Routing, Channel, Key-Rate and Time-Slot Assignment for QKD in Optical Networks

Qiaolun Zhang, Omran Ayoub, Alberto Gatto, *Member, IEEE*, Jun Wu, *Senior Member, IEEE*, Francesco Musumeci, *Senior Member, IEEE*, and Massimo Tornatore, *Fellow, IEEE*

Abstract—**Quantum Key Distribution (QKD) is currently being explored as a solution to the threats posed to current cryptographic protocols by the evolution of quantum computers and algorithms. However, single-photon quantum signals used for QKD permit to achieve key rates strongly limited by link performance (e.g., loss and noise) and propagation distance, especially in multi-node QKD networks, making it necessary to design a scheme to efficiently and timely distribute keys to the various nodes. In this work, we introduce the new problem of joint Routing, Channel, Key-rate and Time-slot Assignment (RCKTA), which is addressed with four different network settings, i.e., allowing or not the use of optical bypass (OB) and trusted relay (TR). We first prove the NP-hardness of the RCKTA problem for all network settings and formulate it using a Mixed Integer Linear Programming (MILP) model that combines both quantum channels and quantum key pool (QKP) to provide an optimized solution in terms of number of accepted key rate requests and key storing rate. To deal with problem complexity, we also propose a heuristic algorithm based on an auxiliary graph, and show that it is able to obtain near-optimal solutions in polynomial time. Results show that allowing *OB* and *TR* achieves an acceptance ratio of 39% and 14% higher than that of *OB* and *TR*, respectively. Remarkably, these acceptance ratios are obtained with up to 46% less QKD modules (transceivers) compared to *TR* and only few (less than 1 per path) additional QKD modules than *OB*.**

Index Terms—**Quantum key distribution, quantum key pool, trusted relay, optical bypass, key rate.**

I. Introduction

Fifth-generation (5G) and beyond communication networks will distribute a large amount of private and sensitive data to support new applications (e.g., e-health applications) that need to be encrypted [1]. However, the rapid development of quantum computing technologies is threatening traditional cryptography [2]–[5], making data exchange over communication networks no longer secure against the attack of a large-scale quantum computer. To address this challenge, Quantum Key Distribution (QKD) schemes are being investigated and deployed in optical networks, in order to provide keys for the application layer, IP layer, or Optical Transport Network (OTN) layer [1], [6]. Since QKD is based on the transmission of single-photon states, this technology holds the potential to share Information-Theoretic Secure (ITS) symmetric keys, thanks to the fundamental principles of quantum physics [3]. Unlike classic bits, in fact, the no-cloning theorem prevents passive eavesdropping of the quantum signal, leading to an unconditionally secure information exchange, which is theoretically immune to any algorithmic cryptoanalysis [3], [7].

A QKD network consists of multiple QKD nodes, QKD links, and Quantum Key Pools (QKPs) [3]. A QKP is a repository, maintained in each QKD node, of the keys generated in QKD networks. Each node has several QKD modules, each of which can work either as a transmitter or as a receiver. Each link has multiple quantum channels where qubits are transmitted at different wavelengths. Each quantum channel requires a QKD module at its end nodes to transmit the quantum signal. The most limiting factor in the field deployment of QKD network is the low key rate, defined as the amount of secret bits distributed between two nodes per second. The low key rate derives from fiber attenuation, which strongly impacts the transmitted single-photon states, reducing their number at the receiver, and hence, the achievable key rate.

Since no optical manipulation is permitted at the intermediate nodes, the quantum information exchange process is intrinsically limited to point-to-point connections between adjacent nodes, which is a significant limitation when secret keys need to be shared in a real network scenario. To enable the sharing of quantum keys in such cases, as shown in the physical topology of Fig. 1, two practical approaches exist[1]: (1) Using a *trusted relay*, i.e., an intermediate and uncompromised node, which is *trusted* to relay the keys between two other nodes. The main limitation of this approach is that it is resource-consuming as it requires 2 QKD modules for every intermediate node in a path of a connection. In addition, trusted relay fails if the intermediate node is compromised. (2) Adopting an *optical bypass*, which allows establishing a quantum channel between non-adjacent nodes bypassing any intermediate nodes in the optical domain (N.B. the optical bypass node does not require any QKD module in the intermediate node). The main limitation of the optical bypass approach is that it introduces an

This journal paper is an extension of the conference version, which has been accepted by Globecom 2022. This work was supported by the Italian Ministry of University and Research (MUR) and the European Union (EU) under the PON/REACT project. It is also partly funded by the National Natural Science Foundation of China under Grants U21B2019 and 61972255 and MUR PON2014_2020 "QUANCOM" Project (MIUR ARS01_00734). Moreover, this work was sponsored by Program of Shanghai Academic Young Research Leader under grant 20XD1422000.

Qiaolun Zhang, Alberto Gatto, Francesco Musumeci, and Massimo Tornatore are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milano, Italy. (email: name.surname@polimi.it)

Omran Ayoub is with the Department of Innovative Technologies, University of Applied Sciences of Southern Switzerland, 6928 Lugano, Switzerland. (email: omran.ayoub@supsi.ch)

Jun Wu is with the Graduate School of Information, Production and Systems, Waseda University, Fukuoka 808-0135, Japan. (email: jun.wu@ieee.org)

[1] Note that another scheme using quantum repeater exists, which creates entanglement to enable key transmission over long distances [8]. However, it is not considered in this paper since its field trials are still not available.

additional loss, thus lowering the key rate and making optical bypass potentially not applicable over long paths.

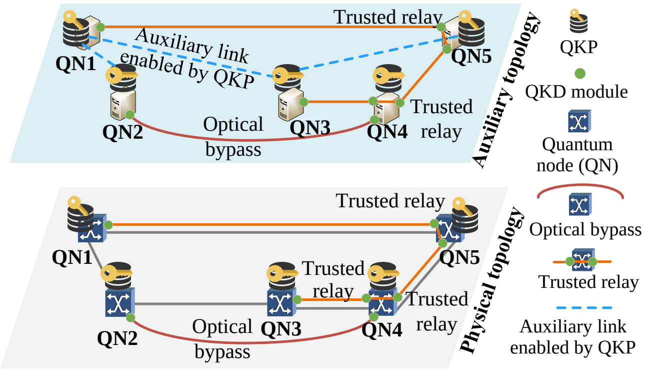


Fig. 1: Overview of key distribution in QKD network.

Since QKD networks have low key rates and both trusted relay and optical bypass have limitations, it is crucial to design a Key Management Scheme (KMS) for efficient key generation, distribution, and usage in QKD network jointly considering trusted relay and optical bypass. QKP is one of the most important mechanisms to effectively manage the keys generated in QKD network. In fact, if we consider dynamically-evolving key rate requests between pairs of quantum nodes, it may occur that some of the generated keys are not immediately used [3], particularly in low-load periods. In these periods, unused keys can be stored in the QKP for later use, i.e., for future key demands. Past works for time-scheduled QKD networks mainly consider storing the keys in the QKP for encrypting and securing data for a given node pair [9], while how to use the stored keys in QKP to share keys for other node pairs has not been systematically defined and discussed in the literature. For instance, as shown in the auxiliary topology of Fig. 1, QKD network can share keys between non-adjacent nodes with a path consisting of an auxiliary link enabled by QKP. In addition, we aim to maximize the key storing rate, defined as the rate of storing keys in the QKP for future requests. Moreover, to the best of our knowledge, there is no existing work considering the joint use of optical bypass and trusted relays for QKD networking.

The main novelties of this work are as follows.

- We define the novel Routing, Channel, Key-rate and Time-slot Assignment (RCKTA) problem to achieve resource-efficient QKD networking with QKP. Specifically, we classify different network settings with optical bypass and trusted relay and prove that for all network settings, the RCKTA problem is NP-hard.
- We formulate a Mixed Integer Linear Programming (MILP) model for all network settings of the RCKTA problem that incorporates the limitations of the secret key rate. Our formulation accounts for the possibility to build QKD paths with not only quantum channels, but also virtual links with QKP.
- We propose a scalable and near-optimal heuristic algorithm for the RCKTA problem, which reduces the execution time of establishing QKD path significantly. Numerical results show that allowing both trusted relay and optical bypass can achieve the highest acceptance ratio, paying off few QKD modules.

The rest of this paper is organized as follows. Section II discusses related work on QKD network. Section III formally states the RCKTA problem, and proves its NP-hardness. Section IV formulates a general MILP model that is applicable to all network settings of the RCKTA problem. In Section V we propose a resource-efficient heuristic algorithm to address the scalability issue of the MILP model. Section VI discusses numerical results obtained for all the proposed network settings and Section VII concludes the paper.

II. Background and Related Work

In this section, we first discuss the enabling technologies for QKD networks. Then, we briefly discuss the resource allocation of QKD networks, motivating the need for a novel resource-efficient QKD networking approach.

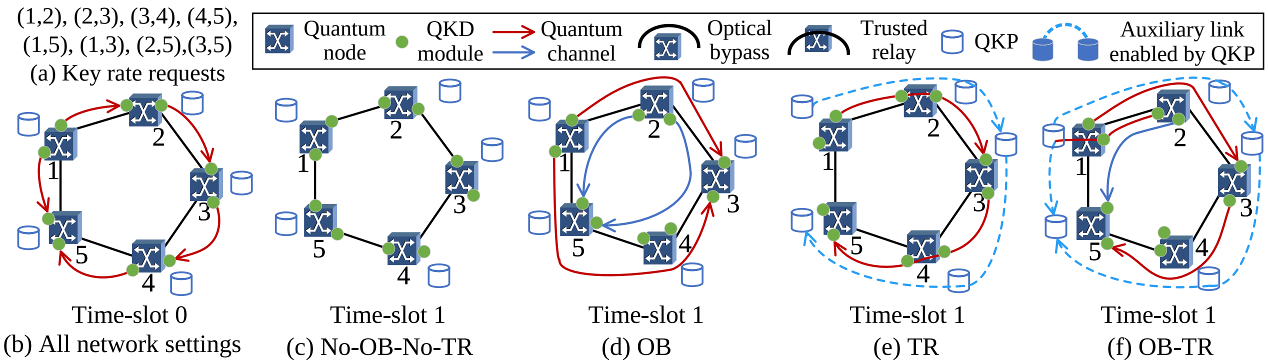
Enabling technologies for QKD networks are developing rapidly, including advances for both point-to-point QKD connectivity as well as for QKD connectivity between non-adjacent nodes. Regarding point-to-point QKD, a state-of-the-art fully connected QKD network has been demonstrated to support a key rate of 49.5 kb/s at a maximum distance of 18 km [10]. Besides, new cutting-edge technologies can support physical links of over 4 600 km [4]. Ref. [5] demonstrates the coexistence of QKD transmission and classical transmissions in already deployed WDM networks, which would enable much more cost-effective QKD deployments. To extend QKD connectivity to non-adjacent nodes, three schemes have been investigated, namely, trusted relay [11], optical bypass [9], and quantum repeater [8].

