Journal of Computer Applications
计算机应用,2013,33(12):3511 -3513
ISSN 1001-9081
CODEN JYIIDU
2013-12-01
http://www.joca.cn

文章编号:1001-9081(2013)12-3511-03
doi:10.11772/j.issn.1001-9081.2013.12.3511

基于低密度奇偶校验码的数据协调技术

张彦煌,郭大波*,王云艳
(山西大学 物理电子工程学院,太原 030006)
(*通信作者电子邮箱 dabo_guo@sxu.edu.cn)

摘　要:低密度奇偶校验码(LDPC)是一种(n,k)线性分组码。当分组码码长较短时,利用常规的编码方法可以完成编码工作。但随着分组码码长的增加,利用常规LDPC的编码方式编码,计算机的内存难以承担。为了解决以上问题,提出两种有效的编译码方案。首先,该数据协调方案不同于传统校验位译码,利用边信息和原始数据产生的校验子进行联合译码;其次,将校验矩阵以稀疏矩阵的形式存储,利用双向十字循环链表只记录1的位置的方式存储校验矩阵,这样可极大地节省内存空间;最后,通过C语言实现可提高编译码的有效性。实验中选取码长10^5的分组长度,译码器误码率(BER)收敛于1.0 dB,每一分组译码时间仅需4 s,译码收敛后速率达到24.85 kb/s,时效性较强。

关键词:低密度奇偶校验码;边信息;校验子;稀疏矩阵;数据协调
中图分类号:TP309　　**文献标志码**:A

Reconciliation technology based on low density parity check code

ZHANG Yanhuang, GUO Dabo*, WANG Yunyan
(*College of Physics and Electronic Engineering, Shanxi University, Taiyuan Shanxi* 030006, *China*)

Abstract: Low Density Parity Check Code (LDPC) is a kind of (n, k) linear block codes. The conventional encoding method can complete the encoding work when the length of the codes is short, but as the codes become longer, the memory of computer is hard to bear when still using the common encoding methods. To solve this problem, two kinds of effective encoding and decoding schemes were proposed. Firstly, different from the traditional data parity bit decoding, the proposed data reconciliation scheme used side information and syndrome produced by the initial data to employ joint decoding. Secondly, the parity matrix was stored in a way that only the positions of 1 in the form of the cross circular list were recorded, which could greatly save memory space. At last, C implementation could improve effectiveness of the codes. Length of a block of codes in this experiment was 10^5. The Bit Error Ratio (BER) of codes was converged above 1.0 dB, only in need of 4 seconds to decode one block, the code rate could reach 24.85 kb/s when the decoder was converged. The results show that the proposed schemes have strong timeliness.

Key words: Low Density Parity Check Code (LDPC); side information; syndrome; sparse matrix; reconciliation

0　引言

量子密钥通信是量子通信领域中最接近实用的研究方向。协调技术是量子密钥通信中必不可少的一个环节,它可判断窃听的存在,纠正量子通信中大量的误码,并通过密性放大实现密钥蒸馏。在经典加密体系中,密钥分配一般分为两种方式:对称密钥体系(私钥密码系统)和非对称密钥体系(公钥密码体系)[1]。其中数据协调是对量子信息传输过程造成的不一致进行纠正的一种协议,而密性放大(或密钥蒸馏)是通过去除在量子传输和协调过程中窃听者得到的信息,从而使Alice和Bob拥有相同的密钥的协议[2-3]。

截止目前,量子密钥分发(Quantum Key Distribution, QKD)[4]已在实际应用方面获得大量进展,成果突出。欧美、日本等地区都投入了大量人力、物力进行相关研究,国内中国科学院、山西大学量子光学与光量子器件重点实验室、中国科学技术大学等单位也致力于这方面的研究,并取得了一定的研究成果。Namekata等[5]利用超低噪声正弦门控的雪崩光电二极管,采用单光子探测器,实现了100 km的传输距离;当改用差分移相的探测方法后,传输距离可达160 km。连续变量量子密钥分发系统方面,Lodewyck等[6]实现了一个25 km的全光纤连续变量QKD系统。Leverrier等[7]提出了一种非高斯调制QKD协议,提高了在密钥协商阶段的效率,该协议还在密钥传输过程中加入了诱骗态,可以抵御任意的集体攻击,证明了它在线性量子信道中是绝对安全的。Zhou等[8]、宋汉冲等[9]提出了连续变量量子确定性密钥分配协议,该协议的主要目的是经由公共信道移交一预先确定的密钥给接收者,其中密钥对发送者而言是确定的,在利用零差探测法的情况下协议的传输效率达到了100%。Wang等[10]在2013年已完成一个30 km全光纤分离调制连续变量QKD系统。

