DOI:10.19651/j.cnki.emt.2210264

伪 RFID 标签信号的谱熵特征提取与选择*

高 威[1] 吴海锋[1,2] 曾 玉[1,2] 普崇荣[1]
(1.云南民族大学电气信息工程学院 昆明 650504;2.云南省高校智能传感器网络及信息系统创新 昆明 650504)

摘 要: 为解决 RFID 通信安全问题,本文提出一种新的物理层标签防伪技术,该技术对原始信号进行信号分离得到期望、噪音和标准化信号,然后对 3 种信号的谱熵特征进行特征提取,最后利用特征选择实现真伪标签分类。另外,本文还提出了一种新的交叉验证来客观测试物理层方法的性能。结果表明,本文方法准确率比传统方法提升近 4%,在新交叉验证下,物理层方法的分类准确率会下降 8%~10%,由此本文得到重要结论,在物理层识别方法中利用谱熵特征实现标签防伪有着重要意义。

关键词: 无线射频识别;防伪;安全;特征选择;交叉验证
中图分类号: TN9 **文献标识码:** A **国家标准学科分类代码:** 510.4

Spectral entropy feature extraction and selection of pseudo RFID tag signals

Gao Wei[1] Wu Haifeng[1,2] Zeng Yu[1,2] Pu Chongrong[1]
(1. School of Electrical and Information Engineering, Yunnan Minzu University, Kunming 650504, China;
2. Innovation of Intelligent Sensor Network and Information System in Yunnan Province University, Kunming 650504, China)

Abstract: In order to solve the problem of RFID communication security, this paper proposes a new physical layer tag anti-counterfeiting technology. This technology separates the original signal to obtain the expected, noise and normalized signals, and extracts the spectral entropy features of the three signals. Finally, feature selection is used to achieve true and false label classification. In addition, this paper proposes a new cross-validation to objectively test the performance of the physical layer method. The results show that the accuracy of the method in this paper is nearly 4% higher than that of the traditional method. Under the new cross-validation, the classification accuracy of the physical layer method will drop by 8~10 percentage points. From this, we get an important conclusion, it is of great significance to use the spectral entropy feature in the physical layer identification method to realize the label anti-counterfeiting.

Keywords: RFID; anti-counterfeiting; security; cross-validation; feature selection

0 引 言

无线射频识别技术(radio frequency identification, RFID)是一种自动识别技术,它通过无线通信对具有唯一身份(identification, ID)信息的电子标签进行数据信息交换[1]。

简单的 RFID 系统可由阅读器、标签和后端数据库组成[2],阅读器通过无线信道向标签发送命令,标签接收到命令后将作出响应发送其 ID 信息,后端数据库可根据 ID 检索相应物品信息。理论上,标签可以响应不同 RFID 阅读器的发送命令,任意拥有阅读器的人员都可以随意读取到缺乏访问控制标签的底层信息,如电子产品码(electronic product code, EPC)[3]由于 RFID 标签的低成本性,读取到的底层信息很容易被克隆到另一标签中,若标签没有设置防克隆的功能,则攻击者就可以对原始标签进行伪造假冒,获取非法利益。基于以上原因,数据的通信安全问题一直是 RFID 研究领域中需要关注的一个重要问题。

在通信安全问题中,对数据加密和使用安全协议[2]是一种较常见的解决方法。然而,RFID 标签的低成本决定了其结构简单、计算能力有限,高性能的加密算法和安全协议会增加标签的复杂性,而使用轻量级加密协议,对于简单的保护方法,一旦密码泄露,数据将被轻易获取,例如 RFID 物流供应链中[4],密码可能在多个环节中被轻易窃取。针对加密和安全协议的问题,一些研究者提出了一些保护标签本身,而不是保护底层数据的方法来提高 RFID 通信的

收稿日期:2022-06-09
* 基金项目:国家自然科学基金(62161052)项目资助

安全性，通过添加一些硬件因素来防御伪造标签，例如利用电感耦合[5-6]或发射天线[7]。然而，使用硬件方式虽提高标签安全性，但会增加标签制作成本。

近年来，大量研究表明电子标签所响应的信号在物理层上有唯一性[8]，具有一个物理不可克隆函数（physical unclonable function，PUF）。由于每个标签的 PUF 均不相同，因此，即使输入相同 PUF 响应也不会相同。更重要的是，PUF 具有不可预测和不可重复性，因此可有效防止标签的伪造，目前已成为一种解决 RFID 通信安全的新方法。基于 PUF 的物理层识别方法对接收到的标签信号提取唯一的特征来作为标签的指纹[8-11]，例如信号的时频统计特性[9-11]。该方法只需在阅读器端增加相应算法，无需在标签端进行加密或更改硬件电路，因此不会提高标签成本，且具有较好的可移植性，适合大规模和低成本标签的防伪场合。

然而，物理层识别方法还有一些问题还需要更进一步研究。首先，从标签信号中提取什么特征未经充分讨论。一些特征值在某几类标签中具有显著性差异，可在其它类中却不一定存在这种差异，如何提取或选择具有显著性差异的特征值是一个需要研究的问题。再次，物理层识别方法的测试方案所得到的结果具有偶然性。传统测试方案使用经典的交叉验证方法[11]来得到真伪标签的分类准确率，但实际工程应用中，攻击者所使用的伪造标签信息也许无法提前掌握，难以在训练库中训练，因此使用这种测试方法所得到的分类准确率并不一定准确。

针对以上问题，本文提出了一种新的提取标签信号特征的方法来提高 RFID 标签安全性，使 RFID 系统免受恶意用户攻击，并且重新设计一种测试方法来评估 RFID 安全性能。在特征提取方法中，不仅直接对阅读器接收到的标签信号提取统计特征，而且经过处理得到了标签本身的 EPC 信号、所携带的噪音信号以及标准化后的接收信号，另外，本文提出了一种新的特征提取方法，对上述 4 种信号进行了交叉谱熵（cross entropy）的统计特征提取。并且，为了得到有效特征，去除冗余特征，本文还使用了特征选择方法。此外，在新的测试方法中，所设计的交叉验证方法更接近工程实际，训练集中所包含的一类伪标签数据，测试集中并不存在该类标签。实验中，本文利用 USRP 软件无线电设备测试了 3 家厂商的 7 类标签，实验结果表明，经 5 倍交叉验证，本文特征提取方法的分类准确率比传统方法提高了 3%～4%。另外，实验还采用了新的测试方法来评判系统的安全性，结果表明，与传统交叉验证相比，无论是传统方法还是本文方法的分类准确率都有所降低。因此，本文可得到一个结论：物理层识别方法在实际应用中，若训练库的伪造数据不完全，防伪性能将会受到影响。

