Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

QKD-XGuard: Physics-Informed Conformal Open-Set Monitoring for Unseen and Adaptive Attacks in Decoy-State BB84

Wail Zita

wailzitazita@gmail.com

Istanbul Medipol University https://orcid.org/0009-0007-0744-7795

Method Article

Keywords: quantum key distribution, BB84, decoy-state QKD, open-set recognition, conformal prediction, anomaly detection, physics-informed machine learning, uncertainty quantification

Posted Date: September 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11078149/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

QKD-XGuard: Physics-Informed Conformal Open-Set Monitoring for Unseen and Adaptive Attacks in Decoy-State BB84

Wail Zita
Department of Computer Engineering, Istanbul Medipol University, Istanbul, Türkiye
wailzitazita@gmail.com

Abstract

Practical quantum key distribution (QKD) is secure only to the extent that a deployed implementation remains compatible with the physical and statistical assumptions used in its security analysis. Scalar quantum bit error rate (QBER) monitoring cannot, by itself, distinguish legitimate channel drift from all implementation-level attacks, and conventional supervised classifiers usually assume that every test-time attack family was represented during training. This work presents *QKD-XGuard*, a simulation-validated monitoring method for decoy-state BB84 that combines physics-informed residuals, temporal feature encoding, metric-structured latent representations, class-conditional conformal prediction, and explicit open-set rejection. The method models expected signal/decoy gains, error rates, single-photon behavior, key-rate consistency, and detector statistics; it then learns from deviations between measured and physically expected telemetry rather than from raw measurements alone.

A reproducible stochastic block-level simulator was used to generate 3,500 development-regime sessions spanning normal operation and four attack families, with 16 telemetry blocks per session. Entire attack families were excluded from training in four leave-one-family-out folds. Across these folds, the physics-residual model achieved known-class macro-F1 of 0.952 ± 0.007, unknown-attack AUROC of 0.945 ± 0.049, unknown recall of 0.876 ± 0.137, and empirical class-conditional conformal coverage of 0.910 ± 0.013 at nominal 90% coverage. Under a disjoint 75–115 km channel regime, macro-F1 remained 0.962 ± 0.004, whereas an otherwise identical raw-telemetry ablation fell to 0.796 ± 0.068. A deliberately stealth-oriented adaptive composite attack was rejected in $69.2\% \pm 5.7\%$ of sessions. These results support physics-residual modeling as a useful mechanism for separating channel variation from unsupported attack behavior, while also showing that open-set performance varies materially by attack family. QKD-XGuard is an operational monitoring layer; it does not replace composable QKD security proofs or finite-key analysis.

Keywords: quantum key distribution; BB84; decoy-state QKD; open-set recognition; conformal prediction; anomaly detection; physics-informed machine learning; uncertainty quantification.

1 Introduction

Quantum key distribution (QKD) enables cryptographic key establishment with security arguments grounded in quantum mechanics rather than computational hardness. BB84 established the prepare-and-measure paradigm [1], while subsequent security analyses treated ideal and imperfect devices, finite-key effects, and realistic decoy-state operation [2–4, 7, 9]. The practical security problem is nevertheless broader than the protocol equations: implementations can deviate from the model through detector mismatch, source imperfections, background noise, timing effects, or active manipulation [5, 6, 10, 16].

Machine learning has therefore been studied as a supplementary monitoring layer for QKD. Existing work has demonstrated multi-attack detection in continuous-variable QKD, automated identification of practical imperfections and attacks, and temporal QBER-based classification in BB84 [12–15, 18]. These results establish that richer telemetry can outperform a scalar threshold,

but they do not remove a structural limitation of ordinary supervised classification: a closed-set classifier assumes that test inputs belong to classes represented during model development.

Let the attack families represented at training time be

$$\mathcal{A}_{\text{train}} = \{A_1, \ldots, A_K\}. \tag{1}$$

Closed-set prediction assumes

$$Y_{\text{test}} \in \mathcal{A}_{\text{train}} \cup \{\mathcal{N}\}, \tag{2}$$

where $\mathcal{N}$ denotes normal operation. Operational monitoring instead requires the possibility

$$Y_{\text{test}} \in \mathcal{A}_{\text{train}} \cup \mathcal{A}_{\text{unknown}} \cup \{\mathcal{N}\}. \tag{3}$$

This is the open-set recognition problem [22]. The security consequence is important: a detector should be allowed to say "unsupported" rather than assign an unfamiliar physical pattern to the nearest known attack class.

