Detection and Masking of Trojan Circuits in Sequential Logic*

A. Matrosova[1], E. Mitrofanov[2], S. Ostanin[3], E. Nikolaeva[4]
Tomsk State University, Russia
{[1]mau11, [2]qvaz, [3]sergeiostanin, [4]nikolaeve-ea}@yandex.ru

Abstract

A technique of finding a set of sequential circuit nodes in which Trojan Circuits (TC) may be implanted is suggested. The technique is based on applying the precise (not heuristic) random estimations of internal node observability and controllability. Getting the estimations we at the same time derive and compactly represent all sequential circuit full states (depending on input and state variables) in which of that TC may be switched on. It means we obtain precise description of TC switch on area for the corresponding internal node v. The estimations are computed with applying a State Transition Graph (STG) description, if we suppose that TC may be inserted out of the working area (out of the specification) of the sequential circuit. Reduced Ordered Binary Decision Diagrams (ROBDDs) for the combinational part and its fragments are applied for getting the estimations by means of operations on ROBDDs. Techniques of masking TCs are proposed. Masking sub-circuits overhead is appreciated.

1. Introduction

The enhanced utilization of outsourcing services for a part of VLSI circuits (IP (Intellectual Property) cores, reprogramming modules based on FPGA and so on) to cut VLSI cost increases risk of implanting Trojan Circuits (TCs) that may demolish VLSI circuit or provide drain of confidential information [1]. TCs act in infrequent operation situations as a rule, therefore they are not detectable neither during VLSI testing nor VLSI verification. TC consists of Trojan trigger and Trojan payload. Trojan trigger is switched on when the defined combination of signals appears on inputs of TC. Trojan payload is operation unit that is switched on by trigger sub-circuit. We need to detect and mask these TCs. In contrast with heuristic estimations precise ones may be very close to zero and be used to determine the most suspicious nodes for injecting malicious circuits.

In [2], authors proposed an automated online approach with low-overhead to service in the detection of TCs. They focus on the recognition of small TC samples (five or less logic gates) that cause logic failure on activation through infrequent internal logic conditions. These conditions are determined by using heuristic estimations of controllability.

In [3] Functional Analysis for Nearly-unused Circuit Identification (FANCI) tool is suggested. FANCI marks suspicious lines in design, which have the potential to be malicious. Approximate Boolean functional analysis for detecting suspicious lines is used.

In this paper in contrast with [2], [3] detection of suspicious nodes is based on using precisely calculated random estimations of observability and controllability of a combinational part internal node. The suggested approach guarantees finding all full states (compactly represented by ROBDD) that may provide triggering the node. The estimations calculations like those in [2], [3] are based on using structural description of the combinational part. In this paper representation of the sequential circuit behavior by State Transition Graph (STG) is additionally used. The calculations are executed with operations on Reduced Ordered Binary Decision Diagrams (ROBDDs, further just BDDs). Techniques of masking TCs are proposed. The experimental results on benchmarks illustrate applicability of the suggested approach and show that overhead for masking TC may be rather small.

In Section 2 techniques of precise calculation of observability and controllability estimations for combinational part nodes of a sequential circuit are briefly described. In Section 3 the way of calculation of precise controllability estimations for combinational part nodes out of sequential circuit working area is given. In Section 4 the techniques of masking TC are proposed and experimental results are considered.

*The reported study was supported by Russian Science Foundation, research project № 14-19-00218.

978-1-5386-3299-4/17/$31.00 ©2017 IEEE

2. Precise calculation of observability and controllability estimations with using description of structural combinational part

A capability of delivering 1(0) value to an internal node will be called 1(0)-controllability, a capability of observation of changing 1(0) value of an internal node on the proper circuit output is observability. Precise calculation of random observability and controllability estimations is based on using of the corresponding BDDs [4] and operations on them. These estimations are obtained for the pair of nodes associated with the input and output of TC, respectively.

Precisely calculation of 1(0)-controllability for internal node v of combinational part C is based on using BDD $R^{cont}(1)$ ($R^{cont}(0)$) derived from the combinational circuit which output is node v, and inputs coincide with circuit C inputs [5]. $R^{cont}(0)$ is obtained from $R^{cont}(1)$ by permutation of terminal nodes.

Precisely observability calculation for internal node v of combinational part C and the proper circuit output is based on using BDD $R(C_v)$ for single-output sub-circuit C_v. The output of the sub-circuit C_v is the proper circuit C output and C_v is obtained from circuit C under the condition that internal node v is an input of sub-circuit C_v [4]. Let root node of BDD $R(C_v)$ is marked by variable v, those variable v is chosen as the first variable of the decomposition.

