DRL-based Progressive Recovery for Quantum-Key-Distribution Networks

MENGYAO LI[1], QIAOLUN ZHANG[1,*], ALBERTO GATTO[1], STEFANO BREGNI[1], GIACOMO VERTICALE[1], AND MASSIMO TORNATORE[1]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy
[*] qiaolun.zhang@mail.polimi.it

Compiled August 18, 2024

With progressive network recovery, operators restore network connectivity after massive failures along multiple stages, by identifying the optimal sequence of repair actions to maximize carried live traffic. Motivated by the initial deployments of quantum-key-distribution (QKD) over optical networks appearing in several locations worldwide, in this work we model and solve the progressive QKD network recovery (PQNR) problem in QKD networks to accelerate the recovery after failures. We formulate an Integer Linear Programming (ILP) model to optimize the achievable accumulative key rates during recovery for four different QKD network architectures, considering different capabilities of using trusted relay and optical bypass. Due to the computational limitations of the ILP model, we propose a deep reinforcement learning (DRL) algorithm based on a twin delayed deep deterministic policy gradients (TD3) framework to solve the PQNR problem for large-scale topologies. Simulation results show that our proposed algorithm approaches well compared to the optimal solution and outperforms several baseline algorithms. Moreover, using optical bypass jointly with trusted relay can improve the performance in terms of key rate by 14% and 18% compared to the cases where only optical bypass and only trusted relay are applied, respectively.

https://doi.org/10.1364/JOCN.526014

1. INTRODUCTION

Quantum key distribution (QKD) secures classical communications against quantum attacks by sharing secret keys between two remote parties with information-theoretic security guaranteed by the fundamentals of quantum physics [1, 2]. QKD was initially deployed over point-to-point links, and only recently some demonstrations of full-scale multi-point QKD connections have been reported [3, 4]. Meanwhile, several studies have investigated and demonstrated the coexistence of quantum signal and classical signal on the same fiber [5], to reduce the deployment cost of QKD networks. A QKD network consists of QKD nodes, QKD modules (transmitters or receivers), and QKD links [1]. These QKD networks are expected to secure very critical services, such as financial and military services. Hence, in case of massive failures (caused, e.g., by natural disasters, such as earthquakes, or by man-made attacks), the services of the QKD network must be recovered as soon as possible. While some existing literature on network resiliency of QKD networks considered resource re-allocation strategies [6, 7], the network recovery of QKD networks is still not investigated.

In this research, we address the problem of Progressive Network Recovery (PNR) in QKD networks, herein referred to as Progressive QKD Network Recovery (PQNR). The PNR problem arises when a network is subject to massive failures, and the network operator has to identify the optimal sequence of repair actions to maximize the performance during the recovery. The PNR problem has already been modeled and solved in traditional networks [8], while PQNR necessitates re-evaluating traditional PNR methodologies to accommodate the unique technologies of QKD networks such as *optical bypass* and *trusted relay*, two technologies that facilitate the key distribution among non-adjacent nodes.

As shown in Fig. 1, using optical bypass quantum channels can be established between non-adjacent nodes by bypassing intermediate nodes, without the need to deploy additional QKD modules, as the quantum signal propagates directly from the transmitter to the receiver end. The trusted-relay technology allows considering some selected intermediate nodes as trusted secure nodes, both from the point of view of cyberattacks and physical-side attacks. Thus, trusted nodes can be used to relay symmetric keys exchanged between two users by encrypting it through a one-time pad (OTP) scheme, i.e., by using additional

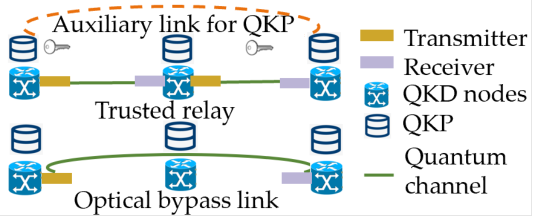
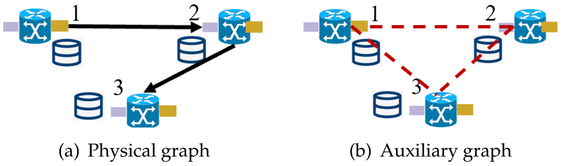


Fig. 1. Three enabled technologies for PQNR.

quantum keys exchanged on each link. This relay mechanism permits to pass through intermediate nodes without additional losses but requires further QKD modules in the intermediate node [1]. Moreover, PQNR can be accelerated in QKD networks thanks to the *quantum key pools (QKPs)*, which represent repositories of keys stored in selected QKD nodes [1], as illustrated in Fig. 1. A QKP allows storing the keys in one stage and consuming them in subsequent stages, which can significantly accelerate the satisfaction of requests. Furthermore, the presence of multiple quantum channels in each QKD link, facilitated by WDM, serves to enhance network resilience and robustness against failures.



(a) Physical graph (b) Auxiliary graph

Fig. 2. Example of auxiliary graph and physical graph

Our work is the first one, to the best of our knowledge, modeling and solving PQNR optimization in response to massive failures. Our study is based on the realistic modeling of the achievable key rates in QKD networks [9], considering both optical bypass and trusted relay. Note that, differently from classical (non-QKD) optical networks [10], the opportunity of caching keys in QKP during recovery and the presence of trusted relay and optical bypass significantly distinguishes the PQNR problem from the traditional PNR problem, as introduced before in [11]. In this paper, we propose a novel deep reinforcement learning (DRL)-based progressive recovery algorithm (DRL-PR) to address the PQNR problem. In addition, we add more results for different numbers of requests and add the results of two large-scale topologies to further evaluate the PQNR problem. The main strength of the DRL-PR algorithm is its versatile approach to sequence-making tasks due to its ability to learn complex policies through the data (i.e., samples), thanks to the interaction between the DRL algorithm and the QKD network environment. By leveraging DRL, networks can autonomously adapt to dynamic environments, and optimize their performance without relying on predefined rules. In our research, we use DRL to optimize recovery sequences across various network scenarios, including different failure scenarios, diverse topologies, and varied requests. Recovery sequences include which nodes and links need to be recovered for each stage and, consequently, the time required to repair the network, given the specific recovery resources. This flexibility enables networks to adapt to diverse conditions, ultimately enhancing performance and scalability.

The main contributions of this work are summarized as:

- We formulate an Integer Linear Program (ILP) model for the PQNR problem to identify the most effective recovery sequences of QKD network elements, which secures critical services as soon as possible.
- We develop a DRL-based algorithm (DRL-PR) to solve the PQNR problem for large topologies, significantly reducing the execution time with a low optimality gap.
- We evaluate the performance of our algorithm by Montecarlo simulation over various network topologies with and without optical bypass and trusted relay, varying the failure rate, acceptance ratio, number of requests. We generalize our model for different architectures, with various numbers of requests, failure rates, and topologies.

The rest of the paper is organized as follows. Sec. 2 discusses the related work with the QKD network and progressive recovery. Sec. 3 introduces the problem statement and the ILP model. Sec. 4 presents the devised DRL-PR algorithm. Sec. 5 discusses the numerical analysis. Sec. 6 concludes the paper.

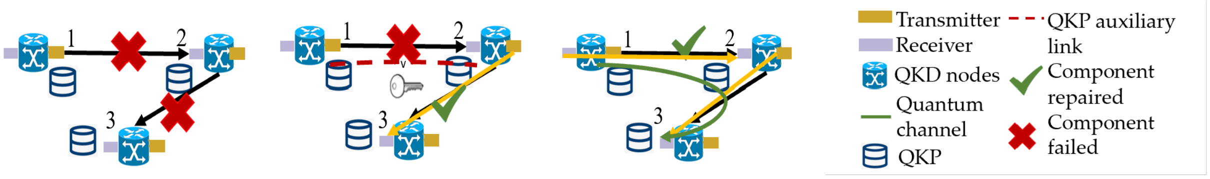
2. RELATED WORK

The integration of Quantum Key Distribution (QKD) technology in classical optical networks is essential to provide information-theoretically secure services[12, 13]. QKD networks have gained attention for securing communication infrastructures, with their inception dating back to the introduction of BB84 by Bennett and Brassard [14]. With the fundamental principles of quantum physics, QKD networks facilitate the sharing of symmetric keys that are information-theoretically secure [10]. QKD networks have been successfully deployed globally, e.g. in Switzerland, Italy, and China [1, 15, 16], and the possibility of the coexistence of QKD with standard optical signals has also been verified.

