基于序列密码的强PUF抗机器学习攻击方法

汪鹏君*① 连佳娜①② 陈 博①
①(温州大学电气与电子工程学院 温州 325035)
②(温州大学计算机与人工智能学院 温州 325035)

摘 要：物理不可克隆函数(Physical Unclonable Function, PUF)在信息安全领域具有极其重要的应用前景，然而也存在其自身安全受机器学习攻击等方面的不足。该文通过对PUF电路和密码算法的研究，提出一种基于序列密码的强PUF抗机器学习攻击方法。首先，通过构造滚动密钥生成器产生随机密钥，并与输入激励进行混淆；然后，将混淆后的激励通过串并转换电路作用于强PUF，产生输出响应；最后，利用Python软件仿真和FPGA硬件实现，并分析其安全性和统计特性。实验结果表明，当建模所用激励响应对(Challenge Response Pairs, CRPs)高达10^6组时，基于逻辑回归、人工神经网络和支持向量机的攻击预测率接近50%的理想值。此外，该方法通用性强、硬件开销小，且不影响PUF的随机性、唯一性以及可靠性。

关键词：硬件安全；强物理不可克隆函数；序列密码；机器学习
中图分类号：TN918.2; TP309 **文献标识码**：A **文章编号**：1009-5896(2021)09-2474-08
DOI: 10.11999/JEIT210726

Sequence Cipher Based Machine Learning-Attack Resistance Method for Strong-PUF

WANG Pengjun① LIAN Jiana①② CHEN Bo①
①(*College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou* 325035, *China*)
②(*College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou* 325035, *China*)

Abstract: Physical Unclonable Function (PUF) has extremely important application prospects to the field of information security, however, there are also shortcomings in its own security from machine learning attacks and other aspects. By studying PUF circuits and cryptographic algorithm, a method based on sequence cipher of strong-PUF is proposed to resist machine learning attacks. Firstly, the random key is generated by constructing a rolling key generator, which is obfuscated with the input challenge; Then the obfuscated challenge is applied to the strong-PUF through a series-parallel conversion circuit to generate the output response; Finally, Python software simulation and FPGA hardware implementation are used to analyze the safety and statistical properties. The experimental results show that the attack prediction rates based on logistic regression, artificial neural network and support vector machine are close to the ideal value of 50% when the CRPs used for modeling are up to 10^6 groups. In addition, this method has high versatile, low hardware overhead and does not affect the randomness, uniqueness and reliability of PUF.

Key words: Hardware security; Strong-Physical Unclonable Function (PUF); Sequence cipher; Machine learning

1 引言

随着物联网技术在智能家居、智能物流和智能医疗等领域的广泛应用，亟需在资源受限的条件下进行安全防护[1]。物理不可克隆函数(Physical Unclonable Function, PUF)作为一种特殊的“芯片指纹”提取技术，通过捕获IC制造过程中无法避免引入的工艺偏差，产生具有随机性、唯一性以及物理不可克隆性的特征密钥，可广泛应用于知识产权保护、设备认证、物联网防护、密钥安全存储等信息安全领域[2–4]。

自Pappu等人[5]首次提出物理单向函数的概念以来，已涌现出许多不同类型的PUF电路，对PUF施加n位激励C将生成m位响应R，从而形成其特

收稿日期：2021-07-19；改回日期：2021-08-20；网络出版：2021-09-06
*通信作者： 汪鹏君 wangpengjun@wzu.edu.cn
基金项目：国家重点研发计划项目(2018YFB2202100)，国家自然科学基金(62174121, 61904125)，温州市基础性科研项目(G20190006, G20210023)
Foundation Items: The National Key Research and Development Program of China (2018YFB2202100), The National Natural Science Foundation of China (62174121, 61904125), The Wenzhou Basic Scientific Research Projects (G20190006, G20210023)

有的激励响应对(Challenge Response Pairs, CRPs)，激励与响应的关系可用函数$R=f(C)$表示。根据产生CRPs能力的不同，PUF分为弱PUF和强PUF。弱PUF仅能产生有限数量CRPs，且一个PUF单元通常只产生一位输出响应，因此输出响应间相互独立，适用于密钥产生和存储。典型的弱PUF包括SRAM-PUF, DRAM-PUF和蝴蝶PUF等[6]。强PUF主要包括仲裁器PUF(Arbiter PUF, APUF)、环形振荡器PUF(RO-PUF)和算术逻辑单元PUF等[5,7,8]，适用于设备认证和目标识别。强PUF通过硬件资源重构产生大量CRPs，输出响应之间不可避免地存在相关性，因此容易受到逻辑回归(Logistic Regression, LR)、人工神经网络(Artificial Neural Network, ANN)和支持向量机(Support Vector Machine, SVM)等多种机器学习(Machine Learning, ML)算法的攻击[9,10]，攻击者可通过对强PUF的部分CRPs进行分析，一旦PUF电路的激励响应行为被精确建模，便可根据输入激励C预测输出响应R。

