基于二元域等效的 RS 码编码参数盲识别

刘 杰[1],张立民[1],钟兆根[2]
(1. 海军航空大学信息融合研究所,山东烟台 264001;2. 海军航空大学电子信息工程系,山东烟台 264001)

摘 要: 现代数字通信中常常进行信道编码识别处理. 目前 RS(Reed-Solomon,RS)码盲识别需对高阶域下所有谱分量进行求取,计算较为复杂,因此提出了一种基于二元域等效的识别方法. 首先根据有限域性质将 RS 码等效为二元域上的线性分组码,然后建立码长、信息分组长度、生成多项式和本原多项式的关联模型. 通过遍历各阶本原多项式,并验证二元线性分组码的校验向量,完成各参数的联合识别. 仿真结果和理论分析表明,该方法在提升抗误码性能的同时有效减少了计算量,可用于智能通信和通信侦察等系统中.

关键词: 智能通信; 通信侦察; 信道编码; RS 码; 盲识别; 有限域
中图分类号: TN911.22 **文献标识码:** A **文章编号:** 0372-2112 (2018)12-2888-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2018.12.010

Blind Parameter Identification of RS Code Based on Binary Field Equivalence

LIU Jie[1], ZHANG Li-min[1], ZHONG Zhao-gen[2]
(1. *Institute of Information Fusion, Naval Aeronautical University, Yantai, Shandong* 264001, *China*;
2. *Department of Electronics and Information Engineering, Naval Aeronautical University, Yantai, Shandong* 264001, *China*)

Abstract: Channel coding identification is often required in modern digital communications. As the existing blind identification methods for RS code are complicated for the reason that all spectral components in high-order domain need to be calculated, a method based on binary equivalence is proposed. Firstly, RS code is equivalent to binary linear block code according to finite field properties, with the establishment of the association model between code length, information block length, generator polynomial and primitive polynomial. Then, the various parameters of RS code are identified by traversing primitive polynomials and verifying parity check vectors of the binary linear block code. Simulation results and theoretical analysis show that the proposed method can improve error-resilient performance and reduce computation cost as well, which proves its availability in intelligent communication and communication reconnaissance systems.

Key words: intelligent communication; communication reconnaissance; channel coding; RS code; blind identification; finite field

1 引言

为保证数据传输稳定性,信道编码技术被广泛应用于无线通信、深空通信和卫星通信等领域. 在智能通信中,接收方需根据接收数据判断系统采用的调制编码形式。在通信侦察中,需根据截获信号完成信道编码参数和相关协议的识别,进而获取原始信息. 因此,信道编码盲识别技术应运而生,并迅速成为国内外的研究热点之一[1,2].

RS(Reed-Solomon,RS)码是一种常用的多进制线性分组码,具有纠错能力强和编码结构简单等特点. 针对 RS 码的识别,现有方法主要包括基于矩阵分析的方法[3~5]、基于欧几里得的方法[6]和基于伽罗华域傅里叶变换(Galois field Fourier transform,GFFT)的方法[7~11]. 文献[5]统计分析矩阵秩的变化规律,进而估计码长等参数. 文献[6]通过欧几里德算法计算码字的最大公约式,并对其分解以得到生成多项式. 上述方法的不足之处在于,对误码的适应性较差. 文献[7]在 GFFT 后根据频谱累积量的分布估计编码参数. 文献[8]对 GFFT 谱分量进行非线性变换和中值滤波,以区分零频分量. 文

收稿日期:2017-08-25;修回日期:2017-11-08;责任编辑:覃怀银
基金项目:国家自然科学基金重大研究计划(No. 91538201);泰山学者工程专项经费(No. st201511020)

献[9]在 GFFT 后引入平方欧几里德测度,用于衡量谱分量概率分布的差异性. 文献[10]分析了 GFFT 的最优判决门限,以提升识别性能. 文献[11]首先利用 Gröbner 基估计码长,以降低 GFFT 过程的运算量. 基于 GFFT 的方法抗误码性能好,但存在以下不足:首先,该方法基于符号码元,而实际 RS 码均以比特形式传输. 因此,需要先将截获的比特序列映射为符号码元序列再进行处理. 在本原多项式未知的情况下,需要对映射关系进行多次验证,而现有文献并未对此进行分析. 其次,该方法需要对编码域中所有元素进行检验,当码长较大时,计算量急剧增加.

针对以上不足,本文提出了一种基于二元域等效的 RS 码识别方法,通过对等效二进制分组码的校验向量进行判决,完成码长和信息分组长度等参数的联合识别,进而利用连续公共码根计算生成多项式. 仿真结果表明,相比传统的识别方法,在保证识别性能的同时有效降低了计算复杂度.

2　RS 码盲识别问题描述

定义 1[12]　GF(q)($q\neq 2$)上生成多项式 $g(x)$ 包含 $\alpha^{l_0},\alpha^{l_0+1},\cdots,\alpha^{l_0+2t-1}$ 等 $2t$ 个连续根的本原 BCH 码称为 RS 码.

RS 码一般用数学符号表示为(n,k) RS 码,其中 n 表示码长,k 表示信息分组长度. 在 RS 码作为纠错编码的应用中,均取 $q=2^m$,且 $3\leqslant m\leqslant 8$. 对 GF(2^m)上纠 t 个错误的 RS 码,其存在以下特点:

(1)码长 n 和信息分组长度 k 满足

$$n=2^m-1 \tag{1}$$

$$k=2^m-2t-1 \tag{2}$$

(2)码元和生成多项式的根均取自 GF(2^m),对于 $g(x)$ 的根 $\alpha^{l_0+i}(0\leqslant i\leqslant 2t-1)$,其最小多项式 $\phi_i(x)=x+\alpha^{l_0+i}$,从而生成多项式可表示为

$$g(x)=\prod_{i=0}^{2t-1}\phi_i(x)=\prod_{i=0}^{2t-1}(x+\alpha^{l_0+i}) \tag{3}$$

通常情况下取 $l_0=1$,且 α 为 GF(2^m)上的本原元. 此时,$g(x)=x^{2t}+g_{2t-1}x^{2t-1}+\cdots+g_1x+g_0$,其中 $g_i\in$ GF(2^m),$0\leqslant i\leqslant 2t-1$.

