

QubitHammer Attacks: Qubit Flipping Attacks in Multi-tenant Superconducting Quantum Computers

Yizhuo Tan

*Yale University
New Haven, USA*

Navnil Choudhury

*University of Texas at Dallas
Richardson, USA*

Kanad Basu

*University of Texas at Dallas
Richardson, USA*

Jakub Szefer

*Northwestern University
Evanston, USA**yizhuo.tan@yale.edu**Navnil.Choudhury@utdallas.edu**Kanad.Basu@utdallas.edu**jakub.szefer@northwestern.edu*

Abstract—Quantum computing is rapidly evolving its capabilities, with a corresponding surge in its deployment within cloud-based environments. Various quantum computers are accessible today via pay-as-you-go cloud computing models, offering unprecedented convenience. Due to its rapidly growing demand, quantum computers are shifting from a single-tenant to a multi-tenant model to enhance resource utilization. However, this widespread accessibility to shared multi-tenant systems also introduces potential security vulnerabilities. In this work, we present for the first time a set of novel attacks, named together as the QubitHammer attacks, which target state-of-the-art superconducting quantum computers. We show that in a multi-tenant cloud-based quantum system, an adversary with the basic capability to deploy custom pulses, similar to any standard user today, can utilize the QubitHammer attacks to significantly degrade the fidelity of victim circuits located on the same quantum computer. Upon extensive evaluation, the QubitHammer attacks achieve a very high variational distance of up to 0.938 from the expected outcome, thus demonstrating their potential to degrade victim computation. Our findings exhibit the effectiveness of these attacks across various superconducting quantum computers from a leading vendor, suggesting that QubitHammer represents a new class of security attacks. Further, the attacks are demonstrated to bypass all existing defenses proposed so far for ensuring the reliability in multi-tenant superconducting quantum computers.

Index Terms—quantum computing, security, attacks, crosstalk, interference

1. Introduction

In recent years, there has been a significant surge in the development and deployment of Noisy Intermediate-Scale Quantum (NISQ) computers, typically in the form of quantum processing units (QPUs) [21]. These QPUs are constructed using qubits, which serve as the fundamental units of quantum information, interconnected by quantum gates and couplers that enable entanglement and coherent quantum operations across the qubit array. Popular technologies for building QPUs include superconducting qubits [11], trapped ions [24], neutral atoms [25], silicon spin qubits [18], photons [35], and diamond NV centers [19]. Quantum systems are poised to revolutionize

computation, achieving complex tasks that would take classical computers exponentially longer in comparison.

For example, IBM unveiled a 1,121-qubit quantum computer in 2023, showcasing the rapid progress in quantum hardware, and by 2029, they are projected to release quantum computers capable of running 100 million gates [11]. As the qubit count in quantum computers continues to grow, so does the potential to tackle increasingly complex problems, paving the way for breakthroughs that were once thought impossible with traditional computing. Significant progress in the development of error-corrected quantum computers has been demonstrated in recent years, bringing us to the threshold of deploying these advanced systems [28] [4].

Access to these quantum computers can be obtained either by purchasing an individual system or, more commonly, through cloud-based platforms. Cloud services like Microsoft Azure, qBraid, Amazon Braket, IBM Quantum, and others enable users to run jobs on larger quantum processors [8], [20], [22]. With increased accessibility, the demand for quantum computers accessible through the cloud has increased significantly. However, existing quantum computers operate in a single-tenant mode, where only one job can run at a time on the QPU, leading to low throughput and the underutilization of valuable quantum resources.

To address this inefficiency and improve the utilization of valuable quantum resources, researchers have explored methods for enabling multi-tenancy in quantum computers, where multiple quantum circuits can be executed simultaneously on different parts of the quantum machine [5] [14]. By partitioning the quantum computer into separate segments that execute quantum circuits in parallel on non-overlapping sets of qubits, they aim to maximize the utilization of these cloud-based quantum systems. Currently, three promising approaches exist for robust multi-tenancy [5]. The Fair and Reliable Partitioning (FRP) algorithm allocates reliable qubits to each program based on machine calibration error data. To address the impact of qubit measurement operations on other simultaneously executed operations of varying lengths [5], the Delayed Instruction Scheduling (DIS) policy reschedules the start times of these programs [5]. Lastly, when multi-tenancy significantly affects program reliability, quantum hardware providers can switch to isolated execution mode using an Adaptive Multi-Programming (AMP) design [5].

While the proposed multi-tenancy enhances the utilization of quantum computers significantly, it also intro-

This work was supported in part by NSF grants 2223046 and 2312754.

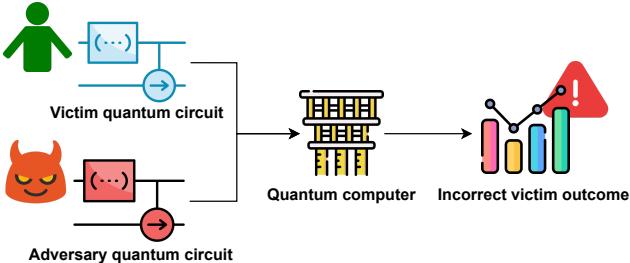


Figure 1: Illustration of victim and attacker sharing a quantum computer, resulting in victim’s computation being incorrect.

duces new challenges when multiple quantum programs are executed concurrently. Maintaining low error rates among qubits is a critical priority for quantum hardware providers, as it directly influences the performance and reliability of their technology. However, the simultaneous execution of multiple jobs on the same quantum hardware can cause complex quantum mechanical interactions, or crosstalk, between qubits allocated to two distinct circuits, resulting in degraded performance.

Quantum computing finds applications in various critical fields of research, including drug discovery, molecular chemistry, and financial modeling. These applications depend on the security and reliability of underlying quantum operations, and disruptions in the same by malicious users can result in significant damages. In a multi-tenant environment, a potentially malicious user can exploit such undesired interactions to cause significant disruptions in the quantum operations of an unsuspecting user. For example, consider the scenario depicted in Figure 1, where an adversary and victim are using the same quantum computer, where the victim circuit output is incorrect and unreliable due to the adversarial attack. To exemplify the threat posed by such an attack, consider a user encoding a portfolio optimization into a quantum circuit for execution. A successful attack by a malicious user compromises the reliability of underlying quantum operations, causing inaccurate risk evaluations or flawed investment strategies based on incorrect output. This would ultimately result in significant financial losses for institutions relying on quantum computing for high-stakes decision-making.

In this paper, we present the first study on what we call QubitHammer attacks aimed at disrupting quantum operations in a multi-tenant environment, assuming a realistic threat model where the victim and attacker are not necessarily nearby in the QPU. We are first to show qubit flipping and disturbance of victim circuits when the attacker circuit and victim circuit are far apart on the same QPU. In our setting, the adversary possesses the same level of access as any ordinary user. Our findings reveal that the adversary can cause significant disruption to the victim’s quantum circuit execution by deploying specially designed attack pulses on their own qubits, which cause far-away victim qubits to be disturbed. We explore and experimentally validate multiple attack scenarios, all of which demonstrate high attack efficacy, underscoring the severity of the threat posed by our proposed attack model.

1.1. Contributions

Our contributions can be summarized as:

- This work designs a set of novel QubitHammer attacks that do not require elevated privileges and only use widely available pulse level control.
- It proposes various attack strategies on real-life superconducting quantum computers to determine the most vulnerable qubits and best attack pulse configurations.
- We identify reproducible, but previously undocumented, high sensitivity to disturbance of physical qubit 0 on range of Eagle r3 quantum computers from IBM.
- We demonstrate the impact of the QubitHammer attacks on real-world quantum algorithms such as Grover’s and QAOA, running on different Eagle r3 machines, showing the ability to change computation outcomes even when the attacker and the victim are not adjacent to each other. Upon extensive evaluation, our attack furnishes a variation distance of up to 0.938.
- We demonstrate that countermeasures like dynamical decoupling are ineffective against our attack, which furnishes a variational distance of up to 0.837, despite these mitigation techniques.

2. Background

This section presents background on quantum computation, as well as on NISQ quantum computers and today’s cloud-based deployments of quantum computers.

2.1. Quantum Computing Principles

Unlike the classical bit, which can be either 0 or 1, a quantum bit (qubit for short) can be a linear combination of two basis states $|0\rangle$ and $|1\rangle$. A qubit state $|\psi\rangle$ can be represented as:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (1)$$

where α and β are complex numbers which satisfy $|\alpha|^2 + |\beta|^2 = 1$. More generally, n -qubit state $|\phi\rangle$ can be expressed as follow:

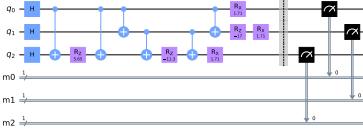
$$|\phi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle \quad (2)$$

where $|i\rangle$ is one of 2^n basis states from $|0\dots0\rangle$ to $|1\dots1\rangle$, α_i satisfies $\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1$.

