scientific reports

Check for updates

OPEN

A novel quantum private query protocol and its application in private set intersection

Zeping Deng[1,2,3], Hongwei Sun[3,4], Kejia Zhang[3,4,5,7]✉, Long Zhang[1,2,3,7]✉, Sujuan Qin[2,7] & Tingting Song[6,7]

As an important topic in quantum secure multiparty computing, quantum privacy query (QPQ) can solve the problem of information query between users and database owners without compromising privacy. In the existing research, most of them focus on its functions, not its applications. In this paper, a new QPQ protocol based on two-particle states is proposed. In our protocol, we use two-particle states combined with the Ctrl or Shift operations to give the user a partial key sequence and the database cannot recognize the user's key sequence. Meanwhile, we conduct simulation experiments at the IBM quantum experience and show that it is feasible. We discuss its application in the private set intersection (PSI), which is derived from QPQ. According to our analysis, the QPQ protocol and PSI protocol are secure.

Keywords Quantum private query, Private set intersection, Quantum secure multiparty computing

With the rapid development of technology, information interaction becomes increasingly frequent and information queries have become an indispensable part in daily life. However, different participants how to protect each other's privacy while querying information is a serious problem. To solve this challenge, secure multiparty computation (SMC) which is a cryptographic technique has emerged. It allows multiple participants to jointly compute a function while protecting data privacy[1–3]. An important scenario in SMC is the private information query. In classical cryptography, the approaches to realize private information query are based on obvious transfer (OT)[4] or symmetrically private information retrieval (SPIR)[5]. But the security of the above methods is ensured based on mathematically difficult problems and computational complexity[6,7], which is insecure in the context of quantum cryptography[8–10].

In this situation, quantum privacy query (QPQ) is proposed to provide a more secure and efficient solution for information queries. Participants can protect their privacy during the information query process by utilizing the superposition and entanglement of quantum states, thus achieving a higher level of privacy protection and security. Users and the database owner share the symmetric keys in the QPQ protocol. While the user only has the partial key, the database owner knows the entire key. In 2008, Italian scholars Giovannetti et al. proposed the first QPQ protocol (GLM protocol)[11], which used database operations (i.e., oracle operation) on the bistate to encrypt Alice's query and detected if Bob was spoofing, and was the first QPQ protocol based on quantum computing. In 2011, Polish scholars Olejnik et al. improved the shortcomings of the GLM protocol and proposed the O protocol[12], which used database manipulation and encryption methods to make it possible to obtain the desired information through a single query. Compared to the classical OT protocols, the above two protocols not only have advantages in terms of security (i.e., the protocols are designed based on physical principles rather than computational complexity) but also significantly reduce communication complexity.

However, it is difficult to implement the two protocols mentioned above because the dimensionality of oracle operations will increase as the database becomes larger. In order to address this issue, a new private query protocol (J-protocol) was introduced by Jakobi et al.[13]. Its foundation was the SARG04[14] system, which addressed the issue of computing challenges with big datasets. From that point on, QKD-based QPQ gets more and more attention from researchers. In 2012, Gao et al. improved the J-protocol and proposed the G-protocol[15], which

[1]School of Mathematical Sciences, Heilongjiang University, Harbin 150080, China. [2]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China. [3]Institute for Cryptology and Network Security, Heilongjiang University, Harbin 150080, China. [4]School of Computer and Big Data, Heilongjiang University, Harbin 150080, China. [5]State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550000, China. [6]College of Information Science and Technology, Jinan University, Guangzhou 510632, China. [7]These authors contributed equally: Kejia Zhang, Long Zhang, Sujuan Qin, and Tingting Song. ✉email: zhangkejia@hlju.edu.cn; lzhang@hlju.edu.cn

provided better privacy protection for both users and database owners. The length of the key in the G-protocol is constrained by changing the value of parameter θ. A QPQ protocol based on counterfactual quantum key distribution was proposed by Zhang et al. in 2013[16]. Both the efficiency and security of this protocol can be controlled by adjusting some parameters. But the above QPQ protocols can only query a single bit during post-processing. Therefore, in 2014, Wei et al. proposed a practical QPQ protocol based on the unbalanced Bennett-Brassard-1984 QKD protocol, which preserves user privacy while enabling users to retrieve many bits in a single query[17]. In 2014, Gao et al. focused their research on the post-processing of oblivious keys, analyzed the shortcomings of existing post-processing methods, and proposed post-processing of the oblivious key in quantum private query[18]. However, the security vulnerability lies in the fact that users can obtain secret information from the database through multiple queries. All malevolent attackers can obtain more data from the database to reveal information because the joint measurement attack is not taken into account. Therefore, Wei et al. introduced a QPQ protocol in 2016[19], where Bob might create quantum states to resist the JM attack. The protocol is resistant to quantum memory attacks and loss-tolerant.

Meanwhile, with the improved performance of the QPQ protocol[20], its application also needs urgent attention. Zhang et al. announced an enhanced transfer technique in 2020 and presented various specific decision problems, including set-member decision and point-inclusion[21]. However, there are currently fewer applications for the QPQ protocol. In this context, privacy set intersection (PSI) has become a topic of great interest. The PSI protocol is to compute the intersection of two parties' sets without each of the participants revealing any additional information about the other party's set beyond the intersection. In 2004, Freedman et al. laid the foundation of PSI research by designing privacy-preserving data-matching algorithms by first proposing a PSI protocol based on blind signatures and hash functions[22]. In 2005, Kissner and Song extended this work by proposing a generalized PSI protocol based on Secure Multiparty Computing (SMC) that implements PSI under multiple participants[23]. Subsequently, in order to solve the communication and computational complexity problems, researchers introduced homomorphic encryption into PSI protocols. In 2012, Huang et al. proposed a PSI protocol based on Paillier homomorphic encryption, which achieves a dynamic balance between security and computational efficiency[24]. However, the computational overhead of homomorphic encryption-based methods is still large and difficult to adapt to large-scale data scenarios. For this reason, Pinkas et al. proposed a PSI protocol based on pseudo-random functions (PRFs) and encrypted data exchange in 2014, which significantly reduces its communication complexity and provides good privacy protection at the same time[25]. In 2016, Kolesnikov et al. designed another batch-processing pseudo-random function, which was applied to the PSI protocol to achieve efficient intersection computation[26]. In 2018, Chen et al. proposed a PSI solution with low communication complexity based on homomorphic encryption[27]. But it still has a higher computational overhead compared to the OT-based scheme. Therefore, in the same year, Pinkas et al. proposed a significant optimization of the previous PSI protocols using the obvious transfer to design an efficient PSI protocol[28]. One well-established cryptographic tool that has been proven to address the PSI problem efficiently is oblivious transfer (OT). OT allows one party to send information to another in a way that ensures the receiver learns only the selected data, which directly solves the issue of privacy in PSI protocols. Given that QPQ is essentially a quantum variant of OT, it retains the ability to securely transfer information while preserving privacy. This makes QPQ a promising candidate for enhancing PSI protocols, as it can be used to securely compute the set intersection while ensuring that no additional information is revealed to the parties involved. Despite the potential applications of PSI, there are still problems such as the relative difficulty of implementation. Therefore, in this paper, we first propose a novel QPQ protocol based on two-particle states and use the key generated by the QPQ protocol as the basis for designing the PSI protocol which can solve the data alignment problem in vertical federated learning.

