激光与光电子学进展

基于 K 近邻的相位编码连续变量量子密钥分发安全性分析

赵常兰，王天一*

贵州大学大数据与信息工程学院，贵州 贵阳 550025

摘要 提出了一种基于 K 近邻算法的相位编码连续变量量子密钥分发量子态识别方法。算法利用接收量子态的相位特征实现识别，首先由已知相干态构成的训练集进行学习，再根据未知量子态提取出的相位特征进行分类。推导了基于 K 近邻的识别方法在集体攻击与反向协调下的安全码率，比较了该方法应用于四态协议和八态协议下，在不同传输距离、调制方差、过量噪声下的性能。数值仿真结果表明，该方法能够有效生成安全密钥，当安全码率为 10^{-5} 比特每符号时，传输距离可达到 250 km。

关键词 连续变量；量子密钥分发；相位编码；K 近邻算法

中图分类号 O431.2 **文献标志码** A **DOI**：10.3788/LOP222511

Security Analysis of Phase Encoding Continuous-Variable Quantum Key Distribution Based on *K*-Nearest Neighbor

Zhao Changlan, Wang Tianyi*

College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, *Guizhou, China*

Abstract This paper proposes a phase encoding, continuous-variable quantum key distribution quantum-state identification method based on the K-nearest neighbor algorithm. The algorithm achieves recognition using the phase features of the accepted quantum states, which are first learned from a training set consisting of known coherent states and then classified according to the phase features extracted from the unknown quantum states. Moreover, this paper derives the secure code rate of the K-nearest neighbor-based identification method under collective attack and reverse coordination and compares the performance of the method applied to four-state and eight-state protocols under different transmission distances, modulation variance, and excess noise. The results of numerical simulation results show that the method can effectively generate secure keys with a transmission distance of 250 km when the secure code rate is 10^{-5} bit per symbol.

Key words continuous variables; quantum key distribution; phase encoding; K-nearest neighbors

1 引 言

量子密钥分发(QKD)是量子信息领域的研究热点之一。基于量子力学的基本定理，QKD能够使两个距离遥远的合法双方Alice和Bob之间共享安全密钥。连续变量量子密钥分发(CVQKD)是QKD实现的重要手段之一[1]，CVQKD通过改变量化电场的正交振幅来编码信息，使用零差或外差检测方案代替单光子检测器来恢复密钥。

目前，CVQKD协议中有两种调制方式，即高斯调制和离散调制[2]。高斯调制CVQKD的安全性由海森堡不确定性原理保证，协议执行过程可概括为：Alice随机选择两个离散信号，平均值为0，数值大小服从高斯分布，分别调制到相干光的正交振幅和相位上，并通过不完美信道发送给Bob；Bob从收到的量子态中，随机选择一个正交分量进行平衡零拍探测，并将自己的测量方式告诉Alice；最后，通信双方Alice和Bob将未测量的正交分量和被破坏的信息丢弃，并通过反向协调纠正错误编码和密钥放大[3]。

由于高斯CVQKD调制在低信噪比条件下无法获

收稿日期：2022-09-12；**修回日期**：2022-10-17；**录用日期**：2022-10-20；**网络首发日期**：2022-11-01

基金项目：贵州省科技计划项目(黔科合基础-ZK[2021]一般304)、贵州大学培育项目(贵大培育[2021]56号)

通信作者：*ty.wang@gzu.edu.cn

得高协调效率，因此提出了离散调制CVQKD，其典型协议包括基于相位编码的四态协议[4]和八态协议[5]，其更适合实现长距离的通信过程且对信道噪声的容忍度较高。离散调制CVQKD协议的通信过程分为两个阶段：首先是量子通信的传输阶段，由Alice制备光源，并从等概率的n个相干态$\left|\alpha\exp\left[\mathrm{i}(2k+1)\pi/4\right]\right\rangle$（$\alpha$是正实数，$k\in\{0,1,\cdots,n\}$）中随机选择一个发送给接收端Bob，在探测到所发量子态后进行零差探测处理；第二阶段经典数据的后处理过程，即误码校正、保密增强，最后生成所需的安全密钥。离散调制协议在理论上有望可达到数百公里，并对量子信道产生的过量噪声表现出很高的耐受性。它们的特征是，对于每个脉冲和任何实际距离，它们总是产生少于比特的信号，这种特性使得它们更适合长距离工作。

基于相干态的CVQKD能够和标准的光通信器件相匹配，因而受到研究者的广泛关注。由于收发端激光器发射噪声和中心频率抖动的影响，在自参考连续变量量子密钥分发协议中提出了利用光放大器补偿参考脉冲引入的相位噪声的影响[6]；在离散调制CVQKD通信系统中，接收端采用平衡零拍探测，由于信道干扰、探测器电子噪声等的影响，量子态相位的测量结果不可避免地叠加了相位噪声，造成收发双方密钥速率的影响。

本文提出了基于K近邻算法（KNN）的相位编码CVQKD方案。在该方案中，将量子通信系统分为两个过程，即量子态的学习和预测，由已知相干态构成的训练集进行学习，再对未知量子态提取相位特征进行分类进而做出判决，设计一个量子态识别分类器，更精确地预测未知相干态提高了离散调制CVQKD的系统性能。

2　相位编码CVQKD协议

目前，相位编码协议主要包括四态协议和八态协议。在四态协议中，Alice发送n个随机相干态，这些相干态来自

$$S_4=\left\{\left|\alpha\exp(\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(3\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(5\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(7\mathrm{i}\pi/4)\right\rangle\right\},\tag{1}$$

$$S_8=\left\{|\alpha\rangle,\left|\alpha\exp(\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(\mathrm{i}\pi/2)\right\rangle,\left|\alpha\exp(3\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(\mathrm{i}\pi)\right\rangle,\left|\alpha\exp(5\mathrm{i}\pi/4)\right\rangle,\left|\alpha\exp(3\mathrm{i}\pi/2)\right\rangle,\left|\alpha\exp(7\mathrm{i}\pi/4)\right\rangle\right\},\tag{2}$$

