基于人工神经网络特征向量提取的FF-APUF攻击方法

马雪娇[①] 李 刚[*②]

[①](温州理工学院数据科学与人工智能学院 温州 325035)

[②](温州大学电气与电子工程学院 温州 325035)

摘 要：为评估物理不可克隆函数(PUF)的安全性，需针对不同的PUF结构设计相应的攻击方法。该文通过对强PUF电路结构和工作机理的研究，利用人工神经网络(ANN)提出一种针对触发器-仲裁器物理不可克隆函数(FF-APUF)的有效攻击方法。首先，根据FF-APUF电路结构，利用多维数组构建电路延时模型；然后，对FF-APUF的二进制激励进行邻位划分，将划分后的激励转换为十进制并表示为行向量，实现特征向量提取；最后，基于提取的特征向量利用ANN构建攻击模型并通过后向传播算法获得最优参数。实验结果表明，相同条件下攻击预测率均高于其他3种常用的机器学习方法，尤其当激励响应对(CRP)数量较少、激励位数较多时，优势更加明显。当激励位数为128、CRP个数为100和500时，平均攻击预测率分别提高36.0%和16.1%。此外，该方法具有良好的鲁棒性和可扩展性，不同噪声系数下攻击预测率与可靠性相差最大仅0.32%。

关键词：物理不可克隆函数；触发器-仲裁器物理不可克隆函数；人工神经网络；特征向量提取

中图分类号：TN918; TP331 **文献标识码**：A **文章编号**：1009-5896(2021)09-2498-10

DOI: 10.11999/JEIT210614

ANN Feature Vector Extraction Based Attack Method for Flip-Flop Based Arbiter Physical Unclonable Function

MA Xuejiao[①] LI Gang[②]

[①](*School of Data Science and Artificial Intelligence*, *Wenzhou University of Technology*, *Wenzhou* 325035, *China*)

[②](*College of Electrical and Electronic Engineering*, *Wenzhou University*, *Wenzhou* 325035, *China*)

Abstract: In order to evaluate the security of Physical Unclonable Function (PUF), it is necessary to put forward corresponding attack methods for different PUF structures. By studying the structure and working mechanism of Flip-Flop based Arbiter Physical Unclonable Function (FF-APUF), an effective attack method against FF-APUF is proposed based on Artificial Neural Network (ANN) in this paper. Firstly, according to the circuit structure, the delay model of FF-APUF is established by using multidimensional array. Secondly, all binary challenge bits are divided by two adjacent bits which are converted to a decimal, and then the challenges are expressed as a row vector to extract the feature vector. Finally, based on the extracted feature vectors, the attack model is constructed by ANN, and the optimal parameters are obtained by back propagation algorithm. The experimental results show that the prediction accuracy of the proposed method is higher than other three common machine learning methods under the same conditions. The attack advantage is more obvious, especially when the number of Challenge Response Pairs (CRP) is less and the bit number of challenges is large. For example, when the number of challenge bit is 128, and the number of CRPs is 100 and 500, the average attack prediction accuracy increased by 36.0% and 16.1% respectively. In addition, the proposed method has good robustness and scalability, and the maximum difference of attack prediction rate and reliability is only 0.32% under different noise.

Key words: Physical Unclonable Function (PUF); Flip-Flop based Arbiter Physical Unclonable Function (FF-APUF); Artificial Neural Network (ANN); Feature vector extraction

收稿日期：2021-06-22；改回日期：2021-08-12；网络出版：2021-08-23

*通信作者： 李刚 ligang@wzu.edu.cn

基金项目：国家重点研发计划(2018YFB2202100)，国家自然科学基金(61874078, 61904125)，温州市基础性科研项目(G20190006, G20190003)

Foundation Items: The National Key Research and Development Program of China (2018YFB2202100), The National Natural Science Foundation of China (61874078, 61904125), The Wenzhou Basic Scientific Research Projects (G20190006, G20190003)

1　引言

物理不可克隆函数(Physical Unclonable Function, PUF)作为一种新兴的硬件安全原语，通过捕获芯片制造过程中不可避免引入的纹理特征，可产生具有随机性、唯一性以及防篡改特性的特征密钥[1]。由于PUF输出的特征信息根植于芯片制造过程中的随机工艺偏差，具有高安全性和轻量型特点，因此在信息安全领域具有广泛的应用前景，如设备认证、密钥存储、物联网抗攻击等[2–4]。

PUF激励与响应之间具有单向性，根据产生激励响应对(Challenge Response Pairs, CRP)的能力不同，可分为弱PUF和强PUF两大类。弱PUF和强PUF的CRP空间与激励位数分别呈多项式和指数增长关系。由于强PUF信息熵源复用，输出密钥之间不可避免地存在相关性，易受到支持向量机(Support Vector Machine, SVM)、逻辑回归(Logistic Regression, LR)、人工神经网络(Artificial Neural Network, ANN)等多种机器学习(Machine Learning, ML)算法的攻击。攻击者只需收集暴露在信道上的少量CRP，便可模拟与目标PUF具有几乎相同的激励响应行为[5]。早期由Lim[6]提出的Arbiter PUF (APUF)是一种典型的强PUF，同时引入SVM攻击算法，攻击预测率可达90%以上。Ruhrmair等人[7]采用LR算法对64位APUF进行攻击，利用650个CRP便可获得高达95%的预测率。为提高强PUF抗攻击能力，多种基于APUF的抗攻击强PUF被相继提出，如XOR APUF[8]、可配置环形振荡器PUF[9]、轻量级安全PUF[10]、iPUF(interpose PUF)[11]、多路复用器PUF(Multiplexer PUF, MPUF)[12]和Flip-Flop APUF(FF-APUF)[13]等。

