Enhanced Local Differential Privacy Protection against Crosstalk Attacks in Quantum Computing

Hui Zhong
Dept. of Electrical and Computer Engineering
University of Houston
Houston, United States
hzhong5@uh.edu

Xinyue Zhang
Dept. of Computer Science
Kennesaw State University
Kennesaw, United States
xzhang48@kennesaw.edu

Hao Wang
Dept. of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, United States
hwang9@stevens.edu

Shucheng Yu
Dept. of Graduate Computer Science and Engineering
Yeshiva University
New York, United States
shucheng.yu@yu.edu

Yu Wang
Dept. of Computer and Information Science
Temple University
Philadelphia, United States
wangyu@temple.edu

Miao Pan
Dept. of Electrical and Computer Engineering
University of Houston
Houston, United States
mpan2@uh.edu

2025 IEEE International Conference on Quantum Computing and Engineering (QCE) | 979-8-3315-5736-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/QCE65121.2025.00019

***Abstract*—Quantum computing has gained widespread interest due to its exponential computational capabilities. In practical scenarios, users often access real quantum computers indirectly through cloud-based platforms (e.g., IBM Quantum), which requires transmitting data to third-party servers. Quantum-specific attacks, such as crosstalk attacks, have demonstrated high success rates in inferring the output of legitimate users. These issues raise serious privacy concerns. To protect client-side privacy, quantum local differential privacy (QLDP) has been proposed, where legitimate users perturbed their true output by adding quantum noise to the circuits. However, we observe that the classical local differential privacy (LDP) properties have not been fully adapted to the quantum domain, and the information can still be inferred from the perturbed output if attackers access the noise type added by legitimate users. To fill this gap, we propose a novel QLDP-based approach to protect the true output of legitimate users. We find that QLDP can be achieved using only simple quantum noise, but not all types of quantum noise can effectively perturb the output under different quantum measurements. In addition, to prevent advanced attackers who have partial user information, we introduce a probabilistic noise addition mechanism. To allow legitimate users to accurately estimate the true output of a quantum circuit, we also propose a new quantum frequency estimation. Our approach is validated using real quantum computers and quantum simulators, achieving 94% accuracy and 90% utility to estimate the true output from the perturbed output.**

***Index Terms*—Quantum computing, Quantum local differential privacy, Quantum privacy**

I. Introduction

Quantum computing is a rapidly growing field that has exponential computational capability over classical computing. It is particularly effective for solving large-scale optimization problems and big data analytics [1]–[3]. As quantum technologies mature, they are widely applied in domains such as cryptographic security [4], chemistry science [5], and finance risk analysis [6]. However, quantum computing faces significant privacy concerns. Quantum attacks such as crosstalk attacks [7] and qubit reset attacks [8] have been demonstrated to effectively extract output of quantum computing, leading to unintended data exposure [9]. For instance, in a financial scenario where qubits are used to evaluate investment strategies, an output of "1" may indicate an investment decision, while "0" indicates rejection. If attackers can infer users' outputs, it could compromise strategic decisions and financial information. Thus, it is critical to find methods to protect privacy in quantum computing.

In classical computing, differential privacy (DP) is a widely adopted framework for preserving data privacy [10]–[12]. DP primarily protects continuous data by introducing artificial noise (e.g., Gaussian noise) to blur the output distribution. The purpose of DP is to ensure that the inclusion or exclusion of individual data points does not significantly affect the output, thus preventing adversaries from inferring sensitive information according to the output. Moreover, DP has rigorous mathematical guarantees, offering a quantifiable measure of privacy protection by parameter ϵ. It has been successfully applied in domains such as data mining [13], smart devices [14], and healthcare [15]. A notable extension of DP is local differential privacy (LDP), which is particularly suited for protecting discrete data [16]–[18]. Unlike traditional DP, which assumes a trusted server, LDP can operate under an untrusted server model. In this setting, noise (e.g., randomized response) is applied directly to user inputs before they are sent to the server, ensuring that even the insecure server cannot directly obtain the original user data. Additionally, LDP allows legitimate users to estimate the true distribution of the data through frequency estimation, enabling useful data analysis while maintaining privacy.

For quantum computing, there are several existing methods to protect data privacy. Saki et al. [19] proposed a simple protection mechanism against crosstalk attacks. An additional X gate is applied before measurement to flip the true output,

979-8-3315-5736-2/25/$31.00 ©2025 IEEE
DOI 10.1109/QCE65121.2025.00019

misleading potential attackers. Extending this idea to multi-qubit circuits, Maurya et al. [20] suggested a protection approach in which an X gate is applied to a randomly chosen qubit, which can flip part of the output and introduce uncertainty. However, these methods have several limitations. The first drawback is that attackers may still be able to infer the true output. If an advanced attacker gains knowledge of the noise type introduced by legitimate users, they can reverse the final result to recover the correct output. The second drawback concerns the security of the registers used by legitimate users. Legitimate users must record the positions of the X gates to retrieve the correct output when needed. However, if an advanced attacker gains access to or manipulates the register storing X gates' information, they can reconstruct or change the intended results, further compromising output privacy.

Quantum differential privacy (QDP) is also a promising framework to protect server-side data based on classical DP [21]–[23]. For example, Zhou et al. [22] proposed three quantum noise mechanisms to realize privacy protection in quantum computing. Similarly, Li et al. [24] demonstrated that quantum measurement noise could serve as a protective mechanism against privacy leakage. In practical quantum computing, users typically access quantum computers remotely through cloud platforms, submitting their data to unfamiliar servers. Moreover, quantum states are frequently transmitted between different user data nodes [25]. These factors highlight the critical need to address client-side data privacy. Consequently, quantum local differential privacy (QLDP) has been proposed as a variant of QDP, though it has been explored in only a limited number of works [25]–[27]. The input of QLDP can be a broader set of quantum states, which can even include all possible quantum input states [25], [27]. It means that QLDP can protect not only the initial quantum state input of the client but also the quantum state after quantum computation or transmitted through a quantum communication channel. To realize QLDP, various noise mechanisms have been proposed. Angrisani et al. [26] found that different quantum measurement operators can perturb user data. Nuradha et al. [28] proposed that adding depolarizing noise in quantum computing can achieve a similar effect as the classical randomized response (RR) mechanism. However, we have identified the third drawback of existing protection methods: these approaches have not fully translated the classical LDP properties into the quantum domain. For example, the depolarizing noise mechanism results in a complete transition of the quantum state into the maximally mixed state, preventing the use of frequency estimation which is crucial for extracting meaningful information from noisy output.

Based on the above, we propose a new QLDP-based quantum privacy protection method. We first demonstrate that the fundamental quantum noise gates (X, Y, or Z gate) can effectively perturb outputs and develop a new RR mechanism, thereby protecting user information and achieving QLDP. Furthermore, our mathematical derivations reveal that not all types of quantum noise can be used for effective data perturbation. We also make improvements for the drawbacks mentioned above. For the first one, we try to add noise gates with varying probabilities, making it more difficult for attackers to infer the true output even if they obtain the type of additional noise. For the second and third drawbacks, we take advantage of frequency estimation in classical LDP and extend this concept to the quantum domain, enabling legitimate users to estimate the true output without registers while ensuring data protection. Our contributions can be summarized as follows:

- We find that simple quantum noise (X, Y, Z gate) can perturb the output and implement a new RR mechanism of QLDP. We also demonstrate that not all types of quantum noise can effectively perturb outputs under different measurement operators.
- We further enhance privacy protection by adjusting the probability of adding noise gates to prevent advanced attackers with partial user information. Moreover, we derive the frequency estimation in the quantum domain, allowing legitimate users to estimate the true output.
- We conduct extensive experiments using both real quantum hardware and quantum simulators, demonstrating the effectiveness of our QLDP-based privacy-preserving approach, achieving 94% accuracy and 90% utility to help estimate the true output.