The architecture of a QKD network has three layers consisting of 1) an infrastructure layer, 2) a control and management layer, and 3) an application layer [1]. In a QKD network, a key manager coordinates the key distribution [17, 18] by controlling the QKD nodes and QKD modules and by leveraging capabilities of trusted relay and optical bypass. Trusted relays are prevalent in QKD testbeds and have been instrumental in extending the total transmission distance of QKD networks [1]. Chen et al. [19] extended the total transmission distance of the QKD network to 4,600 kilometers using trusted relays; to achieve multi-path transmission routing, a hybrid-trusted QKD network approach consisting of trusted nodes and semi-trusted nodes has been proposed [20]. Yu et al. [21] simulated the co-existence of trusted and untrusted relays, considering different conditions in terms of initial secret keys present in the quantum key pools (QKPs), and traffic load. Grillo et al. [22] used GEO and LEO satellite nodes as trusted relays, and proposed a centralized routing algorithm to select trusted relays to forward the secret keys between pairs of ground stations.

Regarding optical bypass, just a few works [23–26] have already considered it. Dong et al. [23] designed a novel quantum node structure that can achieve optical bypass. Auxiliary graphs are constructed to describe the adjacency of quantum nodes at different levels, influenced by the physical distance. Sun et al. [24] proposed experimental results about the introduction of a bypass structure scheme to improve the signal-to-noise ratio of QKD. Zhang et al. [25] theoretically analyzed the network-wide optimization for a QKD network with optical bypass and trusted relay. Yu et al. [26] addressed multi-dimensional routing, wavelength, and time slot allocation (RWTA) problems in short-distance quantum key distribution optical networks with optical bypass.

With the initial demonstrations of complex QKD networks, many works have started to consider advanced resource allo-



(a) Topology after failure. (b) Stage 1. (c) Stage 2.

Fig. 3. An example of how to use three technologies during PQNR.

cation problems, possibly also leveraging DRL. Cao et al. [17] presented a workflow, protocol extensions, and an on-demand secret-key resource allocation strategy for multi-tenant provisioning. Yu et al. [27] introduced quasi-real-time key provisioning (QRT-KP) to address the tradeoff between quantum key storage and the degree of security, and propose a multi-path based QRT-KP algorithm with QKPs. Chen et al. [28] proposed a new routing scheme for QKD based on application demand adaptation. The routing management center is set up based on the idea of software-defined networking, used to collect the network topology information and complete the path selection. Sharma et al. [29] proposed a DRL algorithm to provide an appropriate routing and the efficient utilization of network resources for the establishment of QKD light path requests. Reiß et al. [30] examined secret key rates of key distribution based on quantum repeaters in a broad parameter space of the communication distance with DRL.

Despite the progress in resource allocation and management for QKD networks, only few works have considered the resilience of QKD networks. Wang et al. [31] proposed three methods for reallocation of secret keys in working QKPs to recover failure-affected key provisioning services effectively. Tang et al. [32] have shown three practical SDN applications to improve the resilience of QKD-enabled microgrids under the control of SDN. Zhu et al. [33] presented protection or recovery of the key supply in the event of a QKD failure using the path protection scheme and re-routing restoration scheme. In our work, we investigate, for the first time, the problem of progressive recovery to restore the QKD network during multiple stages after a large-scale failure, considering different network architectures.

3. PROGRESSIVE QKD NETWORK RECOVERY PROBLEM AND ILP MODEL

In this section, we provide description for the PQNR problem, and an ILP model with constraints.

A. Problem Statement

We model the QKD network as a weighted undirected graph $G_p = (N_p, E_p)$, where N_p and E_p are the sets of physical nodes and links, respectively. Assume that the operators have a set of stages (defined as the amount of time to repair with a given amount of recovery resources) to recover the network. Recovery resources are, for instance, maintenance personnel or dedicated vehicles, which determine the number of nodes and links that can be repaired at each stage [8]. A quantum channel is where qubits are transmitted at different wavelengths, and a sequence of quantum channels allows to provision a quantum path. By reserving a sequence of quantum channels, a quantum path is provisioned. The key distribution between adjacent nodes may use either a physical quantum channel or a logical auxiliary link enabled by QKP. The capacity of QKP is elaborated in Subsection. 3. B. The key distribution between non-adjacent nodes can use the quantum channel with optical bypassing or the auxiliary link provided by QKP. To model the key distribution with QKP and quantum channel, we construct a fully-connected auxiliary graph (N_p, E_a) where each node pair is connected with an auxiliary link, which is enabled with keys stored in the QKP between the end nodes and the quantum channel by generating keys. The capacity of the auxiliary link is the key rate enabled by QKP, plus the key rate generated in the quantum channel. For instance, as shown in Fig. 2, we have a three-node topology and two physical links in the physical graph as in Fig. 2(a), while there are three links in the auxiliary graph as in Fig. 2(b), and keys can be distributed through these links.

The PQNR problem can be stated as: **given** a QKD network topology, a set of failed nodes[1] and failed links, the number of QKD modules per node, the number of quantum channels per link, the maximum number of recovery stages, the key rate of requests, **decide** the recovery sequence for nodes and links, and routing, channel, key-rate assignment for requests at each stage, **constrained** by the achievable key rates, the number of quantum channels, and the maximum number of quantum modules, with the **objective** to maximize the cumulative weighted key rates (CWKR) (defined as the sum of key rates of served requests at all stages, formula shown as 1, which is used to maximize the keys provided by this network during recovery. We solve the PQNR problem for four different network architectures depending on the availability of optical bypass and trusted relay as in Ref. [25]: (1) *OB-TR*: both trusted relay and optical bypass permitted; (2) *OB*: only optical bypass permitted; (3) *TR*: only trusted relay permitted; (4) *No-OB-No-TR*: optical bypass and trusted relay denied, only allow key distribution between adjacent nodes.

$$CWKR = \sum_{t}\sum_{d} \alpha_d \cdot k_d \cdot y_d^t \qquad (1)$$

α_d, k_d denotes the weight and the required key rate of request d, respectively. y_d^t shows if request d is accepted at stage t.

B. Achievable Key Rate and QKP Capacity

Table 1. Key rate for different reaches

Reach (km)	10	20	30	40	50
Key rate (kb/s)	23	13	7	3.5	1.9

Our model for estimating the achievable key rate is derived from Ref. [9]. With this model, we can calculate the maximum achievable key rate for different reaches reported in Table 1. It is worth noting that the key rates decrease by 11% for each passing node when using optical bypass.

The QKP capacity is dimensioned as follows. Assuming that AES256 encryption and Cipher Block Chaining (CBC) mode (a typical block cipher mode of operation using a block cipher algorithm) are adopted to encrypt the data, and that the same AES

[1] we assume that a QKP fails together with its hosting nodes.

Table 2. Sets used in the ILP model.

Set	Description
N_p	Set of QKD nodes in the network
E_p	Set of physical links in the network
E_a	Set of links in the fully-connected graph
E_f	Set of failed links in the network
N_f	Set of failed nodes in the network
P	Set of node pairs for paths in topology
W	Set of channels
T	Set of stages
Φ	Set of physical routes use the same end nodes as auxiliary link $e \in E_a$
D	Set of requests
$S^{+}(i)$	Set of outgoing links from node i
$S^{-}(i)$	Set of incoming links for node i

Table 3. Parameters used in the ILP model.

Param.	Description
O_n	Integer, number of QKD modules on node n
$h_{\phi,e}$	Binary, equals to 1 if link $e \in E_p$ in the route ϕ
k_d	Integer, the required key rate of request $d \in D$
l_ϕ	Integer, the key rate can supplied by route ϕ
α_d	Real, the weight of request $d \in D$
v_e	Integer, recovery resources needed by link $e \in E_p$
ω_n	Integer, recovery resources needed by node $n \in N_p$
δ	Integer, nodes recovery resources available at each stage
ϵ	Integer, links recovery resources available at each stage
θ	Integer, the number of second for one stage

key can be used to encrypt multiple messages, the number of messages that can be encrypted with the same key without violating semantic security is 2^{48} AES blocks, which equals 36000 Tb [34]. Considering a single-mode fiber containing 100 channels, each capable of transmitting data at 1 Tb/s [35], and with each channel encrypted using a distinct key, we deduce that the network must rotate keys (256 bits) approximately every 360 seconds to maintain the required security. Given that a stage spans one hour, the minimum QKP capacity required to encrypt data within a single stage is calculated to be 2560 bits.