设信息序列多项式为 $u(x)$,经信道传输产生的误码多项式为 $e(x)$,则实际截获编码序列可表示为

$$v(x)=u(x)\cdot g(x)+e(x) \tag{4}$$

对 RS 码进行盲识别,就是在仅知道 $v(x)$ 的情况下,估计编码参数进而还原出信息序列. 若假定通过帧同步处理已经获得(n,k) RS 码的起始位置,则需要识别的编码参数包括码长 n,信息分组长度 k 和生成多项式 $g(x)$.

3　基于二元域等效的 RS 码盲识别方法

3.1　RS 码的二元域等效

GF(2^m)是 GF(2)的扩域,可通过多项式剩余类环 $F_2[x]/p(x)$ 构造得到,其中 $p(x)$ 为 m 阶本原多项式,它对应本原元 α. 根据文献[12],GF(2^m)的所有元素均可以用 GF(2)上的 m 维二元向量表示. 因此,GF(2^m)上的(n,k) RS 码可以等价为 GF(2)上的一个(mn,mk)线性分组码. RS 码的校验矩阵一般表示为

$$\boldsymbol{H}=\begin{bmatrix}\alpha^{n-1} & \alpha^{n-2} & \cdots & \alpha & 1\\ \alpha^{2(n-1)} & \alpha^{2(n-2)} & \cdots & \alpha^2 & 1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \alpha^{2t(n-1)} & \alpha^{2t(n-2)} & \cdots & \alpha^{2t} & 1\end{bmatrix} \tag{5}$$

可以看出,$\boldsymbol{H}$ 的每一行对应生成多项式 $g(x)$ 的一个根. 令向量 $\boldsymbol{a}=(\alpha^{m-1},\alpha^{m-2},\cdots,\alpha,1)^{\mathrm{T}}$,其中 T 表示转置. 将 $\boldsymbol{H}$ 中的元素分别与向量 $\boldsymbol{a}$ 相乘,并把结果中的元素转化为 m 维二元行向量. 如果用$(\cdot)_2$ 表示这一过程,则与 RS 码等价的 GF(2)上的(mn,mk)线性分组码校验矩阵为

$$\boldsymbol{H}'=\begin{bmatrix}\boldsymbol{H}'_1\\ \boldsymbol{H}'_2\\ \vdots\\ \boldsymbol{H}'_r\\ \vdots\\ \boldsymbol{H}'_{2t}\end{bmatrix}$$

$$=\begin{bmatrix}(\alpha^{n-1}\boldsymbol{a})_2 & (\alpha^{2(n-1)}\boldsymbol{a})_2 & \cdots & (\alpha^{2r(n-1)}\boldsymbol{a})_2 & \cdots & (\alpha^{2t(n-1)}\boldsymbol{a})_2\\ (\alpha^{n-2}\boldsymbol{a})_2 & (\alpha^{2(n-2)}\boldsymbol{a})_2 & \cdots & (\alpha^{2r(n-2)}\boldsymbol{a})_2 & \cdots & (\alpha^{2t(n-2)}\boldsymbol{a})_2\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ (\alpha\boldsymbol{a})_2 & (\alpha^2\boldsymbol{a})_2 & \cdots & (\alpha^{2r}\boldsymbol{a})_2 & \cdots & (\alpha^{2t}\boldsymbol{a})_2\\ (\boldsymbol{a})_2 & (\boldsymbol{a})_2 & \cdots & (\boldsymbol{a})_2 & \cdots & (\boldsymbol{a})_2\end{bmatrix}^{\mathrm{T}} \tag{6}$$

其中,$\boldsymbol{H}'_r=(((\alpha^{2r(n-1)}\boldsymbol{a})_2)^{\mathrm{T}},((\alpha^{2r(n-2)}\boldsymbol{a})_2)^{\mathrm{T}},\cdots,((\alpha^{2r}\boldsymbol{a})_2)^{\mathrm{T}},((\boldsymbol{a})_2)^{\mathrm{T}})$,$1\leqslant r\leqslant 2t$.

经以上过程,RS 码的识别问题就转化为等价的(mn,mk)二元线性分组码的识别问题,当矩阵 $\boldsymbol{H}'$ 为该二元线性分组码的校验矩阵时,对应 $\boldsymbol{H}$ 也为 RS 码的校验矩阵.

3.2　等效二元线性分组码的识别

由第 2 节可知,码长 n 和信息分组长度 k 取决于本原多项式阶数 m 和纠错个数 t,而计算生成多项式 $g(x)$ 需要知道本原元 α 和 $g(x)$ 的连续根,即本原多项式 $p(x)$ 和纠错个数 t. 因此,决定识别的关键在于本原多项式 $p(x)$ 和纠错个数 t. 根据文献[13],各阶数下的本原多项式分布如表 1 所示,其中本原多项式取值用十进

制表示,例如 $11=2^3+2+1$ 表示 $p(x)=x^3+x+1$. 相应的,整体的识别思路是:首先根据表 1 遍历 $p(x)$,按相应参数划分编码序列,并对 $g(x)$ 的根,即式(6)中的 $\boldsymbol{H}'_r$ 进行检验,当通过判决的连续根个数为偶数时,证明所选取的 $p(x)$ 正确,进而可得到 t;然后,将 m 和 t 依次带入式(1)、(2),计算码长 n 和信息分组长度 k,将 $g(x)$ 的连续根带入式(3),并利用 $p(x)$ 化简,得到生成多项式 $g(x)$. 下面对具体的遍历过程进行说明.