Trusted relays can be adopted to extend the key transmission distance [11] to tackle the issue of decreasing key rate due to quantum signal degradation over long distances. Ref. [12] demonstrates a QKD system integrated with a commercial-grade encrypted DWDM system. To achieve high scalability, in [11], a novel routing scheme is proposed for quality-of-service provisioning by minimizing the consumption of cryptographic keys. As an alternative to trusted relays, approaches such as device-independent QKD (DI-QKD) can also relay keys using untrusted nodes [13], but these realizations are either not mature or not available in practice [1]. Thus, the approaches using untrusted relays are not considered in this work.

Optical Bypass: Optical bypass by means of optical switches has also been validated in QKD networks [9]. Specifically, using optical bypass, QKD transmission can be established between non-adjacent nodes with a switching time of a few milliseconds [14].

Quantum Repeater: Quantum repeater is not yet a mature technology for large-scale deployment of quantum networks, and hence quantum repeater is not considered as its field trial is still not available.

Novel techniques for resource allocation in a QKD network have been also investigated. Ref. [3] proposes a layered QKD network and utilizes a KMS layer for efficient resource management. In [9], the authors solve a routing, wavelength, and time-slot assignment (RWTA) problem to store keys in



Time-slot 0
(b) All network settings

Time-slot 1
(c) No-OB-No-TR

Time-slot 1
(d) OB

Time-slot 1
(e) TR

Time-slot 1
(f) OB-TR

Fig. 2: Example of how requests are served under different settings of trusted relay and optical bypass.

the QKP. However, these works do not consider physical layer restrictions (e.g., limitations of QKD modules) neither it considers distributing keys with the cached keys in QKP. In our previous work [15], we define a novel problem of joint routing, channel, and key-rate assignment to achieve resource-efficient QKD networking. However, it does not model a realistic achievable key rate, or time-slot division multiplexing, and does not contain a scalable heuristic algorithm. Other works focused on overcoming the capacity limitations of QKD networks. Ref. [16] proposes to combine Multi-Core Fibers (MCFs) and Space Division Multiplexing (SDM) techniques to increase the secret key rate. Due to the cost of QKD deployment, Ref. [17] proposes solutions for key management in networks with partial QKD deployment.

Different from previous work, in this paper we systematically compare different network settings exploiting trusted relays and/or optical bypass. Based on the proposed scheme, we define a novel RCKTA problem for resource-efficient QKD networking. Moreover, for the first time, we model the achievable secret key rate with trusted relay and optical bypass and incorporate it into MILP for different network settings. Another significant achievement of our work is that we utilize keys in the QKP to establish a QKD path, which improves the efficiency of sharing keys.

III. Routing, Channel, Key-rate and Time-slot Assignment Problem

A. Problem Statement

The RCKTA problem for QKD networking can be stated as follows: **given** a QKD network topology, keys stored in QKP, a set of key rate requests during a given time period, and a set of time-slots constituting the period, **decide** the routing and channel, and key rate assignment for each key request and how each request uses trusted-relay and/or optical-bypass at each time-slot, **constrained to** a maximum number of QKD modules per node, amount of keys stored in QKP, key rate capability of quantum channels, and quantum channel uniqueness, with the **objective** of maximizing the number of served requests and the key storing rate. The secret keys stored in the QKP are managed in a pairwise fashion between any two nodes. For instance, if node 1 distributes keys to node 2, nodes 1 and 2 both maintain a copy of keys in their QKP.

Our proposed solution to the RCKTA problem could be applied in a real QKD network through a centralized Software-Defined Networking (SDN) controller, which is the current state-of-the-art technology to control QKD networks [3], [18]. Before serving requests, the SDN controller gathers information about the set of key rate requests and the amount of keys stored in the QKP from the network to solve the RCKTA problem. After solving the RCKTA problem, the SDN controller controls the quantum nodes to distribute keys along the determined QKD path from the solution of the RCKTA problem. Specifically, to generate keys in quantum channels, the SDN controller propagates the information about quantum channel usages (generation rates and the bypassing decisions) obtained from the RCKTA problem to quantum nodes. Moreover, to distribute keys for requests, the SDN controller propagates the key relay decisions in each node for each QKD path.

TABLE I: Classification of different network settings

Network settings	Optical bypass	Trusted relay
With optical bypass and with trusted relay (OB-TR)	Allowed	Allowed
With optical bypass (OB)	Allowed	Not allowed
With trusted relay (TR)	Not allowed	Allowed
No optical bypass and no trusted relay (No-OB-No-TR)	Not allowed	Not allowed

We classified different network settings according to optical bypass and trusted relay as shown in Tab. I. The key rates between node pairs are calculated considering the link performance and propagation distance, which is described in detail in Appendix A. The maximum achievable key rates for different reaches are listed in Tab. II. Note that the key rate in Tab. II does not consider any bypassed node. When we use optical bypass, one intermediate node introduces 11% decrement in key rate due to additional optical bypass loss.

TABLE II: Key rate for different reaches

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

B. An Illustrative Example for Different Network Settings

To clarify the role of optical bypass, trusted relays, and QKP in the provisioning of secret keys over a QKD network, let us consider the example in Fig. 2, which refers to PoliQi 5-node ring topology, a QKD network currently being deployed in the city of Milan, Italy [19]. Fig. 2(a) shows a sequence of key rate requests, where notation (s, t) means that a key stream with a given key rate must be set up between nodes s and t for a time period of 20 seconds with key rate of 11 kb/s. The length of all edges is 5 km. The key rate achievable is calculated according to Tab. II. Specifically, the achievable key rate between adjacent nodes is 23 kb/s and the achievable key rates between node pairs with one intermediate node and two intermediate nodes are equal to 20 kb/s and 10 kb/s, respectively. Assume we are given a time period of 2 time-slots to serve key rate requests. The key rate achieved in one time-slot is averaged with the total number of time-slots in the time period. Hence, 23, 20, and 10 kb/s key rates are averaged to 11.5, 10, and 5 kb/s key rates. Assume that, before the requests arrive, 60 kb of key bits are already maintained in the QKPs of both end nodes of each adjacent node pair for them to communicate with each other. Each node is equipped with 2 QKD modules and each link has 2 quantum channels in both directions. The quantum channels are indicated with red and blue solid arrows and the auxiliary link enabled by QKP is shown with blue dotted lines in this figure.

We show how these requests are served with different network settings over the 5-node ring topology in Fig. 2(b)-(f). Fig. 2(b) shows how these requests are served in the first time-slot (i.e., time-slot 0). The achievable key rate for all adjacent nodes in time-slot 0 is 11.5 kb/s, which is larger than the requested key rate. Fig. 2(c) represents the case of point-to-point quantum communications in time-slot 2, i.e., when neither trusted relay nor optical bypass is used. In this case, all requests between adjacent nodes are already served in time-slot 0 and no more requests can be served as point-to-point quantum communication can not serve requests between non-adjacent nodes. Fig. 2(d) represents the case when optical bypass is allowed. In this case, 2 requests can be served, namely, $(1, 3)$ (one path optically bypassing node 2 with 10 kb/s key rate, the other path optically bypassing nodes 5 and 4 with 5 kb/s key rate), $(2, 5)$ (one path optically bypassing node 1 with 10 kb/s key rate, the other path optically bypassing node 3 and 4 with 5 kb/s key rate). Request $(3, 5)$ is not served since it can not be served with either quantum channels (all two quantum modules in nodes 3 and 5 are used) or QKP (no key is stored in QKP for non-adjacent node pair (3,5)). Fig. 2(e) shows the case where only trusted relays are used (optical bypass is not allowed). By using trusted relays, 2 requests are served, $(1, 3)$ and $(3, 5)$. In this case, only node 1 and node 5 have vacant quantum modules, making it impossible to serve request $(2, 5)$ with trusted relay. Finally, Fig. 2(f) shows the case when both optical bypass and trusted relays are exploited. In this case, all key rate requests between non-adjacent nodes are served. Specifically, requests $(1, 3)$ and $(3, 5)$ are both served with two QKD paths (one uses optical bypass, the other uses auxiliary links enabled by QKP). Request $(2, 5)$ is also served with two paths (one uses optical bypass, the other uses trusted relay and QKP). In summary, optical bypass can decrease the number of consumed QKD modules and permit bypassing the untrusted nodes at the cost of a lower key rate. On the other hand, trusted relays allow us to achieve a higher key rate with additional QKD modules.

C. Quantum Key Pool and QKD Path

Fig. 3 shows an example of QKD path, which utilizes quantum channels and QKP to establish a path to distribute keys in QKD network. Each link in a QKD path is called *QKD relay link* and can be either a quantum channel or an auxiliary link enabled by QKP. Fig. 3 shows an example of 6 nodes in which nodes 2 and 4 serve as trusted relays. In the example, node 1 distributes keys to node 6. The QKD path between nodes 1 and 6 consists of three QKD relay links, i.e., $(1, 2)$, $(2, 4)$, and $(4, 6)$. The QKD relay link $(2, 4)$ is an auxiliary link, which uses the keys in the QKP to relay keys directly. QKD relay link $(4, 6)$ uses optical bypass at node 5 and it does not consume any QKD modules in the nodes it traverses, in this case, node 5. Note that, although optical bypass may reduce the key rate of the quantum channel, it allows traversing untrusted nodes, which makes the process of distributing keys more secure. Moreover, note that the maximum key rate of the QKD path is less than or equal to the minimum key rate of all the QKD relay links in the QKD path. For instance, the maximum key rate between node pair $(1, 6)$ in Fig. 3 is 8 kb/s, which is equal to the minimum key rate of all the traversed links (link (2,4)).

