PN 序列估计与扩频隐藏信息分析

谢春辉,程义民,陈扬坤
(中国科学技术大学电子科学与技术系,安徽合肥 230027)

摘 要: 在非协作信息侦测情况下,提出了一种直接序列扩频(DS-SS)信号 PN 序列的估计方法,在此基础上实现了扩频隐藏信息的盲提取.该方法以估计序列和扩频信号的累积相关值为目标函数,建立估计序列长度及其构成的两变量优化模型,通过遗传算法求解,可较好地估计出 PN 序列.通过获取的 PN 序列与藏密信号的相关性分析,可实现扩频隐藏信息的盲提取.该方法已在微机上进行了实验,实验结果表明,该方法在截获信号信噪比较低情况下仍具有较好的估计性能,提取的秘密信息具有较好的可辨认性.

关键词: 直接序列扩频; PN 序列; 扩频隐藏; 隐藏分析; 遗传算法
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2011) 02-0255-05

PN Sequence Estimation and Spread-Spectrum Steganalysis

XIE Chun-hui, CHENG Yi-min, CHEN Yang-kun
(*Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, Anhui* 230027, *China*)

Abstract: An efficient method that can estimate PN sequence of direct sequence spread spectrum (DS-SS) signal is presented in a non-cooperative context, on this base, spread-spectrum hidden message can be extracted blindly. At first, the optimization model with the cumulative correlation value between estimate sequence and SS signal as the objective function is established, and then the PN sequence can be estimated by Genetic Algorithm. Finally, by analyzing the correlation between PN sequence and stego-signal, the hidden messmate can be extracted. The method has been carried out on PC, the experimental results show that, PN sequence estimation performance of the low-SNR DS-SS signal is good, and the extracted hidden message can be identificated.

Key words: direct sequence spread-spectrum; pseudo-noise sequence; spread-spectrum steganography; steganalysis; generic algorithm

1 引言

直接序列扩频(Direct Sequence Spread-Spectrum, DS-SS)通信具有低截获率和抗干扰特性,使直扩信号的检测和盲估计成为现代通信侦测的研究难点之一.扩频通信接收端利用 PN 序列解扩接收信息,因此,估计 PN (Pseudo-Noise)序列是非协作信息截获的关键.

2004 年,C Boudert 等人通过对分帧信号协方差矩阵的特征分析实现 PN 序列重构[1];2006 年,Jiang L 等人提出了一种基于二阶循环统计量的 PN 序列估计方法[2];2008 年,Chen Y 等人以最大似然估计方法为基础,采用禁忌搜索算法实现 PN 序列估计[3].

扩频隐藏[4,5]采用扩频技术,将秘密信息通过 PN 序列扩展后,隐藏在载体感知重要成份之中,可以增强隐藏系统的抗干扰性能,是一种常用的信息隐藏手段[6].随着扩频隐藏方法的广泛应用,针对扩频隐藏信息的侦测及提取成为隐藏分析[7]领域的研究热点之一.2003 年,Chandramouli 等提出了一种基于盲信号分离的扩频隐藏信息的提取方法,提取正确率可达 70%[8];2006,Saeed 等在已知扩频序列长度的前提下,实现了扩频隐藏信息的提取[9].

本文首先提出了一种非协作情况下的 DS-SS 信号 PN 序列的盲估计方法,该方法采用遗传算法,以估计序列与扩频信号的累积相关值作为目标函数,通过求解 PN 序列长度及构成的两变量优化模型,获得 PN 序列的最佳估计.在此基础上通过对文献[9]的改进,实现了扩频序列长度未知时扩频隐藏信息的盲提取.