表 1 各阶数下的本原多项式

m 值	n 值	本原多项式 $p(x)$
3	7	11,13
4	15	19,25
5	31	37,41,47,55,59,61
6	63	67,91,97,103,109,115
7	127	131,137,143,145,157,167,171,185,191,193,203,211,213,229,239,241,247,253
8	255	285,299,301,333,351,355,357,361,369,391,397,425,451,463,487,501

从接收序列编码起点开始,按长度 mn 对其划分并构造 $L\times mn$ 阶矩阵 $\boldsymbol{V}$,然后在选取的本原多项式 $p(x)$ 下,将式(5)中 $\boldsymbol{H}$ 第 r 行 $\boldsymbol{h}_r$ 转化为其二元矩阵形式 $\boldsymbol{H}'_r$. 令 $\boldsymbol{h}'_{r,i}$ 为 $\boldsymbol{H}'_r$ 第 i 行,$1\leqslant i\leqslant m$,当 $\boldsymbol{h}'_{r,i}$ 为该(mn,mk)线性分组码的校验向量时,矩阵 $\boldsymbol{V}$ 中每一行都对应该分组码一个码字,此时 $\boldsymbol{V}\cdot(\boldsymbol{H}'_r)^{\mathrm{T}}=0$,即 $\boldsymbol{h}'_{r,i}$ 属于 $\boldsymbol{V}$ 的对偶空间 $V^{\perp}$. 实际传输中由于误码的存在,$\boldsymbol{V}\cdot(\boldsymbol{H}'_r)^{\mathrm{T}}=0$ 不一定成立,又 m 个 $\boldsymbol{h}'_{r,i}$ 彼此相互独立,因此我们先考虑 $\boldsymbol{V}\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0$ 的概率. 当对任意 i 值,$\boldsymbol{V}\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0$ 均以很大的概率成立时,则可以认为 $\boldsymbol{V}\cdot(\boldsymbol{H}'_r)^{\mathrm{T}}=0$ 成立.

令误比特率为 ε,$\boldsymbol{h}'_{r,i}$ 汉明重量为 $\omega_{r,i}$,b_λ 为 $\boldsymbol{h}'_{r,i}$ 中第 $\lambda(1\leqslant\lambda\leqslant w_{r,i})$ 个 1 所在的位置,$\boldsymbol{v}_j=(v_{j,1},v_{j,2},\cdots,v_{j,mn})$ 为矩阵 $\boldsymbol{V}$ 第 $j(1\leqslant j\leqslant L)$ 行. 当 $\boldsymbol{h}'_{r,i}\in V^{\perp}$ 时,若 $\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0$,则 $\{v_{j,b_\lambda}\}$ 中无错误或有偶数个错误比特;反之,则 $\{v_{j,b_\lambda}\}$ 中有奇数个错误比特. 因此

$$\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0]=\sum_{d=0}^{\lfloor\omega_{r,i}/2\rfloor}C_{\omega_{r,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{r,i}-2d} \tag{7}$$

$$\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=1]=1-\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0] \tag{8}$$

其中,$\Pr(\cdot)$ 表示概率,$\lfloor\cdot\rfloor$ 表示向下取整. 当 $\boldsymbol{h}'_{r,i}\notin V^{\perp}$ 时,$\boldsymbol{v}_j$ 和 $\boldsymbol{h}'_{1,i}$ 可以看成是两个相互独立的 mn 维随机向量,从而

$$\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=1]=\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0]=\frac{1}{2} \tag{9}$$

令 $\boldsymbol{y}_{r,i}=\boldsymbol{V}\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}$,将 $\boldsymbol{y}_{r,i}$ 中的 1 映射为 -1,0 映射为 1,得到 $\hat{\boldsymbol{y}}_{r,i}=(\hat{y}_{r,i}^{(1)},\hat{y}_{r,i}^{(2)},\cdots,\hat{y}_{r,i}^{(L)})^{\mathrm{T}}$,然后把 $\hat{\boldsymbol{y}}_{r,i}$ 中的元素十进制相加,即

$$y_{r,i}=\sum_{j=1}^{L}\hat{y}_{r,i}^{(j)} \tag{10}$$

可以看出,$y_{r,i}$ 的值越大,$\boldsymbol{V}\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0$ 成立的概率越大. 由于矩阵 $\boldsymbol{V}$ 各行相互独立,因此 $\hat{\boldsymbol{y}}_{r,i}$ 中各元素也相互独立,且 $\hat{y}_{r,i}^{(j)}$ 服从伯努利分布. 根据独立同分布情况下的中心极限定理[14],当 L 足够大时,$y_{r,i}$ 趋于高斯分布 $\mathrm{N}(L\mu,L\sigma^2)$,其中 μ 和 σ^2 分别表示 $\hat{y}_{r,i}^{(j)}$ 的均值和方差。若用 $\mathrm{E}(\cdot)$ 表示期望,则有

$$\begin{aligned}\mu&=\mathrm{E}(\hat{y}_{r,i}^{(j)})=\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0]\times1+\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=1]\times(-1)\\&=\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0]-\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=1]\end{aligned} \tag{11}$$

$$\begin{aligned}\sigma^2&=\mathrm{E}[(\hat{y}_{r,i}^{(j)})^2]-[\mathrm{E}(\hat{y}_{r,i}^{(j)})]^2\\&=4\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=0]\cdot\Pr[\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{r,i})^{\mathrm{T}}=1]\end{aligned} \tag{12}$$

因此,根据 $\boldsymbol{h}'_{r,i}$ 是否属于矩阵 $\boldsymbol{V}$ 的对偶空间 $V^{\perp}$,有

$$\begin{aligned}y_{r,i}&\sim\mathrm{N}(0,L\sigma_1^2),\boldsymbol{h}'_{r,i}\notin V^{\perp}\\y_{r,i}&\sim\mathrm{N}(L\mu_2,L\sigma_2^2),\boldsymbol{h}'_{r,i}\in V^{\perp}\end{aligned} \tag{13}$$

其中,

$$\sigma_1^2=1 \tag{14}$$

$$\mu_2=2\sum_{d=0}^{\lfloor\frac{\omega_{r,i}}{2}\rfloor}C_{\omega_{r,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{r,i}-2d}-1 \tag{15}$$

$$\begin{aligned}\sigma_2^2=4\sum_{d=0}^{\lfloor\frac{\omega_{r,i}}{2}\rfloor}&[C_{\omega_{r,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{r,i}-2d}\\&\cdot(1-\sum_{d=0}^{\lfloor\frac{\omega_{r,i}}{2}\rfloor}C_{\omega_{r,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{r,i}-2d})]\end{aligned} \tag{16}$$