1　相关工作

RFID 安全问题本质上是无线通信安全问题，解决该问题的较常用方法是使用安全协议。在流行的 EPC C1 Gen2 标准中，其通信协议规定可设置密码来对标签的访问进行控制，不过这种协议安全级别较低，一旦密码泄露，数据很容易被窃取。一个完整的 RFID 安全认证协议应该能具有防止标签被跟踪、克隆、窃听和泄密等功能，较成熟的认证协议采用对称和非对称的加密算法[12-14]，它能抵御大部分常见攻击，然而由于这些算法的高复杂度，将其应用于 RFID 必将提高标签成本。为此，一些轻量级的认证协议[15-16]被提出来，这些协议支持随机数和单向散列函数的认证算法[17-18]。然而在应用于 RFID 安全问题时都应有一个折衷。复杂的认证算法更安全，但会提高标签成本，相反，简单的算法更适用于低成本标签，但安全度降低。

若将加密认证算法称为软件保护方法，那么还有一些提升 RFID 安全性能的硬件保护方法，所谓硬件保护方法是利用 RFID 系统的硬件因素抵御攻击的方法。另外一种方法是在不更改标签本身电路的情况下，添加一些额外的硬件因素来实现 RFID 的安全保护。如一些基于电感耦合的方法[4-5]，它们使用一个额外的保留标签放置在一个需要读取标签的附近，耦合效应就会创建一个指纹系统。无论是对于更改标签电路的内置方法还是添加硬件的外置方法，设计者均应考虑新系统的移植性，对系统所修改的硬件是否能适用于原有通用的系统，如 EPC C1 Gen2 协议支持下的 RFID 系统。即使新系统能够兼容原系统，也需要考虑新的硬件是否会带来更高的成本，从而影响其推广。

由于标签在制造过程中存在差异，导致生产出标签的硬件电路也有差异，因此标签在响应的信号上也就体现出差异。根据该原理，可以通过物理层上读取的标签信号来识别真伪标签。根据从物理层信号所提取的特征，可将物理层识别方法分成多个种类。一类是直接采用信号的一些物理量作为特征，例如最小功率响应[19-20]和电压相关量[21]等。另外一类是提取信号上的统计量作为识别的特征，有提取信号时域上的均值、方差、偏度、峰度、自相关等统计量的方法[8-11]，有提取信号频域上的特征的方法[21]，也有提取信号时频特征的方法[8-11]。物理层识别方法的性能严重依赖于所提取的特征，具有显著性差异的特征将会有较好的识别效果。然而，由于标签种类多种多样，不同的特征在不同类的标签上有不同的表现，难以找到统一的特征能够区分所有种类的标签。另外，如何客观地测试物理层识别方法的性能也是一个问题。如前所述，攻击者使用的标签种类也许无法提前预知，使用传统机器学习的交叉验证方法所得到的分类准确率存在一定偏差。

2　问题提出

从机器学习角度看，物理层识别方法所解决的问题本质是一分类问题。在分类问题中，无论采用何种分类器，其性能均依赖所选取的特征。另外，采用何种测试方法对分类的性能评价也非常重要。下面，将对这两个问题展开讨论。

2.1　特征提取与选择

不同厂商的标签在制造过程中所生产的硬件电路会有所差异,因此由这些电路所响应的信号也将有所差异。从标签在物理层所响应信号中提取的特征就是硬件电路差异的体现,提取合适的特征就能分辨不同类标签。然而,标签间差异往往多种多样,例如,信号的相位偏差在某些类标签中存在差异,但在其他类标签中就不存在。当然,可采用多特征联合分类,通过分类器训练后,较重要特征将被赋予较大权重,而较次要特征将被赋予较小权重。然而在这种情况下,确定特征数量仍是要解决的一个问题。从描述分类对象的特性看,特征数量越多越好,因为将尽可能地展现对象的各个特性。但是,当特征数目较多时将不可避免产生冗余,这部分冗余特征不仅在分类时是多余的,而且会降低分类性能[22]。

由于真伪标签的 EPC 码一样,所以标签信号间差异就更多地体现在各自噪音信号,并且各标签归一化后的 EPC 信号也会有所差异,例如频率漂移[23-24]。在实际测试中,本文发现选用较多特征分类器的性能并不比选用较少特征分类器的性能好。另一方面,特征数量更少的分类器,其分类性能也会下降。由此可知,选择特征的数量其实存在一个最优值,如图 1 所示,太多将产生冗余,太少又会导致信息丢失。因此,本文要解决的第一个问题是,应保留哪些有效特征和去除哪些冗余特征。

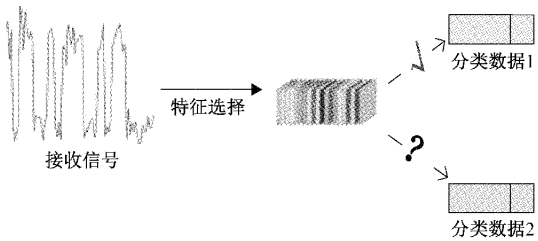

图 1　特征选择问题图

2.2　分类测试

传统测试物理层识别方法的分类性能大多采用经典的交叉验证方法,其中,将每一类数据均随机分成若干份,一部分作为训练集,一部分作为测试集,训练集中具有的数据类型在测试集中也应该有。如图 2 所示给出了一个 5 倍交叉验证示意图,其中类 1 和类 2 分别是真标签和攻击标签。可以看到,训练集中有攻击类标签,测试集中也有该类攻击标签。然而,这种情况在实际应用中不一定能够保证。由于攻击者采用何种标签无法预知,虽然能在训练库中包含尽可能多的各类标签,但也难免缺少攻击者所使用的一类标签。如果采用图 2 的测试模型对物理层的识别方法进行性能评判,得到的结果就不一定准确。既然训练集中已存在攻击标签,那么在测试集中能够识别出该攻击类标签的可能性就较大。从这一个角度看,用传统 5 倍交叉验证的方法来测试物理层的识别方法得到的分类准确率会虚高。因此,本文需要解决的第三个问题是如何采用一个能较准确评判物理层识别方法性能的测试模型。

