COSADE 2015

A Biased Fault Attack on the Time Redundancy Countermeasure for AES

Sikhar Patranabis, Abhishek Chakraborty, Phuong Ha Nguyen and Debdeep Mukhopadhyay

Department of Computer Science and Engineering,

Indian Institute of Technology Kharagpur

Outline

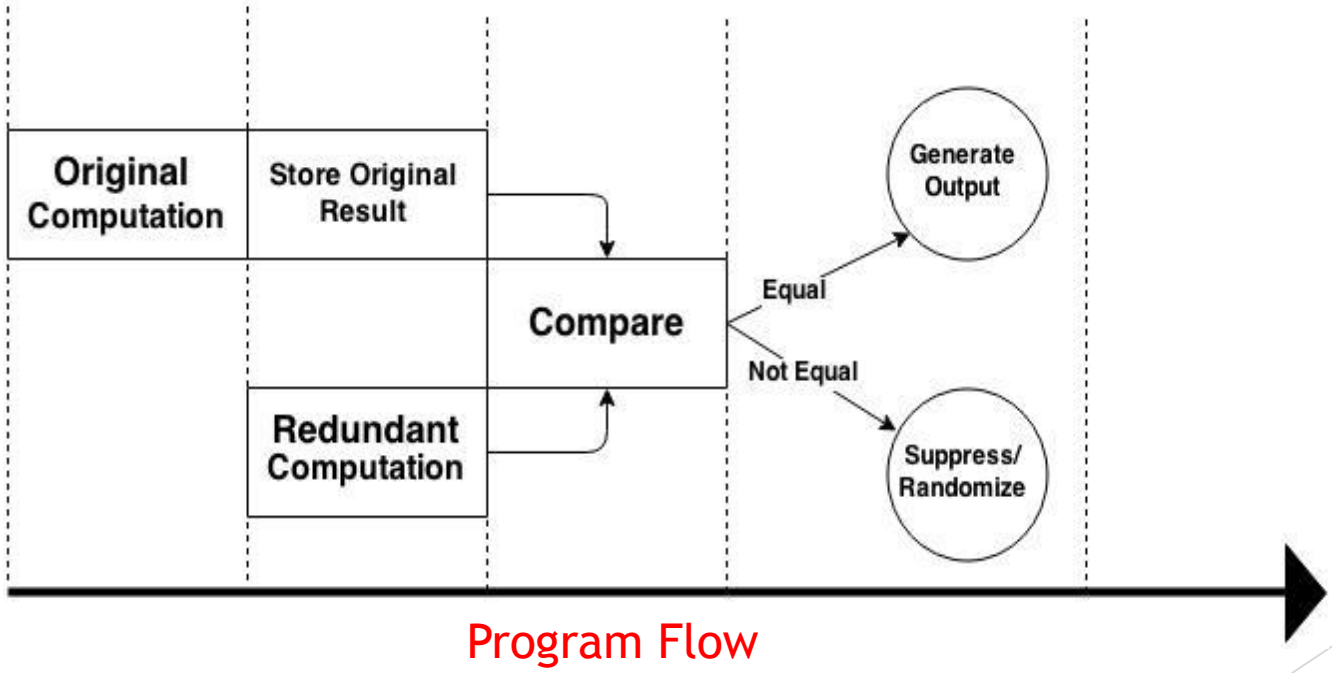
- Objectives
- The Time Redundancy Countermeasure
- Bias Quantification and Adversarial strategy
- Fault Model and Fault Injection Set-Up
- Performed Attacks
- Simulation Studies
- Experimental Results
- Conclusions

Objectives

1. To develop a formulation for the degree of bias in a fault model
2. Propose biased fault models to attack the time redundancy countermeasure for AES-128
3. Establish the feasibility of the proposed attacks via simulations and real life experiments

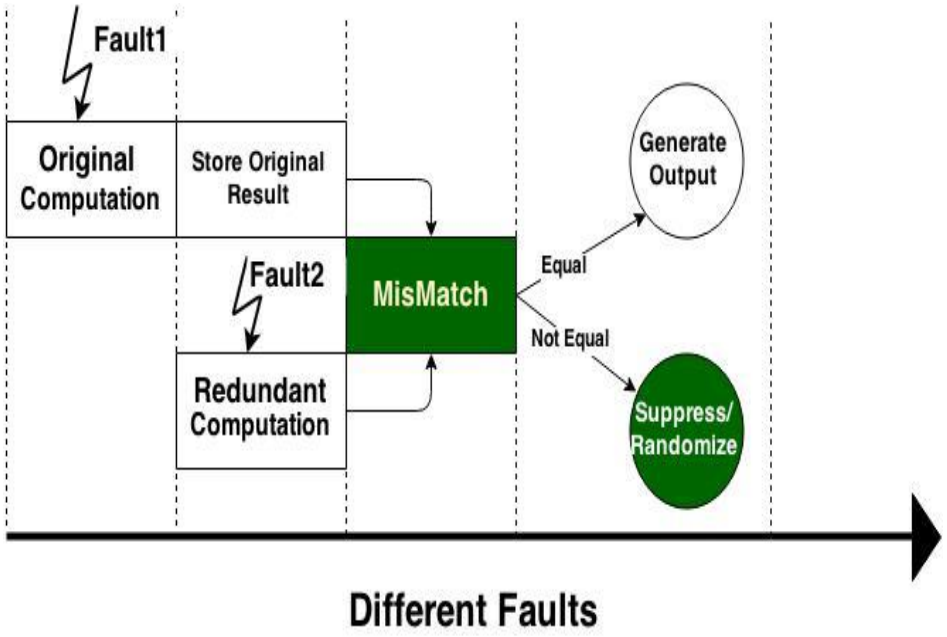
Introduction: Time Redundancy

- Time Redundancy - A Classical Fault Tolerance Technique
  - Each operation is followed by a redundant operation and outputs are matched
  - Output is suppressed or randomized in case of a mismatch

