Abstract—Owing to severe path loss and unreliable transmission over a long distance at higher frequency bands, this paper investigates the problem of path selection and rate allocation for multi-hop self-backhaul millimeter wave (mmWave) networks. Enabling multi-hop mmWave transmissions raises a potential issue of increased latency, and thus, this work aims at addressing the fundamental questions: “how to select the best multi-hop paths and how to allocate rates over these paths subject to latency constraints?”. In this regard, a new system design, which exploits multiple antenna diversity, mmWave bandwidth, and traffic splitting techniques, is proposed to improve the downlink transmission. The studied problem is cast as a network utility maximization, subject to an upper delay bound constraint, network stability, and network dynamics. By leveraging stochastic optimization, the problem is decoupled into: (i) path selection and (ii) rate allocation sub-problems, whereby a framework which selects the best paths is proposed using reinforcement learning techniques. Moreover, the rate allocation is a non-convex program, which is converted into a convex one by using the successive convex approximation method. Via mathematical analysis, a comprehensive performance analysis and convergence proof are provided for the proposed solution. Numerical results show that the proposed approach ensures reliable communication with a guaranteed probability of up to 99.9999%, and reduces latency by 50.64% and 92.9% as compared to baseline models. Furthermore, the results showcase the trade-off between latency and network arrival rate.

Index Terms—Ultra-low latency and reliable communication (URLLC), self-backhaul, mmWave communications, multi-hop scheduling, ultra-dense small cells, stochastic optimization, reinforcement learning.

I. INTRODUCTION

The fifth generation (5G) wireless systems are expected to reach multiple gigabits per second (Gbps) and to serve a massive number of wireless-connected devices [2], [3]. In this regard, both academia and industry have paid tremendous attention to the underutilized mmWave frequency bands (30–300 GHz) due to the current scarcity of wireless spectrum [2], [4], [5]. Moreover, the above challenges can be achieved by: (i) advanced spectral-efficient techniques, e.g., massive multiple-input multiple-output (MIMO) [6]; and (ii) ultra-dense self-backhauled small cell (SC) deployments [5], [7], [8]. Indeed, massive MIMO has been recognized as one of the promising 5G techniques, which allows to form highly directional beamforming to utilize the mmWave frequency bands and to provide wireless backhaul for SC deployment [7], [8]. Ultra dense SC effectively increases network capacity and coverage in which advanced full-duplex (FD) potentially doubles spectral efficiency and reduces latency [5], [9], [10].

In addition to the unprecedented growth of data traffic and devices, the issues of low-latency and high-reliability represent other important concerns in 5G networks and beyond [2], [11], [12], [13], [14], [15], [16]. This paper investigates the above 5G enablers, namely mmWave communication, massive MIMO, and ultra-dense SC deployment, envisaged as the key promoters for providing Gbps data rate, low latency, and highly reliable communication [2], [5]. In particular, an in-band access and wireless backhauling are considered to enable the ultra-dense SC deployment [7], [8], [17], [18] by combining massive MIMO and mmWave to provide Gigabits capacity for both access and wireless backhaul [6], [19]. Owing to the short wavelength, mmWave frequency bands allow for packing a massive number of antennas into highly directional beamforming over a short distance [19], [20], [21]. Besides that, transmitting over a long distance, mmWave communication requires higher transmit power and is very sensitive to blockage [4], [8], [5]. Hence, instead of using a single hop [8], [11], a multi-hop self-backhauling architecture is a promising solution to enable transmissions over long distances in 5G mmWave networks [3], [22]. However, using multi-hop transmissions raises the critical issue of increased delay, which has been generally ignored [22], [23], [24], [25], [26], [27], [28]. Unavoidably, ultra-dense SC network is mainly operated based on the multi-hop multi-path transmission fashion [5], [9], [3]. Hence, there is a need for fast and efficient multi-hop scheduling with respect to traffic dynamics and channel variances in 5G self-backhauled mmWave networks [3], [29]. These previous works focused on addressing one or few issues, have not studied the problem a joint path selection (PS) and rate allocation (RA) in mmWave networks to ensure Gbps data rate and low latency with reliable communications. Thus far, to the best of our knowledge, we are perhaps the first to provide a theoretical and practical framework for addressing all these above concerns.

