Detecting Laser Fault Injection Attacks via Time-to-Digital Converter Sensors

Mohammad Ebrahimabadi*, Suhee Sanjana Mehjabin*, Raphael Viera†,
Sylvain Guilley‡, Jean-Luc Danger‡, Jean-Max Dutertre†, Naghmeh Karimi*.
*University of Maryland Baltimore County, United States
†Mines Saint-Etienne, CEA, Leti, Centre CMP, F - 13541 Gardanne France
‡Telecom Paris, Institut polytechnique de Paris, and Secure-IC S.A.S., France

***Abstract*—Fault Injection Attacks (FIA) have received a lot of attention in recent years. An adversary launches such an attack to abusively take control over the system or to leak sensitive data. Laser illumination has been considered as an effective technique to launch FIA. The laser-based FIAs are mainly used when the adversary opts to target a specific location in the target circuit. However, thanks to the miniaturization of transistors and moving towards smaller feature size, even small laser spots may illuminate more than one gate; making the attack more detectable when the circuitries are equipped with embedded fault detection mechanisms such as digital sensors. In this paper, we use time-to-digital convertors, aka digital sensors, to detect the laser shots. We show that by embedding these digital sensors in the target circuitry, the IR drop caused by the laser illumination can be sensed with a high accuracy. An alarm will be raised when the fault is detected. The simulation results show the high accuracy of the proposed scheme in detecting laser-based FIAs.**

I. Introduction

Sensing abnormal behaviors is highly important in the systems requiring a high level of safety and security. Such mission critical systems need to be equipped with appropriate sensors that raise an alarm if the system is under a physical attack, e.g., a Fault Injection Attack (FIA).

Fault Injection attacks mainly opt to leak sensitive data or provoke system malfunction. They can be implemented both at hardware and at software levels. In this paper, we target the hardware FIAs launched by laser light injection. The transient fault generated due to a laser light injection can highly jeopardize the security of the system via bypassing a security process (such as authentication), corrupting the data used to enforce security (such as privilege escalation in modern microprocessors), toggling the value of a specific signal at run time resulting an embedded cryptographic module to leak its encryption/decryption key, etc. (refer to MITRE CWE-1247 & CWE-1332).

Although relatively expensive, laser-based fault injection attacks have received the lion's share of attention by the adversaries in recent years; owing to the reproducibility of such attacks, the ease of attack, and their high precision in controlling where to inject faults [1]. Accordingly monitoring the security- and safety-critical systems during the runtime to protect them against laser-based FIAs is of utmost importance. One such monitoring can be provided with the Time-to-Digital Converters (TDC); the so-called Digital sensors (DS) hereafter.

Thanks to their low-cost calibration, high portability among different technologies, high accuracy, and resiliency against removal attacks, digital sensors have been broadly deployed to detect operating conditions such as temperature in recent years [2]; replacing the traditional analog counterparts [3]. In this paper, we leverage these sensors to detect laser-based FIAs and we show how such an embedded DS can raise an alarm when a laser attack is launched; thus promoting the safety and security of the underlying system. The contribution of this paper is as follows:

- A model to simply represent the impact of laser-based FIA in the target logic;
- A method to characterize the outcome of the embedded sensor and in turn the status of the underlying system in terms of attacks;
- Experimental results in terms of false and missed alarm rates for the considered FIAs.

II. Preliminaries

A. Fault Injection Attacks

FIAs mainly opt to force system malfunction or to leak sensitive data. Laser fault injection can be elucidated as a transient current resulting in a single event transient (SET) or single-event upset (SEU) in the targeted point; thus toggling its value. Indeed the laser shot induces parasitic current [4] and in turn an abnormal voltage drop.

Figure 1 depicts the impact of laser illumination on an inverter. When the inverter's output is '1' (Fig. 1(a)) the induced transient current (I_{gate}) surges from the drain of the NMOS (the laser-sensitive part of an inverter is the drain of its OFF transistor, where there is a reverse biased PN junction between drain and substrate). The output capacitance is then discharged by the induced transient current which reduces the output voltage resulting a change in the output value from '1' to '0' (shown as '1' $>>$ '0') subsequent to the laser shot [5]. Similarly, if the inverter's output is '0' (Fig. 1(b)) when the laser is injected, the induced transient current charges the inverter output capacitance; as a result the output toggles from '0' to '1'.

