Detection of Photon-Number Splitting Attack in Quantum Key Distribution Networks

Darsh Shani
Dept of CSIS
BITS Pilani, Hyderabad, India
darshshani7@gmail.com

Vishwas Vedantham
Dept of CSIS
BITS Pilani, Hyderabad, India
vishwav1410@gmail.com

Prachi Shah
Dept of CSIS
BITS Pilani, Hyderabad, India
pks14vadodara@gmail.com

Geethakumari G
Dept of CSIS
BITS Pilani, Hyderabad, India
geetha@hyderabad.bits-pilani.ac.in

***Abstract*—Current cryptographic systems rely on the computational difficulty of factoring large prime numbers for both key encapsulation and encryption. Shor's quantum algorithm exploits quantum mechanical principles to efficiently factor large numbers, enabling quantum computers to compromise most contemporary cryptographic algorithms. However properties of quantum physics like quantum entanglement can be used to strengthen security through post-quantum cryptographic approaches that provide quantum-resistant key distribution mechanisms. Quantum Key Distribution (QKD) uses quantum mechanical properties to enable secure key exchange between parties, providing protection against both classical and quantum computational attacks. Despite its promise, QKD faces a significant vulnerability: the Photon-Number Splitting (PNS) attack. In QKD, a Photon Number Splitting(PNS) attack is a mechanism used by an adversary to steal information by diverting a portion of multi-photon pulses sent by the legitimate user. Practical QKD implementations face technological limitations in producing true single-photon sources, which are essential for the theoretical security proofs of quantum cryptographic protocols. Instead they rely on weak laser pulses containing multiple photons per qubit. Adversaries can exploit this limitation by intercepting multi-photon pulses, retaining one photon for analysis while forwarding the remaining photons to the intended recipient. This approach enables attackers to extract substantial key information while introducing minimal transmission disturbance, making detection extremely challenging. This paper addresses the critical need for PNS attack detection and proposes a novel algorithm specifically designed to identify such attacks in QKD networks. The approach in this paper provides a foundation for enhancing the security and reliability of quantum key distribution systems against this sophisticated threat.**

***Index Terms*—Post-Quantum Cryptography, Quantum Key Distribution, Photon-Number Splitting Attack, Decoy states**

I. Introduction

Contemporary cryptographic protocols such as RSA (Rivest-Shamir-Adleman) and Elliptic Curve Cryptography (ECC) derive their security from computationally intractable mathematical problems. Specifically, RSA utilizes the integer factorization problem for cryptographically significant numbers, while ECC relies on the discrete logarithm problem in elliptic curve groups. These algorithms maintain their security through the exponential time complexity required for classical computers to solve these underlying mathematical challenges.

However, Shor's quantum algorithm [4] fundamentally disrupts this security paradigm by solving these problems in polynomial time on sufficiently powerful quantum computers [4]. For instance, contemporary classical computers would require millions of years to compromise a 2048-bit RSA key using the most efficient classical algorithms such as the General Number Field Sieve (GNFS). The computational record demonstrates this complexity: the largest number ever factored was a 250 decimal digit (829-bit) RSA encryption, accomplished in 2020 after requiring 2700 CPU core-years using 2.1 GHz Intel Xeon Gold 6130 processors as reference. In contrast, quantum computers with sufficient fault-tolerant qubits could potentially compromise a 2048-bit RSA key within hours or days [5]. Although current quantum computers remain in the noisy intermediate-scale quantum (NISQ) era, the imminent threat to existing encryption schemes necessitates immediate attention, with experts projecting the emergence of cryptographically relevant quantum computers within the next decade.

