Self-Consistency and Consistency-based Detection and Diagnosis of Malicious Circuitry

Sheng Wei and Miodrag Potkonjak

***Abstract*—Hardware Trojans (HTs) have become a major concern in the modern IC industry, especially with the fast growth in IC outsourcing. HT detection and diagnosis are challenging due to the huge number of gates in modern IC designs and the high cost of testing. We propose a scalable and efficient HT detection and diagnosis scheme based on segmentation and self-consistency analysis of gate-level properties. Also, we employ a new technology named variable elimination to create subsegments from a fixed set of power measurements of the entire IC in order to minimize the number of power measurements. We evaluate our HT detection and diagnosis schemes on a set of ISCAS and ITC benchmarks.**

I. Introduction

Hardware trojans (HTs) [22] are malicious attacks on integrated circuits (ICs) that modify the functionality or impact the performance of the ICs. HTs are possibly embedded by attackers during the manufacturing process in the form of additional gates or resized gates compared to the design specification. As IC outsourcing has become more and more popular recently, HT detection and diagnosis have become a necessity for IC designers and users, because the ICs are exposed to anyone who is in charge of the manufacturing process for potential HT attacks.

A typical procedure of handling HT attacks would include two strategic steps: HT detection and HT diagnosis. HT detection is the process that determines whether any HTs exist in the circuit. If there are any, a HT diagnosis approach is required to locate the HTs in the IC in terms of their types, locations, and input pins.

Although many HT detection approaches have been proposed recently [2][29][4], the HT diagnosis problem was seldom discussed and addressed. Also, in the existing HT detection approaches, scalability has become a major concern, especially with the fast development of submicron technologies. It is challenging for the detection procedure to determine whether there is any HT embedded among millions of gates in the circuit. The cost of testing and running time have become a major concern in conducting this type of detection. Also, even if one can detect the presence of HTs accurately, it will cost much more effort to determine their exact locations.

We develop a complete and scalable solution of HT detection and diagnosis using a consistency-based gate characterization scheme. In particular, we partition a large IC into several overlapping segments and analyze the gate-level properties in each segment. We detect the HTs in the case where the overlapping gates exhibit inconsistent characterized properties in different segments. After confirming the existence of HTs, we employ a consistency-based HT diagnosis scheme to refine the scope of the HTs to a small segment.

During the process of HT detection and diagnosis, we note the major source of cost is the number of power measurements. Therefore, we develop a variable elimination scheme to generate the sub-segments from a fixed set of power measurements of the entire circuit. By reusing the existing measurements in various sub-segments, we are able to reduce the HT detection cost arising from power measurements. Also, we develop a sub-segment selection algorithm to ensure that the minimum number of sub-segments are generated and characterized, to further reduce the computational complexity of gate-level characterization.

In summary, our main technical contributions in this paper include the following:

- a scalable and efficient HT detection method based on segmentation and consistency analysis;
- a HT diagnosis scheme to determine the locations of the HTs in the circuit; and
- a self-consistency based approach to minimize the required number of power measurements in HT detection and diagnosis.

II. Related Work

In this section, we briefly review the directly related work in HT research and the supporting techniques regarding gate-level characterization (GLC).

A. Hardware Trojan Detection

Agrawal et al. [2] proposed one of the first HT detection techniques in 2007. They construct fingerprints using side channels (e.g., power and temperature) of the circuit for a specific design and authenticate the IC instances based on the fingerprints. The technique is based on the assumptions that there is no process variation, ICs are available for reverse engineering, and there are no measurement errors in the side channels.

Several early HT detection approaches employed functional test techniques. Functional tests simulate the input vectors on the circuit and monitor the outputs to see whether they match the expected patterns. For example, Wolff et al. [29] proposed the generation of test vectors that maximize the likelihood of detecting rarely switching HT gates. Also, Banga et al. [4]

Sheng Wei and Miodrag Potkonjak are with the Computer Science Department, University of California, Los Angeles (UCLA), Los Angeles, CA 90095 (Email: {shengwei, miodrag}@cs.ucla.edu). An earlier version of this paper [27] was presented at the 5th International Conference on Network and System Security (NSS'2011). This work was supported in part by the NSF under Awards CNS-0958369 and CNS-1059435.

proposed automatic test pattern generation (ATPG) techniques that employ the divide-and-conquer paradigm. Recently, HT detection methods using side channel-based analysis have been developed [17][20][12][10] . They characterize the target ICs for their manifestational properties, such as delay and power, in order to detect the embedded HTs. For example, two types of HT detection techniques analyzed pertinent ICs in terms of their delay from one flip-flop to another using either deterministic [12] or statistical methods [10]. A number of HT detection techniques advocate the use of switching power measurements[5]. Researchers from UCLA [16] advocate leakage current-based HT detection techniques.

Tehranipoor et al. [22] presented a comprehensive survey of HT detection. There are two most common conceptual mistakes in the existing HT detection approaches: (1) the authors assume that both an IC with and without HT are available, and (2) all the gates have the same PV properties. In this paper, we do not impose these two assumptions for HT attacks and detections. Furthermore, we employ segmentation-based gate characterization into the process of HT detection, which ensures the scalability of the approach.

In our previous work, we have developed a thermal conditioning based HT detection approach by checking a HT variable in the power measurement equations [26]. Also, in the work of [28], we developed a segmentation-based technique for the same HT detection method to ensure scalability of the approach. However, in [26] and [28], the characterization of all the gates in the circuit or in each segment is required, while in this paper, only the overlapping gates between segments need to be characterized. Therefore, we improve the scalability and performance of the HT detection method by using consistency checking.