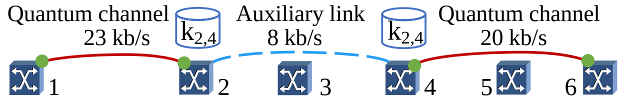


Fig. 3: Example of a QKD path using QKP.

D. NP-hardness of the RCKTA problem

We now prove that the RCKTA problem, even in its *No-OB-No-TR* version, which is considered to be the simplest among the different scenarios, is NP-hard. The proof of NP-hardness of RCKTA problem is shown in theorem 1 as follows.

Theorem 1: The RCKTA problem is NP-hard for all the network settings in Tab. I, even with only one time-slot.

Proof: We first show that the RCKTA problem for *OB-TR*, *OB*, and *TR* is NP-hard by showing that a simplified version of the RCKTA problem can be reduced to the RWA problem. We consider a simplified RCKTA problem for *OB-TR*, *OB*, and *TR* as the RCKTA problem by neglecting the constraints relative to QKD module capacity, splittable flow, caching of keys, and the loss during transmission (i.e., loss due to link performance, propagation distance, and optical bypass). Since the simplified RCKTA problem neglects the loss during transmission, *TR* and *OB* have the same generation rate in all the quantum channels, and hence a quantum channel can be considered as a wavelength with a fixed capacity (key generation rate). Moreover, since we do not consider any limit of QKD module capacity in the simplified RCKTA problem, the difference in quantum module usage for *OB-TR*, *OB*, and *TR* is not considered. Hence, the QKD path found for all three network settings can be considered as a lightpath using the same wavelength. Therefore, the simplified RCKTA problem can be stated as follows: **Given** a QKD network topology, a set of key rate requests, **decide** the routing and channel (i.e., wavelength) assignment, **constrained to** the number of channels in each link, with the **objective** of maximizing the number of served requests. Since the problem statement of the simplified RCKTA problem for *OB-TR*, *OB*, and *TR* is the same as the RWA problem, which is proven to be an NP-hard problem [20], we can conclude that the RCKTA problem for these network settings is NP-hard and even harder than the original RWA problem.

Then, we prove that for *No-OB-No-TR*, the RCKTA problem is NP-hard with only one time-slot. For *No-OB-No-TR*, to establish one QKD path between node pair (i, j), when we assign one quantum module in node i to the QKD path with node j, node j must also assign one quantum module to the QKD path with node i, which is defined as *quantum module assignment restriction* constraint. We show as follows that the multiple knapsack problem [21], a known NP-hard problem, can be reduced to the decision problem of the RCKTA problem with *No-OB-No-TR*, one time-slot, and no *quantum module assignment restriction* constraint, which is referred to as *simplified RCKTA problem* with *No-OB-No-TR*. Thus, the RCKTA problem for *No-OB-No-TR* is NP-hard.

The multiple knapsack problem is described as follows: Given a set of knapsacks $M = \{1, 2..., m\}$ with maximum capacity $W_i, ..., W_m$, a set of items $S = \{1, 2, ..., n\}$ with weights $w_1, w_2, ..., w_n$, value v_i^j of assigning item i to knapsack j, and an integer V, does there exist a set of subsets $I = \{S_1, S_2, ..., S_m\} \in S$, such that any two subsets do not contain the same item, $\sum\limits_{i \in S_j} w_i^j \leq W_j$ for any knapsack $j \in M$, and $\sum\limits_{j \in M} \sum\limits_{i \in S_j} v_i^j \geq V$.

Given an instance of the multiple knapsack problem, we can construct an instance of the decision version of the simplified RCKTA problem with *No-OB-No-TR* as follows. We use M to denote the set of nodes. The number of quantum modules in node i corresponds to the capacity of bins W_i. For each node j, each node i connected to it corresponds to an item and the number of QKD modules needed to serve the request between node i and j is denoted with weight w_i. In addition, the key rate of the request between node i and j corresponds to the value v_i^j. The corresponding question becomes whether there exists a solution to the simplified RCKTA problem such that the sum of the values of the selected items in all the knapsacks is greater or equal to V.

The simplified RCKTA problem reaches the optimum if and only if the multiple knapsack problem reaches the optimum as these two problems have identical objective functions with the settings above. In conclusion, the RCKTA problem is NP-hard for all the network settings in Tab. I.

IV. Mixed Integer Linear Programming Model for RCKTA problem

This section presents the proposed MILP model for RCKTA problem and then extends it considering different settings of optical bypass and trusted relay.

A. Decision Variables and Objective Function

Sets and parameters, as well as variables are reported in Tab. III and Tab. IV, respectively. We denote $\delta^+(i)$ and $\delta^-(i)$ as the set of outgoing links and the set of incoming links from i, respectively. For the QKD path between a node pair or one link e, the initial and the end nodes are denoted with $a(e)$ and $b(e)$, respectively. Besides, the opposite direction of link e is denoted with $\overline{e}$. The objective function is to maximize the weighted number of served requests. Besides, the key storing rate is also included in the objective function, which breaks ties of different solutions that serve the same number of requests, giving priority to the solution that has a higher key storing rate.

TABLE III: Sets and Parameters for the MILP Model

Parameter	Description
N_p	Set of physical nodes
N_t	Set of trusted relays in N_p
E_p	Set of unidirectional physical links
P	Set of all possible physical node pairs
R	Set of node pairs with key rate requests
W	Set of quantum channels
Φ_e	Set of paths for node pair $e \in P$
G_p	Available keys in QKP for node pair $p \in P$ before serving requests
Q_i^m	Capacity of the QKP of node $i \in N_p$
C_i	Number of QKD modules in node $i \in N_p$
L	Set of indexes for possible key rates of a QKD path
T	Set of all time-slots
θ	Interval of one time-slot
r_d	Key rate of key rate request $d \in R$
$\delta_{\phi,e}$	Equal to 1 if link $e \in E_p$ is in the path ϕ
h_ϕ	Maximum key rate in path ϕ
M	A large constant, equal to the maximum key rate of a path
ψ_l	Value of key rate $l \in L$
α	Weight for serving key rate requests

TABLE IV: Variables for the MILP Model

Variable	Description
$q_{e,p}^{w,t}$	Binary, equal to 1 if QKD path $p \in P$ uses link $e \in P$ in quantum channel $w \in W$ at time-slot $t \in T$
$f_p^{w,t}$	Binary, equal to 1 if one QKD path $p \in P$ uses quantum channel $w \in W$ at time-slot $t \in T$
$u_{e,p}^{w,t}$	Binary, equal to 1 if link $e \in P$ in QKD path $p \in P$ uses quantum channel $w \in W$ at time-slot $t \in T$
$\overline{u}_{e,p}^{w,t}$	Binary, equal to 1 if link $e \in P$ in QKD path $p \in P$ uses QKP at time-slot $t \in T$
$x_{e,\phi,p}^{w,t}$	Binary, equal to 1 if path $\phi \in \Phi_e$ is selected for connection between the end nodes of link $e \in P$ in QKD path $p \in P$ at quantum channel $w \in W$ at time-slot $t \in T$
$\zeta_p^{w,t}$	Key rate provided from quantum channel $w \in W$ of $p \in P$ at time-slot $t \in T$
$z_{e,p}^{w,t}$	Key rate provided from QKP for auxiliary link $e \in P$ in QKD path $p \in P$ in quantum channel $w \in W$ at time-slot $t \in T$
λ_d^t	Key rate used to serve the request $d \in R$ at time-slot t
k_p^t	Key storing rate in QKP for QKD path $p \in P$ at time-slot $t \in T$
y_d	Binary, equal to 1 if key rate request between node pair $d \in R$ is served
g_p^t	Amount of stored keys in QKP for node pair $p \in P$ at the end of time-slot $t \in T$
$\overline{q}_{e,p}^{w,t,l}$	Binary, equal to 1 if QKD path $p \in P$ uses link $e \in P$ in quantum channel $w \in W$ at time-slot $t \in T$ with $l \in L$
$\tau_p^{w,t,l}$	Binary, equal to 1 if QKD path $p \in P$ uses quantum channel $w \in W$ at time-slot $t \in T$ with key rate $l \in L$

Objective: Maximizing the number of served requests, then the key storing rate. Weight α is set to a large value such that the minimum variation of the number of served requests is larger than the maximum value of key storing rate, to give higher priority to serve key rate requests.

$$\max \quad \alpha * \sum_{d\in R} r_d * y_d + \sum_{p\in P}\sum_{t\in T} k_p^t \tag{1}$$

B. Constraints

We first describe the constraints for the RCKTA problem with optical bypass and trusted relay.