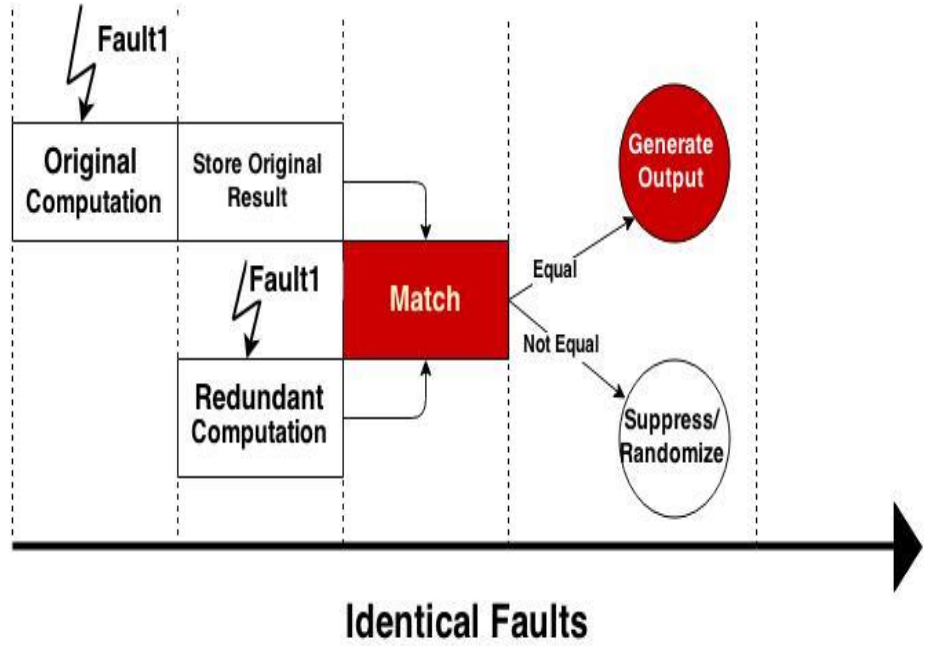


Against Fault Attacks : Detection

Success



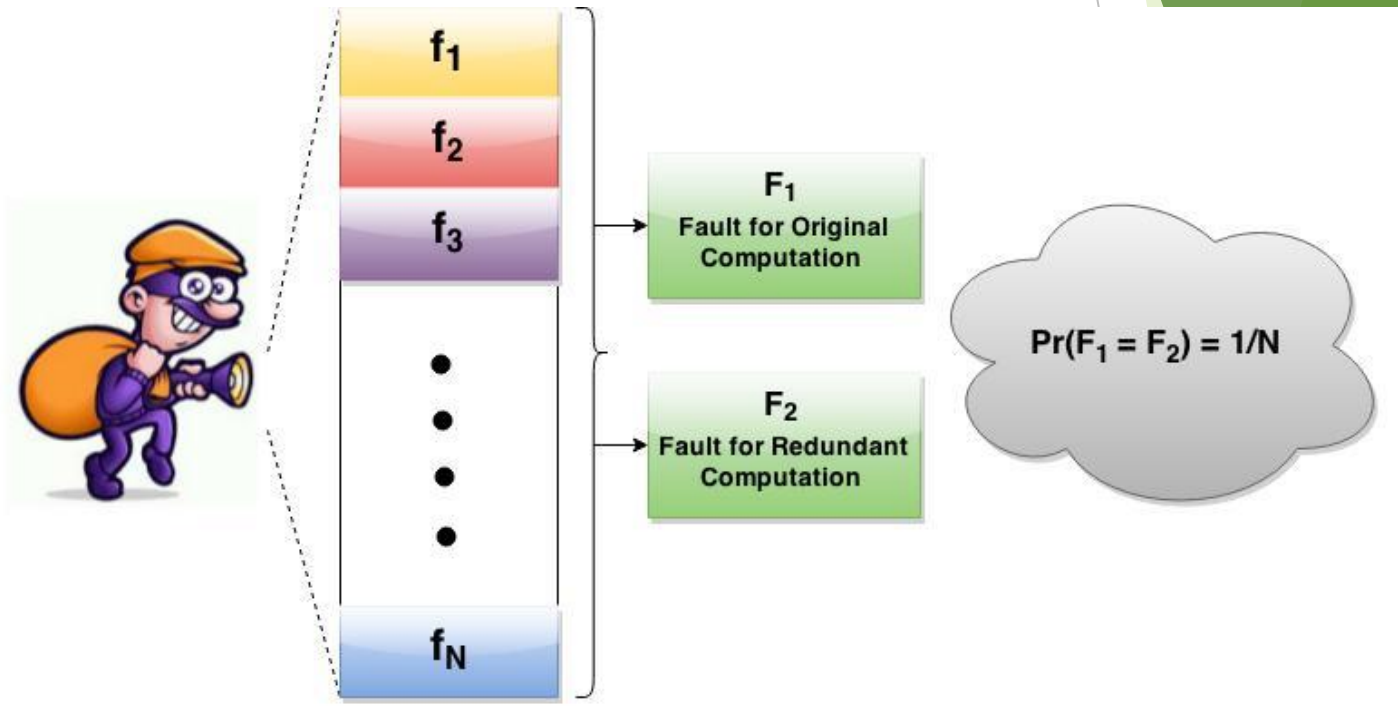
Failure



Identical Faults in both rounds goes undetected

Uniform Fault Model

- All faults under the fault model are equally likely
- Fault collision probability for two random fault injections is low
- Large number of fault injections necessary to get a *useful ciphertext*