式中：α为正实数。接收端Bob对接收到的量子态执行零差探测，得到随机变量$y_i\{i\in1,2,\cdots,n\}$，执行反向协调操作利用y_i的符号b_i对原始密钥位进行编码：当$y_i\geqslant0$时，$b_i=1$；当$y_i<0$时，$b_i=0$。Bob会发送相关信息如y_i的绝对值以及Alice和Bob事先约定好的线性纠错码帮助Alice恢复$\boldsymbol{b}=\{b_1,b_2,\cdots,b_n\}$的值，其中$\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$对应于它发送状态的Bob正交测量的符号。

Bob收到的量子态混合态密度矩阵[7]记为$\boldsymbol{\rho}_d$，其形式为

$$\boldsymbol{\rho}_d=\sum_{i=0}^{d}\lambda_i\left|\phi_i\right\rangle\left\langle\phi_i\right|,d\in\{4,8\}。\tag{3}$$

对于四态协议：

$$\boldsymbol{\rho}_4=\sum_{i=0}^{3}\lambda_i\left|\phi_i\right\rangle\left\langle\phi_i\right|。\tag{4}$$

其中，

$$\begin{cases}\lambda_{0,2}=\dfrac{1}{2}\exp(-\alpha^2)\left(\cosh\alpha^2\pm\cos\alpha^2\right)\\ \lambda_{1,3}=\dfrac{1}{2}\exp(-\alpha^2)\left(\sinh\alpha^2\pm\sin\alpha^2\right)\end{cases},\tag{5}$$

$$\left|\phi_k\right\rangle=\frac{\exp(-\alpha^2/2)}{\sqrt{\lambda_k}}\sum_{n=0}^{\infty}\frac{(-1)^n\alpha^{4n+k}}{\sqrt{(4n+k)!}}\left|4n+k\right\rangle,\quad k\in\{0,1,2,3\}。\tag{6}$$

对于八态协议：

$$\boldsymbol{\rho}_8=\sum_{i=0}^{7}\lambda_i\left|\phi_i\right\rangle\left\langle\phi_i\right|。\tag{7}$$

其中，

$$\begin{cases}\lambda_{0,4}=\dfrac{1}{4}\exp(-\alpha^2)\left(\cosh\alpha^2+\cos\alpha^2\pm2\cos\dfrac{\alpha^2}{\sqrt{2}}\cosh\dfrac{\alpha^2}{\sqrt{2}}\right)\\ \lambda_{1,5}=\dfrac{1}{4}\exp(-\alpha^2)\left(\sinh\alpha^2+\sin\alpha^2\pm\sqrt{2}\cos\dfrac{\alpha^2}{\sqrt{2}}\sinh\dfrac{\alpha^2}{\sqrt{2}}\pm\sqrt{2}\sin\dfrac{\alpha^2}{\sqrt{2}}\cosh\dfrac{\alpha^2}{\sqrt{2}}\right)\\ \lambda_{2,6}=\dfrac{1}{4}\exp(-\alpha^2)\left(\cosh\alpha^2-\cos\alpha^2\pm2\sin\dfrac{\alpha^2}{\sqrt{2}}\sinh\dfrac{\alpha^2}{\sqrt{2}}\right)\\ \lambda_{3,7}=\dfrac{1}{4}\exp(-\alpha^2)\left(\sinh\alpha^2-\sin\alpha^2\mp\sqrt{2}\cos\dfrac{\alpha^2}{\sqrt{2}}\sinh\dfrac{\alpha^2}{\sqrt{2}}\pm\sqrt{2}\sin\dfrac{\alpha^2}{\sqrt{2}}\cosh\dfrac{\alpha^2}{\sqrt{2}}\right)\end{cases},\tag{8}$$

$$|\phi_k\rangle=\frac{\exp(-\alpha^2/2)}{\sqrt{\lambda_k}}\sum_{n=0}^{\infty}\frac{\alpha^{8n+k}}{\sqrt{(8n+k)!}}|8n+k\rangle,$$
$$k\in\{0,1,2,\cdots,7\}。\tag{9}$$

经过接收端Bob执行零差探测操作和经典后处理过程[8]，可计算它的协方差矩阵$\boldsymbol{\Gamma}_d$，计算表达式为

$$\boldsymbol{\Gamma}_d=\begin{pmatrix}V\boldsymbol{I} & Z_d\boldsymbol{\sigma}_z\\ Z_d\boldsymbol{\sigma}_z & V\boldsymbol{I}\end{pmatrix},\tag{10}$$

式中：$\boldsymbol{I}=\mathrm{diag}(1,1)$；$\boldsymbol{\sigma}_z=\mathrm{diag}(1,-1)$；$V=1+2V_{\mathrm{A}}$，$V_{\mathrm{A}}=2\alpha^2$表示Alice的调制方差；$Z_d=\mathrm{V}_d\sum_{i=0}^{d}\frac{\lambda_{i-1}^{3/2}}{\lambda_i^{1/2}}$；在四态和八态协议中$V_4=V_8=2\alpha^2$。

3　基于相位特征的KNN分类算法

KNN算法能够解决监督学习中的分类问题，其核心思想是在给定已知标签的数据集空间中对未知数据样本进行分类，首先基于距离度量计算出与数据集空间最靠近的K个已知标签数据样本，然后依据投票原则确定该测试数据样本的类别。