The rest of this paper is organized as follows. Section II introduces the basic background of quantum computing, QDP, and QLDP. Section III presents the threat model. Section IV provides a mathematical analysis of simple quantum noise that can achieve output perturbation under different quantum measurements. Section V proposes a frequency estimation method for quantum domain under QLDP. Section VI demonstrates the effectiveness of our methods through experiments on real quantum computers and quantum simulators. Finally, we discuss future research directions and conclude the paper.

II. Preliminaries

In this section, we first introduce the basic architecture of quantum computing, including quantum states, quantum gate operations, and quantum measurements. Next, we review the concepts of DP and QDP under classical computing and quantum computing, respectively.

A. Quantum Computing

A quantum computer is a computing device based on the principles of quantum mechanics, including digital quantum computers, topological quantum computers, etc [29]–[31]. In this work, we focus on digital quantum computers, which use quantum gates to construct circuits to perform corresponding quantum algorithms. These systems consist of three fundamental components: quantum states, quantum gates, and quantum measurements [32].

Digital quantum computers use quantum bits (qubits) $|0\rangle$ and $|1\rangle$ as the basic units, corresponding to bits 0 and 1 in classical computers. A quantum state is the mathematical description of qubits and can be categorized as pure and mixed quantum states. A pure quantum state describes a completely known state of the quantum system. In a two-dimensional

Hilbert space, the pure quantum state can be a single basic state ($|0\rangle = \begin{bmatrix}1 & 0\end{bmatrix}^T$ or $|1\rangle = \begin{bmatrix}0 & 1\end{bmatrix}^T$) or a superposition of $|0\rangle$ and $|1\rangle$: $|\psi\rangle = \alpha_1 |0\rangle + \alpha_2 |1\rangle = \begin{bmatrix}\alpha_1 & \alpha_2\end{bmatrix}^T \in \mathbb{C}^2$, where the complex numbers α_1 and α_2 satisfy $|\alpha_1|^2 + |\alpha_2|^2 = 1$. In a 2^n-dimensional Hilbert space (n is the number of qubits), the pure quantum state can be represented as $|\psi\rangle = \sum_{i=1}^{2^n} \alpha_i |i\rangle \in \mathbb{C}^{2^n}$, which satisfies $|\alpha_1|^2 + ... + |\alpha_{2^n}|^2 = 1$. The density matrix can also describe the pure state, denoted as $\rho = |\psi\rangle \langle\psi|$. $\langle\psi|$ is the conjugate transpose of $|\psi\rangle$ (e.g., $\langle 0| = \begin{bmatrix}1 & 0\end{bmatrix}$) and $\mathrm{Tr}(\rho) = 1$, which means the sum of the diagonal elements of the density matrix ρ is 1. A mixed quantum state describes a non-completely known state of the quantum system and is denoted by the density matrix $\rho = \sum_{i=1}^{2^n} p_i |\psi\rangle \langle\psi|$. We can find that the mixed quantum state is represented by a mixture of multiple pure states with different probabilities. In an ideal environment, a quantum system remains in a pure state. When the actual system is affected by the environment, it leads to decoherence and produces a mixed state.

Quantum gates are similar to logic gates in classical computers but with quantum properties. Quantum gates consist of single qubit gates and multiple qubits gates. Quantum gates can be expressed as the product of unitary matrices U, where U satisfies $U^\dagger U = UU^\dagger = I$. $U^\dagger$ denotes the conjugate transpose of U, and I is the unit matrix. Single qubit gates act on individual qubits, e.g., an X-gate corresponds to an amplitude flip of a qubit; a Z-gate corresponds to a phase flip of a qubit. These gates can all be represented as distinct 2×2 unitary matrices. Multiple qubits gates act on multiple qubits, e.g., the CNOT gate is a two qubits gate that conditionally flips the target qubit based on the state of the control qubit. It can be represented as a 4×4 unitary matrix. By combining quantum gates in specific sequences, quantum states can be transformed into desired states, realizing various quantum algorithms. For example, when a quantum state ρ passes through a quantum gate U, it evolves into a new quantum state $\rho' = U\rho U^\dagger$.

Quantum measurements are used to convert the quantum state into classical information after it passes through the circuit. The measurement results are the probability distributions over the possible output values of the circuit. A general framework for quantum measurements is the Positive Operator-Valued Measure (POVM), which consists of a set of positive semi-definite matrices [33]–[35]. These matrices are called measurement operators $\{M_m\}_{m \in O}$ and satisfy $\sum_m M_m = I$, where I is the identity operator, m is the classical results, O is the set of possible outcomes. Since a single measurement provides only one random result, repeated measurements are necessary to infer the probability distribution. So for a quantum state ρ measured several times, the probability of obtaining the outcome m is given by $p_m = \mathrm{Tr}\left(M_m^\dagger M_m \rho\right)$. As an example, consider Z-basis measurements. The measurement operators in this case are M_0 and M_1. When a quantum state ρ is measured, the probabilities of obtaining the classical outcomes 0 and 1 are given by $p_0 = \mathrm{Tr}\left(M_0^\dagger M_0 \rho\right)$ and $p_1 = \mathrm{Tr}\left(M_1^\dagger M_1 \rho\right)$. These results correspond to the projection of the quantum state onto the measurement operators. If the circuit uses projective measurements (e.g. Pauli measurements), a special case of POVM, the measurements can also be expressed as $p_m = \mathrm{Tr}(M_m \rho)$.

In this paper, we utilize Pauli measurements at the end of the quantum circuit. Pauli measurements are the most fundamental class of measurements which consists of three Pauli operators (X, Y, Z) [36]. A more detailed mathematical analysis of Pauli measurements is provided in Section IV.

B. Differential Privacy

By artificially adding noise, classical DP perturbs the computing outputs and effectively prevents attackers from inferring a particular user's privacy. A notable advantage is that DP has a rigorous mathematical derivation, and can quantify privacy protection level by parameter ϵ. The definition of DP can be expressed as follows [37].

*Definition 1 (***Classical Differential Privacy):** A randomized function $\mathcal{K}$ satisfies (ϵ, δ)-differential privacy if the data sets D and D′ differ by only one participant, and every subset S of outcomes satisfy

$$\Pr[\mathcal{K}(D) \in S] \leq e^\epsilon \cdot \Pr[\mathcal{K}(D') \in S] + \delta, \tag{1}$$

where ϵ is the privacy budget and δ is the failure probability.

Similar to DP, QDP perturbs the output by adding either classical or quantum noise. The definition of QDP can be expressed as follows [38].

*Definition 2 (***Quantum Differential Privacy):** Given two quantum datasets ρ and σ with $\tau(\rho, \sigma) \leq d$, where $d \in (0, 1]$. $\tau(\rho, \sigma) = \|\rho - \sigma\|_1$ indicates the trace distance, which is used to define quantum neighboring datasets. A quantum operation $\mathcal{E}$ satisfies (ϵ, δ)-differentially private if every $POVM$ (Positive Operator-Valued Measure) $M = \{M_m\}$ and every subset S of outcomes satisfy,

$$\Pr[M(\mathcal{E}(\rho)) \in S] \leq e^\epsilon \cdot \Pr[M(\mathcal{E}(\sigma)) \in S] + \delta, \tag{2}$$

where ϵ is the privacy budget and δ is the failure probability.

Specifically, ϵ represents the level of privacy protection, where a smaller ϵ indicates stronger privacy protection.