A. Main contributions

In this work, a new system design, which exploits multi-hop transmission, multiple antenna diversity, mmWave bandwidth,
and dynamic PS with traffic splitting techniques, is proposed to overcome the severe path loss and mitigate the impact of blockage. The main contributions of the work are listed as follows:

- A joint PS and RA optimization for multi-hop multi-path scheduling is formulated, whereby self-backhauled FD SCs act as relay nodes to forward data from the macro BS to the intended UEs. Multi-hop transmission technique enables reliable mmWave communications over a long distance. However, there is a probability that the mmWave signal can be blocked by the human body. Hence, we also introduce the multi-path selection scheme in which the transmitter smartly selects a subset of the best paths among the possible paths.

- In the proposed system design, leveraging massive array antenna, hybrid beamforming is adopted to provide Gbps data rate at mmWave bands. In addition, we impose a probabilistic latency bound to ensure URLLC with high data rate. For this purpose, the studied problem is cast as a network utility maximization (NUM), subject to a bounded latency constraint and network stability.

- Leveraging stochastic optimization framework [30], the studied problem is decoupled into two sub-problems, namely PS and RA. By utilizing the benefits of historical information, a reinforcement learning (RL) is used to build an empirical distribution of the system dynamics to aid in learning the best paths to solve PS [31], [32]. Therein, the concept of regret strategy is employed, defined as the difference between the average utility when choosing the same paths in previous times, and its average utility obtained by constantly selecting different paths [31], [32]. The premise is that regret is minimized over time so as to choose the best paths. Second, to solve a non-convex RA sub-problem, the concept of successive convex approximation (SCA) method is applied due to its low complexity and fast convergence [33], [34].

- The proposed approach answers the following fundamental questions: (i) over which paths should the traffic flow be forwarded? and (ii) what is the data rate per flow/subflow?, while ensuring a probabilistic delay constraint, and network stability. By using a mathematical analysis, a comprehensive performance of our proposed stochastic optimization framework is scrutinized. It is shown that there exists an $O(1/\nu), O(\nu)$ utility-queue backlog trade-off, which leads to an utility-delay balancing [30], where $\nu$ is a control parameter. In addition, a convergence analysis of both two sub-problems is studied. Finally, the performance of the proposed solution is validated by extensive set of simulations.

B. Related work

A tractable rate model was proposed to characterize the rate distribution in self-backhauled mmWave networks [35]. Few efforts have been made to study the mmWave network operation regime, noise-limited or interference limited, depending on the density of interferers, transmission strategies, or channel propagation models [36], [37], [38]. A large body of research work has attempted to study the joint RA, congestion control, routing, and scheduling for multi-hop wireless networks, incorporating the proportional delay based on the sum of queue backlogs [23], applying the concept of back-pressure algorithm [39], [40], exploiting the potential of multiple gateways [25].

The authors in [41] considered a problem of joint scheduling and congestion control in a multi-hop mmWave network using a NUM framework in which the proposed solution is verified under three interference models, namely graph-based actual interference, free-interference (IF), and the worse-case interference. [41] also showed that the IF model provides a tight upper bound for a realistic system evaluation in mmWave cellular networks as long as the optimal throughput can be guaranteed. However, [41], [42] was concerned only with the network capacity maximization and single path streaming, a tight latency and reliable constraint should be investigated together with dynamic path diversity. Moreover, the authors in [43] designed a multi-hop wireless backhaul scheme with delay guarantee in which a link activation scheme was proposed to avoid interference and minimize the latency. A rate allocation problem to minimize the application layer video/end-to-end distortion subject to quality of service constraints (delay, backhaul) was considered in [44], [45] for multi-path networks. However, other important aspects in 5G networks such as low-latency and high-reliability are generally ignored when maximizing the network performance (capacity, energy efficiency and spectral efficiency) [28], [35], [46], [47].