度量空间中点的距离，有好几种度量方式，比如常见的曼哈顿距离计算、欧氏距离计算等。

K值是KNN算法重要的参数，表示根据参考态的K个最近邻数据样本做出判决，K值的选择对最近邻算法的结果有很大的影响，因此K值的选择是关键问题。若选择较小的K值，决策就在较小的范围内进行判决，会导致估计误差增大、近似误差减小，容易发生过拟合现象；若K值选择较大，则相反。因此本文采用了交叉验证法[9]选择最优的K值。

3.1　用于相位识别的KNN算法

为了便于分析，将分析限制在线性量子信道上，通过该量子信道的单向量子密钥分配可表示为

$$p'=\sqrt{T}\,\alpha\cos\varphi+N(0,T\varepsilon),\tag{11}$$
$$q'=\sqrt{T}\,\alpha\sin\varphi+N(0,T\varepsilon)。\tag{12}$$

由于信道损耗和噪声的影响，传输状态以一定的概率分布在相空间中。图1是经过100 km量子信道传输后，4PSK和8PSK调制相干态的12000个数据点在相空间的表示。从图中可以观察到混沌点难以区分，因此，我们需要对数据进行特征提取，计算传输态与参考态之间的相位差，通过KNN算法进行判别。

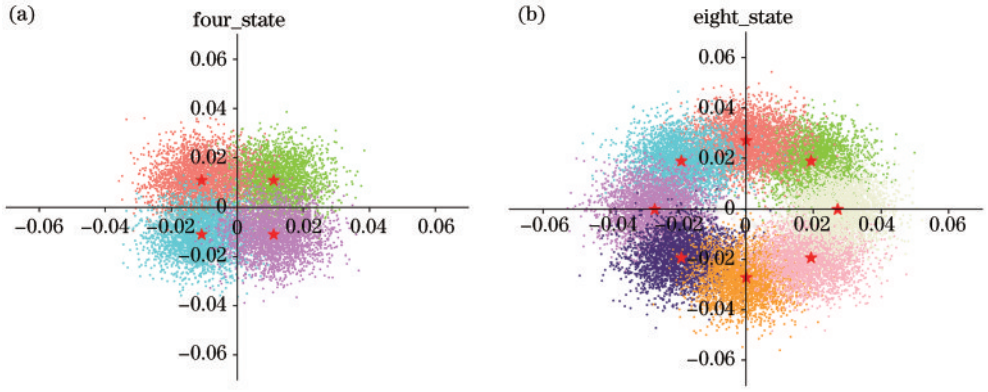

图1　调制相干态在相空间的分布。(a) 4PSK；(b) 8PSK

Fig. 1　Distribution of modulated coherent states in phase space. (a) 4PSK; (b) 8PSK

对于离散调制连续变量量子密钥分发，Alice通过不受信任的量子信道发送相干态，然后Bob接收传输的相干态。但是在传输过程中，由于信道噪声和损耗会导致相位漂移和能量衰减，传输状态不再与初始的调制状态相同。本文提出的方案是，首先发送端发送的训练数据是已知的量子态，系统经过对训练集的训练和学习，对未知的量子态提取相位特征，计算未知量子态与训练数据的相位差，选取前K个数据对其所属类别做出统计，若大多数样本属于某一类别，则此未知量子态也属于该类别。在笛卡尔坐标系中，点a和b的坐标分别为(x_1,y_1)和(x_2,y_2)，其相位差表示为

$$\theta_{a,b}=\arctan\frac{y_1}{x_1}-\arctan\frac{y_2}{x_2}。\tag{13}$$

根据KNN在特征空间对未知数据的分类[10]，可以对其相干态在相空间中做类似的处理，在选择近邻的时候，判断的依据是一维的相位值，绿色圆圈代表的是未知数据点，图中不同的颜色代表不同的类别，从图2(a)可以观察到，QPSK调制的每一个相干态都属于一个类别，当$K=5$时，与未知数据点距离最近的有三个属于第一类、两个属于第二类，所以绿色圆圈将属于第一类。图2(b)[11]是8PSK调制相干态的在相空间的表示，可以观察到其中一些相干态，如$|\alpha_2\rangle$、$|\alpha_4\rangle$、$|\alpha_6\rangle$、$|\alpha_8\rangle$属于多个类别，当$K=9$时，与未知数据点距离最近的有两个属于第一类、四个属于第二类、三个属于第三类，所以未知数据点绿色圆圈将属于第二类。

3.2　基于KNN的相位编码CVQKD协议

如图3所示，对于本文提出的以相位为特征参数的最近邻离散调制连续变量量子密钥分发协议，需要构造一个性能良好的分类器，首先需要一组训练数据集来学习，随后，再利用一组测试集数据来评估分类器的性能，若通过测试，则用于预测未知量子态。将本文方案分为两个部分，即状态学习和状态预测。

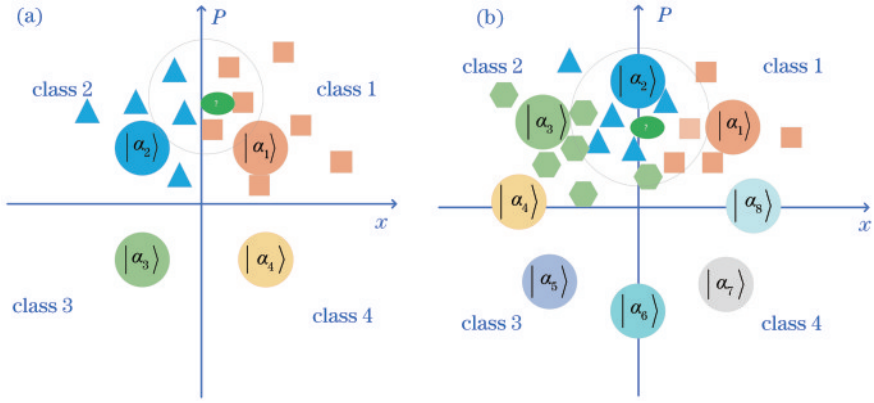

图 2　KNN 分类算法在相空间中的调制表示。(a)QPSK;(b)8PSK

Fig. 2　Modulation representation of KNN classification algorithm in phase space. (a) QPSK; (b) 8PSK