Beating the Countermeasure

- Improving fault collision probability
  - Enhancing the probability of identical faults in original and redundant rounds
- Two major aspects
  - The size of the fault space
  - The probability distribution of faults in the fault space
- A smaller fault space enhances the fault collision probability
- A non-uniform probability distribution of faults in the fault space also enhances the fault collision probability

Biased Fault Model

- Different faults have unequal probability of occurrence
- Biasness of the fault model can be quantified by the variance of fault probability distribution
- Higher the variance, higher is the degree of bias of the fault model

A Hypothetical Fault Model

Fault	1	2	3	4	5	6	7	8
Probability	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125

Variance = 0

Fault	1	2	3	4	5	6	7	8
Probability	0.225	0.175	0.200	0.125	0.100	0.075	0.050	0.050

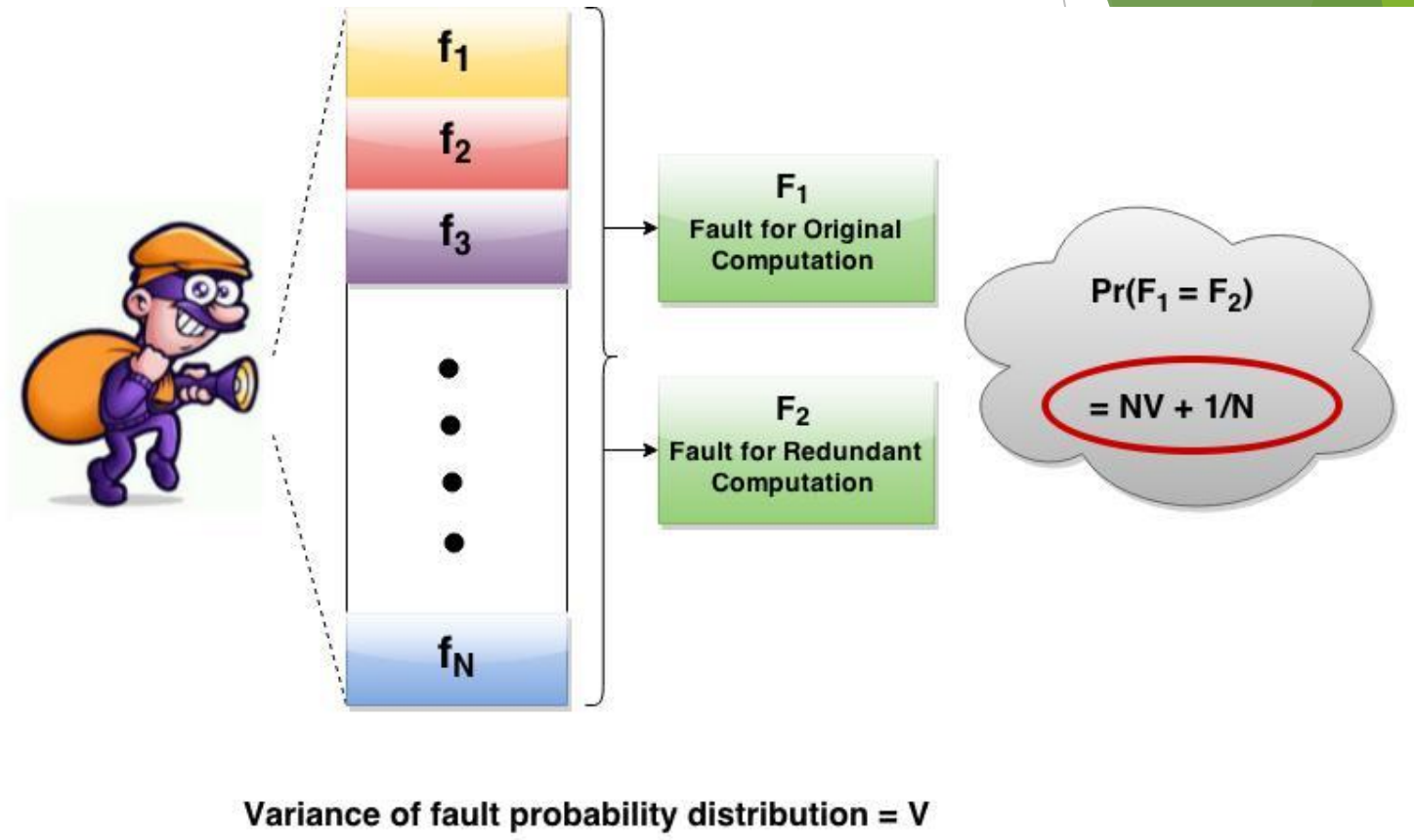
Variance = 0.004

Fault	1	2	3	4	5	6	7	8
Probability	0.5	0.25	0.125	0.05	0.05	0.025	0	0

Variance = 0.026

The Fault Collision Probability

- With increase in variance, the fault collision probability increases
- Requires fewer number of fault injections per *useful ciphertext*



Long Story Short

The Adversarial Perspective

Precise Fault Models

Biased Fault Models

Yes!!
It is practically feasible

Proposed Fault Model

- All faults are restricted to a single byte
- Two kinds of fault models
  - Situation-1: Attacker has control over target byte
  - Situation-2: Attacker has no control over target byte
- Control over target byte makes fault model more precise but is costly to achieve

Fault Classification

Suitable

Symbol	Fault Model
FF	Fault Free
SBU	Single Bit Upset
SBDBU	Single Byte Double Bit Upset
SBTBU	Single Byte Triple Bit Upset
SBQBU	Single Byte Quadruple Bit Upset
OSB	Other Single Byte Faults
MB	Multiple Byte Faults