A recent work in [26] has studied the multi-hop relaying transmission challenges for mmWave systems, aiming at maximizing overall network throughput, and taking account of traffic dynamics and link qualities. In our work, we also study the NUM optimization problem, while considering channel variations and network dynamics. Another recent work in [48] has addressed the problem of traffic allocation for multi-hop scheduling in mmWave networks to minimize the end-to-end latency, in which the minimum latency is derived based on the channel capacity to determine the portions of traffic over channels such that all traffic fractions arrive simultaneously at the destination. In addition, the problem of PS and multi-path congestion control for data transfers was studied in [27] in which the aggregate utility is increased as more paths are provided. One important suggestion is to re-select randomly from the set of paths and shift between paths with higher payoff. However, splitting data into too many paths leads to increased signaling overhead and causes traffic congestion. While interesting, the preceding works do not address the problem of high-data rate, low-latency and reliability communication in multi-path mmWave networks. In this respect, our proposed solution is to select the best paths to maximize the network throughput, subject to a delay bound violation constraint with a tolerable probability (reliability). Our previous work [11] studied URLLC-centric mmWave networks for single hop transmission, and [8] proposed an integrated access and backhaul architecture for two-hop relay without considering the delay-sensitive constraint. Hence, in this work the authors extend to the multi-hop wireless backhaul scenario, and study a joint PS and RA problem focusing on URLLC. Via mathematical analyses and extensive simulations, the authors provide insights into the performance analysis of our proposed...
algorithm and the convergence characteristics of the learning algorithm and the SOCP based iterative method.

The rest of the paper is organized as follows\(^1\). Section II describes the system model and Section III provides the problem formulation for a joint PS and RA optimization. Section IV introduces a stochastic optimization framework to decouple our studied problem, whereby two practical solutions are proposed. A mathematical analysis of the proposed framework is discussed in Section V. Section VI provides extensive numerical results to compare again other baselines. Conclusions are drawn in Section VII.

II. SYSTEM MODEL

A. Network Model

Let us consider a downlink (DL) transmission of a multi-hop heterogeneous cellular network (HCN) which consists of a macro base station (MBS), a set of BS self-backhauled small cell base stations (SCBSs), and a set of K of K user equipments (UEs) as shown in Fig 1. Let \( \mathcal{B} = \{0, 1, \ldots, B\} \) denote the set of all BSs in which index 0 refers to the MBS. The in-band wireless backhaul is used to provide backhaul among BSs [17], [49]. A full-duplex (FD) transmission protocol is used to deliver via disjoint paths and aggregated at UEs [52], [53].

We consider a queuing network operating in discrete time \( t \in \mathbb{Z}^+ \). There are \( F \) independent data flows at the MBS. Each data traffic is destined for only one UE, whereas one UE can receive up to \( R_k \) multiple data streams, i.e., \( F \geq K \). The number of total data streams at the MBS is no greater than the number of RF chains, such that \( F \times R_k \leq N_k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B} \) [20], [21], [51]. Hereafter, we refer to data traffic as data flow. We denote the set of all directional edges \((i, j)\), in which nodes \( i \) and \( j \) are the transmitter and the receiver, respectively.

We consider a queuing network operating in discrete time \( t \in \mathbb{Z}^+ \). There are \( F \) independent data flows at the MBS. Each data traffic is destined for only one UE, whereas one UE can receive up to \( R_k \) multiple data streams, i.e., \( F \geq K \). The number of total data streams at the MBS is no greater than the number of RF chains, such that \( F \times R_k \leq N_k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B} \) [20], [21], [51]. Hereafter, we refer to data traffic as data flow. We use \( \mathcal{F} \) to represent the set of \( F \) data flows/sub-flows. The MBS can split each flow \( f \in \mathcal{F} \) into multiple sub-flows which are delivered via disjoint paths and aggregated at UEs [52], [53].

