# The MALICIOUS Framework:
# Embedding Backdoors into Tweakable Block Ciphers

Thomas Peyrin, Haoyang Wang

August 2020

School of Physical and Mathematical Sciences,
Nanyang Technological University, Singapore

# Introduction

# Backdoors

- Most of time, backdoors of an encryption system refer to those weakness intentionally created in the implementation level, such as protocols of key management and key escrow.
- The other type is the cryptographic backdoor, which is embedded during the design phase of a cryptographic algorithm.

  Known examples:

  - `Dual_EC_DBRG`.
  - The suspicious S-box of `Kuznyechik` and `Streebog`.

# Academic Research

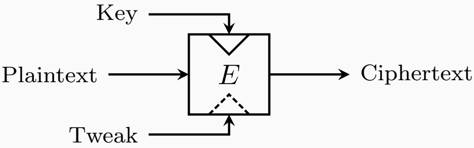
Limited number of works focus on the research of cryptographic backdoors. Almost all designs were either broken or can't provide solid security proof.

**Our contributions**

- We propose the **MALICIOUS framework** to embed backdoors into tweakable block ciphers.
- We show that our backdoor is efficient.
- We provide a concrete security bound for our backdoor.
- We provide a cipher example **LowMC-M**, and give proofs of its backdoor security and classical cipher security.

# The MALICIOUS Framework

# Tweakable block ciphers



A tweakable block cipher accepts an additional input, so-called **tweak**, in order to select the permutation computed by the cipher even if the key is fixed.

- No need to keep the tweak secret.
- An attacker could even have full control over the tweak, i.e., choosing whatever value he wants.

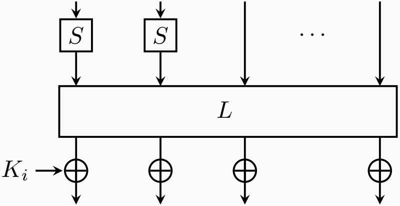
# Block ciphers with partial non-linear layers

**Substitution-Permutation Network (SPN)**

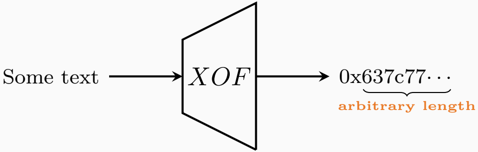
SPN is a method of designing iterated block ciphers, an SPN round consists of a linear layer and a non-linear layer.

**Partial non-linear layer**: the non-linear layer (S-boxes) is only applied to a subpart of the internal state.

- Typical ciphers: `ZORRO`, `LowMC`.



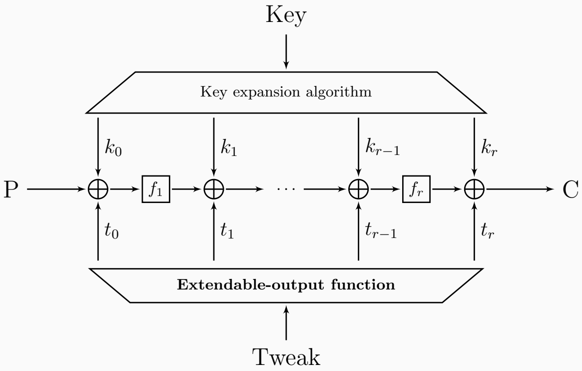
# Extendable-output functions



An extendable-output function (XOF) is a generalization of a hash function which maps an arbitrary length input to an arbitrary length output.

- Security properties: **collision resistance**, **preimage resistance** and **second preimage resistance**.
- Typical algorithms: `SHAKE128`, `SHAKE256`.

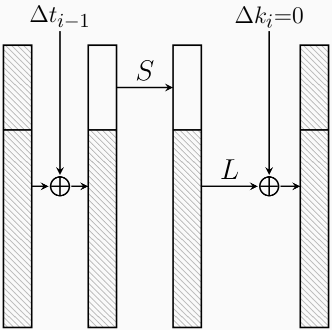
# The MALICIOUS framework



**The MALICIOUS construction** is a framework to build a key-alternating tweakable block cipher with the following special features:

- The non-linear layer of each round function $f_i$ is partial.
- The sub-tweaks are obtained from the original tweak $T$ by an XOF: $\mathsf{XOF}(T) = t_0 || t_1 \cdots || t_r$.

# The backdoor



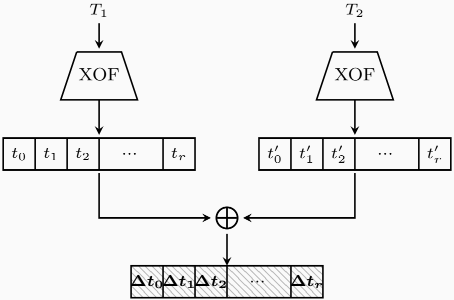
**Related-tweak differential characteristic with probability 1**

- Difference of the non-linear part is canceled by the sub-tweak addition.
- The differential characteristic is built with a secret tweak pair, we call it malicious tweak pair.
- The attack using the backdoor is under the chosen-tweak scenario.