1) Flow conservation constraints for QKD path: Eqn. (2) is the flow constraint for QKD path for all node pairs $p \in P$ in quantum channel $w \in W$ at time-slot $t \in T$.

$$\sum_{e\in\delta^+(i)} q_{e,p}^{w,t} - \sum_{e\in\delta^-(i)} q_{e,p}^{w,t} = \begin{cases} f_p^{w,t} & i = a(p) \\ -f_p^{w,t} & i = b(p) \\ 0 & otherwise \end{cases} \tag{2}$$
$$\forall p \in P, i \in N_p, w \in W, t \in T$$

2) Link formation of QKD path: Eqn. (3) ensures that the QKD path p can use a QKD relay link $e \in P$ only if either the quantum channel or QKP can provide keys.

$$q_{e,w}^{p,t} = u_{e,p}^{w,t} \wedge \overline{u}_{e,p}^{w,t} \quad \forall p \in P, e \in P, w \in W, t \in T \tag{3}$$

3) QKD module capacity constraint: Eqn. (4) ensures that the number of quantum modules used in node $i \in N_p$ can not exceed the number of QKD modules in the node.

$$\sum_{p\in P}\sum_{e\in\delta^+(i)\cup\delta^-(i)}\sum_{w\in W} u_{e,p}^{w,t} \leq C_i \quad \forall i \in N_p, t \in T \tag{4}$$

4) Trusted relay constraint: Eqn. (5) ensures that the QKD relay link can only use trusted relay as an intermediate node.

$$q_{e,p}^{w,t} = 0 \quad \forall p \in P, e \in P, w \in W, t \in T : \\ (a(p) \neq a(e) \wedge a(e) \notin N_t) \vee (b(p) \neq b(e) \wedge b(e) \notin N_t) \tag{5}$$

5) Routing of quantum channels: Eqn. (6) determines the route $\phi \in \Phi_e$ for a QKD relay link $e \in P$ using quantum channel. Eqn. (7) ensures that for a QKD relay link e, the quantum channel w of the path $\phi \in \Phi_e$ can only be assigned once. Eqn. (8) ensures that two different QKD relay links do not use the same channel of the same physical link.

$$\sum\nolimits_{\phi\in\Phi_e} x_{e,\phi,p}^{w,t} = u_{e,p}^{w,t} \quad \forall p \in P, e \in P, w \in W, t \in T \tag{6}$$

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

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

6) Achievable key rate constraints: Eqn. (9) selects a key rate $l \in L$ for a path p on quantum channel w at time-slot t. Eqn. (10) ensures that one key rate $l \in L$ is selected for an edge $e \in P$ if the edge e is used for path $p \in P$. Eqn. (11) and Eqn. (12) ensure that the key rate of the edge e in the path p must be consistent with the key rate of path p.

$$\sum\nolimits_{l\in L} \tau_p^{w,t,l} = f_p^{w,t} \quad \forall p \in P, w \in W, t \in T \tag{9}$$

$$\sum\nolimits_{l\in L} \overline{q}_{e,p}^{w,t,l} = q_{e,p}^{w,t} \quad \forall p \in P, e \in P, w \in W, t \in T \tag{10}$$

$$\overline{q}_{e,p}^{w,t,l} \leq \tau_p^{w,t,l} \quad \forall p \in P, e \in P, w \in W, t \in T, l \in L \tag{11}$$

$$\tau_p^{w,t,l} \leq \sum\nolimits_{e\in P} \overline{q}_{e,p}^{w,t,l} \quad \forall p \in P, w \in W, t \in T, l \in L \tag{12}$$

7) Key supply constraints: Eqn. (13) ensures that the key rate of an edge $e \in P$ in path p can not exceed the key rate provided from the quantum channel and the QKP from the end nodes of e. Eqn. (14) calculates the key rate of path p on quantum channel w at time-slot t. Eqn. (15) ensures no key can be provided from the QKD path if this path is not enabled. Eqn. (16) ensures QKP only provides keys when it is used.

$$\sum\nolimits_{l\in L} \overline{q}_{e,p}^{w,t,l}\psi_l \leq \sum\nolimits_{\phi\in\Phi_e} x_{e,\phi,p}^{w,t} h^\phi + z_{e,p}^{w,t} \\ \forall p \in P, e \in P, w \in W, t \in T \tag{13}$$

$$\zeta_p^{w,t} \leq \sum\nolimits_{l\in L} \tau_q^{w,t,l}\psi_l \quad \forall p \in P, w \in W, t \in T \tag{14}$$

$$\zeta_p^{w,t} \leq M f_p^{w,t} \quad \forall p \in P, w \in W, t \in T \tag{15}$$

$$z_{e,p}^{w,t} \leq M\overline{u}_{e,p}^{w,t} \quad \forall p \in P, e \in P, w \in W, t \in T \tag{16}$$

8) Request serving constraint: Eqn. (17) ensures that key request d is served if the sum of key rates for d in all quantum channels and time-slots is greater or equal to the required key rate r_d during the time period to serve requests.

$$r_d y_d \leq \sum\nolimits_{t\in T} \lambda_d^t / |T| \quad \forall d \in R \tag{17}$$

9) Key storing constraint: Eqn. (18) ensures that the key storing rate cannot exceed the difference between the key rate of generating keys and the key rate of using keys. Eqn. (19) ensures that the key rate provided by QKP between node pair p (working as the auxiliary links p and $\overline{p}$) for the QKD paths between all the node pairs p' at the time-slot 2 and the following time-slots should be smaller or equal to the key rate achievable with the keys stored in QKP at the end of the previous time-slot. Eqn. (20) ensures that the key rate provided by the QKP between node pair p (working as the auxiliary links p and $\overline{p}$) for the QKD paths between all the node pairs p' at the first time-slot (time-slot 1) should be smaller or equal to the key rate achievable with the keys stored in QKP before serving requests. Eqn. (21) and Eqn. (22) obtain the keys stored in QKPs for node pair (i, j) at each time-slot. Eqn. (23) ensures that the keys stored in QKP of node $i \in N_p$ at each time-slot can not be negative and do not exceed the maximum capacity of QKP.

$$k_p^t \leq \sum_{w\in W}(\zeta_p^{w,t} + \zeta_{\overline{p}}^{w,t}) - \sum_{p'\in P}\sum_{w\in W}(z_{p,p'}^{w,t} + z_{\overline{p},p'}^{w,t}) \\ - \lambda_p^t - \lambda_{\overline{p}}^t \quad \forall p \in P, t \in T \tag{18}$$

$$\sum_{p'\in P}\sum_{w\in W}(z_{p,p'}^{w,t} + z_{\overline{p},p'}^{w,t})\theta \leq g_p^{t-1} \; \forall p \in P, t \in T \setminus \{1\} \tag{19}$$

$$\sum_{p'\in P}\sum_{w\in W}(z_{p,p'}^{w,1} + z_{\overline{p},p'}^{w,1})\theta \leq G_p \quad \forall p \in P \tag{20}$$

$$0 \leq g_p^t \leq g_p^{t-1} + k_p^t\theta \quad \forall p \in P, t \in T \setminus \{1\} \tag{21}$$

$$0 \leq g_p^1 \leq G_p + k_p^1\theta \quad \forall p \in P \tag{22}$$

$$\sum_{p\in P: i=a(p)\vee i=b(p)} g_p^t \leq Q_i^m \quad \forall i \in N_p, t \in T \tag{23}$$

C. Extension of the MILP formulation to Different Settings

Here, we extend the MILP model above to the following different settings.

1) No optical bypass and no trusted relay (No-OB-No-TR): Eqn. (24) and Eqn. (25) ensure that optical bypass and trusted relay are not allowed, respectively.

$$q^{w,t}_{e,p} = 0 \quad \forall p \in P, e \in P, w \in W, t \in T : \\ a(e) \neq a(p) \text{ or } b(e) \neq b(p) \tag{24}$$

$$u^{w,t}_{e,p} = 0 \quad \forall p \in P, e \in P - E_p, w \in W, t \in T \tag{25}$$

2) With optical bypass (OB): This case only requires Eqn. (24) to ensure that trusted relay is not allowed.

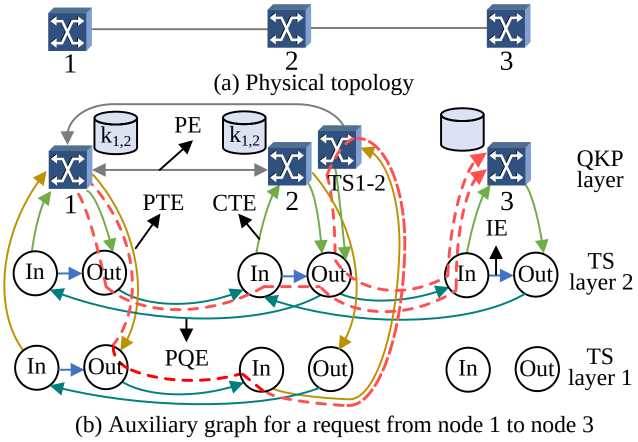
3) With trusted relay (TR): This case only requires Eqn. (25) to ensure that optical bypass is not possible.

V. Auxiliary-graph-based Resource-efficient RCKTA Algorithm

Since the previous MILP model is computationally intractable, we propose a novel Resource-Efficient RCKTA (RE-RCKTA) algorithm based on an auxiliary-graph model with multiple time-slots. As the MILP model, the RE-RCKTA algorithm works for all the network settings by assigning different weights for the auxiliary graph. Different from previous works, the solution proposed in this paper is the first to construct an auxiliary graph for QKD network that incorporates the constraint of the limited number of QKD modules and the QKP to establish a QKD path.

A. Auxiliary Graph Model with QKP

Fig. 4 shows an example of an auxiliary graph with QKP. Fig. 4 (a) and Fig. 4 (b) are the physical topology with three nodes and the corresponding auxiliary graph, respectively. We consider two time-slots in this example. Node 1 in time-slot 1 and all the nodes in time-slot 2 have 2 available QKD modules. In addition, node 2 and node 3 have 2 and 0 available QKD modules in time-slot 1, respectively.



(a) Physical topology

(b) Auxiliary graph for a request from node 1 to node 3

Fig. 4: Example of an auxiliary graph with QKP.

The auxiliary graph consists of two parts: 1) **QKP layer;** 2) **Time-slots (TS) layers**. An auxiliary graph is constructed when serving a request with time-slot t. Each node in the physical node has a corresponding node in the QKP layer, which is defined as *QKP node*. The *QKP node* has outgoing edges to the same physical node with at least one QKD module in all the TS layers. For example, *QKP node* 2 in QKP layer of Fig. 4 corresponds to physical node 2, and node 2 in QKP layer has outgoing edges to nodes in both TS layer 1 and TS layer 2. The links in QKP layer are enabled with the keys stored in the QKP. Each node in the physical topology corresponds to two nodes (*in node* and *out node*) in each TS layer to model the limitations of the numbers of QKD modules. Each TS layer corresponds to 1 time-slot and the TS layer in time-slot t is named as *TS layer* t. As shown in Fig. 4, to serve the key request between the node pair $(1, 2)$, the auxiliary graph contains the current TS layer 2 and the previous TS layer 1. Moreover, if node i has exactly one QKD module in time-slot t, we introduce an additional *constrained TS node*, *TS-t-i* in the QKP layer. Node *TS-t-i* only has incoming edges from the *in node* of node i from TS layer t, and does not have any outgoing edges to TS layer t, which ensures that no path traversing node i can use more than 1 QKD modules if node i has only one QKD module in time-slot t.