Fig. 3 shows how to use QKPs during recovery on a simple network, consisting of 3 nodes (nodes 1, 2, 3) and 2 links (link (1,2), and (2,3)). Fig. 3(a) shows the topology after the failure of links (1,2) and (2,3). Let us assume one key-rate request (r_1 between nodes 1 and 3 requiring 20 kb/s) is present in the network before the failure, and that we can only recover one link at each stage. After the failure, r_1 has to be interrupted. At stage 1, as shown in Fig. 3(b), link (2,3) is repaired, while link (1,2) remains failed. Now r_1 can be served along two different links: on link (1,2), we use the keys already stored in QKPs, and on link (2,3), we can now establish a quantum channel to generate the required keys. At stage 2, link (1,2) is repaired. Now keys can be distributed on link (1,2), and link (2,3), as shown in Fig. 3(c), and r_1 can be served using *OB* or *TR*. In conclusion, this example shows that QKP can accelerate the recovery of a QKD network, and it must be incorporated into the recovery sequence.

C. Integer Linear Programming (ILP) model for PQNR

The sets, parameters, and variables for the ILP model are listed in Table 2, Table 3, and Table 4, respectively.
Objective function: maximize the cumulative weighted number of served key rates (CWKR), to maximize the keys provided by this network, considering the importance of requests.

Table 4. Variables used in the ILP model.

Var	Description
$f_{e,w}^{p,t}$	Binary, equals to 1 if quantum channel w on link $e \in E_a$ is allocated for path between node pair p at stage t
$x_{e,w}^{p,t,\phi}$	Binary, equals to 1 if route ϕ is selected for connection between the end nodes of link $e \in E_a$ in QKD path between node pair p at channel w at stage t
$p_{e,w}^{p,t}$	Binary, equals to 1 if link $e \in E_a$ used quantum channel in between node pair $p \in P$ on channel $w \in W$ at stage $t \in T$
$q_{e,w}^{p,t}$	Binary, equals to 1 if path p contains auxiliary link $e \in E_a$ based on QKP on channel w at stage t
$u_{p,w}^{t}$	Integer, key rate generated for path p on channel w at stage t
$z_{p,w}^{t}$	Binary, equals to 1 if QKD path between node pair $p \in P$ uses quantum channel $w \in W$ at stage t
B_i^t	Binary, equals to 1 if node $i \in N_p$ works at stage t
C_e^t	Binary equals to 1 if link $e \in E_a$ works at stage t
$B_{i(e)}^t$	Binary equals to 1 if end node $i \in N_p$ of link $e \in E_p$ works at stage t
g_p^t	Integer, stored keys in QKP for path between node pair p at stage t
y_d^t	Binary equals to 1 if request d is served at stage t
$\gamma_{e,w}^{p,t}$	Integer, key rate provided from QKP for link e in QKD path between node pair p in channel w
η_t	Integer, available recovery resources for links at stage t
β_t	Integer, available recovery resources for nodes at stage t

$$\max \sum_{t} \sum_{d} \alpha_d \cdot k_d \cdot y_d^t \tag{2}$$

We extend the ILP model for resource allocation in the QKD network in Ref. [25] to the PQNR problem.

C.1. Flow, link, key rate, and modules constraints

Eqn (3), (4) shows the flow constraint for QKD path, and it can use either a quantum channel or a QKP auxiliary link. Eqn (5) ensures that used modules on each node can not exceed the number of modules's upper bound. Eqn (6) determines the physical route $\phi \in \Phi$ between the end node pair of links $e \in E_a$. We use ϕ to represent the physical routing which tends to have the same end nodes as auxiliary link $e \in E_a$ in the auxiliary graph. Eqn (7) ensures that for a link e, quantum channel $w \in W$ of route $\phi \in \Phi$ can only be used once in stage $t \in T$. Eqn.(8) ensures that multiple QKD paths cannot use the same channel $w \in W$ in $e \in E_p$. Eqn (9) ensures the key rate of QKP path p should be less than the sum of the key rate provided by the quantum channel and QKP in each link of the path. Eqn (10) ensures keys can only be distributed if path p is available.

$$\sum_{e \in S^{+}(i)} f_{e,w}^{p,t} - \sum_{e \in S^{-}(i)} f_{e,w}^{p,t} = \begin{cases} z_{p,w}^{t} \; if \; i = a(p) \\ -z_{p,w}^{t} \; if \; i = b(p) \\ 0 \; others \end{cases} \tag{3}$$

$$\forall p \in P, i \in N_p, t \in T, w \in W$$

$$f_{e,w}^{p,t} = q_{e,w}^{p,t} \vee p_{e,w}^{p,t} \quad \forall e \in E_a, p \in P, t \in T, w \in W \tag{4}$$

$$\sum_{\substack{p \in P \\ e \in S^{+}(n) \\ w \in W}} p_{e,w}^{p,t} + \sum_{\substack{p \in P \\ e \in S^{-}(n) \\ w \in W}} p_{e,w}^{p,t} \leq O_n \quad \forall n \in N_p, t \in T \tag{5}$$

$$\sum_{\phi \in \Phi_e} x_{e,w}^{p,t,\phi} = p_{e,w}^{p,t} \quad \forall p \in P, e \in E_a, w \in W, t \in T \tag{6}$$

$$\sum_{p \in P} x_{e,w}^{p,t,\phi} \leq 1 \quad \forall e \in E_a, t \in T, w \in W, \phi \in \Phi_e \tag{7}$$

$$\sum_{p\in P}\sum_{e'\in E_a}\sum_{\phi\in\Phi_e} x_{e,w}^{p,t,\phi} * h_{\phi,e'} \leq 1 \quad \forall e\in E_p, w\in W, t\in T \tag{8}$$

$$u_{p,w}^{t} \leq \sum_{\phi\in\Phi_e}(x_{e,w}^{p,t,\phi}\cdot l_\phi) + \gamma_{e,w}^{p,t} + M\cdot(1-f_{e,w}^{p,t}) \qquad \forall e\in E_a, p\in P, w\in W, t\in T \tag{9}$$

$$u_{p,w}^{t} \leq M\cdot z_{p,w}^{t} \quad \forall p\in P, t\in T, w\in W \tag{10}$$

C.2. Recovery constraints

The QKD network has pre-allocated recovery resources for links and nodes for each stage. Eqns (11) and (12) ensure that the recovery resources consumed by nodes $i \in N_p$ and links $e \in E_p$ at stage t are less than the sum of remaining resources at stage $t-1$ and allocated resource δ and ϵ for stage t. Eqn (13), Eqn (14) and Eqn (15) ensure that we can use a route $\phi \in \Phi$ for a connection between two end nodes of link $e \in E_a$ only if all the traversed nodes and edges in route $\phi \in \Phi$ are working (i.e., not failed).

$$\beta_t \leq \beta_{t-1} + \delta - \sum_{i\in N_p}(B_i^t - B_i^{t-1})\cdot\omega_i \quad \forall t\in T \tag{11}$$

$$\eta_t \leq \eta_{t-1} + \epsilon - \sum_{e\in E_p}(C_e^t - C_e^{t-1})\cdot v_e \quad \forall t\in T \tag{12}$$

$$x_{e,w}^{p,t,\phi}\cdot h_{\phi,e'} \leq B_{a(e')}^{t} \ \forall e'\in E_p, \phi\in\Phi, e,p\in E_a, w\in W, t\in T \tag{13}$$

$$x_{e,w}^{p,t,\phi}\cdot h_{\phi,e'} \leq B_{b(e')}^{t} \forall e'\in E_p, \phi\in\Phi, e,p\in E_a, w\in W, t\in T \tag{14}$$

$$x_{e,w}^{p,t,\phi}\cdot h_{\phi,e'} \leq C_{e'}^{t} \ \ \forall e'\in E_p, \phi\in\Phi, e,p\in E_a, w\in W, t\in T \tag{15}$$