B. Gate-level Characterization

Gate-level characterization (GLC) is the process of identifying the process variation in the manufactured IC [26][30][11]. Recently, there are two classes of GLC methods that have been proposed for IC synthesis and analysis. The first class conducts physical measurements of transistor parameters [9]. The second class employs nondestructive methods that measure the manifestational properties (e.g., delay, leakage power, and switching power) of the entire IC and characterize the gate-level properties. For example, some of these techniques use sophisticated mathematical techniques, such as single value decomposition and compressive sensing [13], while others rely more on statistical processing of data obtained from systems of linear or nonlinear equations [26].

III. Preliminaries

In this section, we introduce the preliminaries of our HT detection and diagnosis approaches, including process variation and gate-level characterization.

A. Process Variation

Process variation (PV) during IC manufacturing causes IC key parameters to vary from their nominal design specifications. For example, PV may vary leakage power by up to 20X and frequency by 30% on a single wafer [6]. In particular, there are two physical level properties that are major sources of PV: threshold voltage and effective channel length. For example, the effective channel length of a manufactured gate can be expressed by Equation (1), where L_{nom} is the nominal design value of the effective channel length, and ΔL is the variation caused by PV.

$$L_{eff} = L_{nom} + \Delta L \tag{1}$$

Although the accuracy of the PV models have been verified with the experimental data, they are only effective from the perspective of statistical properties, i.e., when a large number of chips are presented. For IC applications based on specific IC instances, such as those in hardware security, the PV models are not appropriate because of their lack of control over the individual chips.

B. Gate-level Characterization

In the process of GLC [25], the PV is represented as a scaling factor towards the gate-level manifestational properties, such as delay and power. Then, a system of linear equations can be obtained by summing up the gate-level properties and measuring the total power and delay. Taking leakage power as an example, the system of linear equations can be formulated as follows:

$$\tilde{p_j} = e_{sj} + e_{rj} + \sum_{i=1}^{n} K_{ij}\ \alpha_i \tag{2}$$

where $\tilde{p_j}$ is the total leakage power of the entire IC when input vector j is applied; α_i is a variable that represents the PV scaling factor of gate i; K_{ij} is the nominal leakage power of gate i when input vector j is applied; and e_{sj} and e_{rj} represent the systematic and random measurement errors, respectively. Following Equation (2), we obtain a system of linear equations by applying different input vector j and measuring the leakage power of the entire circuit. Then, we solve the system of linear equations using a linear programming (LP) solver, with an objective function that minimizes the measurement errors, to obtain the α values.

IV. Consistency-Based Hardware Trojan Detection

Our goal in this section is to address the detection of HTs using consistency analysis given the results of GLC. Our idea is based on the fact that a circuit containing HT would cause systematic bias in the total leakage power consumption, no matter where the HT is, how it is constructed, and even whether it is activated or not. With our GLC process, since there are no variables in the system of equations (shown in Equation (2)) to represent the HT, the systematic bias in the total leakage power would create inconsistencies in the equations, and the bias would be reflected in the scaling factors of regular gates in the circuit. By observing the bias in the leakage power scaling factors, we are able to detect HTs embedded in the circuit.

There are two key challenges with the consistency-based HT detection approach. Firstly, we do not assume that we have a clean circuit that does not have any HTs. Therefore, it

is difficult to observe the bias in the scaling factors caused by HTs, as there is no standard scaling factors to compare with. Secondly, since the number of gates in modern IC designs is up to the magnitude of millions, the size of the system of equations would easily exceed the computational limit of the LP solvers.

We address both challenges using segmentation. The segmentation of an IC is based on the divide-and-conquer paradigm, in which we divide a large IC into multiple small segments and characterize each of them using GLC.

Segmentation can be implemented using the SAT approach introduced in Section 3.4, where we set the set the SAT objectives as freezing a subset of the signals in the circuit. If the SAT problem is solvable, only the gates out of the frozen subset would be possible to change their coefficients in the system of linear equations. Therefore, we can represent all the frozen gates using a single variable in the system of linear equations. Consequently, only the gates controlled by the varying inputs would change their coefficients in the system of linear equations, while the other gates would have identical coefficients in all the equations. Therefore, we can represent all the frozen gates using a single variable in the system of linear equations. Furthermore, the complexity of the SAT problem can be significantly reduced by using a pre-processing step, where we divide the inputs of the circuit into several independent groups and formulate SAT problems for the trasitive fan-out gates of each individual group. In this way, we ensure that the SAT solving process is scalable to large industry designs.

In this way, the size of the LP is greatly reduced, to the extent that can be handled by LP solvers.

Furthermore, there are overlapping gates across segments. This provides us with an opportunity to characterize a single (overlapping) gate in multiple sub-circuits (segments), and thus observe possible bias in scaling factors due to the presence of HT. For example, suppose there are two segments A and B with a overlapping gate X, we can characterize the scaling factors of X in both segment A and B, namely α_a and α_b. Our idea is that α_a and α_b will be consistent if there is no HT present in either segment A or segment B, as ensured by the accuracy of GLC in both segments. In the case where HT exists in either A or B, there exists inconsistencies in the segment that contains the HT, and the resulting scaling factor (α_a or α_b) will be biased to reflect the inconsistencies. In the case where HTs exist in both segments A and B, since the two segments are different in terms of their gates and overall leakage power, the systematic bias caused by the HTs will be different in the two segments, which will again result in different values for α_a and α_b. We use the average discrepancy (d_{avg}) in calculated scaling factors of overlapping segments as an indicator of whether a HT is present or not. d_{avg} is calculated as the average standard deviation of the scaling factors of the same gate in the overlapping segments.

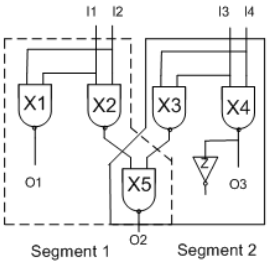
We illustrate our segmentation-based HT detection scheme using an example shown in Fig. 1. For the sake of brevity and clarity, the circuit has only five NAND gates (named $X1$ to $X5$) as shown in Fig. 1(a). We adopt normalized values as shown in Fig. 1(b) for their nominal leakage power. Our goal is to determine whether there is any HT embedded in the circuit. We first partition the circuit into two segments, as shown in Fig. 1(a). We obtain Segment 1 (gates $X1$, $X2$, and $X5$) by freezing inputs 3 and 4 and by applying different input vectors to inputs 1 and 2. Similarly, we obtain Segment 2 (gates $X3$, $X4$, and $X5$) by freezing inputs 1 and 2.

