Revisiting Capacitor-Based Trojan Design

Mohammad Mahdi Bidmeshki, Kiruba Sankaran Subramani and Yiorgos Makris
Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, Texas
E-mail: {bidmeshki, kiruba.subramani, yiorgos.makris}@utdallas.edu

***Abstract*—Among the various strategies for hiding malicious capabilities in integrated circuits (ICs), analog circuit design techniques have recently drawn increased attention due their lower area, power and delay footprints, which make their detection significantly more challenging. Specifically, switched capacitors have been used for creating stealthy trigger circuits based on toggling activity on a victim wire. Various methodologies for detecting such culprits have been investigated; however, recent literature in this area contains several misconceptions or inaccuracies regarding the topologies of these trigger circuits and the effectiveness of previously proposed detection methods. Therefore, in this paper, we first revisit the design of switched capacitor-based trigger circuits and we present several design configurations which are not encompassed by previously demonstrated models, but which can also serve the same malevolent purpose. We, then, discuss the effectiveness and the shortcomings of existing defense methodologies, and we point towards additional research that is needed in this area.**

I. Introduction

Hidden malicious capabilities cteated by hardware Trojans, which are stealthily inserted in ICs, have become a serious threat to the electronics industry [1]. While most hardware Trojan design and detection methods focus on digital approaches [1]–[3], recent efforts have leveraged analog design techniques to create stealthier Trojans, which can better evade detection. Specifically, the A2 Trojan [4] used switched capacitors to implement a trigger mechanism for privilege escalation in a microprocessor. When compared to a digital alternative, the Trojan circuitry in [4] has a smaller area, power and delay footprint, thereby evading detection by functional testing and verification techniques. Beyond A2, a recent study [5] leveraged capacitive coupling in a $45nm$ CMOS process to escalate user privileges and to leak data in an AES core. Notably, the approach reroutes existing resources, incurring no area overhead and making its detection particularly challenging.

As a result, various studies have attempted to detect these Trojan circuits and to evaluate the security and trustworthiness of analog/mixed signal (AMS) designs. In [6], the R2D2 solution introduced additional circuitry for monitoring the toggling frequency of a signal and marking it as suspicious if it exceeds a certain threshold. Alternatively, Information Flow Tracking (IFT) has been used for detecting such stealthy capabilities in AMS designs. Specifically, the analog enhanced VeriCoq-IFT solution [7] is defined at the transistor-level and is capable of tracking information flow crossing the analog and digital domains, such as the A2 stealthy trigger. Along a different direction, [5] proposed a reverse engineering approach for detecting possible security risks in the layout, based on the wire-length characteristics of such capacitive crosstalk.

A recent study [8] sought to model switched capacitor-based Trojan triggers and to develop a detection method which can be applied at the transistor-level netlist generated from the final design layout. The study uses this model along with a desired activation frequency to determine the minimum capacitance that is needed for realizing a trigger circuit. Thereby, any capacitor which exceeds this minimum value in the design and which is connected to a user-controllable source, such as a flip-flop, is marked as a potential trigger circuit. While the overall approach is interesting, the model does not cover or detect other variants of such triggers. Furthermore, it incorrectly characterizes some of the variants as benign or non-trigger circuits, resulting in misrepresentation of the effectiveness of previously proposed methods. Evidently, different variants of Trigger circuits and their inherent threat are not fully understood and, therefore, current methods are not capable of identifying such threats. To help mitigate this situation, in this paper we revisit the topic of switched capacitor-based trigger circuits and we present alternate topologies capable of creating a similar functionality. We also study the effectiveness of existing methods in detecting these circuits and we highlight their strengths and weaknesses.

The rest of this paper is organized as follows. The primary switched capacitor circuit design utilized in A2 [4] and which was modeled in [8], is reviewed in Section II. Alternative circuit topologies which can serve as analog trigger are presented in Section III. Effectiveness of existing methods in detecting these alternative topologies is examined in Section IV. Finally, conclusions are drawn in Section V.

II. Capacitor-based Malicious Circuit Primary Configuration

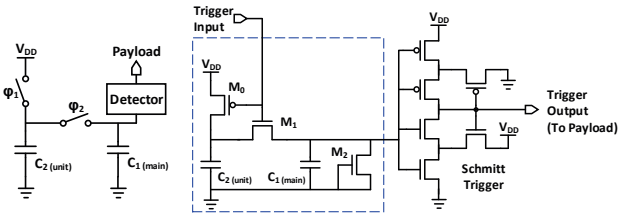
In this section, we review the primary design which uses switched capacitors to create a stealthy trigger mechanism. While the switching activity on the victim wire is below a certain threshold, the trigger is inactive and does not affect the operation of the circuit (e.g., a microprocessor). Upon high-enough toggling activity on the victim wire, however, the malicious behavior is activated and a trigger signal is created.