QKD-XGuard addresses this problem through four coupled mechanisms. First, a decoy-state channel model generates expected telemetry from estimated operating conditions. Second, the learner receives normalized residuals between measured and expected quantities. Third, a latent metric representation is calibrated with class-conditional conformal prediction [21]. Fourth, the decision rule explicitly rejects observations for which the conformal set is empty or latent distance exceeds a held-out calibration threshold.

The principal contributions of this paper are:

- a physics-residual representation for signal/decoy gains, QBER, single-photon proxies, key-rate consistency, and detector-side telemetry;
- a temporal feature encoder and metric-structured latent space designed for open-set decisions rather than closed-set softmax confidence alone;
- a leakage-resistant leave-one-attack-family-out protocol in which the unknown family is excluded from model fitting, calibration, and rejection-threshold selection;
- a disjoint channel-regime test evaluating whether the detector follows physical inconsistency rather than memorized attenuation ranges;
- a reproducible simulation study with all reported tables generated by the accompanying Python script.

The claim made here is deliberately operational. QKD-XGuard does not modify the secrecy or correctness parameters of the underlying QKD protocol. It is a telemetry monitor intended to surface behavior that is poorly supported by the calibrated operating envelope.

2 Related Work and Research Gap

2.1 Practical QKD security

Weak coherent sources have nonzero multiphoton probability and motivate decoy-state parameter estimation [4]. Detector-efficiency mismatch can be exploited through timing manipulation [5], and bright-light experiments demonstrated detector-control vulnerabilities in commercial systems [6]. Measurement-device-independent QKD removes detector side channels at the protocol level [8], while modern reviews emphasize that device imperfections must be explicitly incorporated into security analysis [11, 16].

2.2 Machine learning for QKD monitoring

Neural methods have been investigated for multi-attack detection and attack discovery in QKD [12–14]. Xu *et al.* studied automated recognition of imperfections and attacks in practical QKD systems [15]. More recently, temporal QBER features were shown to provide strong multi-attack discrimination in BB84 under noisy and lossy conditions [18]. Deep-learning-assisted

eavesdropping has also been studied from the adversarial side [17]. These results motivate a monitor that is simultaneously physics-aware and cautious about unsupported inputs.

2.3 Broader cyber-resilience context

Operational security depends on more than the nominal cryptographic primitive. Prior work by Zita and Zita examined supply-chain risks affecting communication devices across hardware, firmware, and deployment layers [19], while Zita studied secure QR-based authentication using replay protection, short-lived cryptographic tokens, and role-based access enforcement [20]. Although these studies do not address QKD directly, they motivate the systems-security perspective adopted here: cryptographic assurances remain dependent on implementation conditions, operational controls, and surrounding trust boundaries. QKD-XGuard applies this perspective to quantum-communication telemetry by treating deviations from the calibrated physical operating envelope as operational security signals rather than as modifications of the underlying QKD proof.

2.4 Research gap

Three gaps motivate the present study. First, high closed-set accuracy does not imply reliable rejection of an attack family absent from training. Second, model confidence is not a statistical coverage statement. Third, raw telemetry can entangle attack identity with channel regime; a detector may appear accurate because attenuation, detector efficiency, or background yield differ across classes. QKD-XGuard addresses these issues through open-set folds, class-conditional calibration, and residualization against an explicit channel model.

3 Decoy-State Physical Model

For fiber length L and attenuation coefficient α in dB/km, the channel transmittance is

$$\eta_{\mathrm{ch}}(L) = 10^{-\alpha L/10}. \tag{4}$$

With detector efficiency η_d, the total transmission efficiency is $\eta = \eta_d \eta_{\mathrm{ch}}$. For a phase-randomized coherent pulse with mean photon number x, photon number follows

$$P_x(n) = e^{-x}\frac{x^n}{n!}. \tag{5}$$

Using background yield Y_0, the gain under the simplified channel model is

$$Q_x = 1 - (1 - Y_0)e^{-\eta x}. \tag{6}$$

Let $e_0 \approx 1/2$ be the random background error rate and e_d the intrinsic alignment/detection error. Then