为抵抗ML攻击，近年来提出了许多抗攻击技术，大致可分为激励响应混淆技术[11–13]和结构非线性技术[14–16]。激励响应混淆技术通过隐藏激励与响应间的映射关系，阻止攻击者收集有效CRPs，从而提高抗攻击能力。Avvaru等人[11]利用m个并行的n级APUF异或产生1位响应，提出异或APUF(XOR-APUF)结构；Gao等人[12]在Controlled PUF中引入FSM状态转换器，提出PUF-FSM结构。然而，这些方法所需硬件开销较大，研究表明，利用改进的ML算法仍然能成功预测PUF激励响应行为。结构非线性技术通过改变PUF电路的内部结构，使激励与响应呈非线性关系以抵抗ML攻击。Santikellur等人[14]利用电流的非线性传输特性提出一种电流镜PUF；Vijayakumar等人[15]利用电压的非线性传输特性提出一种电压传输型PUF；Avvaru等人[16]利用PUF的中间响应值作为激励，提出一种前馈APUF。然而，结构非线性技术会在一定程度上降低输出响应的可靠性。此外，采用进化策略算法也能实现对非线性PUF电路的成功建模。鉴于此，通过对PUF电路结构和攻击方法的研究，利用异或去相关技术，结合序列密码原理，提出一种强PUF抗ML攻击方法。分别利用Python和FPGA对基于该方法的强PUF进行实验验证，采用经典ML攻击算法分析其抗攻击能力，并对随机性、唯一性、可靠性以及硬件开销等关键属性进行评估。

2　理论基础

2.1　APUF结构与原理

APUF结构简单、易于实现，已广泛应用在芯片认证、知识产权保护等多种安全场景中，但存在抗ML攻击能力不足等方面的问题。文献[17]表明，在ML攻击下APUF的预测率高达99%，且对于一个误码率为5%的64级APUF攻击所需的最少CRPs约为650。为提高PUF电路安全性，多种基于APUF的抗攻击方法相继提出，如XOR-APUF, MPUF, IPUF等[7]。近期研究发现，这些方法仍可通过分析目标PUF的部分CRPs实现对激励与响应关系的准确预测。

APUF电路结构如图1所示，由上下两条完全对称的延时路径和交叉耦合与非门组成的仲裁器构成。在激励信号$C=C_1, C_2, \cdots, C_n$的作用下，路径信号平行或交叉通过延时单元，仲裁器通过比较两条路径信号到达的先后顺序，判决输出响应0或1。理论上，两个信号应同时到达仲裁器，但延时路径不可避免地存在工艺偏差，导致信号延时不同，因此产生的响应0或1具有不可预测性。通常用式(1)线性延时模型表示，总延时为每一级延时单元传播延时的累加和：

$$\Delta = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Phi} \tag{1}$$

其中，$\boldsymbol{\omega}$表示APUF中每一个多路选择器传播延时的特征向量。n位激励C的函数表达式用$\boldsymbol{\Phi}$表示：

$$\boldsymbol{\Phi}(\boldsymbol{C}) = (\Phi_1(\boldsymbol{C}), \Phi_2(\boldsymbol{C}), \cdots, \Phi_n(\boldsymbol{C}), 1)^{\mathrm{T}} \tag{2}$$

其中，$\Phi_j(\boldsymbol{C}) = \prod_{i=j}^{n}(1-C_i), j=1,2,\cdots,k$。若$\Delta>0$，APUF的输出响应为1，否则为0。输出响应$t$表示为

$$t = \mathrm{sgn}(\Delta) = \mathrm{sgn}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Phi}) \tag{3}$$

为简化计算，式(3)中的t可用$t=2r-1$ $(t\in\{-1,1\})$代替，其中$r\in\{0,1\}$。

2.2　机器学习攻击方法

机器学习算法依托海量数据，通过在未知条件下对数据进行学习从而构建模型，实现给定任意未知输入，准确预测输出的功能。通常根据训练数据能否线性分类，分为如图2所示的线性可分问题和

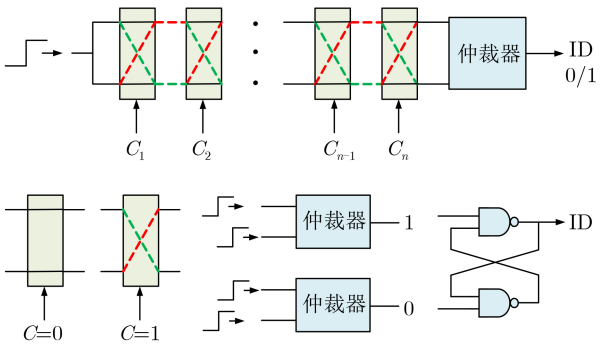

图 1 APUF结构

线性不可分问题。ML算法根据训练数据建立模型进行分类，如图2(a)中的红线所示，可用于预测新数据。对于PUF电路，其输出响应只有0/1两种情况(图2中实心和空心)，因此可用分类算法构建PUF模型。由于APUF的延时差Δ可线性表示，因此其产生的CRPs线性可分，APUF易受ML攻击。常见的ML算法主要包括LR, ANN和SVM等。LR适用于线性模型，具有结构简单、性能强大的特点，通过Sigmoid函数将线性回归得到的预测值作为输入数据，将线性加权和转换为相应的概率值，并据此预测事物类别。为增强算法性能，可利用梯度信息引入迭代算法，如Rprop迭代。ANN适用于线性不可分问题，由简单神经元经过相互连接形成网状结构，可模仿生物神经网络结构和功能，实现感知判断。通常由输入层、隐含层和输出层构成，根据训练集可不断调节各连接权重值改变连接强度，直至达到误差阈值或设定的迭代次数，训练结束。因此，ANN输入的特征向量通过隐含层变换达到输出层，可在输出层得到预测的分类结果。SVM对线性和非线性问题均适用，通过高维或无限维空间构造超平面，将原有限维空间映射到位数更高的空间中，并在该空间进行分类。若训练数据线性不可分，如图2(b)所示，可引入核函数将原始空间中线性不可分的数据集映射到高维空间，从而实现线性可分。常见的核函数有径向基核函数RBF($K(\omega,\boldsymbol{z})=\exp((-||\omega-\boldsymbol{z}||^2/\sigma^2))$、多层感知机核函数(MLP: $K(\omega,\boldsymbol{z})=\tanh(\alpha\boldsymbol{z}^{\mathrm{T}}\omega+\beta)$)和多项式核函数等。