The rest of this paper is organized as follows. In Section 2, the QPQ protocol is described in detail. In Section 3, the correctness and flexibility of this QPQ protocol are analyzed. Simulations are performed through the IBM quantum experience. In Section 4, the security of our QPQ protocol is analyzed. Especially, in Section 5, we propose a generalized model to solve PSI problems with different coding and shifting strategies. Finally, a brief conclusion is summarized in Section 6.

The proposed QPQ protocol

In this section, we propose a new quantum private query protocol based on two-particle states which are from Bell basis and computational basis. Previously, we will describe the participants in the protocol as follows:

1. *The database owner Bob:* Bob has N pieces of data $X = \{x_1, \cdots, x_N\}$ and wants Alice to have access only to the data she retrieves and no other data.
2. *The user Alice:* Alice retrieves a few pieces of data at a time from the database and wants Bob to be unaware of what data she is accessing. In the protocol, Alice has part of the key and Bob has all of the key.

Here, Fig. 1 illustrates our QPQ protocol's simple process.

Initialization phase

Step 1. Alice prepares N qubits which are randomly in one of the states $\{|\psi^+\rangle,|\psi^-\rangle,|01\rangle,|10\rangle\}$. All of these qubits compose the sequence S. For detecting eavesdropping, Alice will generate a new sequence S' by inserting some decoy particles that are chosen from $\{|\psi^+\rangle,|\psi^-\rangle,|01\rangle,|10\rangle\}$. Then Alice sends the sequence S' to Bob.

Step 2. After confirming that Bob has received S', Alice announces the locations and the measurement basis of the eavesdropping detection particles in public. For the decoy states, Bob uses the basis published by Alice to measure and announces the measurement results. If the error rate is less than the predetermined threshold, the protocol will continue. Otherwise, Alice and Bob terminate this protocol and repeat Step1.

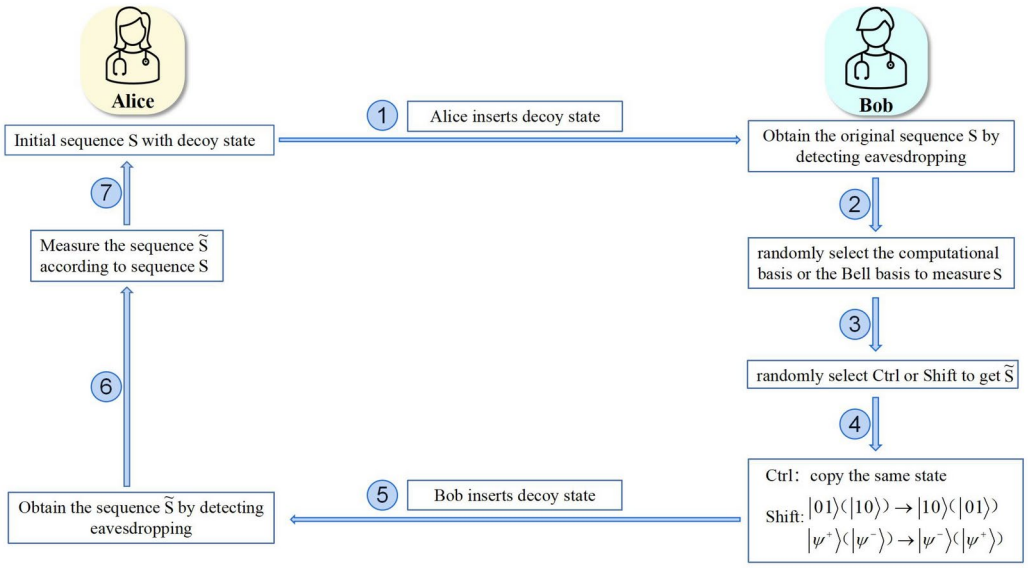

Fig. 1. The summarized process of QPQ protocol.

Original key generation phase

Step 3. Bob extracts the particle sequence S by discarding the decoy states. Then, Bob randomly selects either the computational basis or the Bell basis to measure the quantum state at each position of the sequence S. When Bob uses computational basis measurement, it means that Bob's raw key $k_i = 1$; otherwise, his raw key $k_i = 0$, where Bob gets the raw key $K_r = \{k_1, \cdots, k_N\}$, $k_i \in \{0, 1\}$. At the end of the measurement, Bob randomly chooses the Ctrl or Shift operation for each quantum state to obtain a new sequence $\widetilde{S}$.

Ctrl operation: Bob reproduces the state as the measurement result and returns it to Alice.

Shift operation: If Bob's measurement result is $M_B(N_B)$, he will reproduce $M'_B(N'_B)$ to Alice where

$$M_B = \begin{cases} |\psi^+\rangle \\ |\psi^-\rangle \end{cases}, M'_B = \begin{cases} |\psi^-\rangle \\ |\psi^+\rangle \end{cases}, N_B = \begin{cases} |10\rangle \\ |01\rangle \end{cases}, N'_B = \begin{cases} |01\rangle \\ |10\rangle \end{cases} \tag{1}$$

Subsequently, Bob forms a new sequence $\widetilde{S}'$ by inserting decoy states into $\widetilde{S}$ and sends $\widetilde{S}'$ to Alice. At the same time, Bob announces the operations performed on each bit of the quantum state.