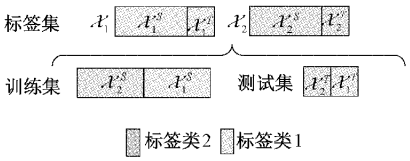

图 2　标签分类的传统交叉验证

3　算　　法

3.1　特征提取

传统方法仅从阅读器接收到的标签响应信号中提取特征,然而,标签的噪音信号、期望的 EPC 信号也会携带其自身信息,因此也可以从上述信号中提取特征。

首先,对阅读器接收到的标签信号进行 IQ 解调,得到 I 路和 Q 路信号,对其求模后得到 $a(n)$,其中 $n=1,2,\cdots N$ 为采样点。然后,对该信号聚类得到聚类中心点向量 $\boldsymbol{V}=[v_0,v_1]^{\mathrm{T}}$,表示为:

$$\boldsymbol{V}=\boldsymbol{clu}[a(n)] \tag{1}$$

其中,$\boldsymbol{clu}[\cdot]$ 为聚类函数,v_0 和 v_1 分别为码元取 0 和 1 时对应的中心点。需要注意的是,聚类后得到的两个中心向量 $\boldsymbol{V}$ 本身并不包含 0 和 1 的信息,然而,标签在发送其 RN16 前会有一段静默期信号[3],该信号在 EPC C1 Gen2 中已被规定,因此可将 EPC 信号与该信号一起聚类,从而确定 v_0 和 v_1。如图 3 所示给出聚类过程,首先由静默期聚类可得到 v_0',再次 EPC 信号聚类得到的两个中心点,离 v_0' 较近的就定为 v_0,另一个就定为 v_1。由聚类中心点,对 $a(n)$ 求欧氏距离判决就得到期望信号:

$$a_e(n)=dec[a(n)] \tag{2}$$

其中,

$$dec(x)=\begin{cases}0, & |x-v_0|<|x-v_1|\\ 1, & |x-v_1|<|x-v_0|\end{cases} \tag{3}$$

理论上,真实标签和克隆标签的期望信号应是一样的,因为两者的 EPC 一样,但是由于存在频率漂移现象,标签间的周期会存在差别,因此从它们提取的相关特征也会存在差别。

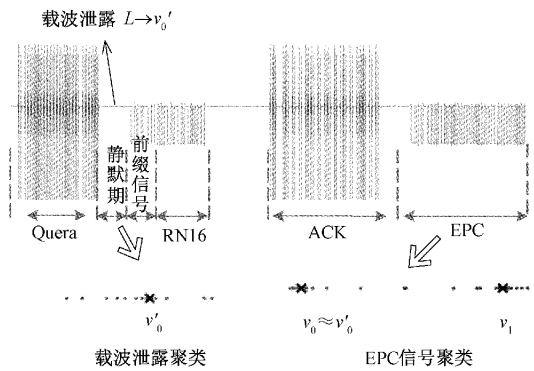

图 3　标签信号的聚类图

再次，由于阅读器在阅读标签时，发射功率、阅读距离和标签灵敏度等不同，标签响应信号的幅度存在差别，为消除这些差别，可将信号标准化为：

$$a_n(n)=\frac{a(n)-v_1}{v_1-v_0} \tag{4}$$

由式(4)得到的信号，大部分采样点均在 0 和 1 附近，因此从该信号提取的相关特征将较少表现出信号的幅度差异。

最后，将标准化信号与期望信号相减就可得到噪音信号，表示为：

$$a_\eta(n)=a_n(n)-a_e(n) \tag{5}$$

若不计频率漂移的差异，各标签信号的 EPC 期望信号应一样，因此标签信号间的差异将更多地体现在由(5)得到的噪音信号上。

经过以上处理，本文将得到期望信号 $a_e(n)$、标准信号 $a_n(n)$ 和噪音信号 $a_\eta(n)$，连同原始接收信号 $a(n)$，一共 4 组信号。

将得到期望信号 $a_e(n)$、标准信号 $a_n(n)$ 和噪音信号 $a_\eta(n)$，连同原始接收信号 $a(n)$，一共 4 组信号。进行交叉谱熵特征提取。交叉谱熵是根据谱熵和维纳-辛钦定理得到，能够有效的表示两个信号之间的相关性程度。

谱熵作为衡量描述了功率谱和熵率之间的关系。表示为：

$$\varepsilon=-\sum_f p_x(f)\log p_x(f) \tag{6}$$

其中，$P_x(f)$ 为各个频段的能量占总能量的比重大小的概率密度函数。

$$p_x(f)=\frac{|P_x(f)|^2}{\sum_f |P_x(f)|^2} \tag{7}$$

由于维纳－辛钦定理，互功率谱密度有两种定义，分别是频域和时域，即：

$$P_{xy}(f)=P_x(f)\overline{P}_y(f)=\mathcal{F}\{R_{xy}(\tau)\} \tag{8}$$

把式(8)代入到式(6)得到交叉谱熵的定义，即：

$$\varepsilon_{xy}=-\sum_f p_{xy}(f)\log p_{xy}(f) \tag{9}$$

交叉谱熵表示两个时间序列的相关性，其相关性越大，独立性越强。交叉谱熵取值在 0～1 之间。为了保障提取特征信息的完整性，又引入新的指标，最大幅度 a_{xy}、最大相位 φ_{xy}、最大频率 $f_{\max}$。其表达式为：

$$a_{xy}=\max|P_{xy}| \tag{10}$$

$$\varphi_{xy}=\max(\angle P_{xy}) \tag{11}$$

$$f_{\max}=\arg\max_f P_{xy}(f) \tag{12}$$

同时，也定义了第二的幅度、相位、和频率。一共 7 个特征。

3.2　交叉验证

对标签真伪识别是一个分类问题，对分类的准确率测试可以采用交叉验证，如前所述，传统交叉验证存在一定偶然性，不能保证总能与实际情况相符。在这儿给出一种更符合实际情况的 K 倍交叉验证方法，其中训练集中具有的攻击类标签，测试集中将不会出现，具体如下。