PUF的攻击与防御是互相竞争发展的，近年来涌现出许多新型建模攻击方法。Awano等人[14]提出基于Xavier初始化和ReLU激活函数的深度神经网络攻击方法，对双仲裁器PUF(Double APUF, DAPUF)的攻击预测率较其他方法提升21.1%；Santikellur等人[15]提出基于张量回归网络的CP分解技术，对激励位数为64的7-XOR APUF仅用2500个CRP便可实现90%的预测率；Shi等人[16]提出基于ANN的近似攻击方法，对MPUF及其变体cMPUF和rMPUF的平均攻击预测率高达96.8%；Chatterjee等人[17]提出概率近似正确理论攻击方法，将iPUF近似表示为线性阈值函数，利用分类噪声和CRP模拟该函数，从而提高对iPUF的预测率；Chakraborty等人[18]提出基于遗传规划的二叉决策图辅助建模攻击方法，对APUF的攻击预测率可达94%，运行时间仅为经典ML攻击方法的1/3。上述攻击方法对DAPUF, XOR PUF, MPUF和iPUF等强PUF的安全性造成了严重威胁。然而，近期出现的FF-APUF具有良好的抗机器学习攻击能力，且在CRP数量较少时攻击者无法以较高预测率对其响应进行预测。鉴此，本文通过对强PUF攻击机理的研究并结合FF-APUF结构特点，提出一种基于ANN特征向量提取的FF-APUF攻击方法。通过软件建模验证FF-APUF在CRP数量较少时依然存在被攻击的可能，并对鲁棒性和可扩展性进行评估。

2　理论基础

2.1　APUF模型

APUF电路结构如图1所示，图1(a)是由上下两条并行的n级多路选择器(MUltipleXer, MUX)和由D触发器组成的仲裁器构成的APUF延时电路。输入信号T 通过APUF的n级偏差延时单元到达仲裁器的输入端，由于延时电路制造过程中不可避免引入的工艺偏差，信号T 到达仲裁器输入端的先后顺序不同，该偏差信号经仲裁器判决产生输出响应。当c_i=0时(i=1, 2, ⋯, n)，信号平行通过延时单元，反之则交叉通过。图1(b)给出了第i级APUF延时单元分别在c_i等于0和1时对应的延时路径及延时值，每位激励将延时单元分成平行或交叉的传播路径。

APUF可用线性累加模型表示，n级APUF结构的延时及输出响应可分别表示为式(1)和式(2)

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

$$R=\begin{cases}1, & \Delta(n)\geq 0\\ 0, & \Delta(n)<0\end{cases} \tag{2}$$

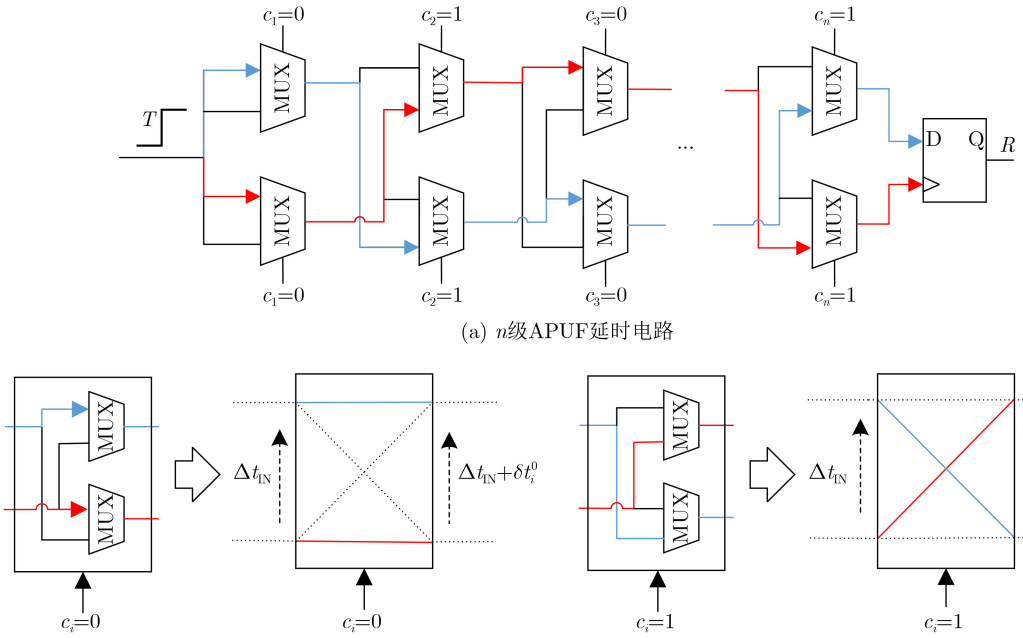
其中，$\boldsymbol{\omega}$ 是由各级延时组成的延时向量，$\boldsymbol{\Phi}$是与原始激励相关的函数构成的列向量，可分别表示为式(3)和式(4)

$$\begin{aligned}\boldsymbol{\omega}=&\frac{1}{2}[\delta_{t_1^0}-\delta_{t_1^1},\delta_{t_1^0}+\delta_{t_1^1}+\delta_{t_2^0}-\delta_{t_2^1},\cdots,\delta_{t_{n-1}^0}\\&+\delta_{t_{n-1}^1}+\delta_{t_n^0}-\delta_{t_n^1},\delta_{t_n^0}+\delta_{t_n^1}]^{\mathrm{T}}\end{aligned} \tag{3}$$

$$\begin{aligned}\boldsymbol{\Phi}&=[\boldsymbol{\Phi}_1(\boldsymbol{C}),\boldsymbol{\Phi}_2(\boldsymbol{C}),\cdots,\boldsymbol{\Phi}_n(\boldsymbol{C}),1]^{\mathrm{T}},\boldsymbol{\Phi}_j(\boldsymbol{C})\\&=\prod_{i=j}^{n}(1-2c_i),\ j=1,2,\cdots,n\end{aligned} \tag{4}$$

2.2　ANN原理

ANN以网络拓扑知识为理论基础，模拟人脑对复杂信息的处理，由大量具有关联的神经元组成。通常第1层的每个神经元代表输入变量，隐藏层及最后一层的神经元代表特定输出函数，称为激活函数。两个神经元之间的连接称为权值，相当于神经网络的记忆量。图2为3层神经网络示意图，其中，$\boldsymbol{X}$=[x_1, x_2]为输入向量，$\boldsymbol{W}^1$=[W_{11}, W_{12}, W_{13},