Partial key generation phase

Step 4. With the similar eavesdropping detection in Step2, Alice obtains quantum particle sequence $\widetilde{S}$ by abandoning the decoy states and declares which instances were successfully received. Alice chooses the measurement basis according to the initial states that she prepares. If her initial state is $|\psi^-\rangle$ or $|\psi^+\rangle$, Alice chooses the Bell states basis for measurement. If her initial state is $|01\rangle$ or $|10\rangle$, Alice chooses the computational basis for measurement.

Step 5. For Ctrl operation: When the measurement result is different from the initial state, this indicates that Bob's measurement basis is different from Alice's. Alice can obtain the value of K_r by analysis. When the measurement result is the same as the initial state, Alice does not know which measurement basis Bob used.

For Shift operation: When the measurement result is the same as the initial state, this indicates that Bob's measurement basis is different from Alice's. Alice can obtain the value of K_r by analysis. When the measurement result is different from the initial state, Alice does not know which measurement basis Bob used.

Step 6. Following the above procedures, Bob and Alice can obtain the raw key, of which Bob is aware of the entirety and Alice only of a portion. It can be seen that all possible cases when the initial state is $|\psi^+\rangle$ and $|01\rangle$ in Table 1.

Key processing phase

Step 7. Bob encrypts his database and Alice obtains the item which she wants with one of her known bits in K_r. The specific steps are as follows:

Firstly, if Alice is the owner of Bob's jth bit k_j of his key K_r and she needs to retrieve the ith record x_i from Bob's database, she will inform Bob of the value $s = j - i$. Then, Bob shifts K_r left circularly with s bits to create a new key K'_r; if s is a negative value, Bob shifts K_r right circularly with $|s|$ bits. Finally, Bob uses K'_r as a one-time pad to encrypt the database. Using her key k_j, Alice decrypts the ith record.

Correctness and flexibility analysis of our QPQ protocol

This section begins with an analysis of the solution's correctness. Then, in order to better illustrate our proposed protocol, simulations are performed through the IBM quantum experience.

Alice's initial state	Bob's measurement basis	Bob's measurement result	Ctrl or shift	Alice's Measurement result
$\|\psi^+\rangle$	Bell Basis	$\|\psi^+\rangle$	Ctrl	$\|\psi^+\rangle$
			Shift	$\|\psi^-\rangle$
	Computational basis	$\|01\rangle$	Ctrl	$\|\psi^+\rangle/\|\psi^-\rangle$
			Shift	$\|\psi^+\rangle/\|\psi^-\rangle$
		$\|10\rangle$	Ctrl	$\|\psi^+\rangle/\|\psi^-\rangle$
			Shift	$\|\psi^+\rangle/\|\psi^-\rangle$
$\|01\rangle$	Computational basis	$\|01\rangle$	Ctrl	$\|01\rangle$
			Shift	$\|10\rangle$
	Bell basis	$\|\psi^+\rangle$	Ctrl	$\|01\rangle/\|10\rangle$
			Shift	$\|01\rangle/\|10\rangle$
		$\|\psi^-\rangle$	Ctrl	$\|01\rangle/\|10\rangle$
			Shift	$\|01\rangle/\|10\rangle$

Table 1. The initial state is the full case of $|\psi^+\rangle$ and $|01\rangle$.

Alice's initial state	$\|\psi^+\rangle$	$\|01\rangle$	$\|10\rangle$	$\|\psi^-\rangle$	$\|\psi^-\rangle$
Bob's measurement basis	Computational Basis	Bell Basis	Bell Basis	Computational Basis	Computational Basis
Bob's measurement result	$\|01\rangle$	$\|\psi^+\rangle$	$\|\psi^-\rangle$	$\|10\rangle$	$\|01\rangle$
Ctrl or shift	Ctrl	Ctrl	Shift	Shift	Ctrl
Alice's measurement result	$\|\psi^-\rangle$	$\|10\rangle$	$\|10\rangle$	$\|\psi^+\rangle$	$\|\psi^-\rangle$
Key	k_1	×	k_3	×	×

Table 2. A specific example.

Correctness analysis

Theorem 1 *The QPQ protocol designed in Section 2 is correct.*

Proof In the protocol, Bob randomly selects the Bell basis or computational basis to measure sequence S and generates the original key K_r based on the selected measurement basis. In this process, Bob does not declare the exact value of the key K_r. Therefore Alice does not get information about the key K_r. In addition, Alice selects the measurement basis to measure the sequence $\tilde{S}$ based on the initial state and obtains a partial key based on the results of the measurement and the operations declared by Bob.

Specifically, suppose Alice initially sends $|\psi^+\rangle$ to Bob. Bob measures the result as $|10\rangle$ and selects the Ctrl operation. Next, Alice measures $|10\rangle$ using the Bell basis and gets $|\psi^+\rangle$ or $|\psi^-\rangle$ with probability $P = \frac{1}{2}$. If Alice's measurement result is $|\psi^-\rangle$, she can judge that the measurement basis chosen by Bob is the computational basis and the corresponding key for that position is 1. Since Bob is unaware of Alice's choice of measurement basis, he has no information about which keys she has recovered. □

Furthermore, we provide a detailed example to explain how Alice obtains a partial key. Assuming that the initial state prepared by the user Alice is $|\psi^+\rangle|01\rangle|10\rangle|\psi^-\rangle|\psi^-\rangle$, according to the protocol flow, Alice and Bob carry out the message transfer, and the specific results are shown in Table 2. From Table 2, it can be seen that the user Alice obtains the keys k_1 and k_3.

If in this protocol the user Alice acquires two keys k_i and $k_j (i \neq j)$, Alice wants to get the data of the database owner Bob x_m. According to Step6 in the protocol, Alice only needs to inform Bob of the number of shifts s. After Bob performs the shift operation, Alice can obtain the data x_m with her key. Although Alice acquires two keys in the process, since Alice can only send one shift s to Bob, then whether she sends $s = i - m$ or $s = j - m$, she will only get the data x_m and will not get the rest of Bob's data. For the database owner Bob, the rest of his data will not be compromised and he does not know which data specifically is accessed by user Alice. The user Alice does not compromise her privacy during the processing.

