周报 向嘉豪 (2025 年 5 月 26 日)

摘要： 本周针对第四篇论文开展故障攻击防护机制的系统性调研，构建了理论证明类、冗余校验类和混合架构类三大防护策略分类框架。建立了形式化故障攻击模型，并分析了当前实验中电压毛刺故障注入成功率过低（0.1%）的技术问题。尝试应用 [BFP19] 参数优化策略未能有效改善注入效率，识别出现有方法在当前实验环境中的适用性限制。为诊断故障注入失效原因，计划通过固件重写技术暴露内部状态进行深入分析。

下周计划： 1) 重写 STM32F303 固件暴露故障注入时的内部状态，诊断注入失效根本原因 2) 完善故障攻击防护机制理论

1 防护机制分类调研

基于对近年来 CHES 会议相关研究的深入分析，特别是 2024 年发表的两篇代表性工作，我们构建了故障攻击防护机制的系统性分类框架。根据防护策略的核心方法论差异，现有技术可归纳为三大类别：理论证明类防护、冗余校验类防护和混合架构类防护。以 [Gen23, DOT24, THN+24] 等近期 CHES 文献为基础，我们对各类防护机制的技术特征和安全保证进行了深入分析。

理论证明类防护基于严格的数学理论构建可证明安全的防护机制。代表性工作 [CM09] 在随机预言模型下证明了 PSS 编码机制对随机故障攻击的可证明安全性，为基于编码的防护策略提供了重要的理论支撑。此类方法的核心优势在于能够通过数学证明建立明确的故障攻击安全边界。

冗余校验类防护通过引入冗余机制和状态校验来检测和缓解故障影响。[Gen23] 提出的基于中间 WOTS+ 缓存的对策通过缓存关键中间状态实现故障检测时的安全回滚，有效控制了故障传播范围。[DOT24] 的 StaTI 方案基于阈值实现和线性编码技术，在非组合攻击场景下同时防护侧信道和故障攻击，展现了统一防护框架的技术可行性。

混合架构类防护结合软硬件多层防护机制构建分层防护体系。[THN+24] 形式化的 k 故障抗性分割概念通过可证明的安全保证减少硬件攻击面，并在此基础上引入软件防护层，实现了软硬件协同的鲁棒性故障防护解决方案。这种分层架构为构建全面的故障攻击防护体系提供了重要的设计参考。

2 故障注入效率问题分析

尽管前期实验成功观察到错误密文输出，但电压毛刺故障注入的成功率仅为 0.1% 且表现不稳定，显著低于 [BFP19] 报告的 STM32F3xx 平台 4% 成功率。这一差异表明当前攻击流程存在系统性问题，需要通过形式化攻击模型深入分析故障注入效率低下的根本原因。

2.1 形式化攻击模型构建

针对故障注入效率问题，我们建立了严格的理论分析框架。考虑目标设备上执行的密码计算 $\mathcal{C} : \mathcal{K} \times \mathcal{M} \to \mathcal{O}$，其中 $\mathcal{K}$、$\mathcal{M}$ 和 $\mathcal{O}$ 分别表示密钥空间、消息空间和输出空间，$\mathcal{S} = \{s_0, s_1, \ldots, s_n\}$ 表示执行期间的内部计算状态序列。

攻击者模型。对手 $\mathcal{A}$ 控制故障注入预言机 $\mathcal{F}(t, \sigma, \phi, \alpha)$,参数化为执行窗口 T 内的时序 $t \in [0, T]$、目标计算域 $\sigma \in \Sigma$（其中 $\Sigma = \{\text{ALU}, \text{控制单元}, \text{存储器}\}$）、注入机制 $\phi \in \{\text{电磁}, \text{激光}, \text{电压}\}$，以及强度 $\alpha \in \mathbb{R}^+$。故障预言机以概率 $P_{\text{fault}}(t, \sigma, \phi, \alpha)$ 诱导状态转换 $s_i \mapsto s_i^{\text{fault}}$，对手通过观察输出对 $(o_{\text{clean}}, o_{\text{fault}})$ 获取攻击信息。

安全假设。内部状态 $s_i \in \mathcal{S}$ 对攻击者 $\mathcal{A}$ 保持不透明，形式化为 $\mathcal{A}(s_i) = \perp$。故障效应根据 $P(\Delta|t, \sigma, \phi, \alpha)$ 概率性表现，其中 Δ 表示故障预言机诱导的计算偏差。攻击者无法确定性控制故障传播，需承认随机故障模型：瞬态比特损坏 $\Delta_{\text{bit}} \sim \text{Bernoulli}(p_\sigma)$、指令干扰 $\Delta_{\text{instr}} \sim \text{Geometric}(q_\sigma)$、数据损坏 $\Delta_{\text{data}} \sim \text{Uniform}(\mathbb{F}_2^w)$，其中成功概率 p_σ, q_σ 依赖于目标域 σ。

2.2 参数优化策略分析

基于电压注入方式，注入时序和持续时间对应 $(t, \alpha) \sim (x, y)$ 参数空间，目标故障为指令干扰 Δ_{instr}。由于 $P(\Delta|t, \sigma, \phi, \alpha)$ 过低导致攻击注入概率不足，我们采用 [BFP19] 提出的参数优化方法对 (t, α) 参数空间进行系统化优化。该文献针对电压毛刺攻击参数空间复杂性提出了两种搜索策略。

我们首先实现半自动监督搜索策略：通过随机生成并插值描述候选毛刺波形的 (x, y) 点集合，迭代选择参数区间内的随机样本并测试组合效果，基于人工评估结果逐步缩减参数空间。然而，该策略在 STM32F303 实验中未达预期效果，经过约 100 次迭代仍未发现有效参数组合。由于监督搜索阶段的失效，依赖其输出作为初始种群基础的遗传算法全自动优化策略亦无法有效实施。

实验结果表明，现有参数优化方法在当前实验环境中存在适用性限制，需要深入分析故障注入失效的根本原因。为此，我们计划通过重写 F303 固件暴露注入时的内部状态，进一步诊断故障注入成功率过低的技术问题，为后续防护机制设计提供更准确的攻击模型基础。

参考文献

[BFP19] Claudio Bozzato, Riccardo Focardi, and Francesco Palmarini. Shaping the glitch: Optimizing voltage fault injection attacks. *IACR TCHES*, 2019(2):199–224, 2019.

[CM09] Jean-Sébastien Coron and Avradip Mandal. PSS is secure against random fault attacks. In Mitsuru Matsui, editor, *ASIACRYPT 2009*, volume 5912 of *LNCS*, pages 653–666. Springer, Berlin, Heidelberg, December 2009.

[DOT24] Siemen Dhooghe, Artemii Ovchinnikov, and Dilara Toprakhisar. StaTI: Protecting against fault attacks using stable threshold implementations. *IACR TCHES*, 2024(1):229–263, 2024.

[Gen23] Aymeric Genêt. On protecting SPHINCS+ against fault attacks. *IACR TCHES*, 2023(2):80–114, 2023.

[THN+24] Simon Tollec, Vedad Hadzic, Pascal Nasahl, Mihail Asavoae, Roderick Bloem, Damien Couroussé, Karine Heydemann, Mathieu Jan, and Stefan Mangard. Fault-resistant partitioning of secure CPUs for system co-verification against faults. *IACR TCHES*, 2024(4):179–204, 2024.