$$E_x = \frac{e_0 Y_0 + e_d(1 - e^{-\eta x})}{Q_x}. \tag{7}$$

A single-photon yield proxy consistent with this model is

$$Y_1 = 1 - (1 - Y_0)(1 - \eta), \tag{8}$$

and the corresponding error proxy is

$$e_1 = \frac{e_0 Y_0 + e_d \eta}{Y_1}. \tag{9}$$

For signal intensity μ, the single-photon gain is approximated by

$$Q_1 = \mu e^{-\mu} Y_1. \tag{10}$$

The asymptotic key-rate proxy used by the simulator is

$$R_{\mathrm{SK}} = \frac{1}{2}\left[-Q_\mu f(E_\mu) h_2(E_\mu) + Q_1(1 - h_2(e_1))\right]_+, \tag{11}$$

where $f(E_\mu) = 1.16$ in the experiments and h_2 is binary entropy. Equation (11) is used only as simulator telemetry; it is not claimed as a substitute for the finite-key calculation of a deployed stack [7, 9].

4 Physics-Informed Residual Representation

Each telemetry block contains

$$\boldsymbol{m}_t = [Q_{\mu,t}, Q_{\nu,t}, E_{\mu,t}, E_{\nu,t}, Y_{1,t}, e_{1,t}, R_{\mathrm{SK},t}, B_t, D_t]^\top, \tag{12}$$

where B_t is a normalized detector-click imbalance and D_t is a double-click proxy. Let

$$\boldsymbol{\theta}_t = [L_t, \alpha_t, \eta_{d,t}, Y_{0,t}, e_{d,t}, \mu_t, \nu_t]^\top \tag{13}$$

be the estimated operating regime. Equations (6)–(11) define a model g that produces expected telemetry $\hat{\boldsymbol{m}}_t = g(\boldsymbol{\theta}_t)$.

The central input is the standardized physical residual

$$z_{t,j} = \frac{m_{t,j} - \hat{m}_{t,j}}{\hat{\sigma}_j(\boldsymbol{\theta}_t) + \epsilon}. \tag{14}$$

The denominator is channel dependent: lower gains at high loss naturally imply different observational scales than short low-loss links. This conditioning is the mechanism by which XGuard attempts to avoid mistaking legitimate regime change for malicious deviation.

For each 16-block window and each of the nine residual channels, five temporal summaries are calculated: mean, sample standard deviation, linear slope, maximum absolute value, and lag-1 autocorrelation. The resulting vector has 45 dimensions.

5 Threat Model and Simulated Attack Families

The proof-of-concept simulator contains normal operation and four named attack families. The parameterizations are not intended to reproduce every hardware implementation; they produce physically interpretable deviations that stress different telemetry coordinates.

Intercept-resend. A partial interception fraction ρ increases error rates toward the ideal full-interception disturbance of 1/4, while slightly reducing gain and key-rate consistency.

Photon-number-splitting-like behavior. The simulator creates signal/decoy gain inconsistency, reduces the single-photon yield proxy, and changes the key-rate residual while introducing only a small QBER perturbation.

Time-shift behavior. The dominant observable is detector imbalance, accompanied by weak QBER, double-click, and gain changes, consistent with the fact that efficiency-mismatch attacks need not create a conspicuous scalar QBER increase [5].

Detector-control behavior. The simulator perturbs click imbalance, double-click rate, signal/decoy gains, and key-rate consistency to represent violations of the assumed detector response model [6].

A fifth, *adaptive composite* family is reserved for stress testing rather than model fitting. It combines weak gain, error, single-photon, key-rate, and detector perturbations while deliberately keeping QBER moderate. It is not presented as a full physical implementation of a reinforcement-learning eavesdropper; it is a controlled test of distributed low-amplitude manipulation. A learned adversarial policy is left for future hardware-aware evaluation.