The auxiliary graph is constructed for a given key rate r_t, which is the minimum key rate on all the edges. The detailed descriptions of edges in the auxiliary graph are as follows.

QKP edge - (PE): Occurs between a node pair p in QKP layer if the keys in the QKP can support the key rate of r_t.

Current transponder edge - (CTE): Exists between the nodes in the QKP layer and the corresponding node in the current TS layer. For both *QKP node* and *constrained TS node*, *CTE* exists only if the corresponding physical node has at least one QKD module at the current time-slot.

Previous transponder edge - (PTE): Exists between the nodes in the QKP layer and the nodes in previous TS layers. Whether a *PTE* exists follows the same rule as the *CTE*.

Potential QKD edge - (PQE): Exists between *out node* and *in node* of different physical nodes at the same TS layer. Note that *PQE* is a *QKD relay link* that uses a quantum channel, as described in Fig. 3. The key rate achievable in *PQE* should be able to support the key rate of r_t.

Inter-node edge - (IE): Exists between the *in node* and *out node* of the same physical nodes with at least 2 QKD modules for all time-slots. A QKD path uses an *IE* if the corresponding node works as trusted relay to distribute keys.

After constructing the auxiliary graph, the RE-RCKTA algorithm finds shortest path in the auxiliary graph to serve requests. For example, to serve the request between nodes 1 and 3, two possible paths are shown in Fig. 4 with red dotted lines. We denote the in(out) node of node i in time-slot t with *t-in(out)-i*. The first path is (*1*, *2-out-1*, *2-in-2*, *2-out-2*, *2-in-3*, *3*), which consumes 4 QKD modules. The second path is (*1*, *1-out-1*, *1-in-2*, *TS-1-2*, *2-out-2*, *2-in-3*, *3*). The auxiliary graph does not contain any outgoing edge from *1-in-2* to node 2. Otherwise, path 2 may select edges (*1-in-2*, *2*) and (*2*, *1-out-2*), which use 2 quantum modules in node 2 at TS layer 1 and contradicts the fact that node 2 only has 1 quantum module at TS layer 1.

B. Weight Assignment Policies for Different Settings

The weight assignment scheme is listed in Tab. V. Assume that h_s denote the smallest number of hops needed in the

TABLE V: Weight assignment scheme for different network settings

Edge	OB-TR	OB	TR	No-OB-No-TR
PE	$10^{-6} \cdot h_s$	$3 \cdot 10^6 +$ $10^{-6} \cdot h_s$	$10^{-6} \cdot h_s$	$3 \cdot 10^6 +$ $10^{-6} \cdot h_s$
CTE	$\beta/2$	$\beta/2 + 10^6$	$\beta/2$	$\beta/2 + 10^6$
PTE	$\beta/2 + 10^{-9}$	$\beta/2 + 10^6$ $+10^{-9}$	$\beta/2 + 10^{-9}$	$\beta/2 + 10^6$ $+10^{-9}$
PQE	$\beta +$ $10^{-6} \cdot h_p$	$\beta + 10^6 +$ $10^{-6} \cdot h_p$	β	$\beta + 10^6$
IE	1	Not exist	1	Not exist

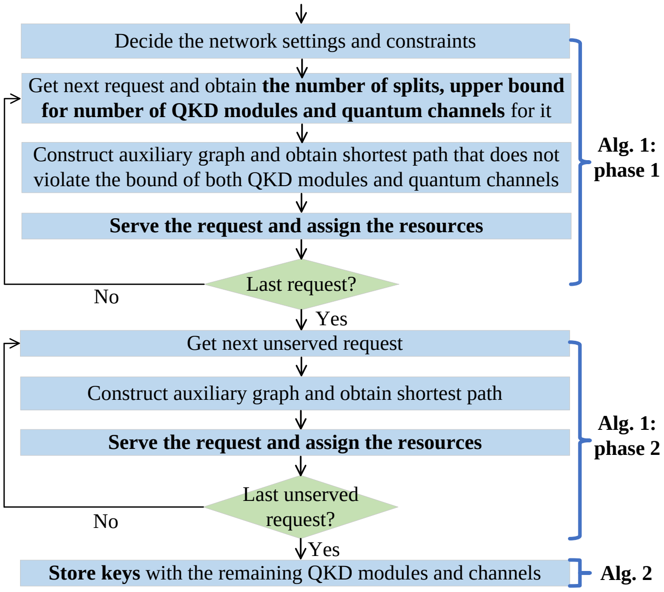

Decide the network settings and constraints
Get next request and obtain **the number of splits, upper bound for number of QKD modules and quantum channels** for it
Construct auxiliary graph and obtain shortest path that does not violate the bound of both QKD modules and quantum channels
Serve the request and assign the resources
Last request?
No
Yes
Alg. 1: phase 1
Get next unserved request
Construct auxiliary graph and obtain shortest path
Serve the request and assign the resources
Last unserved request?
No
Yes
Alg. 1: phase 2
Store keys with the remaining QKD modules and channels
Alg. 2

Fig. 5: Flow chart of the RE-RCKTA algorithm.

physical topology to connect node pair (i, j). For *OB-TR* and *TR*, the weight of *PE* is set to $10^{-6} \cdot h_s$, which is positively correlated with the QKD modules used to generate the keys in QKP. For *OB* and *No-OB-No-TR*, the weight of *PE* is set to $3 \cdot 10^6 + 10^{-6} \cdot h_s$. For *CTE*, *PTE*, and *PQE*, the variable part of the weight, β is set according to the number of used QKD modules in a path. When both nodes i and j have more than one QKD module, β equal to 1. When either node i or j have only one QKD module, β is equal to 2.5, penalizing using a node with the last QKD module. For *OB* and *No-OB-No-TR*, a request can only be served with three different types of paths: 1) A path contains only one *PE*. 2) A path contains two *CTEs* and one *PQE*. 3) A path contains two *PTEs* and one *PQE*. The weight of the shortest path for a request must be less or equal to $4 \cdot 10^6$. *OB* and *No-OB-No-TR* do not have *IE* since these network settings do not allow trusted relay. The weight of *IE* for *OB-TR* and *TR* is equal to 1, which represents that one additional QKD module is used for trusted relay.

C. General Resource-Efficient RCKTA Algorithm

The flowchart of the RE-RCKTA algorithm is shown in Fig. 5. RE-RCKTA consists of two parts, namely Algorithm 1 to serve key rate requests and Algorithm 2 to store keys. Algorithm 1 consists of two phases to serve the requests. Phase 1 aims to serve the requests resource-efficiently. For the requests not served in phase 1, phase 2 tries to serve each request in a greedy manner. For each request, in phase 1, the RE-RCKTA algorithm first obtains the number of splits (paths) required and the upper bound for the number of QKD modules and quantum channels in one path. Then, the algorithm constructs an auxiliary graph and obtains the shortest path that does not violate the bounds of both QKD modules and quantum channels to serve the request. After serving the requests resource-efficiently, the RE-RCKTA algorithm serves requests greedily with the shortest path and does not consider the bound for resources. After serving the requests in two phases with Algorithm 1, Algorithm 2 stores keys with the remaining QKD modules and quantum channels.

The first part of the RE-RCKTA algorithm, namely Algorithm 1, utilizes the auxiliary graph in Fig. 4 to serve the requests with multiple time-slots and QKP as follows. For a given network setting, k_u is the maximum achievable key rate using the shortest path between node pair (i, j), while k_l is the lowest achievable key rate using the shortest path between node pair (i, j). Algorithm 1 first sorts node pairs R according to the length of shortest path between the node pair (in terms of number of hops) in ascending order in the physical topology to give priority to serving requests that require less resources (line 1). Then Algorithm 1 first determines the k_u and k_l for the connections between the node pair of a request d (line 2). For *OB-TR*, k_u and k_l are the maximum achievable key rate of the shortest path using only trusted relay and optical bypass, respectively. For *OB*, *TR*, and *No-OB-No-TR*, k_u and k_l are equal and are also obtained with the shortest path.

After obtaining k_u and k_l (line 3), Algorithm 1 calculates the maximum number of splits, n_u and the minimum number of splits, n_l for the request d (line 4). We denote the minimum number of QKD modules required with key with number of splits n as c_n. For each n ranging between n_l and n_u, an auxiliary graph with only one time-slot with key rate of r_d/n is constructed, and then the algorithm calculates c_n with the shortest path in the auxiliary graph (line 4). After getting the number of splits n_s that uses the least number of quantum modules (line 5), the algorithm calculates the key rate k_s for each split and determines c_{n_s} (line 6). With n_s, the algorithm calculates the required key rate k_s and the bound of the number of QKD modules (c_{n_s}) and quantum channels (l_{n_s}). Then the algorithm loops over each quantum channel and time-slot and finds the shortest path to serve the request (line 7-20). We define $l_d^{w,t}$ as the quantum channel resources required of a path for key rate request d in quantum channel w and time-slot t. $l_d^{w,t}$ contains two parts. The first part is the number of *physical edges whose quantum channels are occupied* (PEQC) in the path. The second part is the number of PEQCs needed in the shortest path (without *PE*) between the end nodes of all *PE* in the path. Then the algorithm obtains the achievable key rate ($\zeta_d^{w,t}$), quantum modules consumed ($c_d^{w,t}$), and $l_d^{w,t}$. If the request can be served without violating the bound of resources, the algorithm assigns resources to serve the request (line 14-15). Otherwise, the algorithm aborts all the resource assignments for d and reverts r_d (line 19).