Next, we conduct GLC for each individual segment. In particular, we apply four input vectors to each segment that provide four sets of nominal leakage values for gates $X1$, $X2$, and $X5$ in Segment 1 and gates $X3$, $X4$, and $X5$ in Segment 2. For HT detection, we show three cases where HT exists or does not exist in Segment 1 and Segment 2. We assume that we do not know whether the circuit has HT in advance, and we form the system of linear equations to conduct GLC for each segment as shown in Fig. 1(c).

In case 1 (where HT is present in neither Segment 1 nor Segment 2), the values of overlapping gate $X5$ in the two segments are identical. In case 2 (where a single HT gate is present in Segment 1), the two calculated values of $X5$ have a 30.8% discrepancy. Finally, in case 3 (where HT is present in both Segment 1 and Segment 2), the values of $X5$ have a 3.7% discrepancy. These results indicate that the discrepancy between overlapping gates in multiple segments can serve as an indicator for the systematic bias in leakage power caused by embedded HTs. Therefore, we check the GLC results of overlapping gates between pairs of segments in the circuit. As long as the segments can cover all the gates in the circuit, our approach can detect any HTs embedded in the circuit. Furthermore, the use of segmentation ensures the scalability of GLC, since the number of gates being characterized in each system of linear equations is drastically reduced.

V. Consistency-based Hardware Trojan Diagnosis

The goal in HT diagnosis is to determine the locations of the HTs in the circuit if any exist, so that one can either remove or mask the HTs from the circuit. We design a scalable HT diagnosis scheme based on our consistency-based HT detection method. We have observed that one can detect the existence of HTs using two segments with overlapping gates. However, the HT detection results do not indicate which segment the HTs may be embedded in, and thus it is difficult for the HT masking process to handle the HTs. In order to diagnose the HTs, we introduce a third segment with the same set or subset of overlapping gates and use it as an arbiter for HT diagnosis. Fig. 2 shows an example of the consistency-based HT diagnosis. We find one more segment (Segment 3) compared to the example in Fig. 1. The three segments have an overlapping gate $X5$. We vary the controlling inputs of each segment and characterize the scaling factor of all the gates. In the case where the HT is embedded in Segment 2, we have the scaling factor of $X5$ consistent in Segment 1 and Segment 3 (e.g., $\alpha 5 = 1.3$), while that in Segment 2 has a different value (e.g., $\alpha 5 = 1.7$). Then, we analyze each combination of the pair of segments following the rule that an inconsistency in the scaling factor of the overlapping gate indicates possible HTs in either of the segments, while a consistent result ensures that both of the involved segments are HT-free. For example,



(a) Segmentation

Gates X1 – X5

Input	Nominal Leakage
00	1
01	3
10	4
11	10

HT Gate Z

Input	Nominal Leakage
0	2
1	5

(b) Leakage power lookup table

Case 1: HT-free vs. HT-free

Input Vectors I1 I2 I3 I4	System of Equations (Segment 1)	Input Vectors I1 I2 I3 I4	System of Equations (Segment 2)
0000	1 α1 + 1 α2 + 10 α5 = 15.3	0000	1 α3 + 1 α4 + 10 α5 = 15.4
0100	4 α1 + 3 α2 + 10 α5 = 21	0001	4 α3 + 3 α4 + 10 α5 = 21.35
1000	3 α1 + 4 α2 + 10 α5 = 21.1	0010	3 α3 + 4 α4 + 10 α5 = 21.45
1100	10 α1 + 10 α2 + 3 α5 = 26.9	0011	10 α3 + 10 α4 + 4 α5 = 29.2
Results	α1 = 1.1; α2 = 1.2; α5 = 1.3	Results	α3 = 1.15; α4 = 1.25; α5 = 1.3

Case 2: HT-free vs. HT-present

Input Vectors I1 I2 I3 I4	System of Equations (Segment 1)	Input Vectors I1 I2 I3 I4	System of Equations (Segment 2)
0000	1 α1 + 1 α2 + 10 α5 = 15.3	0000	1 α3 + 1 α4 + 10 α5 = 19.4
0100	4 α1 + 3 α2 + 10 α5 = 21	0001	4 α3 + 3 α4 + 10 α5 = 25.35
1000	3 α1 + 4 α2 + 10 α5 = 21.1	0010	3 α3 + 4 α4 + 10 α5 = 25.45
1100	10 α1 + 10 α2 + 3 α5 = 26.9	0011	10 α3 + 10 α4 + 4 α5 = 30.8
Results	α1 = 1.1; α2 = 1.2; α5 = 1.3	Results	α3 = 1.15; α4 = 1.25; α5 = 1.7

Case 3: HT-present vs. HT-present

Input Vectors I1 I2 I3 I4	System of Equations (Segment 1)	Input Vectors I1 I2 I3 I4	System of Equations (Segment 2)
0000	1 α1 + 1 α2 + 10 α5 = 18.8	0000	1 α3 + 1 α4 + 10 α5 = 19.4
0100	4 α1 + 3 α2 + 10 α5 = 24.5	0001	4 α3 + 3 α4 + 10 α5 = 25.35
1000	3 α1 + 4 α2 + 10 α5 = 24.6	0010	3 α3 + 4 α4 + 10 α5 = 25.45
1100	10 α1 + 10 α2 + 3 α5 = 28.3	0011	10 α3 + 10 α4 + 4 α5 = 30.8
Results	α1 = 1.12; α2 = 1.21; α5 = 1.64	Results	α3 = 1.15; α4 = 1.25; α5 = 1.7

(c) Formulation of systems of linear equations for HT detection

Fig. 1. Example of the segmentation-based HT detection approach: (a) shows that a circuit with five gates is segmented into two segments, and gate $X5$ is the overlapping gate of the two segments; (b) shows the nominal leakage power values for all the gates in the circuit; and (c) demonstrates the formulation of systems of linear equations and their solutions in three cases regarding whether a HT is present in each segment. The discrepancy of the solutions of overlapping gates ($X5$) in the two segments is an indicator of whether any HT exists or not.

as shown in Fig. 2, we conclude that the HTs are present in Segment 2 (i.e., gates X3 and X4).