令 $\boldsymbol{X}_i=[x_i^{(1)},x_i^{(2)},\cdots x_i^{(M)}]$ 为从第 i 个标签信号提取 M 个特征所组成的矢量，将该矢量和其分类标签 y_i 组成一元胞元素 $\mathcal{X}_i=\langle \boldsymbol{X}_i,y_i\rangle$。若第 l 类标签共有 I 个标签，则将其元胞共同构成第 l 类标签集合 $\mathcal{X}_l=\{\mathcal{X}_i\mid i=1,2,\cdots,I\}$。然后，把第 l 类标签集合 $\mathcal{X}_l$ 划分为两个子集 $\mathcal{X}_l^S$ 和 $\mathcal{X}_l^T$，使其满足：

$$\mathcal{X}_l^S\cup\mathcal{X}_l^T=\mathcal{X}_l \tag{13}$$

$$\mathcal{X}_l^S\cap\mathcal{X}_l^T=0 \tag{14}$$

$$|\mathcal{X}_l^T|/|\mathcal{X}_l^S|=(K-1)/K\times 100\% \tag{15}$$

其中，$l=1,2,\cdots L$。对于 L 类标签，若将第 k 类标签作为真标签，第 j 类标签作为测试的攻击标签，其中 $k,j\in\{1,2,\cdots L\}$ 且 $k\neq j$，那么所得到的训练集和测试集可表示为

$$\mathcal{S}_{k,j}=\tilde{\mathcal{X}}_1^S\cup\cdots\mathcal{X}_k^S\cup\cdots\tilde{\mathcal{X}}_{j-1}^S\cup\tilde{\mathcal{X}}_{j+1}^S\cdots\tilde{\mathcal{X}}_L^S \tag{16}$$

$$\mathcal{T}_{k,j}=\mathcal{X}_k^T\cup\mathcal{X}_j^T \tag{17}$$

其中，$\tilde{\mathcal{X}}_m^S$ 为预训练的假标签集，$m=1,2,\cdots L$ 且 $m\neq k$ 和 j，该集合由从 $\mathcal{X}_m^S$ 中随机抽取部分元素构成，即 $\tilde{\mathcal{X}}_m^S\subset\mathcal{X}_m^S$，并使该集合的势 $|\tilde{\mathcal{X}}_m^S|$ 对每个 m 均相同，且

$$\left|\bigcup_m\tilde{\mathcal{X}}_m^S\right|=|\mathcal{X}_k^T| \tag{18}$$

式(16)确保了训练的真标签集大小和训练的假标签集大小相等。由于 $k\neq j$，因此训练集和测试集共有 $A_L^2=L(L-1)$ 种划分，如图 4 所示。由图可知，测试集中的第 j 类攻击标签在训练集中并没有出现，该情形代表攻击者所采用的克隆标签预先未知，因此也不能预先训练。

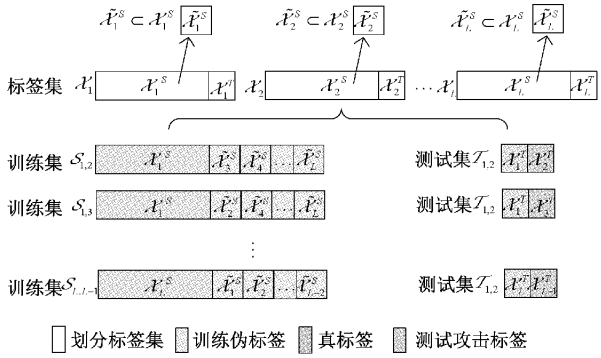

图 4　新型交叉验证

3.3　特征选择

由于在特征提取阶段，所提取特征较多，因此需保留有效特征，去除冗余特征。较常用的一个解决方法是特征选择，特征选择通常分为过滤式、包裹式和嵌入式，后两种特征选择方法涉及到交叉验证的分类模型及分类器。由于本文的交叉验证中，训练集的分类标签和测试集并不一致，采用现存的包裹式和嵌入式特征选择方法易产生过拟合现

象，因此本文将采用过滤式方法。下面，本文将结合所提出的交叉验证，将本文的特征选择做详细介绍。

令训练集 $\mathcal{S}_{i,j}$ 中标签的特征矢量 $\boldsymbol{X}=[x^{(1)},x^{(2)},\cdots,x^{(M)}]$ 和分类标签 y 构成的元胞为 $\chi=\langle\boldsymbol{X},y\rangle$，即 $\chi=\langle\boldsymbol{X},y\rangle\in\mathcal{S}_{i,j}$，则对其特征权重 $\omega_{x^{(m)}}$，$m=1,2,\cdots,M$ 进行排序，并选择权重最大的 W 个特征，表示为：

$$<p_1,p_2,\cdots p_W>=\underset{m}{\operatorname{argmax}}\omega_{x^{(m)}} \tag{19}$$

其中，权重 $\omega_{x^{(m)}}$ 可利用过滤式特征选择方法得到。由式(19)可得到经特征选择后的标签元胞 $\chi^S=\langle\boldsymbol{X}^S,y\rangle$，并得到新的训练集 $\mathcal{S}'_{i,j}$，使其满足：

$$\chi^S=\langle\boldsymbol{X}^S,y\rangle\in\mathcal{S}'_{i,j} \tag{20}$$

其中，$\boldsymbol{X}^S=[x^{(p_1)},x^{(p_2)},\cdots,x^{(p_W)}]$。同理，可得到验证集中标签元胞 $\chi^T=\langle\boldsymbol{X}^T,y\rangle$，并得到新验证集 $\mathcal{T}'_{i,j}$，使其满足：

$$\chi^T=\langle\boldsymbol{X}^T,y\rangle\in\mathcal{T}'_{i,j} \tag{21}$$

其中，$\boldsymbol{X}^T$ 为验证集中权重最大的 W 个特征构成的矢量。

特征选择后，可进行交叉验证。若分类器 $f_{clas}(\cdot)$ 的权重 w 满足：

$$y=f_{clas}(w,\chi^S),\langle\boldsymbol{X}^S,y\rangle\in\mathcal{S}'_{i,j} \tag{22}$$

则训练完成。测试结果由式(23)可得：

$$\hat{y}=f_{clas}(w,\chi^T),\langle\boldsymbol{X}^T,y\rangle\in\mathcal{T}'_{i,j} \tag{23}$$

将测试标签的标签 $\hat{y}$ 与期望标签 y 进行比对，就可得到分类准确率。

在上述特征选择方法中，选择的特征数 W 是影响分类性能的重要参数。若 W 过大，则选择的作用将减小。极端情况 W 等于原特征数，则相当于没有特征选择。若 W 过小，则关键分类信息容易丢失。一种较常用的方法是引入验证集，在测试集中测试不同的 W，将具有最高分类准确率对应的 W 值作为在验证集的值。然而，由前所述，由于本文的交叉验证中，攻击标签的分类信息预先未知，因此不能预训练或预测试，因此采用引入验证集的方法同样会带来过拟合。实际执行过程中，可选取一个中间值，具体参数值的讨论可见实验部分。