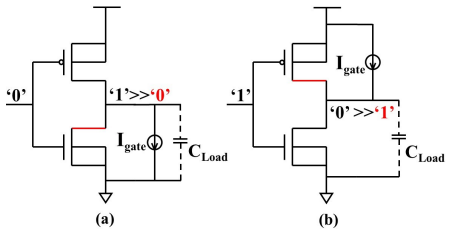


Fig. 1: Electrical model of laser-induced transient currents applied to a CMOS inverter. (a) When the output is '1' before laser illumination; (b) When the output is '0' before laser illumination [5].

In Fig. 2, I_{gate} represents the current induced due to laser illumination (here we show the case when the targeted net has the value of '1' before fault injection) which can drop the voltage of the targeted net. Thus although the adversary may target only one net of interest (an S-Box output in this

figure), such illumination also creates a perturbation in the system voltage source resulting in an IR drop in other parts of the circuit as well [1]. Such IR drop can be sensed by our sensor and the attack is detected accordingly. This will be discussed in more details in Section III.

B. Time-to-Digital Converter

Thanks to their portability among various PDK technologies, low cost calibration, and high failure rate detection, Time-to-digital converters (aka digital sensors) have been broadly used in recent years compared to their analog counterparts to sense environmental conditions such as change of voltage and temperature [6].

A digital sensor can be realized via inserting artificial critical paths (as simple as delay chains) into the chip logic such that if the chip is operated in abnormal conditions, setup time violations occur in the first place on the sensor's intentionally long paths [7]. In practice, the propagation time is not really quantified, rather it is checked if the transition manages to propagate to the end of the delay chain at the considered frequency [7]. In this paper we use such a sensor for detecting laser-based FIAs.

Figure 2 depicts the architecture of the DS used in this paper. It includes n_0 leading buffers followed by n_1 buffers each feeding to a D flip-flop (DFF). A Toggle flip-flop feeds the first leading buffer. All flip-flops operate under the same clock which is the system clock feeding the target circuit (the S-Box of the PRESENT cipher in this paper) as well. Based on the operating conditions, i.e., voltage and temperature, as well as clock frequency the setup time violation occurs in a different sampling DFF, and the sensor outcome is characterized using the so-called Average Flip-Flop Number (AFN) calculated as discussed below.

During the runtime, the sensor is fed with a continuous pulse generated by the toggle flip-flop. The pulse feeds each DFF with an image of the clock at halved frequency. Owing to the propagation delay along the buffers, DFFs 1 to i receive this periodic signal in a given phase ('0' $\rightarrow$ '1' $\rightarrow$ '0' $\rightarrow$ ...), whereas DFF $i+1$ and beyond receive the signal in opposite phase ('1' $\rightarrow$ '0' $\rightarrow$ '1' $\rightarrow$ ...). The index of the pivotal DFF is denoted FN_i [8].

AFN is the average of FN_i values in consecutive clock cycles. For example in Fig. 5, in the first 3 clock cycles the phase change occurs in the 13^{th} DFF resulting in AFN=13 (highlighted in blue). The AFN value changed to 12 after the fault injection in the third clock cycle as will be discussed below. This example confirms that by observing AFN we can detect laser-based FIAs.

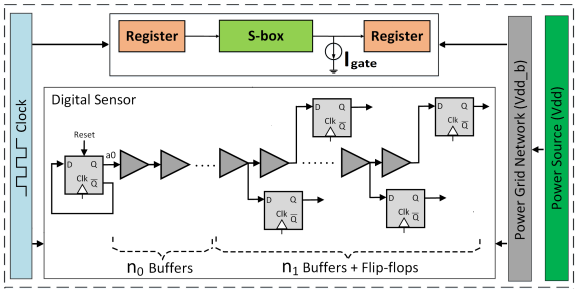


Fig. 2: The architecture of the targetted S-Box, and the fault detection digital sensor IP.

Figure 3 shows the sensitivity of our DS to voltage change. The voltage change can be due to a voltage glitch or a laser-based attack. As shown, the FN value (equal to AFN when one clock cycle is considered) is 14 when the voltage is 1V. This value decreases to 5 when the sensor is operating at 0.6V. We benefit from such sensitivity to detect the laser-based FIA.