After serving the key rate requests resource-efficiently, the

Algorithm 1: RE-RCKTA1 for serving requests

```
   Input: $N_p, E_p, P, N_t, W, C_i, T, \theta, R, Q_p^0, Q_p^m, r_d, L$
   Output: Routing, channel, and key-rate assignment for
           serving requests and storing keys
1  Sort $R$ with the length of shortest path in ascending order
2  for each secret key rate request $d \in R$ do
3      Obtain $k_u$ and $k_l$ for the node pair of request $d$
4      $n_u \leftarrow \lceil r_d/k_l \rceil$, $n_l \leftarrow \lceil r_d/k_u \rceil$, get $c_n, \forall n_l \le n \le n_u$
5      Get minimum $n_s$ such that $n_s c_{n_s} \le n c_n, \forall n_l \le n \le n_u$
6      $k_s \leftarrow r_d/n_s$, determine $c_{n_s}$ and $l_{n_s}$
7      for each quantum channel $w \in W$ do
8          if $r_d = 0$ then Break
9          for each time-slot $t \in T$ do
10             if $r_d = 0$ then Break
11             Construct an auxiliary graph for key rate of
                 $min(r_d, k_s)$ with time-slots 1 to $t$
12             Get shortest path, $\zeta_d^{w,t}$, $c_d^{w,t}$ and $l_d^{w,t}$
13             if path exists $\wedge$ ($c_d^{w,t} \le c_{n_s} \vee l_d^{w,t} \le l_{n_s}$) then
14                 Assign quantum channels and QKD modules
15                 $r_d \leftarrow r_d - \zeta_d^{w,t}$, update $g_p^{t_1}, t_1 \in T$
16             end
17         end
18     end
19     if $r_d > 0$ then Abort all resource assignments for d
20 end
21 for each unserved secret key rate request $d \in R$ do
22     for each quantum channel $w \in W$ do
23         if $r_d = 0$ then Break
24         for each time-slot $t \in T$ do
25             if $r_d = 0$ then Break
26             $\gamma_s = -1$, $\zeta_s = 0$, $c_s = Inf$, $l_s = Inf$
27             for index $\gamma \in L$ of each possible key rate do
28                 Construct an auxiliary graph for key rate of
                     $min(r_d, \psi_\gamma)$ with time-slots 1 to $t$
29                 Obtain shortest path, $\zeta_d^{w,t}$, $c_d^{w,t}$ and $l_d^{w,t}$
30                 if path exists $\wedge$ ($\zeta_d^{w,t}/c_d^{w,t} \ge \zeta_s/c_s$ $\wedge$
                     $\zeta_d^{w,t}/l_d^{w,t} \ge \zeta_s/l_s$) then
31                     $\gamma_s = \gamma$, $\zeta_s = \zeta_d^{w,t}$, $c_s = c_d^{w,t}$, $l_s = l_d^{w,t}$
32                 end
33             end
34             if $\gamma_s > 0$ then
35                 Assign quantum channels and QKD modules
36                 $r_d \leftarrow r_d - \zeta_d^{w,t}$, update $g_p^{t_1}, t_1 \in T$
37             end
38         end
39     end
40     if $r_d > 0$ then Abort all resource assignments for $d$
41 end
```

Algorithm 2: RE-RCKTA2 for storing keys

```
   Input: $N_p, E_p, P, N_t, W, C_i, T, \theta, g_p^t, L$, used QKD modules
          and quantum channels
   Output: Resource assignment for storing keys
1  Sort $P$ with the length of shortest path in ascending order
2  for each node pair $p \in P$ do
3      for each time-slot $t \in T$ do
4          Get the remaining capacity $g_p^r$ of QKP
5          if $g_p^r$ == 0 then Break
6          for each quantum channel $w \in W$ do
7              for index $\gamma \in L$ of each possible key rate do
8                  Construct an auxiliary graph with key rate
                     of $min(g_p^r/\theta, \psi_\gamma)$ with time-slots 1 to $t$
9                  Obtain shortest path and $\zeta_p^{w,t}$
10                 if path found then
11                     Assign quantum channels and QKD
                         modules and update $g_p^t$
12                     Break
13                 end
14             end
15             Determine $\overline{k}_{a(p)}$ and $\overline{k}_{b(p)}$
16             if $\overline{k}_{a(p)}$ == $Q_{a(p)}^m$ or $\overline{k}_{b(p)}$ == $Q_{b(p)}^m$ then
17                 Break
18             end
19         end
20     end
21 end
```

algorithm serves the requests without considering the bound for resources (line 21-41). The resource-efficiency of quantum channels and QKD modules are denoted by $\zeta_d^{w,t}/l_d^{w,t}$ and $\zeta_d^{w,t}/c_d^{w,t}$, respectively. For each quantum channel and each time-slot, Algorithm 1 checks each possible key rate and determines the most resource-efficient key rate (line 26-33). Then the algorithm assigns the resources to serve the request (line 34-37). If the key rate request is not served, the algorithm aborts all resource assignments for the request (line 40).

After serving the requests, Algorithm 2 generates keys to be stored with the remaining quantum channels and QKD modules. Assume the amount of keys stored in the QKP of node i is denoted with $\overline{k}_i$. For each time-slot and quantum channel, Algorithm 2 stores the keys in a greedy manner. The Algorithm 2 first get the remaining capacity, g_p^r for the QKP of the end nodes of the node pair p. Then Algorithm 2 iterates over the possible key rates in descending order to use QKD path with a higher key rate (line 7-14). For the index γ of each possible key rate, Algorithm 2 constructs an auxiliary graph with key rate of $min(g_p^r/\theta, \psi_\gamma)$ to consider the capacity of QKP (line 8). Note that the auxiliary graph does not contain *PE* since it wastes the keys already stored. After finding a shortest path for a key rate, Algorithm 2 assigns quantum channels and QKD modules and updates the amount of keys stored in the QKP (line 11-12). Finally, if the keys stored in the QKP are equal to the capacity of the QKP, the algorithm stops storing keys for node pair p (line 15-18).

Complexity: The complexity of the algorithm is mainly determined by the complexity of constructing an auxiliary graph and finding the shortest path in the auxiliary graph. The complexity of constructing the nodes in the auxiliary graph is $O(|T||N_p|)$. The time complexities to construct *PE*, *CTE*, *PTE*, and *IE* for an auxiliary graph are $O(|N_p|^2)$, $O(N_p)$, $O(|T||N_p|)$, and $O(|T||N_p|)$. The time complexity to construct *PQE* is $O(K|T||N_p|^2)$, where K is equal to the number of pre-defined shortest paths between each node pair. Hence, the time complexity of constructing the auxiliary graph is $O(K|T||N_p|^2)$. Since the number of nodes and edges in the auxiliary graph is $O(|T||N_p|)$ and $O(|T||N_p|^2)$, respectively, the complexity of finding the shortest path in the auxiliary graph using the Dijkstra algorithm is $O(|T||N_p|^2 + |T||N_p|log(|T||N_p|)$.

VI. Numerical Results

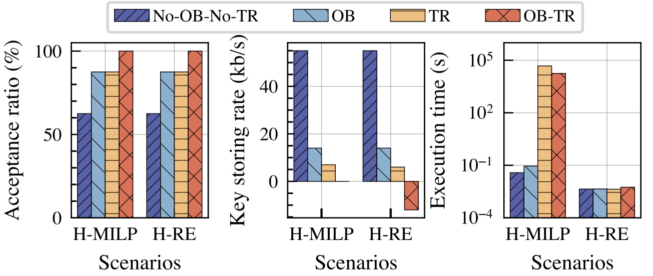
In this section, we first validate the performance of the proposed RE-RCKTA algorithm by comparing its performance to that of MILP model in terms of acceptance ratio, key storing rate, and execution time. Then, we evaluate the performance of different network settings in large network scenarios.

A. Simulation Setup

The simulations are performed on a workstation with Intel(R) Core(TM) i7-4790 CPU (4 Cores @ 3.60GHz) and 16 GB of memory. We implement the MILP formulation using AMPL (A Mathematical Programming Language) and solve it with CPLEX 20.1.0 MIP solver. The RE-RCKTA algorithm is implemented with Python 3.8. In our evaluations, we consider two network topologies, a small network topology (the PoliQi topology in Fig. 2) and a large mesh network topology in Ref. [22] (the USnet topology, 24 nodes and 43 bidirectional links). The link length of the PoliQi network is set, according to the real deployment, to an average link length 5 km. The link length of the large topology is set to values uniformly distributed in [2, 8] km. The number of quantum channels in each link is 5 as described in Appendix A. Note that our proposed approach is extensible to different network topologies. We provide the main insights for the performance of different network settings by providing results on two real-world topologies, specifically, the PoliQi topology and the scaled-down USnet mesh topology. The simulated time duration is 30 seconds. To compare the MILP and the RE-RCKTA algorithm, we consider a high-traffic scenario, in which one unidirectional request is generated between every two nodes (10 requests) and the average requested key rates are set to 12 kb/s. Two time-slots are used to compare MILP and RE-RCKTA algorithm considering the scalability of MILP. To extend our evaluation, we compare different network settings with the RE-RCKTA algorithm in a large network topology, in which one unidirectional request is randomly generated between every two nodes with 80% of probability (on average 221 requests). The number of time-slots to evaluate the large network topology is set to 8 (8 time-slots are enough to have a good performance as the granularity of the key rate in one time-slot is already smaller than the minimum requested key rate in the simulation), and increasing the number of time-slots will increase the execution time. We consider cases under different loads of key rate requests, different capacities of QKD modules, and different amounts of keys stored keys in QKP before serving requests. When evaluating the impact of varying key rates and keys stored in QKP, the average number of QKD modules in each node is 12. All the simulations are averaged over 8 instances, which allows, in our results, to achieve confidence intervals (for the confidence level of 95%) within $\pm 5\%$ of the reported results.

B. Comparison of MILP and the RE-RCKTA Algorithm

We apply the MILP and RE-RCKTA algorithm to a high-traffic scenario (the cases solved with MILP and RE-RCKTA algorithm are named as H-MILP and H-RE, respectively). Keys stored in the QKP are 90 kb for all adjacent node pairs. Fig. 6 (a) shows the acceptance ratio under different network settings. The acceptance ratio of *OB-TR* is 13% higher than *OB* and *TR*, respectively. This is because *OB-TR* combines both the advantages of *OB* (saving QKD modules) and *TR* (higher key rate). Regarding the gap between the RE-RCKTA algorithm and MILP, the RE-RCKTA algorithm and MILP have the same acceptance ratio. In addition, the RE-RCKTA algorithm also has a good performance in terms of key storing rate. As shown in Fig. 6 (b), *No-OB-No-TR* achieves the highest key storing rate since it used the fewest keys to serve requests. The key storing rate of *OB* is 50% higher than *TR* and the key storing rate of *OB-TR* is 0 kb/s. For all the network settings, the gap between the RE-RCKTA algorithm and MILP is below 16 kb/s, which is lower than the key rate of a quantum channel with a distance shorter than 20 km. The performance of the RE-RCKTA is achieved with a significant reduction in execution time, e.g. execution time of *OB-TR* reduces from about 17600 sec to less than 1 sec.



