Reproducible Single-Byte Laser Fault Injection
Jean-Max Dutertre, Amir-Pasha Mirbaha, David Naccache, Assia Tria

To cite this version:
Reproducible Single-Byte Laser Fault Injection

Jean-Max Dutertre*, Amir-Pasha Mirbaha*, David Naccache† and Assia Tria‡

*Département Systèmes et Architectures Sécurisées (SAS)
†École Nationale Supérieure des Mines de Saint-Étienne (ENSMSE), Gardanne, France
dutertre, mirbaha}@emse.fr
‡Équipe de cryptographie, École normale supérieure, Paris, France
david.naccache@ens.fr
‡CEA-LETI, Gardanne, France
assia.tria@cea.fr

Abstract—This note describes laser fault experiments on an 8-bit 0.35µm microcontroller with no countermeasures. We show that reproducible single-byte faults, often considered unfeasible, can be obtained by careful beam-size and shot-instant tuning.

I. INTRODUCTION

Fault attacks consist in using hardware malfunction to infer secrets from the target’s faulty outputs. Within fault attacks, Differential Fault Analysis [2] (DFA) is a particular analysis technique exploiting differences between correct and faulty outputs. We refer the reader to [13] for more information on fault injection techniques.

This note describes laser faults experiments on an 8-bit 0.35µm RISC microcontroller with no countermeasures. We show that reproducible single-byte faults, often considered unfeasible, can be obtained by careful beam-size and shot-instant tuning. Moreover, we obtain such faults even when the beam’s impact area exceeds a single SRAM cell. This underlines the need to protect small data objects, such as pointers, counters or flags, against “surgical” faults targeting a specific byte in memory and nothing else.

II. THE ADVANCED ENCRYPTION STANDARD

We assume that the reader is familiar with the AES [10] that we recall here for the ease of reference.

The AES-128 encrypts 128-bit plaintexts into 128-bit ciphertexts using a 128-bit key K. The algorithm performs 10 rounds (after a short initial round) and consists of two separated processes: a key schedule that derives round keys and the encryption routine itself. During decryption key schedule is reversed and encryption is replaced by a very similar decryption process.

The initial round uses $K_0 = K$ as a round key; for all subsequent rounds, new round keys $K_i$ are calculated from their predecessors $K_{i-1}$. Figure 1 illustrates the AES’ structure.

In most implementations the $K_i$s are computed and stored in memory before encryption starts. Encryption treats the 16-byte plaintext $M$ as a $4 \times 4$ byte matrix. Each round, except the first and the last, includes four steps: A substitution of the matrix’s contents using a lookup table (SubBytes), a rotation of the matrix’s rows (ShiftRows) and a linear transform in GF($2^8$) (MixColumns) combining each matrix element with other column elements weighted by different coefficients (1, 2 or 3). At the end of each round $K_i$ is xored with MixColumns’s result (an operation called AddRoundKey).

III. LASER FAULT INJECTION

Laser (Light Amplification by Stimulated Emission of Radiation) is a stimulated-emission electromagnetic radiation in the visible or the invisible domain. Laser light is monochromatic, unidirectional, coherent and artificial (i.e. laser does not spontaneously exist in nature). Laser light can be generated as a beam of very small diameter (a few µm). The beam can pass through various material obstacles before impacting a target during a very short duration.

Laser impacts on electronic circuits are known to alter functioning. In particular, SRAM (Static Random Access Memory) laser exposure is known to cause bit-flips [12], [6], [5], a phenomenon called Single Event Upset (SEU). By tuning the
Differential Fault Analysis [2] (DFA) is an analysis technique exploiting differences between correct and faulty outputs. Several byte-level and bit-level AES DFA variants exist (e.g. [8], [9], [7], [3]). Given the dependence of these attacks on precise “surgical” fault injection, the feasibility of bit/byte-level DFA remained somewhat unclear.

[11] describes a byte-level DFA on AES (the so-called Piret-Quisquater’s DFA). This attack requires the injection of a single-byte fault into the temporary ciphertext between the MixColumn exit of the antepenultimate round and the MixColumn input of the penultimate round to be successful. A means to meet this requirement is to inject a single-byte fault into the antepenultimate round key (namely $K_8$). As a consequence, a faulty ciphertext with four faulty bytes is obtained (see Figure 2). Then, the attack scheme described in [11] allows to infer some informations on the four corresponding bytes of $K_{10}$ by processing the correct and faulty ciphertexts and checking over the list of all the related possible single-byte faults.

By repeating this process twice (i.e. by iterating the attack for a different plaintext) the exact value of the four bytes of $K_{10}$ is found with a success rate of about 98% [11]. The procedure is repeated to target $K_{10}$‘s remaining bytes. Finally, $K = K_0$ is inferred by reversing the key schedule. We show that this attack can be implemented, even when the laser spot is wider than the SRAM’s cell.

V. Practical Single-Byte Fault Injection

Outline: After chip decapsulation and a mapping of the chip’s components, we selected a large target area, given our knowledge of the implementation. Using automated search on the chip’s front-side, we modified the impact’s coordinates, the beam parameters and timing until a one byte fault was obtained. Finally, Piret-Quisquater’s attack was performed.