Flexibility and efficiency analysis

We assume that Alice prepares three initial states $|\psi^+\rangle$, $|\psi^-\rangle$, $|10\rangle$. Bob receives the quantum state sent by Alice and randomly selects the measurement basis and operation. Then he prepares the quantum state to send to Alice. We assume that the quantum states that Bob sends to Alice are $|\psi^+\rangle$, $|01\rangle$, $|10\rangle$. Alice selects the corresponding measurement basis to measure the received particle sequence based on the initial state. She will

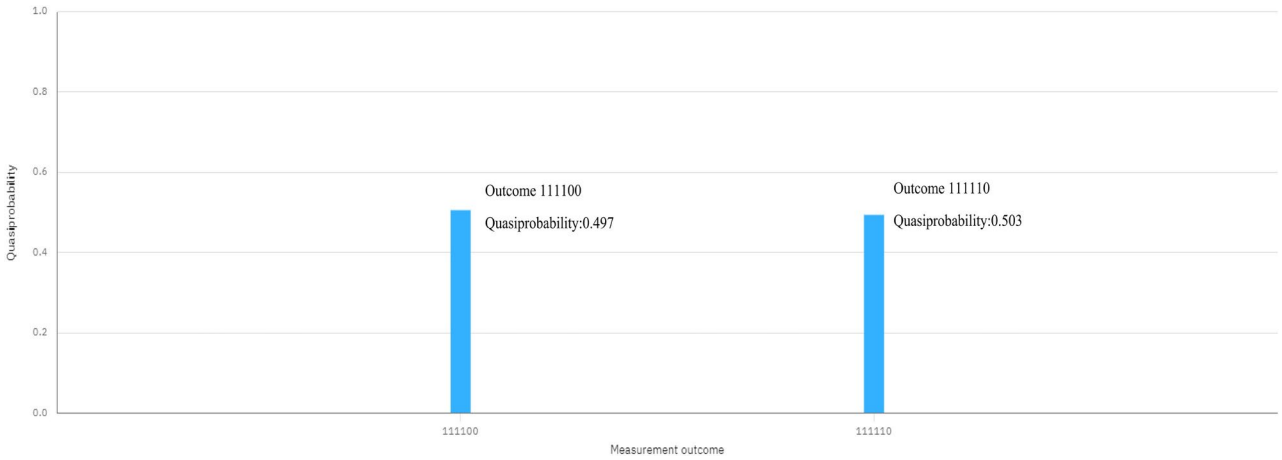

Fig. 2. The probabilities of the two outcomes.

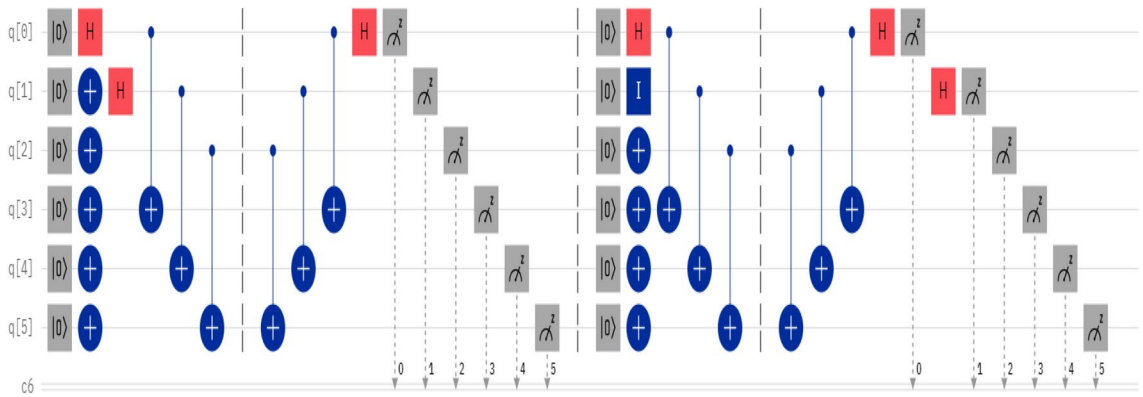

Fig. 3. The circuit diagram.

Protocols	Ref.[13]	Ref.[29]	Ref.[15]	Our protocol
Quantum resources	Single states	Single states and Bell states	Single Photons	Two-particle states and Bell states
Transmisson direction	One	One	One	Two
Query key number	1	1	1	1
Probability of guess	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{\sin^2\theta}{2}$	$\frac{1}{4}$
Quantum efficiency	$\frac{\left(\frac{1}{4}\right)^N}{N+N}$	$\frac{\left(\frac{1}{4}\right)^N}{2N+2N}$	$\frac{\left(\frac{\sin^2\theta}{2}\right)^N}{N+N}$	$\frac{\left(\frac{1}{4}\right)^N}{2N+N}$

Table 3. Comparison between our protocol and some existing protocols.

obtain $|\psi^+\rangle|\psi^-\rangle|10\rangle$ with 49.7% or $|\psi^+\rangle|\psi^+\rangle|10\rangle$ with 50.3%. The probabilities of the two outcomes are shown in Fig. 2. Only when Alice gets $|\psi^+\rangle|\psi^-\rangle|10\rangle$, according to the protocol flow, she can know Bob's second key. The circuit diagram of the above simulation is shown in Fig. 3.

Meanwhile, we also analyze the efficiency of our protocol and compare it with other similar protocols in the section. The comparative results are shown in Table 3. According to Ref.[2], the quantum efficiency is defined as

$$\eta = \frac{s}{q+b} \tag{2}$$

where s is the amount of valid information successfully retrieved, q is the total number of quantum resources utilized, and b is the number of operations performed on the quantum state in the protocol.

In our protocol, since the probability of recovering the correct key is $\frac{1}{4}$ and Ctrl or Shift operation is performed for each quantum state, the efficiency of our protocol is $\frac{\left(\frac{1}{4}\right)^N}{2N+N}$. As shown in Table 3, our protocol ensures efficiency while realizing two-way privacy queries with better security.

Security analysis of our QPQ protocol

In this section, the security of QPQ is analyzed from three aspects: database privacy, user attack and external attack. Our protocol is fully protected against all three of these attacks. The security analysis of the protocol is shown as follows.

