Accepted Manuscript

Utility-optimized bandwidth and power allocation for non-orthogonal multiple access in software defined 5G networks

Xiaohong Huanga, Tingting Yuan, Yan Zhang

PII: S1084-8045(18)30093-6

DOI: 10.1016/j.jnca.2018.03.012

Reference: YJNCA 2093

To appear in: *Journal of Network and Computer Applications*

Received Date: 29 October 2017

Revised Date: 25 January 2018

Accepted Date: 12 March 2018

JOURNAL OF NETWORK AND COMPUTER APPLICATIONS

Utility-optimized Bandwidth and Power Allocation for Non-orthogonal Multiple Access in Software Defined 5G Networks

Xiaohong Huang[a], Tingting Yuan[a], Yan Zhang[b*]

[a]*Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China*

[b]*Department of Informatics, University of Oslo, Norway*

Elsevier use only: Received date here; revised date here; accepted date here

Abstract

The fifth Generation (5G) wireless network will be in very high carrier frequencies with massive bandwidths, extreme base stations and device densities. For optimization of 5G network, it is of vital importance to design end to end resource allocation mechanism across the radio access network and core network. In this paper, we propose an algorithm for joint resource optimization across wireless links and core networks. Our goal is twofold. First, Non-Orthogonal Multiple Access (NOMA) technique is considered for performance enhancement in 5G network, and Software Defined Network (SDN) technique is considered for resource allocation in core network. Second, the joint network utility optimization algorithm with multipath is proposed through proper power allocation across access network in conjunction with bandwidth allocation across core network, in which utility is an important indicator to reflect the level of user or application satisfaction and fairness of the allocated bandwidth. Extensive simulation has been done to evaluate the performance of proposed algorithm. The results show that the proposed joint optimization algorithm is able to achieve significant improvement in terms of network utility with more SDN devices deployed.

Keywords: 5G; NOMA; SDN; Bandwidth Allocation; Power Allocation; Multipaths

1. Introduction

The fifth Generation (5G) (Andrews et al., 2014) is a promising technology to offer significant improvements in terms of network coverage and user experience. The research about the key technologies, significant features, frameworks and key challenges of 5G are being identified in (Agiwal et al., 2016; Chen and Zhao, 2014; Wang et al., 2014; Demestichas et al., 2013). With a large number of BSs connecting to the core network, the success of optimization for 5G network will mainly depend on the joint resource provision of radio access network and core network. Thus, it is important to study the problem to offer end to end resource allocation mechanism jointly considering power allocation in

* Corresponding author. e-mail: yanzhang@ifi.uio.no.

wireless links and bandwidth allocation over the core network.

From the access network point of view, in order to support high data rate communications, huge number of connected devices, ultra-low latency and high reliability application, Non-Orthogonal Multiple Access (NOMA) (Islam et al., 2017) is adopted in next generation wireless network instead of Orthogonal Multiple Access techniques (OMA). With NOMA techniques, the resources can be shared by all the users simultaneously, which however will lead to inter user interference. Therefore, efficient algorithm is required to allocate subcarriers and power to achieve optimal performance.

From the core network point of view, the Software Defined Network (SDN) (Nunes et al., 2014; Masoudi and Ghaffari, 2016) has been identified as a promising architecture. With the separation of control plane and data plane, SDN offers flexible and convenient way for fine-grained network resource allocation with multipath. Software defined 5G network integrating SDN and 5G has been proposed in (Trivisonno et al., 2015; Sun et al., 2015). However, due to operational and economic constraints, SDNs will not be fully deployed in a short-term. This is particular true in a large-scale core network (Hong et al., 2016). The hybrid SDN (Vissicchio et al., 2014; Caria et al., 2015) with the legacy forwarding devices and SDN-enabled devices co-existing is an important scenario. In order to optimize the performance of Software Defined core network, the network utility (Kuo and Liao, 2008; Tan et al., 2005) is adopted to reflect the level of satisfaction (Kuo and Liao, 2008; Tan et al., 2005), fairness (Jain et al., 2013; Mo and Walrand, 2000; Feng et al., 2015) of the allocated bandwidth. It should be noticed that the proper bandwidth allocation in hybrid SDNs will improve the network utility. Thus, efficient algorithm is required to allocate the bandwidth to achieve optimized performance.

Different from the classical problem in the access network and the core network, end-to-end path in a wireless cellular network consists of a wireless air-interface and a path in wired core network. So far, most of the existing work considers the resource allocation in access network and core network separately. Only some work studies joint resource allocation mechanisms taking both wireless links and core network paths into account. In (Chiang and Bell, 2004), an end to end resource allocation has been proposed to handle the joint resource allocation problem across OMA based access network and the traditional IP core network. Some other end to end resource allocation mechanisms have been proposed in (Liao et al., 2014), (Lee et al., 2007) and (Lei et al., 2015) for software defined radio access network, NOMA-based access network and wireless ad-hoc network respectively. However, none of the papers consider the joint resource allocation problem across the NOMA-based radio access network and SDN-based core network.

In this paper, we propose a joint optimization algorithm considering network bandwidth allocation in core network and power allocation in access network. The objective is to maximize the network utility over the joint solution of power and bandwidth allocation. To the best of our knowledge, this is the first optimization solution for bandwidth and power allocation in the software defined 5G network. Our major contributions can be summarized as follows.

• A flow-level optimization algorithm for core bandwidth allocation and wireless power allocation over subcarries is proposed which jointly considering the capacity of access network and core network in the software defined 5G network.

• Joint optimization problem of flow control and routing is considered since multiple paths are available in SDN network. The algorithm works in both pure and hybrid SDN.

• The end to end problem decomposition method is proposed to solve the joint optimization problem. For the first subproblem, a greedy water-filling algorithm is proposed to schedule the power allocation over subcarries. For the second subproblem, the bandwidth allocation of the core network is solved using the Karush-Kuhn-Tucker (KKT) conditions by a gradient way.

The rest of the paper is organized as follows: related work about the resource allocation problem in the software defined 5G network is summarized in section 2. In section 3, system framework is described. Besides, a case study is given as an example. Section 4 formulates the resource allocation problem with end to end multipath. Section 5 presents problem decomposition method to solve the joint optimization problem. Based on it, two algorithms are proposed respectively to solve two subproblems. Section 6 shows the performance evaluation with different topologies in terms of utility. In section 7, a conclusion is derived with a summary.

2. Related Work

The related work about resource allocation in access network, core network, and joint resource allocation mechanisms are summarized in this section.

2.1. Resource Allocation Problem in the Access Network

The resource allocation problem in access network is closely related with the wireless access technologies. In (Tan et al., 2015), authors proposed resource allocation in wireless access network with multiple users in a single cell, and the different type of applications are taken into their consideration. The interference is ignored in single-cell when OMA is used. In (Chiang and Bell, 2004), the authors summarized utility maximization methods over powers and rates in wireless cellular network, including single-cell and multi-cells. For muti-cells, the interference is between the base stations. However, when the NOMA is used in a 5G network which removes the resource allocation exclusivity and allows more than one user to share the same sub-carrier, the inter flow interference can't be ignored. In (Al-Imari et al., 2014), a NOMA-based iterative subcarrier and power allocation scheme for uplink is proposed. But this scheme doesn't consider the capacity of the core network.

2.2. Bandwidth Allocation Problem in the Core Network

In traditional communication core networks, the bandwidth allocation problem to maximize the network utility has been extensively studied considering single path and multipath scenarios (Habib et al., 2016). On one hand, the single path utility maximization problem, where only one path can be used by each user or source-destination pair, has been studied in (Kelly et al., 1998; La and Anantharam, 2002; Chiang et al., 2007). On the other hand, utility maximization problem with multipath, where more than one path can be used by some users or source-destination pairs, has also been studied in (Lin and Shroff, 2006; Jin et al., 2009; Wang et al., 2003). Bandwidth allocation with multipath usually has better performance than that with single path, nevertheless, the bandwidth allocation in traditional communication core networks with multipath is not feasible and controllable enough.

