针对密码芯片数据搬移能量曲线的机器学习攻击

张亮亮[1,2*]，唐 有[2]，张翌维[2]，王新安[1]

(1. 北京大学信息科学技术学院 北京 海淀区 100871；2. 国民技术股份有限公司博士后科研工作站 广东 深圳 518057)

【摘要】机器学习和传统侧信道攻击技术中的模板攻击有类似的处理过程，它们都由学习和测试两个阶段组成。模板攻击可以看作是有监督学习的分类技术，而机器学习领域也有很多这样的分类算法。为了探索机器学习算法在侧信道攻击中的应用，以实际密码芯片中的数据搬移操作为攻击对象，研究了一些机器学习算法利用已知搬移数据值的能量曲线，对新的能量曲线的搬移数据值的预测效果。结果表明，在采用一条能量曲线进行测试时，一些机器学习算法比模板攻击预测的正确率更高。

关 键 词 密码硬件； k近邻算法； 机器学习； 朴素贝叶斯； 支持向量机； 模板攻击
中图分类号 TP309.7 **文献标志码** A doi:10.3969/j.issn.1001-0548.2019.03.017

Machine Learning Attack to Power Traces of Data Movement in Cryptographic Chip

ZHANG Liang-liang[1,2*], TANG You[2], ZHANG Yi-wei[2], and WANG Xin-an[1]

(1. School of Electronics Engineering and Computer Science, Peking University Haidian Beijing 100871;
2. Post-doctoral Scientific Research Station, Nations Technologies Inc. Shenzhen Guangdong 518057)

Abstract Machine learning and template attack in the traditional side channel attack techniques have similar procedures, they all consist of two phases: learning and testing. Template attack can be considered as a classification technique for supervised learning, and there are many such classification algorithms in the machine learning field. In order to explore the application of machine learning algorithms in side channel attack, using the data movement operation in actual cryptographic chip as the attack target, the forecasting effect of some machine learning algorithms is investigated. These algorithms make use of power traces with known value of the moved data, then predict the value of the moved data for some new power traces. The results show that, when employing only one power trace in the testing stage, some machine learning algorithms have higher correctness rate than template attack.

Key words cryptographic hardware; k-nearest neighbors algorithm; machine learning; Naive Hayes; support vector machine; template attack

对于密码芯片等密码硬件，侧信道攻击(side channel attack, SCA)技术[1]是一种低成本、威胁大的物理攻击手段。侧信道攻击技术通常不需要对芯片进行破坏，也不会干扰芯片的正常工作。它通过采集大量的能量曲线(trace)并对它们进行统计分析，从而推断出硬件中正在执行的密码算法所使用的密钥等秘密值。可以按照采用的统计分析手段对侧信道攻击进行分类。简单功耗分析(simple power analysis, SPA)通常从能量曲线的形状出发来分析秘密信息的一些特性。差分功耗分析(differential power analysis, DPA)需要采用某种功耗模型对功耗进行建模，然后利用实际能量曲线反映的功耗和模型预测的理论功耗之间的相似程度来区分出不同的秘密值，并且高阶DPA攻击可以攻破具有防护方案的硬件设备。而模板攻击[2](template attack, TA)与SPA和DPA不同，TA会事先利用大量的能量曲线对不同的密钥或秘密值建立相关模板。在进行攻击的时候，利用新采集到的曲线跟已有的模板进行匹配，找出概率最大的密钥或秘密值。TA攻击等以上几种攻击方式都可用于从电流功耗或电磁辐射中提取的能量曲线，但是电磁辐射能量曲线的采集会受到很多因素的限制，本文的攻击过程采用的都是电流功耗能量曲线。

收稿日期：2017 – 12 – 12；修回日期：2018 – 11 – 01
基金项目：国家核高基科技重大专项(2014ZX01032-204)
作者简介：张亮亮(1986 –)，男，博士，主要从事侧信道分析、硬件物理不可克隆技术方面的研究. E-mail: zhang.liangliang@nationz.com.cn