3 抗机器学习攻击方法

3.1 序列密码结构与原理

密码学中根据加解密的密钥是否相同分为对称密钥密码体制和非对称密钥密码体制[18]。对称密钥密码体制中发送方和接收方使用相同的加密密钥和解密密钥。非对称密钥密码体制中发送方和接收方分别使用不同的加密密钥和解密密钥。对称密钥密码体制又为分组密码和序列密码两种，分组密码将明文划分为m个等长分组，使用同一密钥或算法对每一分组进行加密，加密后的分组合并作为密文进行输出；序列密码也称流密码，对明文按码元逐位加密，伴随加密过程的进行，每次仅输出1个密文，即1次只加密1个比特或1个字节，具有软硬件实现简便、转换速度快、误差传播低等优势。加密过程如图3所示，设密钥流$Z=z_0, z_1, \cdots, z_n$，明文$X=x_0, x_1, \cdots, x_n$，加密后的密文可表示为$Y=y_0, y_1, \cdots, y_n=Ez_0(x_0), Ez_1(x_1), \cdots, Ez_n(x_n)$。密钥流$Z$由滚动密钥生成器$f(\cdot)$产生：$z_i=f(k,s_i)$，其中$s_i$是密钥生成器中的记忆元件(存储器)在时刻$i$的状态，函数$f(\cdot)$受密钥$k$和$s_i$控制。初始滚动密钥$z_0=f(k,s_0)$由函数$f(\cdot)$、密钥$k$和初态$s_0$决定，此后由于输入明文$X$会影响内部记忆元件的存储状态$s_i$，因而$z_i$依赖于$k$, s_0, $\cdots$,s_{i-1}, x_0, $\cdots$, x_{i-1}等参数。

3.2 总体结构

综合密码算法原理、预处理结构以及强PUF，提出如图4所示的基于序列密码的强PUF抗ML攻击总体结构。该方法应用于强PUF可构成抗ML攻击

(a) 线性可分　　(b) 线性不可分

图 2 机器学习中线性与非线性问题

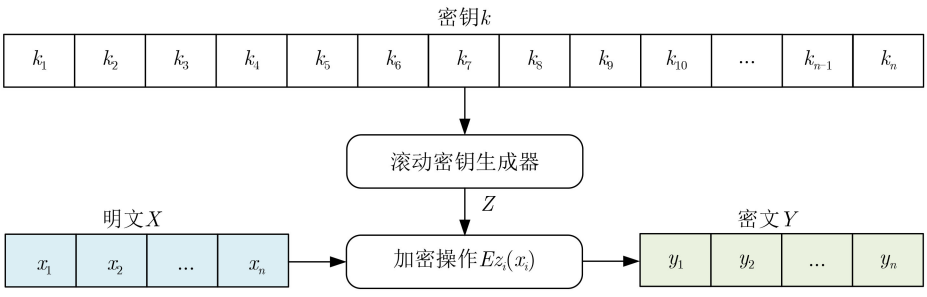

图 3 序列密码结构

PUF，主要包括混淆电路、串并转换(Serial Input Parallel Output, SIPO)电路和强PUF电路。其中，混淆电路将原始激励C复杂化处理后生成新的激励C'，并通过SIPO电路转化为并行的n位激励作用于强PUF上以生成最终响应R。

3.3 抗攻击方法实现

本文利用Xilinx Artix-7 FPGA开发板实现。该开发板由可编程逻辑单元、可编程I/O单元和布线资源3部分构成，其中可编程逻辑单元由2个Slice构成，1个Slice包含4个6输入查找表(Look Up Table, LUT)、3个数据选择器、1个进位链和8个触发器。为充分利用FPGA开发板Slice内部资源，混淆电路采用6个触发器和2个LUT构成，提出如图5所示的针对强PUF的抗攻击实现电路。在混淆电路中，滚动密钥生成器由1个LUT和6个触发器(F1～F6)构成，其中F1～F6的状态作为LUT的地址输入端(A1～A6)，LUT的数据端存放64位密钥k，触发器作为记忆元件控制滚动密钥Z。加密操作采用由一个LUT实现的XOR逻辑门，滚动密钥Z与原始激励C经过异或后生成Z'，Z'作为F1的输入影响下一时刻滚动密钥生成器的输出Z，触发器F6的数据作为混淆后的激励C'。具体过程如下：初始滚动密钥$z_0=f(k,s_0)$由函数$f(\cdot)$、密钥k和滚动密钥生成器初始状态$\boldsymbol{s}_0(a_5,a_4,a_3,a_2,a_1,a_0)^{\mathrm{T}}$决定，在时钟脉冲信号CLK的作用下，记忆元件由初始状态$\boldsymbol{s}_0$更新为下一状态$\boldsymbol{s}_1$，可表示为

$$(a_5,a_4,a_3,a_2,a_1,a_0)^{\mathrm{T}}=(c_0\oplus z_0,a_5,a_4,a_3,a_2,a_1)^{\mathrm{T}} \tag{4}$$