(a) n级APUF延时电路

(b) APUF的第i级延时单元

图 1 APUF电路结构

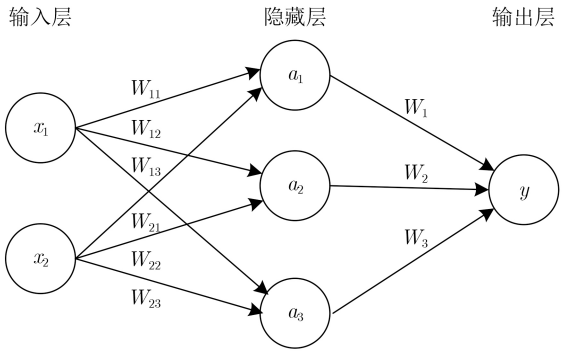


图 2 3层神经网络结构

W_{21}, W_{22}, W_{23}]为输入层与隐藏层之间的权值向量，$\boldsymbol{W}^2$=[W_1, W_2, W_3]为隐藏层与输出层之间的权值向量，该网络中未设置偏置项。神经网络的输入向量经过加权、求和、加入偏置项，采用前向传播算法最终输入到激活函数中，以产生输出。在神经网络的训练过程中，神经元根据训练集模型计算的预测误差，采用反向传播(Back Propagation, BP)算法不断更新权值和偏置项，最终得到全局最优解。

3 FF-APUF攻击方案

3.1 FF-APUF电路结构

FF-APUF产生响应的机理与APUF相似，均是通过判断路径延时差来产生唯一标识，由上下两路n级延时单元组成的延时电路和一个由RS触发器组成的仲裁器构成，结构如图3所示。对于上下两路延时单元，每一级延时单元由4个D触发器和3个二选一选择器构成。对于每一个D触发器，复位端CLR与CLEAR信号相连接，输入端与高电平“1”相连，输出端与两个MUX的输入端相连，相应MUX的输出端与后一个MUX的输入端相连，后一个MUX的输出端与下一级的4个D触发器的时钟输入端相连。

输入信号START由U_0和D_0组成的第1级延时单元的时钟输入端引入，在时钟上升沿到来之前，D触发器由CLEAR信号清“0”。由此，U_0和D_0的4个D触发器输出端分别产生一个上升沿信号，通过两位激励信号控制，该信号经两级MUX选择其中一路进行输出。同理，其他每一级的上下延时单元分别受两位激励信号配置，从而均可从4条不同的延时路径中选择一路输出。由于工艺偏差的存在，经D触发器和MUX产生的延时偏差不同。上下两路延时信号再经后面n–1级延时单元最终到达仲裁器的输入端。仲裁器则根据上下两路延时信号上升沿到来的先后顺序不同产生判决输出R。与APUF不同，为增加对比延时路径的多样性，FF-APUF上下两条延时路径输入激励信号的顺序恰好相反(U为c_0～c_{2n-1}, D为c_{2n-1}～c_0)。因而每一级延时单元的对比延时项由激励信号完全对称的4种增加到不完全对称的16种，可有效增加FF-APUF的信息熵和抗攻击能力。

3.2 FF-APUF延时模型

为构建延时模型，将FF-APUF的电路结构简化为图4，并在延时单元中标注激励的取值情况，如第i级延时单元(由U_i和D_i组成)的激励取值为$c_{2i}c_{2i+1}$=11和$c_{2(n-i)-1}c_{2(n-i-1)}$=00。T^{U}和T^{D}分别代表上下两路延时单元的总延时，则总的延时差ΔT可表示为