C.3. QKP storage constraints

Eqns (16) and (17) ensure that a QKP between node pair of path p can be used to distribute keys only when the end nodes of the p ($a(p) \in N_p$ and $b(p) \in N_p$) work at that stage t. Eqn (18) means that the stored keys in QKP should be not less than 0. Eqn (19) calculate the key rate at each stage. The keys in QKP at stage t are the residual keys at the last stage $t-1$ plus the key supplied by quantum channel, minus the keys used by other paths $p' \in E_a$ and the keys used by requests. Eqn (20) ensures that the variable $\gamma_{e,w}^{p,t}$ equals to 1 only when QKP is being used for path $p \in P$.

$$g_p^t \leq M\cdot B_{a(p)}^{t} \quad \forall p\in E_a, t\in T \tag{16}$$

$$g_p^t \leq M\cdot B_{b(p)}^{t} \quad \forall p\in E_a, t\in T \tag{17}$$

$$g_p^t \geq 0 \quad \forall p\in E_a, t\in T \tag{18}$$

$$g_p^t \leq g_p^{t-1} + (\sum_{w\in W} u_{p,w}^t - \sum_{p'\in E_a}\sum_{w\in W}(\gamma_{p,w}^{p',t} + \gamma_{\bar{p},w}^{p',t}) - k_p\cdot y_p^t)\cdot\theta \qquad \forall p\in P, t\in T \tag{19}$$

$$\gamma_{e,w}^{p,t} \leq M\cdot p_{e,w}^{p,t} \quad \forall p\in P, e\in E_a, t\in T, w\in W \tag{20}$$

4. DEEP REINFORCEMENT LEARNING FOR PROGRESSIVE RECOVERY (DRL-PR)

We present the proposed DRL-PR algorithm, which consists of two parts, namely, deciding the recovery sequence and a resource-efficient routing algorithm to serve requests.

A. TD3-based DRL Framework for Recovery Sequence

For our DRL-PR algorithm, we adopt twin delayed deep deterministic policy gradients (TD3) for recovery sequence, since TD3 has been shown effective in addressing the over-fitting and over-estimation issues for value functions [36]. In the following, we first define DRL elements for the PQNR problem. Then, a TD3-based DRL-PR algorithm with fully connected neural networks (FCNN) is presented for feature extraction, in which a mask mechanism is applied to reduce the candidate actions and speed up the training process.

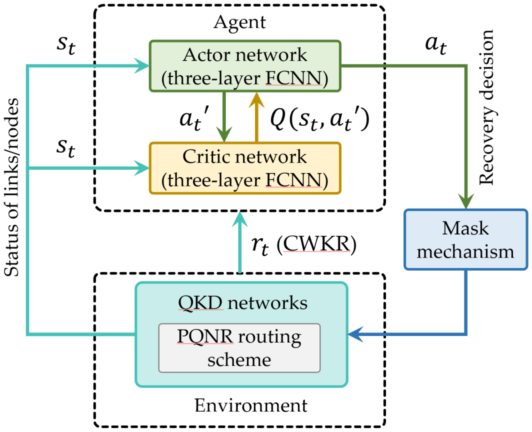


Fig. 4. TD3-based DRL-PR framework.

A.1. Definition of DRL Elements

There are three main components in a DRL framework, which are state, action, and reward, as shown in Fig. 4.

State s_t**:** The state represents the information that the DRL agent gathers from the network environment. In our work, the state represents the status (failed or not) of each component (links and nodes). In stage t, we flatten the total nodes and links into a binary vector, and each value in the vector denotes the status of one component, where "1" denotes non-failed components and "0" denotes failed components. Note that, in order to improve its generalization, we take a long vector whose length is larger than the number of nodes and links of the network as the state. As a consequence, the DRL-PR algorithm can generalize on a larger topology than the topology used to train it. The length of this binary vector is first filled according to the number of topology components, and the remaining slots are padded with "0".

Action a_t**:** The action space indicates the total recovery sequence, in which action a_t represents a recovery sequence at stage t. In this study, a mask mechanism is employed to narrow down the candidate action spaces, thus reducing training time. It is used as a way to tell sequence-processing layers that certain timesteps in some input should be skipped when processing the data. Once action a_t is determined, we update topology G_{t-1} according to the recovery action a_t, and then obtain G_t.

Reward r_t**:** The reward function r_t defines the real-time feedback provided to the DRL agent based on its action a_t, guiding it towards a learning optimal behavior in the given environment. CWKR values are adopted as the reward function r_t, with its equation specified in Eq. (1).

Algorithm 1. DRL-PR algorithm: recovery sequence

```
Input: $s_t$, episode, $N$, $T$
Output: $a_t$, $G_t$
1: Initialize critic primary networks $Q_{\theta_1}$, $Q_{\theta_2}$ and actor primary network $\pi_\phi$ with random parameters $\theta_1, \theta_2, \phi$.
2: Initialize two critic target networks $\theta_1 \leftarrow \theta'_1, \theta'_2 \leftarrow \theta_2$, and one actor target network $\phi' \leftarrow \phi$
3: Initialize replay buffer $\mathcal{B}$ with an empty set
4: for each episode do
5:     for $t = 1$ to $|T|$ do
6:         Select an action $a_t \sim \pi_\phi(s_t) + \epsilon, \epsilon \sim \mathcal{N}(0,1)$ and observe the reward CWKR $r_t$ and update topology to state $s_{t+1}$
7:         Store transition tuple $(s_t, a_t, r_t, s_{t+1})$ in $\mathcal{B}$
8:         Sample mini-batch of $N$ transitions $((s_t, a_t, r_t, s_{t+1})$ from $\mathcal{B}$
9:         $\widetilde{a} \leftarrow \pi_{\phi'}(s) + \epsilon', \epsilon' \sim$ clip $(\mathcal{N}(0,1), -c, c)$
10:        $y \leftarrow r + \gamma \min_{\theta_{i=1,2}} Q_{\theta'_i}(s', \widetilde{a})$
11:        Update critics $\theta_i \leftarrow \text{argmin}_{\theta_i} N^{-1} \sum (y - Q_{\theta_i}(s', a))^2$
12:        if $t$ mod $d = 0$ then
13:            Update $\phi$ by the deterministic policy gradient:
14:            $\nabla_\phi \mathcal{J}(\phi) = N^{-1} \sum \nabla_a Q_{\theta_1}(s,a)|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s)$
15:            Update target networks:
16:            $\theta'_i \leftarrow \tau\theta_i + (1-\tau)\theta'_i$
17:            $\phi' \leftarrow \tau\phi + (1-\tau)\phi'$
18: return $a_t$, $G_t$
```

Algorithm 2. DRL-PR Algorithm: Routing

```
Input: $N_p$, $E_p$, $G_t$, $T$, $D$, $k_d$, $J^p$.
Output: CWKR
1: Sort requests $d \in D$ according to their weight $\alpha_d$.
2: Find the upper bound of the number of modules $\theta_d$.
3: for $t \in T$. At stage t=0, input $J_d^p = 0$ do
4:     Repair according to the Algorithm 1, get $G_t$.
5:     for each quantum channel w = 1 to $|W|$ do
6:         for each unserved request $J_d^p < k_d$, $d = 1$ to $|D|$ do
7:             Get the shortest path $P_d$ and routing with the Dijkstra algorithm
8:             if Path $P_d$ for request $d$ routing from $d_s$ to $d_d$ exists. then
9:                 if the modules used by $P_d$ less or equal to $\theta_d$ then
10:                    Distribute keys $g_d^p$ from path $P_d$, and stored keys in QKP
11:                    Update quantum channel and QKD modules usage.
12:                    $J_d^p = J_d^p + g_d^p$
13:    for each quantum channel w = 1 to $|W|$. do
14:        for each unserved request $J_d^p < k_d$, $d = 1$ to $|D|$. do
15:            Get the shortest path $P_d$ and routing with the Dijkstra algorithm
16:            if Path $P_d$ for request $d$ routing from $d_s$ to $d_d$ exists. then
17:                Distribute keys and store them in QKP.
18:                Update quantum channel and QKD module usage.
19:                $J_d^p = J_d^p + g_d^p$
20:    for each served request $J_d^p > k_d$ do
21:         $CWKR = CWKR + \alpha_d * k_d$
22:        serve request $d$, let $J_d^p = J_d^p - k_d$
23: return CWKR
```

A.2. DRL-PR Algorithm: Recovery Sequence for PQNR

Following the DRL framework, we approximate the Q-value function using three-layer fully connected neural networks (FC-NNs) as a DRL agent. We use TD3 framework [36] with state, action, and rewards proposed by our PQNR problem. Alg. 1 has inputs: the state s_t, the number of episodes N, the training batch size d, a fixed number of updates T, the set of stages, the action a_t and the topologies for each stage G_t, respectively.