其中，c_0为第1位输入激励，a_0作为处理后的第1位激励，此时的滚动密钥$z_1=f(k,s_1)$。在下一时钟脉冲信号CLK的作用下，记忆元件的状态s_1更新为状态s_2，可表示为

$$\begin{aligned}&(c_0\oplus z_0,a_5,a_4,a_3,a_2,a_1)^{\mathrm{T}}\\&\quad=(c_1\oplus z_1,c_0\oplus z_0,a_5,a_4,a_3,a_2)^{\mathrm{T}}\end{aligned} \tag{5}$$

其中，$z_0=f(k,s_0)$，$z_1=f(k,s_1)$，c_1为第2位输入激励，a_1为处理后的第2位激励，此时的滚动密钥$z_2=f(k,s_2)$。因此在多个时钟脉冲信号CLK作用下，s_i可表示为

$$\begin{aligned}&(c_{i-1}\oplus z_{i-1},c_{i-2}\oplus z_{i-2},c_{i-3}\oplus z_{i-3},c_{i-4}\\&\quad\oplus z_{i-4},c_{i-5}\oplus z_{i-5},c_{i-6}\oplus z_{i-6})^{\mathrm{T}}\end{aligned} \tag{6}$$

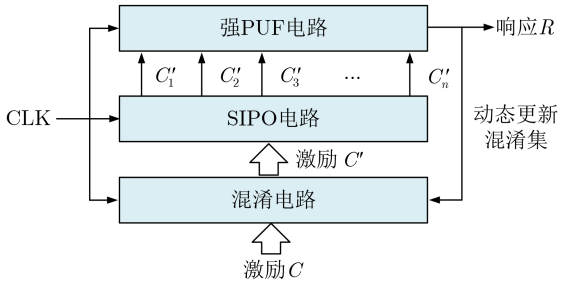

图 4 基于序列密码的强PUF抗ML攻击框图

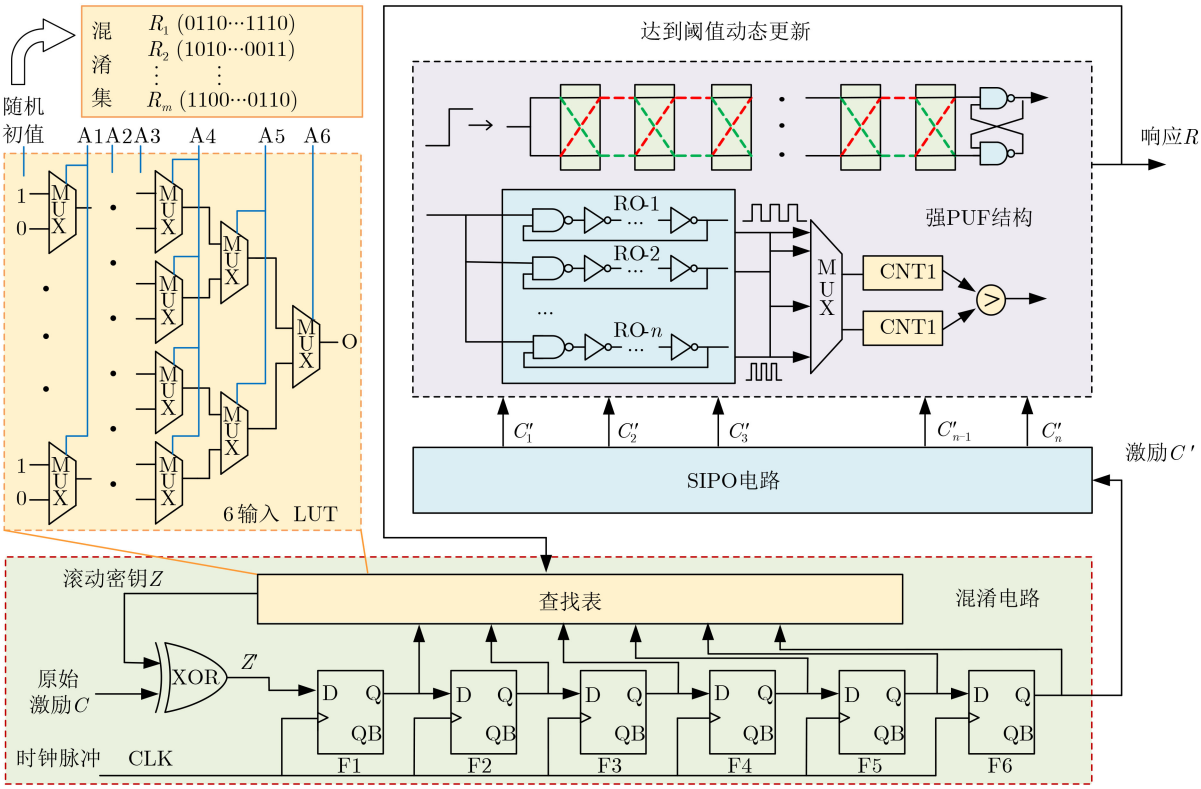

图 5 基于序列密码的强PUF抗ML攻击电路

此时的滚动密钥$z_i=f(k,s_i)$, c_{i-1}为第i位输入激励，$c_{i-7}\oplus z_{i-7}$作为处理后的第i位激励。

在每个时钟脉冲信号CLK到来时，滚动密钥生成器的输出z_i受记忆元件状态s_i控制，z_i与激励c_i异或实现逐位加密，其结果作为触发器F1的输入再次参与混淆。因此，当前滚动密钥生成器的输出z_i不仅受记忆元件状态s_i的控制，还受到先前输出$z_{(i-j)}$(其中j=1, 2, …, 6)的影响，整个过程循环迭代，类似序列密码的1次1密体制，在加密过程中每次产生1个输出密文，该密文即为新的激励C'，可表示为