Database security analysis

According to our procedure, Alice is thought to be dishonest and will attempt to gain more keys through illegal ways. In order to get more items, Alice will try every way. In this section, we will analyze two kinds of attacks from Alice, such as the fake signal attack and the joint-measurement (JM) attack.

Fake single state attack

Since Alice prepares initial states, she can threaten the database security by sending fake quantum states. However, our proposed protocol can defend against that attack, proving that the operation of Alice is useless. The following is a description of the detailed analysis:

We presume Alice is not being truthful with our protocol, as she can create a fake quantum state.

$$|\eta\rangle = \alpha|00\rangle + \beta|11\rangle + \gamma|10\rangle + \mu|01\rangle \tag{3}$$

and send it to Bob. Upon receiving the quantum state, Bob randomly chooses either the Bell or computational basis for the measurement. Because Bob doesn't know anything about the measured particles, He was unable to select an appropriate measure basis that would have been more advantageous for him. However, there is a $\frac{1}{2}$ chance that Bob selects two basises. As a result, we can compute Bob's four possible outcomes: the potential for obtaining $|10\rangle$ is $\frac{1}{2}\gamma^2$, $\frac{1}{2}\mu^2$ for $|01\rangle$, $\frac{1}{2}(\frac{\gamma+\mu}{\sqrt{2}})^2$ for $|\psi^+\rangle$, $\frac{1}{2}(\frac{\gamma-\mu}{\sqrt{2}})^2$ for $|\psi^-\rangle$. Then, Bob randomly chooses two operations, each of which satisfies the loss tolerance. The operations themselves have no effect on Alice's result. Bob is assumed to select the Ctrl operation as a result.

If Alice decides to measure the qubit she got from Bob using the Bell basis, the outcome will be either $|\psi^-\rangle$ or $|\psi^+\rangle$. If $|\psi^+\rangle$ is the measurement result, Alice could deduce the following probability on Bob's choice of Bell basis:

$$\begin{aligned} P_1 &= P(Bell|result = |\psi^+\rangle) \\ &= \frac{P(Bell, result = |\psi^+\rangle)}{P(result = |\psi^+\rangle)} \\ &= \frac{\frac{1}{2}(\frac{\gamma+\mu}{\sqrt{2}})^2}{\frac{1}{2}(\frac{\gamma+\mu}{\sqrt{2}})^2 + \frac{1}{4}\gamma^2 + \frac{1}{4}\mu^2} \\ &= \frac{\gamma^2+\mu^2+2\gamma\mu}{2\gamma^2+2\mu^2+2\gamma\mu} \end{aligned} \tag{4}$$

The image of this function is shown in the Fig. 4. We know that when $\gamma = \mu$, P_1 can reach the maximum value. Moreover, according to the normalization principle, we can solve that only when $\gamma = \mu = \pm\frac{1}{\sqrt{2}}$, P_1 can reach the maximum value of $\frac{2}{3}$ and the probability that Bob uses computational basis is $P_2 = 1 - P_1 = \frac{1}{3}$.

Given the measurement result of $|\psi^-\rangle$, Alice can calculate the likelihood that Bob will select Bell basis as follows:

$$\begin{aligned} P_3 &= P(Bell|result = |\psi^-\rangle) \\ &= \frac{P(Bell, result = |\psi^-\rangle)}{P(result = |\psi^-\rangle)} \\ &= \frac{\frac{1}{2}(\frac{\gamma-\mu}{\sqrt{2}})^2}{\frac{1}{2}(\frac{\gamma-\mu}{\sqrt{2}})^2 + \frac{1}{4}\gamma^2 + \frac{1}{4}\mu^2} \\ &= \frac{\gamma^2+\mu^2-2\gamma\mu}{2\gamma^2+2\mu^2-2\gamma\mu} \end{aligned} \tag{5}$$

The image of this function is shown in the Fig. 5. We know that when $\gamma = -\mu$, P_3 can reach the maximum value of $\frac{2}{3}$ and the probability that Bob uses computational basis is $P_4 = 1 - P_3 = \frac{1}{3}$.

In a nutshell, if Alice wants to obtain more keys, she must know the measurement basis that Bob uses with the biggest probability. She will use $\gamma = \mu = \pm\frac{1}{\sqrt{2}}$ or $\gamma = 0$ or $\mu = 0$. Alice prepares for four states in our protocol, this being one of them. As a result, our protocol could effectively defend the fake single state attacks.

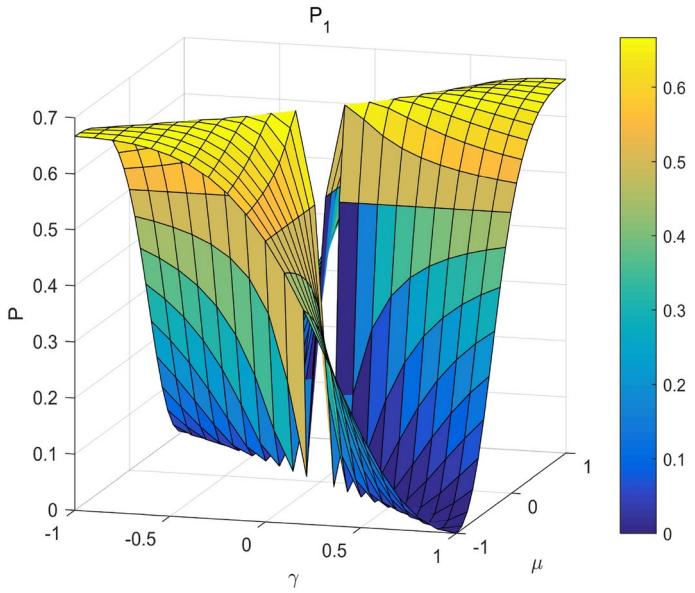

Fig. 4. The image of the function of P_1.