Pseudocode 1 describes the detailed procedure of the consistency-based HT diagnosis. In each round of the diagnosis, we first characterize three segments with at least one overlapping gate. Then, we compare the scaling factor values of the overlapping gate obtained from the three segments. The one that has a large difference compared to the other two values is in the segment that is possibly HT-present. In the case where all three scaling factor values have large difference compared to the others, we conclude that multiple HTs are embedded in at least two segments and find more segments that cover the overlapping gates to further diagnose the HTs.

Furthermore, within one segment where we have confirmed that a HT exists, we employ a variable elimination technique to determine the location of the HT at the gate level. Our intuition is that there must be a gate that drives the HT in the segment, and more importantly, the switching pattern of the HT gate is correlated with the normal gate that drives it. Therefore, they often have linearly dependent coefficients in the system of linear equations. In this case, if we conduct linear transformation and eliminate the driving gate, the HT gate will be eliminated as well due to the linear dependency in the coefficients. On the other hand, if we conduct this variable

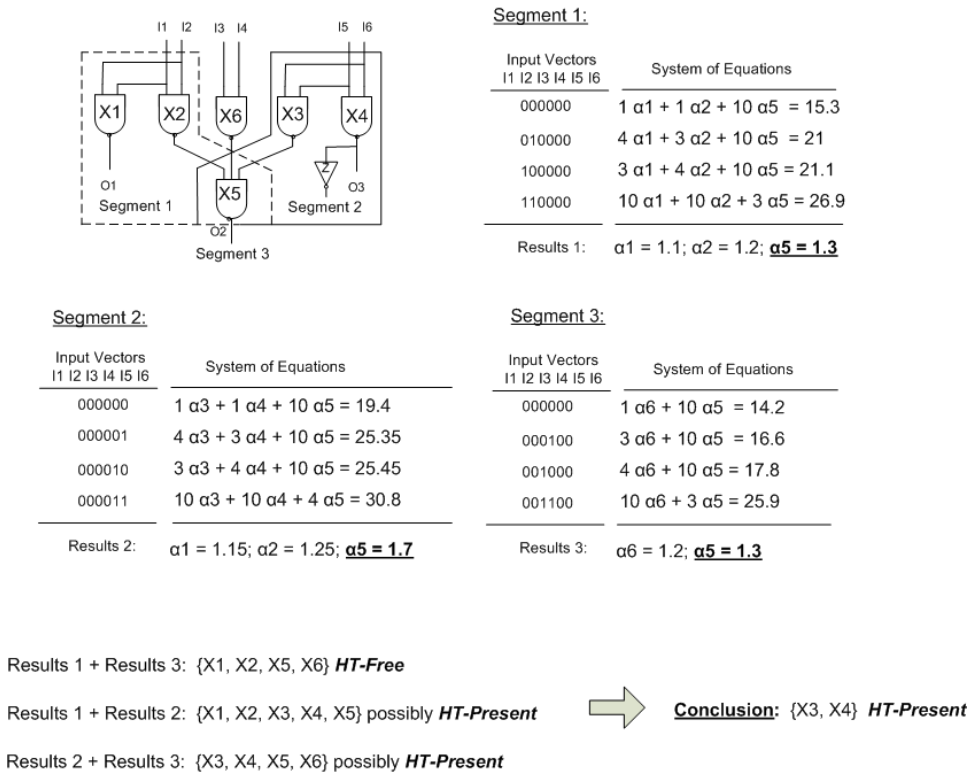


Fig. 2. Example of consistency-based HT diagnosis. We demonstrate the gate characterization in three segments with overlapping gates. The consistency in Segment 1 and Segment 3 exposes the possible HTs in Segment 2.

Pseudocode 1 Consistency-based HT diagnosis.

Input: Target circuit with embedded HTs;
Ouput: Segment set Seg, which contains all the segments that are HT-present;

```
Detect the existence of HTs;
Search for S, the three-segment set that covers all the gates in the circuit;
for each S_i in S do
    for j = 1 → 3 do
        Characterize Segment S_ij and obtain scaling factor α_j for the overlapping gate;
    end for
    d_1 = min{|α_1 − α_2|, |α_1 − α_3|};
    d_2 = min{|α_2 − α_1|, |α_2 − α_3|};
    d_3 = min{|α_3 − α_1|, |α_3 − α_2|};
    h = argmax{d_1, d_2, d_3};
    Insert S_ih into Seg;
end for
return Seg;
```

elimination procedure for each individual gate in the segment, and evaluate whether the segment is HT-free after each round of elimination, we are able to conclude which gate is driving the HT.

VI. Self-consistency Analysis via Optimal Subsegments Creation

A. Motivation

From the discussions in Sections IV and V, we note that the major source of cost in HT detection and diagnosis is the leakage power measurements required in each of the overlapping segments. Despite of the non-instrumentation to the target circuit and the high accuracy, the power measuring devices or techniques [19][18][14] would introduce additional delay or hardware cost to the IC operations.

Therefore, our goal in achieving a highly efficient HT detection and diagnosis scheme is to minimize the number of power measurements that are required to obtain accurate detection and diagnosis conclusions. In order to achieve this goal, we regard the creation of multiple overlapping segments as the key issue in both HT detection and diagnosis processes, since the strategy of segmentation directly impacts the number of measurements that are required to solve the systems of linear equations. In the next subsections, we formulate the problem of minimizing the power measuring cost and propose our solution that delivers a highly efficient segmentation strategy.