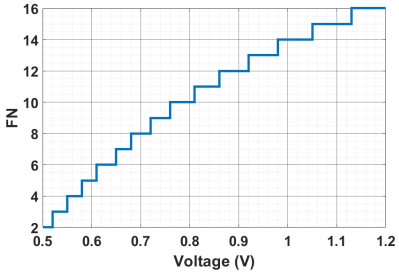


Fig. 3: Sensor output (FN) at different voltage supply values.

III. Detecting Laser-based Fault Injection Attacks

As mentioned earlier, to detect the laser-based FIA, we deploy the DS discussed in Section II. The sensor is embedded in the chip along with the targeted circuitry (S-Box module of the PRESENT cipher in this paper as shown in Fig. 2). The FN index of the sensor is changed when it is operated in different operating condition (e.g., voltage and temperature). On the other hand as mentioned earlier, as the target undergoes a laser shot, several components of currents are induced: the one that is the root cause of a fault (I_{gate} in Fig. 1), and several others adding up to induce an IR drop effect [1]. These current are mainly collected in the Psub-Nwell junctions of the target which total area is much bigger than that of the PN junction of a transistor's drain diffusion. As a result, the IR drop current is significantly bigger than I_{gate} (the laser-induced current magnitude is proportional to the area of the PN junction where it is collected [1]). Such an important IR drop current will affect the target's power supply, hence it shall be visible through variations of the FN index. Thus the idea is to monitor the value of FN in different clock cycles and raise an alarm if this index's change is beyond a threshold value.

Figure 4 illustrates the parasitic model of the power grid network (PGN) while the effect of laser illumination is modeled with the current source I_{PGN} in the PGN. Note that in case of no fault injection (i.e., no laser illumination) I_{PGN}=0. In this figure as we only show the RC model of the PGN, the induced current (modeled by current source I_{gate}) has not been shown. That current is induced in the target point of injection in the circuitry as shown in Fig. 2. In sum, in case of no fault, $I_{gate} = I_{PGN} = 0$, but when the fault is injected, we have both I_{PGN} and I_{gate} currents where the latter relates to the direct impact of laser injection and the former is the indirect impact due to the illumination of other parts of the circuit as a side effect of targeting a specific point.

In Fig. 4, the IR drop-induced voltage that each gate is fed with is shown as Vdd_b. The Vdd_b value in the absence and presence of FIA can be extracted via Eq. 1. Indeed the difference of the Vdd_b before and after fault injection is what affects our sensor outcome and results in raising an alarm when the fault is detected.

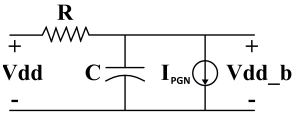


Fig. 4: RC circuitry modeling the laser-induced IR drop.

$$
\begin{aligned}
Vdd_b &= Vdd(1 - e^{-\frac{t}{R \times C}}) \approx Vdd \\
Vdd_{b(faulty)} &= (Vdd - R \times I_{PGN})(1 - e^{-\frac{t}{R \times C}}) \\
&\approx Vdd - R \times I_{PGN}.
\end{aligned} \tag{1}
$$

In practice, based on Eq. 1, $R \times I_{PGN}$ is the magnitude of the laser-induced voltage drop. This drop results in the decrease of the FN index (as the voltage is decreased). In our fault detection method, we monitor the difference between FN of the current clock cycle (FN_i) with that of previous clock cycle (FN_{i-1}). In case of observing any changes in FN, an alarm is raised. However in order to decrease the rate of false alarms, instead of using FN_{i-1}, we use the average value of FN indexes (i.e., AFN) in the last 8 clock cycles:

$$
AFN_{i-1} = \frac{1}{8} \sum_{j=i-9}^{i-1} FN_j. \tag{2}
$$

Finally in each clock cycle C_i, an alarm is raised if Eq. 3 is met.

$$
Alarm = \begin{cases} \text{'1'} & \text{when } \lceil FN_i - AFN_{i-1} \rceil \geq 1, \\ \text{'0'} & \text{otherwise .} \end{cases} \tag{3}
$$

IV. Experimental Results and Discussions

A. Experimental Setup

The sensor-integrated system targeted in this paper is the S-Box module of the PRESENT cipher [9]. The sensor and the S-Box modules were implemented at transistor level using the 45nm NANGATE CMOS technology. Synopsys HSpice was used for the simulations. Our sensor includes n_0=6 leading buffers and n_1=35 sampling flip-flops and related buffers. The sensor dimensioning was performed based on [2] assuming the circuit is supposed to work under $Vdd \in$[0.65V, 1.4V] and the temperature in [-10°C,150°C] range.