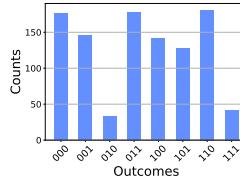
This superposition principle allows quantum computers to explore many states simultaneously and provides massive potential computational power. The n -qubit quantum states can also be represented using n -dimensional vectors. For example, $|00\dots0\rangle = [1, 0, \dots, 0]^T$ and $|1, \dots, 1, 1\rangle = [0, \dots, 0, 1]^T$. As a result, the unitary quantum gate U , which satisfies $UU^T = U^TU = I$, operating qubits like $U|\phi\rangle$ can be represented by $2^n \times 2^n$ matrix. Here are some examples of frequently used gates:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, SX = \frac{1}{2} \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}, CX = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Most quantum gates, such as the Hadamard gate, need to be decomposed into some basis gates before submitting to the real quantum computer hardware. For the IBM



(a) An example quantum circuit with three qubits.



(b) Example output obtained from the circuit.

Figure 2: Example of quantum circuit and it’s output probabilities: a 3 qubit circuit will have $2^3 = 8$ output states each with some probability. The classical “result” of the computation is determined from the output probabilities, different algorithms output is interpreted differently.

quantum machines we use for experiments, the basis gates are ECR, ID, RZ, SX, and X.

2.2. Quantum Circuit

A quantum circuit is composed of a series of quantum gates, mentioned in Section 2.1, which are applied to a set of qubits and the measured. After building a logic-level quantum circuit using a quantum development kit such as Qiskit [13], a series of operations needs to be done to transform such logic-level circuits to hardware-specific instructions. An example quantum circuit is shown in Figure 2a, where quantum gates are applied to three qubits, following which they are measured. The input quantum circuit needs to undergo transpilation to be translated to the basis gates and to better fit the topology of the wanted quantum device. As microwave pulses are typically used to control superconducting qubits, a sequence of control pulses corresponding to each of the basis gates of the transpiled quantum circuit will be generated through the scheduling stage [13], which transforms gate-level circuits to pulse-level circuits before being sent to quantum hardware. Finally, the qubits are measured after all the gates have been executed. An example of measurement probabilities is shown in Figure 2b. In general n qubit circuit will have 2^n states. In NISQ computers, the circuit is executed multiple times, so that output probabilities for each state can be measured. Each time the circuit is executed it is called a *shot*, thus execution of a circuit constitutes many shots. The “output” of the quantum computation has to be determined from the state probabilities. If the probabilities are changed, then the output is changed as well – thus a basis for a security attack is to change the output probabilities.

2.3. NISQ Computers

Noisy Intermediate Scale Quantum (*NISQ*) refers to the current era of quantum computing characterized by devices that contain a moderate number of qubits (typically between 50 and a few hundred) and are prone to noise and errors. The term “*NISQ*” is introduced to describe the state of quantum computing where hardware is powerful enough to perform certain tasks that classical computers struggle with, but is still limited by the challenges of single-qubit and two-qubit errors, readout errors, decoherence, and crosstalk. Figure 3 shows the topology of a NISQ device from IBM [11], which is their

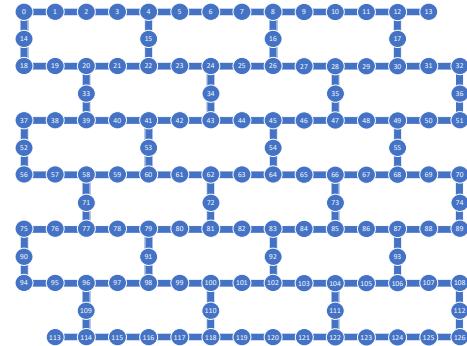


Figure 3: IBMQ Eagle r3 topology for 127 qubits. Circles represent qubits, thick lines represent fixed couplings between the qubits.

127-qubit Eagle r3 processor based on the transmon superconducting qubit architecture. In this heavy-hexagonal qubit layout [11], qubits are connected to either two or three neighboring qubits, resembling the arrangement of edges and corners in tessellated hexagons. In a multi-tenant setting, different qubits can be assigned to different users or circuits. At the logical level, operations on disjoint sets of qubits and couplings should not have an effect on each other, but, as we demonstrate, they do, which leads to the QubitHammer attacks presented in this work.

2.4. Qiskit Pulse

IBMs superconducting quantum computers employ microwave pulses typically characterized by the envelope, frequency, and phase to control qubits [13]. The envelope determines the shape of the signal generated by an arbitrary waveform generator. The frequency and phase define a periodic signal that modulates the envelope. Together, these two signals create the output signal that is sent to the qubit. As each superconducting qubit is different in their frequency, these pulse parameters of the same gate also vary between qubits. The pulses for IBM Quantum’s native gates are all predefined and well-calibrated, while custom-defined pulses are also allowed.

Qiskit Pulse [13] provides such a low-level interface for designing and executing custom quantum gate operations as a sequence of pulses. One efficient way is to use parametrized pulses, which are described by predefined shapes, requiring only a few parameters to be stored. These parameters include the duration, which indicates the pulse length, the amplitude, which indicates the relative pulse length, and other parameters that define the pulse’s shape.

There is no limitation that prevents pulses applied to one qubit from having a configuration that is correct for a different qubit. This is actually the basis of our attack, implying that qubit q_{attacker} can be actuated with pulses that have parameters of q_{victim} , and experimentally we observe that the victim qubit is affected during the time that the attacker qubit(s) are driven with the pulses.

2.5. Quantum-as-a-Service

Similar to Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS),

Quantum-as-a-Service (QaaS) [12] refers to cloud-based delivery of quantum technologies, quantum computing services and quantum computing solutions, without needing to own or maintain the hardware. QaaS is accelerating the adoption of quantum computing by making it more accessible to researchers, developers, and businesses that want to explore the potential of quantum technology without the significant investment in quantum hardware. The growing demand for QaaS attracts many companies to race for their quantum services. Superconducting quantum modalities like IBM [11] and Rigetti [27], ion trap systems like Quantinuum [24], and neutral atom machines like QuEra [25] are all available through cloud-based platforms now.

2.6. Multi-tenant Quantum Computers

Even though a plethora of companies are providing online quantum platforms to access their single-tenant quantum computers nowadays, the growing number of quantum computer users has caused a demand-supply imbalance, resulting in substantial wait times for accessing quantum computing resources [26]. Hence, a multi-tenant quantum computing system could be preferred in terms of cost efficiency, resource utilization, and accessibility. Through mapping multiple quantum programs onto a single quantum hardware simultaneously, multi-tenant Quantum computers can be shared by multiple users at the same time [5] [14]. This is crucial for making quantum computing more accessible and efficient, especially given the high cost of quantum hardware and long waiting time for queues. However, multi-tenant environments may affect the reliability of quantum programs for multiple reasons. The limited availability of high-fidelity qubits in NISQ devices restricts the equal distribution of quantum resources. Crosstalk from simultaneous gate operations can degrade the performance of involved qubits and can introduce security vulnerability in a multi-programming environment as well [1] [30].

3. Threat Model

Here we present the threat model. We detail the assumptions we make about the adversary, and present the attack objectives that we consider.

3.1. Assumptions about Adversary’s Access

We consider a scenario in which both the victim and the adversary execute their circuits on the same quantum computer, albeit on separate, non-overlapping sets of qubits. This is shown in Figure 4, where, e.g., three users (two victims and one adversary) are concurrently executing their quantum programs on a 127-qubit quantum computer.

The adversary is assumed to have the same level of access as any standard user, with no additional privileges. Operating within the constraints of a multi-tenant environment, the adversary is assumed to have the ability to manipulate the state of the qubits allocated to their own attacker circuit, but of course not the victim’s qubits. We assume the attacker, like victims, has access to pulse level control and can use custom pulses in their circuits.

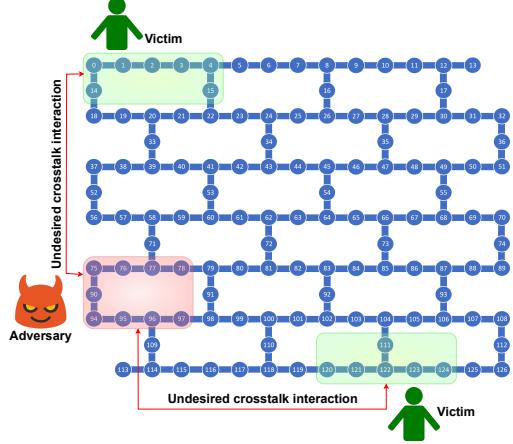


Figure 4: Overview of proposed threat model. An example of multi-tenant quantum computer with three users, where an adversary is exploiting undesired long-distance disturbance to perturb the output of the victim circuit’s output.

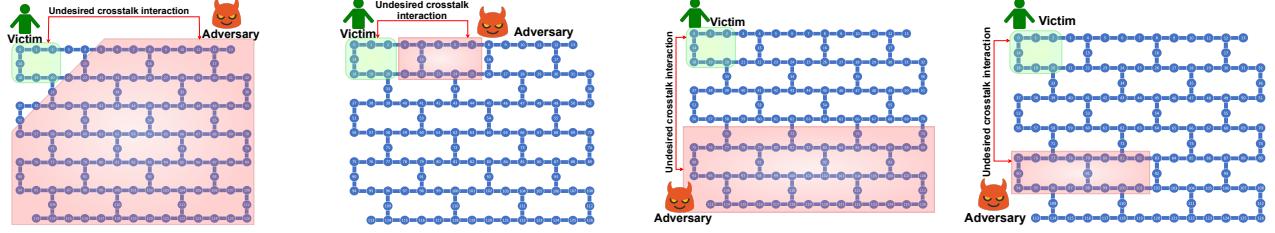
We make two key but minimal assumptions: (1) First, attacker is assumed to be able to execute their circuits using qubits in the same quantum computer as the victim circuit. This is a justified assumption in a multi-tenant quantum computer, where multiple quantum circuits are executed using qubits belonging to a single quantum processor. (2) Secondly, the attacker is assumed to be aware of the start timing or execution phases of the victim circuit. This assumption is reasonable since any quantum circuit is executed thousands of times (*shots*) on the QaaS cloud to obtain the expected output. This allows the attacker to concurrently execute their circuit along with the victim’s.