To address the quantum computational threat, Quantum Key Distribution (QKD) emerged as a promising solution. QKD protocols leverage the fundamental quantum mechanical principles to enable unconditionally secure key exchange between communicating parties. Charles Bennett and Gilles Brassard established the theoretical foundation in 1984 with the BB84 protocol [1], which has subsequently inspired numerous optimized variants targeting specific aspects of quantum key distribution [10]. The core mechanism of QKD protocols is based on the no-cloning theorem which deems it impossible to create an exact copy of an unknown quantum state. The security proofs of various Quantum Key Distribution algorithms are based on the assumption of using single-photon pulses since multi-photon states are not fundamentally protected by the no-cloning theorem. However, producing single-photon pulses is practically challenging because of the inherent properties of light sources and the technological limitations in controlling quantum states of light. So weak coherent pulses that reduce

the probability of emitting multiple photons are most commonly used in these applications. This exposes the system to Photon Number Splitting (PNS) attack, which was first mentioned in [6]. In a PNS attack, the eavesdropper intercepts the transmission and stores a photon from these multi-photon pulses while forwarding the rest of the photons. This allows the eavesdropper to gain partial information about the key without causing detectable errors [2].

The PNS attack hinges on the practical limitations of a QKD system to not be able to generate a single-photon pulse for each quantum state. With the PNS attack, an eavesdropper can gain partial information about a shared key without being detected. The detection of PNS attacks represents a critical requirement for maintaining trust in QKD-based secure communications. Thus, detection of PNS attacks is crucial for facilitating secure financial transactions, confidential military communications and protecting critical data shared over a QKD network.

II. Related work

Several research efforts have addressed the challenges and vulnerabilities in QKD systems. Xu et al. [10] provide a comprehensive survey of QKD protocols and post-quantum cryptographic approaches, establishing theoretical foundations but lacking specific solutions for PNS attack detection in basic QKD implementations. Sabottke et al. [3] proposed an Entanglement Enhanced BB84 (EE BB84) protocol to mitigate PNS attacks using time-entangled photon pulses and quantum non-demolition (QND) measurement detection. Their approach employs Chernoff distance and symmetric hypothesis testing to quantify eavesdropper detection confidence. While EE BB84 can detect attacks with fewer pulses under moderate loss conditions, it requires significant network augmentation with entangled decoy states and specialized beam splitters. Critical limitations include: (1) the method does not universally outperform existing coherent decoy state protocols, (2) it requires complex hardware modifications unsuitable for basic QKD networks. Existing approaches primarily focus on attack mitigation through network enhancement rather than developing detection algorithms applicable to standard QKD implementations. Current literature lacks efficient detection mechanisms that can identify PNS attacks in basic QKD networks without requiring additional hardware or protocol modifications. Our work addresses this gap by proposing a detection algorithm specifically designed for standard QKD networks, focusing on identifying PNS attacks through statistical analysis of transmission characteristics.

III. Photon-Number Splitting attack on Quantum Key Distribution networks

Consider the BB84 protocol [13]. Alice and Bob wish to exchange keys using this protocol. Owing to the implementation difficulty in generating true single photon sources, Alice will use weak coherent pulses to generate photons which will cause multi-photon pulses to be generated with high intensity. This is shown in Fig. 1(i).

When the distance between Alice and Bob is lesser, the probability of the PNS attack happening is low, and the system is relatively secure. However,if the distance is increased, to compensate for attenuation in the quantum channel, Alice will increase the intensity of the pulses leading to multi-photon pulses being generated with higher probability.

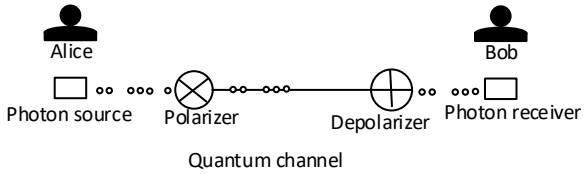

Fig. 1(i). Weak coherent pulses transmitted from Alice to Bob for QKD via BB84 protocol.

Eve, the eavesdropper, intercepts the quantum channel between Alice and Bob. She splits off one photon from a multi-photon pulse whenever she detects it. Further, she stores this in quantum memory and forwards the rest of the photons to Bob as depicted below in Fig. 1(ii) and Fig. 1(iii) respectively.