QKD在实际应用方面成果突出,但系统中很少涉及具体的数据协调方案与协调速率,本文将从低密度奇偶校验码(Low Density Parity Check Code, LDPC)校验矩阵的合理构造和协调方案的改进两方面阐述协调速率提高的原因。其中数据协调步骤为:利用Alice传递过来的边信息[11]根据校验矩阵计算出校验子(syndrome)并通过经典信道传给Bob端,Bob端译出Alice端的密钥信息。该方法即为数据协调中的正向协调。

收稿日期:2013-06-08;**修回日期**:2013-08-21。　**基金项目**:量子光学与光量子器件国家重点实验室开放基金资助项目(KF201003)。
作者简介:张彦煌(1987 -),男,山西朔州人,硕士研究生,主要研究方向:量子密钥分发;　郭大波(1963 -),男,山西阳泉人,副教授,博士,主要研究方向:量子密钥分发;　王云艳(1987 -),女,山西吕梁人,硕士研究生,主要研究方向:量子密钥分发。

1 LDPC 的构造

1.1 LDPC 应用概述

低密度奇偶校验码(LDPC)最早是由 Gallager 在其 1962 年发表的博士论文中提出。1996 年,MacKay 等[12]发现 LDPC 的性能具有超越 Turbo 码的趋势,从此,LDPC 的研究进入新的阶段,引起了人们的广泛关注与研究。LDPC 利用稀疏的校验矩阵进行编译码,校验矩阵构造合适时,可以实现低信噪比下的完美译码。其码率灵活可变,译码算法简单,且可并行计算,这些特点非常适用于数据协调要求。

1.2 LDPC 的构造方法

不同方法构造的校验矩阵性能各异,编译码复杂度也互不相同。因此,如何构造一个兼顾低复杂度和优秀性能的 LDPC 成为研究者们一直追求的目标[13]。本方案中构造 LDPC 的主要工作是:首先寻找一个稀疏矩阵 $\boldsymbol{H}$,使以 $\boldsymbol{H}$ 为校验矩阵的线性分组码有较好的纠错能力,并且解码的复杂度和码长 n 呈线性关系。信息论的研究成果表明,稀疏校验矩阵 $\boldsymbol{H}$ 越大,1 的分布越随机,LDPC 的解码性能越好,越接近于香农限。本方案采用 C 语言实现,其核心技术是 $\boldsymbol{H}$ 矩阵的内存占用问题,随着分组码长增加到百万级别,计算机的内存难以承担负荷。解决方法是利用 $\boldsymbol{H}$ 矩阵的稀疏性,即只记录 1 的位置的方式存储 $\boldsymbol{H}$ 矩阵,这样可极大地降低空间复杂度。本文采用双向循环链表的方式,研究如何提高矩阵元素的定位速度,实现整体系统仿真优化。

图 1 中 $\boldsymbol{H}$ 矩阵中只记录了非零元素 1 的位置 (j,i),相邻两个非零元素之间的 0 省去,其中 N_f、N_x 分别为该非零元素在该行、该列的邻域。这样 $\boldsymbol{H}$ 矩阵中的 1 的元素根据其所在的位置与相邻的非零元素构成了十字双向循环链表,这样既提高了非零元素的定位速度,又大大降低了 $\boldsymbol{H}$ 矩阵的空间复杂度。