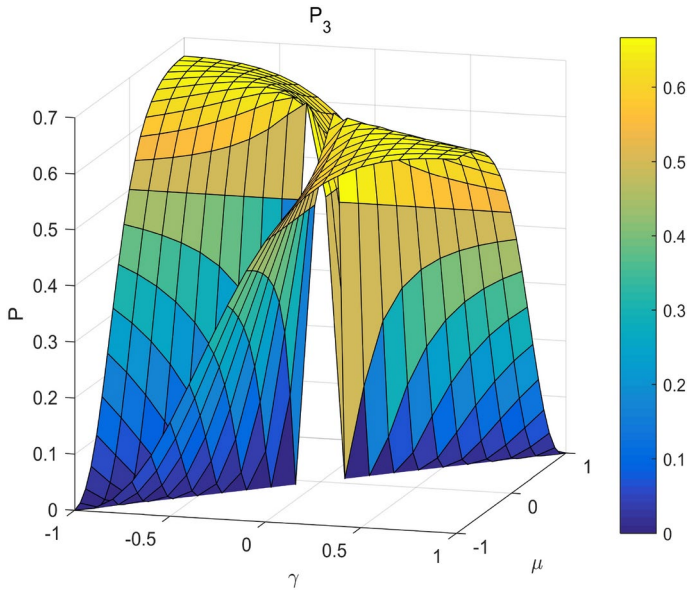

Fig. 5. The image of the function of P_3.

The joint-measurement attack

The joint-measurement (JM) attack allows an attacker to prepare quantum states in advance knowing which quantum bits contribute to the protocol and measure them jointly. The JM attack is considered to be one of the most powerful attacks on database security to date. To gain more access to Bob's original key, Alice will prepare a new sequence of quantum bits through the JM attack and send it to Bob. However, Alice's malicious behavior will be easily detected by Bob.

$$\begin{aligned} |\sigma\rangle &= \frac{1}{\sqrt{2}}[|001\rangle + |110\rangle]_{123} \\ &= \frac{1}{\sqrt{2}}[(|0\rangle + |1\rangle)(|01\rangle + |10\rangle) \\ &\quad + (|0\rangle - |1\rangle)(|01\rangle - |10\rangle)]_{123} \end{aligned} \tag{6}$$

For example, Alice will prepare this quantum state in Step1. Alice keeps the first quantum bit and sends particles 2 and 3 to Bob. When Bob receives the particles, he will randomly choose the computational basis or Bell basis to measure them. We assume that Bob uses the Bell basis for his measurements and his measurement result is $|\psi^+\rangle$. Bob will then ask Alice to disclose the basis and outcome of her measurement. Alice must declare Bob's measuring basis at random because she is unaware of his preference. As an illustration, if she announced "Bell basis and $|\psi^+\rangle$", her malicious behavior would not have been noticed by Bob. But if her announcement is "Bell basis and $|\psi^-\rangle$", Bob will aware that Alice is lying. So the JM attack will be detected easily by Bob. Therefore our protocol can resist the JM attack.

User privacy analysis

Cheat-sensitive means that Bob, the data owner, is dishonest. He will want to get Alice's data even at the risk of being detected. Bob who is malicious may announce some fake quantum states to obtain Alice's information. Our approach is likewise cheat-sensitive in this study. Alice will be able to uncover Bob's malevolent behavior with ease, though. Here's the explanation.

Assume Bob will announce $|01\rangle$ to Alice, even though his outcome state is $|10\rangle$. Then the deception will be detected by Alice with the probability of $\frac{1}{2}$ when the initial state of her preparation is $|10\rangle$. We provide a comprehensive description of how to determine Bob's detected cheat likelihood. Bob, for instance, declares state $|\psi^+\rangle$ but the outcome state or the initial of Alice is $|\psi^-\rangle$. Then, Alice will find the malicious behavior. We assume that the initial state of Alice's preparation is $|\psi^-\rangle$ for event A and Bob's announcement of the result is $|\psi^+\rangle$ for event B. Next, we can figure out how likely it is that Alice discovers Bob's malicious behavior:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(AB) \\ &= \frac{1}{4} + \frac{1}{4} - \frac{1}{8} = \frac{3}{8} \end{aligned} \tag{7}$$

Similarly, if Alice prepares the initial state of $|\psi^+\rangle$ and Bob announces $|\psi^-\rangle$, then Alice will discover Bob's dishonesty with the probability of $\frac{3}{8}$.

In summary, for each particle, Bob's dishonest behavior will be detected by Alice with the probability of $\frac{3}{8}$. For N particles, the probability that Alice discovers Bob's malicious behavior is:

$$P = 1 - \left(1 - \frac{3}{8}\right)^N \tag{8}$$

This probability is approximate to 1 when N is large enough. Therefore our protocol has perfect user privacy. The image of this function is shown in the Fig.6.

In addition, it is important to note that in the above security analysis, we have demonstrated in detail that the participating parties are unable to act maliciously in order to obtain information about each other. However, for Eve, the attacker of the external attack, there are fewer malicious behaviors he can take than the participating parties. Indeed, intercept-measure-resend attack and entangle-measure attack are similar to the internal attacks used in this paper. Therefore, they are not described in detail in this section.

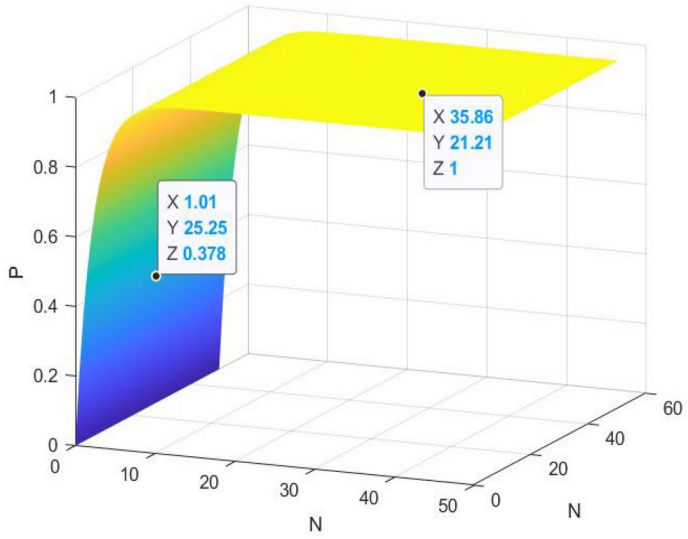

Fig. 6. The image of the function of P.