Let function f is implemented by BDD $R(C_v)$. Derive BDDs $R(f^{v=0})$ and $R(f^{v=1})$ from $R(C_v)$). Roots of $R(f^{v=0})$ and $R(f^{v=1})$ are children nodes of root node of BDD $R(C_v)$. Functions $f^{v=0}$ and $f^{v=1}$ are implemented by BDDs $R(f^{v=0})$ and $R(f^{v=1})$ accordingly. BDD R^{obs} represents observability for internal node v of combinational part C and the proper circuit output. R^{obs} is calculated by formula:

$$R^{obs} = R(f^{v=0})\overline{R(f^{v=1})} \vee R(f^{v=1})\overline{R(f^{v=0})}\,. \quad (1)$$

To obtain $\overline{R(f^{v=0})}$, ($\overline{R(f^{v=1})}$) from $R(f^{v=0})$, ($R(f^{v=1})$) it suffices to swap of terminal nodes of the corresponding BDDs. Note that operations over BDDs have a polynomial complexity.

Calculating precise observability and controllability estimations we suppose that 1 value probabilities of all input variables are equal to 0.5. BDDs $R^{cont}(1)$ and R^{obs} are used for calculating observability and 1-controllability random estimations for node v.

Thus random estimations are obtained by using a structural description of a combinational part. But the behavior of this part as a rule is wider than the working area represented by a State Transition Graph (STG). The point is that a TC may be triggered just out of the working area (out of the specification). If we know the STG description (the specification) from which the combinational part of the sequential circuit is obtained, we may calculate precisely random estimations of controllability out of the working area. As for precise random observability estimations they are always calculated by using only description of structural combinational part.

3. Deriving precise controllability estimations out of working area

Let we have STG description of a behavior of a sequential circuit. To obtain a sequential circuit we have to encode internal states of STG. As a result we get the system of incompletely specified Boolean functions that represents the working area of a sequential circuit. Changing this system into system of completely specified Boolean functions we make easier capabilities of TCs insertion. It's because that getting minimized system of completely specified Boolean functions we enlarge, as a rule, set-off and set-on areas of these functions in comparison with the initial system of incompletely specified Boolean functions. After these operations can appear the full states (depending on input and state variables) that don't belong to the working area. Such full states cannot reach during sequential circuit testing and verification in the working area. These full states may be used for triggering TCs. Therefore, we propose to calculate 1(0)-controllability precise estimations for internal nodes out of the working area.

STG description of Finite State Machine behavior has already encoded symbols of input and output alphabets. One approach to deriving a combinational part of a sequential circuit is as follows.

Encode internal states by unordered code words and obtain system of incompletely specified Boolean functions F.

Replace symbol «0» by symbol «-» (don't care) in code words of internal states and obtain the system of completely specified Boolean functions F^*. This special way of minimization is possible because we use the unordered code for encoding of internal states [6].

Each function f^* of system F^* is presented by Sum of Products (SoP) generated by the proper cubes. The system F^* is used to derive a combinational part of a sequential circuit comprising from gates.

Note that the working area of the sequential circuit is represented by system F of incompletely specified Boolean functions. Make the SoP from all cubes of system F. Derive BDD R^w from the SoP. Let R^{nw} be an inversion of R^w. Calculate 1(0)-controllability for node v within working area using BDDs:

$$R^{cont\ w}(1) = R^{cont}(1)R^w, \quad (2)$$

$$R^{cont\ w}(0) = R^{cont}(0)R^w, \quad (3)$$

and 1(0)-controllability out of working area may be calculated using BDDs:

$$R^{cont\ nw}(1) = R^{cont}(1)R^{nw}, \quad (4)$$

$$R^{cont\ nw}(0) = R^{cont}(0)R^{nw}. \quad (5)$$

If we have only structural description of sequential circuit and know nothing about circuit working area, we derive estimations applying $R^{cont}(1)$ (R^{cont} (0)).

We suggest including the internal nodes into set V of suspicious nodes if the chosen 1(0)-controllability estimations are less than the given threshold. In case if the controllability for node v equals 0 then the node v is excluded from further consideration.

If we know a possible type of TC, we can cut the set V using a precise estimation of observability. Let the internal node v^* is connected with TC output. If the precise estimation of node v^* observability is more than the proper threshold, we exclude corresponding node v from further consideration.