Fig. 1a depicts the primary switched capacitor configuration employed in [4] and modeled in [8]. On ϕ_1 activation, $C_{2\ (unit)}$ is charged to V_{dd}. Then, on ϕ_2 activation, $C_{2\ (unit)}$ transfers its charge to $C_{1\ (main)}$. Switching consecutively between ϕ_1 and ϕ_2 increases the voltage on $C_{1\ (main)}$ and, if this voltage exceeds a threshold, the detector activates the trigger output. The voltage increase at each step can be calculated by Equation 1, where V_0 is the initial voltage on $C_{1\ (main)}$ before ϕ_2 activation [4]:

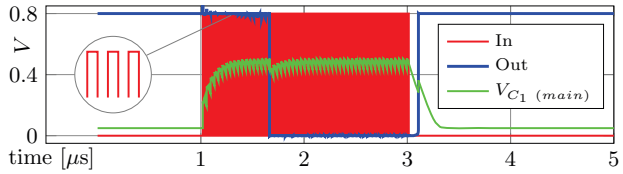
978-1-5386-6648-7/19/$31.00 ©2019 IEEE
DOI 10.1109/ICCD46524.2019.00047

TABLE I: Characteristics of switched capacitor circuit configurations in simulations

Configuration	$C_{1\ (main)}$ (Quantized W)	$C_{2\ (unit)}$ (Q. W)	Estimated Total Area (Q. W)	Activation Delay at Trigger Frequency (ns)	Trigger Frequency (MHz)	Retention Time (ns)
Primary	20	2	28	665	26.3	93
Discharging	20	2	34	366	45.0	147
Series	20	2	28	577	33.8	107
Distributed	10+10	2	32	390	60.2	58
Many-phase	10+10	2	32	323	85.5 (33% Duty Cycle)	61
Charge-pump	20	2	36	700	17.9	89



(a) Configuration (b) Circuit example



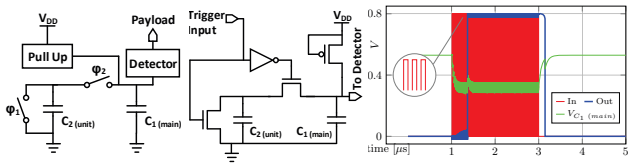
(c) Simulation results

Fig. 1: Primary switched capacitor design for a stealthy trigger

$$\Delta V = \frac{C_{2\ (unit)}}{C_{2\ (unit)} + C_{1\ (main)}}(V_{dd} - V_0) \tag{1}$$

Specifically, $V_0 = 0$ before the input toggling begins. Fig. 1b shows a circuit which implements such a configuration, as presented in [4], and which uses a Schmitt trigger inverter as a detector. To minimize the loading effect on the input (victim) wire and to also minimize the area overhead, $C_{2\ (unit)}$ is set to a unit size capacitor, hence the name. In this circuit, M_2 provides an additional leakage path for discharging $C_{1\ (main)}$, thereby prolonging the time required for the trigger to be activated (i.e., activation time) and reducing the time that the trigger remains activated after the input toggling ends (i.e., retention time). Evidently, the trigger activation time, activation frequency and retention time depend on the capacitor and transistor sizing and the inherent process leakage current.

Fig. 1c shows the practical behavior of this circuit in simulation using the FreePDK15 [9] which models a $15nm$ FinFET process. In this simulation, $C_{2\ (unit)}$ is set to the minimum NMOS transistor size in this process, i.e., the quantized width (FIN multiplier) is set to 2, and $C_{1\ (main)}$ is set to 10 times the size of $C_{2\ (unit)}$. Transistors in the Schmitt trigger circuit and all other transistors are set to the minimum size. The result shown in Fig. 1c corroborates that such a circuit can be used as a Trojan trigger, even in modern, state-of-the-art technologies, where the leakage current might be a concern. The first row of Table I shows the characteristics of this trigger circuit, including its trigger activation frequency, based on these settings. The area of the circuit is estimated based on the unit quantized width and does not include the detector (Schmitt trigger) circuit. Retention time in this example is small, mainly due to the small $C_{1\ (main)}$.



(a) Configuration (b) Circuit example (c) Simulation results

Fig. 2: Discharging switched capacitors

While this primary configuration seems to be the simplest model that achieves the realization of a stealthy trigger, there are many other switched capacitor configurations that produce similar results, some of which we present in the next section.