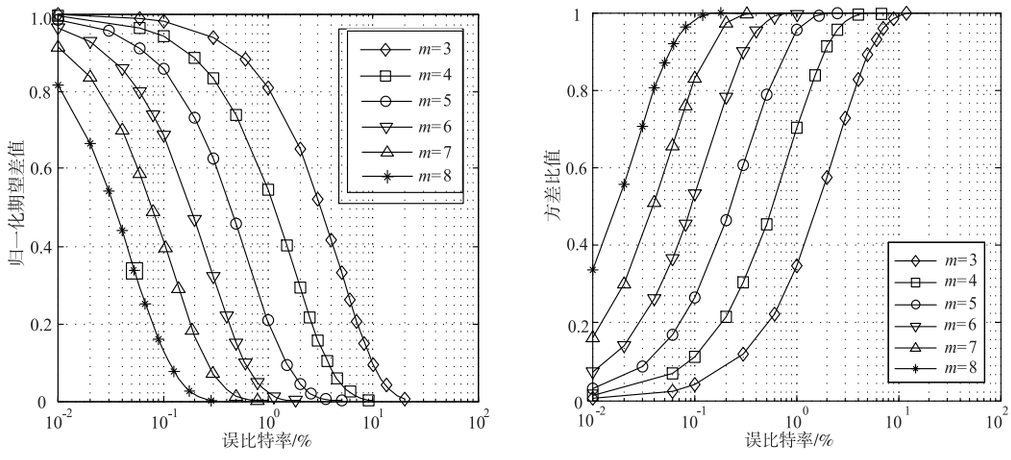
通过设置合适的门限,就能根据 $y_{r,i}$ 值判断对应的 $\boldsymbol{h}'_{r,i}$ 是否属于矩阵 $\boldsymbol{V}$ 的对偶空间. 经分析可知,两个分布的差异越大,识别概率越大. 当 $\boldsymbol{h}'_{r,i}\in V^{\perp}$ 时,$y_{r,i}$ 分布主要取决于参数 m 和误比特率 ε,为此计算不同参数组合下两个分布的归一化期望差值和方差比值,即

$$\Delta\mu=\frac{L\mu_2-0}{L}=\mu_2 \tag{17}$$

$$\Delta\sigma^2=\frac{\sigma_2^2}{\sigma_1^2} \tag{18}$$

为了简化计算,取 $\omega_{r,i}=\lfloor mn/2\rfloor$,最终结果如图 1 所示. 不难看出,在固定 m 值下,误比特率越大,两个分布越接近,识别也越困难. 由于 $\boldsymbol{H}'_1$ 对应本原元 α,而 α 对应本原多项式,因此进行两次判决以提升高误比特率下的识别概率. 首先设定一个较低的门限 T_1 对 $\boldsymbol{H}'_1$ 进行初步判决,避免本原多项式的漏检;然后再设定一个较高的门限 T_2 筛选正确结果,并确定生成多项式其它根.

设门限 T_1 对应虚警概率为 $P_{\text{fa}}^{(1)}$. 若对 $\boldsymbol{H}'_1$ 中任意 $\boldsymbol{h}'_{1,i}(1\leqslant i\leqslant m)$,均有 $y_{1,i}>T_1$,则暂时可认定 α 为 RS 码编码域的本原元,此时再进行第二次判决以进一步确定本原多项式 $p(x)$ 是否选取正确. 因此,在第一次判决下,$p(x)$ 非真实而被误判为正确的概率为



(*a*) 归一化期望差值　　(*b*) 方差比值

图1　两个分布的归一化期望差值和方差比值

$$P_e=(P_{\text{fa}}^{(1)})^m \tag{19}$$

若选择门限 $T_1=0$,则对应虚警概率 $P_{\text{fa}}^{(1)}=50\%$. 又 m 取值不小于 3,故此时

$$P_e=(0.5)^m\leqslant(0.5)^3=0.125 \tag{20}$$

以此可以排除大部分非真实的本原多项式 $p(x)$. 然后仅保留 $\boldsymbol{V}$ 中对任意 $i(1\leqslant i\leqslant m)$ 均满足 $\boldsymbol{v}_j\cdot(\boldsymbol{h}'_{1,i})^{\mathrm{T}}=0$ 的行,得到一个 $\tilde{L}\times mn$ 的矩阵 $\tilde{\boldsymbol{V}}$. 若通过第一次判决的 $p(x)$ 是真实值,则此时 $\tilde{\boldsymbol{V}}$ 中错误比特得到有效剔除;反之,$\tilde{\boldsymbol{V}}$ 中行向量相对于 $\boldsymbol{h}'_{r,i}$ 仍可以看作随机向量. 此时设置一个更高的门限 T_2,按相同的方式检验矩阵 $\boldsymbol{H}'_r(r\geqslant 2)$,对得到的 $\tilde{y}_{r,i}$ 按门限 T_2 进行判决. 若采用文献[8]中的不可能事件概率 0.00135 作为虚警概率,则门限为 $T_2=3\sqrt{\tilde{L}}$,当 $\tilde{y}_{r,i}\geqslant T_2$ 时,可认为 $\boldsymbol{h}'_{r,i}\in V^{\perp}$,反之则 $\boldsymbol{h}'_{r,i}\notin V^{\perp}$. 若对任意 $i(1\leqslant i\leqslant m)$ 均有 $\boldsymbol{h}'_{r,i}\in V^{\perp}$,则称 $\boldsymbol{H}'_r$ 通过判决,并可以确定 α^r 为生成多项式 $g(x)$ 的根. 实际识别时,为减小计算量,并不需要将 RS 码校验矩阵 $\boldsymbol{H}$ 完全转化为二元形式 $\boldsymbol{H}'$,而是每检验一行 $\boldsymbol{h}_r$ 则转化得到一个 $\boldsymbol{H}'_r$. 如果从 $\boldsymbol{H}'_1$ 开始通过判决的连续 $\boldsymbol{H}'_r$ 个数为 R,且 R 为偶数,则说明通过第一次判决的 $p(x)$ 选取正确,此时再进行后续计算处理.

