Scalable Segmentation-based Malicious Circuitry Detection and Diagnosis

Sheng Wei Miodrag Potkonjak
Computer Science Department
University of California, Los Angeles (UCLA)
Los Angeles, CA 90095
Email: {shengwei, miodrag}@cs.ucla.edu

Abstract—Hardware Trojans (HTs) pose a significant threat to the modern and pending integrated circuit (IC). Several approaches have been proposed to detect HTs, but they are either incapable of detecting HTs under the presence of process variation (PV) or unable to handle very large circuits in the modern IC industry. We develop a scalable HT detection and diagnosis scheme by using segmentation techniques and gate level characterization (GLC). In order to address the scalability issue, we propose a segmentation method which divides the large circuit into small sub-circuits by using input vector control. We propose a segment selection model in terms of properties of segments and their effects on GLC accuracy. The model parameters are calibrated by sampled data from the GLC process. Based on the selected segments we are able to detect and diagnose HTs correctly by tracing gate level leakage power. We evaluate our approach on several ISCAS85/ISCAS89/ITC99 benchmarks. The simulation results show that our approach is capable of detecting and diagnosing HTs accurately on large circuits.

I. INTRODUCTION

Hardware Trojans (HTs) [1][2] are malicious hardware components imposed by adversaries in order to make the IC design malfunction or leak confidential information. Recently, HT attacks have drawn a great deal of attention in the hardware security community, as IC manufacturing continues to be outsourced.

The goal of HT diagnosis is to detect and locate the malicious HTs on the target circuit, so that they can be either masked or removed from the hardware. Among all the HT detection and diagnosis approaches, side channel analysis has been widely adopted because of its low instrumentation to the IC product [3][4][5][6]. Side channel analysis detects HTs by observing the variations in IC manifestation properties such as delay and power. Since the presence of HTs would make at least one of the properties (e.g. leakage power) vary from its nominal specification, it is detectable by monitoring the delay/power profile of the circuit. However, due to the inability to access each gate in the circuit, and the attempt to reduce the instrumentation to the IC product, the side channels being used are properties of the entire circuit. In this case, the adversaries may put ultra small or low power HTs so that they can be hidden under the measurement errors in the HT detection process. Moreover, process variation (PV) [7] widely exists due to the nature of IC manufacturing process. Although PV can be used to facilitate some IC security applications [8], it complicates the HT detection scheme,

since HTs can be easily hidden under the unrecognized PV.

Some recent research work [9][10][11][12] proposed gate level characterization (GLC) approaches in order to capture HTs by tracing the side channels at gate level under the presence of PV. The existing approaches formulate a system of equations in terms of the gate level properties and the overall delay/power measurements, and HTs are detected if some variations on gate level properties are observed from the solution of the equations. There are two unsolved issues which may cause the failure of HT diagnosis. First, in the system of equations, it usually happens that a large group of gates (variables) are correlated with each other in the sense that they have collinear coefficients due to the same gate type and same switching activities. Second, with the fast development of deep submicron technologies, the transistor density keeps increasing and it is common that a single IC chip has millions of gates. The huge number of gates result in a large size of equations which are impossible to solve. Both of the issues make the system of equations unsolvable and thus HTs cannot be detected correctly on large modern ICs.

We propose a new approach to HT detection and diagnosis by using segmentation techniques. Our key idea is to divide the large circuit into small sub-circuits by using input vector control, so that the segmented circuits have good properties (e.g. few correlations and small number of gates) for obtaining accurate HT diagnosis results. We propose a segment selection model in terms of properties of segments and their effects on GLC accuracy. The model parameters are calibrated by sampled data from the GLC process. Then, we apply the segmentation approach into the thermal conditioning based GLC, which solves the collinear correlation problem by changing the thermal property of the circuit.

II. PRELIMINARIES

A. Gate Level Characterization

Gate level characterization [10][12] is the process of characterizing each gate of an IC in terms of its physical properties (e.g. scaling factors) or manifestation properties (e.g. propagation delay or leakage/switching power). The objective of GLC is to capture the impact of PV so that the gate level characteristics can be well predicted and used in many applications.

In the existing works of GLC [10][12][13], the processes are based on the gate-level power models [14] and formulate the impact of PV as a scaling factor which indicates the increased/decreased power value due to PV. Then, a system of linear equations can be achieved in terms of the full-chip power value and gate level power values:

$$\tilde{p}_j = e_{sj} + e_{rj} + \sum_{\forall \text{gate } i=1, \dots, n} K_{ij} s_i \quad (1)$$

where $\tilde{p}_j$ is the full-chip leakage power at input state j ; s_i is the PV scaling factor of gate i that the GLC process is trying to achieve; K_{ij} is the nominal leakage power for gate i at input state j ; and e_{sj} and e_{rj} are systematic and random measurement errors, respectively. By solving the system of equations with an objective function of minimizing the measurement errors, the gate level PV scaling factor can be characterized.