3.2. Assumptions on Adversary’s Objective

We assume that the primary objective of the adversary is to disrupt the performance and integrity of the victim quantum circuit, leading to a denial-of-service (DoS) attack or incorrect outputs being generated. Specifically, the adversary seeks to introduce errors or perturbations in the victim circuit’s operations by exploiting crosstalk and other effects between the qubits they control and those of the victim which they do not control.

4. Attack Scenarios

Due to the shared electrical coupling between different components in current NISQ-era quantum computers, pulses intended to drive one set of qubits may inadvertently influence the state of other qubits. Moreover, applying pulses to the attacker’s qubit, pulses that resonate at the frequency of a different victim target qubit can significantly impact the quantum circuits that involve the victim target qubit during execution, which is the basis for our attacks. Especially, in this work, for the first time, we explore different possible attack scenarios using single qubit pulses, based on the different locations and number of adversarial qubits relative to the victim qubits. The exploration of attacks is performed experimentally as we have no data about the actual physical layout of the superconducting chips, only their logical topology and the public data about qubit properties, such as their lifetimes,



(a) Example of our Attack Scenario 1 (b) Example of our Attack Scenario 2 (c) Example of our Attack Scenario 3 (d) Example of our Attack Scenario 4

Figure 5: Four attack scenarios introduced in this work. The figures illustrate examples, and the numbers of victim qubits (green) and attacker qubits (red) are only for illustrative purposes.

frequencies, etc. Consequently, we introduce four attack scenarios as follows:

- 1) Attack Scenario 1: *Concentrated Adversarial Impact with Many Adversarial Qubits*
- 2) Attack Scenario 2: *Proximal Qubit Interference with Few Adversarial Qubits*
- 3) Attack Scenario 3: *Long-Range Adversarial Impact with Distant Qubit Allocation*
- 4) Attack Scenario 4: *Targeted Impact with Distant and Few Adversarial Qubit*

4.1. Attack Scenario 1: Concentrated Adversarial Impact with Many Adversarial Qubits

In this attack scenario, we consider qubits allotted in a manner that the victim has relatively few qubits, and the adversary has access to a major portion of the quantum processor and can control a large number of qubits. An example of this scenario is shown in Figure 5a. Under these conditions, the pulses applied to the qubits controlled by the adversary (adversarial qubits, highlighted in red in the Figure) are shown to have a significant impact on the victim's qubits. The adversary's control over a large number of qubits enables a concentrated attack, where synchronized pulses can cause substantial interference with the victim qubits. These pulses, tuned to frequencies near the victim qubits' operational frequency, can induce unwanted transitions or phase shifts, degrading their quantum state.

Additionally, the proximity of adversarial qubits to the victim further increases the risk of crosstalk, where electromagnetic interactions can disrupt the victim's state, leading to errors such as decoherence, bit-flips, or phase-flips. This interference not only affects the immediate state of the victim qubit but can also introduce errors that persist and propagate through subsequent quantum operations. The effectiveness of this attack scenario, including its potential to consistently disrupt quantum computations, which is presented in Section 7.

4.2. Attack Scenario 2: Proximal Qubit Interference with Few Adversarial Qubits

In the second attack scenario, we consider a qubit allocation where the adversary has access to a limited number of qubits that are positioned in close proximity to the victim's qubits on the quantum hardware. An example of such a case is depicted in Figure 5b.

Although the adversary controls fewer qubits in this situation, the physical proximity of these qubits to the victim qubit on the quantum processor introduces a unique set of challenges.

The close spatial arrangement increases the likelihood of crosstalk and electromagnetic interference, as the qubits in the same segment are more susceptible to mutual interactions. Even with a reduced number of qubits, the adversary can strategically apply pulses that exploit these interactions, potentially causing significant disruptions to the victim qubit's state. This includes inducing phase shifts, bit-flip errors, or other forms of decoherence that compromise the accuracy of quantum operations.

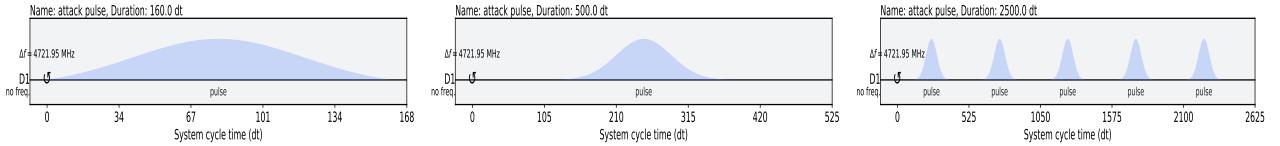
Furthermore, the adversary's ability to concentrate their attack within a limited region of the quantum processor allows for more targeted interference. This focused approach can be particularly effective in scenarios where the victim qubit is involved in critical quantum operations, as even minor disruptions could propagate through the computation, leading to amplified errors. The impact and effectiveness of this attack configuration is thoroughly examined in Section 7.

4.3. Attack Scenario 3: Long-Range Adversarial Impact with Distant Qubit Allocation

For the third attack scenario, we consider a scenario where the qubit allocation is performed in a manner that the victim qubits are allocated far away from the attack, and attacker has many qubits. An extreme example of this setup is displayed in Figure 5c.

Despite the spatial separation between the victim and adversarial qubits, the adversary's control over a large portion of the processor still poses a significant threat. The adversary can deploy pulses across their allocated qubits in a coordinated manner, potentially creating long-range interference effects. These effects, while less direct due to the physical distance, can still induce errors in the victim's qubits, particularly through mechanisms such as resonant frequency overlap, crosstalk across the processor, or even indirect interactions mediated by the quantum processor's control systems.

It is worth noting that this scenario raises important considerations about the effectiveness of attack strategies when the qubits under adversarial control are physically distant from the victim qubits in the quantum hardware. The extent to which long-distance perturbations can influence the victim's operations, and how these interactions



(a) Single attack pulse.

(b) Extended single attack pulse.

(c) Repeated attack pulses.

Figure 6: An illustration of the pulse-level description of attack pulses.

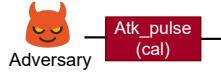


Figure 7: Gate-level representation of single attack pulse.

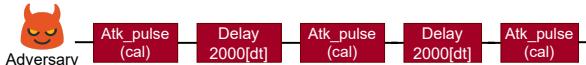


Figure 8: Gate-level representation of repeated attack pulses. Inserted delays allow for control of the density of pulses and duration of the whole sequence. Various types of delays are possible, and $2000\ dt$ is used as an example. The number of the attack pulses can also be varied from 2 to many, if only one pulse is used then we consider this single attack pulse method shown in Figure 7.

can be exploited to degrade performance, are key points of analysis that are explored in Section 7.

4.4. Attack Scenario 4: Targeted Impact with Distant and Few Adversarial Qubit

In this scenario, we consider the victim and adversary to have access to a relatively small number of qubits (10 to 20), which are physically separated by a significant distance in the quantum processor. An extreme example of this configuration is shown in Figure 5d.

This separation, combined with the limited number of qubits available to the adversary, presents a unique challenge.

Although the significant physical separation between the qubits reduces the likelihood of direct interference, the adversary can still exploit long-range interactions, such as resonant frequency overlap or indirect coupling, to affect the victim's qubits. Even with fewer qubits, the adversary can focus their efforts, using precise pulses to induce subtle but impactful disruptions, such as phase shifts, decoherence, or timing errors, that can propagate across the processor.

This scenario illustrates the possibility that even with a moderate number of qubits and significant separation, the adversary might be able to pose a threat to the victim's quantum operations. The effectiveness of this long-distance attack and its potential to disrupt the victim's computation is presented in detail in Section 7.

5. Attack Methods

In this section, we delineate our approach to designing attack pulses which are subsequently used to drive the adversarial qubit to manifest our proposed threat model. There are two broad categories considered in this process, namely, a single attack pulse, and repeated attack pulses.

5.1. Single Attack Pulse Method

For a single attack pulse, the adversarial qubits are driven at the frequency and amplitude of a target qubit (victim in this case). An example of the pulse sequencing representation at the gate level is shown in Figure 7, where, one attack pulse is shown. An example of the pulse-level representation of the attack pulse of this is shown in Figure 6a, where the attack pulse has a duration of $160\ dt$ ¹. It should be noted that for each attack scenario considered in Section 4, the adversary can freely manipulate the duration of the attack pulse used to drive the adversarial qubits. Figure 6b depicts a single attack pulse with an increased duration of $500\ dt$.