综上,基于二元域等效的 RS 码盲识别流程如图 2 所示.

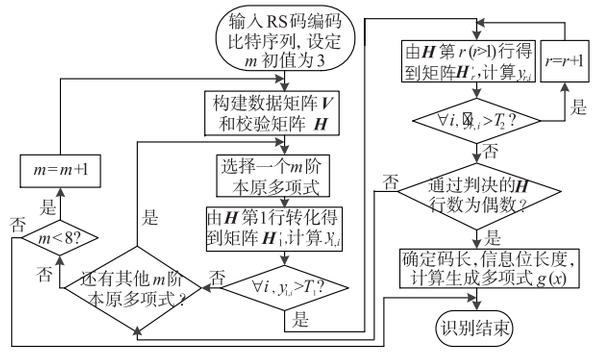


图2 RS码识别流程图

4　仿真验证与分析

4.1　仿真验证

以(63,57)RS 码为例进行仿真,此时 $m_0=\log_2(63+1)=6$,选取本原多项式为 $p_0(x)=x^6+x+1$,生成多项式为 $g_0(x)=x^6+\alpha^{59}x^5+\alpha^{48}x^4+\alpha^{43}x^3+\alpha^{55}x^2+\alpha^{10}x+\alpha^{21}$. 首先生成 1000 组码字,将其转化为二进制比特流形式,并加入误比特率为 $\varepsilon=0.005$ 的错误比特,然后按图 2 所示的流程进行识别. 对本原多项式 $p(x)$ 进行遍历,结果如表 2 所示. 可以看出,当 $m=6$,$p(x)=x^6+x+1$ 时,所有 $y_{1,i}$ 值均大于门限 T_1. 因此,选取这一组参数进行后续第二次判定.

剔除矩阵 $\boldsymbol{V}$ 中的含错码字后,按相同步骤依次处理 $\boldsymbol{H}$ 其它行转化得到的二元矩阵 $\boldsymbol{H}'_r$,分别计算 m 个 $\tilde{y}_{r,i}$ 值并进行判定,结果如表 3 所示. 结合表 2 可以看出,从 $\boldsymbol{H}'_1$ 开始共有连续 6 个 $\boldsymbol{H}'_r$ 通过判决,进而可以确定 $p(x)=x^6+x+1$ 为正确的本原多项式. 此时,生成多项式 $g(x)$ 的连续根为 $\alpha,\alpha^2,\alpha^3,\alpha^4,\alpha^5,\alpha^6$,因此 $g(x)=(x+\alpha)(x+\alpha^2)(x+\alpha^3)(x+\alpha^4)(x+\alpha^5)(x+\alpha^6)=x^6+\alpha^{59}x^5+\alpha^{48}x^4+\alpha^{43}x^3+\alpha^{55}x^2+\alpha^{10}x+\alpha^{21}$,与真实结果相同. 将 $m=6$ 和 $t=3$ 带入式(1)、(2),可得到码长 $n=2^6-1=63$,信息分组长度 $k=2^6-6-1=57$,识别正确.

表 2 第一次判决结果

m 值	本原多项式	$y_{1,i}$值	门限 T_1	判决结果
3	11	$y_{1,1}=30, y_{1,2}=110, y_{1,3}=-16$	0	×
	13	$y_{1,1}=96, y_{1,2}=16, y_{1,3}=-72$		×
4	19	$y_{1,1}=-120, y_{1,2}=-18, y_{1,3}=112, y_{1,4}=6$		×
	25	$y_{1,1}=-58, y_{1,2}=-16, y_{1,3}=80, y_{1,4}=-108$		×
5	37	$y_{1,1}=36, y_{1,2}=14, y_{1,3}=-82, y_{1,4}=-34, y_{1,5}=-36$		×
	41	$y_{1,1}=-66, y_{1,2}=4, y_{1,3}=-30, y_{1,4}=-22, y_{1,5}=-10$		×
	47	$y_{1,1}=-8, y_{1,2}=-24, y_{1,3}=16, y_{1,4}=-34, y_{1,5}=-110$		×
	55	$y_{1,1}=46, y_{1,2}=94, y_{1,3}=70, y_{1,4}=-28, y_{1,5}=-8$		×
	59	$y_{1,1}=18, y_{1,2}=42, y_{1,3}=44, y_{1,4}=-72, y_{1,5}=-30$		×
	61	$y_{1,1}=-46, y_{1,2}=60, y_{1,3}=-14, y_{1,4}=12, y_{1,5}=-8$		×
6	67	$y_{1,1}=170, y_{1,2}=134, y_{1,3}=138, y_{1,4}=160, y_{1,5}=98, y_{1,6}=156$		√

表 3 第二次判决结果

矩阵 $\boldsymbol{H}'_r$	$\tilde{y}_{r,i}$值	门限 T_2	判决结果
$\boldsymbol{H}'_2$	$\tilde{y}_{2,1}=144, \tilde{y}_{2,2}=144, \tilde{y}_{2,3}=146, \tilde{y}_{2,4}=144, \tilde{y}_{2,5}=146, \tilde{y}_{2,6}=144$	37.2	√
$\boldsymbol{H}'_3$	$\tilde{y}_{3,1}=146, \tilde{y}_{3,2}=144, \tilde{y}_{3,3}=144, \tilde{y}_{3,4}=144, \tilde{y}_{3,5}=144, \tilde{y}_{3,6}=144$		√
$\boldsymbol{H}'_4$	$\tilde{y}_{4,1}=144, \tilde{y}_{4,2}=148, \tilde{y}_{4,3}=146, \tilde{y}_{4,4}=144, \tilde{y}_{4,5}=144, \tilde{y}_{4,6}=146$		√
$\boldsymbol{H}'_5$	$\tilde{y}_{5,1}=144, \tilde{y}_{5,2}=146, \tilde{y}_{5,3}=144, \tilde{y}_{5,4}=144, \tilde{y}_{5,5}=146, \tilde{y}_{5,6}=144$		√
$\boldsymbol{H}'_6$	$\tilde{y}_{6,1}=144, \tilde{y}_{6,2}=144, \tilde{y}_{6,3}=146, \tilde{y}_{6,4}=146, \tilde{y}_{6,5}=144, \tilde{y}_{6,6}=144$		√
$\boldsymbol{H}'_7$	$\tilde{y}_{7,1}=-6, \tilde{y}_{7,2}=14, \tilde{y}_{7,3}=8, \tilde{y}_{7,4}=22, \tilde{y}_{7,5}=-14, \tilde{y}_{7,6}=4$		×