6 Latent Metric Model and Conformal Inference

Let $\boldsymbol{x} \in \mathbb{R}^{45}$ be the temporal residual vector. The implemented encoder is a compact multilayer network

$$45 \to 64 \to 32 \to 16, \tag{15}$$

with ReLU activations and dropout 0.1 after the first hidden layer. The 16-dimensional output $\boldsymbol{v} = f_\Theta(\boldsymbol{x})$ is classified among the supported classes. Training minimizes cross-entropy plus a small metric regularizer that contracts same-class samples toward batch centroids and penalizes insufficient separation between batch centroids.

After training, the class-k prototype is

$$\boldsymbol{c}_k = \frac{1}{N_k} \sum_{i:y_i=k} \boldsymbol{v}_i. \tag{16}$$

With shrinkage-regularized covariance $\widetilde{\Sigma}_k$, the nonconformity score is the Mahalanobis distance

$$a_k(\boldsymbol{x}) = d_k(\boldsymbol{v}) = (\boldsymbol{v} - \boldsymbol{c}_k)^\top \widetilde{\Sigma}_k^{-1} (\boldsymbol{v} - \boldsymbol{c}_k). \tag{17}$$

For a held-out class-conditional calibration set $\mathcal{C}_k$, the conformal p-value is

$$p_k(\boldsymbol{x}) = \frac{1 + \sum_{i \in \mathcal{C}_k} \mathbb{I}[a_i \geq a_k(\boldsymbol{x})]}{|\mathcal{C}_k| + 1}. \tag{18}$$

At significance $\alpha_c = 0.10$, the prediction set is

$$\Gamma_{0.10}(\boldsymbol{x}) = \{k : p_k(\boldsymbol{x}) > 0.10\}. \tag{19}$$

A separate distance gate is calibrated as the 99th percentile of the minimum calibration distance,

$$\tau_{\text{ood}} = Q_{0.99}\left(\min_k d_k(\boldsymbol{v}_i) : \boldsymbol{x}_i \in \mathcal{C} \right). \tag{20}$$

The operational rule is

$$D(\boldsymbol{x}) = \begin{cases} \arg\max_k p_k(\boldsymbol{x}), & \Gamma_{0.10}(\boldsymbol{x}) \neq \varnothing \ \wedge \ d_{\min} \leq \tau_{\text{ood}}, \\ \text{unknown}, & \Gamma_{0.10}(\boldsymbol{x}) = \varnothing \ \vee \ d_{\min} > \tau_{\text{ood}}. \end{cases} \tag{21}$$

Conformal coverage for supported classes should not be interpreted as a theorem about unseen-attack recall. The latter remains an empirical open-set property.

7 Experimental Methodology

7.1 Simulation ranges

Table 1 gives the operating envelope. Development sessions use 10–65 km links. Cross-regime testing uses 75–115 km, with no overlap in link distance.

Within each session, L, e_d, Y_0, detector efficiency, and optical intensities undergo small random drift. Measurement noise is weakly temporally correlated. This construction forces normal sessions to occupy a nontrivial stochastic envelope rather than a fixed point.

Table 1: Simulation parameters.

Parameter	Development regime	Cross regime
Fiber length L	$\mathcal{U}(10, 65)$ km	$\mathcal{U}(75, 115)$ km
Attenuation α	$\mathcal{N}(0.20, 0.008^2)$, clipped $[0.17, 0.23]$	same
Detector efficiency η_d	$\mathcal{U}(0.12, 0.25)$	same
Background yield Y_0	log-uniform $[10^{-7}, 10^{-5.25}]$	same
Misalignment e_d	$\mathcal{U}(0.006, 0.026)$	same
Signal intensity μ	$\mathcal{U}(0.45, 0.60)$	same
Decoy intensity ν	$\mathcal{U}(0.08, 0.15)$	same
Window length	16 blocks	16 blocks

7.2 Dataset and splitting

The development dataset contains 700 sessions per class for five classes (normal plus four conventional attack families), giving 3,500 sessions and 56,000 telemetry blocks. A separate cross-regime set contains 220 sessions per class (1,100 sessions). The adaptive stress set contains 260 sessions.

For each leave-one-family-out fold, one attack family is removed entirely from model fitting and calibration. For each supported class, 55% of sessions are used for training, 20% for conformal calibration, and 25% for known-class testing. Consequently, each fold contains 700 known-class test sessions and 175 sessions from the unseen family. The held-out attack is never used to select τ_{ood} or any network parameter.