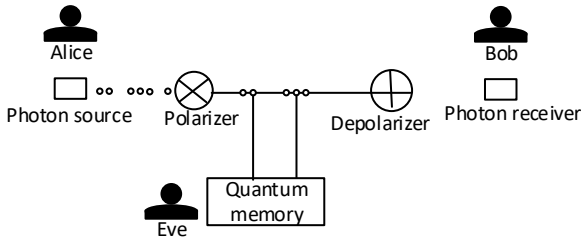

Fig. 1(ii). Eve splitting a photon from a multi-photon pulse and storing it in Quantum memory.

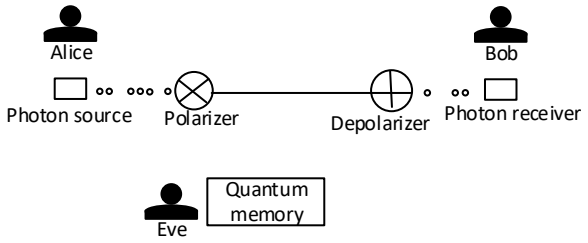

Fig. 1(iii). Eve sends rest of the photons forward.

When Alice and Bob communicate via the classical channel,sharing the basis they used for measurements, Eve measures the stored photon with the correct basis. This way, she gains knowledge of the key shared between Alice and Bob without being detected by them. She can then use the key to decrypt the information shared between Alice and Bob.

In the traditional sense, we constructed the notion that eavesdropping is a passive attack. However, here we can see that an attempt to eavesdrop alters the pulses thus making it an active attack.

This attack can be launched not only on the BB84 protocol but also on other protocols.For instance, B92, SARG04, and

Six State Protocol are also susceptible to the PNS attack [7]–[9]. In [10], [11], researchers have discussed the vulnerabilities of these protocols.

A. Characteristics of the Photon-Number Splitting attack

1) Dependence on exploitation of multi-photon pulses: The attack is possible since weak coherent sources that emit photons whose count follows a Poisson distribution are used. It means that some pulses will contain multiple photons while the rest are single photon pulses. The eavesdropper leverages this statistical feature to selectively target only multi-photon pulses, minimizing detection risk. The attack's success depends on how frequently multi-photon pulses occur in the system. If the probability of multi-photon events is high enough, the eavesdropper's attack becomes more effective. The attack will also be highly effective if the eavesdropper can figure out the pattern in which the sender is transmitting pulses to the receiver [3].

2) Interception without disturbance: The fundamental feature of a Photon-Number Splitting attack is that it can theoretically allow Eve to obtain information about the key without introducing detectable errors. This is because the eavesdropper's interception does not necessarily disturb the quantum states in a way that the photon loss or the disturbance is detected. Hence, it is an active form of eavesdropping.

3) Impact of the attack: A Photon-Number Splitting attack effectively reduces the secure key rate as Eve gains partial information about the transmitted key. The Quantum Key Distribution system must discard or adjust the final key length to account for this leakage, which can degrade the efficiency of the key generation process.

IV. Proposed detection method for Photon-Number Splitting attack

Quantum Bit Error Rate (QBER) is a critical metric in QKD systems that quantifies the error rate in the transmitted qubits between the two communicating parties. It is defined as the ratio of the number of erroneous bits to the total number of bits measured.

When there is a PNS attack, detection rate for single photon pulses is reduced at the receiver end because the eavesdropper selectively blocks single-photon pulses. Also, the ratio of multi-photon pulses to single-photon pulses that are detected at the receiver end increases significantly because less number of single-photon pulses reach the receiver end. If we correlate these changes with the subtle increase in QBER, we can detect with certainty the occurrence of the PNS attack. The method proposed for the detection of the attack is formalized in algorithm 1. The following assumptions have been made while defining it

- A real-world QKD network has been considered where the quantum channel is subject to noise.
- We model the quantum channel noise as inducing only bit-flip errors, and these errors constitute the QBER observed in the network.
- All the bases chosen by both parties match; so QBER is calculated over all the pulses.
- For demonstration, it has been assumed that the attacker will remove only 1 photon from a multi-photon pulse.
- The eavesdropper completely blocks single-photon pulses.
- We choose the total number of pulses to be large enough such that the ratio of the number of single-photon pulses to multi-photon pulses is approximately equal to the ratio of their probabilities as per the Poisson distribution.