B. Correlations and Thermal Conditioning

One of the major issues in GLC is that a large number of gates are correlated with each other due to the collinearity in their nominal power values [12]. This happens very often in modern IC designs because it is usually the case that some specific types of gates are used very often due to their small sizes or low power/delay. The correlation problem can be solved by thermal conditioning, where thermal control is applied on the correlated gates and their nominal values are thus varied to break the collinear correlation. The idea is based on the observation that gate level leakage power increases exponentially with temperature. We have developed a thermal conditioning based GLC approach in [12]. However, it did not consider the scalability issue which limits the application of the proposed approach in large modern IC designs.

III. SEGMENT SELECTION MODEL

We solve the scalability issue in GLC and HT diagnosis using segmentation techniques. Our main idea is to segment the large circuit into multiple small pieces and solve each piece of the circuit for gate level characteristics.

A. Problem Definition

Our goal in segment selection is to find good segments that are likely to obtain accurate GLC results. Our definition of segments is based on the controllability of a set of primary inputs over a set of gates.

Definition 1. *Controllability.* A set of primary inputs have control over a set of gates if the output states of the gates can be switched by varying the set of primary inputs and freezing any other primary inputs.

Definition 2. *Segment.* A segment is a set of gates that are controlled by a set of primary inputs.

We measure the GLC accuracy by relative characterization error, namely the mean error of the characterized gate level scaling factor compared to the real value:

$$E_{avg} = \frac{1}{n_g} \sum_{i=1 \dots n_g} |s_{calc_i} - s_{real_i}| / s_{real_i} \quad (2)$$

where E_{avg} represents the GLC accuracy (the smaller the better); n_g is the number of gates in the circuit; and s_{calc_i} and s_{real_i} are the calculated scaling factor of gate i and its real value, respectively.

In order to evaluate a segment in terms of its resulting GLC accuracy, we define the following properties of a segment by considering the key factors that may affect the GLC process:

- **Controllability Ratio.** We define the controllability ratio property as the ratio between the number of different equations and the number of gates that are available in a segment. The number of different equations of a segment can be calculated as the number of input patterns that can be obtained by varying the set of primary inputs. The controllability ratio is an indicator of the controllability in the segment (n_{input} is the number of primary inputs, and n_g is the number of gates):

$$P_{ctrl} = \frac{2^{n_{input}}}{n_g} \quad (3)$$

- **Correlation Ratio.** We define the correlation ratio property as the ratio between the number of correlated gates and the total number of gates in a segment. The correlation ratio indicates the degree of correlation in the segment (n_{corr} is the number of correlated gates):

$$P_{corr} = \frac{n_{corr}}{n_g} \quad (4)$$

- **Gini Coefficient.** We use the Gini coefficient [15] as a property which indicates the level of imbalance of the gate level scaling factors in a segment. The Gini coefficient can be calculated as the relative mean difference of a set of unordered data. In our case it is the difference between every possible pair of the scaling factors in the segment (μ is the average value of the scaling factor over all gates):

$$P_{Gini} = \frac{\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} |s_{real_i} - s_{real_j}|}{2n_g^2 \mu} \quad (5)$$

- **Number of Gates.** We also consider the number of gates (n_g) in a segment as one of the properties. The intuition is that the size of the segment would impact the GLC accuracy because of the limitation of LP solver.

B. Model Build-up

Our goal in segment selection modeling is to build up a prediction model for GLC accuracy in terms of the properties of segments. Then we can use the model to predict the resulting GLC accuracy for segment selection.

The basic flow of our model build-up process is shown in Fig. 1. We first select some segments from the target circuit by freezing a certain set of the primary inputs and varying the others. We then conduct GLC on each selected segment

and observe the GLC accuracy under the properties of each segment. After collecting the value of properties and GLC accuracy, we build up a regression model to predict the GLC accuracy for the other unknown segments in the circuit. The model parameters are calibrated by solving a system of equations using linear programming (LP). The model calibration is a repeated process in which we use different sets of segments and keep calibrating the parameters. Then, we post-process the calibration results using statistical methods (e.g. maximum likelihood estimation) to find out the parameter values for the model.

Fig. 1. Flow of Segment Selection Model Build-up

C. Segment Selection Algorithm

With the prediction model obtained in the training process, we are able to predict the GLC accuracy according to the properties of a segment. We use the predicted GLC accuracy as a metric for segment selection. Our goal is to obtain a minimum GLC accuracy by segment selection. Our segment selection algorithm is a heuristic algorithm which keeps combining one-input segments in order to improve GLC accuracy. The combining operation stops and the algorithm starts with a new segment until there are no improvements in GLC accuracy according to the prediction model. The segment selection algorithm is shown in Algorithm 1.