(a) Acceptance ratio (b) Key storing rate (c) Execution time

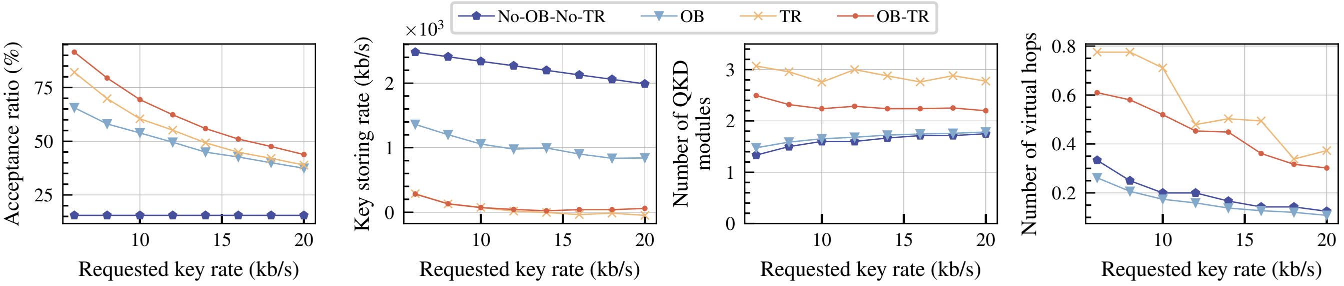
Fig. 6: Comparison of MILP and the RE-RCKTA algorithm.

In conclusion, RE-RCKTA can serve the key rate requests and store keys with small optimality gaps. Besides, *OB-TR* serves more key requests compared to *OB* and *TR*.

C. Performance Evaluation of RCKTA on a Large Topology

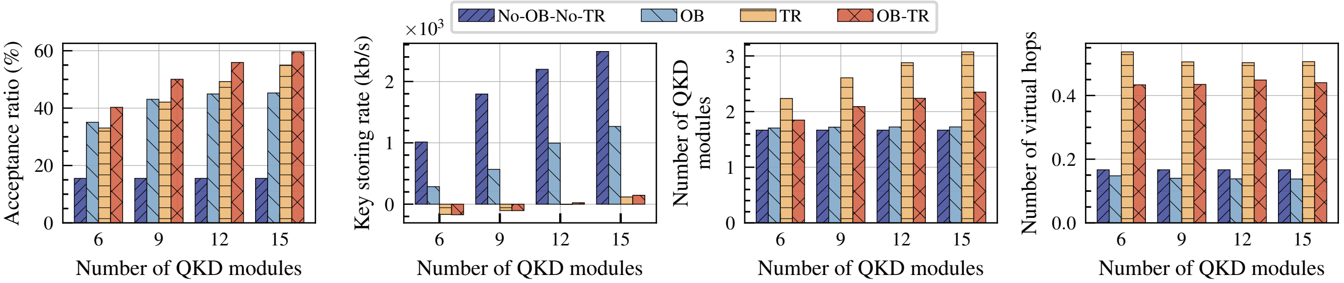
1) Evaluation with Varying Secret Key Rates: We first evaluate the impact of traffic loads on the performance of different network settings. Keys stored in the QKP are 30 kb for all node pairs. Fig. 7 shows the performance of different network settings for varying average requested key rates. Fig. 7(a) shows that *OB-TR* has the highest acceptance ratio while *No-OB-No-TR* has the lowest acceptance ratio. When using *OB-TR*, the acceptance ratio can be improved with a value ranging from 17% to 39% compared to *OB* and a value ranging from 11% to 14% compared to *TR*. Although *OB* and *No-OB-No-TR* have lower acceptance ratios compared to *TR* and *OB-TR*, *OB* and *No-OB-No-TR* can store more keys since fewer keys are used to serve the request. In addition, the key storing rate of *TR* and *OB-TR* is close to 0 as most generated keys are consumed to serve the requests.

The average number of QKD modules used and the average number of virtual hops in each path of serving requests, are shown in Fig. 7(c) and Fig. 7(d), respectively. Fig. 7(c) shows the acceptance ratios of *OB-TR* are obtained with up to 31% fewer QKD modules compared to *TR* and up to 41% additional QKD modules (less than 1 QKD module per path) than *OB*. In addition, *OB-TR* also saves the number of virtual hops used in each path with a value ranging from 5% to 27% compared to *TR*. The average number of quantum modules used in *OB*



(a) Acceptance ratio (b) Key storing rate (c) Number of physical hops (d) Number of virtual hops

Fig. 7: Performance of different network settings vs. load.



(a) Acceptance ratio (b) Key storing rate (c) Number of QKD modules (d) Number of virtual hops

Fig. 8: Performance of different network settings vs. number of QKD modules on each quantum node.

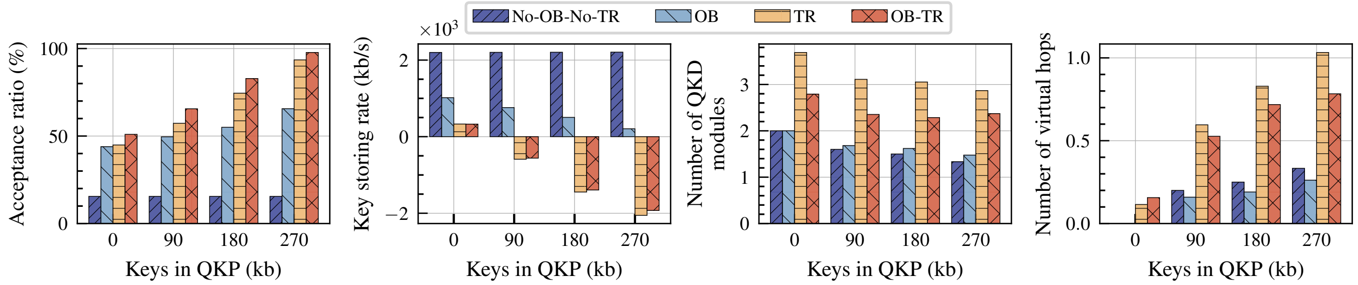
and *No-OB-No-TR* is slightly fewer than 2 since some paths are enabled by keys in the QKP as shown in Fig. 7(d), which do not consume any QKD modules.

In conclusion, by using both *OB* and *TR*, the acceptance ratio of *OB-TR* is up to 39% and 14% higher than *OB* and *TR*, respectively, with respect to *TR*, saves up to 31% of QKD modules and 27% of virtual hops.

2) Evaluation with Varying QKD Modules: We consider an average requested key rate of 15 kb/s. Keys stored in the QKP are 30 kb for all node pairs. Fig. 8(a) evaluates the impact of the number of QKD modules on the acceptance ratio, showing that *OB-TR* achieves the best performance. Note that *OB* works better than *TR* with a small number of QKD modules (6 and 9), while *TR* works better than *OB* when increasing the number of QKD modules. In fact, when the number of QKD modules is small, *TR* requires more QKD modules for each QKD path than *OB*, resulting in fewer QKD paths established to serve requests than *OB*. When the number of QKD modules increases, *TR* accepts more requests because more QKD paths can be established and the achievable key rate of *TR* is higher than *OB*. Specifically, *OB-TR* has a higher acceptance ratio (15% to 32%) and (8% to 22%) higher to *OB* and *TR*, respectively. Note that the acceptance ratio of *OB* is 6% higher than *TR* with 6 QKD modules, while the acceptance ratio of *TR* becomes 22% higher than that of *OB* when the number of QKD modules increases to 15 because *TR* can achieve a higher key rate when the number of QKD modules is not a limiting factor. Apart from the acceptance ratio, the key storing rate increases as the number of QKD modules grows as in Fig. 8(b) since more QKD modules are not used after serving the key rate requests. The reason why the acceptance ratio increases with a larger number of QKD modules is that requests between nodes with longer hops (requiring more QKD modules) can be established. Specifically, as shown in Fig. 8(c), the number of QKD modules of *TR* and *OB-TR* increases by 37% and 27%, respectively when increasing the number of QKD modules from 6 to 15. The average number of QKD modules of *OB* is similar in all cases as *OB* has only one hop. Additionally, Fig. 8(d) shows *OB-TR* can reduce up to 19% of virtual hops compared to *TR* in terms of the average number of virtual hops.

In summary, *OB* shows a better performance than *TR* with a limited number of QKD modules, while *TR* becomes better than *OB* when the number of QKD modules is not a limiting factor because *TR* has a higher key rate.

3) Evaluation with Varying Amount of Keys in QKP: As shown in Fig. 9(a), the acceptance ratio of *OB*, *TR*, and *OB-TR* increases when increasing the keys in QKP, while the acceptance of *No-OB-No-TR* remains the same since the available resources can serve all the adjacent requests without consuming keys in the QKP. Specifically, the acceptance ratio of *OB-TR* is higher than *OB* with a value ranging from 16% to 51%. Moreover, the acceptance ratio of *TR* is only 2% higher than *OB* when no key is stored, while the acceptance ratio becomes 43% higher than *OB* when there are 270 kb keys stored in the QKP of each node pair. This is because, with the *PE* enabled by QKP, more QKD modules can be saved to serve requests between node pairs with longer distances. Regarding the key storing rate, *TR* and *OB-TR* have negative values when the keys stored in QKP before serving requests are 90, 180, and 270 kb since the number of keys used for serving requests is larger than that of the generated keys.



(a) Acceptance ratio

(b) Key storing rate

(c) Number of physical hops

(d) Number of virtual hops

Fig. 9: Performance of different network settings vs. keys stored in QKP for each node pair.