$$\boldsymbol{C}'=\begin{pmatrix} Ez_{n-1}(c_{n-1}) \\ Ez_{n-2}(c_{n-2}) \\ Ez_{n-3}(c_{n-3}) \\ \vdots \\ Ez_0(c_0) \end{pmatrix}=\begin{pmatrix} c_{n-1}\oplus z_{n-1} \\ c_{n-2}\oplus z_{n-2} \\ c_{n-3}\oplus z_{n-3} \\ \vdots \\ c_0\oplus z_0 \end{pmatrix} \tag{7}$$

为减小硬件开销，仅利用一个混淆电路对激励C进行处理。由于激励C与C'为串行输入串行输出，为实现C'并行施加到强PUF产生响应R，需额外增加一个常规SIPO电路。

因此，所提强PUF抗ML攻击是通过混淆电路实现对输入激励C的逐位加密，由于混淆依赖于预置混淆电路的随机密钥k，故实际施加到目标PUF的激励将被彻底扰乱，且无法从原始激励通过逆变换获得，影响激励与响应的真实映射关系，提高PUF抗ML攻击能力。此外，还可增设动态更新机制进一步提高PUF安全性，一旦攻击者收集到的CRPs数量达到设定阈值，混淆电路中存储的密钥k将更新，促使更新前后相同输入地址生成的滚动密钥Z不同。

4 实验结果与分析

所提抗ML攻击方法具有通用性，适用于任何强PUF。由于APUF抗攻击性能存在严重不足，本文将以APUF作为实验对象，评估该方法的抗ML攻击能力、硬件开销以及输出响应的随机性、唯一性、可靠性等。利用Python和Xilinx Artix-7 FPGA分别对基于序列密码的APUF(Sequence Cipher base on APUF, SC-APUF)、原始APUF和XOR-APUF进行仿真和实现，并分析抗LR, ANN, SVM攻击能力(CRPs取100万组)。其中，为使Python仿真结果更真实，引入服从正态分布$N(\mu,\sigma^2)$的模拟噪声信号。

4.1 抗攻击能力

为了评估所提方法抗ML攻击效果，基于LR, SVM和ANN 3种常见的ML攻击方法分别对Python软件仿真和FPGA硬件实现的PUF电路分析抗攻击能力，攻击预测率与实验所用CRPs数量之间的关系如图6所示，可以发现，SC-APUF的抗ML攻击能力优于传统APUF和n-XOR-APUF(n=2,3,4)。且攻击结果显示当CRPs数量小于10^4组时预测率增长速度快；CRPs数量超过10^5组后，攻击预测率缓慢增长并趋于稳定。此外，3种攻击方式下基于Python和FPGA数据的预测率增长趋势基本一致，最终预测率差值不超过2.9%。LR攻击预测率如图6(a)和图6(b)所示，可以发现SC-APUF的预测率相比APUF, 2XOR-APUF和3XOR-APUF较低，抗攻击能力优于三者，且其最终预测率接近4XOR-APUF和5XOR-APUF。ANN攻击预测率如图6(c)和图6(d)所示，可以发现在ANN攻击下4XOR-APUF的预测率提高至72.05%，因此与除5XOR-APUF外的其他PUF结构相比，SC-APUF抗攻击能力强。SVM的攻击预测率如图6(e)和图6(f)所示，攻击效果与LR, ANN类似。综上，无论采用Python还是FPGA获得的CRPs数据进行抗攻击能力分析，在CRPs数量高达10^6组时，攻击预测率依然接近随机猜测。因此，所提方法具有良好的抗ML攻击能力。

4.2 随机性、相关性、唯一性、可靠性

随机性主要通过观察PUF电路输出响应中逻辑0和逻辑1的概率统计分布衡量[10]。理想情况下，逻辑0和逻辑1均匀分布，都接近50%。实验测得SC-PUF的随机性为50.44%，接近理想值50%，可用灰度图直观表示，图7(a)中逻辑0和逻辑1分布均匀、无明显逻辑偏向性，图7(b)为PUF输出响应的平均灰度映射，可以发现平均值在0.5附近波动，无明显空间相关性。

相关性用于衡量PUF输出响应间的相关程度，相关值越小，相关性越低。本文采用自相关函数(Autocorrelation Function, ACF)对SC-APUF的输出响应进行相关性分析，如图8所示。测试结果表明，在95%置信区间内，SC-APUF的ACF值为0.0063(接近理想值0)，因此具有良好的空间独立性。