We assume that there exists \( Z_f \) number of disjoint paths from the MBS to the UE for flow \( f \). For any disjoint path \( m \in \{1, \ldots, Z_f\} \), we denote \( Z_f^{m} \) as the path state, which contains all path information such as topology and queue states for every hop. Let \( Z_f = \{Z_f^{1}, \ldots, Z_f^{m}, \ldots, Z_f^{Z_f}\} \) denote the path states/tables observed by flow \( f \). We use the flow-split indicator vector \( z_f = (z_f^{1}, \ldots, z_f^{Z_f}) \) to denote how the MBS splits flow \( f \), where \( z_f^{m} = 1 \) means path \( m \) is used to send data for flow \( f \); otherwise, \( z_f^{m} = 0 \). Let \( N_f^{od} \) denote the set of next hops from node \( i \) via a directional edge. We denote the next hop and the previous hop of flow \( f \) from and to BSs \( i \) as \( i_f^{n} \) and \( i_f^{b} \), respectively. Table I shows the notations, used throughout this paper.

B. mmWave MIMO Channel Model

Due to limited spatial scattering in mmWave MIMO propagation [4], [51], we assume that there are \( L_{(i,j)} \) clusters between transmitter \( i \) and receiver \( j \), such that \( L_{(i,j)} < \min(N_i, N_j) \). The channel matrix \( \mathbf{H}_{i,j} \) of link \((i, j)\) can be modelled as [51], [54], [55]
\[
\mathbf{H}_{i,j} = \frac{N_i \times N_j}{L_{(i,j)}} \sum_{l=1}^{L_{(i,j)}} h_{i,j}(l) \mathbf{A}_j(\alpha_{j,l}) \mathbf{A}_i^\dagger(\alpha_{i,l}),
\]
where \( h_{i,j}(l) \) denotes the small-scale fading coefficient of the cluster \( l^{th} \), \( \alpha_{j,l} \) and \( \alpha_{i,l} \) denote the azimuths of arrival and departure, respectively. Here, \( \mathbf{A}_j(\alpha_{j,l}) \) and \( \mathbf{A}_i(\alpha_{i,l}) \) represent the transmitter and receiver response vectors, respectively (Please refer [54], [55] for more details). We denote \( \mathbf{H} = \{\mathbf{H}_{i,j}(i, j) \in \mathcal{L}\} \) as the network channel matrix.

C. Rate Formulation

We denote \( p_{i,j}^f \) as the transmit power of node \( i \) assigned to node \( j \) for flow \( f \), such that \( \sum_{f \in \mathcal{F}} \sum_{j \in N_i^{od}} p_{i,j}^f \leq p_{i}^{\text{max}} \), where \( p_{i}^{\text{max}} \) is the maximum transmit power of node \( i \). We have the following power constraint
\[
\mathbf{P} = \left\{ p_{i,j}^f : 0, i, j \in N_i, \sum_{f \in \mathcal{F}} \sum_{j \in N_i^{od}} p_{i,j}^f \leq p_{i}^{\text{max}} \right\}.
\]

Vector \( \mathbf{p} = (p_{i,j}^f)_{i, j \in N_i, \forall f \in \mathcal{F}} \) denotes the transmit power over all flows.