Note that implanting TC changes values both on pole v^* and the proper output. In case if the observability for node v^* equals 0 then the node v is excluded from consideration. Otherwise BDDs used for calculating random estimations for nodes v, v^* may be applied for evidence of the existence of an activating sequence that provides a harmful effect of the TS and finding the sequence itself if it is necessary.

Node v may also be eliminated from the consideration if there is no rather short transfer sequence triggering TC with input v and output v^*.

Execute multiplication BDD $R^{cont\ nw}$ (1) or $R^{cont\ nw}$ (0) for node v and BDD R^{obs} for node v^*. The multiplication result is represented by BDD R^f. Here we consider that a TC is inserted out of working area (out of specification).

Products created by paths from R^f root to the terminal node 1 represent sets of full states of the sequential circuit. A harmful effect of the TS can be provided by reaching any state from these sets [5].

Then we may find the transfer sequence itself for each node of the obtained set V using algorithm [7]. Applying the derived transfer sequences for set V we may detect node v in which TC is inserted. Based on the result we may mask TC attack.

4. Trojan Circuit masking

If we suppose that Trojan Circuit is inserted not out of working are, we may mask it in the following way (Fig. 1).

Here masking sub-circuit together with MUX and XOR are out of sequential circuit area. The sub-circuit implements the same function that the sub-circuit of the combinational part with output v^*. When Trojan Circuit is triggered the proper output keeps the correct value.

In the case of injecting Trojan circuit into out of working area we suggest the more simple way of masking (Fig. 2). The masking sub-circuit implements the function represented by BDD R^f. Connecting the proper output with MUX we keep the correct behavior of a sequential circuit.

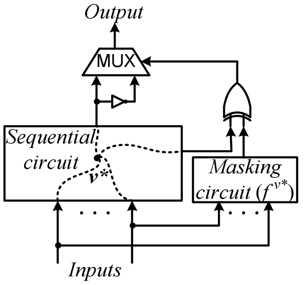

Figure 1. Masking TC scheme

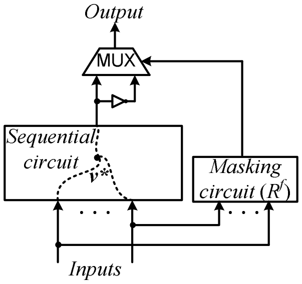

Figure 2. Masking TC scheme inserted into out of working area

In this case we need STG description of the sequential circuit behavior and so we use MCNC [8] sequential benchmark circuits in KISS2 format for experiments.

The set of circuits has been made from KISS2 format (from STG description) by 1-hot (unordered) encoding of states. Then replace symbol «0» by symbol «-» (don't care) in code words of internal states and obtain the system of completely specified Boolean functions F^*. After that logic synthesis and optimization in ABC system is used [9].

For experiments we have limited to TCs which can be inserted into internal nodes with low controllability estimations without taking into consideration observability estimations. This approach is suited for any type of TC. When we know the type, we may use more simple BDDs R^f and consequently to cut overhead.

Experiments show that for each internal node with low controllability there exists rather short transfer *sequence* [7] *triggering TC. For the benchmark circuits considered the transfer sequence lengths are not more than 8 (in average 1.1).*

Calculations of controllability estimations for internal nodes of combinational part of sequential circuits in and out of working area are represented in Tables I, II. In these tables overhead estimations of masking sub-circuits corresponding to 10 nodes with lesser controllability estimations for each circuit are also presented. There are the names of benchmarks (Circuits), numbers of gates (N_Gs), minimum nonzero values of controllability estimations (Min_VC), parts of gates (their output nodes) with values of controllability less or equal to 0.05 (%Gs^1), 0.005 (%Gs^2) and 0.0005 (%Gs^3) in percentage, sizes of minimum masking sub-circuits as a percentage from initial circuit (%Min) and sizes of maximum masking sub-circuits as a percentage from initial circuit (%Max) for 10 internal nodes with lesser controllability estimations.

Benchmark circuits and masking sub-circuits are received in ABC and they consist of 2-input logic-gates.