In the software defined core network, bandwidth allocation problem can be solved at a more fine-grained level over multiple paths with a centralized control. The bandwidth allocation and the chosen paths are under the management of the SDN controllers. The existing work on bandwidth allocation of the SDNs focuses on two aspects. The first aspect is to achieve high network resource utilization and network throughput from the network' point of view. In (Agarwal et al., 2013; Levin et al., 2013; Guo et al., 2015) the routing and bandwidth allocation is to minimize the maximum utilization of the network over multipath. The second aspect is to achieve fairness in the bandwidth allocation from the flows' point of view. The fair criterion is used for bandwidth allocation in SDNs which includes max-min fairness policy (Jain et al., 2013), α-fair policy (Eghbali and Wong, 2015) and proportional fairness policy (Feng et al., 2015) etc. Thus, bandwidth allocation in the software defined core network with multipath can offer a good chance for utility improvement.

2.3. Allocation Problem with End to End Paths

Each end-to-end path in a wireless cellular network consists of a wireless air-interface and a path in wired core network. The end to end resource allocation has been proposed in (Chiang and Bell, 2004) which considers both the OMA based access network and the traditional IP core network together. In the traditional IP core network, only one end-to-end path of one source-destination pair is taken into consideration. In (Liao et al., 2014), the min flow rate maximization is proposed with software defined radio access network. In (Lee et al., 2007), the power scheduling and end-to-end rate control for wireless ad-hoc networks is proposed based on the assumption that each user has its fixed routing path for communication. However, the core network considered in these papers are traditional IP network, thus single routing path is mainly considered. With SDN technology, the multipaths routing of the core network should be considered.

The software defined based 5G network consists radio access network and core network. Two main characteristics should be considered in resource allocation algorithm. First, in access network, resource should be shared by all the flows and users. Second, multipath scenarios should be considered when SDN is deployed, even though SDN is partly deployed. In order to achieve end to end resource allocation, the resource allocation with NOMA in radio access network and bandwidth allocation in core network are highly correlated problem. The base stations allocate subcarriers and power to flows, which will in-turn affect the bandwidth allocation in core network. So far, this joint optimization problem has not been well studied yet.

3. System framework and Problem Analysis

In this paper, a resource allocation problem in the software defined based 5G network is considered which consists of mobile terminals, base stations and the SDN-based core network. The software defined based 5G network scenario, bandwidth allocation process and a study case are described in this section.

3.1. Framework of Software defined 5G network

The software defined based 5G network consists of two parts including the access network and the core network as shown in Figure 1. In this paper, NOMA technique is used to handle resource allocation in wireless access network. Hence, the access network is made up of base stations which are assumed to be out of control of SDN controllers. The core network is a hybrid SDN, in which the legacy forwarding devices and SDN-enabled devices are co-existing. Compared with the legacy OSPF-based core networks, multiple paths can be used to allocate bandwidth in the hybrid software-defined core network. A SDN-enabled device is able to choose several next-hops for any given destination under the control of the SDN controller. However, the legacy devices can only use the next hop of the least-cost path because they are not SDN enabled.

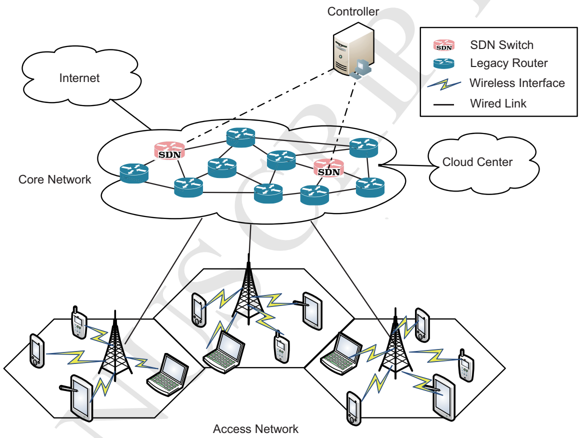


Figure 1. A framework of software defined based 5G network

In the access network, NOMA is adopted instead of OMA. With NOMA technique, the resources can be shared by all the flows simultaneously, which however will lead to inter flow interference. In this paper, we only discuss the case of data sending from mobile users. In the core network, the topology discovery is indispensable in the SDN controller. In this scenario, to exchange link status information with legacy nodes, the SDN-enabled devices will run legacy routing algorithm to forward link-state advertisements (LSA) message. In this way, the legacy devices can detect the links of SDN-enabled devices. Then, the legacy network will do hop-by-hop routing using a standard routing protocol like OSPF, using the information of the links in the OSPF link state Data Base (LSDB). With the Link Layer Discovery Protocol (LLDP), the Broadcast Domain Discovery Protocol (BDDP) and the link information of the legacy routing protocol, such as link-state advertisements (LSAs), the SDN controller has the ability to obtain entire network information, including the network topology and the metrics of links. Then, the controller will calculate the paths for SDN enabled devices. The detailed process has been summarized in (Hong et al., 2016; Pakzad et al., 2016).

3.2. Resource Allocation Problem in the Software defined 5G Network

The resource allocation in software defined 5G network is based on flows of mobile devices. Each flow is with a source, a destination and a demand expressed by an upper bound and lower bound of its data transmitting rate as a commodity flow. The end-to-end paths consists of a wireless access point and one or more wired paths. So, the problem is how to allocate power in wireless links over subcarriers and bandwidth in wired paths.

Two issues shall be addressed for end-to-end joint optimization. The first issue is how to allocate the subcarriers to flows and the power of flows in each subcarrier. As addressed above, the subcarrier is shared by mobile devices. The second issue is how to select a proper path or paths and how much bandwidth should be allocated for each flow in the core network. With hybrid SDNs, multipath scenario shall be considered in bandwidth allocation. These two issues are highly correlated, which also presents a trade-off between the rate-power allocation of wireless links and bandwidth allocation of wired links. If the sending rate of access network exceed the resource limitation of core network, congestion will happen in the core network. However, if the sending rate of access network is smaller than the allocated bandwidth in core network, the resources will be wasted. Consequently, joint optimization model is required to achieve optimal tradeoff.

3.3. A Case Study of Optimization with Multipath in SDN Network

Figure 2 depicts the topology of the network for a case study. There are 9 forwarding devices with 2 SDN ones in the core network. Node C and node K are SDN forwarding devices which are under the control of the SDN controller. The links are unidirectional, and bandwidth capacity is marked on it as shown in Figure 2. There are four flows: *f1*, *f2*, *f3* and *f4*. The flow *f1* and the flow *f2* are from the base station S1 to the cloud center which is connected with node D. The flow *f3* is from the base station S3 to node D, and the flow *f4* is from the base station S2 to the base station S1.

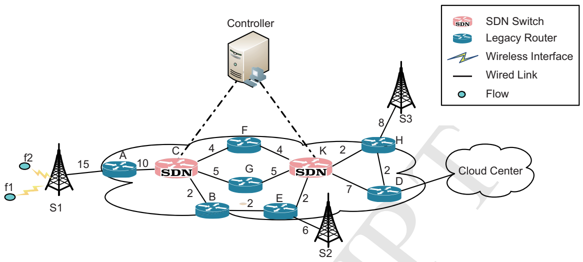


Figure 2. A hybrid SDN network

For this case study, if Figure2 is a legacy network, the traffic from the base station S1 to node D can only use the OSPF path (S1 → A → C→ G → K → D). If the node C is a SDN-enabled device, it is able to choose the next-hops from nodes {F, G, B}. So, two new controllable paths, S1→A→C→F→K→D and S1→A→C→B→E→K →D can be used for flows from base station S1 to node D. The maximum bandwidth that can be allocated to the flows from node S1 to node D is increased from 5-units to 7-units when node C is a SDN forwarding device. Similarly, if both the C and the node K are SDN enabled, three more paths S1 → A →C → G →K → H →D, S1 → A → C → F →K → H → D and S1 → A → C → B → E → K → H → D can be used for flows from the base station S1 to the node D. Therefore, the maximum bandwidth is increased to be 9-units for the divisible flows from the base station S1 to the node D. For example, if only the flow *f1* exits in the network, its maximum allocated bandwidth is 9-units. If the flow *f1* and the flow *f2* are co-existing, they will share the bandwidth with each other based on their demand and utility functions. However, the maximum bandwidth which can be allocated is still 9-units.