Algorithm 1 PNSDetect Algorithm

START
$B \leftarrow$ randomBitString(N)
$nvals \leftarrow \{1, 2, 3, 4\}$
$QBER \leftarrow 0$
$p\_noise \leftarrow 0.05$
Pr $\leftarrow$ []
Ps $\leftarrow$ poisson_sampling(nvals, μ, N)
for $j = 0$ **to** $N - 1$ **do**
  Photon_s $\leftarrow$ createPulse$(B,$ Ps$, j)$
  Photon_r $\leftarrow$ PNSdemo(Photon_s)
  Photon_r $\leftarrow$ addSimpleNoise(Photon_r$, p_noise$)
  $nr \leftarrow$ LENGTH(Photon_r)
  Pr$[j] \leftarrow nr$
  Q$[j] \leftarrow$ computeQBER(Photon_r$, B$)
end for
Ratio_s $\leftarrow$ computeRatio(Ps)
Ratio_r $\leftarrow$ computeRatio(Pr)
$QBER \leftarrow$ AVG(Q)
if Ratio_s $\neq$ Ratio_r **and** $QBER > 0$ **then**
  RETURN "PNS Attack Detected"
end if
END

The algorithm begins by generating a random bit string B of length N, which simulates the key bits transmitted during the QKD process. A set of predefined photon number values, $nvals$, is used to model the photon distribution at the transmitter end based on Poisson sampling with a mean photon number μ.The parameter μ acts as a balancing factor between the key generation rate and the security of the QKD network. A higher value of μ will lead to a higher key generation rate but also makes the network more susceptible to the PNS attack as it leads to an increase in the probability of multi-photon pulses. On the other hand, a lower value of μ will decrease the probability of multi-photon pulses thus enhancing the security of the network but will also lead to a decrease in the key generation rate [15]. Pr and Ps are initialized to store the photon distributions at the receiver and transmitter, respectively. For each bit position in B, a quantum pulse $Photon_s$ is generated using the sampled photon distribution Ps. The pulse is simulated under potential attack conditions using a PNS demonstration function, $PNSdemo$. This produces a set of received photons $Photon_r$. Due to channel noise, some bits may be flipped in the received photons. The number

of received photons is indicated by nr. Simultaneously, the QBER for the current bit is computed by comparing the received photon values with the original bit $B[j]$, and the result is stored in $Q[j]$. The transmission photon ratio ($Ratio_s$) from Ps, the reception photon ratio ($Ratio_r$) from Pr, and the average QBER are calculated. Using these metrics, the algorithm evaluates whether the photon ratio at the receiver matches the photon ratio at the transmitter or not, combined with a non-zero QBER. If the ratios do not match and there is a non-zero QBER, it concludes that a PNS attack has occurred and outputs a warning message indicating the detection of the attack. The important utility functions used in algorithm 1 are defined in Algorithms 2-4.

Algorithm 2 createPulse Algorithm

Require: B: Bitstring, Ps: Poisson sampling distribution, j: Index
Ensure: $pulse$: Generated pulse
1: $pulse \leftarrow$ " "
2: **for** $i = 1$ **to** $Ps[j]$ **do**
3: $\quad pulse \leftarrow pulse + B[i]$
4: **end for**
5: **return** $pulse$

The createPulse algorithm is used to simulate the generation of multi-photon pulses. It takes the photon bit string, Poisson Sampling distribution parameters, and the index of the current photon, and creates a pulse with as many number of photons as required based on the value of Poisson sampling.

