design flow simply shows the stages and assets relevant to this article. System design framed with the dotted lines represents how the forged and poor-quality parts are inserted into the supply chain. Designing an IC covers supplying IP from third-party design houses, joining together IP and in-house designed components, and composing the IC layout. Afterward, the overall design is dispatched to the foundry that produces a mask and fabricates the ICs. Then, the ICs are tested at the fabrication site and test plants. Finally, flawless ICs are packaged up and sent to market. As it can be suspected, there are lots of stages in the IC supply chain, where malicious activities can happen. For instance, a malicious employee, who can reach the design at an untrusted foundry, could insert IC Trojans into any of the photomasks.

Figure 3. IC supply chain (Rostami et al. 2013; Rostami et al. 2014)

4. Known IC Attack Vectors

Table 1 shows the known security problems arising out of untrusted design, fabrication, and test stages of an IC. Attack vectors are instruments or ways for adversaries to reach ICs for malicious intention. One example is to exploit the IC to get the valuable/confidential data stored. Some attack vectors provide the ability for exploiting implementation problems, by physical tampering and side-channel attacks. Trojan attacks use the lack of control at IC fabrication stage, as an advantage. The Trojan attack vector is shown in yellow in Table 1. Attack surface can be defined as the total exposures of all possible security risks. It can also be described as the total of all known, unknown, and possible vulnerabilities. An adversary can utilize one or more vulnerabilities and start an attack, ending with obtaining confidential data from the system the IC is a part of. The smallest possible attack surface is considered as a primary target for countermeasure developers.

At first glance, IC Trojan attacks may be considered as part of IC trust problems. But the consequences of a Trojan attack on an IC may result in the leakage of secret data in that IC. Therefore, the impact factor of IC Trojan Attacks may be considered greater compared to the other IC attacks. This is the reason why we focused on IC Trojans, in this article.

Table 1. Attack vectors for the stages in an IC’s lifespan. (ICS: IC security related attack, ICT: IC trust related attack)
5. IC Trojan Attacks
IC Trojans are hostile alterations in the functional behaviors of an IC (Rostami et al. 2013; Rostami et al. 2014; Bhunia 2018). These alterations are foreign, unknown, and unplanned by the IC designer, which could have harmful impacts on the IC. IC Trojans have three basic features: adversarial purpose, avoidance of detection, and sparseness of activation (Bhunia et al. 2014; Belous & Saladukha 2020). An IC Trojan always has the same purpose: application of an unintentional activity to compromise the confidentiality, integrity, or authentication of the underlying IC. Confidentiality, integrity, authentication triplet is referred to as CIA-Triad in the literature (Parker 2010).

In Figure 4(a), an unaltered IC having two inputs and one output is symbolized. In Figure 4(b), an IC Trojan that has a trigger and a payload is symbolized. In Figure 4(c), the Trojan implanted into an IC is shown. In such a scenario, whenever the Trojan becomes active, its payload inverts the output of the IC, such as from O to O'. The size of IC Trojans may be tiny or big relative to the remaining original circuitry, changing from a few transistors to millions of transistors (Bhunia et al. 2014). Trojans may be in various shapes. They are mostly activated by a sequential, combinational, or hybrid digital circuit. But, they can also be activated by an analog signal. The Trojan payload can be analog or digital, but each is specially designed to produce malicious results when triggered.