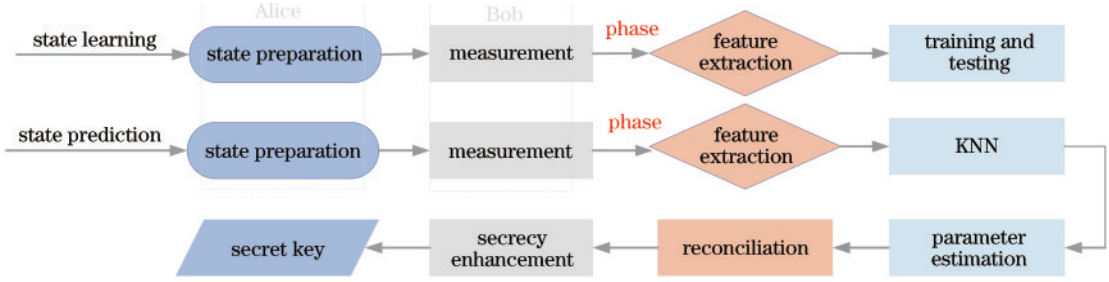

图 3　基于 KNN 的相位编码 CVQKD

Fig. 3　Phase coded CVQKD based on KNN

3.2.1　状态学习

1)Alice 对相干态进行相位编码调制,并通过不完美的传输信道将调制态发送给 Bob,同时把每个调制态的类别信息通过辅助信道发送给 Bob;

2)Bob 对接收到的相干态进行外差测量,从获得的标记相干态中提取特征,为训练和测试分类器做准备;

3)在获得足够多的特征数据之后,将其分为训练集和测试集,前者用于训练分类器,后者用于测试分类器的性能是否已达到要求。

如果经过训练的分类器通过测试,则不需要再重复前面的状态学习,系统已准备好生成密钥。

3.2.2　状态预测

1)Alice 对信号态进行相位编码调制,通过不可信的量子信道发送给 Bob;

2)Bob 对接收到的量子态进行外差探测[12],并从这些数据中提取特征,用作经过测试的分类器的输入数据;

3)Bob 利用 KNN 分类器对信号态进行分类并得到预测结果,经过多轮预测,Alice 和 Bob 之间将共享一串初始密钥;

4)Bob 向 Alice 通过公开一部分初始密钥信息进行量子信道的参数估计,并计算量子系统的协方差矩阵;

5)Alice 和 Bob 利用纠错码对数据进行反向协调[13];

6)Alice 和 Bob 对纠错后的数据进行保密增强,得到最终密钥。

3.3　基于 KNN 的相位编码 CVQKD 协议分类器

在提取出鲁棒性特征后用作状态学习分类器的输入数据,假设 $X=\boldsymbol{R}^d$ 是数据空间,$Y=\{y_1,y_2,\cdots,y_l\}$ 是包含 l 个类别的标签空间,训练集为 $U=\{(x_i,y_i)|1\leqslant i\leqslant m\}$,其中 $x_i\in X$ 是 d 维属性向量 $(x_{i1},x_{i2},\cdots,x_{id})^{\mathrm{T}}$,$y_i\subseteq Y$ 是 x_i 所属的一个类别。状态学习的任务是找到一个分类器 $h(\cdot)$,给定一个阈值函数 $t:X\to R$,使得 $h(x)=\{y|f(x,y)>t(x),y\in Y\}$。

若 $|x\rangle$ 表示未分类相干态,$N(|x\rangle)$ 表示在训练集中 $|x\rangle$ 的 k 个最近邻相干态子集,则 C_j 是 $|x\rangle$ 的邻居数属于第 j 类 $y_j(1\leqslant j\leqslant l)$。

$$C_j=\sum_{(|x^*\rangle,Y^*)\in N(|x\rangle)}[\![y_j\in Y^*]\!],\tag{14}$$

式中:$(|x^*\rangle,Y^*)$ 表示训练集中属于 $N(|x\rangle)$ 的已知类别相干态。假设 H_j 表示相干态 $|x\rangle$ 的类别为 y_j,则 $P(H_j|C_j)$ 表示后验概率,在 C_j 属于类别 y_j 的条件下 H_j 为真,$P(\bar{H}_j|C_j)$ 为后验概率 C_j 属于类别 y_j 的条件下 H_j 为假,令 $f(|x\rangle,y_j)=P(H_j|C_j)/P(\bar{H}_j|C_j)$,则基于 KNN 的相位编码 CVQKD 协议分类器为

$$h(|x\rangle)=\left\{y_j|P\left(H_j|C_j\right)/P\left(\bar{H}_j|C_j\right)>t(|x\rangle),1\leqslant j\leqslant l\right\}。\tag{15}$$

当$P\left(H_j|C_j\right)$后验概率大于$t(|x\rangle)\cdot P\left(\bar{H}_j|C_j\right)$时，未知类别的相干态$|x\rangle$属于$y_j$类。