上述传统的攻击手段，尤其是模板攻击，可以看作是一个分类问题，即将标签(密钥或秘密数据)未知的功耗曲线进行分类，找出可能性最大的标签。如果分类效果好，则可能性最大的标签就会是正确的秘密值。许多机器学习算法[3-4]可以用来处理分类和回归问题，并且机器学习算法可以利用原始的大量数据进行学习，然后将学习的结果或经验模型用于处理新的数据，这个过程和模板攻击的过程十分类似。模板攻击或机器学习的两个阶段在文献里有不同的称呼，例如profiling和validation，training和classification，或者learning和testing。本文采用learning和testing，即称为学习和测试阶段。机器学习两个阶段涉及的主要操作包括预处理、特征提取选择、学习/建模、预测分类。预处理阶段包括对数据进行标准化、变换范围等操作，或将数据进行二元化、处理某些数据缺失项等。特征选取通过主成分分析等算法进行特征维度约减，用来改善模型的最终精度或者提高建模的速度等。经过这些处理之后的数据会更加适合机器学习和建模。当知道已有数据集的正确标签，并且机器学习的任务也是预测未知数据的标签，这种任务称为监督学习(supervised learning)。若没有这样已知的标签可以利用，机器学习可以进行聚类或寻找描述数据的统计量等无监督学习(unsupervised learning)任务。由于模板攻击在建立模板阶段通常是容易知道正确秘密值的，所以它属于监督学习。机器学习需要大量的数据进行学习才能提高模型的准确程度，而在分析密码硬件设备的时候可以采用自动化的信号触发、功耗记录和存储，可以得到的功耗曲线数量能够达到百万量级，这也是可以将机器学习用于进行侧信道分析任务的一个有利因素。

本文首先简单介绍模板攻击，然后介绍 k 近邻算法、朴素贝叶斯和支持向量机这3种较常见的监督学习分类算法。尽管已经有文献使用机器学习算法用于分析密码算法的密钥[5-9]，但是数据搬移操作在密码硬件设备的工作过程中十分常见，一旦忽视对它的保护，容易造成意想不到的秘密泄露，并且数据搬移操作足够简单，它的过程和结果直观并且容易解释。所以将模板攻击和机器学习算法用于数据搬移操作的分析，比较了机器学习算法和模板攻击的效果和正确率。结果显示，利用相同的数据集进行建模和学习之后，在使用一条能量曲线进行测试时，机器学习算法的正确率会高于模板攻击。

1 模板攻击和机器学习算法

1.1 模板攻击

传统侧信道分析技术中的模板攻击会利用能量曲线集的均值和协方差进行建模[2]。假设对于每个秘密值 i 都采集了 m 次功耗曲线，其中要处理的曲线部分 T 由 n 个时间点的功耗值组成。则每个秘密值相应的功耗曲线均值 μ_i 和协方差 $\boldsymbol{\Sigma}_i$ 矩阵定义如下：

$$\mu_i = \frac{1}{m}\sum T$$

$$\boldsymbol{\Sigma}_i = \frac{1}{m}\sum \left(T-\mu_i\right)^t \left(T-\mu_i\right)$$

建模完成之后，对于新采集的未知秘密值的曲线 T' 来说，利用已知的 μ_i 和 $\boldsymbol{\Sigma}_i$ 估计 T' 对应某个秘密值的概率为：

$$P(T';\mu_i,\boldsymbol{\Sigma}_i) = \frac{1}{\sqrt{(2\pi)^n\left|\boldsymbol{\Sigma}_i\right|}}\exp\left(-\frac{1}{2}(T'-\mu_i)(\boldsymbol{\Sigma}_i)^{-1}(T'-\mu_i)^t\right)$$

利用这个概率可以处理 T' 包含一条功耗曲线或多条功耗曲线的情形，多条功耗曲线进行攻击成功概率通常会更高[10]。

1.2 k近邻算法

将能量曲线想象成 n 维空间中的点，并且可以知道这些点的标签。对于新的测试数据，若它和已知标签的某个点很接近，可以认为它和这个点相似，则测试数据的标签也有可能是跟这个点的标签一样。k近邻算法(k -nearest neighbors algorithm)会计算测试数据和所有点的距离，然后找出和测试数据距离最近的 k 点，统计这 k 个点的标签中出现次数最多那个标签作为测试数据的标签。对于 n 维空间的点，可以采用欧式空间距离，对于两个点 $x=(x_1,x_2,\cdots,x_n)$ 和 $y=(y_1,y_2,\cdots,y_n)$，其定义如下：

$$d_{x,y} = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}$$

显然k近邻算法十分简单，易于理解。通常 k 的取值应以实际训练结果为准。因为需要计算测试数据和所有已知数据之间的距离，并且找出前 k 个最大的距离，如果数据的维度和数量都很大，则保存这些数据的存储空间和测试时候的计算量都会很巨大。尽管它很简单，k近邻算法在识别手写数字等应用方面十分有效。