4.2 性能分析

首先对所需数据量进行分析. 由于矩阵 $\boldsymbol{V}$ 的行数 L 越大,式(13)中近似得到的高斯分布越准确,当接收数据量无穷大时,理论上总是能完成识别. 但在实际应用环境下,截获数据量总是有限的. 因此,需要对不同误比特率下识别所需的数据量进行分析.

根据统计学上“3 倍标准差”准则,在第一次判决中为保证 $p(x)$ 正确时的 $y_{1,i}$ 均通过判决,需满足

$$L\mu_2 - 3\sqrt{L}\sigma_2 \geqslant 0 \tag{21}$$

整理得

$$L \geqslant \frac{36\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{1,i}-2d}\left(1-\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{1,i}-2d}\right)}{\left(2\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d}\varepsilon^{2d}(1-\varepsilon)^{\omega_{1,i}-2d}-1\right)^2} \tag{22}$$

因此第一次判决所需数据量

$$Q_1 = mnL \tag{23}$$

设 H_0 表示在误比特率 ε 下,码字中无错误;H_1 表示在相同误比特率下,码字中包含错误比特;D_0 表示 m 个 $y_{1,i}$ 均通过第一次判决,则矩阵 $\tilde{\boldsymbol{V}}$ 中码字包含错误比特的概率为

$$\begin{aligned}\tilde{P}_w &= \frac{\Pr(D_0|H_1)\Pr(H_1)}{\Pr(D_0|H_1)\Pr(H_1)+\Pr(D_0|H_0)\Pr(H_0)} \\ &= \frac{2^{-m}P_w}{2^{-m}P_w+1-P_w} = \frac{P_w}{P_w+2^m(1-P_w)}\end{aligned} \tag{24}$$

其中 $P_w = 1-(1-\varepsilon)^{mn}$ 表示矩阵 $\boldsymbol{V}$ 中码字含有错误的概率. 设矩阵 $\tilde{\boldsymbol{V}}$ 的等效误比特率为 $\tilde{\varepsilon}$,则

$$\tilde{P}_w = 1-(1-\tilde{\varepsilon})^{mn} \tag{25}$$

带入式(24)可得

$$\tilde{\varepsilon} = 1-\sqrt[mn]{\frac{2^m(1-\varepsilon)^{mn}}{1+(2^m-1)(1-\varepsilon)^{mn}}} \tag{26}$$

将 $\tilde{\varepsilon}$ 值带入式(15)、(16)即可得到 $\hat{y}_{r,i}^{(j)}$ 新的均值 $\tilde{\mu}_2$ 和方差 $\tilde{\sigma}_2^2$. 根据式(13),在判决门限 T_2 下,虚警概率 $P_{\mathrm{fa}}^{(2)}$ 和检测概率 $P_{\mathrm{de}}^{(2)}$ 分别为

$$P_{\mathrm{fa}}^{(2)} = \int_{T_2}^{\infty}\frac{1}{\sqrt{2\pi\tilde{L}}\sigma_1}\exp\left(-\frac{z^2}{2\tilde{L}\sigma_1^2}\right)\mathrm{d}z \tag{27}$$

$$P_{\mathrm{de}}^{(2)} = \int_{T_2}^{\infty}\frac{1}{\sqrt{2\pi\tilde{L}}\tilde{\sigma}_2}\exp\left(-\frac{(z-\tilde{L}\tilde{\mu}_2)^2}{2\tilde{L}\tilde{\sigma}_2^2}\right)\mathrm{d}z \tag{28}$$

其中,$\exp(\cdot)$表示以自然常数为底的指数函数。根据“3 倍标准差”准则,为了获得较高的检测概率,应保证

$$\frac{T_2-\tilde{L}\tilde{\mu}_2}{\sqrt{\tilde{L}}\tilde{\sigma}_2} = \frac{3\sqrt{\tilde{L}}-\tilde{L}\tilde{\mu}_2}{\sqrt{\tilde{L}}\tilde{\sigma}_2} \leqslant -3 \tag{29}$$

整理后可得

$$\tilde{L} \geqslant \frac{\left(6\sqrt{\left(\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d} \tilde{\varepsilon}^{2d}(1-\tilde{\varepsilon})^{\omega_{1,i}-2d}\right)\left(1-\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d} \tilde{\varepsilon}^{2d}(1-\tilde{\varepsilon})^{\omega_{1,i}-2d}\right)}+3\right)^2}{\left(2\sum_{d=0}^{\lfloor \omega_{1,i}/2 \rfloor} C_{\omega_{1,i}}^{2d} \tilde{\varepsilon}^{2d}(1-\tilde{\varepsilon})^{\omega_{1,i}-2d}-1\right)^2} \tag{30}$$

因此第二次判决所需比特数据量

$$Q_2 \geqslant \frac{mn\tilde{L}}{(1-\varepsilon)^{mn}} \tag{31}$$

综上,整个识别过程所需数据量

$$Q = \max(Q_1, Q_2) \tag{32}$$

其中,max(·)表示取最大值.

图 3 给出了 m 取值为 3 ~ 8 时在不同误比特率下完成识别所需的数据量. 可以看出,在相同数据量下,m 越小,误码适应能力越好;在相同误比特率下,m 越小,识别所需数据量越小.