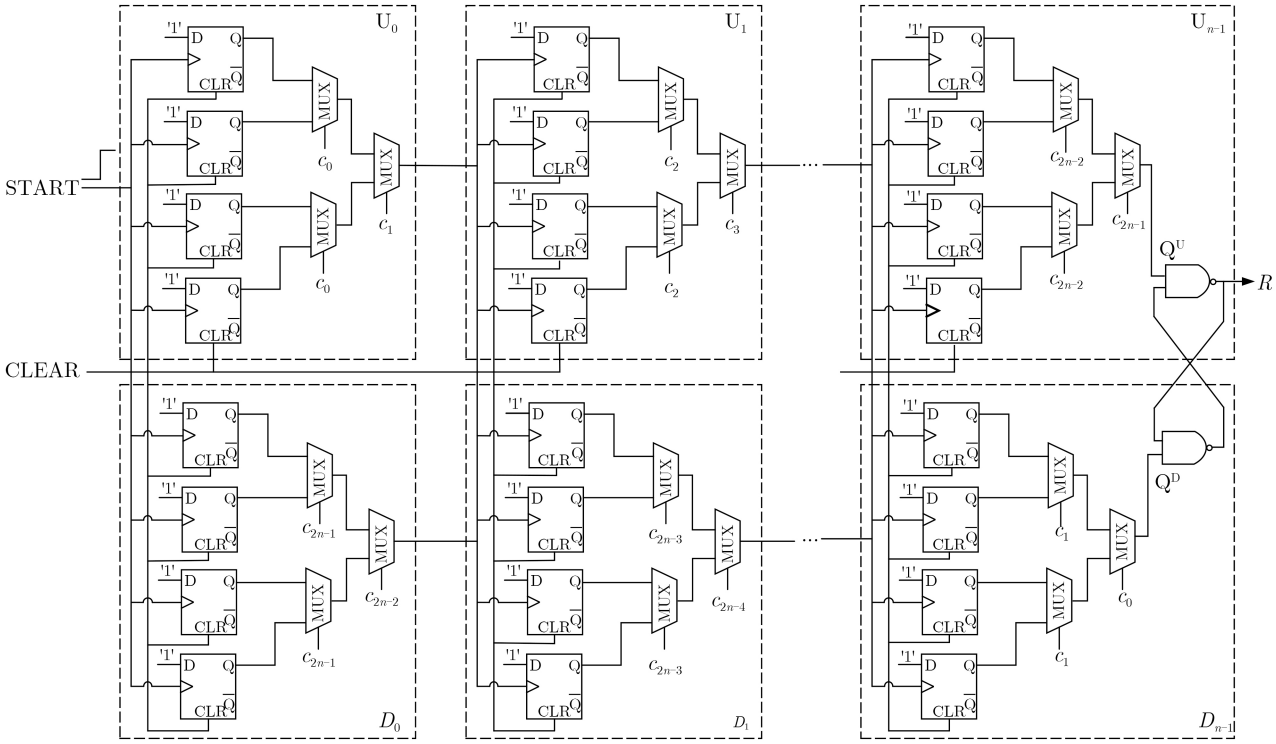


图 3 FF-APUF电路结构

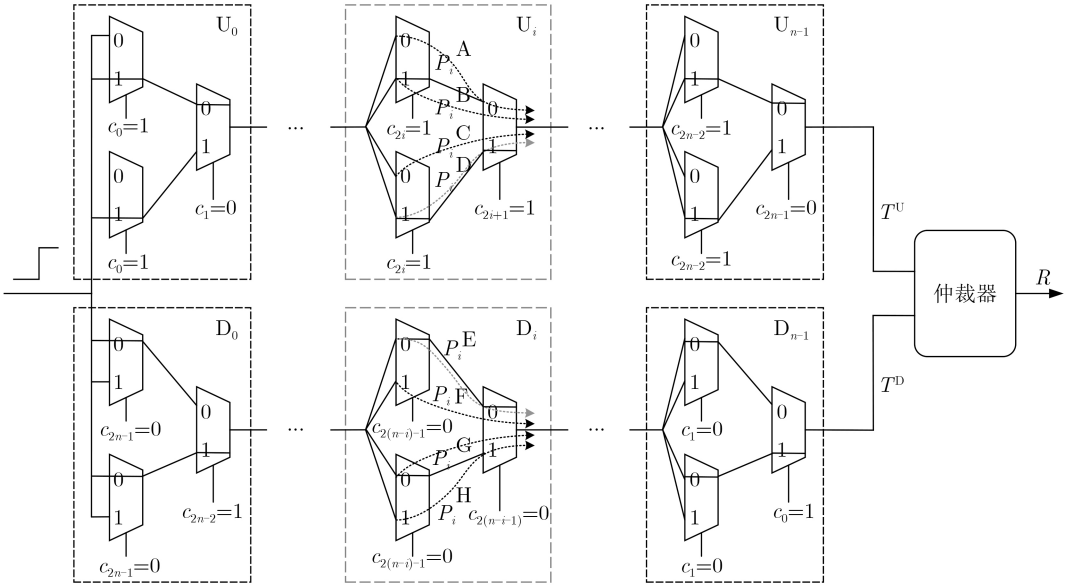


图 4 简化的FF-APUF电路结构

$$\Delta T = T^{\mathrm{U}} - T^{D} = \sum_{i=0}^{n-1} T_i^{\mathrm{U}} - \sum_{i=0}^{n-1} T_i^{\mathrm{D}} \tag{5}$$

其中，T_i^{U}和T_i^{D}分别为U_i和D_i的单元延时。$\Delta T>0$时，表明T^{U}比T^{D}大，仲裁器输出响应为1，反之输出响应为0。为方便表达，将输出0用–1表示，因此R=sgn(ΔT)。

将第i级上下两路共8条路径的延时表示为P_i^{alp}(alp=A, B, C, D, E, F, G, H)，则T_i^{U}或T_i^{D}对应的激励各自具有4种不同的组合，即上下每一级延时单元有4种可选延时路径。为便于处理与计算，将激励中的0用–1表示，则T_i^{D}、T_i^{D}可分别表示为式(6)和式(7)

$$\begin{aligned} T_i^{\mathrm{U}} =& \frac{1}{4}(1-c_{2i})(1-c_{2i+1})P_i^{\mathrm{A}} \\ &+ \frac{1}{4}(1+c_{2i})(1-c_{2i+1})P_i^{\mathrm{B}} \\ &+ \frac{1}{4}(1-c_{2i})(1+c_{2i+1})P_i^{\mathrm{C}} \\ &+ \frac{1}{4}(1+c_{2i})(1+c_{2i+1})P_i^{\mathrm{D}} \end{aligned} \tag{6}$$

$$\begin{aligned}T_i^{\mathrm{D}} =&\frac{1}{4}(1-c_{2(n-i)-1})(1-c_{2(n-i-1)})P_i^{\mathrm{E}}\\&+\frac{1}{4}(1+c_{2(n-i)-1})(1-c_{2(n-i-1)})P_i^{\mathrm{F}}\\&+\frac{1}{4}(1-c_{2(n-i)-1})(1+c_{2(n-i-1)})P_i^{\mathrm{G}}\\&+\frac{1}{4}(1+c_{2(n-i)-1})(1+c_{2(n-i-1)})P_i^{\mathrm{H}}\end{aligned}\tag{7}$$

由式(5)—式(7)进一步得出总延时为

$$\begin{aligned}\Delta T =&\frac{1}{4}\left((\boldsymbol{P}^{\mathrm{A}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{A}}+(\boldsymbol{P}^{\mathrm{B}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{B}}\right.\\&\left.+(\boldsymbol{P}^{\mathrm{C}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{C}}+(\boldsymbol{P}^{\mathrm{D}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{D}}\right)\\&-\frac{1}{4}\left((\boldsymbol{P}^{\mathrm{E}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{E}}+(\boldsymbol{P}^{\mathrm{F}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{F}}\right.\\&\left.+(\boldsymbol{P}^{\mathrm{G}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{G}}+(\boldsymbol{P}^{\mathrm{H}})^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{H}}\right)\end{aligned}\tag{8}$$

其中，$\boldsymbol{\Phi}^{\mathrm{alp}}=[f_0(\boldsymbol{C}),f_1(\boldsymbol{C}),\cdots,f_{n-1}(\boldsymbol{C})]$，$\boldsymbol{P}^{\mathrm{alp}}=[\boldsymbol{P}_0^{\mathrm{alp}},\boldsymbol{P}_1^{\mathrm{alp}},\cdots,\boldsymbol{P}_{n-1}^{\mathrm{alp}}]$，alp=A, B, C, D, E, F, G, H。alp与$f_i(\boldsymbol{C})$的关系可表示为

$$\mathrm{alp}=\begin{cases}\mathrm{A},f_i(\boldsymbol{C})=(1-c_{2i})(1-c_{2i+1})\\\mathrm{B},f_i(\boldsymbol{C})=(1+c_{2i})(1-c_{2i+1})\\\mathrm{C},f_i(\boldsymbol{C})=(1-c_{2i})(1+c_{2i+1})\\\mathrm{D},f_i(\boldsymbol{C})=(1+c_{2i})(1+c_{2i+1})\\\mathrm{E},f_i(\boldsymbol{C})=(1-c_{2(n-i)-1})(1-c_{2(n-i-1)})\\\mathrm{F},f_i(\boldsymbol{C})=(1+c_{2(n-i)-1})(1-c_{2(n-i-1)})\\\mathrm{G},f_i(\boldsymbol{C})=(1-c_{2(n-i)-1})(1+c_{2(n-i-1)})\\\mathrm{H},f_i(\boldsymbol{C})=(1+c_{2(n-i)-1})(1+c_{2(n-i-1)})\end{cases},\ i=0,1,\cdots,n-1\tag{9}$$