唯一性用于评估不同PUF输出响应间的差异，可用平均片间汉明距离(Hamming Distance, HD)衡量。实验测得SC-APUF平均片间HD的拟合曲线如图9所示(红线部分)，归一化平均片间HD为0.5038，唯一性为50.38%，接近理想值50%，具有良好的唯一性。

可靠性采用平均片内HD衡量，对同一PUF多次

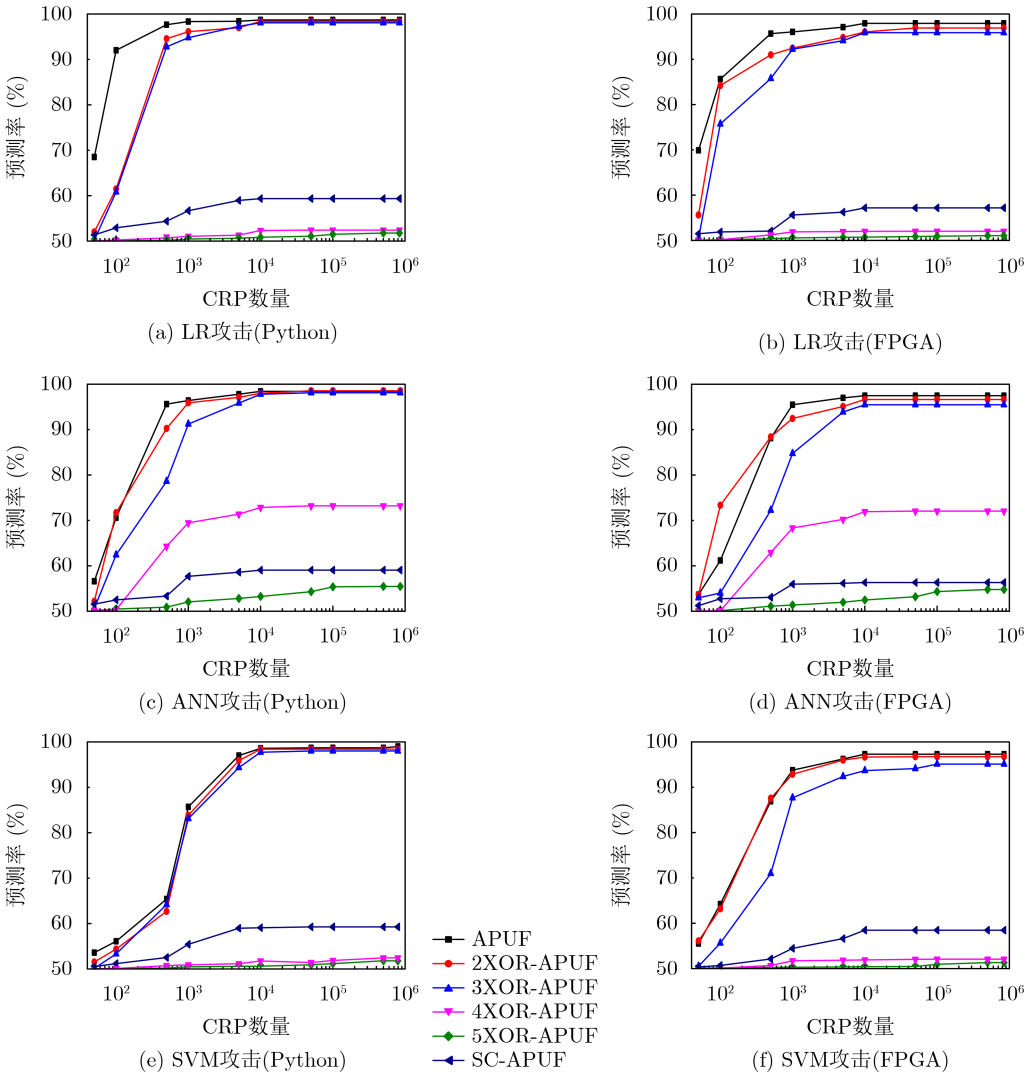

图 6 LR，ANN，SVM的攻击预测率

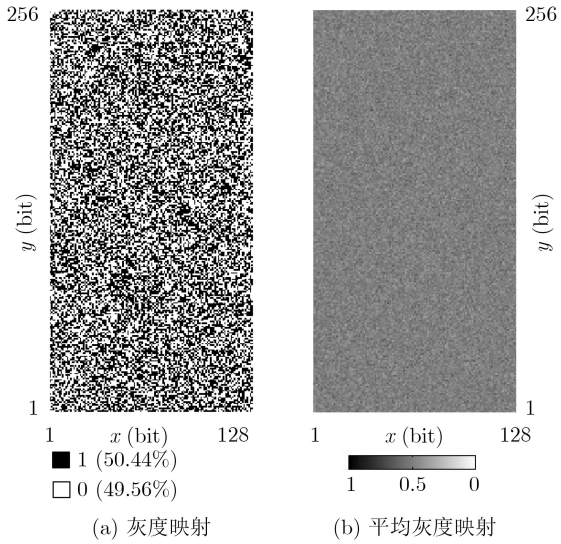

图 7 PUF输出响应分布情况

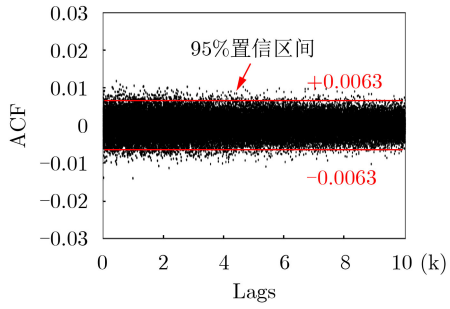

图 8 自相关性测试结果

施加相同激励其输出响应不发生变化。对SC-APUF在常温常压下进行循环读取，平均片内HD的拟合曲线如图9所示(蓝线部分)，归一化片内HD为0.0132，对应的可靠性为98.68%，接近理想值100%，具有良好的可靠性。