7.3 Optimization

Inputs are standardized from the training partition only. The model is trained with AdamW (learning rate 2×10^{-3}, weight decay 2×10^{-4}), batch size 96, and 28 epochs. The implementation seed is 20260917. All values reported below are generated by the accompanying `xguard_simulation.py` script.

7.4 Baselines and ablations

The main ablation replaces physics residuals with raw telemetry while keeping the temporal summaries, network size, calibration procedure, and open-set rule unchanged. Additional open-set scores are obtained from Isolation Forest and one-class SVM trained on normal residual sessions only. A maximum-QBER score is included to illustrate the information available from a scalar error statistic. These baselines are evaluated on the same question: discriminate the held-out attack family from the set of supported known classes.

8 Results

8.1 Leave-one-attack-family-out open-set detection

Table 2 reports the four open-set folds. The method is strongest for intercept-resend and time-shift behavior. PNS-like behavior is the most difficult held-out family, with unknown recall 0.703 and AUROC 0.880, illustrating that open-set security performance cannot be summarized by a single best-case number.

Known-class balanced accuracy was 0.910 ± 0.013. The known false-reject rate was 0.090 ± 0.013, consistent with the deliberately conservative $\alpha_c = 0.10$ calibration setting. Empirical class-conditional coverage was 0.910 ± 0.013, close to the nominal 90% target.

Table 2: QKD-XGuard physics-residual results by held-out attack family.

Held-out family	Macro-F1	Unknown recall	AUROC	AUPR	Cross-F1
Intercept-resend	0.952	1.000	0.997	0.988	0.956
PNS-like	0.955	0.703	0.880	0.710	0.960
Time-shift	0.943	0.971	0.955	0.777	0.966
Detector-control	0.960	0.829	0.948	0.835	0.964
Mean ± SD	0.952 ± 0.007	0.876 ± 0.137	0.945 ± 0.049	0.828 ± 0.119	0.962 ± 0.004

8.2 Physics residuals versus raw telemetry

Table 3 shows that the main value of residualization appears under distribution shift. In the development regime, raw telemetry still provides substantial predictive signal. In the disjoint 75–115 km regime, however, raw-input macro-F1 falls by approximately 0.165, while the residual model remains stable.

Table 3: Physics-residual ablation averaged over four held-out-family folds.

Metric	Physics residuals	Raw telemetry
Known-class macro-F1	0.952 ± 0.007	0.894 ± 0.037
Unknown recall	0.876 ± 0.137	0.633 ± 0.402
Unknown AUROC	0.945 ± 0.049	0.837 ± 0.240
Unknown AUPR	0.828 ± 0.119	0.699 ± 0.354
Cross-regime macro-F1	0.962 ± 0.004	0.796 ± 0.068
Cross-regime false reject	0.073 ± 0.007	0.140 ± 0.032

The result supports the intended role of Equation (14): the learner sees deviations from what the channel model predicts, rather than being forced to learn absolute gain and error levels that vary systematically with distance.

8.3 Open-set baselines

Table 4 compares unknown-versus-supported-class ranking. Normal-only anomaly detectors identify deviation from normal operation, but this is not equivalent to distinguishing an unseen attack from already supported attacks. Their AUROC is therefore much lower in the open-set test.

Table 4: Unknown-family ranking baselines, mean ± SD over held-out families.

Method	AUROC	AUPR
Maximum-QBER score	0.530 ± 0.270	0.369 ± 0.344
Isolation Forest	0.588 ± 0.157	0.314 ± 0.146
One-class SVM	0.618 ± 0.238	0.341 ± 0.126
QKD-XGuard	$\mathbf{0.945 \pm 0.049}$	$\mathbf{0.828 \pm 0.119}$

Under the intentionally stealth-oriented simulation parameters, a literal 11% maximum-QBER alarm did not trigger on any generated session. This should be interpreted narrowly: it demonstrates that the simulated attacks were configured below that scalar alarm level, not that an 11% rule is universally ineffective.

8.4 Adaptive composite stress test

The adaptive composite family was not used for training or calibration. Across the four conventional hold-out models, the residual detector rejected 0.692 ± 0.057 of adaptive sessions. This result is materially below the strongest conventional unknown-family folds and therefore exposes a remaining weakness rather than supporting a claim of universal zero-day detection. The detector is useful against distributed low-amplitude manipulation, but a substantial fraction of such behavior remains within the calibrated support of known classes.