The application of our QPQ in PSI problem

Above we have given a secure QPQ protocol, however, as a basic protocol, considering its application has been an important issue. In this section, we propose a new PSI protocol for solving the data alignment problem in VFL. Specifically, VFL is a federated learning paradigm where multiple parties collaboratively train a model on datasets with overlapping sample IDs but disjoint feature sets. In VFL, data alignment is a fundamental step that securely identifies the overlapping sample IDs between parties while preserving data privacy. The proposed QPQ protocol serves as a foundational building block for the PSI protocol. It ensures the correctness and integrity of the intersection computation while mitigating risks from malicious parties. By integrating QPQ into PSI, we achieve a secure and efficient data alignment process that underpins VFL. This alignment allows VFL to perform collaborative model training on the overlapping samples, leveraging each party's unique features without revealing any private data. Fig. 7 shows the specific implementation of data alignment in VFL.

Protocol assumption

We assume that two companies, Alice and Bob, expect to jointly train a machine learning model to predict whether the user Charlie is interested in the product. Alice has the purchase data x_C of the user Charlie, while Bob has Charlie's information browsing data $X = \{x_1, \cdots, x_N\}(N \leq n-1)$. Alice wants to know if her data x_C belongs to Bob's database, while neither Alice nor Bob will receive any information from each other. Before model training, Alice and Bob performed cross-domain sample alignment to find the intersection due to different data sources. To make this protocol easier to understand, x_C and x_i are chosen from $Z_n^* = \{1, 2, \cdots, n-1\}$.

Specific implementation of our protocol

Step 1. With the presents QPQ protocol in Section 2, Bob owns the raw key $K_r = \{k_1, \cdots, k_n\}(n \geq N)$ and Alice owns k_i.

Step 2. Firstly, Bob is informed by Alice of a shift value S according to her private data x_C and k_i, i.e., $S = (x_C - i) \mod n$. Next, Bob generates a new shifted key sequence $K_r' = \{k_1', \cdots, k_n'\}$. by cyclically shifting his key string by S bits. Moreover, Bob encrypts K_r' into $K_r'' = \{k_1'', \cdots, k_n''\}$ according to X and sends the sequence K_r'' to Alice.

Encrypting rule

For the subscript j of K_r', if $j \in X$, the matching key bit will change to $k_j'' = k_j' \bigoplus 1$, otherwise $k_j'' = k_j'$. The XOR operation is represented here by $\bigoplus$. Table 4 shows the encoding rules for K_r''.

Step 3. Alice computes the corresponding $q = k_{x_C}'' \bigoplus k_i$. If $q = 1$, Alice's data x_C will belong to Bob's database X; otherwise, $x_C \notin X$.

We will explain our protocol with a specific example below. We suppose that Alice possesses private information $x_C = 6$, Bob owns a database $X = \{1, 2, 4, 5, 7, 9\}$. The raw key sequence is established by $K_r = \{k_1, \cdots, k_{10}\}$, where Alice only knows k_4 but Bob knows the entire key sequence.

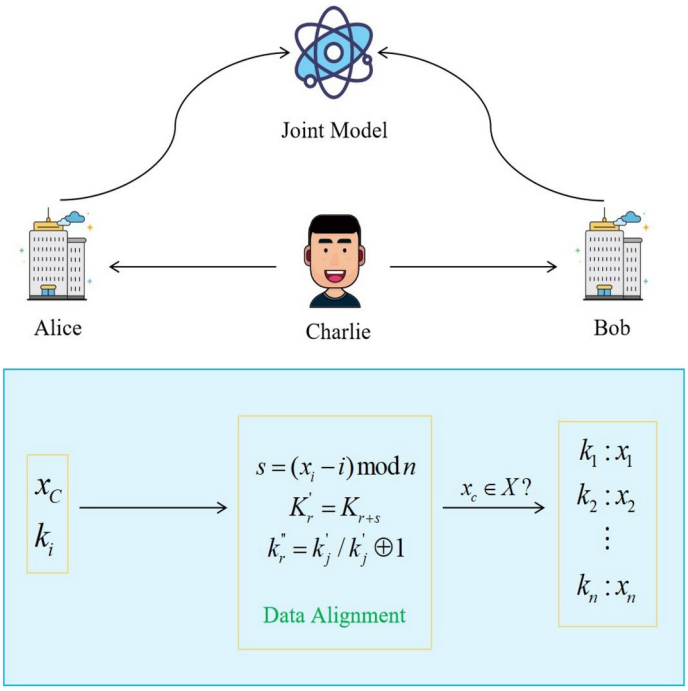

Fig. 7. Specific implementation of data alignment in VFL.

k''_j	k'_j or $k'_j \bigoplus 1$
$j \in X$	$k''_j = k'_j \bigoplus 1$
$j \notin X$	$k''_j = k'_j$

Table 4. Encoding rules for K''_r.

In Step2, Bob is informed by Alice of a shift value of $S = 6 - 4 = 2$. Next, Bob creates a new shifted key sequence $K'_r = \{k'_1, \cdots, k'_{10}\}(k'_{i+2mod10} = k_i)$ by cyclically shifting his key string with 2 bits, here $k'_6 = k_4$. Meanwhile, K'_r is encrypted into $K''_r = \{k''_1, \cdots, k''_{10}\}$ with the presented rules above, where

$$k''_1 = k'_1 \bigoplus 1, k''_2 = k'_2 \bigoplus 1, k''_4 = k'_4 \bigoplus 1 \tag{9}$$

$$k''_5 = k'_5 \bigoplus 1, k''_7 = k'_7 \bigoplus 1, k''_9 = k'_9 \bigoplus 1 \tag{10}$$

$$k''_3 = k'_3, k''_6 = k'_6, k''_8 = k'_8, k''_{10} = k'_{10} \tag{11}$$

In Step 3, Alice computes

$$q = k''_6 \bigoplus k_4 = k'_6 \bigoplus k_4 = 0 \tag{12}$$

and judge her private data 6 does not belong to Bob's database.

Analysis

From the above description, it is clear that the PSI problem requires that no information other than the intersection of the participants be disclosed. Therefore, the key to realizing PSI lies in encrypting the information of both parties. In our proposed protocol, the shift value

$$S = (x_c - i) mod n \tag{13}$$

is determined by Alice's data x_c and the key k_i. Bob encrypts the database through the shift sequence K_r. Furthermore, the most important prerequisite for the PSI problem is that the participant Alice generates a partial key. Fortunately, in Section 2, we propose a new QPQ protocol and prove its security and correctness, which ensures that it does not compromise the privacy of the participating parties. Consequently, the PSI method is also correct and secure.