B. Problem Formulation

The problem of finding and creating multiple overlapping segments that minimize the number of required measurements can be formulated as the following:

Given a netlist of circuit C, find a set of segments Seg such that Seg covers all gates in C and that the total number of

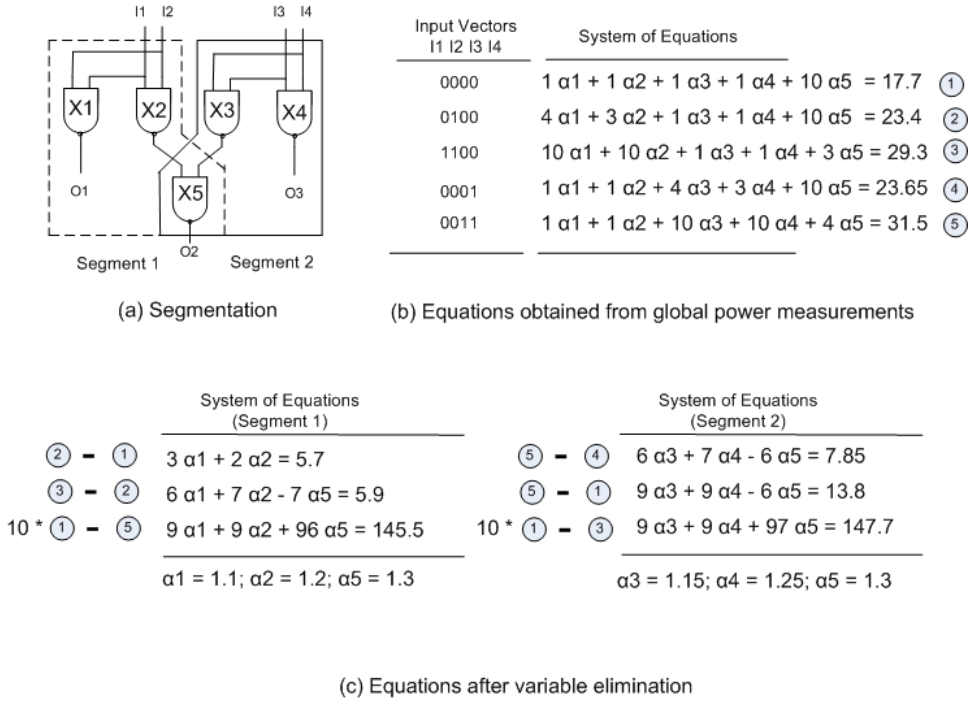


Fig. 3. Example of variable elimination that reduces the number of measurements from 8 to 5.

power measurements is minimized in order to solve the system of linear equations for each segment s in Seg. In particular, the total number of required equations (measurements) can be formulated as the following:

$$N = \sum_{s \in Seg} (q_s - \sum_{i=1}^{q_s} r_i) \tag{3}$$

where q_s is the number of measurements that are required to solve the system of linear equations for segment s and, therefore, $q_s \geq n_s$ (n_s denotes the number of gates in segment s); r_i is the number of times measurement i is reused by other segments except segment s.

From Equation (3) we conclude that we cannot reduce N by simply reducing the number of segments or the number of gates in each segment (i.e., n_s), because all of the gates in the circuit have to be covered by the segment set Seg. Thus, our idea for reducing N is to reuse the global power measurements in various segments, i.e., increasing r_s to the greatest extent. Furthermore, we note that the best way to reuse the measurements is by creating a set of sub-segments from the same larger segment. Consequently, the power measurements of the larger segments can be used in each of the sub-segments since the larger segment is a superset of the corresponding sub-segments.

C. Variable Elimination

In the extreme (ideal) case of sub-segment creation, we only measure a small set of leakage power values of the entire circuit, and generate all the sub-segments that are required for HT checking solely from this set of measurements. In this way, the power measurements are reused in the greatest extent. Also, it reduces the complexity of conducting power measurements, since only global measurements are required.

The remaining issue in reusing the power measurements is how we can accommodate the measurements of the larger segment to those of the sub-segments. We address this issue by leveraging a variable elimination technique in the system of linear equations. In particular, we only conduct one set of leakage power measurements toward the entire circuit. Then, we apply linear transformation to the obtained equations so that a specific set of variables can be eliminated from the equations, leaving only the rest of the variables appearing in the equation. Note that this transformation is equivalent to the process of creating a sub-segment from the global segment, which does not require additional leakage power measurements. In this way, the global power measurements can be reused in various sub-segments and thus that the total number of measurements is reduced.

Fig. 3 shows a motivating example regarding the variable elimination technique to create sub-segments from a global segment. In order to compare our technique with the initial non-optimized approach, we use the same example as Fig. 1 in terms of HT detection. As shown in the example, we start with a system of equations that include all the gates in the circuit and that are obtained from a fixed set of global power measurements of the entire circuit. Then, we extract two separate sets of equations from the global measurements by conducting linear transformations and eliminating the unneeded variables. The two smaller systems of equations cover the gates in the two sub-segments, which are the same as the segments in Fig. 1. However, the advantage in the linear transformation-based segmentation method is that it requires much less power measurements, e.g., 60% less comparing Fig. 3 to Fig. 1, in

order to characterize all the gates in the segments.

Furthermore, in large ICs where the partitioned segment is still large and beyond the processing ability of GLC, we employ a variable grouping technique, in which we simultaneously consider two or more variables (gates) during both HT detection and diagnosis. In particular, we can group two or more variables into a single new variable, in the case that these variables have in all equations exactly the same coefficients. This situation is rather common and its effectiveness can be further enhanced by intentional selection of a subset of equations that satisfy this requirement. From the theoretical point of view, this technique implies that one does not have to characterize each individual gate for accurate HT detection, which ensures the scalability of the approach.