Firstly, we initialize two primary critic networks Q_{θ_1} and Q_{θ_2}, and one primary actor network π_ϕ, with random parameters θ_1, θ_2, and ϕ, respectively. Then, we initialize three corresponding target networks using the same parameters as the primary networks, and an empty buffer $\mathcal{B}$ (line 1-3). Subsequently, we introduce t and $|T|$, which are the current stage and maximum number of stages, respectively. To approximate expected actions, we inject a random normal distribution noise $\epsilon \sim \mathcal{N}(0,1)$ (line 4-6). Next, we take the interaction data (current state s_t, action a_t, reward r_t and next state s_{t+1}) as a sample, and then store it into the buffer $\mathcal{B}$. We randomly select some interaction samples with the number of N from $\mathcal{B}$ and add clipped noise $\epsilon' \sim \mathcal{N}(0,1)$ to the target action. These two critic primary networks both have Q_{θ_i}, and we use the smallest Q-value to calculate the target value y with the Bellman equation. We update the critic network by calculating θ_i with y (lines 7-11). The actor network is updated for each d stage, employing the deterministic policy gradient. Target networks are updated by the primary network. Finally, after looping through all the episodes, we obtain the model, action a_t, and topology G_t for each stage (lines 12-18).

B. DRL-PR Algorithm: Resource-Efficient Routing for PQNR

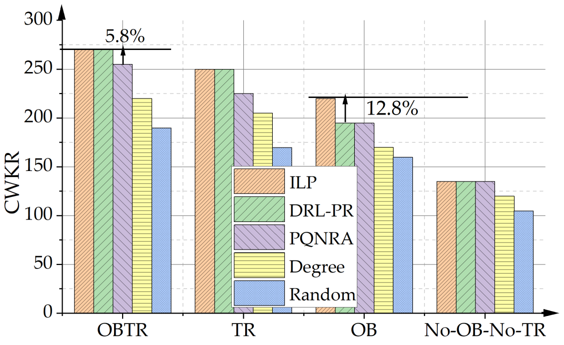


Fig. 5. CWKR for different architecture in German-7.

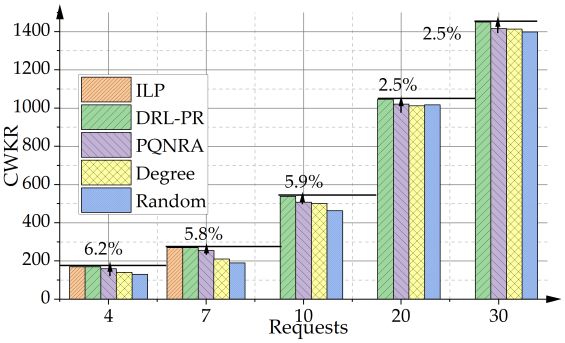
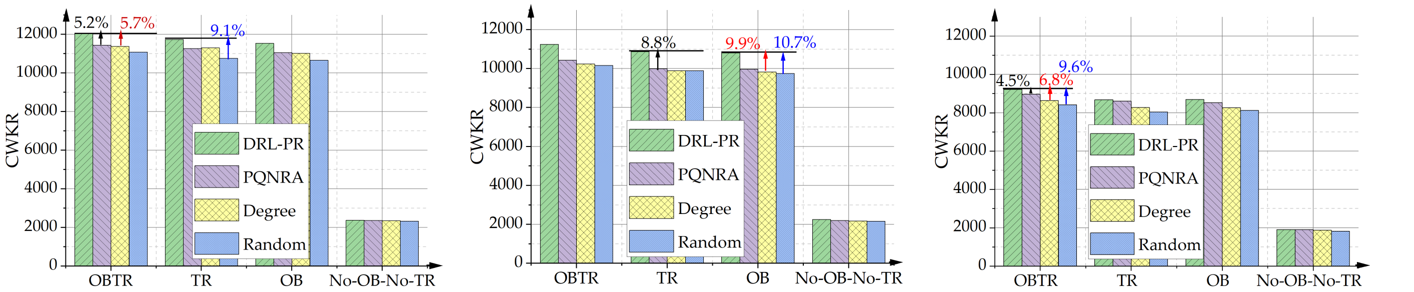


Fig. 6. CWKR vs. different number of requests in German-7.

To address the PQNR problem, we devised Alg. 2 for deciding the routing and key distribution. Alg. 1 outputs a topology G_t for each stage t, which only contains working nodes and links. J^p is an array whose elements contain the achievable key rate for each request d. θ_d is the upper bound of modules, which is decided by the number of modules cost by the shortest path for request d in TR according to the topology before the failure.

Alg. 2 is used to serve requests for all stages, which involves a resource-efficient routing of all requests. Initially, requests are sorted based on their importance, followed by topology repair guided by the output G_t from Alg. 1 (lines 1-4). The first phase focuses on serving requests using the shortest path with a



(a) 20% failure in NSF (b) 30% failure in NSF (c) 40% failure in NSF

Fig. 7. CWKR in NSF-14 topology

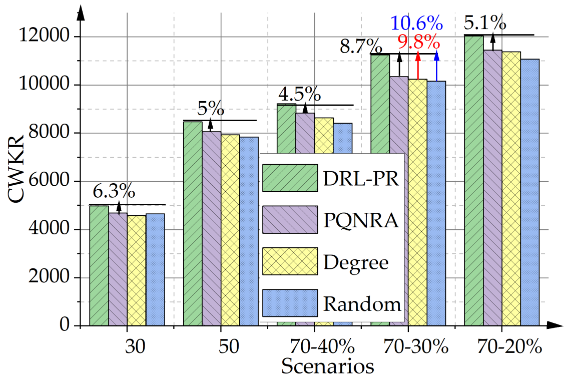


Fig. 8. CWKR vs. different number of requests in NSF-14.

limited number of QKD modules, denoted by the threshold θ_d (lines 6-12). Subsequently, unserved requests from the first phase are addressed greedily, without considering the QKD module threshold (lines 13-19). Finally, after finishing the routing for all the stages, Alg. 2 returns the CWKR (line 20-23).

5. NUMERICAL RESULTS

In this section, we first compare the performance of our proposed DRL-PR algorithm to other heuristic algorithms and ILP in a small topology. Then we evaluate the recovery performance on different network architectures in a large topology.

A. Simulation settings

We perform simulations on a machine with an Intel® Core™ i7-9700 CPU. We evaluate the performance of DRL-PR across four distinct architectures (*OB-TR*, *OB*, *TR*, *No-OB-No-TR*), and considering four topologies, namely, 7-node German topology (German-7), 14-node NSF topology (NSF-14), 17-node German topology (German-17), and 24-node USNET topology (USNET-24). The link lengths of the topologies have been scaled down to suit the current reach limitation of QKD network. Specifically, the link lengths are distributed as follows: German-7 topology: [10, 15] km; NSF-14 topology: [3, 48] km; German-17 topology: [4, 35] km; USNET-24 topology: [10, 40] km. These distributions ensure that the network configurations align with the specified QKD network constraints while accommodating the key rate requirements outlined in Table 1. Moreover, we consider different failure rates where the nodes and links in the network fail with a probability of 20%, 30%, and 40%. Due to current limitations in QKD technologies, the number of QKD modules was restricted, and it varies across different network topologies. Specifically, for the main results, the respective numbers are 4, 40, 50, and 70 for the German-7, NSF-14, German-17, and USNET-24 topologies. However, for Fig. 6 and Fig. 8, we employ different numbers of modules to accommodate the significant disparity in the number of requests.