Based on the hybrid beamforming and combining model [21], [51], with \( c_{i,j} \in \mathbb{C}^{N_i \times 1} \) as the RF combining and baseband equalizer and \( v_{i,j} \in \mathbb{C}^{N_i \times 1} \) as hybrid analog/digital

\(^{1}\text{Notations: Throughout the paper, the lowercase letters, boldface lowercase letters, (boldface) uppercase letters and italic boldface uppercase letters are used to represent scalars, vectors, matrices, and sets, respectively. For a matrix \( \mathbf{X} \), we use \( \mathbf{X}^T, \mathbf{X} \) and \( \text{Rank}(\mathbf{X}) \) to denote its transpose, Hermitian and rank, respectively. \( \mathbb{E} \) denotes the expectation operator.} \)
precoding, the Ergodic achievable rate is calculated as (4) in which $g_r^{(k)}$ denotes the spatial channel gain of link $(i, j)$ [54], [55], [60]. Note that after the beam-searching and alignment are done [54], [60], [62], [63] the receiver broadcasts pilot sequences to the transmitters, each transmitter estimates the channel to the corresponding receiver and precodes transmit signal in the DL. With multiple $N_j$ antennas and $R_j$ RF chains, each receiver is capable of receiving multiple data streams from different transmitters using either the main beam or the side lobe beam. We assume that the traffic split and aggregation are done ideally, the multiple data streams can be transmitted via different paths.

D. Network Queues

Let $Q_f^i(t)$ denote the queue length at a BS $i$ at time slot $t$ for flow $f$. The queue length evolution at the MBS $i = 0$ is

$$Q_f^i(t + 1) = Q_f^i(t) - \sum_{m \in Z_f} r_f^{(i,m)}(t) + a_f(t),$$

where $a_f(t)$ is the data arrival at the MBS during slot $t$, which is i.i.d over time with a mean rate $\bar{a}_f$ and is bounded by $a_f(t) \leq a_{\max} f < \infty$. Due to the disjoint paths, for each flow $f$ the incoming rate from the previous hop $I^i_f$ at the SCBS $i$ is either from another SCBS or the MBS, and thus, the queue evolution at the SCBS $i = \{1, \cdots, B\}$ is given by

$$Q_f^i(t + 1) = Q_f^i(t) - r_f^{(i,m)}(t) + a_f(t).$$

Definition 1. For any vector $\mathbf{x}(t) = (x_1(t), \ldots, x_K(t))$, let $\mathbf{\bar{x}} = (\bar{x}_1, \cdots, \bar{x}_K)$ denote the time average expectation of $\mathbf{x}(t)$, where

$$\bar{x} = \lim_{t \to \infty} \frac{1}{t} \sum_{t=0}^{t-1} \mathbb{E}[\mathbf{x}(t)].$$

Definition 2. For any discrete queue $Q(t)$ over time slots $t \in \{0, 1, \ldots\}$ and $Q(t) \in \mathbb{R}_+$,

- $Q(t)$ is strongly stable if $\lim_{t \to \infty} Q(t) = 0$. A queueing network is stable if each queue is stable.

III. Problem Formulation

Assume that the MBS determines which paths to split data flow $f$ with a given probability distribution, i.e., $\pi_f$ =
\[ p(i,j) = \mathbb{E}[\mathbf{H}_p] \left[ \log \left( 1 + \frac{\mathbf{p}(i,j)\mathbf{c}(i,j)\mathbf{H}_{i,j}^T\mathbf{v}(i,j)}{\sum_{l \neq i} \sum_{j' \in} \mathbf{p}(i,j')\mathbf{c}(i,j')\mathbf{H}_{i,j'}^T\mathbf{v}(i,j') + \sigma_j^2}\right) \right] . \]  

\[ p(i,j) = \mathbb{E}[\mathbf{H}_p] \left[ \log \left( 1 + \frac{\mathbf{p}(i,j)\mathbf{g}(i,j)\mathbf{H}_{i,j}^T\mathbf{g}(i,j)}{\sum_{l \neq i} \sum_{j' \in} \mathbf{p}(i,j')\mathbf{g}(i,j')\mathbf{H}_{i,j'}^T\mathbf{g}(i,j') + \sigma_j^2}\right) \right] . \]