9 Discussion

The experiments support three conclusions. First, physics residuals are more robust than raw telemetry when channel distance moves outside the training range. This is expected because the residual representation factors out a large part of the predictable variation in gain and error statistics. Second, calibrated open-set rejection can detect attack families excluded from training, but performance is attack dependent; PNS-like behavior remains notably harder than intercept-resend. Third, conformal calibration and unseen-attack detection must remain conceptually separate. Near-nominal coverage on supported classes does not imply near-nominal unknown recall.

The strongest evidence in this study is the cross-regime result. Raw telemetry performs reasonably in-distribution yet degrades substantially when link distance changes. The residual model remains stable because the expected telemetry is recomputed from the current physical regime. This is precisely the failure mode that a physics-informed monitor is intended to address.

The study also clarifies what QKD-XGuard does *not* establish. An "unknown" label is an operational statement about model support, not proof that a novel cyberattack exists. Hardware faults, model mismatch, inaccurate channel-parameter estimation, or unmodeled legitimate conditions can also produce rejection. Likewise, a monitoring decision cannot upgrade the composable secrecy of a QKD protocol.

10 Limitations

This work is a simulation-validated method study rather than a hardware experiment. The analytical block simulator does not reproduce every optical, detector, timing, or finite-key phenomenon of a deployed system. The attack parameterizations are interpretable stress models, not calibrated replicas of particular commercial devices. External validity therefore requires experimental traces or a hardware-in-the-loop QKD testbed.

The proof-of-concept uses temporal summary features rather than a recurrent sequence model. This choice improves reproducibility and isolates the effect of physics residualization, but richer sequence encoders may improve adaptive-attack detection. The present adaptive family is a constrained composite perturbation, not a reinforcement-learning policy that observes and reacts to a deployed detector. Learned attacker–defender co-evolution is a natural next step, particularly in light of recent work on learning-assisted QKD attacks [17].

Finally, results are reported across the four leave-one-family-out folds from one deterministic experimental campaign. Although fold-to-fold variation is explicitly reported, future work should repeat the complete simulation and optimization pipeline over multiple independent seeds and hardware-informed parameter priors.

11 Reproducibility and Data Availability

All data used in this study are synthetic and are generated from the equations and parameter ranges specified in this manuscript. No private or human-subject data are used. The accompany-

ing file `xguard_simulation.py` generates the development, cross-regime, and adaptive datasets, trains the models, and writes the result tables to comma-separated-value files. The supplied result directory contains the exact numerical outputs used in Tables 2–4. The random seed is fixed to 20260917.

12 Conclusion

QKD-XGuard reframes QKD attack monitoring as an open-set, physics-conditioned inference problem. The implemented proof of concept learns temporal summaries of residuals between observed telemetry and an explicit decoy-state channel model, constructs metric-aware latent classes, applies class-conditional conformal calibration, and rejects unsupported observations. In leave-one-attack-family-out simulation, it achieved 0.945 ± 0.049 unknown-attack AUROC and 0.876 ± 0.137 unknown recall while retaining approximately nominal conformal coverage on supported classes. More importantly, its macro-F1 remained 0.962 ± 0.004 under a disjoint long-distance channel regime, whereas the raw-telemetry ablation fell to 0.796 ± 0.068.

These results justify further investigation of physics-residual open-set monitoring for QKD, but they do not justify a universal zero-day claim. Hardware-trace validation, finite-key integration, repeated-seed evaluation, and a genuinely adaptive learned adversary are required before deployment-oriented conclusions can be drawn.

Declarations

Ethics approval and consent to participate. Not applicable; this study uses only synthetic simulation data.
Consent for publication. Not applicable.
Data and code availability. The simulator, model implementation, and result files accompany this manuscript.

References

[1] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical Computer Science*, vol. 560, pp. 7–11, 2014, reprint of the original 1984 conference paper, doi: 10.1016/j.tcs.2014.05.025.

[2] P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," *Physical Review Letters*, vol. 85, no. 2, pp. 441–444, 2000, doi: 10.1103/PhysRevLett.85.441.