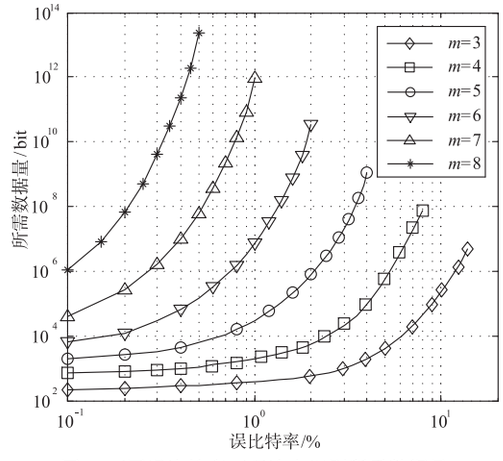


图3　不同误比特率下理论上识别所需数据量

下面分析 m 取值为 3 ~ 8 时在不同误比特率下的识别率,选取的码型分别是(7,3)RS 码、(15,11)RS 码、(31,27)RS码、(63,57)RS码、(127,119)RS码和(255,239)RS 码. 每种码型生成 1000 组码字,然后按不同误比特率加入错误比特,并在每种误比特率下进行 500 次蒙特卡洛仿真,结果如图 4 所示. 可以看出,当误比特率小于 0.001 时,对 6 种 RS 码的识别概率都能达到 90% 以上. 结合上面分析可知,当误比特率更高时,可以通过增加数据量来进一步提高识别率.

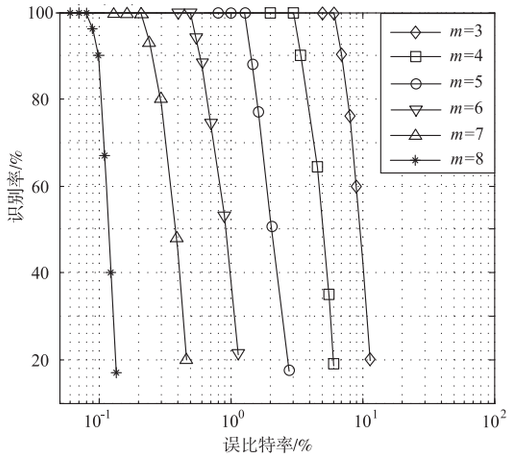


图4　RS码识别概率曲线

4.3　与其他识别方法的对比

首先将本文方法所需数据量与文献[5]中基于矩阵分析的方法、文献[10]中基于 GFFT 的方法和文献[11]中基于 Gröbner 基改进的 GFFT 方法进行对比,m 取值为 3 ~ 8,仍然选取与图 4 中相同的 6 种 RS 码进行研究,且每种 RS 码对应误比特率分别为 0.02、0.01、0.005、0.002、0.001 和 0.0004. 本文方法在各种条件下所需数据量可由式(32)获得;对于文献[5]中方法,其分析矩阵必须满足行数大于列数,因此所需数据量至少为 m^2n^2;基于 GFFT 的方法至少需要 50 组完整码字,相应的数据量为 $50mn$. 最终,得到对比结果如表 4 所示. 可以看出,相同条件下本文方法所需数据量更小.

表 4　四种方法所需数据量比较(bit)

m 值	3	4	5	6	7	8
本文方法	596	1873	5229	1.21×10^4	3.94×10^4	7.56×10^4
文献[5]方法	441	3600	2.40×10^4	1.43×10^5	7.90×10^5	4.16×10^6
文献[10]方法	1050	3000	7750	1.89×10^4	4.45×10^4	1.02×10^5
文献[11]方法	1050	3000	7750	1.89×10^4	4.45×10^4	1.02×10^5

下面对识别所需计算量进行比较. 设实际码长值 $n_0 = 2^{m_0} - 1$,为了统一,将所有方法的计算量均转化为模 2 加运算量. 文献[5]中进行单次高斯列消元的计算量为 $ab(b-1)/4$,其中 a、b 分别为分析矩阵的行数和列数,因此总计算量为 $\sum_{m=3}^{m_0} amn(mn-1)/4$;文献[10]中计算单个码字频谱需要 $n^2 - n$ 次 GF(2^m)加法和 n^2 次 GF(2^m)乘法,转化为模 2 加计算量为 $\sum_{m=3}^{m_0} M[3m(m$

$-1)n^2+mn(n-1)]$,其中 M 表示码字个数;文献[11]中基于 Gröbner 基进行分量码处理的运算量为 mn 次 GF(2^m)加法,转化为模2加运算量为 m^2n 次,因此加上 GFFT 处理的总计算量为 $\sum_{m=3}^{8} m^2n + M[3m_0(m_0-1)n_0^2 + m_0n_0(n_0-1)]$;根据第3节分析可知,本文方法最多所需计算量为 $\sum_{m=3}^{m_0} L\phi_m m(2mn-1)+(2t-1)\tilde{L}m_0(2m_0n_0-1)$,其中 ϕ_m 表示阶数为 m 的本原多项式个数. 利用各方法对前面的6种 RS 码进行识别,其计算量变化曲线如图5所示. 可以看出,当 $m\geqslant 5$ 时,本文方法的计算量小于三种对比方法,且 m 值越大,优势越明显,表明本文方法更适用于码长较大情况下的识别.