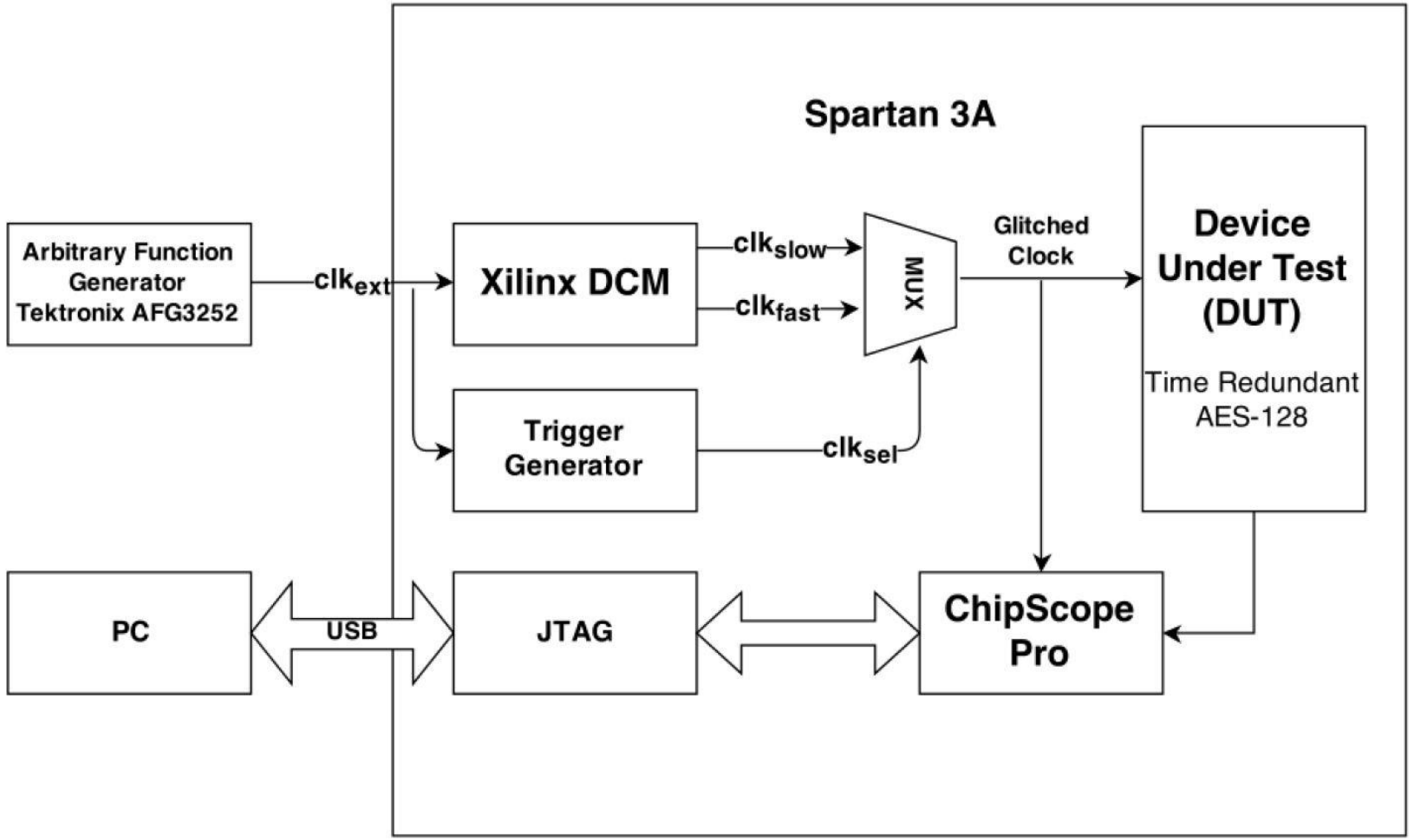
Fault Precision

Fault Model	Faults Possible(n) (Situation-1)	Faults Possible(n) (Situation-2)
SBU	8	128
SBDBU	28	448
SBTBU	56	896
SBQBU	70	1120
OSB	93	1488