2 PN 序列估计

2.1 扩频

设待传输的信息码序列 $X=\{x(i)=\pm 1, 0\leqslant i\leqslant R-1\}$, R 为信息码长度,若采用长度为 N 的 PN 序列 $P=\{p(i)=\pm 1, 0\leqslant i\leqslant N-1\}$,对 X 扩频后得到扩频信号 W(见图 1),其长度为

$$W=\{w(i\cdot N+j)=x(i)p(j), 0\leqslant i\leqslant R-1, 0\leqslant j\leqslant N-1\} \quad (1)$$

传输过程中受到高斯白噪声 G 的污染,得到信号 S

$$S=W+G \quad (2)$$

收稿日期:2009-10-17;修回日期:2010-11-05

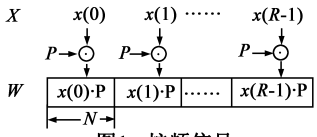


图1 扩频信号

2.2 估计序列与扩频信号的相关性分析

若伪随机序列 P' 是 P 的估计序列,其长度为 N'. 为了描述问题的方便,设 $\overleftrightarrow{P}'$ 是 P' 的周期扩展序列,$\overleftrightarrow{p}'_i$ 表示第 i 个周期序列,$\overleftrightarrow{p}'_i(j)=\overleftrightarrow{p}'(i\cdot N'+j)$ 表示 $\overleftrightarrow{p}'$ 第 i 个周期第 j 个比特

$$\overleftrightarrow{p}'_i(j)=\overleftrightarrow{p}'(i\cdot N'+j)=p'_j \tag{3}$$

同样,设 $\overleftrightarrow{p}$ 是 PN 序列 P 的周期扩展序列,$\overleftrightarrow{p}_i$ 表示第 i 个周期序列,$\overleftrightarrow{p}_i(j)=\overleftrightarrow{p}(i\cdot N+j)$ 表示 $\overleftrightarrow{p}$ 第 i 个周期第 j 个比特

$$\overleftrightarrow{p}_i(j)=\overleftrightarrow{p}(i\cdot N+j)=x(i)\cdot p(j) \tag{4}$$

由于 P' 的伪随机性,可以认为 $\overleftrightarrow{P}'$ 与 G 不相关,则 $\overleftrightarrow{P}'$ 的前 M 个周期与 S 的绝对相关值的平均为

$$\begin{aligned}cor(S,\overleftrightarrow{P}')&=\frac{1}{MN'}\Big[\sum_{i=0}^{MN'-1}g_i\cdot\overleftrightarrow{p}'(i)+\sum_{i=0}^{MN'-1}\alpha\cdot\overleftrightarrow{p}(i)\cdot\overleftrightarrow{p}'(i)\Big]\\&\approx\frac{\alpha}{MN'}\sum_{i=0}^{MN'-1}\overleftrightarrow{p}(i)\cdot\overleftrightarrow{p}'(i)\end{aligned} \tag{5}$$

记 $cor(\overleftrightarrow{p},\overleftrightarrow{p}')=\frac{1}{MN'}\sum_{i=0}^{MN'-1}\overleftrightarrow{p}(i)\cdot\overleftrightarrow{p}'(i)$,则

$$cor(S,\overleftrightarrow{P}')\propto cor(\overleftrightarrow{P},\overleftrightarrow{P}') \tag{6}$$

令 $\Delta N=N'-N$,不失一般性,下面讨论 $0<\Delta N<N$ 的情况. 如图 2 所示,若 $i\cdot\Delta N<N$

$$\begin{aligned}cor(\overleftrightarrow{P},\overleftrightarrow{P}')=\frac{1}{MN'}\sum_{i=0}^{M-1}\Big|&\sum_{j=0}^{N'-(i+1)\cdot\Delta N-1}\overleftrightarrow{P}'_i(j)\cdot\overleftrightarrow{P}_i(j+i\cdot\Delta N)\\&+\sum_{j=N'-(i+1)\cdot\Delta N}^{N'-1}\overleftrightarrow{P}'_i(j)\cdot\overleftrightarrow{P}_{i+1}(j-(N'-(i+1)\cdot\Delta N))\Big|\end{aligned} \tag{7}$$

令

$$\overleftrightarrow{P}_i(j)=\begin{cases}\overleftrightarrow{P}_i(j+i\cdot\Delta N),0<j<N'-(i+1)\cdot\Delta N-1\\\overleftrightarrow{P}_{i+1}(j-(N'-(i+1)\cdot\Delta N)),\\\qquad N'-(i+1)\cdot\Delta N<j<N'-1\end{cases} \tag{8}$$