C. Local Differential Privacy

LDP is a variant of DP that protects the privacy of user input data. By artificially adding noise to the input, LDP perturbs the input data before transmission, effectively preventing untrusted servers from obtaining true data. The definition of LDP can be expressed as follows [16].

*Definition 3 (***Local Differential Privacy):** For any pairs of input values D and D′, a randomized noise mechanism $\mathcal{K}$ satisfies ϵ-LDP if and only if for any subset Y of outcomes, it holds

$$\Pr[\mathcal{K}(D) \in Y] \leq e^\epsilon \cdot \Pr[\mathcal{K}(D') \in Y], \tag{3}$$

where $\Pr[\cdot]$ is the probability, ϵ is the privacy budget.

Randomized Response (RR) is a classical technique to implement LDP. An example is given below to understand the relationship between RR and LDP. Legitimate data collectors seek to determine the true number of smokers among N users.

Each user is required to consider the question "Are you a smoker?" and can respond with either "yes" or "no". To protect the users' response, each user employs an RR mechanism before answering. Specifically, the user flips a biased coin: with probability p, the user tells the truth; with probability $1-p$, the user tells a lie. Suppose that the total number of users saying "yes" is N_1 after applying RR, we can now estimate the true number of smokers $\hat{f}$ by frequency estimation,

$$\hat{f} = \frac{p - 1 + N_1/N}{2p - 1}. \tag{4}$$

Similar to LDP, QLDP perturbs the input by adding quantum noise. The definition of QLDP can be expressed as follows [25].

*Definition 4 (***Quantum Local Differential Privacy):** For any quantum states ρ and σ, a quantum operation $\mathcal{E}$ satisfies ϵ-QLDP if and only if every POVM $M = \{M_m\}$ and every subset Y of outcomes satisfy,

$$\Pr\left[M\left(\mathcal{E}\left(\rho\right)\right) \in Y\right] \leq e^{\epsilon} \cdot \Pr\left[M\left(\mathcal{E}\left(\sigma\right)\right) \in Y\right], \tag{5}$$

where $\Pr\left[\cdot\right]$ is the probability, ϵ is the privacy budget.

Specifically, ϵ represents the level of privacy protection, where a smaller ϵ indicates stronger privacy protection.

Multiple types of quantum noise have been shown to implement RR mechanisms and thereby achieve QLDP, but none have introduced frequency estimation to the quantum domain and combined it with defense mechanisms against output attacks [25]–[27]. Hence, further research is necessarily needed. A detailed comparison between our work and previous work is presented in Section VI-A and Table II.

III. Threat Model

Existing works have demonstrated that specific quantum attacks can successfully infer user output with high accuracy, such as crosstalk attacks [19]. Crosstalk refers to unintended interactions between qubits in the same quantum computer due to physical coupling. These interactions can lead to some problems, such as introducing noise into idle qubits or causing state leakage during qubit reset. Such effects can reduce the performance of quantum computing and pose risks to the data privacy of legitimate users [39], [40].

In our work, we focus on measurement-induced crosstalk [9], where the attacker infers the legitimate user's output information by comparing their own output to the legitimate user's. This specific attack process can be found in [19], where experiments have shown that the successful rate of the crosstalk attack is 96%. To establish a favorable scenario for the attacker, the threat model adopts the following assumptions as [19]. The attacker knows the number of qubits of the legitimate user's circuit. Both the legitimate user and attacker circuits use the same measurement basis. Since the output of quantum computing is highly likely to be leaked, it is crucial to explore appropriate privacy-preserving methods.

We consider two attack scenarios as [8] that can be defended to make our privacy-preserving approach more practical. First, the attacker has full knowledge of the legitimate user's algorithm. For example, the legitimate user is implementing Shor's algorithm for integer factorization but does not know the input number. Second, the attacker has only partial information about the legitimate user's algorithm. For example, the attacker knows that the legitimate user is executing a quantum error correction protocol but lacks details about the specific error model. Our approach can protect privacy in both cases.

IV. Defense Framework

As we mentioned in Section III, with the crosstalk attack, the attacker can infer the legitimate user's output with high probability and leak the user's privacy. One defense against crosstalk attack was proposed in [19]. The legitimate user perturbs the final output distribution by adding an X gate with a 100% probability after their circuit, before the measurement. This perturbation ensures that the attacker can only infer the legitimate user's perturbed output, which does not reveal the user's true information, thereby protecting the user's data privacy. However, we found that the X-gate is not always effective when different measurements are used. To ensure robust privacy protection, other types of noise gates must be added to protect output under different quantum measurements.

To address the limitations of existing defense methods, we conduct a more comprehensive analysis of the relationship between noise and user output. Unlike previous work that primarily focuses on Z-basis measurements [19], we consider three of the most common and fundamental measurement bases: X, Y, and Z-basis measurements, which can also be called Pauli measurements [41]. For perturbing the legitimate user's output, we analyze the different effects by adding simple noise (e.g., X, Y, or Z noise gates) after the legitimate user circuits and before the measurements. Notably, this approach simultaneously satisfies the requirements of QLDP. The legitimate users can recover its true output through frequency estimation and the details can be found in Section V.

We summarize the results after perturbation in Table I. The first column represents the type of noise added after the legitimate users' circuit and before the measurements, while the first row specifies the measurement basis used by the legitimate user and attacker. For each measurement type, $P\left(\cdot\right)$ denotes the probability distribution of the measurement outcomes, and the column "Protect data" indicates whether the data is successfully protected under the given setup. Table I demonstrates that a simple noise gate can effectively perturb the output under different measurements, thereby protecting privacy. However, under the same measurement settings, not all noise gates can achieve the output perturbation. The detailed analysis for Table I is explained in Section IV-A- IV-C.

We first compute the generic quantum state used in Section IV-A- IV-C. We denote the quantum state of the legitimate user after passing through the circuit as $|\rho\rangle = a\,|0\rangle + be^{i\phi}\,|1\rangle$, where a and b are real numbers satisfying $a^2 + b^2 = 1$. Here, $e^{i\phi}$ represents the relative phase, with ϕ being the phase

angle. ρ can also be expressed as a density matrix for ease of computation, i.e.

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} a \\ be^{i\phi} \end{bmatrix} \begin{bmatrix} a & be^{-i\phi} \end{bmatrix} = \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}.$$

When an X, Y, or Z gate noise is applied after the quantum state ρ, the new quantum state can be represented as the following matrix respectively,

$$\begin{aligned} \rho_1^{'} = X\rho X^{\dagger} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} b^2 & abe^{-i\phi} \\ abe^{i\phi} & a^2 \end{bmatrix}, \\ \rho_2^{'} = Y\rho Y^{\dagger} &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\ &= \begin{bmatrix} b^2 & -abe^{-i\phi} \\ -abe^{i\phi} & a^2 \end{bmatrix}, \\ \rho_3^{'} = Z\rho Z^{\dagger} &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} a^2 & -abe^{-i\phi} \\ -abe^{i\phi} & b^2 \end{bmatrix}. \end{aligned}$$

Next, we calculate the output distribution of legitimate users under different measurement operators.