3.4. Resource Allocation Implementation in Software Defined 5G Network

The way to implement the flexible flow-level bandwidth allocation in the software defined 5G network is introduced in this part. Figure 3 is used for illustration of the framework of resource allocation in software defined 5G network, which is flow-level based, flexible and end-to-end multipath allocation.

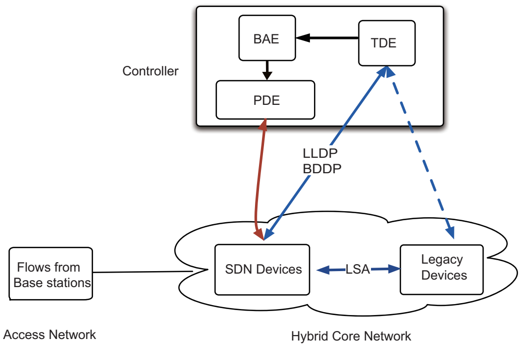


Figure 3. The framework of resource allocation in the software defined based 5G network

The process of resource allocation in the software defined 5G network contains four important steps. First, the topology discovery element (TDE) uses LLDP, BDDP and LSA to discover the entire topology. Second, the band- width allocation element (BAE) finds all the candidate paths using the topology information from the TDE, and computes how much bandwidth should be allocated to each flow in each candidate path. The SDN nodes that the candidate path passes through shall be recorded as well. Third, the path deployment element (PDE) reserves bandwidth for flows in the SDN devices. In this framework, the flows from the base station will decide its sending rate with power and sub carriers allocation, and these will be the input of the system. Then, the bandwidth allocation of the core network, and the power and sub carriers allocation of the access network should be outputted from the system according to the input.

4. Problem Formulation

In SDN-based 5G network, multiple paths can be used for some flows to improve network utility according to the analysis above. In this section, we formulate the end to end multipath bandwidth allocation problem as an optimization problem to maximize the network utility in both access and core network. The utility function can be used to illustrate the bandwidth allocation performance which is hypothesized to be continuous and concave with the allocated bandwidth. Each end-to-end path in a software defined 5G network consists of a wireless air-interface and a wired path of the core network. Thus, the constraints of this problem consist of two aspects: the access network and the core network. The parameters and variables used in our model are summarized in Table 1.

Table 1 List of notations

Parameters	Meaning
N, E	The set of all forwarding devices and the set of physical links between them.
W	The set of base stations.
F	The set of flows. The number of flows is m.
$\tilde{T}_f$	The total power for flow f.
K_f	The set of subcarriers which can be used by the flow f. Its number is $\mid K_f \mid$.
B_e	The bandwidth capacity of the link e.
B_w	The system bandwidth of w ∈ W.
$h_{f,k}$	The channel gain of the flow f on the k-th subcarrier.
σ^2	The noise power per sub carrier.
P_f	The set of candidate paths for the flow f.
n_f	The number of candidate paths for the flow f.
$\pi^p{}_e$	Boolean to determine if link e is in the path p.
d_f, D_f	The lower bound and the upper bound bandwidth of flow f.
Variables	**Meaning**
$x^p{}_f$	The bandwidth allocated to the flow f of the path p.
$T_{f,k}$	The power allocated to flow f on the subcarrier k.

From the access network point of view, the main constrains are the limited power and the shared subcarries when NOMA is used. The vector of the allocated power is defined by:

$$\mathbf{T}=\left\{T_{f,k} \mid f \in F, k \in K_f\right\}=[T_{f_1,k_1}, T_{f_1,k_2}, \ldots, T_{f_m,k_{|K_{f_m}|}}].$$

where $T_{f,k}$ is the allocated power of flow f in subcarrier k. In the access network, each flow has a power budget that should be divided into multiple parts to different subcarriers within its range. So, it should be ensured that the allocated transmit power is not beyond its budget.

$$\sum_{k \in K_f} T_{f,k} \le \tilde{T}_f, \forall f \in F. \tag{1}$$

From the core network point of view, the constrains are the shared and limited capacity of wired links. As analyzed above, multiple paths can be used in SDN-based core network. Thus, the first step is to find all available paths. However, not all these paths can be used to route flows. The reason is that in the future 5G systems the delay requirements are stringent. It's justified with the fact that a flow routed by a path with more hops or more path cost implies more bandwidth occupation, decreased network utility and higher user satisfaction. According to the QoS requirement and delay requirement of 5G network, some paths are chosen. The set of candidate paths of a flow f is defined by $\forall f \in F \ : P_f \ = \left[p_{f,1}, \ p_{f,2}, ..., \ p_{f,n_f} \right]$. Based on the candidate paths, the vector of the allocated bandwidth to all flows of all their candidate paths is defined by:

$$\mathbf{x} = \left\{ x_f^p \mid f \in F, p \in P_f \right\}$$
$$= [x_{f_1}^{p_{f_1,1}}, ..., x_{f_1}^{p_{f_1,n_{f_1}}}, ...x_{f_n}^{p_{f_n,1}}, ..., x_{f_n}^{p_{f_n,n_{f_n}}}].$$

In the wired core network where links have fixed 'sizes', the main issue is to avoid overloading the links' capacity. The goal is to prevent any flows from 'pumping' so much data into the network that the total traffic exceeds the link capacity. In the core network, every link has a limited capacity of bandwidth. The inequality (2) ensures that the total amount of the traffic over a link is less than its capacity.

$$\sum_{f \in F} \sum_{p \in P_f} x_f^p \pi_e^p \le B_e, \forall e \in E. \tag{2}$$

When jointly considered the resource allocation in software define 5G network, the inequality (3) is defined to ensure the reservation bandwidth for flows should be fully used to avoid the surplus bandwidth allocation in the core network. The inequality is:

$$\sum_{p \in P_f} x_f^p \le R_f, \forall f \in F. \tag{3}$$

where R_f is the sending rate of flow f from the access network. In NOMA-based 5G access network, it can be assumed that flows are decoded in an increasing order of their indices. Hence, when the first flow ($s_f = 1$) is decoded, its interference is from all the other flows $j = 2, ..., S_f$. Similarly, the second flow ($s_f = 2$) to be decoded will see interference from the flows $j = 3, ..., S_f$, and so on. Thus, the interference ($I_{f,k}$)of each flow on each sub-carrier with this decoding order will be:

$$I_{f,k} = \sum_{j=s_f+1}^{S_{w_f}} T_{j,k} h_{j,k},$$

where w_f denotes the base station of f, S_{w_f} denotes the number of flows under this base station and S_f is the index of flow f. Considering the interference of flows, the sending rate of flow f:

$$R_f = \sum_{k \in K} B_k \log(1 + \frac{T_{f,k} h_{f,k}}{I_{f,k} + \sigma^2}), \forall f \in F,$$

where B_k is the bandwidth of the subcarrier k.

Utility function is always used to illustrate the satisfaction of the allocated bandwidth. According to (Jin et al., 2009), it is assumed that all the utility functions of flows are continuous, increasing and strictly concave. According to the analysis above, the joint problem for core bandwidth allocation and wireless power allocation is formulated as follows:

$$\max \sum_{f \in F} a_f U_f (\sum_{p \in P_f} x_f^p)$$

Subject to

$$(1), (2), (3),$$
$$d_f \le \sum_{p \in P_f} x_f^p \le D_f, \forall f \in F, \tag{4}$$
$$x_f^p \geqslant 0, \forall f \in F, p \in P_f, \tag{5}$$
$$T_{f,k} \geqslant 0, \forall f \in F, k \in K_f. \tag{6}$$

In the formulation, the variables are $\mathbf{x}$ and $\mathbf{T}$. The a_f of utility function can be viewed as the priority of flow f. The bandwidth which is reserved and allocated to a flow should be in its range of the bandwidth demand (4). Besides, the allocated bandwidth on a candidate path is non-negative (5). The transmit power of flow over a subcarrier is non-negative (6). So, three more restrictive conditions are added in this formulation.