3.3 FF-APUF攻击方法

由2.1节可知，APUF每级两条延时信号的传播路径仅由一位激励决定。若APUF前后两组激励中仅一位激励不同，则产生相同输出响应的概率极大。由3.1节可知，与APUF不同，FF-APUF每一级的上下两路延时信号的传播路径分别由两位激励决定。即使前后两组激励中仅有一位不同，输出响应相同的概率也很小，因此输出结果较难预测。当上下两路每个延时单元的触发器个数为m时，FF-APUF信息熵是APUF的m^n倍，训练集相同的情况下FF-APUF的攻击预测率远低于APUF。本文提出针对m-FF-APUF的通用攻击模型，由于m=4时FPGA最小逻辑单元Slice的资源利用率最高，因此下面以4-FF-APUF为例介绍所提方法并构建FF-APUF攻击模型。

3.3.1 特征向量提取

对于APUF，基于ANN攻击的基本思路是将n位激励表示的行向量$\boldsymbol{C}_{\mathrm{APUF}}=[c_1,c_2,\cdots,c_n]$作为ANN的输入层向量，经隐藏层采用前向传播算法求得输出层响应，并采用BP算法对CRP进行训练得出算法的预测率。然而，在此过程中若将输入向量$\boldsymbol{C}_{\mathrm{APUF}}$作为特征向量并不能准确地描述输入与输出之间的映射关系。因此，根据APUF电路结构将提取的特征向量表示为式(3)并作为神经网络的输入向量。对于FF-APUF，若将$2n$位激励表示的行向量$\boldsymbol{C}=[c_0,c_2,\cdots,c_{2n-2},c_{2n-1}]$作为ANN的输入向量，也无法准确描述输入与输出之间的映射关系。因此，需要对FF-APUF进行特征向量提取：由于第i级延时由U_i和D_i对应的两位相邻激励共同决定，因此将FF-APUF的激励向量$\boldsymbol{C}$相邻两位进行组合并转换为十进制数，从而输入向量可表示为式(10)所示的行向量

$$\begin{aligned}\boldsymbol{C}^{\mathrm{S}}=&[2^0\times c_0+2^1\times c_1,2^0\times c_2+2^1\times c_3,\cdots,2^0\times c_{2n-2}\\&+2^1\times c_{2n-1}]=[c_0^{\mathrm{S}},c_1^{\mathrm{S}},\cdots,c_{n-1}^{\mathrm{S}}]\end{aligned}\tag{10}$$

为建立输入与输出之间的映射关系，根据c_i^{S}取值不同将$\boldsymbol{C}^{\mathrm{S}}$的每个值表示为向量$C_i^{\mathrm{S}}$

$$\boldsymbol{C}_i^{\mathrm{S}}=\begin{cases}[1,\ 0,\ 0,\ 0],\ c_i^{\mathrm{S}}=0\\[0,\ 1,\ 0,\ 0],\ c_i^{\mathrm{S}}=1\\[0,\ 0,\ 1,\ 0],\ c_i^{\mathrm{S}}=2\\[0,\ 0,\ 0,\ 1],\ c_i^{\mathrm{S}}=3\end{cases},\ i=0,1,\cdots,n-1\tag{11}$$

则输入层特征向量为

$$\boldsymbol{C}^{\mathrm{S}}=[C_0^{\mathrm{S}},C_1^{\mathrm{S}},C_2^{\mathrm{S}},\cdots,C_{n-1}^{\mathrm{S}}]\tag{12}$$

3.3.2 ANN攻击模型

式(12)和式(8)中的$\boldsymbol{P}^{\mathrm{alp}}$分别作为ANN的输入向量和权值向量，两者维数均为$n$。但实际上，输入层的输入向量仍然利用位宽为$2n$的激励表示，维数为$2n$。因此，需在ANN中增加转换层，将式(12)作为ANN转换层的输入向量。相应地，式(8)中的权值向量$\boldsymbol{P}^{\mathrm{alp}}$转换为

$$\boldsymbol{P}=[\boldsymbol{P}_0,\boldsymbol{P}_1,\cdots,\boldsymbol{P}_{2n-1}]\tag{13}$$

其中，$\boldsymbol{P}_i=\left[P_i^{\mathrm{A}},P_i^{\mathrm{B}},P_i^{\mathrm{C}},P_i^{\mathrm{D}},P_i^{\mathrm{E}},P_i^{\mathrm{F}},P_i^{\mathrm{G}},P_i^{\mathrm{H}}\right]^{\mathrm{T}}$，$i=0,1,\cdots,2n-1$。

输出层采用Sigmoid激活函数，并根据式(8)—式(13)构建ANN攻击模型如图5所示。其中，$a_1=\sum_{i=0}^{n-1}C_i^{\mathrm{S}}\boldsymbol{P}_i$，$\boldsymbol{P}_i=[P_i^{\mathrm{A}},P_i^{\mathrm{B}},P_i^{\mathrm{C}},P_i^{\mathrm{D}}]^{\mathrm{T}}$，$a_2=\sum_{i=n}^{2n-1}$