Fault Injection Set-Up

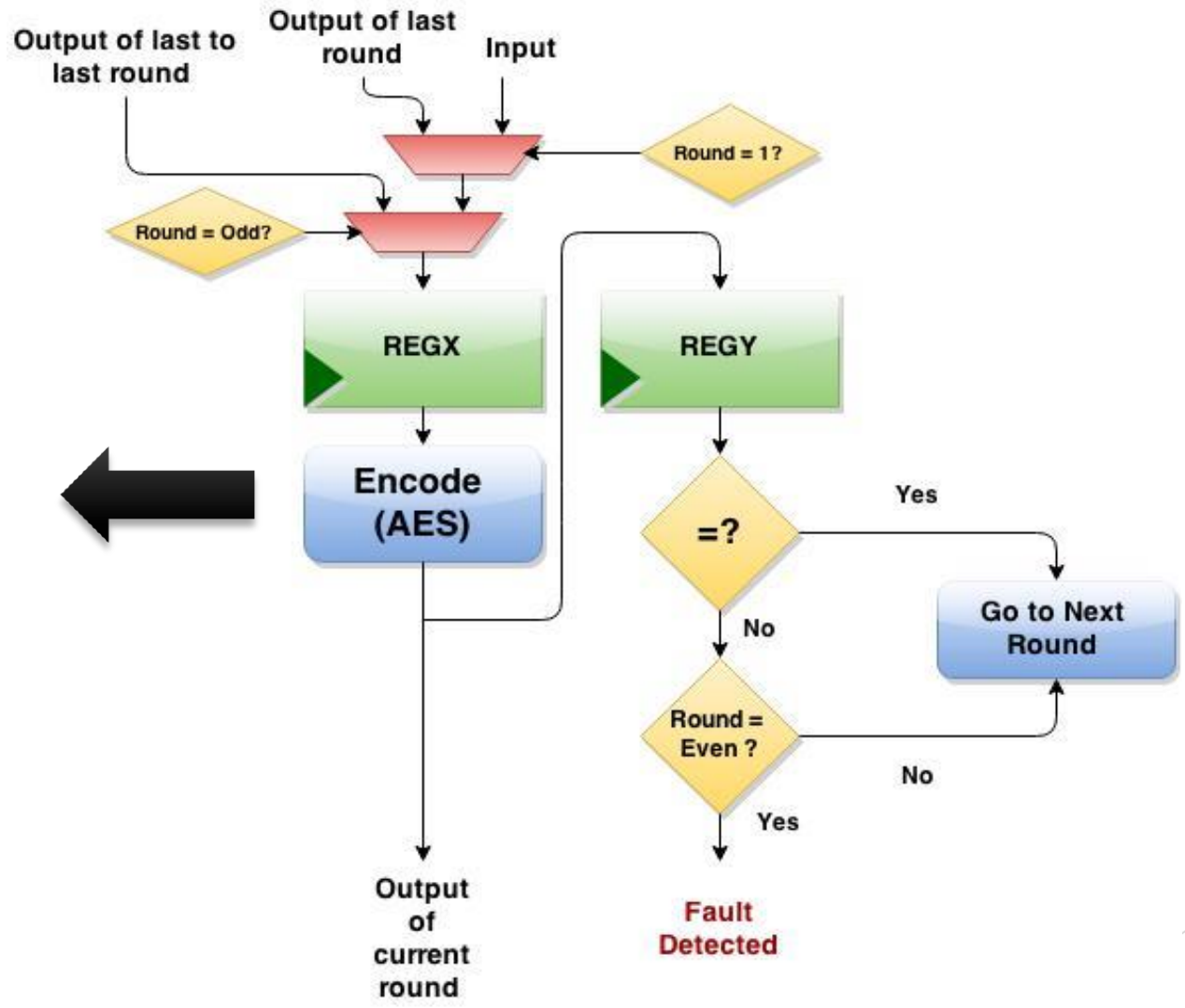
- Time redundant AES-128 implemented in Spartan 3A FPGA
- Fault injection using clock glitches at various frequencies
- Xilinx DCM to drive fast clock frequency
- Internal state monitoring using ChipScope Pro 12.3

Fault Injection Technique



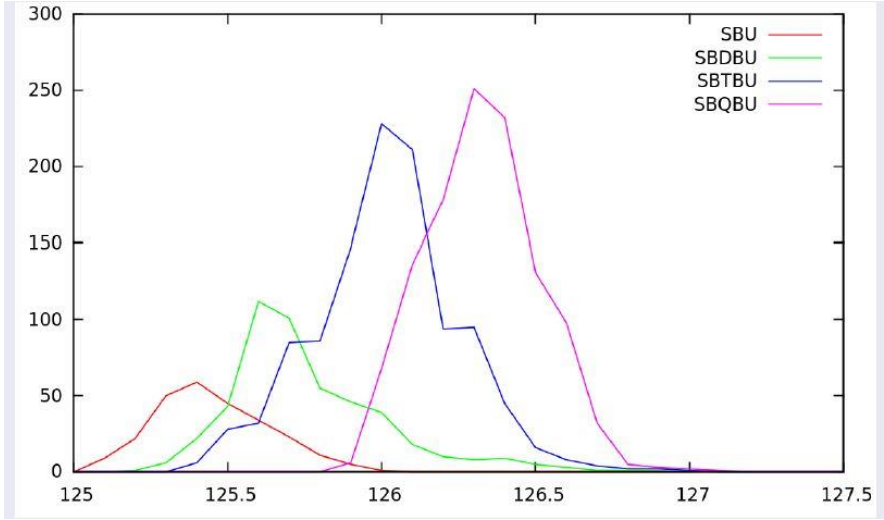
Time Redundant AES-128

AES-128 Encryption Module



Fault Distribution Patterns

Distribution pattern checked over
512 random fault injections



Frequency Ranges

Fault Model	Frequency range for both original and redundant rounds (MHz)
SBU	125.3-125.4
SBDBU	125.6-125.7
SBTBU	126.0-126.1
SBQBU	126.3-126.4