A. Z-basis Measurement

First, we analyze the effect of the X, Y, or Z noise gates on the legitimate user's output when both the legitimate user and the attacker perform Z-basis measurements. The basis vectors for Z-basis measurements are $|0\rangle$ and $|1\rangle$. The measurement operators are

$$M_0 = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, M_1 = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

In the initial case, where no noise is added after the legitimate user circuit, the probabilities of the final output (i.e., the classical outcomes 0 and 1) are given by

$$\begin{aligned} P_0 &= \mathrm{Tr}\left(M_0\rho\right) = \mathrm{Tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}\right) = a^2, \\ P_1 &= \mathrm{Tr}\left(M_1\rho\right) = \mathrm{Tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a^2 & abe^{i\phi} \\ abe^{-i\phi} & b^2 \end{bmatrix}\right) = b^2. \end{aligned}$$

We now add an X, Y, or Z noise gate after the original circuit and before the measurement. When an X gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_1^{'} = X\rho X^{\dagger} = \begin{bmatrix} b^2 & abe^{-i\phi} \\ abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes 0 and 1 are given by

$$\begin{aligned} P_0 &= \mathrm{Tr}\left(M_0\rho_1'\right) = \mathrm{Tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) = b^2, \\ P_1 &= \mathrm{Tr}\left(M_1\rho_1'\right) = \mathrm{Tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) = a^2. \end{aligned}$$

When a Y gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_2^{'} = Y\rho Y^{\dagger} = \begin{bmatrix} b^2 & -abe^{-i\phi} \\ -abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes 0 and 1 are given by $P_0 = \mathrm{Tr}\left(M_0\rho_2^{'}\right) = b^2$, $P_1 = \mathrm{Tr}\left(M_1\rho_2^{'}\right) = a^2$. When a Z gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_3^{'} = Z\rho Z^{\dagger} = \begin{bmatrix} a^2 & -abe^{-i\phi} \\ -abe^{i\phi} & b^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes 0 and 1 are given by $P_0 = \mathrm{Tr}\left(M_0\rho_3^{'}\right) = a^2$, $P_1 = \mathrm{Tr}\left(M_1\rho_3^{'}\right) = b^2$.

Based on the above analysis under Z-basis measurements, we can observe that when an X or Y noise gate is added, the final output distribution of legitimate users changes. In other words, even if the attacker obtains the legitimate user's output through an attack, the information is no longer the true output. However, when a Z noise gate is added, the final output distribution of legitimate users remains unchanged. As a result, it fails to protect the information effectively.

B. X-basis Measurement

Second, we analyze the effect of the X, Y, or Z noise gates on the legitimate user's output when both the legitimate user and the attacker perform X-basis measurements. The basis vectors for X-basis measurements are $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$. The measurement operators are

$$M_+ = |+\rangle\langle +| = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, M_- = |-\rangle\langle -| = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

In the initial case, where no noise is added after the legitimate user circuit, the probabilities of the final output (i.e., the classical outcomes $+$ and $-$) are

$$\begin{aligned} P_+ &= \mathrm{Tr}\left(M_+\rho\right) = \mathrm{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2 + b^2 + 2ab\cos\phi\right) = \frac{1}{2}\left(1 + 2ab\cos\phi\right), \\ P_- &= \mathrm{Tr}\left(M_-\rho\right) = \mathrm{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2 + b^2 - 2ab\cos\phi\right) = \frac{1}{2}\left(1 - 2ab\cos\phi\right). \end{aligned}$$

Similarly, we add an X, Y, or Z noise gate after the original circuit and before the measurement. When an X gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_1^{'} = X\rho X^{\dagger} = \begin{bmatrix} b^2 & abe^{-i\phi} \\ abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes $+$ and $-$ are

$$\begin{aligned} P_+ &= \mathrm{Tr}\left(M_+\rho_1^{'}\right) = \mathrm{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2 + b^2 + abe^{i\phi} + abe^{-i\phi}\right) = \frac{1}{2}\left(1 + 2ab\cos\phi\right), \\ P_- &= \mathrm{Tr}\left(M_-\rho_1^{'}\right) = \mathrm{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2 + b^2 - abe^{i\phi} - abe^{-i\phi}\right) = \frac{1}{2}\left(1 - 2ab\cos\phi\right). \end{aligned}$$

	Z-basis measurement			X-basis measurement			Y-basis measurement		
	P(0)	P(1)	Protect data	P(+)	P(-)	Protect data	P(+i)	P(-i)	Protect data
No noise	a^2	b^2		$\frac{1}{2}[1+2ab\cdot\cos\phi]$	$\frac{1}{2}[1-2ab\cdot\cos\phi]$		$\frac{1}{2}[1+2ab\cdot\sin\phi]$	$\frac{1}{2}[1-2ab\cdot\sin\phi]$	
Add X gate	b^2	a^2	✓	$\frac{1}{2}[1+2ab\cdot\cos\phi]$	$\frac{1}{2}[1-2ab\cdot\cos\phi]$	✗	$\frac{1}{2}[1-2ab\cdot\sin\phi]$	$\frac{1}{2}[1+2ab\cdot\sin\phi]$	✓
Add Y gate	b^2	a^2	✓	$\frac{1}{2}[1-2ab\cdot\cos\phi]$	$\frac{1}{2}[1+2ab\cdot\cos\phi]$	✓	$\frac{1}{2}[1+2ab\cdot\sin\phi]$	$\frac{1}{2}[1-2ab\cdot\sin\phi]$	✗
Add Z gate	a^2	b^2	✗	$\frac{1}{2}[1-2ab\cdot\cos\phi]$	$\frac{1}{2}[1+2ab\cdot\cos\phi]$	✓	$\frac{1}{2}[1-2ab\cdot\sin\phi]$	$\frac{1}{2}[1+2ab\cdot\sin\phi]$	✓

✓ means that the output can be protected; ✗ means that the output cannot be protected.

TABLE I
MEASUREMENT RESULTS UNDER DIFFERENT MEASUREMENTS.

When a Y gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_2^{'} = Y\rho Y^{\dagger} = \begin{bmatrix} b^2 & -abe^{-i\phi} \\ -abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes $+$ and $-$ are given by $P_+ = \text{Tr}\left(M_+\rho_2^{'}\right) = \frac{1}{2}(1-2ab\cos\phi)$, $P_- = \text{Tr}\left(M_-\rho_2^{'}\right) = \frac{1}{2}(1+2ab\cos\phi)$. When a Z gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_3^{'} = Z\rho Z^{\dagger} = \begin{bmatrix} a^2 & -abe^{-i\phi} \\ -abe^{i\phi} & b^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes + and - are given by $P_+ = \text{Tr}\left(M_+\rho_3^{'}\right) = \frac{1}{2}(1-2ab\cos\phi)$, $P_- = \text{Tr}\left(M_-\rho_3^{'}\right) = \frac{1}{2}(1+2ab\cos\phi)$.

Based on the above analysis under X-basis measurements, we can observe that when a Y or Z noise gate is added, the final output distribution of legitimate users changes. In other words, even if the attacker obtains the legitimate user's output through an attack, the information is no longer the true output. However, when an X noise gate is added, the final output distribution of legitimate users remains unchanged. As a result, it fails to protect the information effectively.