1.3 朴素贝叶斯

朴素贝叶斯(Naive Bayes)算法利用贝叶斯定理以及样本特征向量不同特征点的概率分布独立这一简单的假设来建立模型。设样本特征向量为

$(x_1,x_2,\cdots,x_n)$，则它属于某一类 C 的概率为：

$$P(C|x_1,x_2,\cdots,x_n)=\frac{P(x_1,x_2,\cdots,x_n|C)P(C)}{P(x_1,x_2,\cdots,x_n)}$$

利用特征向量不同特征点独立的假设，上式变为：

$$P(C|x_1,x_2,\cdots,x_n)=\frac{\prod_{i=1}^{n}P(x_i|C)P(C)}{P(x_1,x_2,\cdots,x_n)}$$

通常 $P(x_1,x_2,\cdots,x_n)$ 是恒定的，所以测试结果可以选取让 $\prod_{i=1}^{n}P(x_i|C)P(C)$ 取最大值的类别 C 即可。$P(C)$ 可以从样本数据集中统计出来。对于 $P(x_i|C)$，若特征点 x_i 取离散值，则所有 $P(x_i|C)$ 可以从样本数据集中统计得到。对于连续值，比如可以假定 x_i 服从高斯分布，然后通过最大似然估计从样本数据集中拟合出相应的分布，这样任意 $P(x_i|C)$ 也可以计算出。尽管这个方法简单，但是朴素贝叶斯算法在文档分类和垃圾邮件过滤等任务中表现很好，需要少量的数据就能得到需要的参数估计值。

1.4　支持向量机

支持向量机(support vector machine, SVM)是一种可以产生二值分类结果的分类方法，即它可以通过在 n 维空间中寻找到一个超平面将数据集分成两类。其中离超平面最近的那些点称为支持向量，决定超平面位置的原则就是让支持向量离超平面的间隔最大，SVM产生的分类器会尽可能的健壮。这可以转化成一个带有约束条件的优化问题，可采用拉格朗日乘子法求解。SVM训练的过程就是对可以表示超平面的参数进行优化，这是训练SVM最费时的部分。另外，对于非线性可分的情形，通过使用核函数技巧(kernel trick)将数据点映射到更高维或无穷维的空间中，而在更高维空间中，数据点可分离性会变得更好。比如文本分类中可显著减少所需的训练集数量，在图片分类问题中提高识别的精度以及用于生物学中蛋白质分类问题。基本的SVM只可用于二值分类，对于数据集具有 $K>2$ 个标签的情形，可以采用将多分类问题转换为多个二值分类问题。转换方式可以有以下两种：1) 将某个标签的数据集看作一类，其他所有标签的数据集看作一类，称作“一对多”(one-vs-rest)，显然这需要训练 K 个模型；2) 对每两个标签的数据进行一次基本的SVM训练，总共需要训练 $K(K-1)/2$ 个模型，这种方式称为“一对一”(one-vs-one)。对于测试数据，选择概率最大的结果来给出预测的类别。通过这样的方式，SVM也可以用于多类别分类问题。

2　数据搬移操作的分析

为了比较模板攻击和第1节介绍的3种机器学习算法，选择了在密码硬件操作中常见的数据搬移操作作为攻击对象。本文以C语言中memcpy内存拷贝函数为例进行数据搬移，并且搬移操作未进行防护。使用memcpy搬移一个字节的数据，数据取值的范围为0～31，这些数据的汉明重量为0～5，但占据9种可能汉明重量的多数，并且可以让学习和测试结果更加简单。针对每个数值，利用Riscure公司的功耗分析平台采集10 000条电流功耗曲线，这些数据用于建模学习。然后再随机挑选一个数值，也采集10 000条曲线用于测试分析。通过对采集的学习功耗曲线进行相关性分析，找出泄露信息量最大的一段能量曲线进行分析，这相当于进行特征选择。图1绘出了一条能量曲线示意图，其中对应memcpy操作的部分用方框标出。从图1可以看出，数据搬移具有功耗低的特点，导致其能量曲线特征跟其他部分相比不明显，直观视觉观察上远不如密码轮运算、蒙哥马利迭代运算等耗能操作所带来的能量特征，容易被忽视，但数据搬移有时会移动非常重要的敏感信息，特别在搬移密钥、子密钥等致命信息时，分析者一旦分析取得搬移的数据值，芯片将会被直接破译。下面使用第1节介绍的模板攻击和机器学习方法对memcpy数据搬移操作进行分析，找出所移动的数据值。