5.2. Repeated Attack Pulse Method

Repeated attack pulses are created by sequencing multiple single attack pulses periodically, with a delay inserted between each attack pulse. An example of the pulse sequencing representation at the gate level is shown in Figure 8, where, three attack pulses are encoded with uniform delays interleaved between successive attack pulses. An example pulse-level representation of repeated attack pulses is depicted in Figure 6c, where the total duration of the repeated attack pulses is $2500dt$. Each attack pulse has a duration of $500dt$ which are separated by inserting uniform delays.

6. Experimental Setup

This section discusses the experimental setup used in this work. In particular the experiments are done on real IBM quantum computers, discussed below.

6.1. Hardware Utilized

We conducted our experiments to evaluate the threat posed by the different attack scenarios described in Section 4 using 127-qubit Eagle r3 processors offered by IBM Quantum. Specifically, we utilize three quantum processors available for public access, namely, *ibm_brisbane*, *ibm_osaka* (retired as of August 13th, 2024) and *ibm_kyoto*.

6.2. Attack Pulses

We design custom X-gate pulses to drive the adversarial qubits. The frequency and amplitude of these pulses

¹. dt is the system cycle time that defines the frequency of quantum operations, and $dt = 2.22ns$ is determined by the backend.

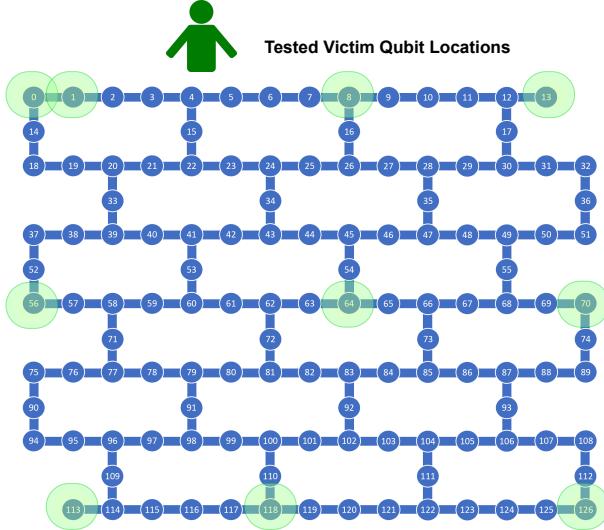


Figure 9: Each of the 10 qubits located in distinct parts of the processor topology, namely, qubits 0, 1, 8, 13, 56, 64, 70, 113, 118, 126, are selected as victims for evaluating which qubits are most vulnerable to our attacks.

are derived from the victim qubit through a two-step calibration experiment:

- First, we perform a frequency sweep on the victim qubit using spectroscopy, based on the default qubit frequency, which depends on the specific backend being used, to precisely estimate the victim qubit's frequency.
- Secondly, we calibrate the pulse amplitude of the victim qubit using a Rabi experiment, where we apply a pulse at the estimated frequency (obtained in the previous step) and identify the amplitude that induces a rotation of the desired angle π [23].

With the obtained frequency and amplitude of the victim's X-gate, we construct our custom X-gate pulses, which are subsequently used to drive the adversarial qubits. For our experiments, a pulse duration of $160dt$ is used, except for the extended pulse attacks. It is important to note that the adversary can manipulate the number of attack pulses driving the adversarial qubits. To address this, we categorize the impact of each attack scenario into two broad categories: single-pulse attacks and repeated-pulse attacks as discussed before.

6.3. Evaluation Metrics

To evaluate the impact of the various attack scenarios discussed in this paper, we use the variational distance. Also known as the total variation distance, this metric quantifies the difference between two probability distributions. It is defined as:

$$D_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{x \in X} |P(x) - Q(x)| \quad (3)$$

In Equation 3, $P(x)$ and $Q(x)$ represent the probabilities of the outcomes x under the distributions P and Q , respectively. The summation is taken over all possible

outcomes x in the sample space X . The factor of $\frac{1}{2}$ ensures that the variational distance ranges from 0 to 1. As an evaluation metric, the variational distance quantifies the maximum discrepancy between two distributions. A distance ranging from 0 to 0.2 indicates a minimal impact of the attack, while a distance between 0.2 and 0.4 suggests a mild influence. A distance of 0.4 to 0.6 implies a significant impact, and a distance from 0.6 to 1 denotes a very high impact of the attack.

6.4. Evaluation Benchmarks

For our evaluation of different attack scenarios, we employ three distinct benchmarks: one for single-pulse attacks and another for repeated-pulse attacks. For single-pulse attacks, we consider a single idle victim qubit which, upon measurement, is expected to produce the output '0' with high fidelity. For repeated-pulse attacks, we consider two victim qubits using which a two-qubit Grover's circuit is executed. The ideal output in this case is '11', also expected with high fidelity. Finally, we also test the attacks on a commonly used QAOA circuit.

6.5. Evaluation Circuits

To evaluate the attacks, we developed custom circuits to emulate the behavior of the victim and the attacker. An example circuit is shown in Figure 12. Each evaluation circuit is composed of two parts. First, the victim benchmark, e.g., idle qubit, Grover's circuit, or QAOA. Second, the attacker circuit, e.g., single pulses, repeated pulses, etc. Although in the circuit diagram, the attacker and victim are shown to be located next to each other, when the circuits are mapped to physical qubits, they will be physically far apart in the quantum chip, as shown in Figure 5 earlier in the paper.

7. Experimental Results

In this section, we present the results of each Attack Scenario. We evaluate the four Attack Scenarios, each for the two attack methods: *single pulse* and *repeated pulse*.

7.1. Evaluation of Attack Scenario 1

In this section, Attack Scenario 1 is examined, where the adversary has access to an extensive number of qubits in the quantum processor.

7.1.1. Single Attack Pulse Method. In this case, a single qubit is designated as the victim qubit, and all other qubits on the quantum processor are allocated to the adversary. Since each qubit has distinct properties, it is important to verify any dependence of attack potency on the victim qubit. To this end, we examine the efficacy of the attack over 10 distinct qubits, situated at different physical locations across the topology of the quantum processor, as depicted in Figure 9. In this figure, the quantum processor shown is *IBM_brisbane*, and the set of victim qubits comprises qubits 0, 1, 8, 13, 56, 64, 70, 113, 118, and 126. The results of our experiments for Attack Scenario 1 for each victim qubit are highlighted

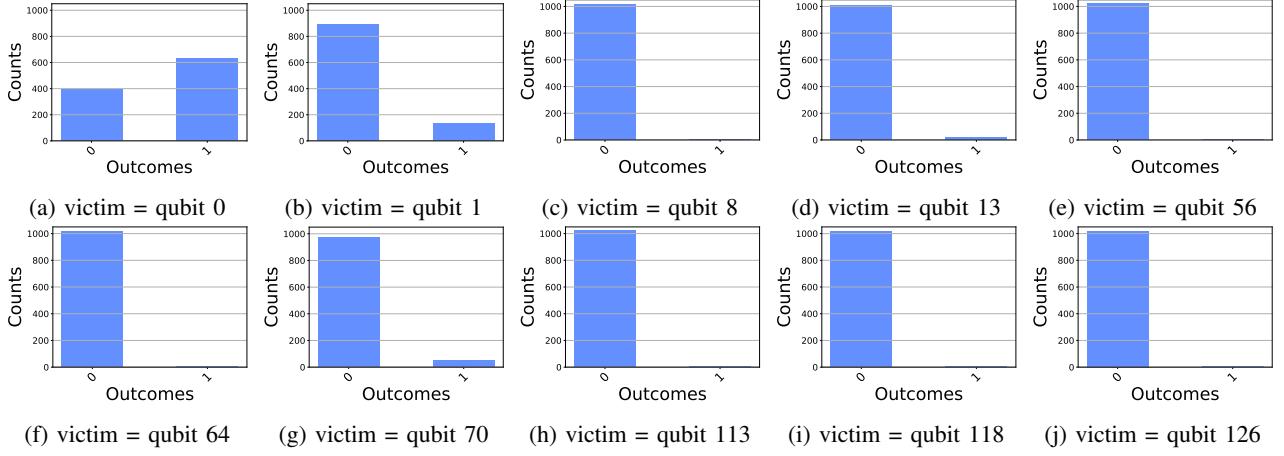


Figure 10: Evaluation of Attack Scenario 1 on *IBM_brisbane* using a single attack pulse on all 126 adversarial qubits, this is extreme version of Attack Scenario 1.

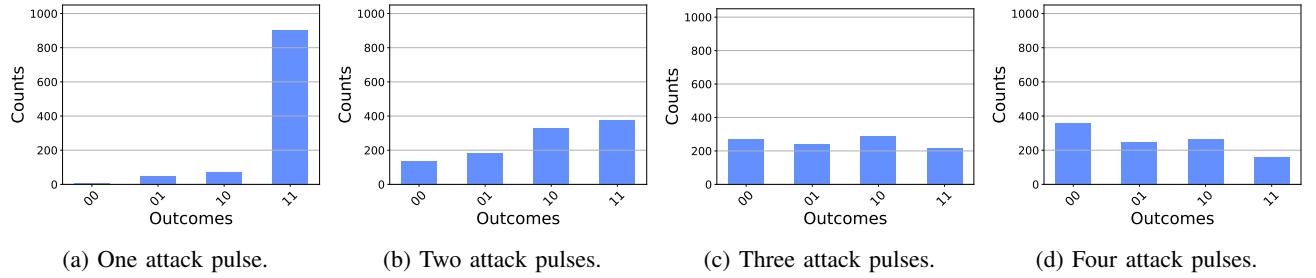


Figure 11: Effect of varying the number of attack pulses used to drive attacker qubits, on attack potency on quantum algorithms with Attack Scenario 1.