Attack Procedure

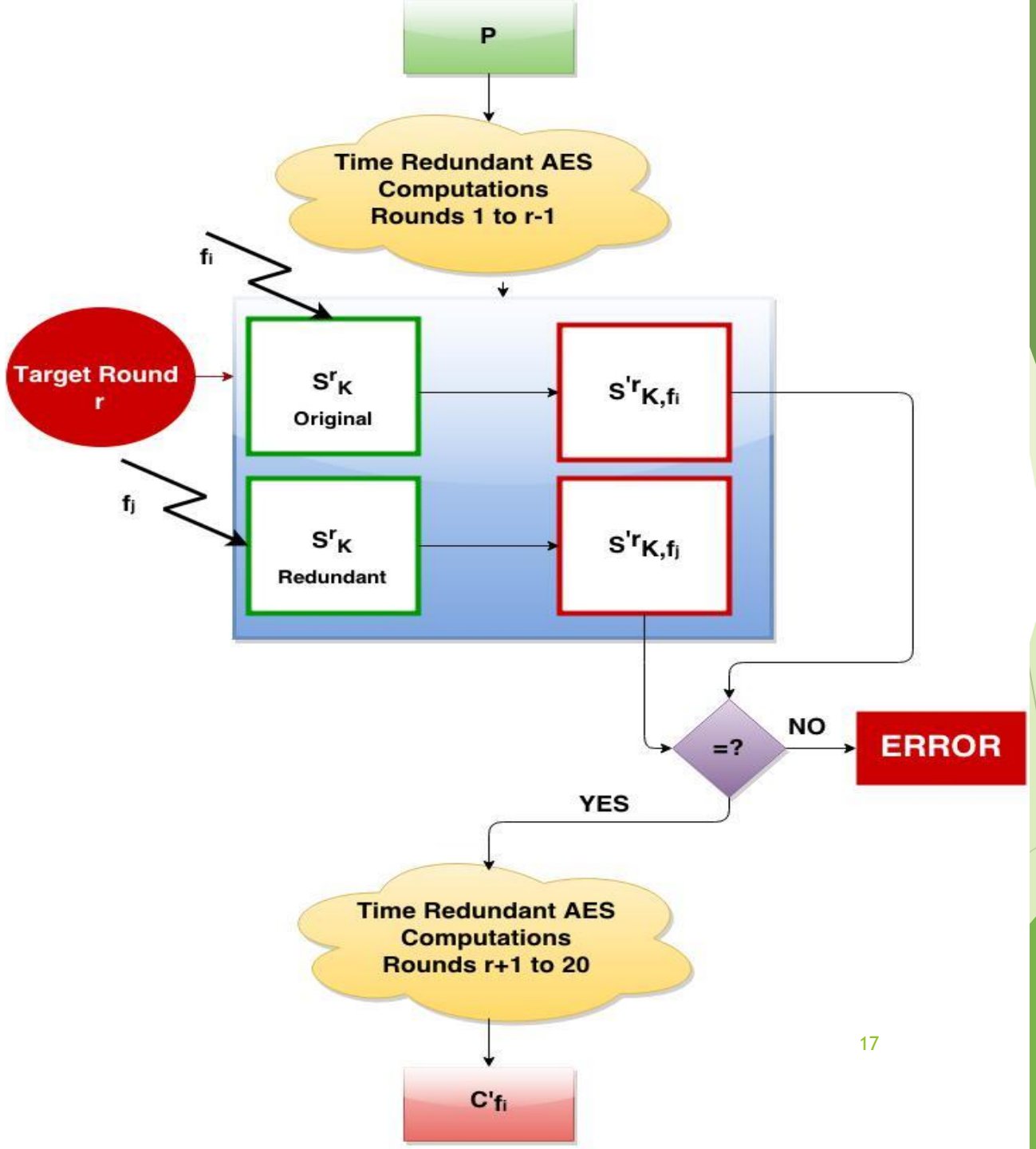
Notations

P	Plaintext
C	Fault-free ciphertext
f_i	A specific fault instance
n	The number of possible faults under the fault model
N_C	The total number of faulty ciphertexts obtained (excluding random ciphertexts generated by the countermeasure)
N_F	The total number of fault injections
C'_{f_i}	The faulty ciphertext under fault f_i
r	A round of AES
k	A key hypothesis
K	The correct key
$S^r{}_K$	The fault free cipher state in round r for key K
$S'^r{}_{k,f_i}$	A guess for the faulty cipher state before the SubBytes of round r under fault f_i and key hypothesis k

Attack Procedure

Fault Injection

Useful Ciphertext obtained if $f_i = f_j$



Attack Procedure

Requires Only Faulty Ciphertexts

- Distinguishers used :
  - Hamming Distance (HD)
  - Squared Euclidean Imbalance (SEI)
- Make a key hypothesis k and evaluate the distinguishers
- Correct hypothesis gives minimum and maximum values respectively

$$H(k) = \sum_{i=1}^{N_C} \sum_{j=1}^{i-1} HD(S'^{r}{}_{k,f_i}, S'^{r}{}_{k,f_j})$$

SEI

$$S(k) = \sum_{i=1}^{N_C} \sum_{\delta=0}^{255} \left(\frac{\#\{b \mid S'^{r}{}_{k,f_i}[b] = \delta\}}{N_C} - \frac{1}{256} \right)^2$$

- Differential Fault Intensity Analysis, Ghalaty et. al., FDTC 2014
- Fault Attacks on AES with Faulty Ciphertexts Only, Fuhr et. al., FDTC 2013

Attack Procedure

Target Rounds

- Round 9 (Rounds 17 and 18 of time redundant AES)
  - Fault is injected before the SubBytes operation of round 9

$$S'^{9}{}_{K,f_i} = SB^{-1}(SR^{-1}(C'_{f_i} \oplus K_{10}))$$

  - Hypothesize on one byte of K_{10} at a time
- Round 8 (Rounds 15 and 16 of time redundant AES)
  - Fault is injected before the SubBytes operation of round 8

$$S'^{8}{}_{K,f_i} = SB^{-1}(SR^{-1}((MC^{-1}((SB^{-1}(SR^{-1}(C'_{f_i} \oplus K_{10})) \oplus K_9))))$$

  - Hypothesize on 4 bytes of K_{10} and one byte of K_9 at a time
- Beyond Round 8 attacks on time redundant AES become infeasible as very large number of fault injections are required

Simulation : Part-1

- Identical faults introduced into both original and redundant rounds
- Target byte chosen at random
  - Same fault for original and redundant computations
  - Each fault injection yields a *useful ciphertext*
- Attacks simulated on rounds 8 and 9
- Performed separately for each fault model