式(8)可转化为

$$\begin{aligned}cor(\overleftrightarrow{P},\overleftrightarrow{P})&=\frac{1}{MN'}\sum_{i=0}^{M-1}\Big|\sum_{j=0}^{N'-1}\overleftrightarrow{p}'_i(j)\cdot\overleftrightarrow{P}_i(j)\Big|\\&=\frac{1}{MN'}\sum_{i=0}^{M-1}\Big|\sum_{j=0}^{N'-1}p'_j\cdot\overleftrightarrow{P}_i(j)\Big|\end{aligned} \tag{9}$$

由式(9)可知,$cor(\overleftrightarrow{P},\overleftrightarrow{P}')$ 取得最大值的充要条件是:对于 $\forall\, i\in\{0,1,\cdots,M-1\}$,均满足 $P'=\overleftrightarrow{p}_i$ 或 $P'=-\overleftrightarrow{p}_i$.

图2 估计序列与扩频序列相关性

要使上述条件成立,首先须满足

$$\overleftrightarrow{p}_{i_1}=\overleftrightarrow{p}_{i_2}\text{或}\overleftrightarrow{p}_{i_1}=-\overleftrightarrow{p}_{i_2},\text{对于}\forall\, i_1,i_2\in\{0,1,\cdots,M-1\} \tag{10}$$

根据式(8):当 $\Delta N=0$ 时,存在 $\overleftrightarrow{p}_i=\overleftrightarrow{p}_i=w_i\cdot P$ 使得式(10)成立. $cor(\overleftrightarrow{P},\overleftrightarrow{P}')$ 取最大值的充要条件可转化为:在 $N'=N$ 条件下满足

$$P'=w_i\cdot P\text{ 或 }P'=-w_i\cdot P,\text{对于}\forall\, i\in\{0,1,\cdots,M-1\} \tag{11}$$

由 $w_i\in\{-1,1\}$,当 $P'=P$ 或 $P'=-P$ 时

$$\begin{aligned}cor(\overleftrightarrow{P},\overleftrightarrow{P}_{\max})&=\frac{1}{MN'}\sum_{i=0}^{M-1}\Big|\sum_{j=0}^{N-1}p'_j\cdot w_i\cdot p_j\Big|\\&=\frac{1}{N}\sum_{j=0}^{N-1}\Big|w(i)\cdot p_j^2\Big|=1\end{aligned} \tag{12}$$

当 $i\cdot\Delta N>N$ 时,令 $k=\big|(i+1)\cdot\Delta N/N'\big|$,$(i'+1)\cdot\Delta N=(i+1)\cdot\Delta N\bmod N'$,

式(7)变为

$$\begin{aligned}cor(\overleftrightarrow{P},\overleftrightarrow{P})=\frac{1}{MN'}\sum_{i=0}^{M-1}\Big|&\sum_{j=0}^{N'-(i'+1)\cdot\Delta N-1}\overleftrightarrow{p}'_i(j)\cdot\overleftrightarrow{p}_{i+k}(j+i'\cdot\Delta N)\\&+\sum_{j=N'-(i'+1)\cdot\Delta N}^{N'-1}\overleftrightarrow{P}'_i(j)\cdot\overleftrightarrow{P}_{i+k+1}(j-(N'-(i'+1)\cdot\Delta N))\Big|\end{aligned} \tag{13}$$

通过类似的分析,可以得到同样的结论,这里不再重复讨论.

综上所述,只有当估计序列与 PN 序长度相等,且估计序列与 PN 序列一致或相反(即 PN 序列按位取反)时,累积相关值取得最大值. 上述问题可等价为:以式(5)为目标函数,N' 和 P' 为未知变量的两变量最优化问题.

2.3 两变量优化模型求解

遗传算法是求解复杂优化问题的一种有效方法,具有全局收敛性、鲁棒性,已得到较广泛应用. 基本遗传算法可定义为一个九元组:

$$GA\triangleq\langle C,E,P_0,Q,\Phi,\Gamma,\Psi,\Pi,T\rangle \tag{14}$$

其中,C:个体的编码方法,E:个体适应度评价函数,P_0:初始种群,Q:群体大小,Φ:选择算子,Γ:交叉算子,Ψ:变异算子,Π:置换算子,T:遗传运算终止条件.