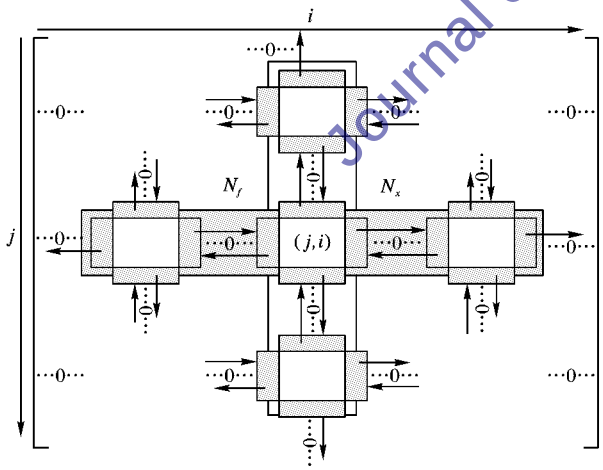

图 1 $\boldsymbol{H}$ 矩阵中非零元素十字双向循环链表结构

2 基于边信息的二进制信源压缩

2.1 QKD 正向协调模型

在实际量子密钥传输系统中,Bob 端收到的信息含有各种噪声、损失和窃听等干扰,为了在原始密钥序列中提取出绝对安全的密钥序列,BB84 和 B92 协议规定,需要在 QKD 系统中引入数据协调(reconciliation)的步骤,即 Bob 对量子态进行测量重新得到经典数据后,利用公开经典信道对筛选后数据进行纠错。数据协调的作用是,通过纠错编码的方法和信道纠错编码技术把 Alice 和 Bob 共同拥有的存在少量不一致的序列变为一致序列。

正向协调是 Alice 根据校验矩阵计算出校验子(syndrome)并通过经典信道传给 Bob 端,Bob 端译出 Alice 端的密钥信息。从信息论的角度来看,量子密钥通信 + 正向协调系统可抽象成图 2 所示模型。

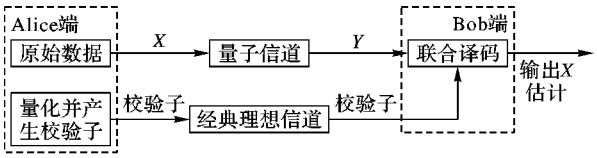

图 2 QKD 正向协调框图

2.2 二进制信源编码和压缩

为了使用现有的信道编码技术,上述模型中的量子信道可用图 3 所示的高斯白噪声信道(Additive White Gaussian Noise,AWGN)模型所代替,X 经过该信道得到 Y,边信息 Y 在联合译码器端与校验子 Z 联合解译得到 $\hat{X}$。

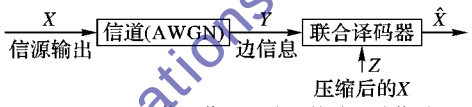

图 3 用于边信息信源压缩的等效相关信道

编码端得到码流为:

$$Z = \boldsymbol{H}X \tag{1}$$

得到长度为 $n-k$ 的校验子 Z,从而得到压缩比为 $n:n-k$ 的压缩码流,与传统信道编码的方式不同,X 是长为 n 的不包含校验位的原始信息码字。从二值图(bipartite)的角度来看,校验子的每一位可看作所有变量节点与同一校验节点连接的加和关系。

图 3 中的各个符号的数学意义为:$X=[X_1,X_2,\cdots,X_n]$,$Y=[Y_1,Y_2,\cdots,Y_n]$,X_i、Y_i 分别是独立等概的二进制随机序列;X_i 和 Y_i 相关且 $\Pr[X_i \neq Y_i]=p<0.5$。解码器端是联合译码器并且其输入是无压缩的 Y 和 Z,其中 Z 是 X 的压缩。Y 的码率是其熵 $nR_2 = n\boldsymbol{H}(Y_i) = n$ 比特,X 压缩的理论下限是 $nR_1 \geqslant n\boldsymbol{H}(X_i \mid Y_i) = n\boldsymbol{H}(p) = n(-p\ln p-(1-p)\ln(1-p))$。

3 基于边信息的解码器

解码器的目标是从 $n-k$ 比特长的校验子和对应的 n 比特长的序列 Y 估计出 n 比特长的 X。以下为本文使用的变量符号:

$x_i, y_i \in \{0,1\}(i=1,2,\cdots,n)$ 分别是 X 和 Y 序列中的当前值,对应于第 i 个变量节点 v_i;

$l_i \in \{2,3,\cdots\}(i=1,2,\cdots,n)$ 是 v_i 的度;