Jointly consider this resource allocation problem in software defined 5G network could improve the network utility. However, this problem is difficult to solve because it's a nonconvex problem with the constrains of (3). So, we proposed a method to decompose the joint problem into two subproblems.

5. Algorithm

In this section, a method to decompose the joint resource allocation problem is proposed. Two algorithms are proposed to solve these two subproblems respectively.

5.1. Problem Decomposition

In order to decompose the joint problem, new dummy variables, which is the total bandwidth reserved for every flow f, are defined:

$$X_f = \sum_{p \in P_f} x_f^p \,. \tag{7}$$

Then, the inequality (3) can be formed as:

$$X_f \le R_f \,. \tag{8}$$

According to the problem formulation, the partial Lagrange function with constraint (8) for every flow is defined by:

$$\begin{aligned} L(\mathbf{x}, \mathbf{T}, \boldsymbol{\varphi}) &= \sum_{f \in F} a_f U_f (X_f) + \sum_{f \in F} \varphi_f (R_f - X_f) \\ &= \sum_{f \in F} \varphi_f R_f + \left(\sum_{f \in F} a_f U_f (X_f) - \sum_{f \in F} \varphi_f X_f \right). \end{aligned}$$

where $\varphi = [\varphi_1, \varphi_2, ..., \varphi_F]^T$ are introduced Lagrange multipliers, which are not nonnegative. φ_f can be seen as the penalty price of flow f for the surplus reserved bandwidth in the core network. To maximize the partial Lagrange, the Lagrange dual function can be obtained:

$$G(\varphi) = \sup_{\mathbf{x}, \mathbf{T}} \{ L(\mathbf{x}, \mathbf{T}, \varphi) \mid (1), (2), (4), (5), (6), (7) \} \,.$$

The Lagrange dual function can be partial decomposed into two parts:

$$G(\varphi) = G_a(\varphi) + G_c(\varphi) \,.$$

where $G_c(\varphi)$ and $G_a(\varphi)$ are the optimized value for the access network and the core network.

The optimization problem can be reformulated into two subproblems. The first subproblem is about access network shown as follows.

$$G_a(\varphi): \quad \begin{array}{ll} \max & \sum_f \varphi_f R_f \\ \text{subject to} & (1), (6) \\ & R_f \ge d_f \end{array} \,.$$

where the last constraints are from (4) and (8). The variables are T and φ. The second subproblem is about core network shown as followed.

$$G_c(\varphi): \quad \begin{array}{ll} \max & \sum_{f \in F} (a_f U_f (X_f) - \varphi_f X_f) \\ \text{subject to} & (2), (4), (5), (7) \end{array} \,.$$

where the variables are x and φ.

According to the problem decomposition, the algorithm outline is shown in Algorithm 1. The penalty price of surplus bandwidth allocation should be updated in a gradient way which is given by line 4. The symbol $[\cdot]^+$ is the projection of [0, ∞) which is defined by $[z]^+ = max\{0, z\}$. r_1 is a sufficiently small positive step-size.

Algorithm 1 Power and bandwidth allocation for end to end path allocation

1: Repeated in parallel by iterations until convergence:

2: The flow mangers (Access Network):

Power and sub-carrier allocated of the access network is done by flow manages (Algorithm 2).

The sending rates of flows R_f are computed according to the allocated power and the selected sub-carriers.

3: The SDN controllers (Core network):

The task of SDN controller is bandwidth allocation for flows with multipaths (Algorithm 4).

4: The penalty price update:

$$\varphi_f (t+1) = [\varphi_f (t) - r_1 (R_f - X_f)]^+ .$$

Figure 4 is used as an example to illustrate the mechanism of the proposed algorithm. When the rate of flow R_f is less than the allocated bandwidth X_f, φ_f is dropped until zero. When R_f is more than X_f, φ_f grows. With grown φ_f, the X_f is decreased as shown in Figure 4 where the logarithmic function is used for example as the utility function. When φ_f grows from 0.1 to 0.2, the X_f which can bring the maximum of $G_a(\varphi)$ drops from X_f^1 to X_f^2, and vice versa. Higher φ_f stands for the fact that more bandwidth has been allocated to this flow which is underused. For optimization $G_a(\varphi)$, the power and the better sub-carriers will be allocated to the flow with high φ_f .When the more rate of flows is achieved, the value of φ_f will be dropped after no overplus of allocated bandwidth for flow *f*.

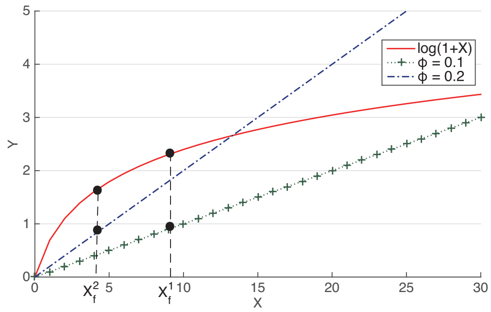


Figure 4. Relationship between φ and the allocated bandwidth

5.2. Algorithm for the Subproblem of Access Network

Because of the interference between flows, it's hard to obtain the optimal power and subcarriers allocation. So, the greedy power allocation is proposed for suboptimal $G_a(\varphi)$. The proposed algorithm includes five main steps: (1) Find the candidate flows. The candidate flows are those need to be allocated with power. The flows with bandwidth which is less than its lower bound (d_f) should be allocated power with priority. If flows with sending rate more than D_f, it shouldn't be allocated during this iteration. (2) Each flow in the candidate flow set $\tilde{F}$ performs the Water-Filling Algorithm (Algorithm 3) to compute the power to be allocated in each iteration. The Water-Filling Algorithm includes two steps. The first step is to compute water-filling level according to the remaining power. The second step is to compute the allocated power over each subcarrier. (3) The sub-carrier which supports the most sending rate should be chosen in each iteration. (4) The flow with the maximum objection $\varphi_f R_f$ will be chosen as the best flow and to be allocated the sub-carrier and power. (5) Finally, the interference should be updated based on NOMA. The process will repeat until no power can be allocated or all the flows has been satisfied with the upper bound (D_f).

Algorithm 2 Iterative Sub-carrier and Power Allocation of the Access Network

1: Initialization: $R_f = 0, T_{f,k} = 0 \quad \forall f \in F$.

2: Repeat:

3: Candidate Flow Set Finding:

The remaining power of flow $f : \omega_f = \tilde{T}_f - \sum_{k\in K_f} T_{f,k}$.

The set of flows still have power to allocated:

$$F' = \{ f \mid f \in F, \omega_f > 0 \}.$$

The sets of flows are defined according to the bandwidth:

$$F_d = \{ f \mid f \in F', R_f(t) < d_f \},$$

$$F_{dD} = \{ f \mid f \in F', d_f \le R_f(t) < X_f \},$$

$$F_D = \{ f \mid f \in F', R_f(t) \ge X_f \}.$$

The set of candidate flows:

$$\tilde{F} = \begin{cases} F_d & \text{if } F_d \neq \varnothing, \\ F_{dD} & \text{if } F_d = \varnothing, F_{dD} \neq \varnothing, \\ \varnothing & \text{if } F_D == F. \end{cases}$$

4: Allocated Power Computation:

Each flow in $\tilde{F}$ performs Water-Filling Algorithm (Algorithm 3) over all the available sub-carrier.

