Quality Assessment of RFET-based Logic Locking Protection Mechanisms using Formal Methods

Marcel Merten* Sebastian Huhn*† Rolf Drechsler*†

*University of Bremen, Germany
{mar_mer,huhn,drechsle}
@informatik.uni-bremen.de

†Cyber-Physical Systems, DFKI GmbH
28359 Bremen, Germany

***Abstract*—The high distribution of the manufacturing of Integrated Circuits (ICs) over different foundries yields long and untrustworthy supply chains. Logic locking is one prominent protection technique against malicious usage and counterfeit. The emerging technology of Reconfigurable Field-Effect Transistors (RFETs) has recently been utilized to implement new polymorphic logic mechanisms to protect intellectual property. The mechanisms' assessment is important to reinforce the newly introduced protection mechanism and, hence, avoid any weak logic structures. So far, approximate Hamming Distance-based assessment techniques have been used for determining the protection quality while considering combinatorial circuits only. This work proposes a novel method to assess the quality of the RFET-based logic locking structures for sequential circuits. In particular, formal techniques are orchestrated to analyze the circuit's state space to determine whether any incorrect keys exist that unintentionally unlock and exhibit the circuit's correct functional behavior. The experimental evaluation validates that the proposed scheme unveils weaknesses of the protection structure, which remain undetected when using existing techniques.**

I. Introduction

Due to the *Integrated Circuits* (ICs) manufacturing globalization, designers can benefit from access to advanced technology nodes without having the large capital expenditure of operating their own semiconductor foundries. The distribution of the chips' manufacturing is one of the main security challenges. A growing threat prevails about compromising the integrity of once trusted IC processes by unauthorized or untrusted users [1]. During the last decade, *Complementary Metal-Oxide-Semiconductor* (CMOS)-based protection mechanisms have been the dominant technology for implementing various protection measures. However, a trade-off between the achievable protection level and the resulting cost overhead exists.

Recent works like [1], [3], [4] have been focusing on achieving high protection while still preserving low overhead by utilizing reconfigurable silicon nanowire field-effect transistor-based polymorphic logic gates [1]. In [1], an algorithm is proposed that replaces gates with high impact on the original circuit's behavior by reconfigurable polymorphic logic gates. Afterward, the quality of the resulting logic locking functionality is assessed by a metric based on the *Hamming Distance* (HD) of the outputs over certain applied stimuli. The result is considered optimal if the HD is 50% of the maximal HD. Due to the simulation-based nature of the existing approaches, they cannot cover all input and key combinations.

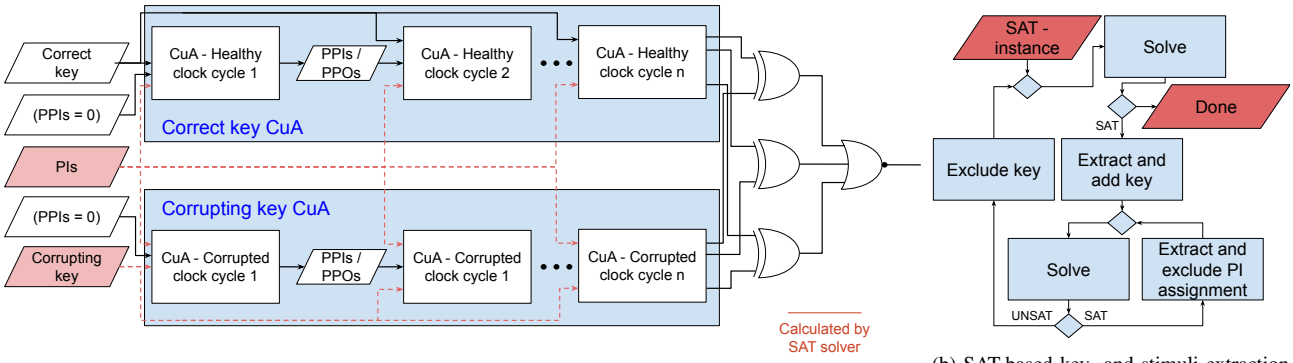
This work proposes a novel technique to assess the quality of introduced *Reconfigurable Field-Effect Transistor* (RFET)-based logic locking protection mechanisms by orchestrating the *Boolean Satisfiability* (SAT) problem to tackle the shortcomings of existing approaches. Various experiments have been conducted on the ITC'99 benchmark set. Compared to the HD-based technique, the proposed approach determines more information about the encrypted circuit's behavior and clearly outperforms the analysis capability for weak logic locking mechanisms, as provided by any other existing technique.

This work was financially supported by the German Federal Ministry of Education and Research BMBF under the framework of VE-CirroStrato and the AI initiative of the Free Hanseatic City of Bremen. We would like to thank Verific Design Automation Inc. for providing the SystemVerilog frontend used for the implementation of our technique.

II. Quality Assessment Framework

At first, an inverted miter circuit is generated from the CuA while considering the a-priori known *correct* key k_c yielding the SAT instance Φ_{k_c} and any incorrect key in $\widehat{\mathcal{K}}$ yielding $\Phi_{\widehat{\mathcal{K}}}$. The basic principle of this construction is given in Sub-figure 1a. The CuA is unrolled for N clock cycles since sequential elements – meaning *Flip-Flops* (FFs) – have to be considered for an exact assessment in terms of sequential circuits' unrolling [5], [6]. Here, the value N has to be adjusted for the CuA characteristics. Furthermore, 0 is assumed as the initialization value for all FFs in cycle $n = 1$. For keeping the resulting miter model small, only the relevant combinatorial logic, i.e., the transitive fan-in, for the corresponding output is calculated, which is done for all cycles n (with $1 \leq n \leq N$).