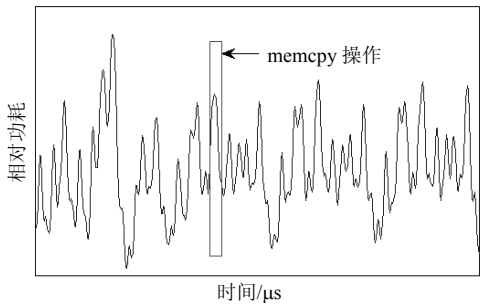

图1　数据搬移操作能量曲线

2.1　模板攻击的结果

利用1.1节介绍的模板攻击算法进行建模和分析，测试的结果如表1所示。表1给出了一次测试使用不同数量的功耗曲线时的相关统计结果。可以看出，若每次测试使用的曲线数量越多，则匹配正确的概率越大。当一次测试使用30条以上的能量曲线，匹配成功的概率可以达到30%以上。当正确值占比排名第二的时候，排名第一的数据值的汉明重量跟

正确值是一样的，这种结果出现的原因是由于汉明重量功耗模型的有效性。另外，实验结果表明在计算功耗曲线和模板的匹配概率时可以将 $\boldsymbol{\Sigma}_i$ 用单位矩阵代替，这样可以避免协方差矩阵的病态问题，而且测试的结果表明这种近似是可以接受的。

表1 模板攻击结果

一次匹配使用曲线条数	正确结果出现次数排名	正确结果占比/%	从第一位到正确结果的占比总和/%
1	6	3.08	95.99
2	4	7.00	92.70
6	3	17.41	81.45
16	2	26.40	64.00
30	2	32.13	70.87
40	2	34.40	76.80

2.2 k近邻算法的结果

通过k近邻算法可以利用已有的能量曲线来直接测试新的能量曲线对应的数据值。在设置不同的 k 值时，得出的结果如表2所示。从表2可以看出，不同的 k 值学习得到的模型会有不同的测试结果。随着 k 值的增大，正确结果出现的次数会逐渐增多，并且在所有测试得到的结果中排到第1位，最好的训练结果接近于 $1/3$，意味着利用相应模型进行一次测试，得到正确数据的概率为 $1/3$，这对应着利用一条能量曲线可以得到的攻击结果。另外，k 值越大并不意味着测试效果越好，例如，$k=101$ 时的结果反而不如 $k=81$。

表2 k 近邻算法的攻击结果

k	正确结果出现次数排名	正确结果占比/%	从第一位到正确结果的占比总和/%
11	2	25.43	62.17
41	2	30.97	62.10
81	1	31.34	31.34
101	1	31.06	31.06
121	1	31.48	31.48

2.3 朴素贝叶斯算法的结果

考虑到不同功耗曲线上同一时间点上的功耗是一个连续值，所以可以对样本特征向量点采用高斯分布假设并使用朴素贝叶斯算法对功耗曲线进行学习和测试。结果见表3。结果表明，高斯型朴素贝叶斯算法的攻击效果不好，正确结果出现的概率只有1/10。

表3 高斯型朴素贝叶斯算法的攻击结果

类型	正确结果出现次数排名	正确结果占比/%	从第一位到正确结果的占比总和/%
高斯分布假设	4	11.62	85.90

2.4 SVM算法的结果

对于SVM算法，本文使用了机器学习软件包scikit-learn中提供的sklearn.svm.SVC和sklearn.svm.LinearSVC类来进行学习和测试[11]。SVC和LinearSVC分别基于开源算法库libsvm和liblinear。Libsvm是一个通用用途的SVM，包含各种核函数，而liblinear实现了快速的线性SVM[12-13]。使用SVC和LinearSVC类中的“一对多”模式进行数据搬移能量曲线的学习和测试。SVC默认使用RBF类型的核函数。由于SVM对数据敏感，不同范围的数据所需的训练时间和结果都会不同，所以将能量曲线同一时间点的值分别变换到 $[0,1]$ 范围再进行训练。注意变换操作应该保存下来，数据也应经过此变换操作才能用于测试。经过训练后的SVM测试结果见表4。可以看出使用线性SVC的效果更好，可以在使用一条功耗曲线进行测试的情况下，正确率达到30%以上。另外，SVC的训练时间是LinearSVC的4倍以上。

表4 SVM的攻击结果

类型	正确结果出现次数排名	正确结果占比/%	从第一位到正确结果的占比总和/%
SVC	4	14.25	85.85
LinearSVC	2	32.68	65.89