The rate of flow f in sub-carrier

$$k : R_{f,k} = B_k \log_2(1 + \frac{T_{f,k}\lambda_{f,k}}{\sigma^2 + I_{f,k}}).$$

5: Sub-carrier Selection:

For each flow, the best sub-carrier k_f^* is found with:

$$k_f^* = \arg\max_{k\in K} R_{f,k}, \forall f \in \tilde{F}.$$

6: Sub-carrier and power Allocation:

$$f^* = \arg\max_{f\in \tilde{F}} \varphi_f R_{f,k_f^*}.$$

Allocate the sub-carrier $k_{f^*}^*$ and power T_{f^*,k^*} to the flow f^*.

7: Update the interference I_{f,b_f} .

8: **until** $R_{f,k} = 0, \forall f \in F, k \in K_f or \tilde{F} = \varnothing$.

Algorithm 3 Water-Filling Algorithm

1: The water-filling level: $\gamma_f = \dfrac{\omega_f + \sum_{k\in K} \dfrac{\sigma_k^2 + I_{f,k}}{\lambda_{f,k}}}{K_f}$.

2: The power of flow f in sub-carrier k:

$$T'_{f,k} = \left[\gamma_f - \frac{\sigma_s^2 + I_{f,k}}{\lambda_{f,k}}\right]^+, \forall k \in K_f,$$

$$T_{f,k} = T_{f,k} + T'_{f,k}.$$

5.3. Algorithm for the Subproblem of Core Network

Firstly, the Lagrange function of the subproblem is formulated. Secondly, the Karush-Kuhn-Tucker (KKT) approach is used to get the optimal result. Thirdly, we introduce the ideas of sub-gradient method used in (Lin and Shroff, 2006) to allocate bandwidth to flows over multiple paths.

According to the subproblem formulation of core network $G_c(\boldsymbol{\varphi})$, the Lagrange function is defined by:

$$L_c(\mathbf{x}, \varphi, \overline{\lambda}, \underline{\lambda}, \xi, \mu) = \sum_{f\in F} a_f U_f(X_f) - \sum_f \varphi_f X_f + \sum_{f\in F} \overline{\lambda}_f (D_f - X_f) + \sum_{f\in F} \underline{\lambda}_f (X_f - d_f) + \sum_{f\in F}\sum_{p\in P_f} \xi_{f,p} x_f^p - \sum_{e\in E} \mu_e \left(\sum_{p\in P} x_f^p \pi_e^p - B_e\right).$$

The μ_e can be viewed as the price of link *e*. The price of a path *p* can be defined as:

$$C_p = \sum_{e\in E} \mu_e \pi_e^p.$$

The allocated bandwidth of a link is denoted as:

$$X^e = \sum_f \sum_{p\in P_f} \pi_e^p x_f^p.$$

The optimal solution of **X** must satisfy the KKT conditions which are shown as follows:

$$a_f U_{f'}(X_f) - \varphi_f - \overline{\lambda}_f + \underline{\lambda}_f = C_p - \xi_{f,p}, \quad (9)$$

$$\overline{\lambda}_f \left(D_f - X_f\right) = 0, \quad (10)$$

$$\underline{\lambda}_f \left(X_f - d_f\right) = 0, \quad (11)$$

$$\xi_{f,p} x_f^p = 0, \quad (12)$$

$$\mu_e \left(X^e - B_e\right) = 0, \quad (13)$$

$$\lambda_f, \varphi_f, \mu_e, \xi_{f,p}, \theta_f, \beta_w \geq 0. \quad (14)$$

From equations (10) and (11), when the X_f is within the region $[d_f, D_f]$, both the lower bound and upper bound prices ($\overline{\lambda}$ and $\underline{\lambda}$) are converged to zero. From equations (9) and (12), it's easy to draw a conclusion that the path which can be used to route the flow f must be the one with the lowest price. The minimum price of the path which can be used by the flow f is defined as: $C^f = \min_{p\in P_f} C_p$. The dual decomposition results of each flow f are also the optimal bandwidth allocated to it when given a C^f.

$$X_f^* = \sum_{p\in P_f} x_f^*(p) = \left[U_f^{'-1}\left(\frac{C^f + \varphi_f}{a_f}\right)\right]_{d_f}^{D_f}, \forall f \in F. \quad (15)$$

which is unique if the utility is the strictly concave and X_f is viewed as the variable. $[z]_a^b = \max(a, min(z,b))$. $x_f^*(p)$ is the optimal bandwidth allocated to flow *f* in path *p*. The total amount of bandwidth allocated to flow *f* is between d_f and D_f.

The objective functions of $G_c(\boldsymbol{\varphi})$ is not strictly concave, because $\sum_{p\in P_f} x_f^p$ is linear if the flows have multiple alternative paths. In other words, once multiple paths are used by flows, the objective function is not strictly concave, even through the utility functions of all flows are strictly concave. So, the first-order Lagrange algorithm usually oscillates. In order to overcome it, an algorithm based on a sub-gradient approach is used. It decomposes the problem into a bandwidth allocation problem and a routing problem. The routing problem is to decide how to split the total data rate among a set of paths for a flow. The proposed algorithm shown in Algorithm 4 includes three main steps:

(1) For each flow, we use the following first-order Lagrange algorithm to update the bandwidth allocate to each flow:

$$X_f(t+1) = \left[U_{f'}^{-1}\left(\frac{C^f(t) + \varphi_f(t)}{a_f}\right)\right]_{d_f}^{D_f},$$

where t is the iteration index.

Algorithm 4 Iterative bandwidth allocation of the core network

1: SDN controller update allocated bandwidth:

2: for each flows *f* do

3: Compute the price of paths:

The price of paths C_p for all $p \in P_f$ is computed using μ.

Find the shortest path p_f^s and its price C^f.

4: Update the allocated bandwidth:

Update the allocated bandwidth of the flow *f*:

$$X_f(t+1) = \left[U_{f'}^{-1}\left(\frac{C^f(t) + \varphi_f(t)}{a_f} \right) \right]_{d_f}^{D_f}.$$

Update the allocated bandwidth of the flow *f* in path p:

$$x_f^p(t+1) = \left[x_f^p - r_2(C_{p_f}(t) - C^f(t)) \right]^+.$$

The allocated bandwidth of the path with the minimum price:

$$x_f^{p_f^s}(t+1) = \left[X_f(t+1) - \sum_{p \in P_f \setminus p_f^s} x_f^p(t+1) \right]^+.$$

5: end for

6: The price of links update:

Update the price of wired links in the core network according to the rate of flows (Algorithm 5).

(2) The way to split the bandwidth among multiple paths.

$$x_f^p(t+1) = \left[x_f^p - r_2(C_{p_f}(t) - C^f(t)) \right]^+.$$

where r_2 is a sufficiently small positive step-size for bandwidth. The allocated bandwidth of the path with the minimum price is updated by:

$$x_f^{p_f^s}(t+1) = \left[X_f(t+1) - \sum_{p \in P_f \setminus p_f^s} x_f^p(t+1) \right]^+,$$

where p_f^s is the shortest path with price μ. The paths with less price means there are more bandwidth can be allocated, while, paths with high price means some congestion may be existed. So, the allocated bandwidth of paths with excess price is decreased, while, the rate of the path with less price is increased.

(3) According to the allocated bandwidth X_e, links should update their price. Then, they will send the new price to SDN controller. The price of links should be updated in a gradient way according to the bandwidth utilization, which is given by:

$$\mu_e(t+1) = \left[\mu_e(t) + r_3(X^e(t) - B_e) \right]^+ \tag{16}$$

where t is the iteration index, r_3 is a sufficiently small positive step-size for link price. If the allocated bandwidth is beyond the link capacity, its price increases and vice versa. The allocated bandwidth of links with more price will be decreased in the subsequent iterations.

Algorithm 5 Price update for a wired link

1: Receives allocated bandwidth x_f^p for all paths that contain link e.

2: Compute the flow rate on link e:

$$X^e = \sum_f \sum_{p \in P_f} \pi_e^p x_f^p.$$

3: Compute a new price of this link:

$$\mu_e = \left[\mu_e + r_3(X^e - B_e) \right]^+.$$

4: Send new prices μ_e to the SDN controller.