$q_{i,m}^{\text{out}}(q_{i,m}^{\text{in}}) \in \mathbf{R}(i=1,2,\cdots,n, m=1,2,\cdots,l_i)$ 是从校验节点到变量节点 v_i 的第 m 条边对数似然比;

$s_j \in \{0,1\}(j=1,2,\cdots,n-k)$ 是 $Z_{1,j}$ 的值,对应于第 j 个校验节点 c_j,即第 j 个校验子分量;

$r_j \in \{2,3,\cdots\}(j=1,2,\cdots,n-k)$ 是 c_j 的度;

$t_{i,m}^{\text{out}}/t_{i,m}^{\text{in}} \in \mathbf{R}(j=1,2,\cdots,n-k, m=1,2,\cdots,r_j)$ 是从变量节点到校验节点 c_j 的第 m 条边对数似然比;

设

$$q_{i,0} = \ln\frac{\Pr[x_i=0 \mid y_i]}{\Pr[x_i=1 \mid y_i]} = (1-2y_i)\ln\frac{1-p}{p} \tag{2}$$

其中:$i = 1,2,\cdots,n$,$p = \Pr[x_i \neq y_i] < 0.5$,变量节点 v_i 沿第 m 条边所送出的对数似然比为

$$q_{i,m}^{\text{out}} = q_{i,0} + \sum_{j=1,j\neq m}^{l_i} q_{i,j}^{\text{in}} \tag{3}$$

其中:$m = 1,2,\cdots,l_i$,$i = 1,2,\cdots n$,这里初始化 $q_{i,j}^{\text{in}} = 0$。

根据二值图的连接关系,$q_{i,m}^{\text{out}}$ 将送给对应的 $t_{j,\pi(i,m,j)}^{\text{in}}$,并用于校验节点,处理如下:

$$\tanh\left(\frac{t_{j,m}^{\text{out}}}{2}\right) = (1-2s_j)\prod_{i=1,i\neq m}^{r_j}\tanh\left(\frac{t_{i,m}^{\text{in}}}{2}\right) \tag{4}$$

其中:$m = 1,2,\cdots,r_j$,$j = 1,2,\cdots,n-k$,$(1-2s_j)$ 项的引入代表校验子信息。

这时,对于二值图的所有边,$q_{i,m}^{\text{in}} = t_{j,\pi(i,m,j)}^{\text{out}}$,可估计 $\hat{x}_i$ 如下:

$$\hat{x}_i = \begin{cases} 0, & q_{i,0} + \sum_{m=1}^{l_i} q_{i,m}^{\text{in}} \geqslant 0 \\ 1, & q_{i,0} + \sum_{m=1}^{l_i} q_{i,m}^{\text{in}} < 0 \end{cases} \tag{5}$$

如此反复循环迭代,直到满足校验方程或达到最大迭代次数。

4　数值仿真结果及分析

本文使用的硬件平台是 CPU 为 Inter Xeon E5620 2.4 GHz和32 GB 内存的双核服务器,选择码率0.5,码长分别为 10^4、10^5、10^6,共 100 个分组进行统计,译码最大迭代次数为 100,实验数据统计如表 1。

表 1　不同码长的实验结果

信噪比/dB	码长 $=10^4$			码长 $=10^5$			码长 $=10^6$		
	误码率	迭代次数	平均耗时/s	误码率	迭代次数	平均耗时/s	误码率	迭代次数	平均耗时/s
0	1.457 E−1	100.0	0.3803	1.430 E−1	100.0	10.1540	1.446 E−1	100.0	125.590
0.5	6.742 E−2	97.6	0.3774	4.120 E−2	99.6	9.6172	3.351 E−2	100.0	116.780
1.0	7.000 E−5	33.8	0.1526	0.000 E+0	34.1	4.0249	0.000 E+0	38.1	49.956
1.5	0.000 E+0	18.8	0.0945	0.000 E+0	21.8	2.6965	0.000 E+0	23.4	32.503
2.0	0.000 E+0	13.5	0.0692	0.000 E+0	15.2	1.9107	0.000 E+0	16.8	24.796