III. Alternative Trigger Designs

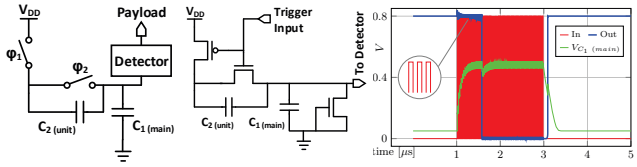
In this section, we study other variants of switched capacitor-based circuits, which can also be utilized for creating a similar stealthy trigger mechanism as A2. Through simulation results, we show that malicious utilization of switched capacitor circuits is not limited to the primary model presented in Section II.

A. Discharging Switched Capacitors

Compared to the primary configuration, wherein $C_{2\ (unit)}$ is charged on ϕ_1 activation and the charge is transferred to $C_{1\ (main)}$ on ϕ_2 activation, in this variant, $C_{1\ (main)}$ is normally charged through a pull-up circuit, as shown in Fig. 2a. Therefore, by toggling the input, $C_{2\ (unit)}$ is discharged on ϕ_1 activation while, on ϕ_2 activation, it depletes charge from $C_{1\ (main)}$. If the toggling frequency is high enough, the pull-up circuit will not be able to keep the $C_{1\ (main)}$ charged and its voltage will drop. This, in turn, can trigger a change of state at the output.

Fig. 2b shows an example circuit for realizing this configuration and Fig. 2c shows the simulation results for this circuit, clearly depicting the trigger activation and deactivation, consistent with the toggling activity on the input. Due to the normally charged capacitor and the Schmitt trigger inverter, the trigger is active high in this case. We note that, while this topology has been examined in [8], it is stated that it is impossible to use it as a malicious trigger. This claim is incorrect, as evidenced by the simulation of Fig. 2c.

The second row in Table I shows the characteristics of this configuration, which has the same capacitor setting as the primary configuration in Section II. Compared to the primary configuration, the discharging model shows higher trigger frequency, lower activation delay and larger retention time.



(a) Configuration (b) Circuit example (c) Simulation results

Fig. 3: Switched capacitors in series

B. Switched Capacitors in Series

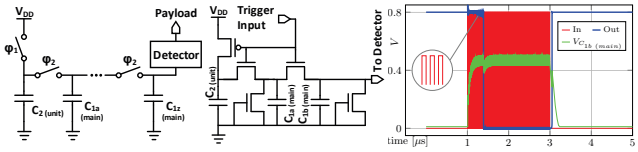
As shown in Fig. 3a, in this variant, $C_{2\ (unit)}$ and $C_{1\ (main)}$ are configured in series and are charged together upon ϕ_1 activation. Activation of ϕ_2 causes $C_{2\ (unit)}$ to discharge. Although, at first glance, this configuration may seem benign, it turns out that this variant also exhibits a behaviour which can be used maliciously. Fig. 3b shows an example implementation of this variant. Successive toggling of the input can increase the voltage on $C_{1\ (main)}$ to a point where the output is triggered, as shown in the simulation results of Fig. 3c. Again, we note that, while this topology has been examined in [8], it is stated that it is benign. This claim is incorrect, as evidenced by the simulation of Fig. 3c.

The third row of Table I shows the characteristics of this variant which, compared to the primary configuration, show higher trigger frequency, slightly lower activation delay and somewhat larger retention time. While this and the following variants are demonstrated using charging of switched capacitors, we note that it may be possible to also design discharging switched capacitor versions.

C. Distributed Switched Capacitors

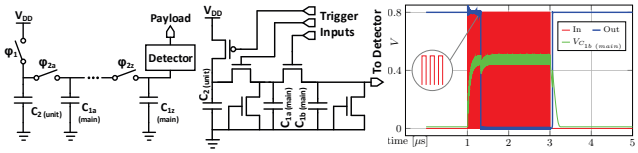
Large capacitor values, especially for $C_{1\ (main)}$, could be a revealing feature for detecting malicious capacitor-based circuits. At the expense of a higher number of switches, this variant distributes large capacitance into several capacitors and connects them together via switches, as shown in Fig. 4a. This can deceive detection methods, such as [8], which look for a large capacitance in the design or the netlist. Especially in digital designs, where such large capacitors are not common and may only be utilized for power supply decoupling, this can be an identifying feature. While only $C_{1\ (main)}$ is broken into several capacitors in Fig. 4a, the same can be done for $C_{2\ (unit)}$ if necessary. However, since normally $C_{1\ (main)}$ is larger than $C_{2\ (unit)}$, this approach is more suitable for distributing $C_{1\ (main)}$. In this variant, switches connecting these partial capacitors are activated at the same time. The effective capacitance for $C_{1\ (main)}$ is the total of the partial capacitors, i.e., $C_{1\ (main)} = C_{1a} + C_{1b} + \cdots + C_{1z}$. Consequently, the voltage increase at each step can be approximated using Equation 1, by replacing this equivalent $C_{1\ (main)}$ value and neglecting the effect of ϕ_2 switches.