This algorithm converges to a unique bandwidth allocation and an equilibrium price vector when the lower bound of the bandwidth of each flow can be offered by the network. The allocated bandwidth of a flow is determined by the minimum price of its allocated paths. So, if the price of links is converged, the bandwidth of flows is converged as well. From (16), the algorithm converges if $\mu_e(t+1) \to \mu_e(t)$ is true. From (13) and (16), the problem is converged if the condition $\mu_e(t) \to 0$ or $X^e(t) - B_e \to 0$ is satisfied. For the condition $\mu_e(t) \to 0$, the algorithm converges when $X^e(t)$ is less than B_e, which means the upper bound demands can be satisfied. Therefore, in this case, the price of the link will be decreased to zero. If $X^e(t)$ is bigger than B_e, the price of the link will grow until $X^e(t) = B_e$. Then, the price of the link is converged to a steady-state value which may be not zero. However, if $X^e(t)$ cannot be reduced and is bigger than B_e even if only lower bound of demands are offered to flows, the algorithm is not converged. Thus, the selection of the lower bound of the bandwidth demands should be within the capacity of the network; otherwise the algorithm cannot be converged.

6. Performance Evaluation

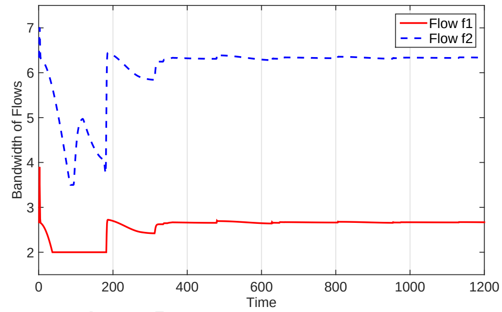
According to the marginal utility theory of microeconomics (Nicholson and Snyder, 2011), there is a diminishing marginal rate of performance enhancement as achievable bandwidth increases, so the utility function is defined as a logarithmic function. In the following experiments, the utility function is given by $U_f(x) = a_f \log(x+1)$. In this section, the topology in Figure 2 is used as a case study to test the performance of the proposed algorithms. Furthermore, topologies from the Survivable fixed telecommunication Network Design library (SNDLib) are used to show the performance with end to end multipath bandwidth allocation and the performance with SDN gradual deployment.

6.1. Simulation for Case Study

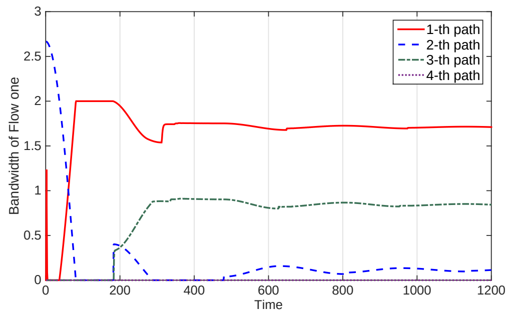
The topology in Figure 2 is studied with only two flows f_1 and f_2 in the network. The footstep r_1, r_2, r_3 are set to be 0.01. For f_1 and f_2, four available paths are taken into bandwidth allocation: S1 → A → C → G →K →D (1-th Path), S1 → A → C → F →K → D (2-th Path), A →C → G →K →H →D (3-th Path) and S1 → A → C → F →K → H → D (4-th Path). The information of two flows is shown in Table 2. As shown in the table, the demand of a flow is defined by an upper bound (D_f) and lower bound (d_f). The lower bound is the minimum demand that the flow requires. Besides, the upper bound is the maximum demand that the flow requires. Therefore, it's assumed that the power budget of flows should provide the lower bound (d_f) rate at the least. The simulation results of case studies are shown in following.

Table 2 Two flows information

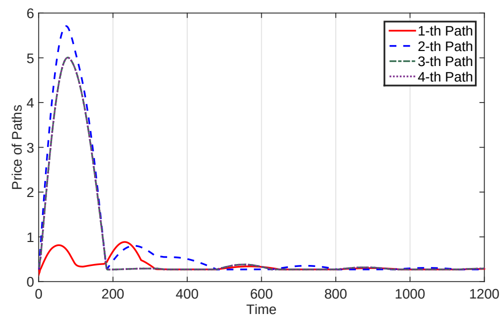
Flow	Source	Destination	D_f	d_f	a_f
f_1	S1	D	8	2	1
f_2	S1	D	7	3.5	2



(a) Allocated bandwidth of f1 and f2



(b) Bandwidth allocation of paths for f1



(c) The price of candidate paths for f1

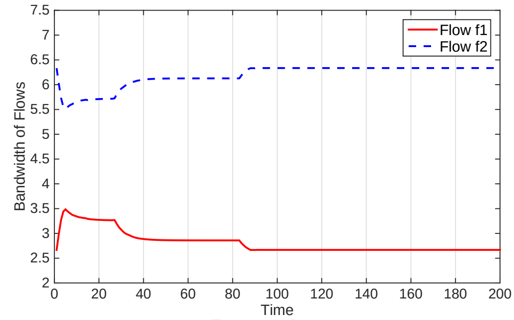
Figure 5. Simulation results of the core network

Firstly, the performance of the algorithm for the core network is shown in Figure 5. The x-axis is iteration times. By observation, we can see this algorithm has superior performance in convergence. Figure 5(a) is the allocated bandwidth of f_1 and f_2 .

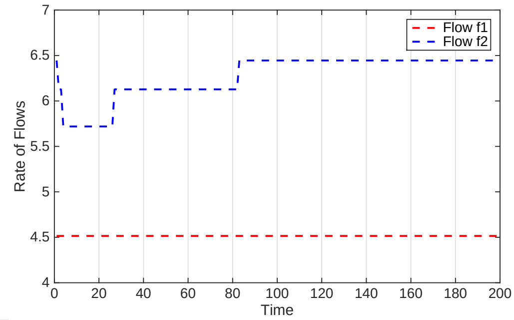
The allocated bandwidth reaches to about 2.66-unit for f_1 and about 6.33-unit for f_2 over four paths. From the topology, it can be easily known that the maximum bandwidth which can be allocated to flows from S1 to D is 9-unit which is almost allocated to f_1 and f2. Because the coefficient a_f of f_2 is more than that of f1, so the bandwidth allocated to f_2 is more than f_1. Figure 5(b) is the bandwidth allocated on every path of flow f_1. The bandwidth allocated to f1 on these paths is converged to 1.71, 0.11, 0.84 and 0, respectively. Figure 5(c) is the price of paths which is also converged.

Secondly, the performance of end to end bandwidth and power allocation is shown in Figure 6. The bandwidth allocation of the core network according to different penalty price (φ) is shown in Figure 6(a). Because f_2 has higher a_f, so the bandwidth allocated to it is more than that of f_1. The sending rate is shown in Figure 6(b) and penalty price is shown in Figure 6(c) with iteration times. The penalty price is sent to be 0.1 as the initial value. The penalty price of f_1 is in a decrease trend until zero because the bandwidth in the core network allocated to f_1 is not more than its sending rate of the access network. In other word, the bandwidth allocated to f_1 is fully used. The allocated bandwidth in core network for f_1 is also in increases trend at first 5 time in order to provide more bandwidth to f_1 to satisfy its sending rate. For flow f_2, the penalty price decreases, but it is slower than that of f_1. So, the bandwidth in the core network allocated to f_2 is in decrease trend at first time. But with increasing sending rate, the penalty price drops and the allocated bandwidth grows after time 5. At time 88, all penalty price decreases to be zero. At the same time, the allocated bandwidth and the sending rate is the optimum solution.

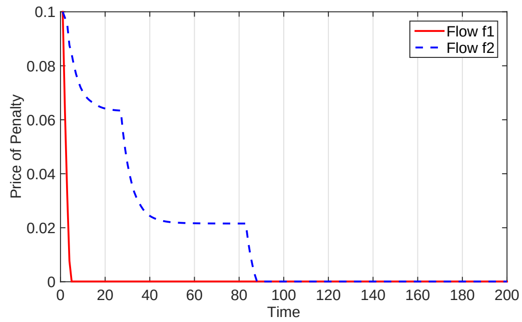
From the case study, it's proved that our algorithms have superior performance in convergence. The capacity of the core network can be fully used.