根据贝叶斯定理可得到

$$f\left(|x\rangle,y_j\right)=\frac{P\left(H_j|C_j\right)}{P\left(\bar{H}_j|C_j\right)}=\frac{P\left(H_j\right)\cdot P\left(C_j|H_j\right)}{P\left(\bar{H}_j\right)\cdot P\left(C_j|\bar{H}_j\right)},\tag{16}$$

式中：$P\left(H_j\right)$和$P\left(\bar{H}_j\right)$分别表示H_j为真或假的先验概率；$P\left(H_j|C_j\right)$和$P\left(\bar{H}_j|C_j\right)$分别代表事件$H_j$为真或假的条件下$C_j$属于类别$y_j$的条件概率。

式(16)中的概率可以通过训练集中的频率计数来估计，先验概率可由下式计算。

$$P\left(H_j\right)=\frac{s+\sum_{i=1}^{m}\left[\!\left[y_j\in Y_i\right]\!\right]}{s\times 2+m},(1\leqslant j\leqslant l),\tag{17}$$

$$P\left(\bar{H}_j\right)=1-P\left(H_j\right),(1\leqslant j\leqslant l),\tag{18}$$

式中：s是一个平滑参数，用于控制概率估计过程中均匀先验分布的权重，对于拉普拉斯平滑通常设置为1。

与先验概率不同，式(16)中的条件概率的计算是复杂的，对于类别y_j计算了数组ς_j和$\bar{\varsigma}_j$，表示为

$$\begin{cases}\varsigma_j[r]=\sum_{i=1}^{m}\left[\!\left[y_j\in Y_i\right]\!\right]\cdot\left[\!\left[\phi_j(|\boldsymbol{x}_i\rangle)=r\right]\!\right],(0\leqslant r\leqslant k)\\ \bar{\varsigma}_j[r]=\sum_{i=1}^{m}\left[\!\left[y_j\notin Y_i\right]\!\right]\cdot\left[\!\left[\phi_j(|\boldsymbol{x}_i\rangle)=r\right]\!\right],(0\leqslant r\leqslant k)\end{cases}。\tag{19}$$

其中，

$$\phi_j(|x_i\rangle)=\sum_{(|x^*\rangle,\,Y^*)\in N(|x^*\rangle)}\left[\!\left[y_j\in Y^*\right]\!\right],\tag{20}$$

式中：$\phi_j(|x_i\rangle)$为统计第i个相干态属于y_j类的邻居的数量；ς_j用于计算属于类别y_j本身的相干态数目，并在k个邻居中精确地计算出属于y_j类别的r个邻居；$\bar{\varsigma}_j$则用于计算不属于y_j的相干态数目并且在k个邻居中精确地计算出属于y_j类别的r个邻居。因此，式(16)的条件概率可由下式计算。

$$\begin{cases}P\left(C_j|H_j\right)=\dfrac{s+\varsigma_j\left[C_j\right]}{s\times(k+1)+\sum_{r=0}^{k}\varsigma_j[r]}\\ P\left(C_j|\bar{H}_j\right)=\dfrac{s+\bar{\varsigma}_j\left[C_j\right]}{s\times(k+1)+\sum_{r=0}^{k}\bar{\varsigma}_j[r]}\end{cases},\tag{21}$$

式中：$1\leqslant j\leqslant l,0\leqslant C_j\leqslant k$。因此，一个基于KNN的相位编码CVQKD协议分类器便通过状态学习获得。

4　安全性分析

具有反向协调的传统离散调制CVQKD在集体攻击下的安全码率[14]可表示为

$$K=\beta I_{\mathrm{AB}}-\chi,\tag{22}$$

式中：β为数据反向协调效率；I_{AB}为Alice和Bob之间的香农互信息。其计算表达式为

$$I_{\mathrm{AB}}=\log_2\frac{\mathrm{V}+\chi_{\mathrm{tot}}}{1+\chi_{\mathrm{tot}}},\tag{23}$$

式中：$V=V_A+1$；$\chi_{\mathrm{tot}}=\xi-1+2(1+\upsilon_{el})/\left(\eta T\right)$为传输信道的总噪声，$\eta$和$\upsilon_{el}$分别是实际探测器的效率和探测器电子器件产生的噪声[15]；χ表示Eve和Bob之间互信息的Holevo界，需要在参数估计中计算。

$$\chi=\sum_{i=1}^{2}G\left(\frac{\lambda_i-1}{2}\right)-\sum_{i=3}^{5}G\left(\frac{\lambda_i-1}{2}\right),\tag{24}$$

式中：$G(x)=(x+1)\log_2(x+1)-x\log_2 x$是冯诺依曼熵。其本征值为

$$\begin{cases}\lambda_{1,2}^2=\dfrac{1}{2}\left[A\pm\sqrt{A^2-4B}\right]\\ \lambda_{3,4}^2=\dfrac{1}{2}\left[C\pm\sqrt{C^2-4D}\right]\end{cases}。\tag{25}$$

其中，

$$\begin{cases}A=V^2+T^2\left(V+\chi_{\mathrm{line}}\right)^2-2TZ^2\\ B=T^2\left(V^2+V\chi_{\mathrm{line}}-Z^2\right)^2\\ C=\dfrac{1}{T^2\left(V+\chi_{\mathrm{tot}}\right)^2}\left\{A\chi_{\mathrm{het}}^2+B+1+2\chi_{\mathrm{het}}\left[V\sqrt{B}+T\left(V+\chi_{\mathrm{line}}\right)+2TZ^2\right]\right\},\\ D=\left[\dfrac{V+\sqrt{B}\,\chi_{\mathrm{het}}}{T\left(V+\chi_{\mathrm{tot}}\right)}\right]^2\end{cases}\tag{26}$$