Algorithm 3 PNSDemo Function

Require: $Photon_s$
Ensure: $newlist$: A new list of strings after the PNS attack
1: $l \leftarrow$ LENGTH$(Photon_s)$
2: $newlist \leftarrow [\,]$
3: **if** $l = 1$ **then**
4: $\quad$ **return** $newlist$
5: **end if**
6: **if** $l > 1$ **then**
7: $\quad rmindex \leftarrow$ RANDOM$(0, l)$
8: $\quad$ REMOVE $Photon\_S[rmindex]$
9: **end if**
10: COPYWITHOUTNULLS$(newlist, Photon_s)$
11: **return** $newlist$ // remaining photons

The PNSDemo function simulates the PNS attack carried out on a pulse. In reality, the eavesdropper obtains information about the number of photons through Quantum Non-Demolition(QND) Measurements, which has been abstracted in the above algorithm.Then, the algorithm checks the number of photons in the pulse and then blocks it if it is a single-photon pulse. Otherwise, it removes a photon from the multi-photon pulse and sends the remaining photons to the receiver. The output list returned by this algorithm represents the distribution of photons after the PNS attack has been carried out by the eavesdropper, after removing the zeros. This is identical to the distribution of photons as seen at the receiver's end. The computeQBER function calculates the Quantum Bit

Algorithm 4 computeQBER Function

Require: $Photon_r$: List of strings where each element represents the string encoded by the photon as detected at the receiver.
B: The original bit string.
Ensure: Q: Quantum Bit Error Rate (QBER), a float value.
1: $newB \leftarrow$ " "
2: $incorrect \leftarrow 0$
3: $Q \leftarrow 0$
4: **for** $i \leftarrow 0$ **to** LENGTH$(Photon_r) - 1$ **do**
5: $\quad newB \leftarrow newB + Photon\_r[i]$
6: **end for**
7: **for** $k \leftarrow 0$ **to** LENGTH$(newB) - 1$ **do**
8: $\quad$ **if** $B[k] \neq newB[k]$ **then**
9: $\quad\quad incorrect \leftarrow incorrect + 1$
10: $\quad$ **end if**
11: **end for**
12: $Q \leftarrow incorrect/$LENGTH$(newB)$
13: **return** Q

Error Rate for a qubit, which is an important metric used to detect the PNS attack in PNSDetect algorithm. For the sake of greater accuracy, the QBER for the network has been reported as an average over N iterations. The PNSDetect Algorithm integrates all the algorithms 2-4 as part of its functionality to detect the PNS attack effectively.

V. Implementation and Results

We performed a code-level implementation of the PNS-Detect algorithm, which was consistent with our initial assumptions. For each simulation, we have considered 1000 pulses, and the probability of bit flips due to noise to be 5% .The results of our implementation for various values of mean photon number(μ) are presented in Table 1 below:

Table 1: Simulation Results for Different μ Values

μ	Attack Simulated	Ratio_s	Ratio_r	Avg. QBER	Attack Detected
0.1	False	31.2581	31.2581	0.043500	False
0.1	True	20.7391	14.3333	0.001000	True
0.3	False	6.5188	6.5188	0.052083	False
0.3	True	6.4627	8.5714	0.007500	True
0.5	False	3.2553	3.2553	0.047000	False
0.5	True	3.5872	5.8125	0.009833	True
0.7	False	2.0395	2.0395	0.050333	False
0.7	True	2.4602	3.5873	0.015167	True
1.0	False	1.4155	1.4155	0.043333	False
1.0	True	1.2173	2.0890	0.023833	True

The implementation accurately reflects our initial assumptions and validates the core detection logic of the PNSDetect algorithm across different values of the mean photon number μ The results obtained are consistent with the algorithm, as it accurately detects the attack when it has been intentionally simulated. In the attack scenario, the sender and receiver ratios

deviate significantly, and the measured QBER remains non-zero, both of which are valid indicators of the attack. In contrast, when no attack is simulated, the sender and receiver ratios are equal, there is minimal QBER due to noise, hence the algorithm does not detect any attack in this case. Moreover, the random nature of Poisson sampling can cause the experimental metrics to differ between repeated simulations. However, this variability does not affect the validity of the qualitative conclusions or the overall detection performance of the algorithm.