6. IC Trojan Threat Patterns
Generally, the design and production procedure of an IC can be separated into three phases: Development of IP core, development of IC, and fabrication of IC. Thus, three types of companies (IP vendors, IC developers, and foundries) have a chance to inject IC Trojans. Trojans can be inserted at any stage of the three phases, by adversaries. The stage of insertion results in various adversarial patterns. Table 2 shows seven potential IC attack patterns for the IC Trojan attacks (Xiao et al. 2016).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Description</th>
<th>IP Vendor</th>
<th>IC Developer</th>
<th>Foundry</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Untrusted IP vendor</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>Trusted</td>
</tr>
<tr>
<td>P1</td>
<td>Untrusted foundry</td>
<td>Trusted</td>
<td>Trusted</td>
<td>Untrusted</td>
</tr>
<tr>
<td>P3</td>
<td>Untrusted EDA tool or rogue employee</td>
<td>Trusted</td>
<td>Untrusted</td>
<td>Trusted</td>
</tr>
<tr>
<td>P4</td>
<td>Commercial off-the-shelf (COTS) component</td>
<td>Untrusted</td>
<td>Untrusted</td>
<td>Untrusted</td>
</tr>
<tr>
<td>P5</td>
<td>Untrusted design house</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>Untrusted</td>
</tr>
<tr>
<td>P6</td>
<td>Fabless IC design house</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>Untrusted</td>
</tr>
<tr>
<td>P7</td>
<td>Untrusted IC developer with trusted IPs</td>
<td>Trusted</td>
<td>Untrusted</td>
<td>Untrusted</td>
</tr>
</tbody>
</table>

**Pattern P1:** Untrusted Third-party IP Vendor—Driven by the demands of global IC industry (lower expenses, rapid marketing, etc.), most IC designers have to obtain some third-party IP cores for their designs. Adversaries at the untrusted vendor have the chance to implant IC Trojans, concealed from the IC designer, into the third party IP.

**Pattern P2:** Untrusted Foundry—Fabless design houses outsource the production of ICs to untrusted (mostly overseas) manufacturers. Adversaries employed in these untrusted foundries can reach the layers of the design and insert Trojans into the photolithography masks.

**Pattern P3:** Untrusted IC Developer—Production of complicated IC designs requires trained IC designers and professional design tools. In this pattern, adversaries are the insiders, who can use untrusted Electronic Design Automation (EDA) and Computer Aided Design CAD tools.
**Pattern P4: Untrusted Commercial off-the-shelf (COTS) Components**—Several COTS elements are inserted into designs. Mostly, COTS products are cheaper than custom-designed ones. However, no development stage can be trusted for Trojans, in COTS.

**Pattern P5: Untrusted Design House**—In this pattern, it is assumed that the whole supply chain is untrusted except foundry, i.e. ICs are produced in a trusted foundry, but third-party IP vendors and the design house are not trusted for disinfected designs.

**Pattern P6: Untrusted Outsourcer**—Pattern P6 is a combination of Pattern P1 and Pattern P2, and it concerns most fabless design houses. The designers utilize third-party IP vendors and untrusted foundries and fabricate these ICs in untrusted third-party foundries.

**Pattern P7: Untrusted System Integrator**—In this pattern, the customers expect to have a supplier that has both design and fabrication capabilities. But, untrusted system integrators exploit this expectation. The developer can use a diversity of resources to meet customer requests, but the completed hardware design may contain some inherent vulnerabilities (Rostami et al. 2013; Rostami et al. 2014; Bhunia 2018; Xiao et al. 2016).

### 7. Countermeasures against IC Trojan Attacks

Several methods for IC Trojan detection have been worked out and proposed, for many years. These methods can be grouped under three main categories, Trojan detection, design-for-trust, and split-manufacturing, as shown in Figure 5. These categories can also further be divided into several subcategories (Bhunia & Tehranipoor 2019).

![Figure 5. Taxonomy of IC Trojan countermeasures (Bhunia & Tehranipoor 2019).](image)