\((\pi_f^1, \cdots, \pi_f^Z_f)\), where for each \(m \in Z_f\) we have \(\pi_f^m = \text{Pr}\left( z_f = z_f^m \right)\). Here, \(\pi_f\) is the probability mass function (PMF) of the flow-split vector, i.e., \(\sum_{m=1}^{Z_f} \text{Pr}\left( z_f^m \right) = 1\). We denote \(\pi = \{\pi_1, \cdots, \pi_f, \cdots, \pi_F\} \in \Pi\) as the global probability distribution of all flow-split vectors in which \(\Pi\) is the set of all possible global PMFs. Let \(\bar{x}_f\) denote the achievable average rate of flow \(f\) such that

\[ \bar{x}_f = \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} x_f(\tau), \]

and

\[ x_f(\tau) = \sum_{m=1}^{Z_f} \mathbb{E}[\mathbf{H}_p] |\pi_f^m(t)| I(\pi_r^m(t)) \bigg|_{t=0}, \]

We assume that the achievable rate is bounded, i.e.,

\[ 0 \leq x_f(\tau) \leq d_{\text{max}}, \]

where \(d_{\text{max}}\) is the maximum achievable rate of flow \(f\) at every time \(t\). Vector \(\bar{x} = (\bar{x}_1, \cdots, \bar{x}_F)\) denotes the time average of rates over all flows. Let \(\mathcal{R}\) denote the rate region, which is defined as the convex hull of the average rates, i.e., \(|\bar{x}| \in \mathcal{R}\). We define \(U_0\) as the utility network function, i.e., \(U_0(\bar{x}) = \sum_{f \in F} U(\bar{x}_f)\). Here, \(U(\bar{x})\) is assumed to be twice differentiable, concave, and increasing \(-\)Lipschitz function for all \(|\bar{x}| \geq 0\). According to Little’s law \([64]\), the average queuing delay is defined as the ratio of the queue length to the average arrival rate. By taking account of the probabilistic delay constraints for each flow/subflow, the network utility maximization (NUM) is formulated as follows:

\[ \text{OP:} \max_{\pi, \bar{x}} U_0(\bar{x}) \]  

subject to

\[ \text{Pr}\left( \frac{\mathbb{E}[\mathcal{Q}_j(i)]}{\bar{a}_f} \geq d_{\text{max}} \right) \leq \epsilon, \forall i, f \in \mathcal{F}, i \in \mathcal{B}, \quad (8b) \]

\[ \lim_{t \to \infty} \mathbb{E}[\mathcal{Q}_j(i)] = 0, \forall f \in \mathcal{F}, \forall i \in \mathcal{B}, \quad (8c) \]

\[ \bar{x}(t) \in \mathcal{R}, \quad (8d) \]

\[ \pi \in \Pi, \quad (8e) \]

and

\[ (2), (7). \]

where \(d_{\text{max}}\) reflects the delay threshold required for UEs, and \(\epsilon \ll 1\) is the target probability for reliable communication\(^3\).

\(\)The UEs can have different delay and reliability requirements.

The probabilistic delay constraint (8b) implies that the probability that the delay for each flow at node \(i\) is greater than \(d_{\text{max}}\) is very small, which captures the constraints of ultra-low latency and reliable communication \([11], [65]\). It is also used to avoid congestion for each flow \(f\) at any point (BS) in the network, since the queue length is ensured less than \(d_{\text{max}}\bar{a}_f\) with probability \(1 - \epsilon\). Hence, (8b) forces the transmission of all BSs without building large queues, and (8c) maintains network stability.