Fig. 4b shows an example circuit implementing this variant by distributing $C_{1\ (main)}$ into two capacitors. As confirmed by the simulation results in Fig. 4c, this variant can be utilized for building a malicious trigger. We note that a similar design can also be created for the discharging capacitor case described in Section III-A.



(a) Configuration (b) Circuit example (c) Simulation results

Fig. 4: Distributed switched capacitors



(a) Configuration (b) Circuit example (c) Simulation results

Fig. 5: Many-phase switched capacitors

The forth row of Table I shows the characteristics of this distributed configuration which, compared to the primary case, shows much higher trigger frequency and lower activation delay. The retention time in this case is smaller due to the lower capacitance value associated with $C_{1b\ (main)}$.

D. Many-Phase Switched Capacitors

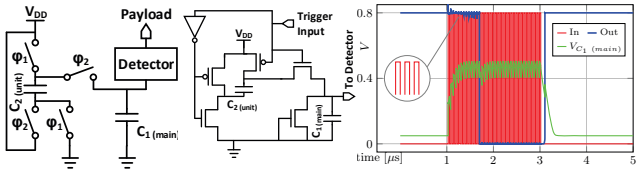
Similar to the distributed variant demonstrated in Section III-C, this variant also distributes $C_{1\ (main)}$ among several capacitors. However, the switches in this variant are independently toggled to activate the trigger, as shown in Fig. 5a. In addition to distributing the capacitance, this variant can provide more flexibility in designing the toggling pattern and frequency leading to the trigger activation, considering that ϕ_1 is not activated at the same time as $\phi_{2a} \ldots \phi_{2z}$. Since the inputs in this variant do not need to be activated all in the same cycle, an attacker may be able to evade detection methods which monitor the switching activity on a signal in a time window, such as [6].

Fig. 5b shows an example circuit implementing this variant. To demonstrate a similar triggering capability, Fig. 5c shows simulation results of this circuit in which In represents one of the trigger inputs only. In this simulation, the trigger inputs were activated in sequence and only one was active at a time.

The fifth row of Table I shows the characteristics of this many-phase variant which, compared to the primary case, shows much higher trigger frequency and lower activation delay. However, we note that compared to the other variants, the toggling activity in this variant has a duty cycle of 33%. Similar to the distributed variant, the retention time in this case is smaller due to the lower capacitance value of $C_{1b\ (main)}$.

E. Charge-Pump Switched Capacitors

This variant employs charge-pump switched capacitors to create a stealthy trigger, as shown in Fig. 6a. In this configuration, on ϕ_1 activation $C_{2\ (unit)}$ is charged to V_{dd}. When ϕ_2 is activated, $C_{2\ (unit)}$ is connected in series with V_{dd}, boosting the voltage on $C_{1\ (main)}$ to $2 \times V_{dd}$. Although this variant requires more switches, since it can boost the voltage on $C_{1\ (main)}$, it can activate the trigger with much less delay and keep it active with much less switching activity on the



(a) Configuration (b) Circuit example (c) Simulation results

Fig. 6: Charge-pump switched capacitors

input. Therefore, it might be useful for cases where the input signal cannot be kept toggling with high frequency. On the other hand, these properties may increase the risk of unwanted trigger activation, which needs to be considered in the design.

Fig. 6b shows an example implementation of this variant. Note that, due to the use of an NMOS transistor as a switch connecting $C_{2\ (unit)}$ and $C_{1\ (main)}$, the voltage on $C_{1\ (main)}$ cannot reach above $V_{dd} - V_{th}$, where V_{th} is the threshold voltage of the NMOS transistor. Alternatively, PMOS transistors can be employed if needed. Similar to the other variants presented herein, the charged-pump based variant can also serve the purpose of a stealthy trigger, as shown in the simulation results of Fig. 6c.

The last row of Table I shows the characteristics of this variant which, compared to the primary case, shows much lower trigger frequency with similar activation delay and retention time.

IV. Detecting Capacitor-Based Malicious Circuits

In this section, we study the effectiveness of three previously proposed methods [6]–[8] in identifying possible capacitor-based malicious circuits in hardware designs.