C. Y-basis Measurement

Third, we analyze the effect of the X, Y, or Z noise gates on the legitimate user's output when both the legitimate user and the attacker perform Y-basis measurements. The basis vectors for Y-basis measurements are $|+i\rangle = \frac{|0\rangle+i|1\rangle}{\sqrt{2}}$ and $|-i\rangle = \frac{|0\rangle-i|1\rangle}{\sqrt{2}}$. The measurement operators are

$$M_{+i} = |i\rangle\langle i| = \frac{1}{2}\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}, M_{-i} = |-i\rangle\langle -i| = \frac{1}{2}\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}.$$

In the initial case, where no noise is added after the legitimate user circuit, the probabilities of the final outputs (i.e., the classical outcomes $+i$ and $-i$) are given by

$$\begin{aligned} P_{+i} &= \text{Tr}\left(M_{+i}\rho\right) = \text{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}\begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2+b^2+2ab\sin\phi\right) = \frac{1}{2}\left(1+2ab\sin\phi\right), \\ P_{-i} &= \text{Tr}\left(M_{-i}\rho\right) = \text{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}\begin{bmatrix} a^2 & abe^{-i\phi} \\ abe^{i\phi} & b^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2+b^2-2ab\sin\phi\right) = \frac{1}{2}\left(1-2ab\sin\phi\right). \end{aligned}$$

Similarly, we add an X, Y or Z noise gate after the original circuit and before the measurement. When the X gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_1^{'} = X\rho X^{\dagger} = \begin{bmatrix} b^2 & abe^{-i\phi} \\ abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes $+i$ and $-i$ are

$$\begin{aligned} P_{+i} &= \text{Tr}\left(M_{+i}\rho_1^{'}\right) = \text{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}\begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2+b^2-iabe^{i\phi}+iabe^{-i\phi}\right) = \frac{1}{2}\left(1-2ab\sin\phi\right), \\ P_{-i} &= \text{Tr}\left(M_{-i}\rho_1^{'}\right) = \text{Tr}\left(\frac{1}{2}\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}\begin{bmatrix} b^2 & abe^{i\phi} \\ abe^{-i\phi} & a^2 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(a^2+b^2+iabe^{i\phi}-iabe^{-i\phi}\right) = \frac{1}{2}\left(1+2ab\sin\phi\right). \end{aligned}$$

When a Y gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_2^{'} = Y\rho Y^{\dagger} = \begin{bmatrix} b^2 & -abe^{-i\phi} \\ -abe^{i\phi} & a^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes $+i$ and $-i$ are given by $P_{+i} = \text{Tr}\left(M_{+i}\rho_2^{'}\right) = \frac{1}{2}(1+2ab\sin\phi)$, $P_{-i} = \text{Tr}\left(M_{-i}\rho_2^{'}\right) = \frac{1}{2}(1-2ab\sin\phi)$. When a Z gate is added after the original circuit, the quantum state can be represented as the matrix $\rho_3^{'} = Z\rho Z^{\dagger} = \begin{bmatrix} a^2 & -abe^{-i\phi} \\ -abe^{i\phi} & b^2 \end{bmatrix}$. The probabilities of obtaining the classical outcomes $+i$ and $-i$ are given by $P_{+i} = \text{Tr}\left(M_{+i}\rho_3^{'}\right) = \frac{1}{2}(1-2ab\sin\phi)$, $P_{-i} = \text{Tr}\left(M_{-i}\rho_3^{'}\right) = \frac{1}{2}(1+2ab\sin\phi)$.

Based on the above analysis under Y-basis measurements, we can observe that when an X or Z noise gate is added, the final output distribution of legitimate users changes. In other words, even if the attacker obtains the legitimate user's output through an attack, the information is no longer the true output. However, when a Y noise gate is added, the final output distribution of legitimate users remains unchanged. As a result, it fails to protect the information effectively.

V. QUANTUM LOCAL DIFFERENTIAL PRIVACY

Our next objective is to ensure that the true output remains inaccessible to the attacker, while allowing the legitimate user to infer the true output. It is very important to determine whether the legitimate user can effectively utilize the data. For random noise gates, a common approach to obtain the true output is to record whether an X-gate was applied during each shot in a register [39], [40]. The legitimate user can then retrieve the true output by reversing the noise operation using the recorded information. However, for crosstalk attacks, since

the legitimate user and the attacker are in the same quantum computer, the attacker can likely access these registers, potentially tampering with the recorded information or inferring the output. For fixed noise gates, the limitation is that due to the simple method of perturbation (adding X-gate), once the attacker discovers the legitimate user's perturbation method, he/she can reverse the flipped output and retrieve the legitimate user's true information.

Therefore, we aim to propose a more advanced defense mechanism. In our method, the legitimate user perturbs the final output by adding noise with a probability p after the circuit. Even if the attacker is aware of the noise type added, they cannot infer the true output without knowing the exact probability p. In this setting, we find that QLDP can be satisfied. Specifically, we take the quantum state after the circuit as the input to the QLDP, apply a noise gate as the noise mechanism, and obtain a final output that satisfies QLDP privacy preservation. Additionally, we take inspiration from classical Local Differential Privacy (LDP) and propose a method that eliminates the need for secure registers. Specifically, we introduce frequency estimation in the quantum domain to infer the true output of the legitimate user circuit while preventing the attacker. QLDP and the corresponding properties (e.g., quantum frequency estimation) are described below.

A. QLDP Model

1) Definition: We follow the Definition 4 for our QLDP-based protection approach.

2) The Noise Mechanisms to Achieve QLDP: Randomized Response (RR), a classical technique to implement LDP, can similarly realize QLDP. The core idea of RR is to add probabilistic noise while collecting data, so that the attacker can not get the real information, but the legitimate user can still infer the useful information from the overall data. In QLDP, we implement RR by adding a probabilistic quantum noise gate (X, Y, or Z gate): with probability $1-p$, no noise is added after the quantum state, while with probability p, a specific noise is added.

3) Quantum Frequency Estimation: As an example, we assume both the legitimate user and the attacker use Z-basis measurements in their respective circuits; the legitimate user adds an X noise gate to perturb outputs. The legitimate user output without noise gates is defined as "the true output", which represents the unperturbed results. In contrast, the legitimate user output after the addition of noise gates is referred to as "the final output", which represents the privacy-protected results. The legitimate user can only obtain the final output, and now they attempt to get the true output distribution after N shots. Each shot requires considering the question "Will the X noise gate be added?", answering "yes" or "no". To protect the distribution of the true output, the legitimate user decides whether to add noise or not by a uniformly distributed random number $[0, 1]$. If the random number is smaller than p, no noise is added; if the random number is larger than p, the noise is added. In this way, the final probabilities of "1" and "0" can be calculated as

$$\begin{aligned} \Pr\left[\text{final} = \text{“1”}\right] &= hp + (1-h)(1-p), \\ \Pr\left[\text{final} = \text{“0”}\right] &= (1-h)p + h(1-p), \end{aligned} \tag{6}$$

where h is the probability that the true output is "1".

In addition, $\Pr\left[\text{final} = \text{“1”}\right] = N_1/N$, where N_1 is the number of final results that get "1". Combined with Eq. (6), we can get an estimate of the true output as "1" is

$$\hat{h} = \frac{N_1/N - p}{2(1-p) - 1}. \tag{7}$$

Legitimate users will now be able to use Eq. (7) to estimate their true output.

4) Protection Level ϵ of QLDP: Now we can obtain the following new QLDP theorem under RR.

*Theorem 1 (***QLDP under Random Response):** For all quantum states $\mathcal{E}(\rho)$ and $\mathcal{E}(\sigma)$ before a random noise mechanism $\mathcal{K}$, this random response in the D-dimension Hilbert space provides ϵ-quantum local differential private, where ϵ is

$$\epsilon = \ln \frac{p}{1-p}. \tag{8}$$

Proof 1: We can find that in Eq. (6), $\Pr\left[\text{final} = \text{“1”}\right]$ varies from p to $1-p$, $\Pr\left[\text{final} = \text{“0”}\right]$ also varies from p to $1-p$. Hence, the ratio of probabilities for different results of one shot can be at most $\frac{p}{1-p}$. According to Definition 4, under the random response, we can obtain that $\epsilon = \ln \frac{p}{1-p} \geq \frac{\Pr[\mathcal{K}(\mathcal{E}(\rho))\in y]}{\Pr[\mathcal{K}(\mathcal{E}(\sigma))\in y]}$.