The improvement of the acceptance ratio in Fig. 9(a) is due to the savings of QKD modules and the usage of virtual hops with QKP. As shown in Fig. 9(c), when no key is stored, the number of QKD modules used for each path is 2 for *No-OB-No-TR* and *OB* since no path with *PE* is used. When trusted relay is allowed, the numbers of QKD modules used for each path are 3.7 and 2.8 for *TR* and *OB-TR*, respectively, which are larger than *No-OB-No-TR* and *OB*. When the keys stored in QKP is 270 kb, the number of QKD modules used in each path reduces by 22% and 15% compared to that of the case with no keys stored before serving requests for *TR* and *OB-TR*, respectively. After discussing the number of QKD modules used, we show the average number of virtual hops with different keys stored in QKP before serving requests in Fig. 9(d). When no key is stored before serving requests, no virtual hop is used for *No-OB-No-TR* and *OB* while *TR* and *OB-TR* use 0.11 and 0.16 virtual hops, respectively. When increasing the amount of keys stored in QKP, the average number of virtual hops also increases. For instance, when there are 270 kb keys stored in the QKP, the average number of virtual hops for *TR* and *OB-TR* is 1.03 and 0.78, respectively.

In conclusion, thanks to the keys stored in the QKP, the acceptance ratio of RCKTA improves since it can save the number of QKD modules for each path. Moreover, *TR* can become increasingly better than *OB* with more keys stored in QKP since fewer QKD modules are required for each request.

VII. Conclusion

With the rapid development of quantum technology, it is urgent to design efficient quantum key distribution network schemes to securely share keys. In this work, we investigated a novel problem of Routing, Channel, Key-rate and Time-slot Assignment (RCKTA) to distribute keys resource-efficiently. To solve the RCKTA, we first identified four network settings (with/without optical bypass (OB) and/or trusted relay (TR)) and modeled the achievable key rate between two nodes. Then, we proposed a QKD path scheme to utilize the keys in QKP to establish paths for serving requests. After, we proved the NP-hardness for each network setting and formulated a Mixed Integer Linear Programming (MILP) model for all different settings of OB and TR. Due to the computational limitations of the MILP model, we proposed an auxiliary-graph-based resource-efficient heuristic algorithm. Results show that by allowing OB and TR, *OB-TR* achieves a higher acceptance ratio than that of *OB* and *TR*, respectively. Nevertheless, *OB* works better than *TR* with a limited number of QKD modules while *TR* outperforms *OB* when the number of QKD modules is not the most limiting factor. In the future, we plan to evaluate the RCKTA in a real-time scenario under dynamic traffic.

Appendix A
Limitations of the Secret Key Rate

A suitable simulation tool has been developed to predict the expected performance of the QKD system to be employed in the PoliQI network in the metropolitan area of Milan.

TABLE VI: General Parameters of the QKD Transmission System

Parameter description	Value
Pulse repetition rate	16 MHz
Average photons in signal pulses	0.6 photons/pulse
Probability of signal pulses	70%
MUX/DEMUX in-band loss	2.5 dB
Fiber attenuation @1550nm	0.25 dB/km
Additional optical bypass loss	0.5 dB
Total loss of receiver module	5 dB
Polarization extinction	20 dB
SPAD efficiency	30%
SPAD gate duration	5 ns
Dark count probability	1.5e-5
Maximum number of QKD channels	5

The main parameters used to estimate the secret key generation rate are reported in Tab. VI. In particular, the estimation of the secret key generation rate follows the model described in [23], which is tailored to the discrete-variable BB84 protocol [24]. Assume that we want to share keys between two nodes (referred as Alice and Bob) with point-to-point QKD connection. The secret key generation rate is defined considering that only single photon pulses emitted by Alice are guaranteed to be secure. This requirement provides therefore a lower bound of the real secure key generation rate, since Bob is not able to select only the actual pulses with one photon only but he considers all the signal states. The secure key rate (per signal state emitted by node A) is given by:

$$R = q\{Q_1[1 - H_2(e_1)] - f_e Q_\mu H_2(E_\mu)\} \tag{26}$$

where q depends on the specific implementation of the protocol (in traditional BB84 protocol $q = 1/2$), Q_1 and e_1 are the gain and the error rate of single-photon states, respectively, $f_e > 1$ depends on the inefficiency of the reconciliation procedure, E_μ is the overall quantum bit error rate (QBER) of the signal and H_2 is the binary Shannon entropy. The gains Q_i are defined as the product of the probability of transmitting a i-photon state (with Poisson distribution) and its yield Y_i, indicating the ratio of the number of Bob's detection events (with the correct basis) to the number of Alice's emitted signals. The overall gain of the signal Q_μ is the sum of the gains of all the possible photon-number eigenstates.

References

[1] Y. Cao, Y. Zhao, Q. Wang, J. Zhang, S. X. Ng, and L. Hanzo, "The evolution of quantum key distribution networks: On the road to the qinternet," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 839–894, 2022.

[2] Y. Cao, Y. Zhao, J. Li, R. Lin, J. Zhang, and J. Chen, "Multi-tenant provisioning for quantum key distribution networks with heuristics and reinforcement learning: a comparative study," *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, pp. 946–957, 2020.

[3] P. K. Tysowski, X. Ling, N. Lütkenhaus, and M. Mosca, "The engineering of a scalable multi-site communications system utilizing quantum key distribution (qkd)," *Quantum Science and Technology*, vol. 3, no. 2, p. 024001, 2018.

[4] Y.-A. Chen, Q. Zhang, T.-Y. Chen, W.-Q. Cai, S.-K. Liao, J. Zhang, K. Chen, J. Yin, J.-G. Ren, Z. Chen *et al.*, "An integrated space-to-ground quantum communication network over 4,600 kilometres," *Nature*, vol. 589, no. 7841, pp. 214–219, 2021.

[5] A. Gatto, M. Brunero, M. Ferrari, A. Tarable, D. Bodanapu, J. P. Brito, R. B. Mendez, R. J. Vicente, F. Bianchi, M. Frittelli *et al.*, "A bb84 qkd field-trial in the turin metropolitan area," in *Photonics in Switching and Computing*, 2021, pp. Tu1A–2.

[6] R.-h. Shi and H. Yu, "Privacy-preserving range query quantum scheme with single photons in edge-based internet of things," *IEEE Transactions on Network and Service Management*, 2023, early access.

[7] C. Lee, I. Sohn, and W. Lee, "Eavesdropping detection in bb84 quantum key distribution protocols," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2689–2701, 2022.

[8] Y.-F. Pu, S. Zhang, Y.-K. Wu, N. Jiang, W. Chang, C. Li, and L.-M. Duan, "Experimental demonstration of memory-enhanced scaling for entanglement connection of quantum repeater segments," *Nature Photonics*, vol. 15, no. 5, pp. 374–378, 2021.

[9] K. Dong, Y. Zhao, X. Yu, A. Nag, and J. Zhang, "Auxiliary graph based routing, wavelength, and time-slot assignment in metro quantum optical networks with a novel node structure," *Optics Express*, vol. 28, no. 5, p. 5936, Mar. 2020.

[10] 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 Information*, vol. 7, no. 1, pp. 1–6, 2021.

[11] M. Mehic, P. Fazio, S. Rass, O. Maurhart, M. Peev, A. Poppe, J. Rozhon, M. Niemiec, and M. Voznak, "A novel approach to quality-of-service provisioning in trusted relay quantum key distribution networks," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 168–181, 2019.

[12] A. Wonfor, C. White, A. Bahrami, J. Pearse, G. Duan, A. Straw, T. Edwards, T. Spiller, R. Penty, and A. Lord, "Field trial of multi-node, coherent-one-way quantum key distribution with encrypted 5× 100g dwdm transmission system," in *European Conference on Optical Communication (ECOC)*, 2019, pp. 1–4.

[13] R. Schwonnek, K. T. Goh, I. W. Primaatmaja, E. Y.-Z. Tan, R. Wolf, V. Scarani, and C. C.-W. Lim, "Device-independent quantum key distribution with random key basis," *Nature Communications*, vol. 12, no. 1, p. 2880, 2021.

[14] X. Tang, A. Wonfor, R. Kumar, R. V. Penty, and I. H. White, "Quantum-safe metro network with low-latency reconfigurable quantum key distribution," *IEEE/OSA Journal of Lightwave Technology*, vol. 36, no. 22, pp. 5230–5236, 2018.

[15] Q. Zhang, O. Ayoub, A. Gatto, J. Wu, X. Lin, F. Musumeci, G. Verticale, and M. Tornatore, "Joint routing, channel, and key-rate assignment for resource-efficient qkd networking," in *IEEE Global Communications Conference (GLOBECOM)*, 2022, pp. 1–6.

[16] E. E. Moghaddam, H. Beyranvand, and J. A. Salehi, "Resource allocation in space division multiplexed elastic optical networks secured with quantum key distribution," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 9, pp. 2688–2700, 2021.

[17] Q. Zhu, X. Yu, Y. Zhao, A. Nag, H. Wang, L. Chen, and J. Zhang, "Auxiliary graph based qkd key provisioning for end-to-end security service in optical networks," in *Optical Fiber Communication Conference*, 2022, pp. Th3D–2.

[18] Y. Cao, Y. Zhao, J. Zhang, and Q. Wang, "Software-defined heterogeneous quantum key distribution chaining: An enabler for multi-protocol quantum networks," *IEEE Communications Magazine*, vol. 60, no. 9, pp. 38–44, 2022.

[19] "Rigene Project - POLIQI - POLItecnico Quantum Infrastructure." [Online]. Available: https://sites.google.com/view/rigene/news/poliqi-politecnico-quantum-infrastructure

[20] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath communications: An approach to high bandwidth optical wan's," *IEEE Transactions on Communications*, vol. 40, no. 7, pp. 1171–1182, 1992.

[21] L. Zhang and S. Geng, "The complexity of the 0/1 multi-knapsack problem," *Journal of Computer Science and Technology*, vol. 1, no. 1, pp. 46–50, 1986.

[22] Y. Cao, Y. Zhao, J. Li, R. Lin, J. Zhang, and J. Chen, "Hybrid trusted/untrusted relay-based quantum key distribution over optical backbone networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 9, pp. 2701–2718, 2021.

[23] H.-K. Lo, X. Ma, and K. Chen, "Decoy state quantum key distribution," *Physical review letters*, vol. 94, no. 23, p. 230504, 2005.

[24] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical Computer Science*, vol. 560, pp. 7–11, 2014.