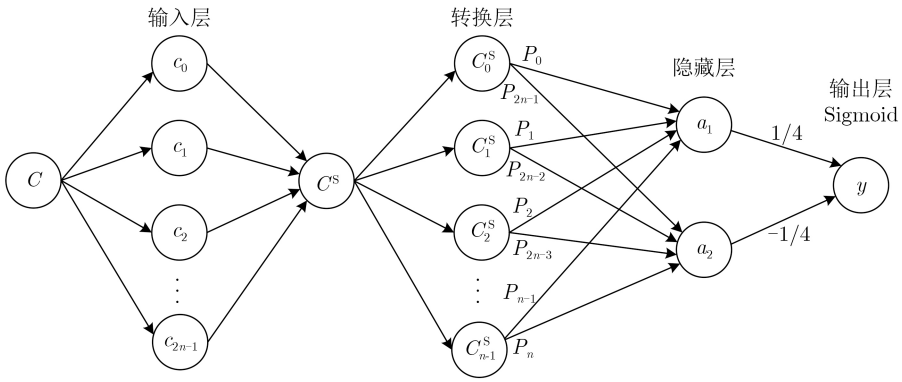


图 5 基于ANN的FF-APUF攻击模型

$C^{\mathrm{S}}_{2n-1-i}\boldsymbol{P}_i$，$\boldsymbol{P}_i=[P_i^{\mathrm{E}},P_i^{\mathrm{F}},P_i^{\mathrm{G}},P_i^{\mathrm{D}}]^{\mathrm{T}}$，$y=\mathrm{Sigmoid}\left(\frac{1}{4}(a_1-a_2)\right)$。输入层为$2n$个变量，代表FF-APUF的激励位数；转换层为提取的n个特征行向量，由$2n$个变量转换得到；隐藏层包含两个神经元a_1和a_2，分别代表上下两路延时电路的总延时；输出层采用激活函数Sigmoid，通过ANN前向传播得出响应。利用ANN的BP算法不断优化神经网络更新模型参数，并根据损失函数确定模型参数。该攻击模型能够较为准确地表示FF-APUF输入与输出的映射关系，从而得到接近实际的响应R。

4　实验结果与分析

实验所采用的硬件平台为Intel i7-8700 HYPERLINK CPU@2.4 GHz，软件平台为基于Python3.7.6的机器学习框架Tensorflow 1.13.1。基于上述软硬件平台，利用所提方法及3种常见的ML算法对其进行攻击。实验过程中，保证4种方法的激励一致且响应均在无噪声环境下获得，FF-APUF每条路径的延时参数δ服从均值μ为10、标准差σ为1的高斯分布$N(10,1)$，所提方法与其他对比方法均采用相同比例的训练集(80%)和测试集(20%)。

4.1　攻击预测率

为验证所提攻击方法的有效性，本文同时采用LR, SVM和高斯朴素贝叶斯(Gaussian Naive Bayes, GNB)算法分别对APUF和FF-APUF进行建模攻击。LR非常适合广义线性模型，SVM对于解决线性不可分问题很有效，而GNB算法将数据集中的各个特征视为完全独立，相比一般线性模型算法运行效率更高。需要说明的是，在使用SVM时需采用分类器分析并选择适当的核函数，并注意软间隔参数和正则化参数的调节匹配。结合特征值数量、训练样本规模以及训练时间，实验中SVM算法选用线性核函数。表1给出了ANN的主要特性参数及参数选取情况。其中，优化器的选择有Adam和RMSProp两种，根据提供CRP数量的不同来选择不同的优化器以达到最佳攻击效果。

表 1　ANN特性参数及选取

特性参数	参数选取
初始化权值	随机正态分布
初始化偏置	零向量
优化器	Adam或RMSProp
损失函数	最小均方误差
正则化	L2
输出层激活函数	Sigmoid

表2和表3分别给出了所提方法与其他3种对比方法的攻击预测率，以及攻击预测率的提升百分比。由表2可知，对于APUF和FF-APUF而言，本文方法的攻击的预测率均高于其他方法，且用于攻击的CRP个数越少优势越明显。具体而言，对于激励位数为128的FF-APUF，当CRP的个数为100时，LR, SVM和GNB的攻击预测率均小于60%，接近随机猜想值50%，而所提方法的预测率高达78.8%；当CRP个数为500时，3种对比方法的攻击预测率均小于80%，而所提方法的预测率高达89.2%。需要说明的是，当CRP数量较多时由于对比方法的攻击预测率已经接近100%，因此所提方法的攻击优势并不明显。由表3可知，对于FF-APUF在CRP个数为100，激励位数分别为32, 64, 128时，相比对比方法平均预测率分别提升7.7%, 10.2%和36.0%；相应的当CRP个数为500时，平均预测率分别提升6.5%, 11.5%和16.1%。此外，FF-APUF的激励位数越多，平均攻击预测率提升比例越明显。综上，所提方法攻击效果显著优势明显。

4.2　鲁棒性

在实际应用中，PUF电路易受到诸如温度、电压变化等噪声因素的影响。为模拟噪声环境中的PUF，同时验证所提方法的抗噪声性能，在PUF

表 2 攻击方法及预测率(%)

CRP个数	激励位数	APUF				FF-APUF			
		LR	SVM	GNB	本文方法	LR	SVM	GNB	本文方法
100	32	81.7	83.7	82.7	90.1	77.3	76.0	73.6	81.4
	64	77.7	74.6	78.6	80.3	66.4	77.0	75.7	80.1
	128	74.0	70.3	73.3	76.1	59.6	57.3	57.0	78.8
500	32	94.0	95.5	87.0	97.1	93.1	92.4	87.4	96.8
	64	93.7	92.4	84.6	96.2	88.3	87.1	82.4	95.7
	128	85.5	85.5	80.9	89.6	79.9	76.8	74.1	89.2
1000	32	97.5	97.9	92.2	99.5	97.0	97.3	90.2	99.0
	64	95.8	95.1	89.9	97.5	94.3	93.1	86.9	95.5
	128	93.4	91.7	86.2	93.5	89.2	87.1	81.5	91.5
2000	32	98.7	98.4	93.4	99.3	98.3	98.4	92.1	99.5
	64	97.8	97.1	92.8	98.8	96.7	96.8	90.3	97.8
	128	95.4	95.4	89.4	97.0	94.2	93.2	85.8	96.5
5000	32	99.4	99.2	95.6	99.7	99.0	99.2	95.5	99.6
	64	99.0	98.8	93.4	99.2	98.5	98.5	94.1	98.9
	128	98.2	98.2	92.5	98.7	97.7	97.1	90.7	97.9
10000	32	99.7	99.4	96.7	99.9	99.4	99.4	94.7	99.7
	64	99.3	99.1	95.8	99.6	99.2	99.2	94.4	99.6
	128	99.0	98.9	95.2	99.4	98.6	98.7	94.0	99.0

表 3 所提方法相比其他方法预测率的提升比例(%)