Trojan detection is the fundamental and most utilized method to fight the IC Trojans. Its purpose is the verification of newly fabricated ICs with the present IC designs. These methods are used at the design phase to confirm IC designs or after the fabrication phase to validate fabricated ICs. They can be grouped under two subcategories, destructive and nondestructive methods (See Figure 5). Reverse-engineering is utilized by destructive methods to open an IC case and get images of layers to rebuild the design-for-trust (DFT) confirmation. Non-destructive methods can be grouped under two subcategories, functional test and side-channel analysis. Functional test techinics apply test vectors to activate Trojans and check against the responses with the correct results (Bhunia et al. 2014; Banga & Hsiao 2009; Chakraborty & Bhunia 2009). IC Trojans are detected by side-channel signal analysis methods using circuit parameters, such as power dissipation (Agrawal et al. 2007; Aarestad et al. 2010), temperature (Forte et al. 2013), delay (Jin & Makris 2008; Xiao et al. 2013), and radiation (Stellari et al. 2014; Zhou et al. 2015). The side effects from Trojan activation alter power and/or heat dissipation, propagation or contamination delays, or radiation patterns due to additional circuit activity. Design stage Trojan detection methods are utilized to support IC developers and designers for validation of IP cores and the final designs. Present design stage detection techniques can be classified into formal verification (Jin et al. 2013; Guo et al. 2015; Rajendran et al. 2015; Rajendran et al. 2016), code coverage analysis (Hicks et al. 2010; Sturton et al. 2011), logic testing, functional analysis (Waksman et al. 2013), and structural analysis (Salmani & Tehranipoor 2013; Salmani et al. 2013; Tehranipoor et al. 2013).

DFT methodologies can be grouped under three subcategories with respect to their goals: Trojan prevention, facilitate detection, and trustworthy computing (See Figure 5). Trojan prevention methods are formed by techniques that aim to prevent IC Trojan implantation by adversaries. The adversaries need to know the function of the design first, in order to able to implant Trojans. Usually, reverse engineering is used by adversaries, who
are not employed in the design house, to determine circuit functionality (Bhunia & Tehranipoor 2019). These techniques can be sub classified into camouflaging (Rajendran et al. 2013; Cocchi et al. 2014), logic obfuscation (Roy et al. 2010; Baumgarten et al. 2010; Wendt et al. 2014), and functional filler cell (Xiao & Tehranipoor 2013). Facilitate detection can be sub classified into runtime monitoring (Forte et al. 2013; Narasimhan et al. 2012), side-channel analysis (Salmani & Tehranipoor 2012; Rajendran et al. 2011; Ramdas et al. 2014) and functional test; which targets triggering a Trojan from inputs and observing the Trojan impact from outputs (Salmani et al. 2012; Zhou et al. 2014). Trustworthy computing is the last class of DFT on untrusted elements (McIntyre et al. 2010; Liu et al. 2014).

Recently, split-manufacturing has been offered as an approach to IC foundries to be able to reduce the risks of Trojan insertion in IC design (Bhunia & Tehranipoor 2019). Current split manufacturing methods trust either 2D integration (Vaidyanathan et al. 2014; Jagasivamani et al. 2014; Hill et al. 2013), 2.5D integration (Xie et al. 2015), or 3D integration (Valamehr et al. 2013).

8. Future Work
The competition among IC vendors in the semiconductor market forced a trade-off, between IC security and the market cost-performance requirements on ICs. This development is expected to result in novel IC trust issues in the global IC supply chain. ULSI (Ultra Large Scale Integration), the successor of VLSI (Very Large Scale Integration) has become the main propulsions of the global IC market. With the advances in ULSI technology, IC Trojans will be more accurate, smaller, more concealed, and more difficult to be detected. Therefore, the adversaries are expected to launch new, more sophisticated, and unexpected, attacks which are even more difficult to be handled by existing countermeasures. Thus, countermeasure techniques against emerging IC Trojan attacks will need continuous further development.

9. Conclusion
The purpose of this article is to demonstrate the latest advances in the IC Trojan attack vectors and the countermeasures against the attacks. At the same time, it is intended to provide a general understanding and guidance to those who want to engage in IC Trojan research. Fighting against the IC Trojan threat will require everlasting and hard endeavor. With proper, progressive, scientific approaches, the difficulty and cost of IC Trojan attack elimination can be achieved.

References


doi: 10.1109/MC.2010.284