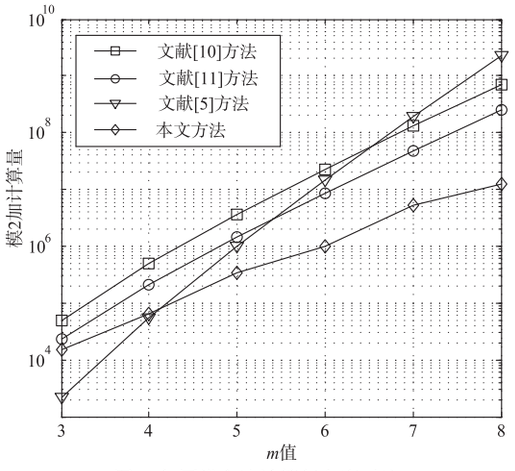


图5 与同类方法计算复杂度对比

最后将本文方法与其它三种方法进行误码适应能力对比. 选取(7,3)和(255,239)两种 RS 码,根据表4结果,为保证矩阵分析法和基于 GFFT 的方法能完成操作,分别生成1050bit 和 4.16×10^6bit 数据,加入错误比特后按不同方法进行识别. 在每组 m 值、误比特率和识别方法组合下进行500次蒙特卡洛仿真,结果如图6所示. 可以

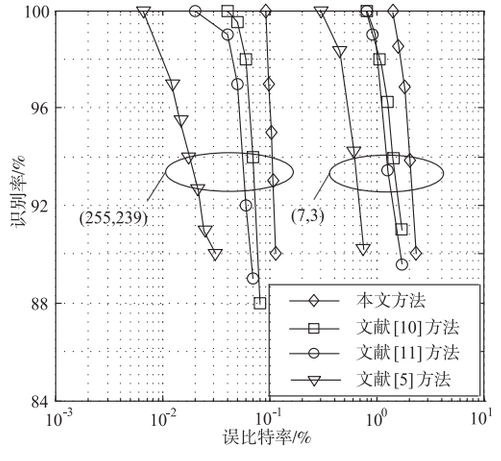


图6 与同类方法识别性能对比

看出,本文识别方法性能明显优于其它三种方法. 综合前面识别所需数据量和计算量对比结果,所提方法具有一定的先进性.

5 结论

本文提出了一种基于二元域等效的 RS 码识别方法,能有效完成码长、信息分组长度和生成多项式的识别. 方法遍历本原多项式,并随之构建二元域分析矩阵和校验向量对其进行检验. 为了提升识别性能,先后进行两次判决,进而确定本原多项式、码长和信息分组长度. 最后,利用连续根分布计算生成多项式. 该方法性能优良、计算量低,易于工程实现. 后续研究将主要针对缩短 RS 码,以进一步完善方法的适用范围.

参考文献

[1] YU P D, PENG H, LI J. On blind recognition of channel codes within a candidate set[J]. IEEE Communications Letters, 2016, 20(4): 736 – 739.

[2] YARDI A D, VIJAYAKUMARAN S, KUMAR A. Blind reconstruction of binary cyclic codes from unsynchronized bit stream[J]. IEEE Transactions on Communications, 2016, 64(7): 2693 – 2706.

[3] KARIMIAN Y, ATTARI M A. Recognition of channel encoder parameters from intercepted bitstream[A]. SHAFIEE M. IEEE 21st Iranian Conference on Electrical Engineering[C]. Mashhad, Iran: IEEE, 2013. 1 – 5.

[4] ZRELLI Y, GAUTIER R, RANNOU E. Blind identification of code word length for non-binary error-correcting codes in noisy transmission[J]. Eurasip Journal on Wireless Communications & Networking, 2015, 2015(1): 1 – 16.

[5] LI T, MIAO C L, and LV J. An improved algorithm of RS codes blind recognition[J]. Applied Mechanics and Materials, 2014, 603 – 605(2014): 2308 – 2312.

[6] 戚林,郝士琦,李今山. 基于有限域欧几里德算法的 RS 码识别[J]. 探测与控制学报,2011,33(2):63 – 67.
QI Lin, HAO Shi-qi, LI Jin-shan. Recognition method of RS codes based on Euclidean algorithm in Galois field[J]. Journal of Detection & Control, 2011, 33(2): 63 – 67. (in Chinese)

[7] XIE H, WANG F H, HUANG Z T. Blind recognition of Reed-Solomon codes based on histogram statistic of Galois field spectra[J]. Advanced Materials Research, 2013, 791 – 793: 2088 – 2091.

[8] 解辉,王丰华. 基于频谱预处理的 RS 码盲检测识别方法[J]. 宇航学报,2013,34(1):128 – 132.
XIE Hui, WANG Feng-hua. Blind detection and recognition of RS code based on spectral preprocessing[J]. Journal of Astronautic, 2013, 34(1): 128 – 132. (in Chinese)

[9] 包昕,陆佩忠,游凌. 基于伽罗华域傅里叶变换的 RS 码识别方法[J]. 电子科技大学学报,2016,45(1):30 – 35.
BAO Xin, LU Pei-zhong, YOU Ling. Recognition of RS coding based on Galois field Fourier transform[J]. Journal of University of Electronic Science and Technology of China,2016,45(1):30 – 35. (in Chinese)

[10] ZHANG X K, WU G, ZHANG B N, et al. Blind recognition of RS codes based on Galois field Fourier transform [A]. LU D. International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery[C]. Washington, DC, USA: CPS, 2016. 429 – 433.

[11] LU O X, GAN L, and LIAO H S. Blind reconstruction of RS codes[J]. Asian Journal of Applied Sciences, 2015, 8(1):37 – 45.

[12] MACWILLIAMS F J, SLOANE N A. The Theory of Error-Correcting Codes[M]. New York, USA: North-Holland Publishing Company, 1981. 294 – 295.

[13] 朱联祥,李荔. RS 码的盲识别方法研究[J]. 电子测量与仪器学报,2013,27(8):781 – 786.
ZHU Lian-xiang, LI Li. Research on blind recognition for RS code[J]. Journal of Electronic and Instrument, 2013, 27(8):781 – 786. (in Chinese)

[14] SOONG T T. Fundamentals of Probability and Statistics for Engineers [M]. Chichester, England: John Wiley &Sons Ltd, 2004. 199 – 201.

作者简介

刘　杰　男,1990 年出生于湖北省宜昌市. 博士. 主要研究方向为通信信号处理技术及应用.
E-mail:iamliu1573@163.com

张立民(通信作者)　男,1966 年出生于辽宁省开原市. 教授,博士生导师. 主要研究领域为卫星信号处理、视景仿真与虚拟现实.
E-mail:iamzlm@163.com

钟兆根　男,1984 年出生于江西省南昌市. 博士,讲师. 主要研究方向为通信侦察和空间信息对抗.