VI. Conclusion and future scope

The proposed approach is an algorithm specifically designed to detect the occurrence of the PNS attack in a basic Quantum Key Distribution network. Existing literature proposes algorithms to detect the PNS attack by augmenting the QKD network with features such as decoy states [12] [14], but these solutions are applicable even to basic QKD networks without any augmentation. In this work, we focus on proposing an efficient detection mechanism for the PNS attack. The proposed solution provides a focused and reliable method for PNS attack detection, while adhering to the stated assumptions. Future research could focus on extending the applicability of PNSDetect by considering adaptive mechanisms to distinguish between attack-induced and non-attack-induced variations in QBER. In future, we plan to emulate the PNS attack scenario on the nodes participating in quantum key distribution via various protocols to formulate an algorithm to mitigate the attack on that specific protocol. Observations can be made by detection of the attack using PNSDetect on different protocols and in varied conditions, comparing the results to produce conclusive outcomes.

References

[1] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. arXiv preprint arXiv:2003.06557, 2020.

[2] L. O. Mailloux, D. D. Hodson, M. R. Grimaila, R. D. Engle, C. V. McLaughlin, and G. B. Baumgartner, "Using Modeling and Simulation to Study Photon Number Splitting Attacks," *Air Force Institute of Technology*, Wright-Patterson AFB, OH, USA; *Naval Research Laboratory*, Washington, DC, USA; *Laboratory for Telecommunications Sciences*, College Park, MD, USA, 2023. Corresponding author: L. O. Mailloux (logan.mailloux@afit.edu).

[3] C. F. Sabottke, C. D. Richardson, P. M. Anisimov, U. Yurtsever, A. Lamas-Linares, and J. P. Dowling, "Thwarting the Photon Number Splitting Attack with Entanglement Enhanced BB84 Quantum Key Distribution," *Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Louisiana State University*, Baton Rouge, LA, 70803, USA, and MathSense Analytics, Altadena, CA, USA, 2023.

[4] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM review, 41(2):303-332, 1999.

[5] C. Gidney and M. Ekera. "How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits. Quantum", 5:433, 2021.

[6] B. Huttner, N. Imoto, N. Gisin, and T. Mor, "Quantum cryptography with coherent states," *Physical Review A*, vol. 51, no. 3, pp. 1863–1869, Mar. 1995.

[7] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," *Physical Review Letters*, vol. 68, no. 21, pp. 3121–3124, May 1992.

[8] D. Bruß, "Optimal eavesdropping in quantum cryptography with six states," *Physical Review Letters*, vol. 81, no. 14, pp. 3018–3021, Oct. 1998.

[9] V. Scarani, A. Acin, G. Ribordy, and N. Gisin, "Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations," *Physical Review Letters*, vol. 92, no. 5, pp. 057901, 2004.

[10] Xu, Guobin and Mao, Jianzhou and Sakk, Eric and Wang, Shuangbao Paul, "An Overview of Quantum-Safe Approaches: Quantum Key Distribution and Post-Quantum Cryptography", *2023 57th Annual Conference on Information Sciences and Systems (CISS)*, pp.1-6, doi: 10.1109/CISS56502.2023.10089619

[11] V. Prakash Rajendran and P. Deepalakshmi, "Mitigating Photon Number Splitting Attacks in Quantum Key Distribution: A Comprehensive Analysis of Security Vulnerabilities," *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India*, pp. 47-53, doi: 10.1109/ICESC60852.2024.10690045,2024.

[12] Lo H, Ma X and Chen K 2005,"Decoy state quantum key distribution", *Phys. Rev. Lett.* 94 230504

[13] Chunduru, Anilkumar and Lenka, Swathi and Neelima, N. and Veerappampalayam Easwaramoorthy, Sathishkumar, "A Secure Method of Communication Through BB84 Protocol in Quantum Key Distribution", *Scalable Computing: Practice and Experience*, vol. 25, doi: 10.12694/scpe.v25i1.2152 , 2024

[14] Logan O. Mailloux and Michael R. Grimaila and Douglas D. Hodson and Ryan D. Engle and Colin V. McLaughlin and Gerald B. Baumgartner, "Optimizing Decoy State Enabled Quantum Key Distribution Systems to Maximize Quantum Throughput and Detect Photon Number Splitting Attacks with High Confidence", arXiv eprint:1606.07313, 2016

[15] D. Pearson and C. Elliott, "On the optimal mean photon number for quantum cryptography," *quant-ph*/0403065, 2004.