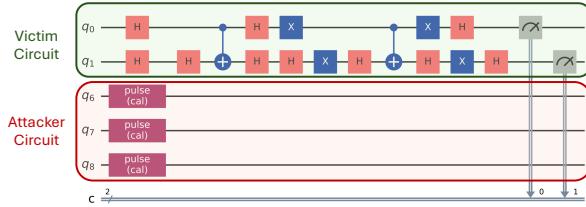


Figure 12: An example quantum circuit used in the evaluation.

TABLE 1: Impact of Attack Scenario 1 with a Single Pulse method, evaluated on *IBM_Brisbane* machine.

Victim qubit	0	1	8	13	56	64	70	113	118	126
Variational Distance	0.609	0.123	0.003	0.009	0.004	0.003	0.042	0.0059	0.0029	0.0039

in Figure 10. In these figures, the possible output states are denoted in the x-axis, and their corresponding counts obtained are shown in the y-axis. From the figure, it can be observed that there is a pronounced effect of the attack on victim qubit 0, shown in Figure 10a. The variation distance was evaluated for each of these experiments, and the impact of the attack is summarized in Table 1. In this table, the first row denotes the victim qubits, and the second row depicts the variational distance from the ideal output. From the table, it can be observed that the attack is highly impactful on qubit 0, with a variational distance of 0.609, but has negligible impact on the other qubits.

To establish the plausibility of this attack across different quantum hardware, we conduct the same attack on

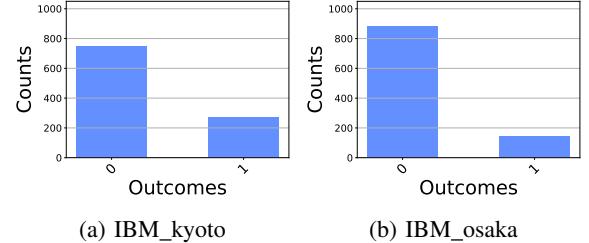


Figure 13: Demonstration of Attack Scenario 1 on additional IBM Quantum machines: *IBM_kyoto* and *IBM_osaka* on victim qubit. Single pulse attack method is tested, targeting victim qubit 0.

backends *IBM_kyoto* and *IBM_osaka*. For both systems, qubit 0 was allocated as the victim qubit, and all other qubits were assigned as adversarial qubits. The results furnished by these experiments are displayed in Figure 13. Upon evaluation of attack potency, variational distances of 0.26 and 0.131 were furnished for *IBM_kyoto* and *IBM_osaka*, respectively, indicating a minimal to moderate attack efficacy. Similar to *IBM_brisbane*, the attack was much less effective on other victim qubits.

7.1.2. Repeated Attack Pulses Method. In this experiment, we allocate qubits 0 and 1 to the victim, and all other qubits on the quantum processor are assigned to the adversary. Furthermore, to accentuate the impact of this attack in real-world scenarios, we consider a two-qubit Grover's circuit being executed using the victim

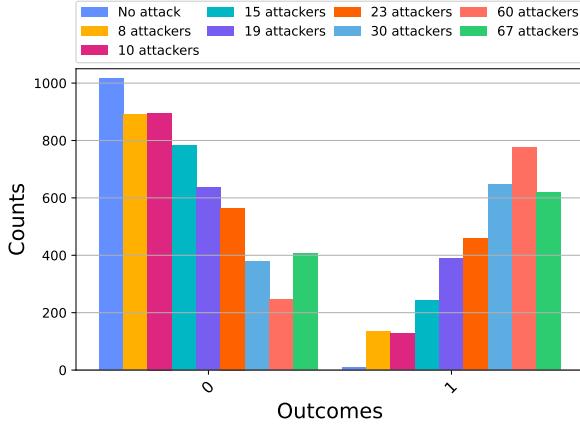


Figure 14: Evaluation of Attack Scenario 2 with a Single Pulse method on *IBM_brisbane*.

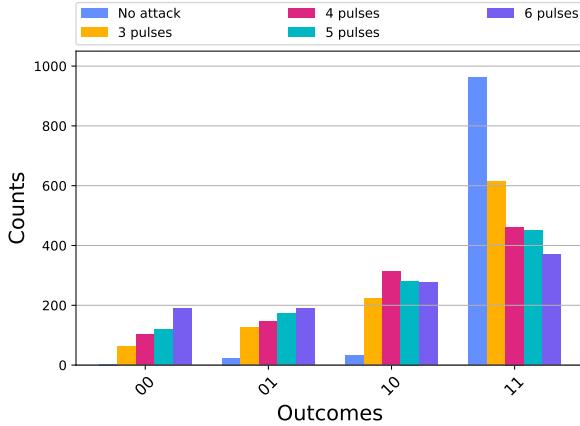


Figure 15: Evaluation of Attack Scenario 2 with Repeated Pulses method on *IBM_brisbane*.

qubits. An example of such a setup is shown in Figure ???. The number of attack pulses on each adversarial qubit is gradually increased from one to four, and their impact on the output fidelity of the victim circuit is displayed in Figure 11. The output is observed to become progressively more skewed as the number of attack pulses applied to adversarial qubits is increased. Upon evaluation, variational distances of 0.074 , 0.586 , 0.741 , and 0.801 were furnished for *one*, *two*, *three*, and *four* attack pulses, respectively. This indicates that attack potency increases from minimal to highly effective with an increase in the number of attack pulses employed by the adversary.

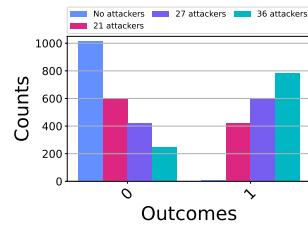
7.2. Evaluation of Attack Scenario 2

In this section, we consider Attack Scenario 2, where the adversarial qubits are limited, but located in physical proximity to the victim’s qubits on the quantum processor.

7.2.1. Single Attack Pulse Method. For this evaluation, we consider that qubit 0 is allocated to the victim, and the number of adversarial qubits varies from 8 to 67. The results furnished by this attack are depicted in Figure 14. From this figure, it can be observed that as the number of attackers increases, the output fidelity decreases noticeably. The evaluation results for these experiments

TABLE 2: Impact of Attack Scenario 2 with a Single Pulse method.

Number of attackers	8	10	15	19	23	30	60	67
Variational Distance	0.122	0.116	0.227	0.37	0.439	0.623	0.75	0.596



(a) Attackers at distance 13.

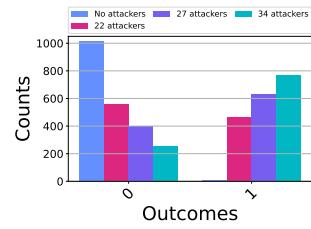


Figure 16: Evaluation results for Attack Scenario 3 with varying number of adversarial qubits on *IBM_kyoto*.

are summarized in Table 2, where the first row denotes the number of attackers, and the second row depicts the evaluated variational distance. From the table, it can be observed that the attack potency increases from minimal (variational distance of 0.122 for 8 adversarial qubits) to highly effective (variational distance of 0.75 for 60 adversarial qubits). These results highlight that a high number of adversarial qubits increases the efficacy of this Attack Scenario.

7.2.2. Repeated Attack Pulses Method. To assess the impact of repeated attack pulses, we allocate qubits 0 and 1 to the victim and qubits 5 to 36 to the adversary. The results of this experiment, illustrated in Figure 15, show the effect of increasing the number of attack pulses from one to three. As the number of attack pulses increases, a noticeable reduction in output fidelity is observed. Upon evaluation, variational distances of 0.342 , 0.489 , 0.501 , and 0.578 were furnished for *three*, *four*, *five* and *six* attack pulses, respectively. This demonstrates that even with a limited number of qubits under the adversary’s control, the victim qubits can be significantly affected by an increased number of attack pulses.

7.3. Evaluation of Attack Scenario 3

Here, we evaluate the impact of Attack Scenario 3, in which the victim and adversarial qubits are significantly separated on the quantum processor, and each can be allocated up to half of the qubits.

7.3.1. Single Attack Pulse Method. In this evaluation, qubit 0 is assigned to the victim, while the adversary controls qubits located at distances of 13 and 17 from the victim. We vary the number of adversarial qubits to assess how this affects the potency of the attack. The results are presented in Figure 16, where the x-axis denotes the outcome of the circuit execution, and the y-axis denotes the fidelity of the quantum circuit executed. To illustrate the impact of varying the number of adversarial qubits, Figures 16a and 16b presents three sets of experiments, where each experiment showcases the corresponding variations in output fidelity based on the

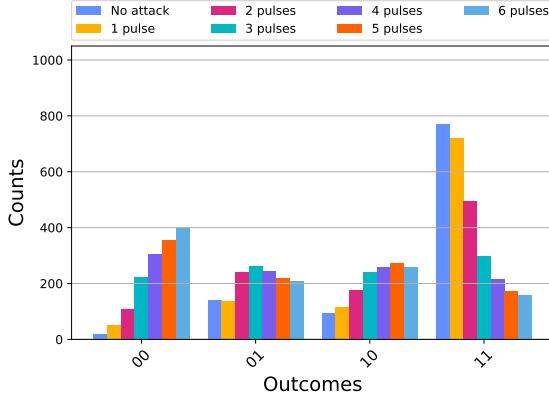


Figure 17: Evaluation of Attack Scenario 3 on *IBM_kyoto*.