D. Gate Cover Problem

Another important issue in HT detection and diagnosis is that it must cover all the gates/locations in the circuit in terms of searching hardware Trojans. Therefore, the selected sub-segments must cover all the gates in the circuit. Meanwhile, in order to reduce the computational complexity in HT detection and diagnosis, we aim to minimize the number of sub-segments that we select for consistency checking, under the condition that the sizes of the segments are well controlled so that the resulting systems of linear equations are solvable using common linear programming solvers. Taking into consideration the above requirements and goals, we formulate the segment selection problem as a set cover problem:

Segment Selection Problem. Given (1) a netlist of circuit C that contains a set of gates $G = \{g_1, g_2, ..., g_m\}$ and (2) k sub-segments obtained from the variable elimination process, identify the smallest number of sub-segments whose union contains all gates (i.e., $g_1, g_2, ..., g_m$) in G.

We solve the set cover problem using integer linear programming and the approximation algorithm discussed in [23], which provides us with the smallest number of sub-segments to minimize the computation complexity.

VII. Simulation Results

A. Consistency-based HT detection

We evaluate the consistency and self consistency-based HT detection methods on a set of ISCAS and ITC benchmarks. We generate the IC instances following the process variation models proposed by Asenov et al. [3] and Cline et al. [8]. In particular, we consider that the threshold voltage follows a Gaussian distribution (e.g., mean is 0.25V, and standard deviation is 0.01V). Also, we assume that the effective channel length follows the quad-tree model [8] that reflects the spatial correlation. For each benchmark, we simulate two cases where HTs are present (i.e., HT-present) and there are no HTs in the circuit (i.e., HT-free). The threat model we consider is the additional gate attack [24], where the attacker embeds one or more small sized gate (e.g. an inverter) into the circuit. The metric we use for identifying HTs is the inconsistency value, i.e., average discrepancy (d_{avg}) of the scaling factors that is defined in Section IV. We select pairs of segments that have overlapping gates and can cover all the gates in the circuit, conduct GLC of each of the segment, and calculate the d_{avg} value over all pairs.

Table I shows the inconsistency values obtained from the consistency-based HT detection. We observe that there are large gaps (more than 15X) in terms of d_{avg} between the HT-free case and the HT-present case. This enables us to draw a decision line between the d_{avg} values in the two cases and use it to determine whether HTs exist or not. In this case, the cost is relatively high in terms of the required number of measurements, since new separate power measurements are required for each segment being analyzed.

We further evaluate the self-consistency based approach on the same set of measurements, as shown in Table I. We note that the number of measurements is greatly reduced because of the reuse of the same set of global power measurements and the creation of subsegments. Meanwhile, we obtain a set of inconsistency values comparable to Table I, which indicates that the effectiveness of the approach is not impacted by the reduced measurements. Furthermore, in Table II, we show the savings in the measurement cost obtained from our divide-and-conquer (i.e., variable grouping) technique. The total number of measurements can be further reduced by 50% to 80% by grouping a subset of the gates, which ensures the scalability of the approach.

In figure 5 we summarize and compare the measurement costs of the two methods we have developed in this paper, including the segment selection using the set cover algorithm and the variable grouping technique. We observe that the ratio of the number of measurements and gates scales with the sizes of the circuit using our variable elimination technique.

Finally, to be more specific, we plot in Fig. 4 the distribution of the inconsistency values over all pairs of segments in four of the benchmarks. We observe that there are no overlaps between the inconsistency values in the two cases, which ensures zero false positives and false negatives in the self-consistency based HT detection. In particular, we can determine a decision line (e.g., at the 50% boundary of the average gap) based on the results obtained from a small set of training benchmarks and use it to diagnose an arbitrary circuit after manufacturing.

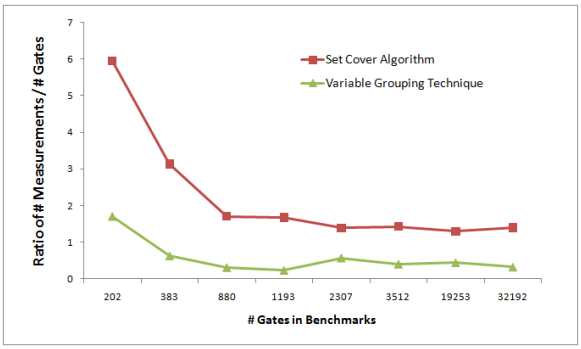


Fig. 5. Comparison of number of measurements.

B. Consistency-based HT Diagnosis

We evaluate the consistency-based HT diagnosis approach on a set of ISCAS benchmarks, as shown in Fig. 6. For each

TABLE I

HT DETECTION RESULTS USING CONSISTENCY AND SELF CONSISTENCY-BASED GLC: THE VALUES IN THE "HT-FREE" AND "HT-PRESENT" COLUMNS REPRESENT THE AVERAGE DISCREPANCY OF THE OVERLAPPING GATES IN TERMS OF THEIR SCALING FACTORS.

Benchmark	# Gates	# Measurements (Consistency)	# Measurements (Self-Consistency)	HT-Free (Consistency)	HT-Present (Consistency)	HT-Free (Self-Consistency)	HT-Present (Self-Consistency)
C499	202	7,200	1200	6.2E-03	2.0E-01	4.1E-03	3.0E-01
C880	383	14,800	1200	5.8E-03	7.3E-02	5.7E-02	4.0E-01
C1908	880	7,500	1500	2.1E-03	2.3E-01	3.8E-03	2.3E-01
C2670	1193	16,700	2000	1.4E-03	1.3E-01	8.1E-04	4.3E-01
C5315	2307	12,800	3200	6.2E-03	1.2E-01	5.9E-02	8.5E-01
C7552	3512	16,600	5000	4.6E-03	9.8E-02	1.2E-02	5.8E-01
S38584	19253	160,000	25000	4.7E-03	2.4E-01	2.3E-03	3.6E-01
b17	32192	280,000	45000	5.9E-03	3.8E-01	1.6E-03	6.4E-01

TABLE II

SAVINGS FROM THE VARIABLE GROUPING TECHNIQUES. COLUMNS 2 AND 3 SHOW THE COMPOSITION OF THE DIVIDED GROUPS, WITH MULTIPLE NODES AND ONE NODE, RESPECTIVELY. COLUMN 4 SUMMARIZES THE TOTAL NUMBERS OF GROUPS. COLUMN 5 SHOWS THE SAVINGS IN THE NUMBER OF MEASUREMENTS, COMPARED TO TABLE II, USING THE VARIABLE GROUPING TECHNIQUE.