为了适应 N' 和 P' 两变量优化问题,采用了适当的个体编码方式,这种编码包括两个部分(见图 3):前 N_1 比特,用于编码序列长度;后 N_2 比特,表示估计序列. 其中 $N_1=\big|\log_2 N_2\big|$,N_2 表示 PN 序列的最大可能长度.

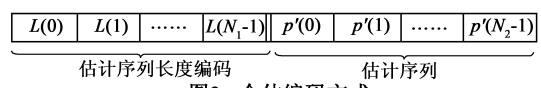


图3 个体编码方式

对个体 x_i,前 N_1 比特通过二进制转十进制,得到序列长度 N_i,则估计序列 $P'=\{p'(0),p'(1),\cdots,p'(N_i-1)\}$,该个体适应度评价函数可表示为:

$$f(x_i)=cor(S,P')=\sum_{i=0}^{M-1}\left|\sum_{j=0}^{N_i-1}p'(j)s(i\cdot N_i+j)\right| \tag{15}$$

选择算子,交叉算子及变异算子采用文献[10]中的自适应遗传算法策略,在保持群体多样性的同时,保证遗传算法的收敛性.

3　扩频隐藏信息的盲提取

设 C 为原始载体信号,P 是嵌入方所使用的扩频序列,长度为 N.对于某一比特待隐藏秘密信息 $w\in\{1,-1\}$,嵌入过程如下[11]

$$S=C+\alpha\cdot w\cdot P \tag{16}$$

其中 α 为嵌入强度.S 为藏密信号.

无先验知识条件下,如何从 S 中提取出秘密信息,主要包括两个阶段:

首先,获取扩频序列的最佳估计 P'.一般地,扩频序列也是 PN 序列,因而可采用第 2 节提出的 PN 序列估计方法来解决.若嵌入秘密信息后的藏密信号为 S,经过维纳滤波后得到 S',个体适应度表达式(15)修改为

$$\begin{aligned}f(x_i)&=cor(S,P')\\&=\sum_{i=0}^{M-1}\left|\sum_{j=0}^{N_i-1}p'(j)(s(i\cdot N_i+j)-s'(i\cdot N_i+j))\right|\end{aligned} \tag{17}$$

由维纳滤波器[12]的性质可知,S'可看成是嵌入信息前载体图像的一个估计,通过分析估计序列 P' 与差值图像 $S-S'$ 的相关性,可以有效减小载体图像对遗传算法目标函数的影响.

其次,根据估计序列长度 N' 对藏密信号分段后,计算 P' 和 S 的相关值

$$\begin{aligned}cor(S,P')&=\sum_{i=0}^{N'-1}s_i\cdot p_i'\\&=\sum_{i=0}^{N'-1}c_i\cdot p_i'+\sum_{i=0}^{N'-1}g_i\cdot p_i'+\sum_{i=0}^{N'-1}\alpha\cdot w\cdot p_i'^2\end{aligned} \tag{18}$$

由于 P'的伪随机性,P'与 C 和 G 是不相关的,式(18)可转化为

$$cor(S,P')\approx\sum_{i=0}^{N-1}\alpha\cdot w\cdot p_i^2=N\cdot\alpha\cdot w \tag{19}$$

由相关值的符号即可获取秘密信息 w.

4　实验及结果

为了验证本文的方法,在微机上进行模拟实验,实验平台为 Windows XP 下 Visual C ++ 6.0 和 Matlab 7.04.实验包括 PN 序列估计方法仿真分析和扩频隐藏信息的盲提取仿真分析两个部分.

4.1　PN 序列估计实验

取信息码长度 $R=200$,PN 序列长度 $N=31$ 和 $N=63$,在不同的信噪比下,对 PN 序列长度及其构成的估计性能进行仿真分析,经过 260 次实验,得到 PN 序列长度正确估计率与信噪比的关系曲线(如图 4).实验中遗传算法参数设定为:$p_{c1}=0.8$,$p_{c2}=0.6$,$p_{m1}=0.01$,$p_{m2}=0.001$,最大遗传代数 $T=1000$,种群大小 $Q=500$.从图中可以看出,当 $SNR>-10$dB 时,PN 序列长度正确估计率接近或达到 100%.图 5 是扩频码序列正确估计率与信噪比的关系曲线,当 $SNR>-10$dB 时,正确率达到 90% 以上,$SNR>-5$dB 时,正确率接近 100%.