The results are reported for (V_{dd},T)=(1V, 85°C). The simulations were conducted for a range of R, C values ($R \in [1\Omega, 100\Omega], C \in [100fF, 1000fF]$) to represent different intensity of the injected faults for different sizing of the PGN. We assume I_{PGN}=1mA and I_{gate}=150uA. Indeed, as mentioned in section III the IR drop induced current is significantly greater than I_{gate}. Thus, based on [1], we considered a case with a moderate ratio between I_{PGN} and I_{gate} to prove our sensor detection capability in a worst case. In fact, I_{PGN} current is defined as a $factor \times I_{gate}$ related to the area of Nwells and the area of transistors drains. It is possible to compute $factor$ to be applied to each instance by analyzing layout files such as .lef and netlist files that contain information regarding each available cell. This allows to estimate the area of the affected PN junction of a particular transistor's drain as well as the area occupied by the Nwell, and thus deduce the value of $factor$.

As mentioned in Section II-A, based on the value of the output during the laser-based FIA, a current is induced between Vdd and the output, or between the output and ground. Without loss of generality, in our experiments we targeted the Least Significate Bit (LSB) of the S-Box module for our fault injection while the S-Box is fed with the input data that results in the value of '1' in its LSB output in case of no-fault. Thus in our simulation we considered the case shown in Fig. 1(a). The results for the cases where the S-Box golden LSB output is '0' follows the similar trend. Note that injecting faults in other S-Box output bits gives a very similar results as well.

B. Experimental Results and Discussion

1) The effect of laser illumination on the S-Box and Sensor outcomes: The first set of results shown in Fig. 5 illustrates the impact of laser illumination in our setup. As depicted, Vdd_b is experiencing a drop during fault injection, resulting in an erroneous S-Box output (Y_0 changes from '1' to '0'). As depicted, these changes affect our sensor's outcome, i.e., its FN index changes from 13 to 12 during the fault injection time window (shown in red on signal Q12).

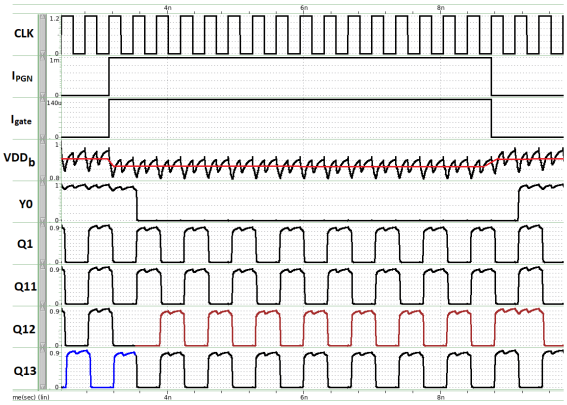


Fig. 5: Sensor and S-Box signals for $R = 50\Omega$ and $C = 500fF$. Y_0 is the LSB of the S-Box. FN changes from 13 (shown in blue) to 12 (shown in red) after laser illumination.

Figure 6 shows the value of Vdd_b for different values of R and C when no fault is injected, i.e., $I_{gate} = I_{PGN} = 0$. As shown, Vdd_b only experiences a slight change from $Vdd = 1V$ due to the IR drop related to the RC circuitry of PGN. As expected, the change of Vdd_b is more significant in case of laser illumination. For the sake of clarity, instead of Vdd_b values, in Fig. 7, we show the IR drop magnitude ($Vdd - Vdd_b$) in different combinations of (R, C) when the laser fault is injected for 1.5 and 16 clock cycles. Comparing Fig. 6 and Fig. 7 depicts that the laser illumination forces a drop greater than $\approx 0.1V$ for higher value of resistance. The drop is at most $0.08V$ (i.e., Vdd_b=0.92) when no fault is injected. Note that very few overshooting may occur in both cases resulted in a Vdd_b exceeding Vdd value.