The primary inputs are equally driven for both unrolled instances (of the CuA) and are kept constant during the unrolling. After the inverted miter has been added, the key is constrained for both instances of the unrolled CuA. For Φ_{k_c} , the correct key k_c is set by adding clauses implying k_c, whereby $\Phi_{\widehat{\mathcal{K}}}$ is extended by a conflict clause excluding k_c. The entire model is stored as one SAT instance Φ_{comp} and processed by a state-of-the-art SAT solver. Sub-figure 1b presents the general approach about how corrupting keys are being evaluated. If Φ_{comp} is unsatisfiable, the correct key is considered as stable. If a satisfiable solution is determined, a corrupting key k_f has been detected yielding a functional equivalent behavior of the CuA given at least one stimuli. For a qualitative assessment of the discovered security threat, every determined corrupting key is evaluated against the number of possible stimuli leading to this breach.



(a) Resulting abstract model of the CuA

(b) SAT-based key- and stimuli extraction

Figure 1: Quality assessment technique

III. Experimental evaluation

This section describes the experimental evaluation of the proposed quality assessment framework for RFET-based logic locking protection mechanisms and discusses the obtained results.

All experiments have been executed on an *AMD 4750U* processor with 32 GB system memory. The proposed technique has been solely implemented in C++. For the evaluation, different benchmark circuits of the *ITC'99* benchmark suite are considered. For each of these circuits, ten of the *NOR*, *NAND*, *XOR* and *XNOR* gates have been randomly replaced by RFETs. Consequently, each circuit has ten control signals resulting in $2^{10} = 1,024$ possible keys. Furthermore, a maximum of 1,024 stimuli (per corrupting key k_f) is captured – for limiting the computation time per corrupting key – if the CuA behaves functionally correct even though a corrupting key is applied. Each circuit has been unrolled for five clock cycles, which has been proven as an appropriate parameter to cover the functional behavior's majority (of the considered benchmark circuits) [7].

Table I shows the detailed results as follows: The average HD, whereby 10,000 randomly chosen stimuli and key combinations were considered. The absolute number of corrupting keys, the minimum, the average, and the maximum number of corrupting stimuli per key detected the proposed approach.

Following the HD-based approach, the circuits b06, b08, b09, b12, and the b14 are close to the $HD_{avg}(S) = 0.5$, considered optimal in [1]. However, the proposed SAT-based quality assessment technique shows that all these circuits have at least one k_f. In particular, b12 is corrupted by about half of the possible incorrect keys and each of those corrupting keys is corrupting all 32 possible stimuli of the circuit. Thus, the chance to randomly select either the correct key or a corrupting key resulting in completely equivalent behavior of the circuit is at 50%. This security breach has a $HD_{avg}(S)$ of 0.431 that is very close to the optimal value. This reflects the weak HD-based assessment quality.

TABLE I: Results

circuit	HD	$\#\{k_c\}$	#stimuli		
			minimum	average	maximum
b05	0.712	3	1	1	2
b06	0.597	127	2	3	4
b07	0.767	15	2	2	2
b08	0.564	127	256	286	512
b09	0.571	15	2	2	2
b10	0.284	63	416	823	1,024
b11	0.012	1,023	126	126	128
b12	0.431	511	32	32	32
b13	0.746	127	512	512	512
b14	0.386	1,023	1,024	1,024	1,024
b15	0.168	1,023	1,024	1,024	1,024
b20	0.807	31	1,024	1,024	1,024
b21	0.837	31	1,024	1,024	1,024

IV. Conclusions

This paper presented a novel method for assessing the quality of RFET-based logic locking protection systems In the end, the proposed framework allows determining corrupting keys and evaluates their threat to the protection system while considering for the first time even sequential circuits. In contrast to other approaches, the assessment is conducted exactly considering the fully functional state space of the circuit. Future work will enhance the SAT-based model by incorporating Pseudo-Boolean Optimization techniques and investigates a compositional approach allowing for processing even larger industrial-sized designs.

References

[1] Q. Alasad, J.-S. Yuan, and Y. Bi, "Logic locking using hybrid CMOS and emerging SiNW FETs," *Electronics*, vol. 6, no. 3, 2017.

[2] S. Rai, S. Srinivasa, P. Cadareanu, X. Yin, X. S. Hu, P.-E. Gaillardon, V. Narayanan, and A. Kumar, "Emerging reconfigurable nanotechnologies: Can they support future electronics?" in *IEEE/ACM International Conference on CAD*, 2018.

[3] Q. Alasad and J. Yuan, "Logic obfuscation against IC reverse engineering attacks using PLGs," in *IEEE International Conference on Computer Design*, 2017, pp. 341–344.

[4] Q. Alasad, J.-S. Yuan, and P. Subramanyan, "Strong logic obfuscation with low overhead against IC reverse engineering attacks," *IEEE Transaction on CAD of Integrated Circuits and Systems*, vol. 25, no. 4, 2020.

[5] R. Arora and M. Hsiao, "Enhancing SAT-based bounded model checking using sequential logic implications," in *International Conference on VLSI Design*, 2004, pp. 784–787.

[6] G. Fey, A. Sulflow, S. Frehse, and R. Drechsler, "Effective robustness analysis using bounded model checking techniques," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 8, pp. 1239–1252, 2011.

[7] A. Finder, A. Sülflow, and G. Fey, "Latency analysis for sequential circuits," in *2011 Sixteenth IEEE European Test Symposium*, 2011, pp. 129–134.