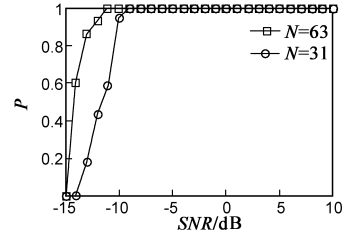


图4　PN序列长度正确估计率与信噪比的关系

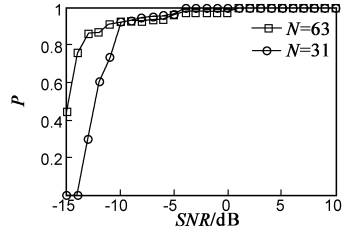


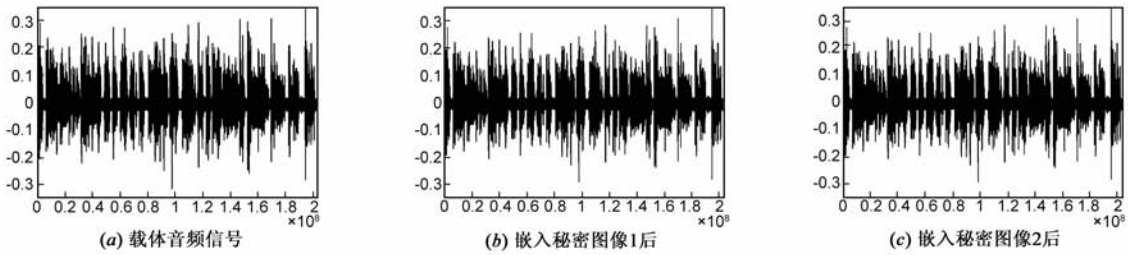
图5　扩频码序列正确估计率与信噪比关系

4.2　扩频隐藏信息盲提取实验

实验中选取音频信号作为载体,从音频隐藏分析数据库中随机选取了 30 段原始音频(包括语音,歌曲,音乐等不同类型),选定扩频序列长度 $N=63$,嵌入强度 $\alpha=0.05$,采用连续嵌入方式,分别嵌入大小为 71 * 86 和 80 * 80 的未加密二值秘密图像,得到一组共 60 段藏密音频.改变 N 和 α 的取值(见表 1),得到其余 11 组共 660 段藏密音频.采用本文的方法从藏密音频中提取秘密信息,获得的数据经过适当调整,可得到秘密图像.

图 6 和 7 给出了部分实验结果.图 6 给出了音频嵌入前后的波形图;图 7 给出了提取的秘密图像与嵌入的秘密图像的对比.由图中可以看出,载体音频波形嵌入前后基本无明显变化,提取的秘密图像虽然带有噪点,但仍然具有较好的可辨认性.而文献[9]的方法由于无法估计扩频序列长度,无法真正实现秘密信息的盲提取.

表 1 是 12 组藏密音频中秘密信息提取结果的统计.从表中可以看出,与文献[8]实现的 70% 提取正确率相比,本文的方法得到了提高.



(a) 载体音频信号 (b) 嵌入秘密图像1后 (c) 嵌入秘密图像2后

图6 音频嵌入前后的波形图

(a) 秘密图像1 (b) 提取的图像1

(c) 秘密图像2 (d) 提取的图像2

图7 盲提取效果

表 1 扩频隐藏信息提取结果

扩频序列长度	嵌入强度	扩频序列长度估计正确率(%)	扩频隐藏信息提取正确率(%)
63	0.05	89.97	81.45
	0.08	90.48	87.96
	0.10	89.93	89.31
	0.12	91.29	94.18
127	0.03	85.25	71.96
	0.05	87.03	80.52
	0.08	87.44	89.19
	0.10	89.40	91.15
255	0.01	80.89	55.97
	0.03	83.92	68.77
	0.05	84.71	79.42
	0.08	86.71	90.73

5 结论

本文首先提出了一种非协作情况下的 DS-SS 信号 PN 序列的盲估计方法,该方法采用遗传算法,以估计序列与藏密信号的累积相关值作为目标函数,通过求解 PN 序列长度及构成的两变量优化模型,获得 PN 序列的最佳估计.在此基础上,实现了扩频序列长度未知时扩频隐藏信息的盲提取.实验结果表明,该方法在截获信号信噪比较低情况下仍具有较好的估计性能,提取的秘密图像信息具有较好的可辨认性.该方法同样适应于其它载体形式(如图像等)的扩频隐藏分析.