# How to build the backdoor?

## Step 1

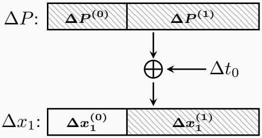
- Choose a pair of tweaks and keep it secret.
- Compute the corresponding sub-tweak differences by the XOF.

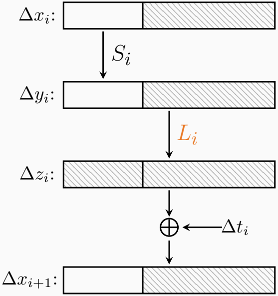


# How to build the backdoor?

## Step 2

- Select a plaintext difference $\Delta P$, satisfying $\Delta P^{(0)} = \Delta t_0^{(0)}$ such that $\Delta x_1^{(0)} = 0$.
- Generate the differential characteristic round by round, by selecting an appropriate linear layer $L_i$ of each round, satisfying $L_i(\Delta y_i)^{(0)} = \Delta t_i^{(0)}$ such that $\Delta x_{i+1}^{(0)} = 0$.





Note: It is possible to embed multiple differential characteristics.

# The Backdoor Security

# Target-difference resistance

**Definition (Target-difference resistance)**

A hash function $H$ is target-difference resistant if it is hard to find two inputs $x$ and $y$ such that $H(x) \oplus H(y) = \Delta$, where $\Delta$ is a non-zero constant.

The complexity is the same as the classical collision resistance (where $\Delta = 0$), that is the birthday bound $O(2^{n/2})$.

Security strength:

- `SHAKE128`: $\min(n/2, 128)$ bits
- `SHAKE256`: $\min(n/2, 256)$ bits

# The backdoor is protected by the XOF

- Finding the malicious tweak pair is difficult even if the differential characteristic is public known. The complexity is the target-difference resistance of the XOF used in the framework.

$$XOF(T_1) \oplus XOF(T_2) = \Delta t_0 || \Delta t_1 || \cdots || \Delta t_r$$

- The complexity will be at most $O(2^{128})$ for `SHAKE128` and $O(2^{256})$ for `SHAKE256`.

# Other potential backdoors

1. Actually, as we did not fix the tweak length, there might be other tweak pairs satisfying the requirement.

$$XOF(T_1') \oplus XOF(T_2') = \Delta t_0||\Delta t_1||\cdots||\Delta t_r$$

2. Furthermore, it is also possible that there is a suitable tweak pair for a randomly given differential characteristic, that is, the value of $\Delta t_0||\Delta t_1||\cdots||\Delta t_r$ is not fixed.

These tweak pairs imply new backdoors, which are not intentionally embedded by us. However, finding these backdoors is still as hard as finding the originally embedded backdoor.

# An Instantiation of MALICIOUS: LowMC-M

# LowMC-M

## LowMC-M

A family of tweakable block ciphers derived from the block cipher **LowMC**.

- The size of the non-linear layer $S$ can be set arbitrarily.
- The linear layer $L_i$ is an invertible $n \times n$ binary matrix which can be chosen randomly, but has to be customized if a backdoor is to be embedded.
- The tweak schedule uses `SHAKE128` or `SHAKE256`.

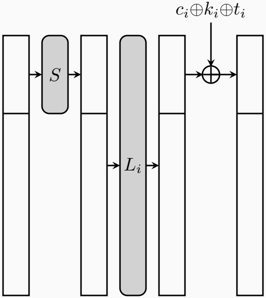


**Figure 1:** A single round of LowMC-M

# Backdoor security

LowMC-M has the following security properties:

- **Undetectable.** The attacker is unable to detect whether an instance of LowMC-M is embedded with backdoors or not.
- **Undiscoverable.** It is computationally difficult for the attacker to recover the backdoors (due to the target-difference resistance of the XOF).
- **Traceable.** If the backdoor is used in an attack, it will reveal the information of the backdoor (since it is chosen-tweak chosen-plaintext attack).

# Classical cipher security

### Attacks without using the tweak

The security of LowMC-M can be reduced to the security of LowMC which remains strong currently.

- Without considering the tweak, LowMC-M is an equivalent representation of LowMC.
- Even if a LowMC-M instance is backdoored, we show that its customized linear layer matrices can be considered as independently and randomly chosen from the view of the attacker.

### Attacks based on the tweak

Since the tweak schedule is an XOF, the attacker can't control its output. Thus, the tweak can't provide additional advantage for the attacker.

# Future Works

# Future works

- Can we use the framework to build other backdoored cryptographic algorithms? Such as hash functions and MACs.
- Is it possible that other cryptanalysis techniques than just a plain differential attack can be used in the framework?
- How to make the backdoored cipher untraceable?

Thank you!