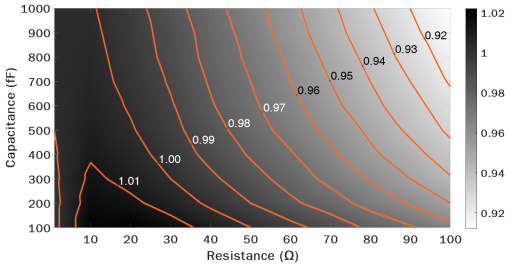
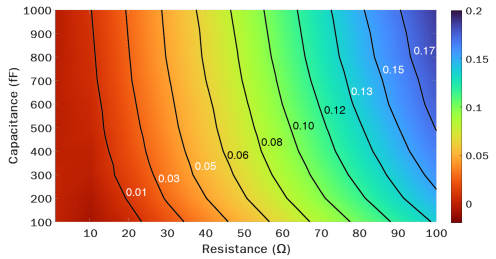
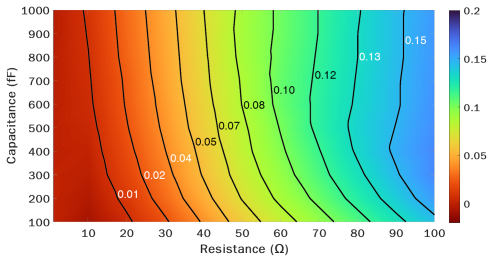


Fig. 6: The heatmap shows the value of Vdd_b in different (R, C) combinations when no fault is injected.

Note that in both cases (with or without fault injection) the circuit experiences a small IR drop related to the PGN (modeled with the RC circuitry shown in Fig. 4), yet I_{PGN} is only induced in case of a fault injection and its value would be '0' otherwise. Thereby, as shown in Fig. 7, the laser illumination increases the magnitude of IR drop. Such decrease of Vdd_b when the injected fault is sensed by our DS and contributes to detecting the erroneous S-Box outputs in most of the cases (this will be discussed by the next set of results).



(a) Laser illumination for 16 clock cycles.

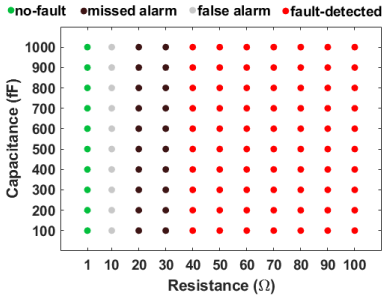


(b) Laser illumination for 1.5 clock cycles.

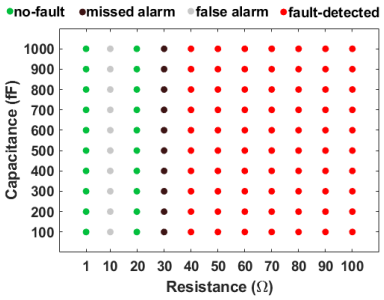
Fig. 7: The heatmaps show the value of voltage drop in Vdd_b (i.e.,$Vdd - max(Vdd_b)$) in different (R, C) combinations.

2) Detecting Laser-based FIA using the embedded DS: This set of results depicts the sensor outcome in different R and C combinations when laser illuminates the S-Box output. Figure 8(a) and Fig. 8(b) show the cases where laser is injected for 16 and 1.5 clock cycles, respectively. As shown in both cases the S-Box output would be faulty for $R \geq 30\Omega$. In the scenario of injecting fault for 1.5 clock cycles, for over 87% of all cases when the fault occurs (i.e., the attack is strong enough to change the output), our sensor can accurately detect the fault. When the fault injected for 16 clock cycle, the sensor is able to detect the error in 78% of all cases it occured. We have some false alarms for lower R values in both cases as in those cases the amount of IR drop was not significant to change the S-Box output, yet sensor could sense it. Note that our sensor doesn't result in any false alarms when there is no fault, i.e., no alarm is raised for any of the (R, C) combinations when there is no FIA (not shown for the sake of space).