Online, In-Field Detection: As mentioned earlier, R2D2 [6] introduces additional hardware to monitor the switching activity of signals considered as potential threats. This solution can detect most of the trigger variants presented here. However, in the case of the many-phase variant, an attacker may be able to design an activation pattern that does not fall into the R2D2 frequency threshold. The hardware overhead (area and power) and the need for fine-tuning of the monitoring parameters to avoid false positives (benign signal transition patterns marked as malicious) and false negatives (malicious signal transition patterns marked as benign), as well as the selection of signals to monitor, which are often unknown, are some of its major shortcomings.

Model-Based Detection: The approach introduced in [8] models the primary configuration we reviewed in Section II and uses this model to define constraints in the form of the switched capacitor circuits that can be utilized as a malicious trigger. In the first step, this methodology selects capacitors larger than a specific threshold by assuming practical limitations on the activation frequency. Then, it searches the design for a configuration similar to Fig. 1a and traces back the activation of switches to user controllable inputs, such as flip-flops, using a process which it terms *taint propagation*. While this methodology considers a variety of logic gate structures that can activate the switches for $C_{2\ (unit)}$, it fails to detect the variants presented in Section III. This is because each one of these variants invalidates at least one of the assumptions that were made in the model employed in [8], as summarized in Table II.

TABLE II: Model-based detection summary

Variant	Detected?	Reason
Primary	✓	Modeled
Discharging	×	Not modeled
Series	×	Not modeled
Distributed	×	Smaller $C_{1\ (main)}$
Many-phase	×	Smaller $C_{1\ (main)}$
Charge-pump	×	Not modeled

Transistor-Level Analog IFT: The analog-enhanced VeriCoq-IFT method, as described in [7], can also be utilized for detecting paths which lead to information leakage in an analog/mixed signal design and can detect capacitor-based malicious circuits with proper labeling of signals. On the down side, due to the low level and conservativeness of this IFT approach, it may produce some false positives.

V. Conclusion

Due to their minimal overhead and challenging detection, switched capacitor-based triggers have become a topic of serious interest. While one form of such circuits has gained great attention, in this work we demonstrated other variants which can be utilized with equal effectiveness for the same purpose. In light of these new variants, we evaluated existing methods for detecting these triggers and we dispelled various misconceptions in the literature regarding their effectiveness. The set of demonstrated variants may not necessarily be exhaustive, as other design strategies may be used to create future variants. Nevertheless, by elucidating their fundamental operating principles and by highlighting the limitations of existing methods, we can steer the community towards developing effective solutions for detecting such malicious circuits.

References

[1] S. Bhunia, M. S. Hsiao, M. Banga, and S. Narasimhan, "Hardware Trojan attacks: Threat analysis and countermeasures," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1229–1247, 2014.

[2] J. Zhang, F. Yuan, L. Wei, Z. Sun, and Q. Xu, "Veritrust: Verification for hardware trust," in *ACM Design Automation Conference (DAC)*, 2013, pp. 1148–1161.

[3] I. Wilcox, F. Saqib, and J. Plusquellic, "GDS-II Trojan Detection using Multiple Supply Pad V_{DD} and GND I_{DDQ}s in ASIC Functional Units," in *IEEE Symposium on Hardware Oriented Security and Trust (HOST)*, 2015, pp. 144–150.

[4] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, "A2: Analog malicious hardware," in *IEEE Symposium on Security and Privacy (S&P)*, 2016, pp. 18–37.

[5] C. Kison, O. M. Awad, M. Fyrbiak, and C. Paar, "Security implications of intentional capacitive crosstalk," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 14, no. 12, pp. 3246–3258, 2019.

[6] Y. Hou, H. He, K. Shamsi, Y. Jin, D. Wu, and H. Wu, "R2D2: Runtime reassurance and detection of A2 Trojan," in *IEEE Symposium on Hardware Oriented Security and Trust (HOST)*, 2018, pp. 195–200.

[7] M.-M. Bidmeshki, A. Antonopoulos, and Y. Makris, "Information flow tracking in analog/mixed-signal designs through proof-carrying hardware IP," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2017, pp. 1703–1708.

[8] X. Guo, H. Zhu, Y. Jin, and X. Zhang, "When capacitors attack: Formal method driven design and detection of charge-domain trojans," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2019, pp. 1706–1711.

[9] K. Bhanushali and W. R. Davis, "FreePDK15: An open-source predictive process design kit for 15nm FinFET technology," in *International Symposium on Physical Design*, 2015, pp. 165–170.