算法步骤

输入：

接收到的标签信号 $a(n)$

输出：

测试集标签 $\hat{y}$

已知条件：

特征选择算法

特征选择数目 W

步骤：

1)信号获取：由式(1)～(5)得到期望信号 $a_e(n)$、标准信号 $a_n(n)$ 和噪音信号 $a_\eta(n)$。

2)信号特征提取：由 4.1 得到信号的特征矢量 $\boldsymbol{X}=[x^{(1)},x^{(2)},\cdots x^{(M)}]$。

3)训练集和测试集划分：由式(13)～(17)得到训练集 $\mathcal{S}_{i,j}$ 和测试集 $\mathcal{T}_{i,j}$。

4)特征选择：由式(19)～(21)得到特征选择后的训练集 $\mathcal{S}'_{i,j}$ 和测试集 $\mathcal{T}'_{i,j}$。

5)测试结果：由式(22)～(23)得到测试结果 $\hat{y}$。

4　数据处理

本实验数据来自 EPC C1 Gen2 规定的无源超高频 RFID 标签，采用了市场上常见的 7 个种类的 140 个标签，其中该 7 类标签分别由三家厂商制造，详细情况如表 1 所示。采集数据前对所有 140 个标签写入相同的 EPC 码，采用的读写器为广州网源电子厂商的 UHF100U 读写器，其系统参数如表 2 所示。

表 1　标签型号和厂商

类别	标签型号	厂商
1	Alien9640	广州网源电子
2	Alien9662	深圳骐宝科技
3	Alien9654	深圳骐宝科技
4	Alien9662	广州网源电子
5	Alien7017	广州网源电子
6	Alien9662	南京陆加壹科技
7	Alien9654	南京陆加壹科技

表 2　RFID 写入器系统参数 UHF100 system setting

参数	描述
型号	UHF 100U
工作频率	865～868 MHz 902～928 MHz
支持协议	EPC C1 Gen2
标签与阅读器距离	0～0.1 m
通讯接口	USB
工作电压	DC+5 V
最大功率	4 W

所有标签写入相同 EPC 码后再对标签数据采集，采集由一个 USRP 软件无线电所构建的超高频 RFID 系统[25]完成，该系统遵循 EPC C1 Gen2 标准，软件采用 GNU Radio 实现，详细参数参如表 3 所示，代码下载地址为 https://github.com/nkargas/Gen2-UHF-RFID-Reader。

每次数据采集时，只有一个标签被放置在天线前，其余标签未在阅读系统磁场范围之内，以降低标签冲突风险。所有的数据采集并未在一个隔离环境中进行，即该环境可能包括热噪声，手机噪声，无线网络和射频噪声等。标签随机放置在两个天线[26]构成的一个矩形区域内，如图 5 所示，该区域面积为 25 cm×25 cm。每个标签记录 10 s 的数

表 3　USRP system setting

参数	描述
主板	USRP N200
子板	RXF900
天线	
数量	2
类型	圆极化天线
增益	7 dBic
标签与阅读器距离	0.5～1.5 m
链路频率	40 kHz
最大查询次数	1 000 次
编码	FM0 编码
传输功率	17.8 dBm
发射振幅	0.1
采样频率	1 MHz

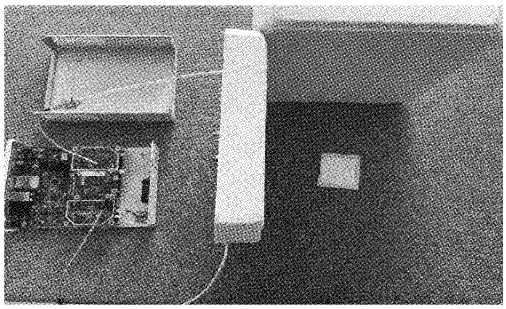

图 5　读取 RFID 标签信号的实验场景图

据，随机截取包含静默期和 EPC 码的信号，并以 MATLAB 格式存储作为要分类的数据，总共得到 140 个标签数据。

本实验采用以下两种 5 倍交叉验证方法来测试分类算法：

1)5 倍交叉验证一：有 L 类标签，若第 k 类标签作为真标签，则分别将其与其余的第 j 类标签分别进行二分类的传统五倍交叉验证，其中 $j \neq k$，如图 6 所示。第 k 类标签的分类准确率将是对各测试集 $\mathcal{T}'_{k,j}j \neq k$ 的二分类结果平均值。

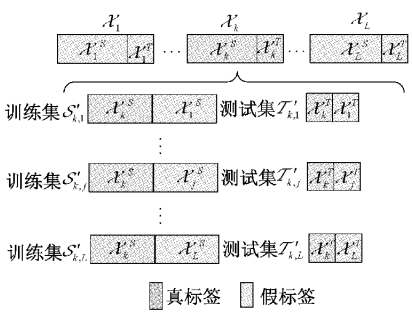

图 6　交叉验证一示意图

2)5 倍交叉验证二：有 L 类标签，若第 k 类标签作为真标签，则分别将其余的第 j 类标签作为攻击标签在测试集中测试，但不在训练集中训练，其中 $j \neq k$，如图 4 所示；第 k 类标签的分类准确率将是对各测试集 $\mathcal{T}_{k,j}j \neq k$ 的二分类结果平均值。

针对不同的特征和特征选择，测试以下几种算法的性能：

1)传统 7 个特征：采用文献[14]的方法，从原始接收信号提取中的 7 个统计量均值，方差，最大自相关，香农熵，第二中心距，偏度和峭度，不进行特征选择，直接进行分类。

2)交叉谱熵 49 个特征：由式(1)～(5)得到期望、标准信号和噪音信号，采用交叉谱熵的方法分别对 4 种信号进行提取 42 个相关性特征，加上原始信号提取中的 7 个统计量，共计 49 个特征，不进行特征选择，直接分类。

3)chi2 特征选择交叉谱熵 12/24/36 个特征：采用表 1 的算法对交叉谱熵特征提取和特征选择，其中特征选择方法采用卡方检验算法，特征选择数为 12，24 或 36，即 $W=12$，24 或 36，其中特征选择算法采用 MATLAB2021a Statistics and Machine Learning Toolbox 的 fscchi2 函数实现。

4)relief 特征选择交叉谱熵 12/24/36 个特征：采用表 1 的算法对交叉谱熵特征提取和特征选择，其中特征选择方法采用 ReliefF 算法，特征选择数为 12，24 或 36，即 $W=12$，24 或 36，其中特征选择算法采用 MATLAB2021a Statistics and Machine Learning Toolbox 的 relieff 函数实现。