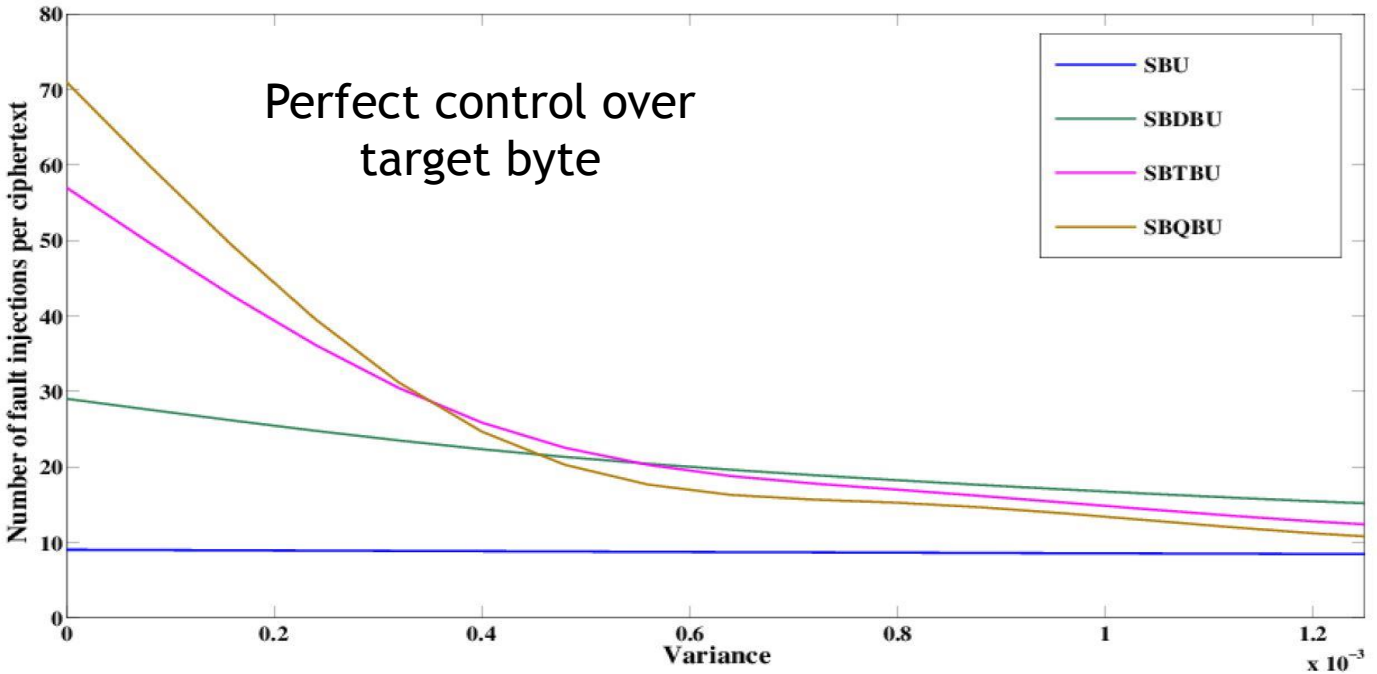
Simulation results

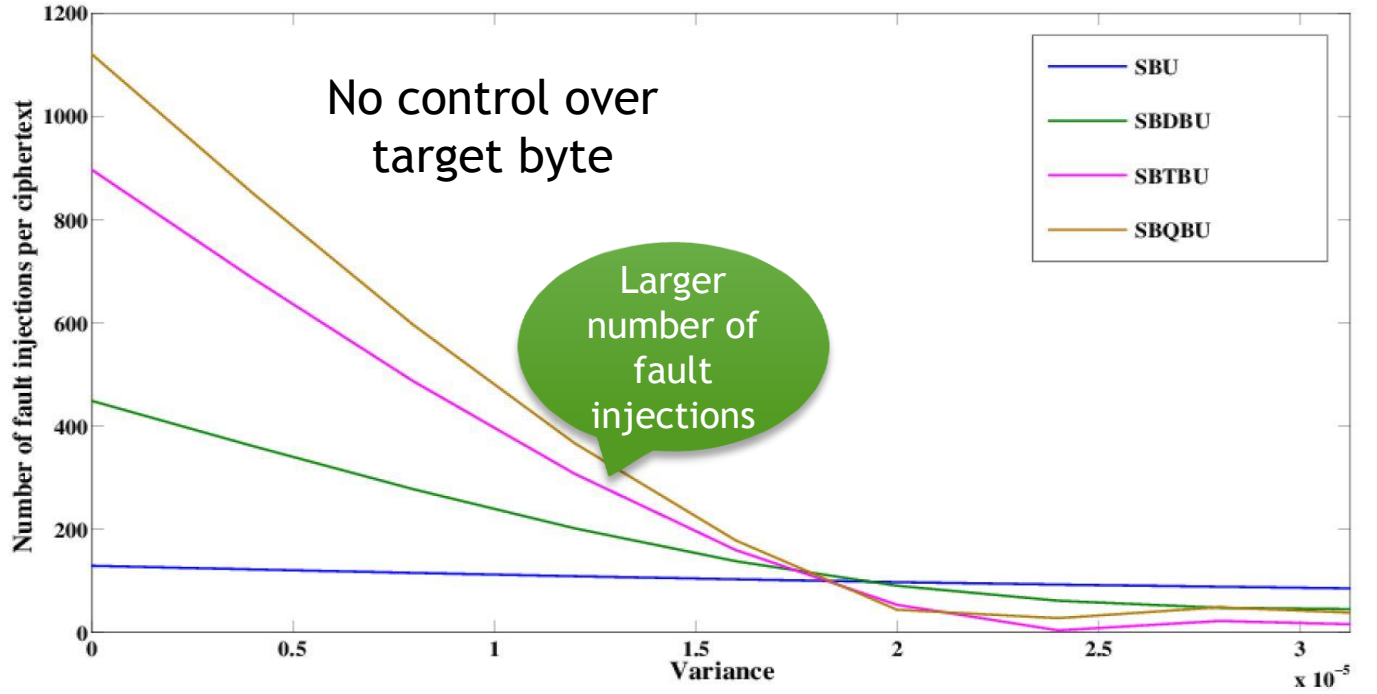
Number of ciphertexts required to guess a key byte with 99% accuracy