Benchmark	# Gates	# multi-node groups	# one-node groups	total # groups	% Savings
C499	202	10	48	58	71.29
C880	383	23	54	77	79.90
C1908	880	46	114	160	81.82
C2670	1193	32	135	167	86.00
C5315	2307	185	767	952	58.73
C7552	3512	170	826	996	71.64
s38584	19253	1548	5022	6570	65.88
b17	32192	1666	5749	7415	76.97

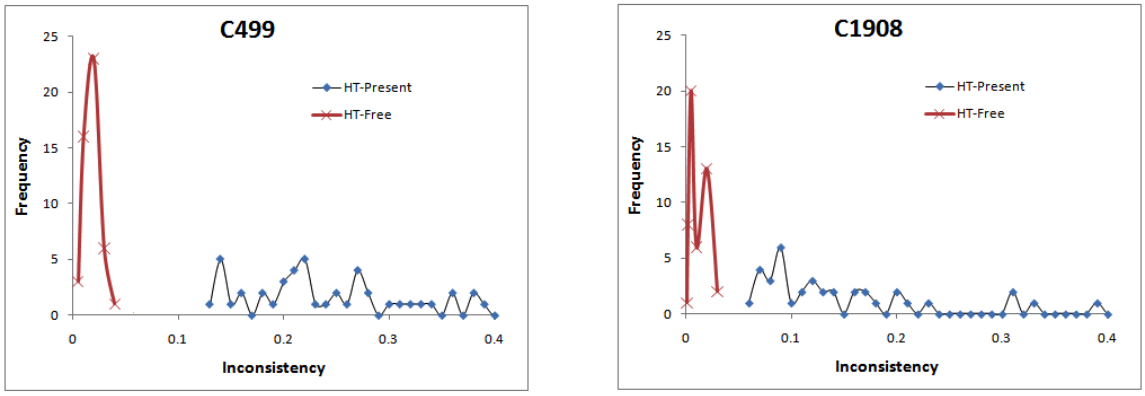


Fig. 4. Distribution of the inconsistency values in the HT-present and HT-free cases.

benchmark, we show the scaling factors of the overlapping gates in three segments, where a single HT is embedded in one of the segments (e.g., Segment 3). We observe from the results that the two values of scaling factors from the HT-free segments are consistent with each other, and that in the HT-present segment is either a very high value or a very low value apart from the two consistent values. These results enable us to conclude that the HT is embedded in Segment 3 with zero false positives and zero false negatives.

Furthermore, we evaluate our variable elimination technique for HT diagnosis using the same set of benchmarks. Table III shows the diagnosis results, which includes the objective function value of the linear program in the two cases: (1) where we eliminated the exact gate (variable) that drives the HT (shown in the second column); and (2) where we did

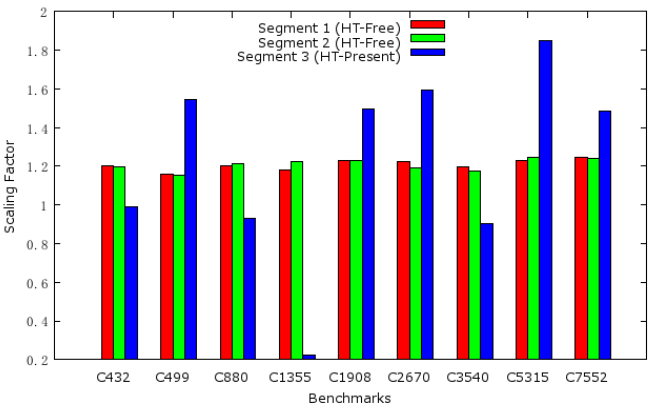


Fig. 6. Simulation results for consistency-based HT diagnosis.

TABLE III
HT DIAGNOSIS RESULTS USING THE VARIABLE ELIMINATION TECHNIQUE.

Benchmark	With Var. Elimination	No Var. Elimination
C499	5.57 E-6	118.73
C880	2.68E-06	93.91
C2670	1.28E-04	17.06
C6288	2.39E-06	137.44
C5315	2.67E-06	140.62
C7552	7.43E-06	21.98
s38584	5.12E-06	48.88
b17	0.00E+00	89.42

not eliminate the gate that drives the HT (shown in the third column). We note that the objective function values have a large gap between these two cases, which provides us with a clear and distinguishable metric to determine the location of the HT in the circuit.

VIII. CONCLUSION

We developed a complete solution of HT detection and diagnosis. We employed segmentation and consistency-based gate characterization to determine the existence of HTs and their locations. Next, we select input vectors to age the HTs embedded in the circuit and disable their functionalities. Furthermore, we developed the variable elimination and variable grouping techniques to reduce the number of measurements and ensure the scalability of our approach toward large designs. Our simulation results on a set of ISCAS and ITC benchmarks indicate that the proposed approach is scalable and capable of detecting and diagnosing HTs accurately.

REFERENCES

[1] M. Agarwal, B.C. Paul, M. Zhang, and S. Mitra. Circuit failure prediction and its application to transistor aging. In *VLSI Test Symposium (VTS)*, pages 277–286, 2007.