B. The Variants of QLDP

The (ϵ, δ)-LDP is a relaxed version of ϵ-LDP, where the mechanism satisfies ϵ-LDP with probability at least $1-\delta$. Similarly, we introduce (ϵ, δ)-QLDP as a relaxed version of QLDP, where the mechanism satisfies ϵ-QLDP with probability at least $1-\delta$. If δ is 0, (ϵ, δ)-QLDP becomes ϵ-QLDP. The following is the definition of (ϵ, δ)-QLDP.

Definition 5 ((ϵ, δ)**-Quantum Local Differential Privacy):** For any quantum states ρ and σ, a quantum operation $\mathcal{E}$ satisfies (ϵ, δ)-QLDP if and only if every POVM $M = \{M_m\}$ and every subset Y of outcomes satisfy,

$$\Pr\left[M\left(\mathcal{E}\left(\rho\right)\right) \in Y\right] \leq e^{\epsilon} \cdot \Pr\left[M\left(\mathcal{E}\left(\sigma\right)\right) \in Y\right] + \delta, \tag{9}$$

where $Pr\left[\cdot\right]$ is the probability, ϵ is the privacy budget, $\mathcal{E}$ is an arbitrary quantum operation, δ is the failure probability for privacy breaches.

C. The Variance of The True Output

We now infer the variance of the true output, which evaluates the accuracy of the estimated true output by a legitimate user over multiple shots. We first redefine Definition 4 in a visual terms.

Definition 6 (ϵ**-Quantum Local Differential Privacy Protocol):** Consider two probabilities, p and q. A local protocol given by the mechanism $\mathcal{M}$ (binary randomized response), in which the legitimate user does not add noise gates to a quantum circuit with probability p, and adds noise with

probability q, satisfies ϵ-QLDP if and only if $p \leq q \cdot e^{\epsilon}$. p and q satisfy $p + q = 1$.

Based on Eq. (13) in [16], the variance for the value v (i.e. $\hat{f}_v$) with random noise among N users will be $\mathrm{Var}[\hat{f}_v] = \frac{q(1-q)}{N(p-q)^2} + \frac{\hat{f}_v(1-p-q)}{N(p-q)}$. Similarly, we can introduce the variance of the number of true output "1" (i.e. $\hat{h}_{\mathcal{E}(v)}$) among N shots will be

$$\mathrm{Var}[\hat{h}_{\mathcal{E}(v)}] = \frac{q(1-q)}{N(p-q)^2} + \frac{\hat{h}_{\mathcal{E}(v)}(1-p-q)}{N(p-q)} \approx \frac{q(1-q)}{N(p-q)^2}. \tag{10}$$

When p is 1 or 0, the estimated output $\hat{h}_{\mathcal{E}(v)}$ becomes more accurate; as p gets closer to 0.5, this estimate output $\hat{h}_{\mathcal{E}(v)}$ becomes increasingly random. As the number of shots N increases, the estimated output $\hat{h}_{\mathcal{E}(v)}$ becomes more accurate.

VI. Experimental Evaluation

Our experiment is divided into three parts to show the effectiveness of our approach. First, we compare our paper with existing works. Second, we verify that simple noise can perturb the output, but not all types of noise are effective under different measurements. Third, we demonstrate that quantum frequency estimation enables legitimate users to obtain reliable true outputs while preserving privacy.

A. Comparison with Existing Works

We summarize the key differences between our work and existing approaches in Table II. All three characteristics in the table are from the perspective of legitimate users, where the "irreversible" means that even if the attacker knows the noise type the legitimate user adds, they still cannot infer the true output. "Memory-independent" means that legitimate users do not need to record where the noise occurs in order to apply inverse operations to recover the true output. "Estimable" refers to the ability of legitimate users to estimate the distribution of the true output even after the noise is added. As shown in Table II, only our method satisfies all three characteristics.

In addition, we compare the privacy budgets of our method with those of methods [19] and [20], that also add X noise gate for output protection, as shown in Fig. 1. Since [19] and [20] apply an X gate with 100%, they can only achieve a fixed privacy budget, represented by the asterisk in Fig. 1. In contrast, our approach adds a probabilistic X noise gate, enabling us to obtain different privacy budgets, represented by the blue line, to satisfy different privacy protection requirements. This also supports the conclusion in Section V.C, where we demonstrate that for X (or Y, Z) noise gate, setting $p = 0.5$ achieves the largest perturbation with the most random output and the smallest privacy budget.

B. The Noise Mechanisms to Realize QLDP

We utilize the IBM real quantum computer "ibm_kyiv" to validate Table I, i.e., whether different types of noise can perturb the output under different quantum measurement operators. The number of shots is set to 1000, the noise gates (i.e., 'Noise', Noise1', 'Noise2' in Fig. 2) are either X or Y, or Z gates. In noisy circuits, the probability of noise occurrence is 100%. Measurements are either X or Y, or Z-basis. The corresponding quantum circuit model is shown in Fig. 2.

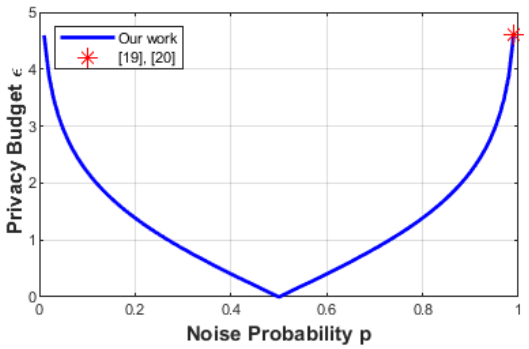

Fig. 1. Comparison of privacy budgets.

For a single-qubit circuit, as shown in Fig. 2(a), we assume that the quantum circuit part consists of only random single qubit gates $U\left(\frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{6}\right)$. The first qubit represents the quantum circuit without additional noise (i.e., the original circuit), and the second qubit represents the quantum circuit with an additional noise gate. The result is shown in Table III. Taking Z-basis measurement as an example, we observe that adding X or Y noise gate can effectively perturb the original output distribution, but adding a Z noise gate results in an output distribution the same as the original, meaning it fails to provide protection. This result is consistent with Table I, so we show that not all types of noise can perturb the output under different measurements in single qubit circuits. Moreover, since the final output of legitimate users is perturbed (i.e., the true output inversion), the crosstalk attack described in Section III becomes ineffective by inferring the true output with only 4% accuracy.

For multi-qubit circuits, we utilize the Bernstein-Vazirani algorithm to be the circuit part. The Bernstein-Vazirani algorithm works by allowing the user to query the algorithm with different inputs to obtain corresponding outputs, ultimately revealing the hidden number encoded in the system [42]. In this experiment, we assume that the hidden number of the system is "11" and the corresponding circuit is shown in Fig. 2(b). We aim to show that adding a specific noise can modify the output and thus protect the hidden number, which is of practical importance. The result is shown in Table IV. Taking Z-basis measurement as an example, we observe that adding an X or Y noise gate on both qubits can effectively perturb the original output distribution, but adding a Z noise gate results in an output distribution the same as the original, meaning it fails to provide protection and has the chance of revealing the hidden number. This result is consistent with Table I, so we show that not all types of noise can perturb the output under different measurements in multi-qubit circuits. Moreover, since the final output of legitimate users is perturbed, the crosstalk attack described in Section III becomes ineffective by inferring the true output with only 4% accuracy.

	Irreversible		Memory-independent		Estimable	
	Noise type added	Support	Register usage	Support	Estimation Method	Support
Saki et al. [19]	100% X gate	✗	Register	✗	Inverse operation	✓
Maurya et al. [20]	100% X gate	✗	Register	✗	Inverse operation	✓
Guan [25]	Depolarizing noise	✓	Register	✗	Not mention	✗
Angrisani et al. [26]	Measurements	✗	Register	✗	Not mention	✗
Nuradha et al. [28]	Depolarizing noise	✓	Register	✗	Not mention	✗
Our Work	Probabilistic X/Y/Z gate	✓	Register-free	✓	Frequency estimation	✓

✓ means that this work has this characteristic; ✗ means that this work does not have this characteristic.