5　实验结果和讨论

5.1　预处理结果

由 USRP 软件无线电平台采集到的信号为，阅读器与标签通信的信号经解调后所得到的实部和虚部的采样点信号。对采集到信号的实部和虚部求模，然后再截取一段完整的通信信号，如图 7 所示。在图中，该信号包含了 Query 信号、RN16 信号、ACK 信号和 EPC 信号，与 EPC C1 Gen2 标准一致。对标签进行分类所采用的主要信号为 EPC 段的信号，由式(1)～(5)对该段信号处理分别得到标准化、期

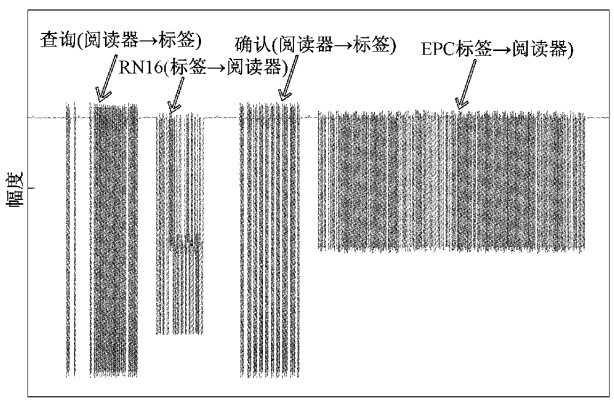

图 7　USRP 采集符合 EPC C1 Gen2 的完整标签响应信号

望和造影信号,如图 8 所示。由图 8 可知,标准化信号的幅值主要集中在 0 和 1 附件,期望信号为二进制信号,噪音信号则是以上两信号的差值。

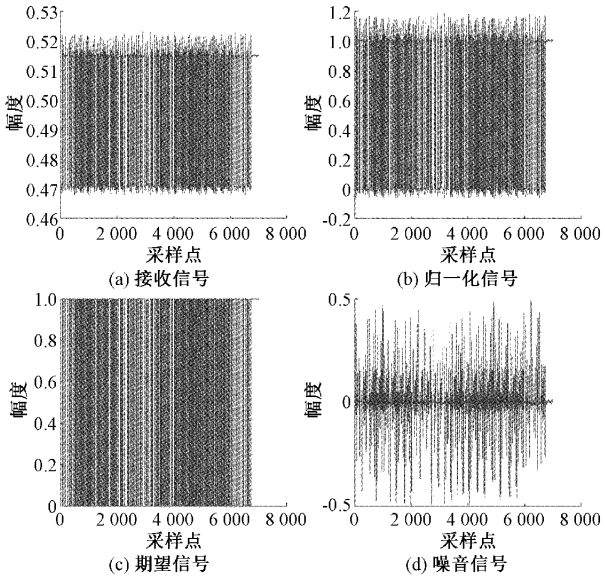

图 8　EPC 段信号经处理后得到的接受信号,标准化信号,期望信号和噪音信号

在传统特征提取方法中,本文先从原始标签响应的 EPC 信号中提取了 7 个常用统计量,包括一阶统计量、二阶统计量甚至熵等统计量。然而,为了能够找到可能的有效特征,或者为了扩大特征可选择的范围,增加了上述期望信号、误差信号和标准化信号的交叉谱熵特征,4 种信号的相关特征。当期望信号在理论上是相同的,通过不同标签在信道环境复杂的情况下,提取每个标签的不同信号的相关特征是合理的。因此本文通过对上述 4 种信号提取交叉谱熵特 42 个特征,再加上原始标签的 7 个特征共计 49 个特征。

5.2　交叉验证一的结果

图 9 给出了在交叉验证一的情况下,采用传统方法 7 个特征,交叉谱熵的 49 个特征的分类准确率曲线图。从图 9 中可以看出,在传统交叉验证的情况下,本文方法的曲线略高于传统特征个特征,除第六类标签外。传统的分类方法下本文的方法仅比 7 个特征的方法高了约 1%,传统交叉验证下未经过特征选择分类准确率提升不太明显。

图 10 给出了在在交叉验证一的情况下。对传统方法 7 个特征进行卡方特征选择的结果图。从图 10 中可以看到,无论选择多少特征,准确率提升也不明显。各方法相差大约只有 0.5%。因此,在没有加入有效特征的情况下,即使经过特征选择,准确率也难以提升。图 11 对 7 个特征采用 ReliefF 方法的结果,结果也同图 10 类似,准确率提升有限。

图 12 给出了给出了在交叉验证一的情况下,采用 ReliefF 的特征选择方法对交叉谱熵做特征选择和传统方法得到的分类准确率曲线图,与高阶统计量特征提取方法类似,除了第六个标签外,分类准确率要高于传统方法,图 13 为采用 chi2 特征选择方法,与 ReliefF 的特征选择结果类似,但第 6 个和第 7 个标签准确率结果略低于 ReliefF 特征选择结果。

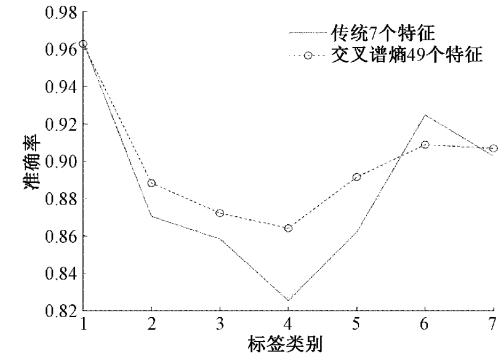

图 9　交叉验证一情况下,未进行特征选择的分类准确率曲线图(标签类别对应表 2)

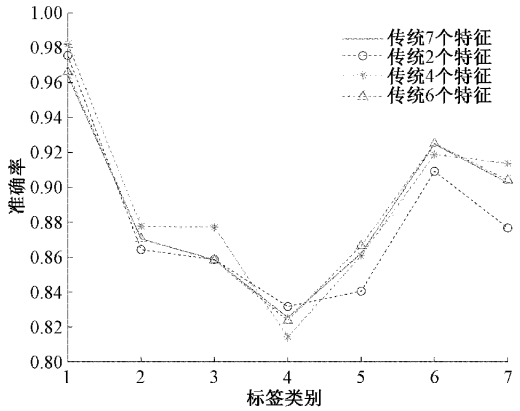

图 10　交叉验证一情况下,对传统算法进行卡方特征选择下对 7 类标签进行分类的分类准确率曲线图(标签类别对应表 2)