4.3 性能比较

在LR，ANN和SVM 3种攻击方式下，分别对Python软件仿真和FPGA硬件实现的SC-APUF及其对比PUF(APUF和n-XOR-PUF)的预测率统计特性如表1所示。表中所有攻击预测率均在10^6组

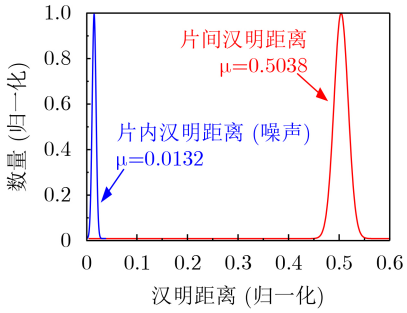

图 9 片间汉明距离与片内汉明距离拟合曲线

CRPs建模分析下获得，两种方式下的攻击预测率基本相同。由表1可知所提方法对PUF输出响应的随机性、唯一性和可靠性无明显影响。在3种攻击方式下，SC-APUF的抗攻击能力均优于APUF和n-XOR-APUF(n=2,3,4)，且仅需在APUF的基础上增加9-Slice的硬件开销，即可将平均预测率从98.08%有效降至58.28%以下；虽然5XOR-APUF(52.69%)抗攻击能力略强于SC-APUF，但5XOR-APUF增加的硬件开销为所提方法的8.56倍，增加了77-Slice；且当n不同时，硬件开销分别为2.22倍(n=2)、4.33倍(n=3)、6.44倍(n=4)和8.56倍(n=5)，因此所提方法硬件开销小，具有轻量型特征。此外，n-XOR-APUF随着异或级数n的增加，虽然其抗ML攻击能力逐渐增加，但输出响应的可靠性受到级数n的影响。由表可知，n-XOR-APUF的可靠性从异或级数n为1的98.84%逐渐降低到异或级数n为5的93.08%，而所提方法并不会对原始PUF的可靠性带来负面影响。因此，所提方法具有抗ML攻击能力强、轻量型和使用前后不影响PUF统计特性的优势。

表 1 预测率与统计特性的实验结果比较(%)

PUF结构	实现方法	硬件开销	随机性	唯一性	可靠性	LR	ANN	SVM
APUF	仿真	–	50.21	49.76	98.86	98.69	98.43	98.71
	FPGA	19-Slice	50.73	49.56	98.82	97.94	97.47	97.28
2XOR-APUF	仿真	–	49.67	50.61	98.92	98.38	98.54	98.41
	FPGA	39-Slice	50.13	49.35	98.74	96.91	96.63	96.74
3XOR-APUF	仿真	–	49.63	49.12	98.37	98.06	98.10	98.01
	FPGA	58-Slice	49.72	49.68	97.22	95.89	95.48	95.11
4XOR-APUF	仿真	–	50.22	49.36	95.18	52.41	73.21	52.43
	FPGA	77-Slice	49.94	50.33	93.38	52.01	72.05	52.07
5XOR-APUF	仿真	–	49.28	49.52	93.61	51.76	55.46	51.79
	FPGA	96-Slice	50.07	50.59	92.54	51.04	54.78	51.32
SC-APUF	仿真	–	50.19	50.69	98.73	59.34	59.06	59.27
	FPGA	28-Slice	50.44	50.38	98.68	57.18	56.33	58.47

5 结论

针对强PUF易受到机器学习攻击的不足，通过对序列密码加解密机理的研究，提出一种针对强PUF的抗ML攻击方法。该方法通过滚动密钥对初始激励进行逐位混淆，扰乱激励与响应之间的映射关系，以防止攻击者获取真正作用于原始PUF上的CRPs进行ML攻击，并可通过混淆集更新进一步提升抗攻击能力。实验结果表明，即便用于建模的CRPs数高达100万组，LR, ANN和SVM的攻击预测率均不超过60%。此外，所提方法硬件开销小、可扩展性强，使用前后不影响PUF的随机性、唯一性和可靠性等关键特性，可广泛用于强PUF抗ML攻击。

参 考 文 献

[1] 汪鹏君, 李乐薇, 郑雁公, 等. 基于气敏传感器的高稳态物理不可克隆函数发生器[J]. 电子与信息学报, 2021, 43(6): 1596–1602. doi: 10.11999/JEIT201104.
WANG Pengjun, LI Lewei, ZHENG Yangong, *et al*. High steady-state physical unclonable function generator based on gas sensors[J]. *Journal of Electronics & Information Technology*, 2021, 43(6): 1596–1602. doi: 10.11999/JEIT201104.

[2] ZHANG Jiliang and QU Gang. Physical unclonable function-based key sharing via machine learning for IoT security[J]. *IEEE Transactions on Industrial Electronics*, 2020, 67(8): 7025–7033. doi: 10.1109/TIE.2019.2938462.