参考文献:

[1] C Bouder, S Azou, G Burel. Performance analysis of a spreading sequence estimator for spread spectrum transmissions [J]. Journal of the Franklin Institute, 2004, 341(7): 594 – 614.

[2] Li Jiang, Hongbing Ji, Lin Li. A blind estimation algorithm for PN sequence in DS-SS signals [A]. Proceedings of 8th International Conference on Signal Processing [C]. Guilin, China, 2006.3.16 – 20.

[3] Mingyan Jiang, Yong Wang, Francisco Rubio. Spread spectrum code estimation by artificial fish swarm algorithm [A]. Proceedings of 2007 IEEE International Symposium on Intelligent Signal Processing [C]. Alcala, Spain, 2007.1 – 6.

[4] Chun-Hsiang Huang, Shang-Chih Chuang, Ja-Ling Wu. Digital invisible-lnk data hiding based on spread-spectrum and quantization techniques [J]. IEEE Trans on multimedia, 2008, 10(4): 557 – 569.

[5] Maria Gkizeli, Dimitris A. Pados, Michael J. Medley. Optimal signature design for spread-specturm steganography [J]. IEEE Trans on image processing, 2007, 16(2): 391 – 405.

[6] 钱振兴,程义民,王以孝,等.一种图像自嵌入方法[J].电子学报,2006,34(7):1347 – 1350.
Qian Zhen-xing, Cheng Yi-min, Wang Yi-xiao, et al. A method of image self-embedding [J]. Acta Electronica Sinica, 2006, 34(7): 1347 – 1350. (in Chinese)

[7] 谢春辉,程义民,汪云路,等.基于统计特征的音频中隐藏信息检测[J].电子与信息学报,2009,31(6):1341 – 1344.
Xie Chun-hui, Cheng Yi-min, Wang Yun-lu, et al. Estimation of secret message in audio based on statistic characteristics [J]. Journal of Electronics & Information Technology, 2009, 31(6): 1341 – 1344. (in Chinese)

[8] Rongrong Ji, Hongxun Yao, Shaohui Liu, et al. A new steganalysis method for adaptive spread spectrum steganography [A]. Proceeddings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing [C]. Pasadena, California, 2006.365 – 368.

[9] Saeed Sedghi, Habib Rajabi Mashhadi, Morteza Khademi. Detecting hidden information from a spread spectrum watermarked signal by genetic algorithm [A]. Proceedings of 2006 IEEE congress on Evolutionary Computation [C]. Vancouver, Cana-

da,2006.173 – 178.
[10] 王小平,曹立明.遗传算法:理论,应用及软件实现[M].西安:西安交通大学出版社.2002.73 – 74.
[11] 邹潇湘,戴琼,黄晁,等.零知识水印验证协议[J].软件学报,2003,14(9):1645 – 1651.
Zou Xiao-xiang, Dai Qiong, Huang Chao, et al. Zero knowledge watermark verification protocols [J]. Journal of Software,2003,14(9):1645 – 1651.(in Chinese)
[12] Imteyaz Ahmad, Partha P Mondal, Rajan Kanhirodan. Hebbian learning based FIR filter for image restoration [A]. Proceedings of 2005 IEEE International Symposium on Signal Processing and Information Technology [C]. Athens, Greece, 2005.726 – 730.

作者简介:

谢春辉　男,1983 年出生于湖南益阳,博士研究生,主要研究领域为信息隐藏,数字水印,网络多媒体等.
E-mail:cvlab@ustc.edu.cn

程义民　男,1945 年出生于陕西西安,教授,博士生导师,主要研究领域为信息隐藏,网络多媒体,计算机视觉,深度图像分析等.
E-mail:ymcheng@ustc.edu.cn