Table I. Experimental results for TC in working area

Circuit	N_Gs	Min_VC	%Gs^1	%Gs^2	%Gs^3	%Min	%Max
cse	145	0.0000305176	17.2	1.4	0.7	3.4	11.0
dk14	102	0.03125	2.9	0.0	0.0	1.0	11.8
dk16	142	0.015625	7.0	0.0	0.0	2.1	16.2
ex1	176	0.0000305176	14.2	4.5	2.8	0.6	10.8
keyb	193	0.00138255	16.1	1.6	0.0	1.0	33.2
kirkman	126	0.0000305176	10.3	1.6	0.8	0.8	30.2
sand	388	0.000000159256	10.1	2.8	0.8	0.3	12.9
sse	88	0.000731945	10.2	1.1	0.0	2.3	23.9
styr	305	0.000000953674	16.7	2.6	2.0	2.0	15.4
tbk	669	0.000000000232831	21.5	3.9	2.2	0.4	6.0
train11	44	0.03125	2.3	0.0	0.0	6.8	13.6

Table II. Experimental results for TC out of working area

Circuit	N_Gs	Min_VC	%Gs^1	%Gs^2	%Gs^3	%Min	%Max
cse	145	0.000000238419	54.5	54.5	54.5	0.7	26.9
dk14	102	0.000976562	58.8	34.3	0.0	2.0	14.7
dk16	142	0.00000000186265	55.6	55.6	55.6	0.7	6.3
ex1	176	0.000000178814	76.7	38.6	19.3	0.6	5.1
keyb	193	0.0000000596046	58.5	58.5	58.5	0.5	9.8
kirkman	126	0.000000476837	17.5	4.0	3.2	0.8	19.0
sand	388	0.000000000009095	52.8	52.8	52.8	0.3	1.5
sse	88	0.0000038147	51.1	51.1	51.1	1.1	8.0
styr	305	0.000000000021828	58.4	58.4	58.4	0.3	11.5
tbk	669	0.000000000003638	60.7	60.7	60.7	0.1	23.8
train11	44	0.000488281	59.1	15.9	2.3	4.5	13.6

Masking TC with using out of working area requires as a rule smaller overhead (in average from 1.1% to 12.8%) in comparison with masking TC with using only structural description of a combinational part (in average from 1.9% to 16.8%).

5. Conclusion

Possibilities of triggering TC are examined. Precise description of switch on TC area together with precise estimations of internal node observability and controllability are derived. They both are obtained by using structural combinational part description. The approach to TCs detection inserted out of working area is suggested. The techniques of masking TCs are proposed.

6. References

[1] R. Karri, J. Rajendran, K. Rosenfeld, and M. Tehranipoor, “Trustworthy Hardware: Identifying and Classifying Hardware Trojans,” *Computer*, vol. 43, no. 10, pp. 39–46, Oct. 2010.

[2] R. S. Chakraborty, S. Pagliarini, J. Mathew, R. S. Ranjani, and M. N. Devi, “A Flexible Online Checking Technique to Enhance Hardware Trojan Horse Detectability by Reliability Analysis,” *IEEE Trans. Emerg. Top. Comput.*, vol. PP, no. 99, pp. 1–1, 2017.

[3] A. Waksman, M. Suozzo, and S. Sethumadhavan, “FANCI: Identification of Stealthy Malicious Logic Using Boolean Functional Analysis,” in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, New York, NY, USA, 2013, pp. 697–708.

[4] A. Matrosova, S. Ostanin, and I. Kirienko, “Generating all test patterns for stuck-at faults at a gate pole and their connection with the incompletely specified Boolean function of the corresponding subcircuit,” in *2014 14th Biennial Baltic Electronic Conference (BEC)*, 2014, pp. 85–88.

[5] A. Y. Matrosova, I. E. Kirienko, V. V. Tomkov, and A. A. Miryutov, “Reliability of Physical Systems: Detection of Malicious Subcircuits (Trojan Circuits) in Sequential Circuits,” *Russ. Phys. J.*, vol. 59, no. 8, pp. 1281–1288, Dec. 2016.

[6] F. Y. Busaba and P. K. Lala, “Self-checking combinational circuit design for single and unidirectional multibit error,” *J. Electron. Test.*, vol. 5, no. 1, pp. 19–28, Feb. 1994.

[7] A. Matrosova, V. Andreeva, and A. Melnikov, “ROBDDs application for finding the shortest transfer sequence of sequential circuit or only revealing existence of this sequence without deriving the sequence itself,” in *2016 IEEE East-West Design Test Symposium (EWDTS)*, 2016, pp. 1–4.

[8] S. Yang, *Logic Synthesis and Optimization Benchmarks User Guide Version 3.0.* 1991.

[9] *ABC: A System for Sequential Synthesis and Verification.* (http://www.eecs.berkeley.edu/~alanmi/abc/).