[3] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," *Quantum Information and Computation*, vol. 4, no. 5, pp. 325–360, 2004, doi: 10.26421/QIC4.5-1.

[4] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, "Practical decoy state for quantum key distribution," *Physical Review A*, vol. 72, 012326, 2005, doi: 10.1103/PhysRevA.72.012326.

[5] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," *Quantum Information and Computation*, vol. 7, no. 1–2, pp. 73–82, 2007, doi: 10.26421/QIC7.1-2-3.

[6] L. Lydersen *et al.*, "Hacking commercial quantum cryptography systems by tailored bright illumination," *Nature Photonics*, vol. 4, pp. 686–689, 2010, doi: 10.1038/nphoton.2010.214.

[7] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, "Tight finite-key analysis for quantum cryptography," *Nature Communications*, vol. 3, 634, 2012, doi: 10.1038/ncomms1631.

[8] H.-K. Lo, M. Curty, and B. Qi, "Measurement-device-independent quantum key distribution," *Physical Review Letters*, vol. 108, 130503, 2012, doi: 10.1103/PhysRevLett.108.130503.

[9] C. C. W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, "Concise security bounds for practical decoy-state quantum key distribution," *Physical Review A*, vol. 89, 022307, 2014, doi: 10.1103/PhysRevA.89.022307.

[10] E. Diamanti, H.-K. Lo, B. Qi, and Z. Yuan, "Practical challenges in quantum key distribution," *npj Quantum Information*, vol. 2, 16025, 2016, doi: 10.1038/npjqi.2016.25.

[11] S. Pirandola *et al.*, "Advances in quantum cryptography," *Advances in Optics and Photonics*, vol. 12, no. 4, pp. 1012–1236, 2020, doi: 10.1364/AOP.361502.

[12] H. Du and D. Huang, "Multi-attack detection: General defense strategy based on neural networks for CV-QKD," *Photonics*, vol. 9, no. 3, 177, 2022, doi: 10.3390/photonics9030177.

[13] H. Luo *et al.*, "Beyond universal attack detection for continuous-variable quantum key distribution via deep learning," *Physical Review A*, vol. 105, 042411, 2022, doi: 10.1103/PhysRevA.105.042411.

[14] C. Ding *et al.*, "Machine-learning-based detection for quantum hacking attacks on continuous-variable quantum-key-distribution systems," *Physical Review A*, vol. 107, 062422, 2023, doi: 10.1103/PhysRevA.107.062422.

[15] J. Xu *et al.*, "Automatically identifying imperfections and attacks in practical quantum key distribution systems via machine learning," *Science China Information Sciences*, vol. 67, no. 10, 202501, 2024, doi: 10.1007/s11432-023-3988-x.

[16] P. Andriolo, E. Vasquez, E. Agudelo, M. Riegler, M. Pivoluska, and G. Murta, "Quantum key distribution with imperfections: Recent advances in security proofs," *Brazilian Journal of Physics*, vol. 56, 157, 2026, doi: 10.1007/s13538-026-02062-2.

[17] T. Lejeune and F. Damanet, "Deep-learning-based continuous attacks on QKD protocols," *Quantum Information Processing*, vol. 25, no. 4, 114, 2026, doi: 10.1007/s11128-026-05133-0.

[18] Isha, D. Singh, D. Kumar, S. K. Pal, P. Hambarde, and A. Shukla, "Beyond the QBER threshold: A temporal QBER based machine learning framework for multi attack detection in BB84 QKD," arXiv:2608.04047, 2026, doi: 10.48550/arXiv.2608.04047.

[19] W. Zita and N. Zita, "Mitigating Supply Chain Risks in Communication Devices Across Conflict Zones," *TechRxiv*, preprint, posted 12 July 2025, doi: 10.36227/techrxiv.175234514.44631243/v1.

[20] W. Zita, "Secure QR Code Login System with Role-Based Access for Web Applications," *TechRxiv*, preprint, posted 26 August 2025, doi: 10.36227/techrxiv.175616741.16408346/v1.

[21] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, 2nd ed. Cham: Springer, 2022, doi: 10.1007/978-3-031-06649-8.

[22] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013, doi: 10.1109/TPAMI.2012.256.