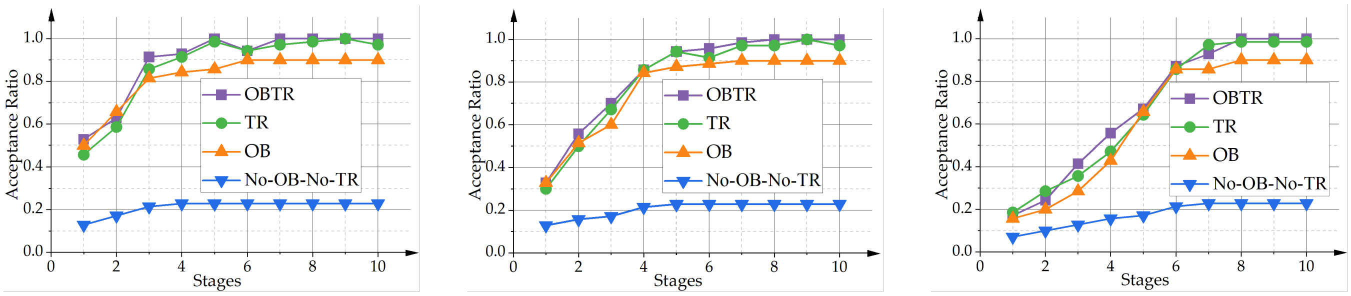
We train DRL-PR algorithm on NSF-14 topology with *OB-TR* architecture under 30% failure rate, and then generalize the model for different architectures, different topologies, different numbers of requests, and different failure rates. We use 2000 episodes for training and 20 episodes for testing. Note that we do not train the DRL-PR algorithm on the considered smallest topology, namely German-7 topology, as German-7 only has a number of nodes and links that are too small to have enough training data for DRL. We not only generalize the DRL model trained with NSF-14 on larger topologies (i.e., German-17 and USNET-24), but also on the smaller topology (i.e., German-7).

The baselines algorithm considered in this paper are *random* algorithm, *degree* algorithm, and the PQNRA algorithm in [11]. *Random* algorithm randomly selects nodes and links to repair. *Degree* algorithm repairs nodes/links with the largest nodal degree/link priority first (and in case of ties, it randomly selects nodes/links among them). PQNRA algorithm improves the *Degree* algorithm by enumerating all candidate recovery sequences in case of ties.

We first evaluate the performance of our proposed DRL-PR algorithm with heuristic algorithms and ILP on the German-7 topology. The key rates uniformly range from 10 to 15 kb/s for all the considered scenarios for German-7 topology. We then evaluate the performance of our proposed DRL-PR algorithm on NSF-14 topology and other larger topologies. For 80% of the requests across all topologies, the key rate uniformly ranges from 10 to 15 kb/s. However, to increase network variability, the remaining 20% of requests are generated with key rates ranging from 30 to 45 kb/s. This setting aims to simulate a diverse range of traffic demands and stress the network under varying conditions. All the results are averaged from 10 instances, except that the results compared to ILP use one instance due to the excessive computational time of ILP.

B. Results on German-7 topology

Let us start by commenting on the CWKR achieved by the different QKD network architectures for topology German-7. We use the result of this topology as a benchmark for comparison with ILP. In the figures, the black arrow means the advantage of ILP compared with DRL-PR or DRL-PR compared to PQNRA, the red arrow means the advantage of DRL-PR compared to *degree* algorithm, the blue arrow means the advantage of DRL-PR compared to *random* algorithm. In Fig. 5, we can see that CWKR of *OB-TR* is 8%, 27%, and 100% higher than *TR*, *OB*, and *No-OB-No-TR*, respectively in German topology. This shows how the utilization of trusted relay and optical bypass can significantly improve the recovery performance and that *TR* is more effective than *OB* when performing network recovery thanks



(a) Acceptance ratio for 20% failure

(b) Acceptance ratio for 30% failure

(c) Acceptance ratio for 40% failure

Fig. 9. Acceptance ratio in NSF-14 topology

to the keys stored in QKP. We then evaluate the optimality gap of our proposed DRL-PR algorithm and the advantages of the proposed DRL-PR algorithm compared to other baselines. Our DRL-PR algorithm can achieve the same value as ILP for most architectures, with only a 12.8% optimality gap observed for the *OB* architecture. However, despite this gap, our DRL-PR algorithm significantly reduces execution time, decreasing from over three hours to just 44.8 seconds. DRL-PR achieves up to 5.8% and 11.1% higher CWKR than PQNRA for *OB-TR* and *TR* architectures, respectively. Moreover, DRL-PR also outperforms other heuristic algorithms. Specifically, the CWKR of DRL-PR is 22.7% and 42% higher than the *degree* algorithm and *random* algorithm, respectively.

Moreover, we show that DRL-PR can generalize across scenarios with different number of requests , as shown in Fig. 6. We test scenarios with 4, 7, 10, 20, 30 requests for German-7 topology. Even when subjected to varying numbers of requests, the DRL-PR algorithm consistently outperforms other algorithms in terms of CWKR. For the 4, and 7 request scenarios, our DRL-PR can achieve the same value as ILP. Moreover, our DRL-PR can achieve a 2.5% to 6.2% improvement in CWKR compared to the PQNRA method.

C. Performance Evaluation on NSF-14 topology

The results of CWKR in the NSF-14 topology used during model training, are presented in Fig. 7. We perform a sensitivity analysis to examine the impact of varying failure rates on network recovery. Specifically, we consider scenarios where 20%, 30%, and 40% of the nodes and links in the network had failed, respectively. Note that we train the model with 30% of failure. We aim to assess the model's ability to generalize across different architectures (train for *OB-TR*, test *OB-TR*, *OB*, *TR*, and *No-OB-No-TR*), different failure rates, and different requests.

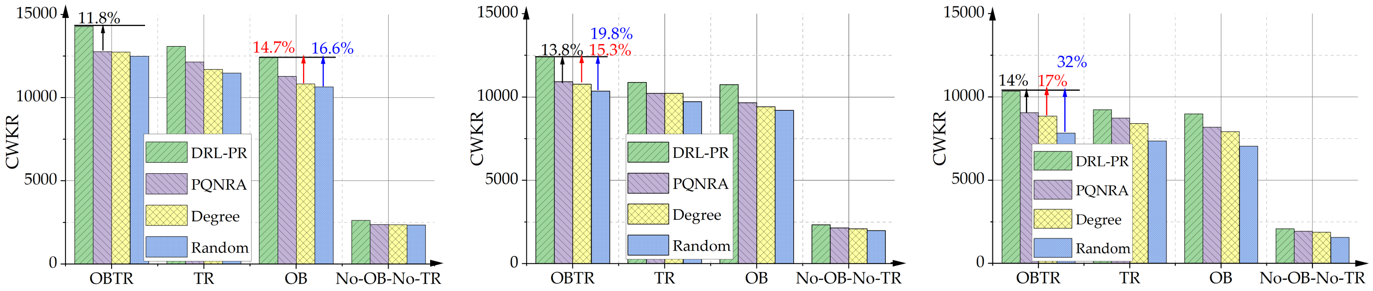
Let us start by commenting on the CWKR achieved by the different QKD network architectures for NSF-14 topology. The DRL-PR algorithm consistently outperforms PQNRA, the *degree* algorithm, and the *random* algorithm across all failure scenarios and architectures. In Fig. 7, we can see that CWKR of *OB-TR* is 6.3%, 6%, and 407% higher than *TR*, *OB*, and *No-OB-No-TR*, respectively in NSF-14 topology, demonstrating the significant improvement achieved through the combination of trusted relay and optical bypass. Let us then compare the CWKR of the proposed DRL-PR algorithm with PQNRA and two baseline algorithms (*degree* algorithm and *random* algorithm) on the NSF topology. In three subgraphs, the CWKR of DRL-PR can achieve 8.8%, 9.9%, and 10.7% higher than PQNRA, *degree* algorithm and *random* algorithm, respectively. Regarding the impact of failure rate, the performance of the DRL-PR algorithm for *OB-TR* architecture surpasses that of other architectures across all failure scenarios. The CWKR for the case of 20% failure (Figure 7(a)) is 8% and 35% higher than the cases of 30% failure (Figure 7(b)) and 40% failure(Figure 8(c), respectively. Notably, the advantage of DRL-PR compared to other algorithms is most significant at a 30% failure rate, that because we train our model on 30% failure scenarios.