3 结 束 语

通过使用机器学习算法对密码芯片中数据搬移操作执行时的电流功耗曲线进行了学习和测试，目的是得到移动的数据值。3种机器学习算法的一次测试都是使用了一条能量曲线，跟模板攻击匹配时也采用一条能量曲线相比，一些机器学习算法的正确率更高。在只能采集少量能量曲线进行攻击的情况下，这些机器学习算法比模板攻击更有优势。其中，k 近邻算法和线性SVM的结果最好，在测试时尽管只使用一条能量曲线，正确率仍然能达到30%以上，这个结果跟模板攻击每次采用30条以上能量曲线进行匹配的正确率很接近。另外，随着一次匹配使用的能量曲线数目增多，模板攻击的效果会更好。如何在机器学习算法一次测试中利用多条能量曲线以及正确率是否有提高是值得研究的问题。

模板攻击也可以看作是一种贝叶斯决策方法[14]，但是效果比朴素贝叶斯好，因为朴素贝叶斯方法采用了不同特征点的概率分布独立这一粗糙假设，而模板攻击在估计特征点的概率分布关系时采用的是多变量正态分布，这是一种更符合实际情况的假设[1]，所以攻击效果更好。

通过利用机器算法对实际密码芯片中的数据搬移能量模板进行攻击，展示了机器学习算法在侧信

道分析中的处理过程，结果表明机器学习算法在某些情况下比传统的模板攻击方法更有效，为继续研究机器学习算法攻击密码设备的其他部分，如带防护措施的密码运算，提供了示例。本文使用的算法都是串行执行的，在学习和测试阶段的数据量都十分巨大的情况下，如何提高机器学习算法的学习效率也是值得进一步研究的问题。

参 考 文 献

[1] MANGARD S, OSWALD E, POPP T. Power analysis attacks: Revealing the secrets of smart cards[M]. New York: Springer, 2007.

[2] CHARI S, RAO J R, ROHATGI P. Template attacks[C]// Cryptographic Hardware and Embedded Systems. Berlin Heidelberg: Springer-Verlag, 2003: 13-28.

[3] MITCHELL T M. Machine learning[M]. New York: McGraw-Hill, 1997.

[4] HARRINGTON P. Machine learning in action[M]. New York: Manning Publications, 2012.

[5] HOSPODAR G, GIERLICHS B, DE MULDER E, et al. Machine learning in side-channel analysis: a first study[J]. Journal of Cryptographic Engineering, 2011, 1(4): 293-302.

[6] LERMAN L, BONTEMPI G, MARKOWITCH O. Power analysis attack: an approach based on machine learning[J]. Int J Applied Cryptography, 2014, 3(2): 97-115.

[7] LERMAN L, BONTEMPI G, MARKOWITCH O. A machine learning approach against a masked AES: Reaching the limit of side-channel attacks with a learning model[J]. Journal of Cryptographic Engineering, 2015, 5(2): 123-139.

[8] LERMAN L, POUSSIER R, BONTEMPI G, et al. Template attacks vs. machine learning revisited (and the curse of dimensionality in side-channel analysis)[C]//Constructive Side-Channel Analysis and Secure Design. Cham: Springer, 2015: 20-33.

[9] MAGHREBI H, PORTIGLIATTI T, PROUFF E. Breaking cryptographic implementations using deep learning techniques[C]//Security, Privacy, and Applied Cryptography Engineering. Cham: Springer, 2016: 3-26.

[10] 崔琦, 王思翔, 段晓毅, 等. 一种AES算法的快速模板攻击方法[J]. 计算机应用研究, 2017, 34(6): 1801-1804.
CUI Qi, WANG Si-xiang, DUAN Xiao-yi, et al. Fast template DPA attack against AES algorithm[J]. Application Research of Computers, 2017, 34(6): 1801-1804.

[11] Scikit-learn. Machine learning in python[EB/OL]. [2017-09-11]. http:// scikit-learn.org/.

[12] CHANG C C, LIN C J. LIBSVM—a library for support vector machines[EB/OL]. [2017-01-30]. https://www.csie.ntu.edu.tw/~cjlin/libsvm.

[13] Machine Learning Group at National Taiwan University. LIBLINEAR — a library for large linear classification [EB/OL]. [2017-01-30]. http://www.csie.ntu.edu.tw/~cjlin/liblinear/.

[14] 刘飚. 基于机器学习的密码芯片电磁攻击技术研究[D]. 北京: 北京邮电大学, 2014.
LIU Biao. The research on the technologies of electromagnetic attack oriented to cryptographic chips based on machine learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.

编　辑　税　红