Fig. 9 shows the sensitivity of the DS to the voltage drop, ΔVdd_b. It shows the average value for the required change in each value of Vdd_b that can be sensed by the sensor to raise an alarm. For example, when the sensor is operating at $Vdd_b = 0.95V$, a voltage drop of 0.045V is required to generate an alarm. This figure can clarify the reason of "missed alarms" when $R = 20, 30\Omega$ in Fig. 8(a). Indeed, when $R = 20, 30\Omega$, the Vdd_b on average would be 0.945V and 0.962V, respectively. Thus according to Fig. 9, a voltage drop of $\Delta Vdd_b > 0.04V$ is required to raise an alarm in this case (This area is shown in red in Fig. 9). However, the laser illumination is able to generate at most 0.037V drop in this range of R. Thus, the sensor cannot raise an alarm and a missed-alarm happens for these values of R. The same reasoning can be considered for the missed alarms shown in Fig. 8(b). For the sake of space we did not show the average values of Vdd_b for all (R,C) combinations and Fig. 7 depicts the maximum values of voltage drop, not average, so cannot be used as a reference for discussing the sensitivity.



(a) Laser illumination for 16 clock cycles.



(b) Laser illumination for 1.5 clock cycles.

Fig. 8: The sensor outcome in case of laser illumination.

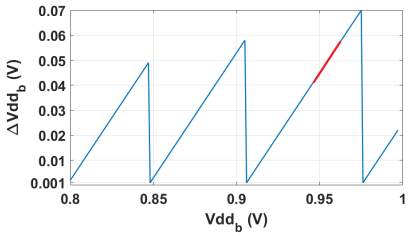


Fig. 9: Voltage sensitivity of the sensor to raise an alarm in different Vdd_b

V. Conclusion and Future Directions

An adversary can benefit from injecting laser shots to the targeted circuitries to leak sensitive data or impose circuit malfunction. Thus, detecting such attacks is of utmost importance. In this paper we deployed digital sensors to detect such faults by sensing the IR drop caused by laser illumination. Our experimental results show the high accuracy of the proposed method in detecting laser-induced fault injections. In the continuation of this study, we will consider the impact of device aging on the success of the attacks, and also on the accuracy of our sensor-based fault detection scheme. We will also investigate the impact of temperature on the proposed detection approach.

References

[1] R. A. C. Viera et al., "Simulation and Experimental Demonstration of the Importance of IR-Drops During Laser Fault Injection," *TCAD*, vol. 39, no. 6, pp. 1231–1244, 2020.

[2] M. T. H. Anik et al., "Detecting failures and attacks via digital sensors," *TCAD*, vol. 40, no. 7, pp. 1315–1326, 2021.

[3] D. Shahrjerdi et al., "Shielding and securing integrated circuits with sensors," in *ICCAD*, 2014, pp. 170–174.

[4] A. H. Johnston, "Charge generation and collection in p-n junctions excited with pulsed infrared lasers," *IEEE Trans. on Nuclear Science*, vol. 40, no. 6, pp. 1694–1702, 1993.

[5] R. A. Camponogara-Viera, "Simulating and modeling the effects of laser fault injection on integrated circuits," Theses, Université Montpellier, Oct. 2018. [Online]. Available: https://tel.archives-ouvertes.fr/tel-02150306

[6] N. Selmane et al., "Security evaluation of Application-Specific Integrated Circuits and Field Programmable Gate Arrays against setup time violation attacks," *IET Information Security*, vol. 5, no. 4, pp. 181–190, 2011, DOI: 10.1049/iet-ifs.2010.0238.

[7] M. Ebrahimabadi et al., "Using digital sensors to leverage chips' security," in *PAINE*, December 15-16 2020, pp. 1–6, DOI: 10.1109/PAINE49178.2020.9337730.

[8] M. T. H. Anik et al., "Reducing aging impacts in digital sensors via run-time calibration," in *JETTA*, 2021, pp. 653–673.

[9] ISO/IEC 29192-2:2012, "Information technology – Security techniques – Lightweight cryptography – Part 2: Block ciphers," p. 41, Publication date: 2012-01, Edition: 1. https://www.iso.org/standard/56552.html.