[3] AMAN M N, TANEJA S, SIKDAR B, *et al*. Token-based security for the internet of things with dynamic energy-quality tradeoff[J]. *IEEE Internet of Things Journal*, 2019, 6(2): 2843–2859. doi: 10.1109/JIOT.2018.2875472.

[4] CHATTERJEE B, DAS D, MAITY S, *et al*. RF-PUF:

Enhancing IoT security through authentication of wireless nodes using *in-situ* machine learning[J]. *IEEE Internet of Things Journal*, 2019, 6(1): 388–398. doi: 10.1109/JIOT.2018.2849324.

[5] PAPPU R, RECHT B, TAYLOR J, *et al*. Physical one-way functions[J]. *Science*, 2002, 297(5589): 2026–2030. doi: 10.1126/science.1074376.

[6] LI Gang, WANG Pengjun, MA Xuejiao, *et al*. A 215-F^2 bistable physically unclonable function with an ACF of $<$ 0.005 and a native bit instability of 2.05% in 65-nm CMOS process[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2020, 28(11): 2290–2299. doi: 10.1109/TVLSI.2020.3014892.

[7] CUI Yijun, WANG Chenghua, LIU Weiqiang, *et al*. Lightweight configurable ring oscillator PUF based on RRAM/CMOS hybrid circuits[J]. *IEEE Open Journal of Nanotechnology*, 2020, 1: 128–134. doi: 10.1109/OJNANO.2020.3040787.

[8] LIU Weiqiang, ZHANG Lei, ZHANG Zhengran, *et al*. XOR-based low-cost reconfigurable PUFs for IoT security[J]. *ACM Transactions on Embedded Computing Systems*, 2019, 18(3): 25. doi: 10.1145/3274666.

[9] 徐金甫, 吴缙, 李军伟, 等. 基于敏感度混淆机制的控制型物理不可克隆函数研究[J]. 电子与信息学报, 2019, 41(7): 1601–1609. doi: 10.11999/JEIT180775.

XU Jinfu, WU Jin, LI Junwei, *et al*. Controlled physical unclonable function research based on sensitivity confusion mechanism[J]. *Journal of Electronics & Information Technology*, 2019, 41(7): 1601–1609. doi: 10.11999/JEIT180775.

[10] ZHANG Jiliang and SHEN Chaoqun. Set-based obfuscation for strong PUFs against machine learning attacks[J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2021, 68(1): 288–300. doi: 10.1109/TCSI.2020.3028508.

[11] AVVARU S V S, ZENG Ziqing, and PARHI K K. Homogeneous and heterogeneous feed-forward XOR physical unclonable functions[J]. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 2485–2498. doi: 10.1109/TIFS.2020.2968113.

[12] GAO Yansong, MA Hua, AL-SARAWI S F, *et al*. PUF-FSM: A controlled strong PUF[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018, 37(5): 1104–1108. doi: 10.1109/TCAD.2017.2740297.

[13] 刘伟强, 崔益军, 王成华. 一种低成本物理不可克隆函数结构的设计实现及其RFID应用[J]. 电子学报, 2016, 44(7): 1772–1776. doi: 10.3969/j.issn.0372-2112.2016.07.036.

LIU Weiqiang, CUI Yijun, and WANG Chenghua. Design and implementation of a low-cost physical unclonable function and its application in RFID[J]. *Acta Electronica Sinica*, 2016, 44(7): 1772–1776. doi: 10.3969/j.issn.0372-2112.2016.07.036.

[14] SANTIKELLUR P and CHAKRABORTY R S. A computationally efficient tensor regression network-based modeling attack on XOR arbiter PUF and its variants[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021, 40(6): 1197–1206. doi: 10.1109/TCAD.2020.3032624.

[15] VIJAYAKUMAR A and KUNDU S. A novel modeling attack resistant PUF design based on non-linear voltage transfer characteristics[C]. Proceedings of 2015 Design, Automation & Test in Europe Conference & Exhibition, Grenoble, France, 2015: 653–658. doi: 10.7873/DATE.2015.0522.

[16] AVVARU S V S and PARHI K K. Effect of loop positions on reliability and attack resistance of feed-forward PUFs[C]. Proceedings of 2019 IEEE Computer Society Annual Symposium on VLSI, Miami, USA, 2019: 366–371. doi: 10.1109/ISVLSI.2019.00073.

[17] XU Yunhao, LAO Yingjie, LIU Weiqiang, *et al*. Mathematical modeling analysis of strong physical unclonable functions[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2020, 39(12): 4426–4438. doi: 10.1109/TCAD.2020.2969645.

[18] 李俊志, 关杰. 非线性反馈移存器型序列密码的完全性通用算法[J]. 电子学报, 2018, 46(9): 2075–2080. doi: 10.3969/j.issn.0372-2112.2018.09.005.

LI Junzhi and GUAN Jie. Universal algorithm of full diffusion of stream cipher based on nonlinear feedback shift register[J]. *Acta Electronica Sinica*, 2018, 46(9): 2075–2080. doi: 10.3969/j.issn.0372-2112.2018.09.005.

汪鹏君：男，1966年生，教授，研究方向为集成电路设计、信息安全等技术及其相关理论.

连佳娜：女，1996年生，硕士生，研究方向为物理不可克隆函数攻击与防御.

陈　博：男，1981年生，讲师，研究方向为密码芯片攻击和防御理论及其VLSI实现.

责任编辑：陈　倩