(a) Bandwidth allocation in the core network



(b) The sending rate of each flow

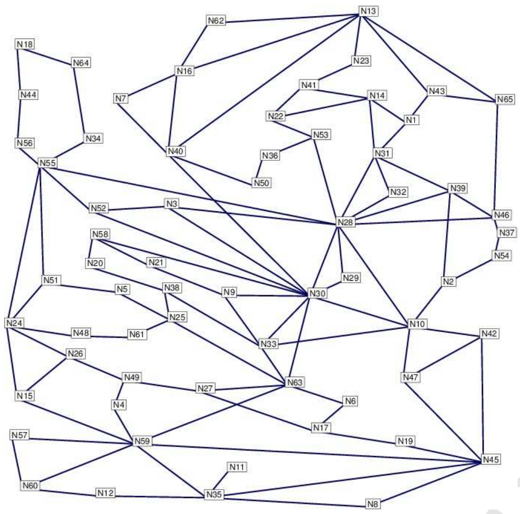


(c) The penalty price of each flow

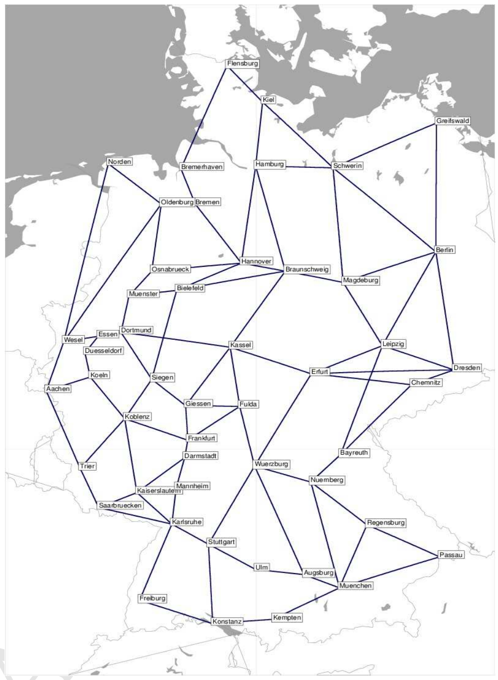
Figure 6. Simulation results of case studies with two flow f1 and f2

6.2. Performance of the End-to-end Multipath Resource Allocation Algorithm

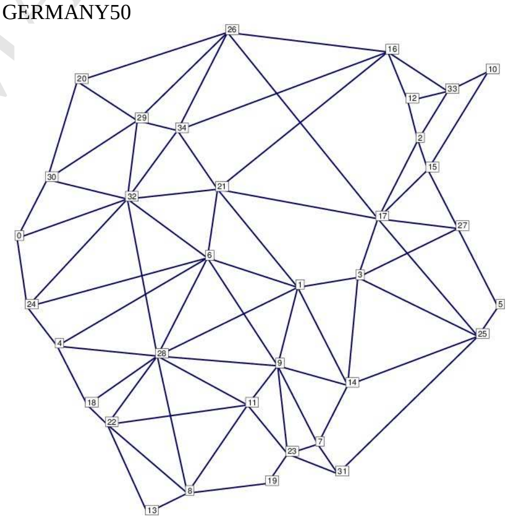
The networks INDIA35 as shown in Figure 7 are used to show the performance of our algorithms. The topology information is shown in TABLE 3. The capacities of links are randomly set ranging from 40 MB/s to 60MB/s. The number of base station is generated randomly from 5 to 10. Their location is also generated randomly which is direct connected with a switch in the core network. a_f is generated between 1 to 5 randomly. Flows select the access point from the set of base station randomly, and the destination of flows is randomly generated. The maximum bandwidth. The maximum bandwidth demands of flows D_f is between 8 MB/s and 30 MB/s and the d_f is between 2 MB/s and 5 MB/s. The footsteps of each iteration are set to be 0.01.



(a) TA2



(b) GERMANY50



(c) INDIA35

Figure 7: Topologies of simulation.

Table 3 Information of topologies

Topology	Nodes	Links	Flow number
TA2	65	108	100
GERMANY50	50	88	50
INDIA35	35	80	100

Figure 8 shows the network utility with different number of flows using the topology of INDIA35. Two bandwidth allocation schemes are used for a comparison. One is named not jointly multipaths for simplicity, in which path allocation optimization is only done in core network and multipath scenario is considered. Another is named end-to- end single path, in which end-to-end resource allocation is considered and single path for one source-destination pair is considered in the core network. From Figure 8, it’s obvious that with more flows the utility is in a growing trend, and our method which is end to end multipath resource allocation scheme has a better performance in utility relative to others. Besides, schemes with multipath have relatively better performance than the schemes with a single path for one source-destination pair.

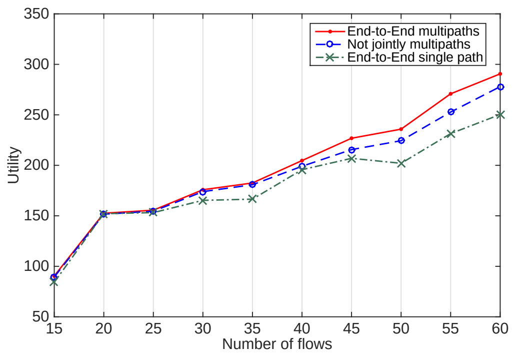


Figure 8. Utility with different schemes.

6.3. Performance with SDN Gradual Deployment

The networks TA2, GERMANY50 and INDIA35 as shown in Figure 7 are used to show the relationship between the SDN deployment and network utility. The SDN deployment sequence is generated based on the betweenness centrality. Betweenness centrality is an indicator of a node’s centrality in a network. It is equal to the number of candidate paths pass through that node. The number of flows is set to be 60.

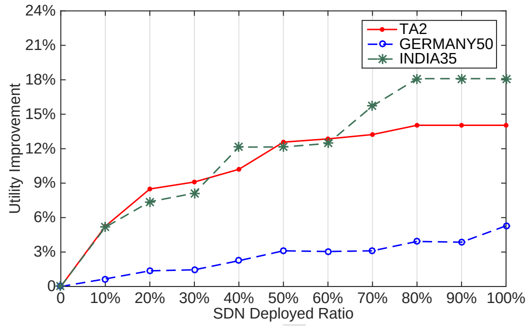


Figure 9. Total utility improvement with different SDN deployed ratio

Figure 9 shows the network utility improvement over the legacy networks with different SDN deployment ratio. The horizontal axis is the ratio of SDN deployment. The vertical axis is the network utility improvement compared with that of legacy networks. Three topologies (TA2, GERMANY50 and INDIA35) are used for illustration. As shown in the figure, the utility improvement is increasing with SDN deployment ratio. It is because more paths can be used for the bandwidth allocation if more devices are upgraded to support SDN. In this way, the utility can be improved. For example, when the 40% devices are deployed by SDN devices, the utility improvement of GER- MANY50, TA2 and INDIA35 are nearly 2.5%, 10% and 12%, respectively. From this experiment, it is easy to draw the conclusion that the utility can be improved with more SDN devices deployed in the network.

In order to test the performance of the utility improvement with different sets of flows, TA2 is used for illustration. The number of flows is generated from 50 to 100 randomly. Twenty sets of flows are generated randomly to test the performance of the hybrid bandwidth allocation in the hybrid SDNs as shown in Figure 10.