The above problem has a non-linear probabilistic constraint (8b), which cannot be solved directly. Hence, we replace the non-linear constraint (8b) with a linear deterministic equivalent by applying Markov’s inequality \([66]\), \([11]\) such that

\[ \text{Pr}(X \geq x) \leq \mathbb{E}[X]/x \text{ for a non-negative random variable } X \text{ and } x > 0. \]

Thus, we relax (8b) as

\[ \mathbb{E}[\mathcal{Q}_j(i)] \leq \bar{a}_f d_{\text{max}}, \quad (9) \]

Assuming that \(a_f(t)\) follows a Poisson arrival process \([66]\), we derive the expected queue length in (5) for \(i = 0\) as

\[ \mathbb{E}[\mathcal{Q}_j(i)] = t \bar{a}_f - \sum_{m=1}^{\bar{x}_f} \sum_{i \in \mathcal{B}} \pi_f^m f_j(i,\bar{x}_f)(\tau), \]

and the expected queue length in (6), for each SCBS, i.e.,

\[ \mathbb{E}[\mathcal{Q}_j(i)] = \sum_{m=1}^{\bar{x}_f} \sum_{i \in \mathcal{B}} \pi_f^m f_j(i,\bar{x}_f)(\tau) - f_j(i,\bar{x}_f)(\tau), \]

Subsequently, combining the constraints (9) and (10), we obtain the following linear constraint (12) of instantaneous rate requirements, which helps to analyse and optimize the URLLC problem \([11], [65]\), for MBS \(i = 0\),

\[ \bar{a}_f(t) - \epsilon d_{\text{max}} \leq -\sum_{m=1}^{\bar{x}_f} \sum_{i \in \mathcal{B}} \pi_f^m f_j(i,\bar{x}_f)(\tau) \leq \sum_{m=1}^{\bar{x}_f} \sum_{i \in \mathcal{B}} \pi_f^m f_j(i,\bar{x}_f)(\tau), \]

Similarly, for each SCBS \(i = 1, \cdots, B\), we have

\[ -\bar{a}_f d_{\text{max}} + \sum_{m=1}^{\bar{x}_f} \pi_f^m f_j(i,\bar{x}_f)(\tau) - f_j(i,\bar{x}_f)(\tau) \leq \sum_{m=1}^{\bar{x}_f} \pi_f^m f_j(i,\bar{x}_f)(\tau) - f_j(i,\bar{x}_f)(\tau), \]

by combining (9) and (11). With the aid of the above derivations, we consider (12) and (13) instead of (8b) in the
original problem (8). In practice, the statistical information of all candidate paths to decide \( \pi_f, \forall f \in \mathcal{F} \), is not available beforehand, and thus solving (8) is challenging. One solution is that paths are randomly assigned to each flow which does not guarantee optimality, whereas applying an exhaustive search is not practical. Therefore, in this work, the stochastic optimization is pertained to characterize the queuing latency in the presence of randomness (mmWave wireless channels and arbitrary arrivals). As a result, (8) is decoupled into subproblems, which can be solved by low-complexity and efficient methods. In particular, RL is leveraged to find the best paths without requiring the statistic information, and SCA method obtains a locally efficient solution for assigning rate over the flows.

IV. PROPOSED PATH SELECTION AND RATE ALLOCATION ALGORITHM

In this section, we propose a Lyapunov stochastic optimization based framework to solve our predefined problem (8) with relaxed latency constraints. To do that, we first introduce a set of auxiliary variables to refine the original problem (8). Next, we convert the constraints into virtual queues and write the conditional Lyapunov drift function. Finally, the solution of the equivalent problem is obtained by minimizing the Lyapunov drift and a penalty from the objective function. Let us start by rewriting (8) equivalently as follows:

\[
\begin{align*}
\text{RP}: \quad & \max_{\bar{\varphi}, \pi, \nu} U_0(\bar{\varphi}) \quad \text{subject to} \\
& \bar{\varphi}_f - \bar{\pi}_f \leq 0, \; \forall f \in \mathcal{F}, \tag{14a} \\
& (2), \; (7), \; (8c), \; (8e), \; (12), \; (13), \tag{14b}
\end{align*}
\]

where the new constraint (14b) is introduced to replace the rate constraint (8d) with new auxiliary variables \( \varphi = (\varphi_1, \ldots, \varphi_F) \). In (14b), \( \bar{\varphi} = \lim_{t \to \infty} \frac{1}{t} \sum_{t=1}^{t-1} E[|\varphi(t)|] \). In order to ensure the inequality constraint (14b), we introduce a virtual queue vector \( Y_f(t) \), which is given by

\[
Y_f(t+1) = [Y_f(t) + \varphi_f(t) - x_f(t)]^+, \; \forall f \in \mathcal{F}. \tag{15}
\]