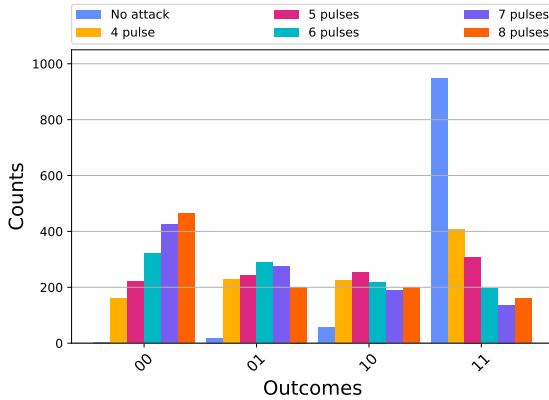


Figure 18: Evaluation of Attack Scenario 3 on *IBM_osaka*

number of adversarial qubits present at a specified distance from the victim qubit. Figure 16a shows the impact of the attack when the adversarial qubits are at a distance of 13 from the victim qubits, and the number of adversarial qubits is varied from 21 to 36. Variational distances of 0.401, 0.581, and 0.753 were obtained upon evaluation of attack efficacy for 21, 27, and 36 adversarial qubits, respectively. Similarly, Figure 16b shows the outcomes when the adversarial qubits are positioned at a distance of 17 from the victim qubits. The number of adversarial qubits is varied from 22 to 34. These experiments yield variational distances of 0.445, 0.604, and 0.743 for 22, 27, and 34 adversarial qubits, respectively. Both sets of results indicate a significant to very high attack efficacy under the proposed attack scenario.

7.3.2. Repeated Attack Pulses Method. For the evaluation of Attack Scenario 3 using repeated attack pulses, qubits 0 and 1 are allocated to the victim, executing a two-qubit Grover's algorithm. The adversary is assigned qubits 71 to 126, positioned 13 units away from the victim qubits. The experiments were conducted on the *IBM_kyoto* and *IBM_osaka* processors, with varying numbers of attack pulses driving the adversarial qubits.

The results furnished by this experiment are depicted in Figures 17 and 18, where the x-axis represents the outcomes of each experiment, and the y-axis represents the output fidelity furnished following each experiment. Figure 17 shows the impact on *IBM_kyoto* as the number

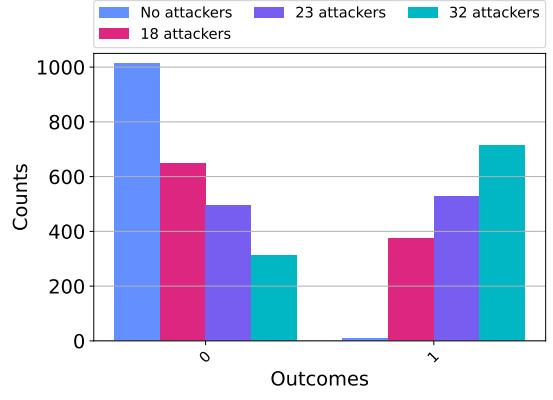


Figure 19: Evaluation of Attack Scenario 4 on *IBM_kyoto* with single attack pulse

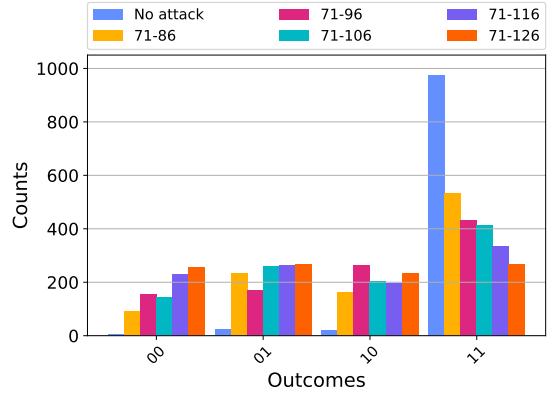


Figure 20: Evaluation of Attack Scenario 4 on *IBM_kyoto* with repeated attack pulses.

of attack pulses increases from zero (no attack) to six. Similarly, Figure 18 illustrates the effects on *IBM_osaka*, where the number of attack pulses is increased from four to eight, compared to the ideal output with no attack. These figures clearly demonstrate that as the number of attack pulses increases, there is a significant reduction in output fidelity. For *IBM_kyoto*, variational distances of 0.052, 0.269, 0.462, 0.541, and 0.597 were obtained for two to six attack pulses, in increasing order. For *IBM_osaka*, variational distances of 0.526, 0.624, 0.734, 0.791, and 0.770 were furnished for four to eight attack pulses, in increasing order, respectively. These results emphasize a high attack efficacy for a high number of attack pulses on adversarial qubits, in addition to demonstrating attack efficacy across different quantum hardware.

7.4. Evaluation of Attack Scenario 4

In this section, we assess the final Attack Scenario, where the adversary controls a moderate number of qubits (fewer than 35) and is positioned at a moderate distance (15-20) from the victim qubits within the quantum processor. This scenario is particularly important because it reflects a situation where the adversary has limited resources but is still capable of mounting a significant attack.

7.4.1. Single Attack Pulse Method. To evaluate the impact of a single attack pulse in this scenario, we assign

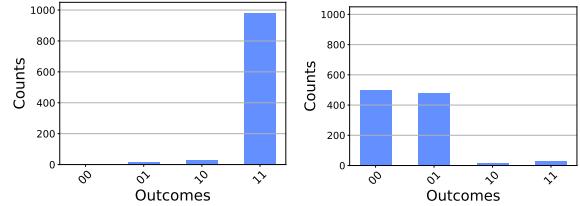
qubit 0 to the victim and allocate adversarial qubits at a distance of 9 from the victim. We then systematically increase the number of adversarial qubits from 18 to 32 to analyze how the number of qubits under the adversary's control influences the effectiveness of the attack. The results, summarized in Figure 19, clearly demonstrate that as the number of adversarial qubits increases, the output fidelity of the victim's quantum operations decreases. Notably, it can be observed that using 23 or more adversarial qubits causes the victim's circuit to produce erroneous output. This is corroborated by the evaluation results, where variational distances of 0.357, 0.508, and 0.688 were obtained for 18, 23, and 32 adversarial qubits, respectively. This indicates that victim circuit output becomes increasingly disrupted as attack efficacy increases due to the increasing number of adversarial qubits.

7.4.2. Repeated Attack Pulses Method. To explore the effects of repeated attack pulses in this scenario, we consider a setup where the victim is allocated qubits 0 and 1, which are used to execute a two-qubit Grover's algorithm. In this experiment, the adversary initially controls qubits 71 to 126, with the number of adversarial qubits gradually reduced to 71 to 86, while five attack pulses are applied. The impact on output fidelity is depicted in Figure 20, where the x and y axes denote possible outcomes and the output fidelities, respectively. The outcomes of all the experiments are grouped together to emphasize the impact of the attack and its dependence on the number of adversarial qubits. As shown in the figure, reducing the number of adversarial qubits diminishes the potency of the attack. Upon evaluation, variational distances of 0.429, 0.501, 0.521, 0.600, and 0.664 were obtained for adversarial qubits in the range 71-86, 71-96, 71-106, 71-116, and 71-126, respectively. This indicates a significant impact of the attack despite a limited number of adversarial qubits (for range 71-86, a variational distance of 0.429 was obtained). Attack Scenario 4 demonstrates that even when the adversary is constrained by both the number of qubits and the distance from the victim, they can still exert considerable influence over the outcome of the victim's quantum operations.

7.5. Additional Evaluation

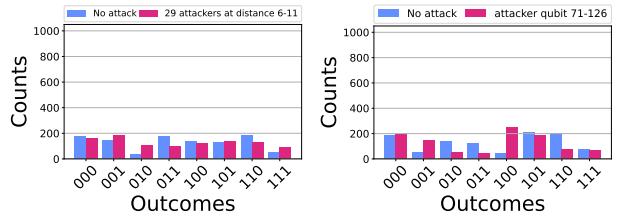
In this section we present addition evaluation of the attacks considering how higher optimization levels used in circuit compilation may help or hurt the attacks. Further, we apply the attack to QAOA, an algorithm with promising applications to real-world problems.

7.5.1. Attacks when Higher Optimization Level is Used. In the previous sections, all attacks were executed on victim circuits transpiled with optimization level 0 to clearly demonstrate the efficacy of the attacks. To further validate our findings, we conducted an additional experiment using optimization level 3 on Grover's circuit. In this experiment, qubits 0 and 1 were assigned to the victim, while qubits 71 to 126 were allocated to the adversary, with five attack pulses applied (attack scenario 3 with repeated pulses, as mentioned in Section 7.3.2). The results of this experiment are presented in Figure 21, where it is evident that the output fidelity is completely skewed



(a) Grover circuit output with (b) Output following attack optimization level 3. on optimization level 3.