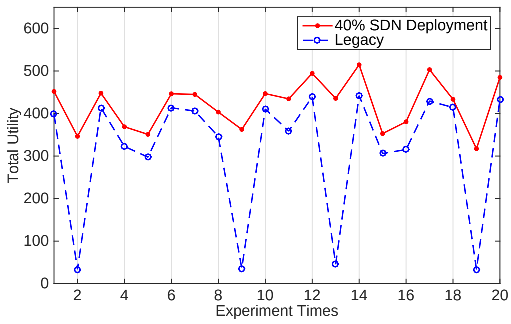


Figure 10: Bandwidth allocation performance with different flows

In Figure 10, the x-axis is the number of experiments. The y-axis is the network utility. There are two curves. One is the network utility of 40% devices deployed by the SDN. Another is the network utility of the legacy network. With different set of flows, the utility improvement is different. But it should be noted that the network utility of 40% devices deployed by the SDN is better than the legacy network in all the experiments.

7. Conclusion

To optimize the utility of software defined 5G network, a joint optimization algorithm is proposed to provide end to end resource optimization across access and core network. In this paper, the NOMA technique is considered in 5G network to ensure the carrier sharing by multiple users, and SDN technique is considered for resource allocation in core network to enable multiple paths routing. In order to solve the optimization model, we also proposed a method to decompose it into two subproblems with the penalty price. The results show that the proposed joint optimization algorithm is able to achieve significant improvement in terms of network utility no matter in pure SDN or hybrid SDN network.

Acknowledgments

This work has been supported by Project U1603261 supported by Joint Funds of National Natural Science Foundation of China and Xinjiang, the Research Fund of Ministry of Education-China Mobile under Grant No. MCM20160304 and the doctoral program for short term study abroad in Beijing University of Posts and Telecommunications.

References

Agarwal S, Kodialam M, Lakshman T. Traffic engineering in software defined networks. In: INFOCOM, 2013 Proceedings IEEE. IEEE; 2013. p. 2211–9.

Agiwal M, Roy A, Saxena N. Next generation 5G wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials 2016;18(3):1617–55.

Al-Imari M, Xiao P, Imran MA, Tafazolli R. Uplink non-orthogonal multiple access for 5G wireless networks. In: Wireless Communications Systems (ISWCS), 2014 11th International Symposium on. IEEE; 2014. p. 781–5.

Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong AC, Zhang JC. What will 5G be? IEEE Journal on selected areas in communications 2014;32(6):1065–82.

Caria M, Das T, Jukan A, Hoffmann M. Divide and conquer: Partitioning ospf networks with sdn. In: Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on. IEEE; 2015. p. 467–74.

Chen S, Zhao J. The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication. IEEE Communications Magazine 2014;52(5):36–43.

Chiang M, Bell J. Balancing supply and demand of bandwidth in wireless cellular networks: utility maximization over powers and rates. In: INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies. IEEE; volume 4; 2004. p. 2800–11.

Chiang M, Low SH, Calderbank AR, Doyle JC. Layering as optimization decomposition: A mathematical theory of network architectures. Proceedings of the IEEE 2007;95(1):255–312.

Demestichas P, Georgakopoulos A, Karvounas D, Tsagkaris K, Stavroulaki V, Lu J, Xiong C, Yao J. 5G on the horizon: key challenges for the radio-access network. IEEE Vehicular Technology Magazine 2013;8(3):47–53.

Eghbali H, Wong VW. Bandwidth allocation and pricing for sdn-enabled home networks. In: Communications (ICC), 2015 IEEE International Conference on. IEEE; 2015. p. 5342–7.

Feng T, Bi J, Wang K. Allocation and scheduling of network resource for multiple control applications in sdn. China Communications 2015;12(6):85–95.

Guo Y, Wang Z, Yin X, Shi X, Wu J, Zhang H. Incremental deployment for traffic engineering in hybrid SDN network. In:

2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC). IEEE; 2015. p. 1–8.

Habib S, Qadir J, Ali A, Habib D, Li M, Sathiaseelan A. The past, present, and future of transport-layer multipath. Journal of Network and Computer Applications 2016; 75:236–58.

Hong DK, Ma Y, Banerjee S, Mao ZM. Incremental deployment of SDN in hybrid enterprise and ISP networks. In: Symposium on Sdn Research. 2016. p. 1–7.

Islam SR, Avazov N, Dobre OA, Kwak KS. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges. IEEE Communications Surveys & Tutorials 2017;

Jain S, Kumar A, Mandal S, Ong J, Poutievski L, Singh A, Venkata S, Wanderer J, Zhou J, Zhu M, et al. B4: Experience with a globally-deployed software defined WAN. ACM SIGCOMM Computer Communication Review 2013;43(4):3–14.

Jin J, Wang WH, Palaniswami M. Utility max-min fair resource allocation for communication networks with multipath routing. Computer Communications 2009;32(17):1802–9.

Kelly FP, Maulloo AK, Tan DK. Rate control for communication networks: shadow prices, proportional fairness and stability. Journal of the Operational Research society 1998;49(\3):237–52.

Kuo WH, Liao W. Utility-based radio resource allocation for QoS traffic in wireless networks. IEEE Transactions on Wireless Communications 2008;7(7):2714–22.

La RJ, Anantharam V. Utility-based rate control in the internet for elastic traffic. IEEE/ACM Transactions on Networking 2002;10(2):272–86. Lee JW, Mazumdar RR, Shroff NB. Joint opportunistic power scheduling and end-to-end rate control for wireless ad hoc networks. IEEE Transactions on Vehicular Technology 2007;56(2):801–9.

Lei L, Yuan D, Ho CK, Sun S. Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems. In: IEEE Global Communications Conference. 2015. p. 1–6.

Levin D, Canini M, Schmid S, Feldmann A. Incremental SDN deployment in enterprise networks. acm special interest group on data communication 2013;43(4):473–4.

Liao WC, Hong M, Farmanbar H, Li X, Luo ZQ, Zhang H. Min flow rate maximization for software defined radio access networks. IEEE Journal on Selected Areas in Communications 2014;32(6):1282–94.

Lin X, Shroff NB. Utility maximization for communication networks with multipath routing. IEEE Transactions on Automatic Control 2006;51(5):766–81.

Masoudi R, Ghaffari A. Software defined networks: A survey. Journal of Network and Computer Applications 2016; 67:1–25.

Mo J, Walrand J. Fair end-to-end window-based congestion control. IEEE/ACM Transactions on Networking (ToN) 2000;8(5):556–67. Nicholson W, Snyder C. Microeconomic theory: Basic principles and extensions. Nelson Education, 2011.

Nunes BAA, Mendonca M, Nguyen XN, Obraczka K, Turletti T. A survey of software-defined networking: Past, present, and future of programmable networks. IEEE Communications Surveys & Tutorials 2014;16(3):1617–34.

Pakzad F, Portmann M, Tan WL, Indulska J. Efficient topology discovery in openflow-based software defined networks. Computer Communications2016; 77:52–61.

Sun S, Kadoch M, Gong L, Rong B. Integrating network function virtualization with SDR and SDN for 4G/5G networks. IEEE Network 2015;29(3):54–9.

Tan L, Zhang X, Andrew LL, Zukerman M. Price-based max-min fair rate allocation in wireless multi-hop networks. In: TENCON 2005-2005 IEEE Region 10 Conference. IEEE; 2005. p. 1–6.

Tan L, Zhu Z, Ge F, Xiong N. Utility maximization resource allocation in wireless networks: Methods and algorithms. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2015;45(7):1018–34.

Trivisonno R, Guerzoni R, Vaishnavi I, Soldani D. SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility. Transactions on Emerging Telecommunications Technologies 2015;26(1):82–92.

Vissicchio S, Vanbever L, Bonaventure O. Opportunities and research challenges of hybrid software defined networks. ACM SIGCOMM Computer Communication Review 2014;44(2):70–5.

Wang CX, Haider F, Gao X, You XH, Yang Y, Yuan D, Aggoune H, Haas H, Fletcher S, Hepsaydir E. Cellular architecture and key technologies for 5G wireless communication networks. IEEE Communications Magazine 2014;52(2):122–30.

Wang WH, Palaniswami M, Low SH. Optimal flow control and routing in multi-path networks. Performance Evaluation 2003;52(2):119–32.