式中：$\lambda_5=1$；$\chi_{\mathrm{line}}=1/T-1+\xi$是通道输入有关的总通道附加噪声；$\chi_{\mathrm{het}}=[1+(1-\eta)+2\upsilon_{el}]/\eta$是Bob外差检测输入相关的检测附加噪声[16]；$Z$是Alice和Bob的相关性。

考虑KNN算法的效率，上述安全码率可修正为

$$K_{\mathrm{knn}}=\Lambda(T)\beta I_{\mathrm{AB}}-\chi_{\mathrm{knn}},\tag{27}$$

式中：$\Lambda(T)$为分类器的效率，其取值将随着传输距离的变化而变化；χ_{knn}表示在集体攻击下窃听者Eve通过与量子态相互作用获得的有用信息的Holevo量[17]。

$$\chi_{\mathrm{knn}}=S(\rho_{\mathrm{E}})-\sum_{y_i}^{m}p(y_i)S(\rho_{\mathrm{E}|y_i}),\tag{28}$$

$$\rho_{\mathrm{E}}=\sum_{y_i}^{m}p(y_i)\rho_{\mathrm{E}|y_i},\tag{29}$$

式中：$S(\rho)=-\mathrm{Tr}(\rho\log\rho)$表示冯诺依曼熵；$p(y_i)$表示Bob测量得到的原始密钥$y_i$的概率，在传统的离散调制CVQKD协议中，$Y=y_i(i=1,2,\cdots,m)$表示Alice随机选择的$m$个有限且完整的编码，每个编码均匀分布，因此$p(y_i)=1/m$。每个离散调制相干态与其二进制之间的表示关系是固定和分开的，四态协议中$p(y_i)=1/4$，八态协议中$p(y_i)=1/8$。在本文方案中，二进制密钥位之间的关系是由Alice随机分配的，在状态学习过程Eve并不参与，因此不知道Alice和Bob之间的编码。对于Eve来说，Y包含的随机编码是无限的，$p(y_i=0)(m\to\infty)$，意味着Eve很难从传输过程中获取相关信息。

5　讨论与分析

5.1　KNN算法中K值的选择

K值是KNN算法重要的超参数，表示判决时选取的最近邻数据样本的数量，其选择对最近邻算法的结果有很大的影响[18]。若选择较小的K值，决策就在较小的范围内进行判决，则会导致估计误差增大、近似误差减小，容易发生过拟合现象；若K值选择较大，则相反。因此，本文采用了交叉验证法选择最优的K值。

在评价模型时，使用了K折交叉验证。K折交叉验证是指将原始的数据分成K组，分别选择每个数据集作为验证集，其余作为训练集，将会得到K个模型并对其验证集的分类准确率求取平均值作为该算法的分类准确率。

对不同调制相干态下的数据样本进行K折交叉验证的仿真，从图4可以看出KNN算法在不同K值下的性能，当K值较小时，随着K值的增加，KNN算法的精确度也增加，但计算复杂度也越高，由于后面随着K的增加精度趋于稳定，所以综合考虑K的最优值应该取10左右，4PSK调制相干态在$K=12$时取最大值0.96，8PSK调制相干态在$K=11$时算法精确度达到最大值0.94。

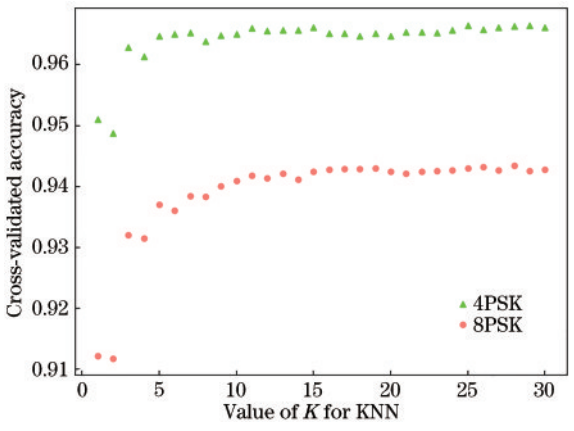

图4　4PSK和8PSK调制相干态下KNN算法在不同K值下的性能

Fig. 4　Performance of KNN algorithm under different K values in 4PSK and 8PSK modulated coherent states

5.2　KNN算法在不同调制方差、传输距离和过量噪声下的性能

量子态在传输过程中会受到各种不同因素的干扰，导致在决策过程中产生错误判决，本文对KNN算法在不同调制方差、传输距离和过量噪声下对4PSK和8PSK两种相位编码方案的分类精度进行仿真。

图5为不同调制方差下KNN算法的性能仿真，设置参数如下：$K=11$，过量噪声$\xi=0.01$，传输距离$d=100$ km，协调效率$\beta=0.98$。可以发现，随着调制方差的增大，KNN算法的性能得到增加，即精确度增加，调制方差增加意味着相邻量子态之间的距离增加，因而也就变得更容易区分，并且在同一调制方差下4PSK调制相干态的算法精确度优于8PSK调制相干态。

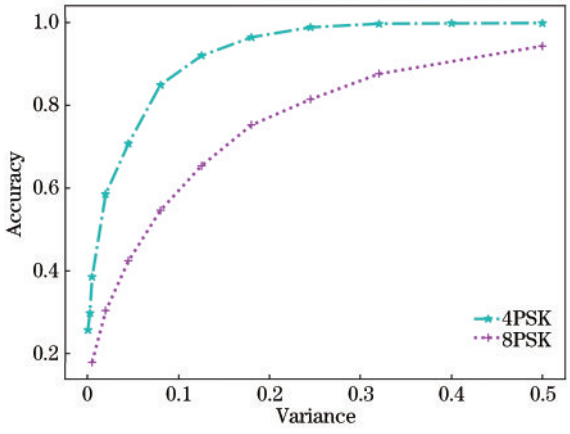

图5　KNN算法在不同调制方差下的性能

Fig. 5　Performance of KNN algorithm under different modulation variances

图6为KNN算法在不同传输距离下的仿真结果，设置参数如下：$K=11$，过量噪声$\xi=0.01$，协调效率$\beta=0.98$，调制方差$V_A=0.35$。随着传输距离的增大，KNN算法的精确度有所下降，在相同的传输距离下，4PSK调制相干态下的算法精确度优于8PSK调制相干态。