Figure 21: Additional evaluation, efficacy of Attack Scenario 3 on maximally optimized Grover's two-qubit circuit.



(a) Attack Scenario 2 on (b) Attack Scenario 3 on QAOA. QAOA.

Figure 22: Additional evaluation, efficacy of Attack Scenario 2 and 3 on a three-qubit QAOA circuit.

following the attack execution. A variational distance of 0.938 was furnished by this experiment, that indicates extremely high attack potency. Additionally, since circuits are optimized to reduce impact of errors, this finding implies existing optimization techniques are rendered useless against our proposed attack. These findings emphasize the urgent need for developing robust countermeasures to ensure a secure and trusted cloud-based quantum computing.

7.5.2. Attacks on QAOA. In addition to exploring various attack scenarios using Grover's two-qubit circuit, we extended our testing to a more complex four-qubit QAOA algorithm on the *IBM_brisbane* quantum processor, utilizing attack scenarios 2 and 3. The results, shown in Figures 22a and 22b, correspond to variational distances of 0.17 and 0.307 for attack scenarios 2 and 3, respectively, indicating nominal to mild impacts. Notably, despite only four adversarial pulses being used in attack scenario 3, the attack still proved moderately effective, underscoring the vulnerability of quantum algorithms even under limited adversarial conditions.

8. Overall Summary of QubitHammer Attacks and Attack Evaluation Results

This section summarizes the novel QubitHammer attacks, presented in this paper. We have presented four attack scenarios, each considering two strategies: *single* and *repeated pulse*.

Table 3 summarizes the effectiveness of the attacks. In this table, the first column summarizes the attack scenario utilized. Columns 2, 4, and 6 denote the maximal variational distance obtained by our experiments for three attack methods, respectively. Columns 3, 5, and 7 show whether the attack is successful or not. The table shows that while Attack Scenario 1 is the most potent, with a

TABLE 3: Overall summary of the attacks and their effectiveness. The best (highest) variational distance for each attack scenario and method is summarized in this table. We assume that if the variational distance is larger than 0.2, then the corresponding attack is successful. The checkmark represents the success of the attack, while the cross means the attack failed.

Attack Scenario	Attack Method					
	Single Attack Pulse on Single Victim Qubit		Repeated Attack Pulses on Grover		Repeated Attack Pulses on QAOA	
	Variational Distance	Attack Succeeded	Variational Distance	Attack Succeeded	Variational Distance	Attack Succeeded
Scenario 1	0.609	✓	0.801	✓	-	-
Scenario 2	0.750	✓	0.578	✓	0.17	✗
Scenario 3	0.753	✓	0.770	✓	0.307	✓
Scenario 4	0.688	✓	0.664	✓	-	-

variational distance of up to 0.801 furnished with repeated attack pulses, it is less feasible since it assumes a high number of qubits under the adversary’s control. However, it is worth noting that the most realistic attack vector, represented by Attack Scenario 4, furnishes variational distances of up to 0.688 and 0.664 for single and repeated pulse attacks, respectively. Moreover, we demonstrate that, even at an optimization level of 3, our attack yielded a variational distance as high as 0.938. As explained in Section 6, a higher variational distance indicates a greater impact of the attack on the victims. We consider the attack to be successful when the variational distance exceeds 0.2. Consequently, we find that all attacks, except for one, are successful, demonstrating the overall effectiveness of our QubitHammer attacks. This presents an unprecedented threat to multi-tenant quantum computers since a run-of-the-mill user would be able to jeopardize computational processes of high importance being executed on the same QPU.

Our evaluations discovered that qubit 0 on three IBM machines, *IBM_brisbane*, *IBM_kyoto*, and *IBM_osaka* is extremely vulnerable to the QubitHammer attacks. Consequently, any circuit using qubit 0 can be manipulated easily. Each of the four scenarios is evaluated using both single victim qubits and the Grover algorithm. Additionally, for attack scenarios 2 and 3, we expanded our testing to include the QAOA algorithm.

9. Effectiveness of QubitHammer Against Existing Countermeasures

Recent research has explored approaches for mitigating crosstalk for superconducting quantum computers. In this section we demonstrate the existing defenses are not sufficient against our novel attacks.

One effective technique is dynamical decoupling, which reduces interqubit crosstalk errors by inserting pulse sequence on idling qubits [3] [32]. However, our experiments demonstrate that the proposed QubitHammer attacks still significantly affect systems protected by dynamical decoupling. This is shown in Figure 23, where a two-qubit Grover’s circuit, with dynamical decoupling, is executed on *IBM_brisbane* and *IBM_kyoto*. In both IBM systems, the circuit is subjected to the proposed QubitHammer attacks, specifically, attack scenario 3, with five attack pulses. The evaluation of the attack revealed a variational distance of 0.747 and 0.837 for *IBM_brisbane* and *IBM_kyoto*, respectively. Utilization of other attack scenarios also demonstrated similar degradation in performance. These results indicate that a well-established error

suppression method like dynamical decoupling fails to mitigate the threat of the proposed QubitHammer attacks.

A naive defense solution would be simply disabling qubit 0 on these machines. However, similar issues could arise with other qubits. Moreover, disabling qubit 0 merely conceals the problem without identifying or addressing the underlying issue. Other potential software-based defenses including readout discriminator [15], randomized compiling [10], frequency-aware compilation [7] and crosstalk-adaptive scheduling [17], can be extended to multi-tenant environments in the future to protect against our QubitHammer attacks.

Crosstalk-aware qubit allocation mechanisms in a multi-tenant quantum computer have been explored in existing research [33]. This approach proposes allocating idle qubits as padding between different quantum programs to minimize the effect of crosstalk. However, as explained in Attack Scenario 3 in Section 4.3, such a defense is ineffective against our proposed attack.

Another potential defense would be to employ active padding on the qubits between the victim and adversary to reduce crosstalk. This involves executing single gate operations on qubits, which result in identity at the end of execution. Examples of such sequences include an even number of X-gates, an even number of XY-gate sequences, or an even number of Y-gates. This prevents qubits from remaining idle, mitigating the risk of idling crosstalk errors. We evaluated our attack in the presence of this defense strategy by employing Attack Scenario 3 where the victim is assumed to be executing Grover’s circuit on qubits 0 and 1. Figure 24 depicts the resulting distribution obtained. This has a variational distance of 0.73, thereby demonstrating the effectiveness of our attack in the presence of such defenses.

It is important to note that simply preventing the adversary from deploying malicious pulses in the cloud environment is not feasible. Firstly, detecting custom pulses as an attack is inherently challenging, as no established precedent exists for such detection. Further, custom pulses have valid applications in quantum machine learning, so disabling such functionality is impractical.

A summary of existing defense strategies, and how QubitHammer bypasses them, is shown in Table 4. The first column in the table depicts the defenses, followed by their description in the second column. Finally, the third column displays the effectiveness of our proposed QubitHammer attacks in the presence of these defensive strategies. We focus on four possible attack scenarios using QubitHammer, and if any one of these scenarios successfully circumvents the defense, we categorize the

TABLE 4: Effectiveness of QubitHammer attacks against existing countermeasures.

Defense	Defense Description	Attack Success
Dynamical decoupling	Add dynamical decoupling sequences to program	✓
Disabling qubit 0	Removing qubit 0 during qubit allocation	✓
Crosstalk-aware qubit allocation	Allocating idle qubits as padding between two quantum circuits	✓
Active padding	Allocating active qubits as padding between two quantum circuits	✓
Disabling custom pulses	Disabling deployment of user-designed custom pulses	—

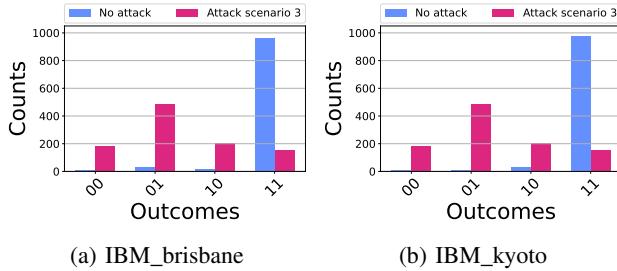


Figure 23: Evaluation of Attack Scenario 3 using dynamical decoupling on *IBM_brisbane* and *IBM_kyoto*.

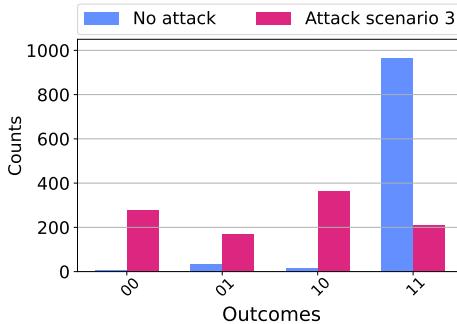


Figure 24: Evaluation of Attack Scenario 3 with padding on idle qubits on *IBM_brisbane*.

attack as a success. It should be noted that in all cases, QubitHammer proves effective in bypassing the existing defensive strategies. This is especially concerning since disabling custom pulses, which could potentially mitigate the attack, is not a viable solution due to the significant restrictions it would place on algorithm deployment, severely impacting the practical utility of quantum computers. Consequently, we do not consider disabling custom pulses as a feasible defense in our evaluation.