Conclusion

In this paper, we propose a novel QPQ protocol based on two-particle states, which effectively ensures both user privacy and data owner privacy. The protocol has been simulated using the IBM quantum experience, and the simulation results confirm its correctness. Security analysis further demonstrates that the protocol is resistant to both internal and external attacks. Additionally, we introduce a new privacy set intersection (PSI) method, which is applied for the first time to vertical federated learning (VFL), effectively solving the data alignment problem. This work provides a new approach for privacy-preserving communication and data processing in federated learning. Looking forward, we aim to explore more efficient and universal solutions based on this research and design alternative QPQ protocols to enhance the flexibility and security of privacy-preserving communication.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 1 August 2024; Accepted: 18 February 2025
Published online: 14 October 2025

References

1. Goldreich, O. Secure multi-party computation. *Manuscr. Prelim. Vers.* **78**(110), 1–108 (1998).
2. Chi, Y. P. et al. A new protocol for semi?quantum private set of intersection and union mixed cardinality for any tripartite based on bell states. *Adv. Quant. Technol.* **7**(9), 2400137 (2024).
3. Zhang, Y. et al. A new hybrid protocol that simultaneously achieves quantum multiparty summation and ranking. *Adv. Quant. Technol.* **7**(6), 2400078 (2024).
4. Rabin, M. O. How to exchange secrets with oblivious transfer. *Cryptology* (2005).
5. Gertner, Y. et al. Protecting data privacy in private information retrieval schemes. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, 151–160 (1998).
6. Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 124–134 (IEEE, 1994).
7. Grover, L. K. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, 212–219 (1996)
8. Song, Y. et al. A quantum federated learning framework for classical clients. *Sci. Chin. Phys. Mech. Astron.* **67**(5), 250311 (2024).

9. Song, Y. et al. A resource-efficient quantum convolutional neural network. *Front. Phys.* **12**, 1362690 (2024).
10. Li, L. et al. An efficient quantum proactive incremental learning algorithm. *Sci. Chin. Phys. Mech. Astron.* **68**(1), 210313 (2025).
11. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum private queries. *Phys. Rev. Lett.* **100**, 230502 (2008).
12. Olejnik, L. Secure quantum private information retrieval using phase-encoded queries. *Phys. Rev. A* **84**(2), 022313 (2011).
13. Jakobi, M. et al. Practical private database queries based on a quantum-key-distribution protocol. *Phys. Rev. A* **83**(2), 022301 (2011).
14. Scarani, V. et al. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Phys. Rev. Lett.* **92**(5), 057901 (2002).
15. Gao, F. et al. Flexible quantum private queries based on quantum key distribution. *Opt. Express* **20**(16), 17411–17420 (2012).
16. Zhang, J. L. et al. Private database queries based on counterfactual quantum key distribution. *Phys. Rev. A* **88**(2), 022334 (2013).
17. Wei, C. Y. et al. Practical quantum private query of blocks based on unbalanced-state Bennett–Brassard-1984 quantum-key-distribution protocol. *Sci. Rep.* **4**(1), 7537 (2014).
18. Gao, F. et al. Postprocessing of the oblivious key in quantum private query. *IEEE J. Sel. Top. Quant. Electron.* **21**(3), 98–108 (2014).
19. Wei, C. Y., Wang, T. Y. & Gao, F. Practical quantum private query with better performance in resisting joint-measurement attack. *Phys. Rev. A* **93**(4), 042318 (2016).
20. Qin, L. et al. Decoy-state quantum private query protocol with two-way communication. *Physica A* **633**, 129427 (2024).
21. Zhang, K., Gao, F., Wen, Q.-Y., Liu, B. & Zhu, F.-C. Privacy-preserving decision protocols based on quantum oblivious key distribution. *Comput. Mater. Contin.* **64**(2), 1689–1701 (2020).
22. Freedman, M. J., Nissim, K., & Pinkas, B. Efficient private matching and set intersection, In *International Conference on the Theory and Applications of Cryptographic Techniques*, 1–19 (Springer, 2004).
23. Kissner, L., & Song, D. Privacy-preserving set operations. In *Annual International Cryptology Conference*, 241–257 (Springer, 2005).
24. Huang, Y., Evans, D., & Katz, J. Private set intersection: Are garbled circuits better than custom protocols? NDSS. (2012).
25. Pinkas, B., Schneider, T. & Zohner, M. Scalable private set intersection based on OT extension. *ACM Trans. Privacy Secur (TOPS)* **21**(2), 1–35 (2018).
26. Kolesnikov, V.., Kumaresan, R., Rosulek, M., & Trieu, N. Efficient batched oblivious PRF with applications to private set intersection. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 818–829 (2016).
27. Chen, H., Laine, K., & Rindal, P. Labeled PSI from fully homomorphic encryption with malicious security. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 1223–1237 (2018).
28. Pinkas, B., Schneider, T. & Zohner, M. Scalable private set intersection based on OT extension. *ACM Trans. Privacy Secur. (TOPS)* **21**(2), 1–35 (2018).
29. Liu, D. M. et al. Two quantum private query protocols based on Bell states and single photons. *Mod. Phys. Lett. A* **36**(02), 2150005 (2021).

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62271234, Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications)(SKLNST-2024-1-04), Open Foundation of State Key Laboratory of Public Big Data (Guizhou University) under Grant No. PBD2022-16, the Fundamental Research Funds for Heilongjiang Universities under Grant 2022-KYYWF-1042, Double First-Class Project for Collaborative Innovation Achievements in Disciplines Construction in Heilongjiang Province under Grant No. LJGXCG2022-054 and LJGXCG2023-028, Advanced Programs of Heilongjiang Province for the overseas scholars.

Author contributions

Zeping Deng: Conceptualization, Methodology, Writing - original draft. Hongwei Sun: Supervision, Methodology, Investigation. Kejia Zhang: Supervision, Project administration, Writing - review & editing. Long Zhang: Project administration, Writing - review & editing. Sujuan Qin: Supervision, Writing - review. Tingting Song: Supervision, Modification.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.Z. or L.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.