图7为KNN算法在4PSK和8PSK调制相干态下，过量噪声对算法精确度的影响，设置参数如下：$K=11$，协调效率$\beta=0.98$，调制方差$V_A=0.35$，传输距离$d=100$ km。从图中可以观察到，随着过量噪声的增加算法的精确度呈下降趋势，尤其是过量噪声从0增加到0.1过程中，算法的精确度急剧下降，当过量噪声为0.03时，算法的精确度低于0.6，因此，过量噪声的范围为0～0.03。

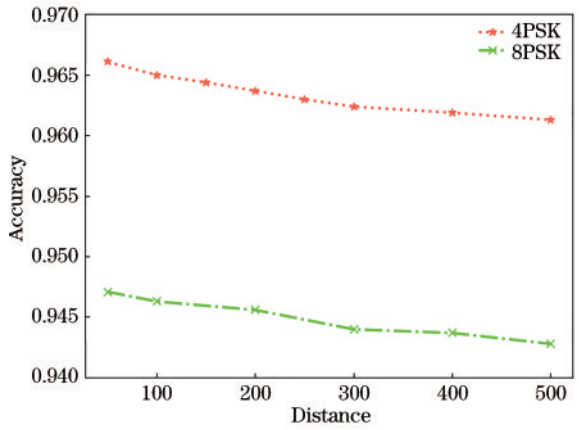

图 6　KNN 算法在不同传输距离下的性能

Fig. 6　Performance of KNN algorithm at different transmission distances

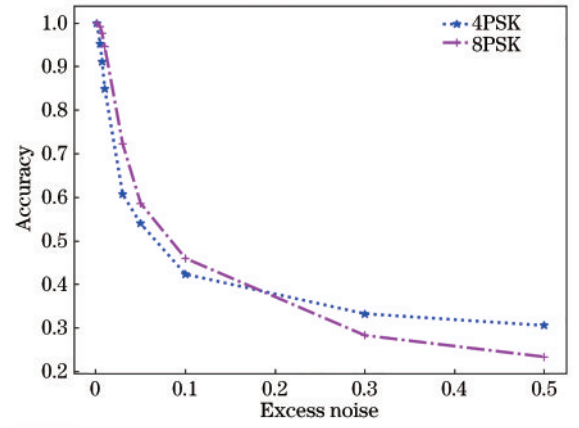

图 7　KNN 算法在不同过量噪声下的性能

Fig. 7　Performance of KNN algorithm under different excess noises

5.3　安全码率仿真

图 8 是在反向协调和集体攻击下基于 KNN 算法的离散调制 CVQKD 协议和几个现有 CVQKD 协议在渐近极限下的性能仿真图，设置参数如下：协调效率 $\beta=0.98$，过量噪声 $\xi=0.01$，$K=11$，调制方差 $V_A=0.35$。四态协议、八态协议和高斯调制 CVQKD 的调制方差在合理的信噪比进行了优化[19]，仿真结果表明，基于 KNN 算法的 QPSK 和 8PSK 调制协议的密钥速率和最大传输距离优于传统的四态协议和八态协议。

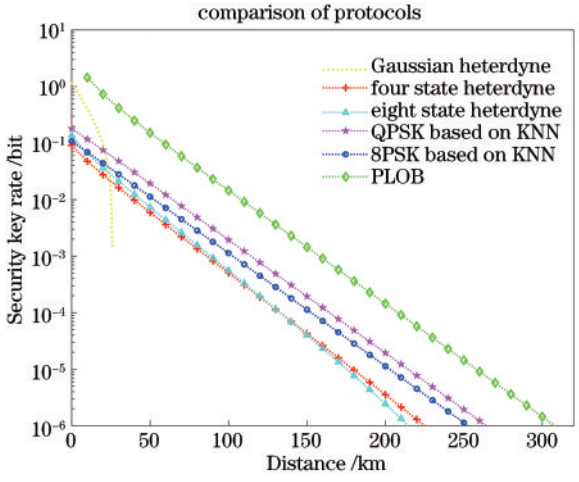

图 8　不同协议的性能比较

Fig. 8　Performance comparison of different protocols

图 9 是对不同调制方差下的基于 KNN 算法的 QPSK 调制和传统的四态协议在渐进极限下的性能比较，设置参数如下：协调效率 $\beta=0.98$，过量噪声 $\xi=0.01$，$K=11$。仿真结果表明，在相同的调制方差下，基于 KNN 算法的 QPSK 调制的最大传输距离优于传统的四态协议。并且，为了确保传统的四态协议的安全码率，其调制方差仅在一定的小范围内，随着调制方差增加到 1.5，其传输距离几乎为 0。此外，基于 KNN 算法的 QPSK 调制随着调制方差的增加密钥速率和最大传输距离均有所增加。