CRP个数	激励位数	APUF				FF-APUF			
		LR	SVM	GNB	平均值	LR	SVM	GNB	平均值
100	32	10.3	7.6	8.9	9.0	5.3	7.1	10.6	7.7
	64	3.3	7.6	2.2	4.4	20.6	4.0	5.8	10.2
	128	2.8	8.3	3.8	5.0	32.2	37.5	38.2	36.0
500	32	3.3	1.7	11.6	5.5	4.0	4.8	10.8	6.5
	64	2.7	4.1	13.7	6.8	8.4	9.9	16.1	11.5
	128	4.8	4.8	10.8	6.8	11.6	16.1	20.4	16.1
1000	32	2.1	1.6	7.9	3.9	2.1	1.7	9.8	4.5
	64	1.8	2.5	8.5	4.3	1.3	2.6	9.9	4.6
	128	0.1	2.0	8.5	3.5	2.6	5.1	12.3	6.6
2000	32	0.6	0.9	6.3	2.6	1.2	1.1	8.0	3.5
	64	1.0	1.8	6.5	3.1	1.1	1.0	8.3	3.5
	128	1.7	1.7	8.5	4.0	2.4	3.5	12.5	6.2
5000	32	0.3	0.5	4.3	1.7	0.6	0.4	4.3	1.8
	64	0.2	0.4	6.2	2.3	0.4	0.4	5.1	2.0
	128	0.5	0.5	6.7	2.6	0.2	0.8	7.9	3.0
10000	32	0.2	0.5	3.3	1.3	0.3	0.3	5.3	2.0
	64	0.3	0.5	4.0	1.6	0.4	0.4	5.5	2.1
	128	0.4	0.5	4.4	1.8	0.4	0.3	5.3	2.0

模型中引入均值为0、标准差为σ_{noise}服从高斯分布的噪声。由于FF-APUF每条路径的延时参数同样服从高斯分布$N(\mu,\sigma)$，且延时参数与噪声相互独立，则FF-APUF含噪声的延时参数服从均值为μ，

标准差为$\sigma+\sigma_{\text{noise}}$的高斯分布$N(\mu, \sigma+\sigma_{\text{noise}})$。因此，将通过设置不同的噪声系数$\alpha=\sigma_{\text{noise}}/\sigma$来验证所提方法在不同噪声环境下的鲁棒性。

理想情况下可靠性为攻击预测率的最大值，即噪声环境下对PUF的攻击预测率不会超过PUF输出响应的可靠性。攻击预测率越接近可靠性，说明攻击方法的抗噪声能力越强。表4给出了FF-APUF在不同激励位数和噪声系数下，输出响应的可靠性和预测率的统计数据。表中所有统计数据均是通过10次实验获得的平均值，选取的CRP个数均为100000。由表4可知，所提方法在不同噪声系数和激励位数下的预测率与输出响应的可靠性非常接近。α=0时为无噪声条件，此时PUF的可靠性为100%，且在相同激励位数条件下攻击预测率最高。α逐步增大时，FF-APUF的可靠性和预测同步减小。此外，所有统计数据中可靠性与预测率之差的均值和标准差分别为0.32%和0.35%。因此，所提方法具有良好的鲁棒性。

表 4　不同噪声条件下FF-APUF可靠性和攻击预测率

CRP个数	激励位数	噪声系数α						
		0	0.0125	0.025	0.050	0.100	0.150	0.200
可靠性(%)	32	100.00	99.67	99.20	98.85	96.38	95.40	93.42
	64	100.00	99.68	99.16	98.33	96.14	95.57	93.39
	128	100.00	99.58	99.21	98.23	96.20	95.14	93.28
预测率(%)	32	99.94	99.57	98.91	98.57	96.11	94.28	93.31
	64	99.86	99.41	98.09	98.08	95.68	95.54	93.12
	128	99.80	99.59	99.10	98.27	95.23	95.06	92.62

4.3　可扩展性

实验过程仅针对FF-APUF的特定CRP个数进行测试，当CRP个数增加或电路规模增大时，攻击方法可能并不适用，因此需要对本文所提方法的可扩展性进行分析。可扩展性分析主要从两方面考虑：达到一定攻击预测率与所需CRP数量的关系以及计算复杂度[7]。文献[7]指出采用LR以一定预测率攻击n级APUF所需的CRP个数(N_{CRP})，以及攻击方法的基本运算数量(N_{BOP})应分别遵循式(14)和式(15)

$$N_{\text{CRP}} = O(n/\varepsilon) \tag{14}$$

$$N_{\text{BOP}} = O\left(\frac{n^2}{\varepsilon}\lg\frac{n}{\varepsilon}\right) \tag{15}$$

其中，ε为预测误差，计算复杂度由N_{BOP}决定，式(14)的经验公式为

$$N_{\text{CRP}} \approx 0.5(n+1)/\varepsilon \tag{16}$$

由3.1节可知，激励位数为n的APUF和FF-APUF对应的级数不同，分别为n和$n/2$，除此之外两者产生响应的机理相同。因此，以一定预测率攻击n级FF-APUF所需CRP个数同样遵循式(16)。为验证这一关系，选取激励位数为16, 32, 64, 128，CRP个数为100, 1000, 5000, 10000，采用线性回归拟合$N_{\text{CRP}}/(n+1)$与预测误差ε之间的关系如图6所示。图6中横纵坐标均为对数坐标，拟合曲线满足的关系为y=0.5109/x，其中x代表自变量$N_{\text{CRP}}/(n+1)$，y代表因变量ε。由此可得，N_{CRP}=0.5109$(n+1)/\varepsilon$亦遵循经验式(16)。

攻击方法的计算复杂度N_{BOP}受模型优化过程中预测值达到最优值时的迭代次数以及基本运算的计算复杂度两方面影响。预测值达到最优值时的迭代次数epoch与N_{CRP}之间的关系如图7所示。图中选取的激励位数为32, 64, 128，CRP个数为100, 1000, 2000, 5000, 10000，采用线性回归拟合得到的表达式为y=3773×lg(x)−19792。式中，x代表N_{CRP}，y代表迭代次数。由此可知，epoch与lg(N_{CRP})成正比，即迭代过程的计算复杂度可表示为O(lg(N_{CRP}))。对于基本运算的计算复杂度，由式(6)—式(12)可知，所有运算都只包含矩阵运算和常数加法运算。假设一个常数运算是一个复杂度为O(1)的初等运算，则方法的总运算复杂度可以通过初等运算的个数来衡量。n级FF-APUF的响应最终由每级延时单