10. Related Work

Research on developing large-scale quantum computers is booming across the world. IBM released an ambitious roadmap for over 1000-qubit system. However, due to imperfections in current quantum computers, exploring security and privacy issues is also a necessary topic.

Malicious users could exploit shared quantum computing environments to deduce the quantum states of other users' qubits by analyzing leaked data from computation results. This type of leakage is termed "horizontal" leakage, where information flows sequentially from one execution to the next. Such "horizontal" leakage has been exploited in various attacks, such as reset

attacks [16] [31], side-channel attacks [2], and higher-energy state attacks [34]. On the other hand, "vertical" leakage occurs across qubits simultaneously, presenting another form of vulnerability, as seen in crosstalk attacks [9] [1] and qubit sensing [29].

In parallel, various multi-tenant schemes aimed to optimize throughput by allowing multiple users to run their circuits on the same quantum computer simultaneously, but utilizing different qubits, are being proposed [5] [14]. While this concept holds significant promise for enhancing the utilization of quantum computing resources, it also exposes quantum computers to security vulnerabilities caused by crosstalk errors [30] [17] [16].

Malicious users can exploit crosstalk in various ways. For instance, the entangling effect of CNOT gates has been shown to induce substantial crosstalk in superconducting quantum computers, allowing an attacker to construct a circuit designed to disrupt adjacent qubits rather than perform legitimate computations [9]. Crosstalk attacks can also potentially extract sensitive information from state-preparation circuits if the attacker is aware of the QPU's co-tenancy at a specific time [9]. Furthermore, crosstalk can be used to target specific quantum algorithms, such as ensuring that Grover's algorithm consistently returns incorrect results, thereby sabotaging more complex quantum jobs [6]. As quantum computers grow in complexity, the risks associated with crosstalk attacks become more prevalent, necessitating extensive defense mechanisms.

11. Conclusion

In this work, we introduced the QubitHammer, a novel set of qubit-flipping attacks targeting state-of-the-art superconducting quantum computers. Our attack was validated across three public access IBM quantum computers, namely, *IBM_brisbane*, *IBM_kyoto* and *IBM_osaka*, highlighting significant security and reliability concerns potentially extending to other cloud-based superconducting quantum systems. Further, the attacks were demonstrated to bypass all existing defenses that have been so-far proposed for defending against crosstalk-based attacks in superconducting quantum computers. This indicates the necessity for anticipating and further addressing such vulnerabilities in the design of current and future quantum computers. In particular, there is a need to develop a fundamental understanding of cross-talk and other physical effects in quantum computers, so that better mitigations of the attacks can be developed.

12. Ethical Considerations

In this research, we have carefully considered the potential risks and harms that could arise from both the

conduct of our study and the publication of its findings on the nascent field of quantum computation. The experiments were developed to not cause physical harm to the quantum computer hardware. No sensitive or private data was used and all algorithms used are publicly known. While the results show new types of security attacks, we believe finding and publishing them early will help make nascent quantum computing systems more secure and direct attention of researchers and industry to protect against these systems against attacks.

References

- [1] A. Ash-Saki, M. Alam, and S. Ghosh, "Experimental characterization, modeling, and analysis of crosstalk in a quantum computer," *IEEE Transactions on Quantum Engineering*, vol. 1, pp. 1–6, 2020.
- [2] B. Bell and A. Trügler, "Reconstructing quantum circuits through side-channel information on cloud-based superconducting quantum computers," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2022, pp. 259–264.
- [3] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, "Noise spectroscopy through dynamical decoupling with a superconducting flux qubit," *Nature Physics*, vol. 7, no. 7, pp. 565–570, 2011.
- [4] C. Chamberland, K. Noh, P. Arrangoiz-Arriola, E. T. Campbell, C. T. Hann, J. Iverson, H. Puterman, T. C. Bohdanowicz, S. T. Flammia, A. Keller *et al.*, "Building a fault-tolerant quantum computer using concatenated cat codes," *PRX Quantum*, vol. 3, no. 1, p. 010329, 2022.
- [5] P. Das, S. S. Tannu, P. J. Nair, and M. Qureshi, "A case for multi-programming quantum computers," in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, 2019, pp. 291–303.
- [6] S. Deshpande, C. Xu, T. Trochatos, Y. Ding, and J. Szefer, "Towards an antivirus for quantum computers," in *2022 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE, 2022, pp. 37–40.
- [7] Y. Ding, P. Gokhale, S. F. Lin, R. Rines, T. Propson, and F. T. Chong, "Systematic crosstalk mitigation for superconducting qubits via frequency-aware compilation," in *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 2020, pp. 201–214.
- [8] C. Gonzalez, "Cloud-based qc with amazon braket," *Digitale Welt*, vol. 5, no. 2, pp. 14–17, 2021.
- [9] 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*, 2024.
- [10] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien *et al.*, "Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor," *arXiv preprint arXiv:2010.00215*, 2020.
- [11] IBM Quantum, "Ibm quantum platform," 2024, <https://quantum.ibm.com/>.
- [12] IQuEra, "Quantum-as-a-service: Definition, advantages and examples," 2023, <https://www.quera.com/blog-posts/quantum-as-a-service-definition-advantages-and-examples>.
- [13] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, "Quantum computing with Qiskit," 2024.
- [14] L. Liu and X. Dou, "Qucloud: A new qubit mapping mechanism for multi-programming quantum computing in cloud environment," in *2021 IEEE International symposium on high-performance computer architecture (HPCA)*. IEEE, 2021, pp. 167–178.
- [15] S. Maurya, C. N. Mude, W. D. Oliver, B. Lienhard, and S. Tannu, "Scaling qubit readout with hardware efficient machine learning architectures," in *Proceedings of the 50th Annual International Symposium on Computer Architecture*, 2023, pp. 1–13.
- [16] A. Mi, S. Deng, and J. Szefer, "Securing reset operations in nisq quantum computers," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 2279–2293.
- [17] P. Murali, D. C. McKay, M. Martonosi, and A. Javadi-Abhari, "Software mitigation of crosstalk on noisy intermediate-scale quantum computers," in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, 2020, pp. 1001–1016.
- [18] S. Neyens, O. K. Zietz, T. F. Watson, F. Luthi, A. Nethewewala, H. C. George, E. Henry, M. Islam, A. J. Wagner, F. Borjans *et al.*, "Probing single electrons across 300-mm spin qubit wafers," *Nature*, vol. 629, no. 8010, pp. 80–85, 2024.
- [19] J. Pena, "Quantum diamond biomarker detection," *PhotonicsViews*, vol. 19, no. 1, pp. 48–50, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/phvs.202270107>
- [20] K. Prateek and S. Maity, "Quantum programming on azure quantum—an open source tool for quantum developers," in *Quantum Computing: A Shift from Bits to Qubits*. Springer, 2023, pp. 283–309.
- [21] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [22] qBraid, "qbraid lab," <https://lab.qbraid.com>, 2024.
- [23] Qiskit Development Team , "Calibrations: Schedules and gate parameters from experiments," 2024, <https://qiskit-community.github.io/qiskit-experiments/tutorials/calibrations>.
- [24] Quantinuum, "Quantinuum," 2024, <https://www.quantinuum.com/>.
- [25] QuEra Computing Inc, "Quera," 2023, <https://www.quera.com/>.
- [26] G. S. Ravi, K. N. Smith, P. Gokhale, and F. T. Chong, "Quantum computing in the cloud: Analyzing job and machine characteristics," in *2021 IEEE International Symposium on Workload Characterization (IISWC)*. IEEE, 2021, pp. 39–50.
- [27] Rigetti Computing, "Rigetti," 2024, <https://www.rigetti.com/>.
- [28] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown *et al.*, "Realization of real-time fault-tolerant quantum error correction," *Physical Review X*, vol. 11, no. 4, p. 041058, 2021.
- [29] A. A. Saki and S. Ghosh, "Qubit sensing: A new attack model for multi-programming quantum computing," *arXiv preprint arXiv:2104.05899*, 2021.
- [30] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, "Detecting crosstalk errors in quantum information processors," *Quantum*, vol. 4, p. 321, 2020.
- [31] J. Tan, C. Xu, T. Trochatos, and J. Szefer, "Extending and defending attacks on reset operations in quantum computers," *arXiv preprint arXiv:2309.06281*, 2023.
- [32] V. Tripathi, H. Chen, M. Khezri, K.-W. Yip, E. Levenson-Falk, and D. A. Lidar, "Suppression of crosstalk in superconducting qubits using dynamical decoupling," *Physical Review Applied*, vol. 18, no. 2, p. 024068, 2022.
- [33] S. Upadhyay and S. Ghosh, "Share: Secure hardware allocation and resource efficiency in quantum systems," 2024. [Online]. Available: <https://arxiv.org/abs/2405.00863>
- [34] C. Xu, J. Chen, A. Mi, and J. Szefer, "Securing nisq quantum computer reset operations against higher energy state attacks," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 594–607.
- [35] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.

Appendix A. Data Availability

All code, data, and materials necessary to reproduce the results of this paper will be made publicly available upon acceptance.

These artifacts are shared in the following link:
<https://anonymous.4open.science/r/QubitHammer-1E04/>.

Any proprietary tools or third-party libraries used will be clearly documented, along with instructions for their acquisition.