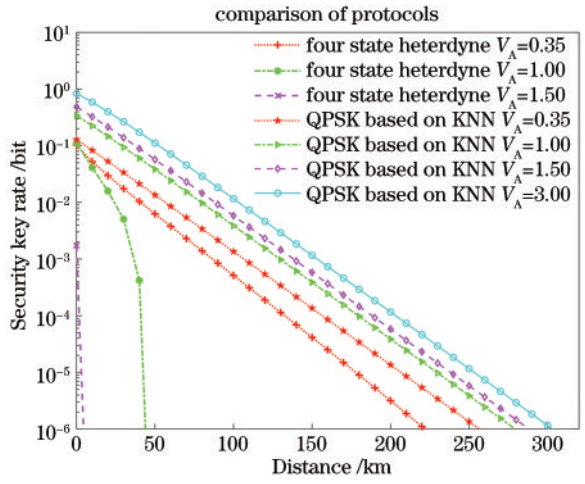

图 9　不同调制方差下协议的安全密钥速率

Fig. 9　Security key rates for protocols with different modulation variances

6　结　　论

本文提出了基于 KNN 的相位编码 CVQKD 协议，可分为两个过程即状态学习和状态预测两个步骤，状态学习用于训练和测试分类器，而状态预测用于生成最终密钥。本文通过 K 折交叉验证找到最佳 K 值，并分析了传输距离、调制方差和过量噪声等因素对算法性能的影响。仿真结果表明：算法的精确度随着调制方差的增加而增加，随着过量噪声和传输距离的增加算法精确度下降；基于 KNN 的相位编码 CVQKD 协议与传统的四态协议和八态协议相比，其传输距离和密钥率的性能随着调制方差的增加有所提升。

参　考　文　献

[1] Wang S Y, Huang P, Wang T, et al. Feasibility of continuous-variable quantum key distribution through fog [J]. Optics Letters, 2021, 46(23): 5858-5861.

[2] Li Z, Zhang H, Liao Q, et al. Ensemble learning for

failure prediction of underwater continuous variable quantum key distribution with discrete modulations[J]. Physics Letters A, 2021, 419: 127694.

[3] Liao Q, Zhong H, Guo Y. Multi-label learning for improving discretely-modulated continuous-variable quantum key distribution[J]. New Journal of Physics, 2020, 22(8): 083086.

[4] Zhao M F, Yuan R Z, Cheng J L, et al. Security of binary modulated continuous variable quantum key distribution using optimally displaced threshold detection [J]. IEEE Communications Letters, 2021, 25(4): 1089-1093.

[5] He D, Wu Y J, Li Y L, et al. Stability improvement enabled by four-state modulation in dual-polarization fiber optic gyroscopes[J]. Optics Communications, 2019, 452: 68-73.

[6] 黄彪, 黄永梅, 彭真明. 连续变量量子密钥分发的参考脉冲相位攻击与探测[J]. 光学学报, 2019, 39(11): 1127001.
Huang B, Huang Y M, Peng Z M. Attack and detection on reference-pulse phase of continuous-variable quantum-key distribution[J]. Acta Optica Sinica, 2019, 39(11): 1127001.

[7] Zhao W, Shi R H, Feng Y Y, et al. Unidimensional continuous-variable quantum key distribution with discrete modulation[J]. Physics Letters A, 2020, 384(2): 126061.

[8] Ghalaii M, Ottaviani C, Kumar R, et al. Discrete-modulation continuous-variable quantum key distribution enhanced by quantum scissors[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(3): 506-516.

[9] Chen S L, Zhang Z H, Liu L Y. Attribute selecting in tree-augmented naive Bayes by cross validation risk minimization[J]. Mathematics, 2021, 9(20): 2564.

[10] Thirani E, Jain J, Narawade V. Enhancing performance evaluation for video plagiarism detection using local feature through SVM and KNN algorithm[J]. International Journal of Image, Graphics and Signal Processing, 2021, 13(5): 41-50.

[11] Pan Y, Wang H, Shao Y, et al. Experimental demonstration of high-rate discrete-modulated continuous-variable quantum key distribution system[J]. Optics Letters, 2022, 47(13): 3307-3310.

[12] Erementchouk M, Mazumder P. Continuous-variable quantum key distribution with discretized modulations in the strong noise regime[J]. Physical Review A, 2020, 101(6): 062313.

[13] Wang Y J, Wang X D, Li J W, et al. Self-referenced continuous-variable measurement-device-independent quantum key distribution[J]. Physics Letters A, 2018, 382(17): 1149-1156.

[14] Guo Y, Li R J, Liao Q, et al. Performance improvement of eight-state continuous-variable quantum key distribution with an optical amplifier[J]. Physics Letters A, 2018, 382 (6): 372-381.

[15] Djordjevic I B. On the discretized Gaussian modulation (DGM)-based continuous variable-QKD[J]. IEEE Access, 2019, 7: 65342-65346.

[16] Guo Y, Ding J Z, Mao Y, et al. Quantum catalysis-based discrete modulation continuous variable quantum key distribution with eight states[J]. Physics Letters A, 2020, 384(12): 126340.

[17] Jain N, Chin H M, Mani H, et al. Practical continuous-variable quantum key distribution with composable security[J]. Nature Communications, 2022, 13: 4740.

[18] 褚荣燕, 王钰, 杨杏丽, 等. 基于正则化KL距离的交叉验证折数K的选择[J]. 计算机技术与发展, 2021, 31(3): 52-57.
Chu R Y, Wang Y, Yang X L, et al. A selection criterion of fold K in cross-validation based on regularized KL distance[J]. Computer Technology and Development, 2021, 31(3): 52-57.

[19] 卢奉宇, 银振强, 王双, 等. 50 km无特征源的测量设备无关量子密钥分发实验[J]. 光学学报, 2022, 42(3): 0327017.
Lu F Y, Yin Z Q, Wang S, et al. Uncharacterized-source measurement-device-independent quantum key distribution experiment with over 50 km fiber[J]. Acta Optica Sinica, 2022, 42(3): 0327017.