The target is an 8-bit 0.35 $\mu$m 16 MHz RISC microcontroller with an integrated 4KB SRAM and no countermeasures. The device runs SOSSE (Simple Operating System for Smartcard Education [4]) to which we added some commands, most notably for feeding-in cleartexts and retrieving ciphertexts. $K$ was embedded in the code. As encryption starts, the $K_8$s are derived and stored in SRAM. The laser, shown in Figures 7 and 8, is equipped with a YAG laser emitter in three different wavelengths: green, infrared and ultraviolet.

The spot’s diameter can be set between 0 and 2500$\mu$m. As the beam passes through a lens, it gets reduced by the lens’ zoom factor and loses a big part of its energy. Our experiments were conducted with a 20$\times$ Mitutoyo lens, a green$^2$ beam of 0.4$\mu$m and $\approx 15pJ$ per shot. The circuit is installed on a programmable Prior Scientific X-Y positioning table$^3$. The X-Y table, card reader, laser and an FPGA trigger board, were connected via RS-232 to a control PC. The FPGA trigger board receives an activation signal from the reader and sends a trigger signal to the laser after a delay defined by the control PC.

Experiments were conducted in ambient temperature and at $V_{CC} = 5V$. These parameters are within the device’s normal operating conditions $2.7V \leq V_{CC} \leq 5.5V$.

The chip was decapsulated by chemical etching using a Nisene JetEtch automated acid decapsulator. The decapsulator can be programmed for the chemical opening of different chip types using different ratios of nitric acid (HNO$_3$) and sulfuric acid (H$_2$SO$_4$), at a desired temperature and during a specified time. For opening our chip, we used only nitric acid at 80°C for 40 seconds. The etched chip (Figure 3) successfully passed functional tests before and during fault injection.

As it is very difficult to target the chip’s (ALU) (Arithmetic Logic Unit) and inject only a single-byte fault during a very specific instant between the end of MixColumn of the 8-th round and before the MixColumn of the 9-th round, we decided to target $K_8$.

Finding the SRAM area containing $K_8$ and properly tuning the laser’s parameters is very time consuming. The number of

\footnote{2532nm wavelength.}

\footnote{Motorized stepper stage for upright microscopes with 0.1 $\mu$m steps.}
faults in the ciphertext ($C$), their position and their contents indicate which round key has been hit. *MixColumns* will amplify any single-byte fault occurring in any $K_i$ preceding $K_8$ and result in a completely faulty ciphertext (we call such a bad event an "early fault"). As shown in Figure 2, a single-byte fault on $K_8$ changes 4 ciphertext bytes while a fault in $K_9$ or $K_{10}$ changes only one byte. However, injected faults are not always limited to a single byte and/or to a single $K_i$.

When more than 4 ciphertext bytes are faulty, it is difficult to determine if the observed result is due to an early fault or to several faults in $K_8$, $K_9$ and $K_{10}$ (Table I). After finding sensitive SRAM areas that affect the ciphertext, and before targeting $K_8$, we tried to find the memory cells containing $K_{10}$. This is necessary for tuning beam-size and energy to limit the fault injection area to very few bytes and preferably one byte. Despite fine-grained energy and spatial control we detected faults in keys neighboring $K_8$. To overcome this problem, we isolated $K_9$ and $K_{10}$ from faults and restricted the shot to a 100 $\mu$s interval between the use of $K_7$ and $K_8$ (Figure 5).

Figure 6 shows how we could confine faults to a single byte of $K_9$. When more than one faulty byte existed, we could obtain single byte faults by controlling the laser’s shooting time. Figure 6 is just a model of the real SRAM (Figure 4) to describe our technique and does not correspond to real address allocation. We could successfully inject faults into 13 bytes of $K_8$. This sufficed to run Piret & Quisquater attack.

As shown in the topmost part of Figure 6, we searched $K_8$’s precise storage area by monitoring the number and the type of faults in the ciphertext. Then (middle part of Figure 6), by a precise beam localization, we managed to inject faults only in $K_8$. This, however, did not turn out to be fully deterministic as sometimes we would also inflict faults to previous round keys. At that point (lowermost part of Figure 6), by fine-tuning spatial and temporal beam localization (just after the use of
Fig. 8. Laser and target (closeup).

We managed to restrict the injected faults only to $K_8$. This is the exact assumption of Piret-Quisquater’s scenario.

VI. CONCLUSION

We implemented Piret-Quisquater’s attack [11] using laser fault injection. [11] is usually regarded as the most effective fault attack on AES as it requires only two faulty ciphertexts. This effectiveness comes at the price of stringent fault injection assumptions. The experiments reported in this paper also apply to other attacks (e.g. [8], [9]) and underline the possibility to precisely modify flags, counters, pointers and other single memory cells that control program flow, in the absence of countermeasures.

In summary, this note’s main conclusions are:

- It is possible to implement a single-byte laser fault attack on an AES round key.
- Even when it is physically impossible to target a single-byte because the beam hits a few other bytes, careful spatial and temporal coordination may allow to deceive the encryption process to consider logically only a single-byte fault that corresponds to Piret-Quisquater’s scheme.
- It is possible to reproduce the same faults on different plaintexts. This assesses the reality of Piret-Quisquater’s scenario on unprotected chips.

As we send this paper to press we can already refer the reader to recent reproducible single-bit fault injection results [1].

REFERENCES