IV. SCALABLE HT DETECTION AND DIAGNOSIS

Our goal in HT detection and diagnosis is to detect the existence of HTs (HT detection), and if there are any, to further identify their types, input pins, and scaling factors (HT diagnosis). We use thermally conditioned GLC of the power profiles to identify HTs on the circuit. The main challenge we face is to make the small variation caused by HTs observable in large circuits. We solve the problem by using the aforementioned segmentation technique.

Our flow of HT detection and diagnosis is the following. We first do preprocessing in which we segment the large circuit into a few segments using Algorithm 1. Then we apply thermal control on each segment and obtain leakage power measurements while varying the primary input signals. After obtaining the system of equations shown in (1), we manipulate the constraints and add one HT variable into all the equations

input : Primary Input Set $PI = \{PI_i | 1 \leq i \leq n_i\}$
 Controlled Gate Set
 $G(PI_{sub}) = \{G_{sub} | G_{sub} \text{ is controlled by } PI_{sub}\}$
output: Selected Segment Set Seg

```

1 while not all gates covered in the circuit do
2   Select starting  $PI_i$  close to the uncovered gates;
3    $PI_{sub} = \{PI_i\}$ ; // selected primary input set
4   repeat
5      $E_{cur} = E(PI_{sub})$ ; // current GLC accuracy
6      $Seg_{cur} = G(PI_{sub})$ ; // current segment
7     Add  $PI_j$  to  $PI_{sub}$ , where  $PI_j \notin PI_{sub}$ ;
8   until  $E_{cur} < E(PI_{sub})$ ;
9   Add  $Seg_{cur}$  to  $Seg$ ;
  
```

Algorithm 1: Segment Selection Algorithm

indicating the variation caused by HTs. By solving the system of equations the same way it is done for GLC, we obtain the characterization results for the HT variable and make conclusions about HTs based on its value.

For HT detection, we first assume some HT exists in the segment and use a single HT variable to represent it in the LP formulation [12]:

$$var_{ht} + K \cdot s = \tilde{p} + e \quad (6)$$

where var_{ht} is the HT variable we add. After solving the modified LP, if var_{ht} has a value close to 0, it indicates that no HTs are detected during the characterization process. Otherwise, we assume that it is the existence of some HTs that has made var_{ht} a large value.

For HT diagnosis, our approach is “guess and verify”, in which we first identify the type and input pins of the HT, and then we verify our identifications by doing additional constraint manipulation based on (6), where we add one more HT item according to our identifications:

$$var_{ht} + k_{ht} \cdot s_{ht} + K \cdot s = \tilde{p} + e \quad (7)$$

where $k_{ht} \cdot s_{ht}$ is the new item we add for HT diagnosis. k_{ht} is the nominal leakage power of the HT gate, which is dependent on the type and input signals of the HT. s_{ht} is the variable representing the PV scaling factor of the HT gate. After solving the LP, if the current identification is correct, we will get var_{ht} close to 0, and s_{ht} to be the estimated scaling factor for the HT. Otherwise var_{ht} is a large value which represents the discrepancy caused by an incorrect identification.

V. SIMULATION RESULTS

We evaluate our HT detection and diagnosis approach on ISCAS85, ISCAS89 and ITC99 benchmarks. For HT detection, we simulate two cases for each benchmark: 1) where there are no HTs embedded; and 2) where some HTs exist on the target circuit. We repeat the leakage power measurement many times so that we can have a large enough sample space

TABLE I
HT DETECTION AND DIAGNOSIS ON ISCAS AND ITC BENCHMARKS

Benchmark	Gates	HT Detection		HT Diagnosis		
		HTHs Exist	No HTHs	Correct Identification	Wrong Identification 1	Wrong Identification 2
C17	6	226.4~420.3	0	0	319.1	189.1
C432	160	608.3~675.4	0~11.0	0	708.9	764.7
C499	202	112.0~297.0	0~11.7	0.228	208.4	217.8
C880	383	62.9~560.7	0~7.2	0	234.1	236.7
C1355	546	229.1~678.1	0~3.0	0	508.2	448.1
C1908	603	591.9~1281.6	0~9.5	0.057	760.5	806.1
C2670	872	234.1~1092.8	0~0.092	0.031	334.5	394.7
C3540	1179	159.9~1006.6	0~0.265	0	673.8	583.2
S526	214	84.2~723.5	0~15.8	0	2602.5	2614.9
S832	292	229.5~1271.5	0	0	2490.1	2494.2
S1423	490	571.0~902.5	0~0.0014	0.0051	257.2	794.6
S5378	1004	290.2~1348.8	0	0.0023	501.1	799.3
S9234	2027	751.9~954.1	0~0.0050	0	422.8	954.1
S13207	7951	667.5~1087.7	0	0	2038.3	420.7
S15850	9772	539.1~857.4	0~0.0087	0	598.4	620.1
S35932	16065	503.2~946.0	0~0.0096	0	692.8	326.3
S38584	19253	313.0	2.1	0	385.2	369.0
b17	32326	868.7	2.3	2.45	629.8	575.2
Smallest		62.9	0	0	208.4	189.1
Largest		1348.8	15.8	2.45	2602.5	2614.9