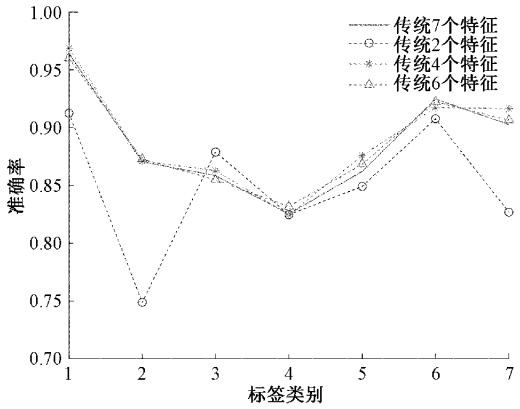

图 11　交叉验证一情况下,对传统算法进行 ReliefF 特征选择下对 7 类标签进行分类的分类准确率曲线图(标签类别对应表 2)

实验结果表明针对不同标签类别进行特征选择时,权重较大或者说显著性差异较大的几类特征的分布会随标签

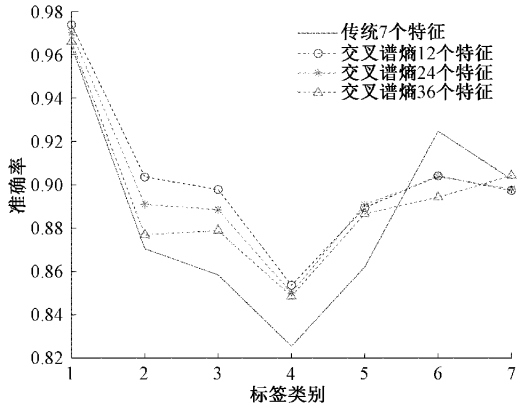

图 12　交叉验证一情况下，chi2 特征选择方法对交叉谱熵特征进行分类的分类准确率曲线图(标签类别对应表 2)

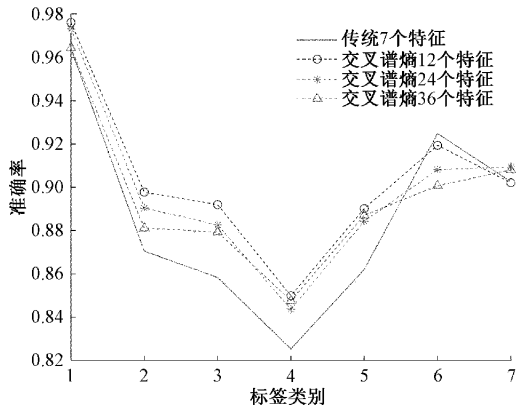

图 13　交叉验证一情况下，ReliefF 特征选择方法对交叉谱熵特征进行分类的分类准确率曲线图(标签类别对应表 2)

类别变化而变化。例如，深圳骐宝科技 Alien9662 标签，其显著特征是接收信号信号的最大自相关特征，而对于南京陆加壹科技 Alien9654 标签，显著特征就变为均值。因此，试图找到几类固定的特征就能识别所有标签类别并不实际。从这一角度看，特征选择方法并不是选择某几个固定特征，而是依靠训练数据选择特征，特征选择结果是由训练数据的结果而定，只要有合适的训练数据就能选择到合适的特征。

另外，实验采取了卡方和 ReliefF 两种特征选择方法去测试方法的性能，主要目的是验证本文所提的方法是否一定要依赖于特定的特征选择方法。实验结果表明，只要进行特征选择，分类准确率均有所提高。

5.3　交叉验证二结果

图 14 给出了在交叉验证一和交叉验证二下采用两种特征选择方法与传统方法的分类准确率曲线对比图，从图中可以看到，无论是否经过特征选择，交叉验证二的结果低于交叉验证一结果 8%左右。

最后，图 15 给出了交叉验证一和交叉验证二下本文特征提取方法的平均分类准确率直方图。传统方法在新的交叉验证情况下，其平均分类准确率要低 8%，本文的方法在

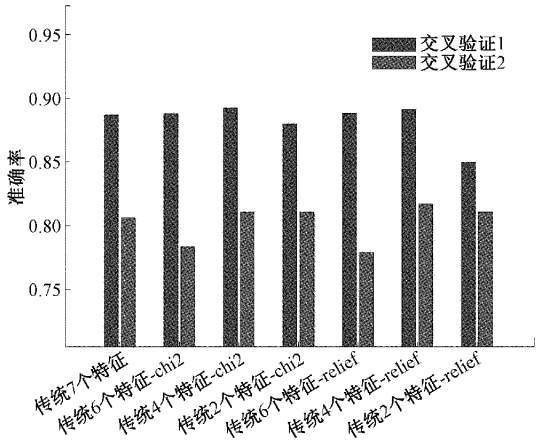

图 14　两种交叉验证情况下特征选择方法的分类准确率

新的交叉验证下，平均分类准确率也要低 10%。而且本文的交叉谱熵特征提取结果在交叉验证二的情况下比传统方法要提高 3%～4%。在面对未知标签的情况下有很好的分类准确率结果。

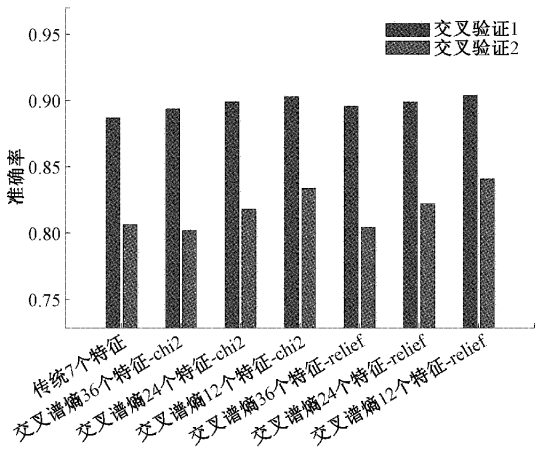

图 15　交叉验证对比平均图

实验结果表明，与传统交叉验证相比，所提取的特征无论是否经过特征选择，其分类准确率均会下降，下降了约 8%～10%。这一结果也在情理之中，因为毕竟攻击标签的特征并未在训练集中学习到。它给了重要启示：任何基于分类学习的物理层标签真伪识别方法如果要获得较好的分类性能，训练库中应可能包含尽可能多的标签类别，一旦攻击标签所在类别未在训练库中，那么假阴发生的概率就会增高。