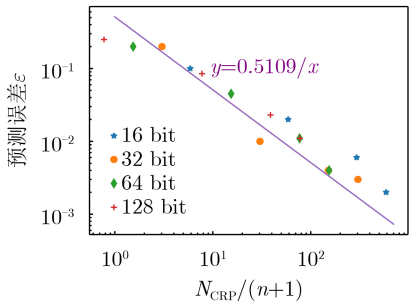


图 6　预测误差与$N_{\text{CRP}}/(n+1)$的关系

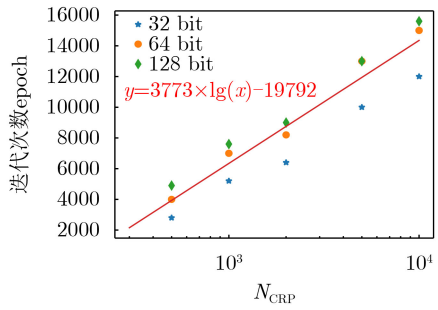


图 7 N_{CRP}与迭代次数的关系

元的6个MUX控制，则总的运算复杂度表示为$O(6n\times N_{\mathrm{CRP}})$。综合迭代过程的计算复杂度和运算复杂度得到的计算复杂度为$O(6n\times N_{\mathrm{CRP}}\times\lg(N_{\mathrm{CRP}}))$，将式(14)代入可得最终计算复杂度为

$$N_{\mathrm{BOP}}=O\left(\frac{6n^2}{\varepsilon}\lg\frac{n}{\varepsilon}\right) \tag{17}$$

与式(15)相比，所提方法的计算复杂度与LR的计算复杂度处在同一数量级。

5 结束语

PUF的攻击与防御是相互竞争发展的，攻击者通过优化现有攻击策略或提出新的攻击方法来评估PUF的抗攻击能力，从而指导PUF设计者提出更加安全可靠的电路结构。本文针对FF-APUF，提出一种基于特征向量提取的ANN攻击方法。所提方法与其他ML算法相比预测率显著提升，在CRP数量较少、激励位数较多时，攻击优势更加明显；不同噪声系数下攻击预测率与PUF稳定性非常接近，鲁棒性良好。此外，所提方法具有可扩展性，计算复杂度与LR处在同一数量级。未来工作将考虑利用ANN结合智能算法攻击其他强PUF结构，在保证预测率的前提下进一步提高攻击效率。

参 考 文 献

[1] LI Gang, WANG Pengjun, MA Xuejiao, *et al*. A multimode configurable physically unclonable function with bit-instability-screening and power-gating strategies[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2021, 29(1): 100–111. doi: 10.1109/TVLSI.2020.3030945.

[2] YAN Wei, TEHRANIPOOR F, and CHANDY J A. PUF-based fuzzy authentication without error correcting codes[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2017, 36(9): 1445–1457. doi: 10.1109/TCAD.2016.2638445.

[3] USMANI M A, KESHAVARZ S, MATTHEWS E, *et al*. Efficient PUF-based key generation in FPGAs using per-device configuration[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2019, 27(2): 364–375. doi: 10.1109/TVLSI.2018.2877438.

[4] 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.

[5] 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.

[6] LIM D. Extracting secret keys from integrated circuits[D]. [Ph. D. dissertation], Massachusetts Institute of Technology, 2004.

[7] RÜHRMAIR U, SÖLTER J, SEHNKE F, *et al*. PUF modeling attacks on simulated and silicon data[J]. *IEEE Transactions on Information Forensics and Security*, 2013, 8(11): 1876–1891. doi: 10.1109/TIFS.2013.2279798.

[8] SUH G E and DEVADAS S. Physical unclonable functions for device authentication and secret key generation[C]. The 44th ACM/IEEE Design Automation Conference, San Diego, USA, 2007: 9–14. doi: 10.1145/1278480.1278484.

[9] MAITI A and SCHAUMONT P. Improving the quality of a physical unclonable function using configurable ring oscillators[C]. 2009 International Conference on Field Programmable Logic and Applications, Prague, Czech Republic, 2009: 703–707. doi: 10.1109/FPL.2009.5272361.

[10] SAHOO D P, NGUYEN P H, MUKHOPADHYAY D, *et al*. A case of lightweight PUF constructions: Cryptanalysis and machine learning attacks[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2015, 34(8): 1334–1343. doi: 10.1109/TCAD.2015.2448677.

[11] NGUYEN P H, SAHOO D P, JIN Chenglu, *et al*. The interpose PUF: Secure PUF design against state-of-the-art machine learning attacks[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2019, 2019(4): 243–290. doi: 10.13154/tches.v2019.i4.243-290.

[12] SAHOO D P, MUKHOPADHYAY D, CHAKRABORTY R S, *et al*. A multiplexer-based arbiter PUF composition with enhanced reliability and security[J]. *IEEE Transactions on Computers*, 2018, 67(3): 403–417. doi: 10.1109/TC.2017.2749226.

[13] GU Chongyan, LIU Weiqiang, CUI Yijun, *et al*. A Flip-Flop

based Arbiter Physical Unclonable Function (APUF) design with high entropy and uniqueness for FPGA implementation[J]. *IEEE Transactions on Emerging Topics in Computing*, To be published. doi: 10.1109/TETC.2019.2935465.

[14] AWANO H, IIZUKA T, and IKEDA M. PUFNet: A deep neural network based modeling attack for physically unclonable function[C]. 2019 IEEE International Symposium on Circuits and Systems, Sapporo, Japan, 2019: 1–4. doi: 10.1109/ISCAS.2019.8702431.

[15] 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.

[16] SHI Junye, LU Yang, and ZHANG Jiliang. Approximation attacks on strong PUFs[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2020, 39(10): 2138–2151. doi: 10.1109/TCAD.2019.2962115.

[17] CHATTERJEE D, MUKHOPADHYAY D, and HAZRA A. Interpose PUF can be PAC learned[OL]. https://www.researchgate.net/publication/343524875_Interpose_PUF_can_be_PAC_Learned?channel=doi&linkId=5f2e86d5458515b7290d567f&showFulltext=true. 2020.

[18] CHAKRABORTY R S, JELDI R R, SAHA I, *et al*. Binary decision diagram assisted modeling of FPGA-based physically unclonable function by genetic programming[J]. *IEEE Transactions on Computers*, 2017, 66(6): 971–981. doi: 10.1109/TC.2016.2603498.

马雪娇：女，1991年生，助教，研究方向为物理不可克隆函数攻击与防御.

李　刚：男，1988年生，讲师，研究方向为密码芯片攻击防御理论及VLSI实现.

责任编辑：马秀强