由表 1 可知码长为 10^4、10^5、10^6 时误码率的收敛信噪比分别为 1.5 dB、1.0 dB、1.0 dB。随着码长的增加,译码效果增强,译码时间也随之增加。本文实验结果验证了 LDPC 随着码长的增加性能加强,但当码长增加到一定值时,其性能也会达到瓶颈。本文中选码长 10^5 为最佳码长,当信噪比达到 1.0 dB后,随着误码率收敛,平均迭代次数减少,耗时也不断减少,译码收敛后速率达到 24.85 kb/s,而同样的条件和平台上,用 Matlab 仿真的速度为 85.46 b/s,所以 C 语言优化后的协调算法是 Matlab 协调算法的近 300 倍。可见 C 语言算法实时性较强,为以后硬件实现奠定了基础。

5　结语

本文为了提高数据协调的可靠性和有效性,提出两种解决方案,首先,协调方案采用数据协调中的正向协调方案,这种方案得益于避开由校验矩阵产生生成矩阵,从策略上加速译码;其次,将 LDPC 校验矩阵以稀疏矩阵的形式存储,以降低空间复杂度来提高译码速度,从根本上加速译码。实验结果表明当码长为 10^5 时,收敛于 1.0 dB,译码收敛后速率达到 24.85 kb/s,时效性较强。

进一步的研究工作是能够在更低信噪比下加速收敛,从优化校验矩阵度分布这一角度进一步降低信噪比要求,最终将在现场可编程门阵列(Field Programmable Gate Array, FPGA)上实现整个协调过程。

参考文献:

[1] 陈晓峰, 王育民. 公钥密码体制研究与进展[J]. 通信学报, 2004, 25(8): 109−118.

[2] van ASSCHE G, CARDINAL J, CERF N J. Reconciliation of a quantum-distributed gaussian key [J]. IEEE Transactions on Information Theory, 2004, 50(2): 394−400.

[3] LEVERRIER A, GRANGIER P. Continuous-variable quantum key distribution protocols with a discrete modulation [J/OL]. ArXiv Preprint, 2010: 1002.4083 (2011−01−24) [2013−02−12]. http://arxiv.org/abs/1002.4083.

[4] MINK A, NAKASSIS A. LDPC for QKD reconciliation [J/OL]. ArXiv Preprint, 2012: 1205.4977 (2012−05−22) [2013−02−16]. http://arxiv.org/abs/1205.4977.

[5] NAMEKATA N, TAKESUE H, HONJO T, *et al*. High-rate quantum key distribution over 100 km using ultra-low-noise, 2-GHz sinusoidally gated InGaAs/InP avalanche photodiodes [J]. Optics Express, 2011, 19(11): 10632−10639.

[6] LODEWYCK J, BLOCH M, GARCÍA-PARTÓN R, *et al*. Quantum key distribution over 25 km with an all-fiber continuous-variable system [J]. Physical Review A, 2007, 76(4): 042305.

[7] LEVERRIER A, GRANGIER P. Continuous-variable quantum key distribution protocols with a non-Gaussian modulation[J]. Physical Review A, 2011, 83(4):042312.

[8] ZHOU N R, WANG L J, GONG L H, *et al*. Quantum deterministic key distribution protocols based on teleportation and entanglement swapping [J]. Optics Communications, 2011, 284(19): 4836−4842.

[9] 宋汉冲, 龚黎华, 周南润. 基于量子远程通信的连续变量量子确定性密钥分配协议[J]. 物理学报, 2012, 61(15): 154206.

[10] WANG X Y, BAI Z L, WANG S F, *et al*. Four-state modulation continuous variable quantum key distribution over a 30 km fiber and analysis of excess noise [J]. Chinese Physics Letters, 2013, 30(1): 010305.

[11] LIVERIS A D, XIONG Z X, GEORGHIADESC N. Compression of binary sources with side information at the decoder using LDPC codes [J]. IEEE Communications Letters, 2002, 6(10): 440−442.

[12] MacKAY D J C, NEAL R M. Near Shannon limit performance of low density parity check codes [J]. Electron Letters, 1996, 33(6): 457−458.

[13] 袁东风, 张海刚. LDPC 码理论与应用[M]. 北京: 人民邮电出版社, 2008: 63−68.