TABLE II
COMPARISON WITH EXISTING WORKS.

	Z-basis measurement			X-basis measurement			Y-basis measurement		
	P(0)	P(1)	Protect data	P(+)	P(-)	Protect data	P(+i)	P(-i)	Protect data
No noise	0.73	0.27		0.82	0.18		0.81	0.19	
Add X gate	0.25	0.75	✓	0.80	0.20	✗	0.22	0.78	✓
Add Y gate	0.25	0.75	✓	0.19	0.81	✓	0.79	0.21	✗
Add Z gate	0.76	0.24	✗	0.18	0.82	✓	0.20	0.80	✓

✓ means that the output can be protected; ✗ means that the output cannot be protected.

TABLE III
SINGLE QUBIT CIRCUIT RESULTS UNDER DIFFERENT GATES PERTURBATIONS.

	Z-basis measurement				
	P(00)	P(01)	P(10)	P(11)	Protect data
No noise	0.00	0.02	0.02	0.96	
Add X gate	1.00	0.00	0.00	0.00	✓
Add Y gate	0.94	0.02	0.04	0.00	✓
Add Z gate	0.00	0.02	0.02	0.96	✗

✓ means that the output can be protected; ✗ means that the output cannot be protected.

TABLE IV
MULTI-QUBITS CIRCUIT RESULTS UNDER DIFFERENT GATES PERTURBATIONS.

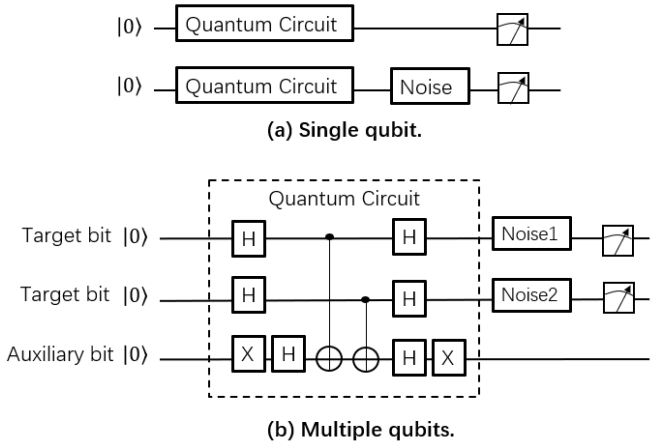

Fig. 2. Quantum circuit models to validate noise mechanisms. (a) The two circuits illustrate the single qubit cases without and with noise, while the rest is the same. 'Quantum circuit' denotes single qubit gates, and 'Noise' denotes an X, Y or Z noise gate. (b) The circuit illustrates the multiple qubits case with noise. 'Noise1' and 'Noise2' both denote an X noise gate.

C. The Quantum Frequency Estimation

We utilize the IBM quantum simulator "qasm_simulator" to validate the frequency estimated result (i.e., 'Result' in Fig 3) is similar to the true output. As an example, we choose X-basis measurements for this simulation. The number of shots is set to 1000 and the noise gates is X-gate noise with occurrence probability 0.2.

The corresponding quantum circuit model is shown in Fig. 3. The simulation was conducted on 50 randomly generated circuits (i.e., 'Quantum Circuit' in Fig 3), where the circuit consists of 5-10 gates taken randomly from X, H, S, R_y gates. The first qubit represents the quantum circuit with additional noise; the distribution after the measurement symbol represents the final output; 'Formula' refers to the frequency estimation Eq. (7); 'Result' represents the estimated result. The second qubit represents the quantum circuit without additional noise (i.e., the original circuit); the distribution after the measurement symbol represents the true output. To assess the similarity between the true output and the estimated output, we used Kullback-Leibler (KL) divergence, where a lower KL divergence indicates a higher similarity between the two distributions. The results, shown in Fig. 4, indicate that 94% of the circuits have a KL divergence below 0.01, suggesting that the estimated and true outputs are highly similar. This confirms the accuracy of our frequency estimation method. In addition, we evaluated the utility of our method by measuring the probability that the estimated and true outputs match (i.e., both output 1 or 0). Across the same 50 circuits, our approach achieved a utility score of 90%, which shows that our method maintains usability while protecting data privacy.

VII. DISCUSSION

We consider exploring the frequency estimation of multi-qubit circuits in the future. For multi-qubit circuits, we typically obtain the joint probability distribution over multiple qubits rather than the probability of individual qubits. If the qubits are uncorrelated, the joint probability distribution can

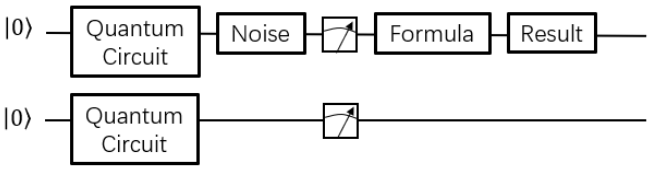

Fig. 3. Quantum circuit models to validate frequency estimation. The two circuits illustrate the cases without and with noise, while the rest are the same. 'Quantum circuit' denotes single qubit gates, 'Noise' denotes an X, Y, or Z noise gate, 'Formula' denotes Eq. (7), 'Result' denotes the estimated results.

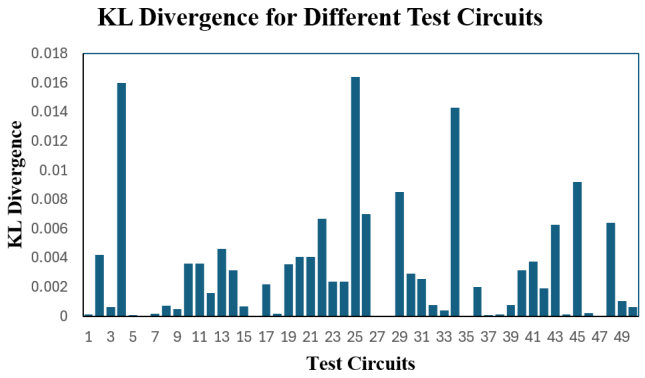

Fig. 4. KL Divergence for different test circuits.

be factorized using the law of total probability, allowing us to determine the individual probability distributions of each qubit. In this way, we can perform the frequency estimation for each qubit, as Eq. (7) derived earlier in our paper, to obtain the true distribution of each qubit and the corresponding joint probability. However, in most practical scenarios, multi-qubit circuits involve entanglement and correlation, such as CNOT gates, CCX gates, etc. In such cases, the joint probability would not be able to obtain the respective distribution of each qubit. Nevertheless, through our preliminary derivation, we have found that the relationship between the true and estimated distributions of multiple qubits quantum circuits can be expressed as a system of 2^n equations, and the exact relationship can still be obtained. The specific details will be explored in the future.

VIII. Conclusion

In this paper, we develop a novel protection method for the quantum computing output based on quantum local differential privacy. Through rigorous mathematical derivation, we demonstrate that under different quantum measurement operators, simple quantum noise can perturb the output and thereby protect the data. However, not all types of noise are effective for this purpose. To further prevent an advanced attacker who has the user's partial information, we change the probability of the noise gate added to the quantum circuit. Additionally, we propose quantum frequency estimation enabling legitimate users to estimate the true output from the perturbed results. We conduct extensive experiments on both real quantum computers and quantum simulators to validate our theoretical findings. Our results also show that the quantum frequency achieves 94% accuracy and 90% utility to estimate the true output, demonstrating the effectiveness of our approach.