[2] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar. Trojan detection using IC fingerprinting. In *IEEE Symposium on Security and Privacy (SP)*, pages 296–310, 2007.

[3] A. Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic" simulation study. *IEEE Transactions on Electron Devices*, 45(12):2505–2513, 1998.

[4] M. Banga and M.S. Hsiao. A region based approach for the identification of hardware Trojans. In *IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, pages 40–47, 2008.

[5] M. Banga and M.S. Hsiao. VITAMIN: Voltage inversion technique to ascertain malicious insertions in ICs. In *HOST*, pages 104–107, 2009.

[6] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. Parameter variations and impact on circuits and microarchitecture. In *Design Automation Conference (DAC)*, pages 338–342, 2003.

[7] S. Chakravarthi, A. Krishnan, V. Reddy, C.F. Machala, and S. Krishnan. A comprehensive framework for predictive modeling of negative bias temperature instability. In *International Reliability Physics Symposium (IRPS)*, pages 273–282, 2004.

[8] B. Cline, K. Chopra, D. Blaauw, and Y. Cao. Analysis and modeling of CD variation for statistical static timing. In *International Conference on Computer-Aided Design (ICCAD)*, pages 60–66, 2006.

[9] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos. Modeling within-die spatial correlation effects for process-design co-optimization. In *International Symposium on Quality of Electronic Design (ISQED)*, pages 516–521, 2005.

[10] Y. Jin and Y. Makris. Hardware Trojan detection using path delay fingerprint. In *IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, pages 51–57, 2008.

[11] F. Koushanfar and A. Mirhoseini. A unified framework for multimodal submodular integrated circuits Trojan detection. *IEEE Transactions on Information Forensics and Security*, 6(1):162–174, 2011.

[12] J. Li and J. Lach. At-speed delay characterization for IC authentication and Trojan horse detection. In *IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, pages 8–14, 2008.

[13] M. Nelson, A. Nahapetian, F. Koushanfar, and M. Potkonjak. Svd-based ghost circuitry detection. In *Information Hiding (IH)*, pages 221–234, 2009.

[14] NI PXI-4130 Power SMU: http://sine.ni.com/nips/cds/view/p/lang/en/nid/204239.

[15] J.H.L. Pang, D.Y.R. Chong, and T.H. Low. Thermal cycling analysis of flip-chip solder joint reliability. *IEEE Transactions on Components and Packaging Technologies*, 24(4):705 –712, 2001.

[16] M. Potkonjak, A. Nahapetian, M. Nelson, and T. Massey. Hardware Trojan horse detection using gate-level characterization. In *Design Automation Conference (DAC)*, pages 688 –693, july 2009.

[17] R.M. Rad, Xiaoxiao Wang, M. Tehranipoor, and J. Plusquellic. Power supply signal calibration techniques for improving detection resolution to hardware trojans. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 632 –639, 2008.

[18] R. Rajsuman. Iddq testing for CMOS VLSI. *Proceedings of the IEEE*, 88(4):544 –568, 2000.

[19] S. Sabade and D. Walker. Iddx-based test methods: A survey. *ACM Trans. Des. Autom. Electron. Syst.*, 9(2):159–198, 2004.

[20] H. Salmani, M. Tehranipoor, and J. Plusquellic. New design strategy for improving hardware Trojan detection and reducing Trojan activation time. In *IEEE International Workshop on Hardware-Oriented Security and Trust (HOST)*, pages 66–73, 2009.

[21] J.H. Stathis. Physical and predictive models of ultra thin oxide reliability in CMOS devices and circuits. In *IEEE International Reliability Physics Symposium (IRPS)*, pages 132 –149, 2001.

[22] M. Tehranipoor and F. Koushanfar. A survey of hardware Trojan taxonomy and detection. *IEEE Design Test of Computers*, 27(1):10–25, 2010.

[23] V. Vazirani. *Approximation Algorithms*. Springer, 2001.

[24] S. Wei, K. Li, F. Koushanfar, and M. Potkonjak. Hardware trojan horse benchmark via optimal creation and placement of malicious circuitry. In *Design Automation Conference (DAC)*, pages 90–95, 2012.

[25] S. Wei, S. Meguerdichian, and M. Potkonjak. Gate-level characterization: Foundations and hardware security applications. In *Design Automation Conference (DAC)*, pages 222–227, 2010.

[26] S. Wei, S. Meguerdichian, and M. Potkonjak. Malicious circuitry detection using thermal conditioning. *IEEE Transactions on Information Forensics and Security*, 6(3):1136–1145, 2011.

[27] S. Wei and M. Potkonjak. Scalable consistency-based hardware Trojan detection and diagnosis. In *International Conference on Network and System Security (NSS)*, pages 176–183, 2011.

[28] S. Wei and M. Potkonjak. Scalable hardware Trojan diagnosis. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(6):1049–1057, 2012.

[29] F. Wolff, C. Papachristou, S. Bhunia, and R.S. Chakraborty. Towards Trojan-free trusted ICs: Problem analysis and detection scheme. In *Design, Automation and Test in Europe (DATE)*, pages 1362–1365, 2008.

[30] W. Zhang, X. Li, and R. Rutenbar. Bayesian virtual probe: minimizing variation characterization cost for nanoscale ic technologies via bayesian inference. In *Design Automation Conference (DAC)*, pages 262–267, 2010.

Sheng Wei is a Ph.D. candidate in Computer Science at the University of California, Los Angeles. His research interests include computer-aided design of VLSI circuits, hardware security, and wireless networking.

Miodrag Potkonjak received his Ph.D. degree in Electrical Engineering and Computer Science from University of California, Berkeley in 1991. He is a professor with Computer Science Department at UCLA. He created first watermarking, fingerprinting, and metering techniques for integrated circuits as well as first remote trusted sensing and trusted synthesis approaches, compilation using untrusted tools, and public physical unclonable functions.