Round	Fault Model	N_C
8	SBU	320-340
	SBDBU	580-600
	SBTBU	1000-1040
	SBQBU	1900-2000
9	SBU	288-320
	SBDBU	608-640
	SBTBU	832-880
	SBQBU	1360-1440

Simulation : Part-2

- Vary the degree of bias in the fault model
  - Control the variance of the fault probability distribution
- Observe the number of fault injections per *useful ciphertext*
- Two adversarial models:
  - Perfect control over target byte
  - No control over target byte

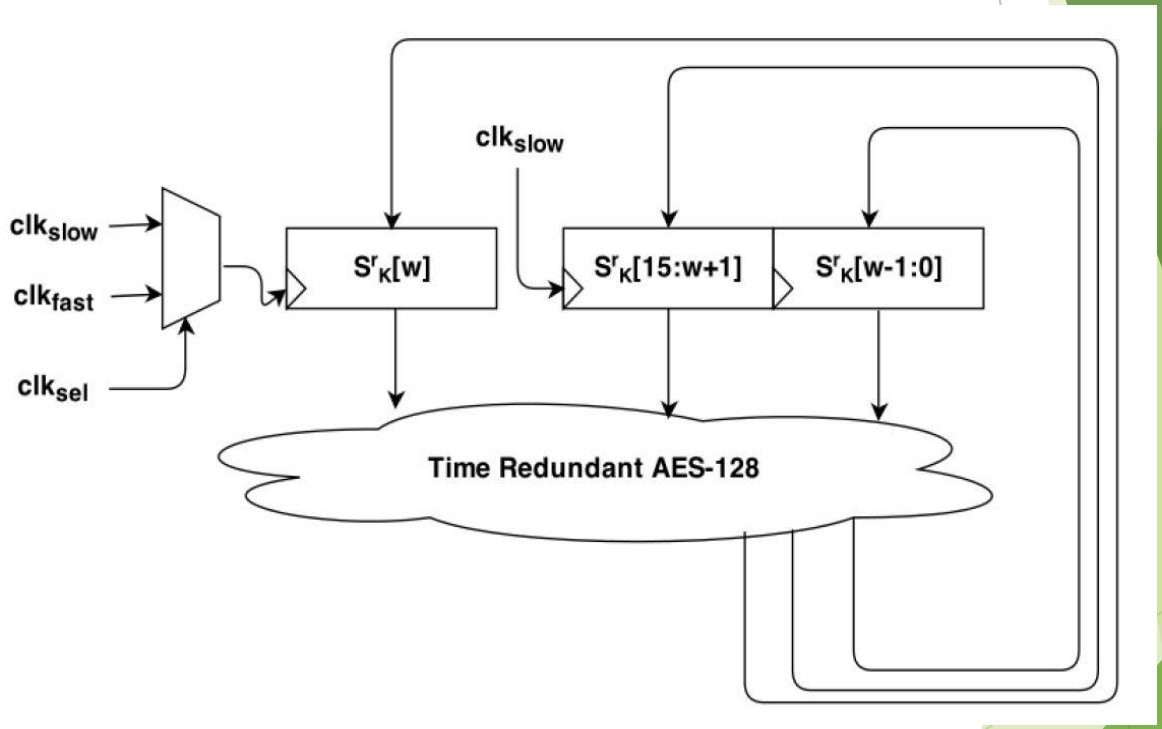




Practical Experiments

- Proposed attack evaluated on a time redundant hardware implementation of AES-128 on Spartan 3A FPGA
  - RTL Verilog definition of time redundant AES
  - A total of 20 rounds with comparison after each even round
- Two types of implementations:
  - Type 1 : Target byte is fixed
  - Type 2 : Target byte is random

Fixing the Target Byte



Experimental Results

Useful ciphertexts

Total Fault Injections

Round	Fault Model	Fault Variance		N_C	N_F(simulation)		N_F(experimental)	
		Type-1	Type-2		Type-1	Type-2	Type-1	Type-2
8	SBU	9.5×10^{-2}	3.6×10^{-3}	304.75	340.48	647.52	387.67	687.91
	SBDBU	1.4×10^{-2}	9.2×10^{-4}	625.12	1456.25	1506.25	1448.45	1652.30
	SBTBU	9.7×10^{-3}	4.9×10^{-4}	1020.49	1815.60	2315.40	1974.86	2395.83
	SBQBU	3.2×10^{-3}	5.9×10^{-5}	1878.55	7868.82	28038.54	8003.14	30201.41
9	SBU	9.2×10^{-2}	3.5×10^{-3}	304.24	385.88	603.11	387.98	632.71
	SBDBU	8.8×10^{-2}	7.9×10^{-4}	624.65	641.18	1487.36	647.82	1556.69
	SBTBU	8.1×10^{-2}	6.7×10^{-4}	832.32	873.56	2054.00	878.23	2489.25
	SBQBU	7.5×10^{-2}	3.5×10^{-5}	1328.22	1788.84	17239.10	1809.25	20145.66

Results presented per byte of key

Conclusions

- Biased fault models weaken the time redundancy countermeasure considerably
- Our experiments demonstrate practically feasible attacks on actual implementations of time redundant AES-128
- Countermeasures based on uniform fault patterns must therefore be revisited in the light of biased fault models

Thank You!