References

[1] K. Ju, H. Zhong, X. Zhang, X. Qin, and M. Pan, "Quantum computing catalyzes future quantum networks: Efficiency and inherent privacy," *IEEE Network*, vol. 39, no. 1, pp. 124–131, May 2024.

[2] A. Steane, "Quantum computing," *Reports on Progress in Physics*, vol. 61, no. 2, p. 117, 1998.

[3] J. L. O'brien, "Optical quantum computing," *Science*, vol. 318, no. 5856, pp. 1567–1570, December 2007.

[4] C. Portmann and R. Renner, "Security in quantum cryptography," *Reviews of Modern Physics*, vol. 94, no. 2, p. 025008, June 2022.

[5] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, "Quantum algorithms for quantum chemistry and quantum materials science," *Chemical reviews*, vol. 120, no. 22, pp. 12 685–12 717, October 2020.

[6] S. Wilkens and J. Moorhouse, "Quantum computing for financial risk measurement," *Quantum Information Processing*, vol. 22, no. 1, p. 51, January 2023.

[7] A. Ash-Saki, M. Alam, and S. Ghosh, "Analysis of crosstalk in nisq devices and security implications in multi-programming regime," in *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, Boston, Massachusetts, August 2020, pp. 25–30.

[8] A. Mi, S. Deng, and J. Szefer, "Securing reset operations in nisq quantum computers," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, Los Angeles, CA, USA, November 2022.

[9] F. Chen, L. Jiang, H. Müller, P. Richerme, C. Chu, Z. Fu, and M. Yang, "Nisq quantum computing: A security-centric tutorial and survey [feature]," *IEEE Circuits and Systems Magazine*, vol. 24, no. 1, pp. 14–32, March 2024.

[10] C. Dwork, "Differential privacy," in *International colloquium on automata, languages, and programming (ICALP)*, Berlin, Heidelberg, July 2006.

[11] ——, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation (TAMC)*, Xi'an, China, April 2008.

[12] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *ACM SIGSAC conference on computer and communications security (CCS)*, Vienna, Austria, October 2016.

[13] A. Friedman and A. Schuster, "Data mining with differential privacy," in *ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, Washington, DC, USA, July 2010.

[14] J. Liu, C. Zhang, and Y. Fang, "Epic: A differential privacy framework to defend smart homes against internet traffic analysis," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1206–1217, February 2018.

[15] F. K. Dankar and K. El Emam, "Practicing differential privacy in health care: A review." *Trans. Data Priv.*, vol. 6, no. 1, pp. 35–67, 2013.

[16] T. Wang, X. Zhang, J. Feng, and X. Yang, "A comprehensive survey on local differential privacy toward data statistics and analysis," *Sensors*, vol. 20, no. 24, p. 7030, October 2020.

[17] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827–5842, November 2019.

[18] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang, "Privacy at scale: Local differential privacy in practice," in *International conference on management of data (SIGMOD)*, Houston, TX, USA, May 2018.

[19] A. A. Saki and S. Ghosh, "Qubit sensing: A new attack model for multi-programming quantum computing," *arXiv preprint arXiv:2104.05899*, April 2021.

[20] S. Maurya, C. N. Mude, B. Lienhard, and S. Tannu, "Understanding side-channel vulnerabilities in superconducting qubit readout architectures," in *IEEE International Conference on Quantum Computing and Engineering (QCE)*, Montreal, QC, Canada, September 2024.

[21] K. Ju, X. Qin, H. Zhong, X. Zhang, M. Pan, and B. Liu, "Privacy preserving quantum search mechanism using grover's algorithm," in *International Conference on Quantum Communications, Networking, and Computing (QCNC)*, Kanazawa, Japan, July 2024, pp. 204–210.

[22] L. Zhou and M. Ying, "Differential privacy in quantum computation," in *IEEE Computer Security Foundations Symposium (CSF)*, Santa Barbara, CA, USA, August 2017, pp. 249–262.
[23] K. Ju, X. Qin, H. Zhong, X. Zhang, M. Pan, and B. Liu, "Harnessing inherent noises for privacy preservation in quantum machine learning," in *IEEE International Conference on Communications (ICC)*, Denver, CO, USA, June 2024.
[24] Y. Li, Y. Zhao, X. Zhang, H. Zhong, M. Pan, and C. Zhang, "Differential privacy preserving quantum computing via projection operator measurements," in *International Conference on Quantum Communications, Networking, and Computing (QCNC)*, Kanazawa, Japan, July 2024.
[25] J. Guan, "Optimal mechanisms for quantum local differential privacy," *arXiv preprint arXiv:2407.13516*, November 2024.
[26] A. Angrisani and E. Kashefi, "Quantum local differential privacy and quantum statistical query model," *arXiv preprint arXiv:2203.03591*, August 2022.
[27] C. Hirche, C. Rouzé, and D. S. França, "Quantum differential privacy: An information theory perspective," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 5771–5787, May 2023.
[28] T. Nuradha and M. M. Wilde, "Contraction of private quantum channels and private quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 71, no. 3, pp. 1851–1873, March 2025.
[29] H. Zhong, K. Ju, M. Sistla, X. Zhang, A. Li, X. Qin, X. Fu, and M. Pan, "Tuning quantum computing privacy through quantum error correction," in *IEEE Global Communications Conference (GLOBECOM)*, Cape Town, South Africa, December 2024, pp. 3986–3991.
[30] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, "Quantum computers," *nature*, vol. 464, no. 7285, pp. 45–53, March 2010.
[31] K. Valiev, "Quantum computers and quantum computations," *Physics-Uspekhi*, vol. 48, no. 1, p. 1, 2005.
[32] Y. Zhao, H. Zhong, X. Zhang, C. Zhang, and M. Pan, "Bridging quantum computing and differential privacy: a survey on quantum computing privacy," *arXiv e-prints*, pp. arXiv–2403, March 2024.
[33] R. A. Somaraju, A. Sarlette, and H. Thienpont, "Quantum filtering using povm measurements," in *IEEE Conference on Decision and Control (CDC)*, Firenze, Italy, December 2013.
[34] W. J. Yun, H. Baek, and J. Kim, "Projection valued measure-based quantum machine learning for multi-class classification," *arXiv preprint arXiv:2210.16731*, November 2022.
[35] H. E. Brandt, "Positive operator valued measure in quantum information processing," *American Journal of Physics*, vol. 67, no. 5, pp. 434–439, May 1999.
[36] Y.-K. Liu, "Universal low-rank matrix recovery from pauli measurements," *Advances in Neural Information Processing Systems*, vol. 24, 2011.
[37] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, January 2010.
[38] J. Guan, W. Fang, M. Huang, and M. Ying, "Detecting violations of differential privacy for quantum algorithms," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, Copenhagen, Denmark, November 2023.
[39] B. Harper, B. Tonekaboni, B. Goldozian, M. Sevior, and M. Usman, "Crosstalk attacks and defence in a shared quantum computing environment," *arXiv preprint arXiv:2402.02753*, February 2024.
[40] S. Bajpayee and I. Mukherjee, "Analysis of the effects of crosstalk errors on various quantum circuits," in *International Conference on VLSI Design and International Conference on Embedded Systems (VLSID)*, Kolkata, India, January 2024, pp. 408–413.
[41] S. T. Flammia and Y.-K. Liu, "Direct fidelity estimation from few pauli measurements," *Physical review letters*, vol. 106, no. 23, p. 230501, June 2011.
[42] Y.-C. Liu and M.-F. Liu, "Implementation of grover's algorithm & bernstein-vazirani algorithm with ibm qiskit," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 76–95, February 2024.