6　结　　论

本文用特征选择方法去解决标签分类中特征问题，试图在不同的标签类别中也能找到有效的特征去提高分类准确率。本文设计了一种训练集中无攻击类标签的交叉验证，在这种新验证情况下评估特征选择的方法甚至是无特征选择的识别方法将非常必要，因为如果分类准确率下降，

那么意味着这些方法性能存在虚高的可能。同时本文的特征选择在攻击标签所在类别未在训练集的交叉验证中有着很好的结果,当选择特征数目为 14 时,比传统方法要提高到 4%。当通过本文方法特征作为信号分类,在面对未知标签的攻击下有着较好的分类性能。当然,本文的一些实验结果存在一定的不足。由于实验条件的限制,本文所采用的 USRP 设备的发射功率有限,标签只能在阅读器较小的磁场范围内被阅读到,然而,实际伪造标签攻击的环境可能较复杂,不仅标签响应信号幅度有所变化,信噪比也会更小。因此,希望在将来的工作中,将标签识别放在一种情况更为复杂多变的条件下进行。

参考文献

[1] GIANMARCO B, GARY S, FRANC D, et al. Experimental identification of smartphones using fingerprints of built-In micro-electro mechanical systems(MEMS)[J]. Sensors(Basel, Switzerland), 2016, 16(6):818.

[2] IBRAHIM A, DALKLC G. Review of different classes of RFID authentication protocols[J]. Wireless Networks, 2019, 25: 961-974.

[3] GLOBAL E P C. EPC radio-frequency identity protocols class-1 generation-2 UHF RFID protocol for communications at 860 MHz~960 MHz[J]. Version, 2008, 1: 23.

[4] LIN C,HSU H,CHENG Y. A cloud-based authentication protocol for RFID supply chain systems[J]. Journal of Network & Systems Management, 2015, 23: 978-997.

[5] XU H, YIN X, ZHU F, et al. An enhanced secure authentication scheme with one more tag for RFID systems[J]. IEEE Sensors Journal, 2021, 21(15): 17189-17199.

[6] WANG G, CAI H, QIAN C, et al. Hu-Fu: Replay-resilient RFID authentication [J]. IEEE/ACM Transactions on Networking, 2020(99):1-14.

[7] ESSAM G, SHEHATA H, KHATTAB T, et al. Novel hybrid physical layer security technique in RFID systems [C]. 2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC), 2019.

[8] REHMAN S, SOWERBY W, COGHILL C. Analysis of impersonation attacks on systems using RF fingerprinting and low-end receivers[J]. Journal of Computer and System Sciences, 2014, 80(3):591-601.

[9] HAN J, CHEN Q, YANG P, et al. GenePrint: Generic and accurate physical-layer identification for UHF RFID tags[J]. IEEE/ACM Transactions on Networking, 2016, 24(2):846-858.

[10] MEHMOOD A, AMAN W, RAHMAN M, et al. Preventing identity attacks in RFID backscatter communication systems: A physical-layer approach [C]. 2020 International Conference on UK-China Emerging Technologies (UCET),IEEE, 2020: 1-5.

[11] BERTONCINI C, RUDD K, NOUSAIN B, et al. Wavelet fingerprinting of radio-frequency identification (RFID) tags[J]. IEEE Transactions on Industrial Electronics, 2012, 59(12):4843-4850.

[12] WEI J, LI N. Privacy-preserving and undeniable authentication for mobile RFID tags[C]. 2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), IEEE, 2019.

[13] FARASH S, NAWAZ O, MAHMOOD K, et al. A provably secure RFID authentication protocol based on elliptic curve for healthcare environments[J]. Journal of Medical Systems, 2016, 40(7):165.

[14] IBRAHIM A, DALKILIC I. An advanced encryption standard powered mutual authentication protocol based on elliptic curve cryptography for RFID, proven on WISP[J]. Journal of Sensors, 2017, 2017:1-10.

[15] KAPS J P. Chai-tea, cryptographic hardware implementations of xTEA [C]. Progress in Cryptology-INDOCRYPT 2008: 9th International Conference on Cryptology in India, Kharagpur, India, December, 2008: 14-17.

[16] HOSSEINZADEH M, AHMED H, AHMED H, et al. An enhanced authentication protocol for RFID systems[J]. IEEE Access, 2020,8:126977-126987.

[17] WANG S, LIU S, CHEN D. Security analysis and improvement on two RFID authentication protocols[J]. Wireless Personal Communications, 2015, 82(1):21-33.

[18] TODD M, BURLESON W, TESSIER R. The design and assessment of a secure passive RFID sensor system[C]. 2011 IEEE 9th International New Circuits and Systems Conference, IEEE, 2011: 494-497.

[19] CAI H, WANG G, SHI X, et al. When tags 'read' each other: Enabling low-cost and convenient tag mutual identification[C]. 2019 IEEE 27th International Conference on Network Protocols(ICNP), IEEE, 2019.

[20] GOUISSEM A, ABUALSAUD K, YAACOUB E, et al. Hybrid physical layer security for passive RFID communication[C]. 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 2020.

[21] ALSAIFY B A, THOMPSON R. Exploiting hidden Markov models in identifying passive UHF RFID tags[C].

Radio & Wireless Symposium, IEEE, 2014.

[22] 郑宝芬,苏宏业,罗林.无监督特征选择在时间序列数据挖掘中的应用[J].仪器仪表学报,2014,35(4):834-840,DOI:10.19650/j.cnki.cjsi.2014.04.017.

[23] 彭章友,任秀方,孟春阳,等.UHF RFID密集标签互耦效应的频移特性研究[J].电子测量技术,2015,38(6):11-15.

[24] 何怡刚,佘培亮,佐磊,等.高频RFID密集标签系统频率偏移预估研究[J].电子测量与仪器学报,2018,32(11):139-146,DOI:10.13382/j.jemi.2018.11.019.

[25] WU H, WU X, LI Y, et al. Collision resolution with FM0 signal separation for short-range random mul-ti-access wireless network[J]. IEEE Transactions on Signal and Information Processing Over Networks, 2021,7:438-450.

[26] 朱文强,张爱军.一种小型化超高频段RFID读写器天线设计[J].国外电子测量技术,2019,38(11):142-146,DOI:10.19652/j.cnki.femt.1901603.

作者简介

高威,硕士研究生,主要研究方向为信号与信息处理。

吴海锋(通信作者),教授,博士,主要研究方向为机器学习、通信与信息系统。

曾玉,博士研究生,主要研究方向为信号与信息处理。

普崇荣,硕士研究生,主要研究方向为信号与信息处理。