Furthermore, our model can generalize across different request scenarios, as shown in Fig. 8. We test the performance with 30 and 50 requests under 30% of rejection rate together with 70 requests under different failure rates (i.e, 70-20%, 70-30% and 70-40% cases with 70 requests under 20%, 30%, and 40% of failure rate, respectively. As shown in Fig. 7, even when subjected to varying numbers of requests and failure rates, the DRL-PR algorithm consistently outperforms other algorithms in terms of CWKR. 70-30% still has the highest advantage compared to PQNRA, because we train with 70-30% scenario.

Finally, Fig. 9 shows the acceptance ratio (*AR*) at different stages across various architectures, indicating that the *OB-TR* architecture achieves an *AR* of up to 100%, while *OB*, *TR*, and *No-OB-No-TR* has 15.6%, 21.4%, and 337% lower AR during recovery, respectively. Notably, the *AR* can reach 100% for several stages with the DRL-PR algorithm, indicating its robust recovery performance. Conversely, architectures like *No-OB-No-TR* struggle to achieve high *AR* due to their inability to serve requests between non-adjacent nodes. The network topology achieves complete restoration at different stages depending on the failure scenario. Specifically, for 20% failure rate, all the nodes and links are recovered at stage 4, and it takes a larger number of stages to recover a network with a higher failure rate. For instance, under 40% failure rate, all the nodes and links are recovered at stage 8. Notably, in the case of a 40% failure rate scenario, the *AR* for stage 1 is notably lower compared to the 20% failure rate scenario, as expected. It is important to highlight that while the AR tends to increase when repairing new links or edges, it doesn't always increase with subsequent stages. This is because earlier stages may consume a significant number of keys in QKP, leaving fewer keys available for subsequent stages to serve requests. In summary, the DRL-PR algorithm demonstrates acceptable recovery performance across ten stages, with the AR consistently stabilizing at 100%. Conversely, the *No-OB-No-TR* architecture struggles to achieve an AR above around 20% due to its inability to serve requests between non-adjacent nodes.

D. Generalizability of Proposed DRL-PR Algorithm on Large Topologies and Different Failure Rates

Next, we generalize our analysis to larger topologies, specifically the German-17 and USNET-24 topologies, as depicted in Fig. 10 and Fig.11. Let us begin by comparing the CWKR for the German-17 topology. The advantage in CWKR of *OB-TR* over *TR* and *OB* increases from 6% to 14% and 6% to 18%, respectively, compared to the results for the NSF-14 topology. However, the

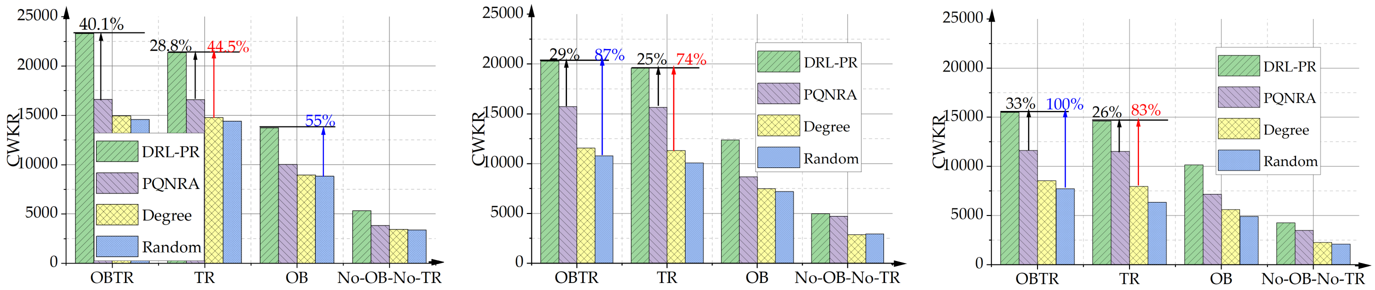


(a) 20% failure in German-17

(b) 30% failure in German-17

(c) 40% failure in German-17

Fig. 10. CWKR in German-17 topology



(a) 20% failure in USNET

(b) 30% failure in USNET

(c) 40% failure in USNET

Fig. 11. CWKR in USNET-24 topology

advantage of *OB-TR* compared to *No-OB-No-TR* decreases from 407% to 334%, depending on how much number of input requests are for the adjacent nodes. Moreover, for the German-17 topology, the DRL-PR algorithm achieves a significantly higher CWKR of 14%, 17%, and 32% compared to PQNRA, *degree* algorithm and *random*, respectively. Besides, the impact of the number of failures is depicted in Figure 10. The performance of DRL in the *OB-TR* architecture surpasses that of other architectures in all three failure scenarios. Now let us discuss the impacts of the failure rate. The CWKR for the case of 20% failure (Figure 10(a)) is 20% and 59% higher than the cases of 30% failure (Figure 10(b)) and 40% failure(Figure 10(c), respectively.

Finally, we evaluate the performance on the USNET-24 topology, as illustrated in Fig. 11. In comparison to the results for the NSF-14 topology, the advantage in CWKR of *OB-TR* over *TR* and *OB* increases from 6% to 21% and 6% to 81%, respectively. However, the advantage of *OB-TR* compared to *No-OB-No-TR* decreases from 407% to 236%. The DRL-PR algorithm achieves a significantly higher CWKR of 40.1%, 83%, and 100% compared to PQNRA, degree algorithm, and random algorithm, respectively. Specifically, the CWKR for the case of 20% failure (Figure 11(a)) is 42% and 85% higher than the cases of 30% failure (Figure 11(b)) and 40% failure (Figure 11(c), respectively. PQNRA costs more than 20 hours to get the result for USNET-24 topology, while DRL-PR costs 12 minutes to get one result, which means that our proposed DRL-PR is more scalable than PQNRA.

In conclusion, our DRL-PR model exhibits impressive generalization capabilities across different dimensions: topologies, architectures etc. Our DRL-PR algorithm consistently outperforms other algorithms, with *OB-TR* consistently delivering superior performance than other architectures. The CWKR experiences a notable decrease with increasing failure rates. As the size of the topology increases, the advantage of our DRL-PR algorithm compared to PQNRA increases. Specifically, the advantage of DRL-PR compared to PQNRA is up to 6.2%, 8.8%, 14,%, and 40.1% for German-7, NSF-14, German-17, and US-NET-24, respectively. These findings highlight the effectiveness and scalability of our DRL-PR approach in network optimization tasks.

6. CONCLUSION

We investigated the PQNR problem under four network architectures (with or without optical bypass, and with or without trusted nodes) using the ILP model and DRL-PR algorithm. Our analysis included comparisons with the PQNRA algorithm proposed in our previous paper [11] and two baseline algorithms. We found that the DRL-PR algorithm demonstrates robust generalization capabilities across different request scenarios, failure rate scenarios, architectures, and topologies. Specifically, DRL-PR results exhibited an optimality gap within 12.8% compared to ILP, while significantly reducing execution time from approximately 3 hours to 44 seconds over a 7-node topology. Notably, the DRL-PR algorithm consistently outperformed the other three algorithms considered in our study. Furthermore, our investigation highlighted the superior CWKR achieved by the *OB-TR* architecture compared to *TR*, *OB*, and *No-OB-No-TR* architectures. Additionally, we examined how varying percentages of failures impact performance, providing insights into the resilience of the proposed algorithms under different failure scenarios. Overall, our findings underscore the effectiveness and efficiency of the DRL-PR approach in addressing PQNR, offering promising avenues for future research and practical implementation in quantum network restoration tasks.

ACKNOWLEDGMENT

A preliminary version of this work was presented at Globecom 2023 conference (title “Progressive Quantum Key Distribution Network Recovery after Massive Failures”). This work was supported by the Italian Ministry of University and Research (MUR) and the European Union (EU) under the PON/REACT project and also partly funded by MUR *PON*2014_2020 "QUANCOM" Project (MIUR *ARS*01_00734) and the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").