to simulate the real measurement errors. By doing this, we expect to have a decision line which enables us to determine whether HTs exist or not. Table I shows the results in terms of the value of the HT variable we obtained for each set of measurements. We can see that for each benchmark the HT variable is always a large value (between 62.9 and 1348.8) when HTs exist, while it is a value close to 0 (between 0 and 15.8) when HTs do not exist. The results show that there is no overlapping and we observe a large enough gap between the two situations. Therefore, a decision line can be achieved to determine whether there are any HTs embedded in the circuit.

After detecting any HTs on the target circuit, we then start the HT diagnosis process to locate the HTs. In our guess and verify diagnosis, we use the HT variable as the indicator of a correct or incorrect identification. The identifications are based on the type and input pins of a HT. The results in Table I show that when the identification correctly identifies the actual properties of a HT, the HT variable is small (from 0 to 2.45), while an incorrect identification always induces a large value (from 189.1 to 2614.9).

VI. CONCLUSION

We propose a scalable HT detection and diagnosis approach based on segmentation and GLC. By freezing a subpart of the primary inputs and varying the others, a large circuit can be segmented into small pieces. We propose a segment selection model which predicts the GLC accuracy of the segments. Then, a segment selection algorithm is used for selecting good segments in terms of GLC accuracy. We then conduct GLC on each segment of the circuit and diagnose HTs by observing the variations on leakage power profiles. The simulation results on several ISCAS85/ISCAS89/ITC99 benchmarks show that HTs can be detected and diagnosed accurately on large circuits.

REFERENCES

- [1] M. Tehranipoor, F. Koushanfar. A Survey of Hardware Trojan Taxonomy and Detection. *IEEE Design and Test of Computers*, Vol. 27, No. 1, 2010, pp. 10-25.
- [2] S. King, J. Tucek, A. Cozzie, C. Grier, W. Jiang, Y. Zhou. Designing and Implementing Malicious Hardware. *Proceedings of the 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats*. Article No. 5, 2008, pp. 1-8.
- [3] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, B. Sunar. Trojan Detection Using IC Fingerprinting. *IEEE Symposium on Security and Privacy*, 2007, pp. 296-310.
- [4] R. Rad, X. Wang, M. Tehranipoor, J. Plusquellic. Power Supply Signal Calibration Techniques for Improving Detection Resolution to Hardware Trojans. *ICCAD 2008*, pp. 632-639.
- [5] F. Wolff, C. Papachristou, S. Bhunia, R. Chakraborty. Towards Trojan-Free Trusted ICs: Problem Analysis and Detection Scheme. *DATE 2008*, pp. 1362-1365.
- [6] D. Rai, J. Lach. Performance of Delay-based Trojan Detection Techniques under Parameter Variations. *HOST 2009*, pp. 58-65.
- [7] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, V. De. Parameter Variations and Impact on Circuits and Microarchitecture. *DAC 2003*, pp. 338-342.
- [8] F. Koushanfar, M. Potkonjak. CAD-based Security, Cryptography, and Digital Rights Management. *DAC 2007*, pp. 268-269.
- [9] Y. Alkabani, F. Koushanfar, N. Kiyavash, M. Potkonjak. Trusted Integrated Circuits: A Nondestructive Hidden Characteristics Extraction Approach. *Information Hiding 2008*, pp. 102-117.
- [10] M. Potkonjak, A. Nahapetian, M. Nelson, T. Massey. Hardware Trojan Horse Detection Using Gate-Level Characterization. *DAC 2009*, pp. 688-693.
- [11] M. Nelson, A. Nahapetian, F. Koushanfar, M. Potkonjak. SVD-Based Ghost Circuitry Detection. *Information Hiding*, pp. 221-234, 2009.
- [12] S. Wei, S. Meguerdichian, M. Potkonjak. Gate-level Characterization: Foundations and Hardware Security Applications. *DAC 2010*, pp. 222-227.
- [13] Y. Alkabani, T. Massey, F. Koushanfar, M. Potkonjak. Input Vector Control for Post-Silicon Leakage Current Minimization in the Presence of Manufacturing Variability. *DAC 2008*, pp. 606-609.
- [14] R. Gu, M. Elmasry. Power Dissipation Analysis and Optimization of Deep Submicron CMOS Digital Circuits. *IEEE Journal on Solid-State Circuits*. Vol. 31, No. 5, 1996, pp. 707-713.
- [15] R. Keeney, H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.