REFERENCES

1. Y. Cao *et al.*, "The evolution of quantum key distribution networks: On the road to the qinternet," IEEE Commun. Surv. & Tutorials **24**, 839–894 (2022).
2. Q. Zhang, O. Ayoub, J. Wu, X. Lin, and M. Tornatore, "Ic-qkd: An information-centric quantum key distribution network," IEEE Commun. Mag. **61**, 148–154 (2023).
3. T.-Y. Chen, X. Jiang, S.-B. Tang, L. Zhou, X. Yuan, H. Zhou, J. Wang, Y. Liu, L.-K. Chen, W.-Y. Liu *et al.*, "Implementation of a 46-node quantum metropolitan area network," npj Quantum Inf. **7**, 134 (2021).
4. O. Alia, R. S. Tessinari, E. Hugues-Salas, G. T. Kanellos, R. Nejabati, and D. Simeonidou, "Dynamic dv-qkd networking in trusted-node-free software-defined optical networks," J. Light. Technol. **40**, 5816–5824 (2022).
5. K. Patel, J. Dynes, I. Choi, A. Sharpe, A. Dixon, Z. Yuan, R. Penty, and A. Shields, "Coexistence of high-bit-rate quantum key distribution and data on optical fiber," Phys. Rev. X **2**, 041010 (2012).
6. H. Wang *et al.*, "Resilient fiber-based quantum key distribution (qkd) networks with secret-key re-allocation strategy," in *OFC,* (IEEE, 2019), pp. 1–3.
7. J. Lv *et al.*, "Recovery scheme with resource abstraction in multi-domain quantum-key-distribution networks," in *2022 OECC,* (2022), pp. 1–3.
8. S. Ferdousi *et al.*, "Joint progressive network and datacenter recovery after large-scale disasters," IEEE T NETW SERV MAN **17**, 1501–1514 (2020).
9. H.-K. Lo *et al.*, "Decoy state quantum key distribution," Phys. Rev. Lett. **94**, 230504 (2005).
10. O. Amer *et al.*, "Efficient routing for quantum key distribution networks," in *2020 IEEE QCE,* (IEEE, 2020), pp. 137–147.
11. M. Li, Q. Zhang, A. Gatto, S. Bregni, Z. Yang, and M. Tornatore, "Progressive quantum key distribution network recovery after massive failures," in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference,* (2023), pp. 2148–2153.
12. Y. Cao, Y. Zhao, J. Wang, X. Yu, Z. Ma, and J. Zhang, "Sdqaas: Software defined networking for quantum key distribution as a service," Opt. express **27**, 6892–6909 (2019).
13. V. Zapatero, T. van Leent, R. Arnon-Friedman, W.-Z. Liu, Q. Zhang, H. Weinfurter, and M. Curty, "Advances in device-independent quantum key distribution," npj quantum information **9**, 10 (2023).
14. A. Gatto *et al.*, "A bb84 qkd field-trial in the turin metropolitan area," in *Photonics in Switching and Computing,* (2021), pp. Tu1A–2.
15. Q. Zhang, F. Xu, L. Li, N.-L. Liu, and J.-W. Pan, "Quantum information research in china," Quantum Sci. Technol. **4**, 040503 (2019).
16. W. Li, L. Zhang, H. Tan, Y. Lu, S.-K. Liao, J. Huang, H. Li, Z. Wang, H.-K. Mao, B. Yan *et al.*, "High-rate quantum key distribution exceeding 110 mb s–1," Nat. Photonics **17**, 416–421 (2023).
17. Y. Cao *et al.*, Y. Zhao, X. Yu, and J. Zhang, "Multi-tenant provisioning over software defined networking enabled metropolitan area quantum key distribution networks," JOSA B **36**, B31–B40 (2019).
18. M. Pereira, G. Currás-Lorenzo, Á. Navarrete, A. Mizutani, G. Kato, M. Curty, and K. Tamaki, "Modified bb84 quantum key distribution protocol robust to source imperfections," Phys. Rev. Res. **5**, 023065 (2023).
19. Y.-A. Chen *et al.*, "An integrated space-to-ground quantum communication network over 4,600 kilometres," Nature. **589**, 214–219 (2021).
20. L.-Q. Chen, J.-Q. Chen, Q.-Y. Chen, and Y.-L. Zhao, "A quantum key distribution routing scheme for hybrid-trusted qkd network system," Quantum Inf. Process. **22**, 75 (2023).
21. X. Yu *et al.*, "Secret-key provisioning with collaborative routing in partially-trusted-relay-based quantum-key-distribution-secured optical networks," IEEE J. Light. Technol. **40**, 3530–3545 (2022).
22. M. Grillo, A. A. Dowhuszko, M.-A. Khalighi, and J. Hämäläinen, "Resource allocation in a quantum key distribution network with leo and geo trusted-repeaters," in *2021 17th International Symposium on Wireless Communication Systems (ISWCS),* (2021), pp. 1–6.
23. K. Dong *et al.*, "Auxiliary graph based routing, wavelength, and time-slot assignment in metro quantum optical networks with a novel node structure," Opt. express **28**, 5936–5952 (2020).
24. W. Sun, L.-J. Wang, X.-X. Sun, Y. Mao, H.-L. Yin, B.-X. Wang, T.-Y. Chen, and J.-W. Pan, "Experimental integration of quantum key distribution and gigabit-capable passive optical network," J. Appl. Phys. **123** (2018).
25. Q. Zhang, O. Ayoub, A. Gatto, J. Wu, F. Musumeci, and M. Tornatore, "Routing, channel, key-rate, and time-slot assignment for qkd in optical networks," IEEE Transactions on Netw. Serv. Manag. **21**, 148–160 (2024).
26. X. Yu, X. Ning, Q. Zhu, J. Lv, Y. Zhao, H. Zhang, and J. Zhang, "Multi-dimensional routing, wavelength, and timeslot allocation (rwta) in quantum key distribution optical networks (qkd-on)," Appl. Sci. **11**, 348 (2020).
27. X. Yu, X. Liu, Y. Liu, A. Nag, X. Zou, Y. Zhao, and J. Zhang, "Multi-path-based quasi-real-time key provisioning in quantum-key-distribution enabled optical networks (qkd-on)," Opt. Express **29**, 21225–21239 (2021).
28. L.-Q. Chen, M.-N. Zhao, K.-L. Yu, T.-Y. Tu, Y.-L. Zhao, and Y.-C. Wang, "Ada-qkdn: A new quantum key distribution network routing scheme based on application demand adaptation," Quantum Inf. Process. **20**, 1–22 (2021).
29. P. Sharma, S. Gupta, V. Bhatia, and S. Prakash, "Deep reinforcement learning-based routing and resource assignment in quantum key distribution-secured optical networks," IET Quantum Commun. **4**, 136–145 (2023).
30. S. D. Reiß and P. van Loock, "Deep reinforcement learning for key distribution based on quantum repeaters," Phys. Rev. A **108**, 012406 (2023).
31. H. Wang, Y. Zhao, X. Yu, A. Nag, Z. Ma, J. Wang, L. Yan, and J. Zhang, "Resilient quantum key distribution (qkd)-integrated optical networks with secret-key recovery strategy," IEEE Access **7**, 60079–60090 (2019).
32. Z. Tang, P. Zhang, and W. O. Krawec, "Enabling resilient quantum-secured microgrids through software-defined networking," IEEE Transactions on Quantum Eng. **3**, 1–11 (2022).
33. Q. Zhu, Y. Zhao, X. Yu, and J. Zhang, "Collaborative resilience in hybrid quantum-classical networks," in *2022 IEEE 14th International Conference on Advanced Infocomm Technology (ICAIT),* (2022), pp. 116–119.
34. D. Boneh and V. Shoup, "A graduate course in applied cryptography," Draft. 0.5 (2020).
35. A. Kumar, "Design and modal analysis of optical fibers with multiple cores and multiple cladding fiber," in *IEEE ICMNWC,* (IEEE, 2022), pp. 1–5.
36. S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning,* (PMLR, 2018), pp. 1587–1596.