Let \( \Xi(t) = (Q(t), Y(t)) \) denote the queue backlogs. We first write the conditional Lyapunov drift for slot \( t \) as

\[
\Delta(\Xi(t)) = \mathbb{E}\left[ L(\Xi(t+1)) - L(\Xi(t)) \| \Xi(t) \right]. \tag{16}
\]

where \( \mathbb{E}[\Xi(t)] = \frac{1}{2} \left[ \sum_{f=1}^{F} \sum_{i=0}^{B} Q_f(t)^2 + \sum_{f=1}^{F} Y_f(t)^2 \right] \) is the quadratic Lyapunov function of \( \Xi(t) \) [30]. By applying the Lyapunov drift-plus-penalty technique [8], [30], the solution of (14) is obtained by minimizing the Lyapunov drift and a penalty from the objective function, i.e.,

\[
\min \quad \Delta(\Xi(t)) - \nu \mathbb{E}[U_0(\bar{\varphi}) | \Xi(t)], \quad \text{subject to} \quad \varphi_f(t) \geq 0, \; \forall f \in \mathcal{F}. \tag{17}
\]

Here, \( \nu \) is a control parameter to trade off utility optimality and queue length [8], [30]. Moreover, the stability of \( \Xi(t) \) ensures that the constraints of problem (8c) and (14b) are held. Noting that \( (a^+)^2 \leq a^2 \) and \( (a \pm b)^2 \leq a^2 \pm 2a b + b^2 \) for any real positive number \( a, b \), and thus, by neglecting other indexes \( t, f, \ldots \), we have the following inequalities:

\[
\left( Y_f - x_f \right)^2 - Y_f^2 \leq 2Y(\varphi - x) + (\varphi - x)^2. \tag{18}
\]

Subsequently, following the calculations of the Lyapunov optimization [30], choosing that \( \varphi \in \mathcal{R} \) and a feasible \( \pi \) and all possible \( \Xi(t) \) for all \( t \), we obtain

\[
\min \quad \psi \quad \text{subject to} \quad \nu U(\varphi_f) \quad \text{subject to} \quad (12), \; (13), \tag{19}
\]

Thus, the solution to (14) can be obtained by minimizing the upper bound in (18).

For every slot \( t \), observing \( \Xi(t) \), we have three decoupled subproblems and provide the solutions for each subproblem as follows. The flow-split vector and the probability distribution are determined by

\[
\text{SP1: min} \quad \pi \quad \text{subject to} \quad (8e), \tag{20}
\]

where

\[
N_f = \sum_{i=1}^{B} Q_f(t)^2 \sum_{m=1}^{M} \pi_f^m \left( r_f^{(i,j)} - r_f^{(i,j)} \right) \tag{21}
\]

Then, we select the optimal auxiliary variables by solving

\[
\text{SP2: min} \quad \varphi_f \quad \text{subject to} \quad \varphi_f(t) \geq 0, \; \forall f \in \mathcal{F}. \tag{22}
\]

Let \( \varphi_f^t \) be the optimal solution obtained by the first order derivative of the objective function of SP2. Assuming a logarithmic utility function, we have

\[
\varphi_f^t = \max \left[ \nu U_f, 0 \right]. \tag{23}
\]

Finally, the RA is done by assigning transmit power, which is obtained by
A. Path Selection

Recall that \( z_f \) represents the flow-split vector given to flow \( f \) and \( z^{m}_{f} = 1 \) when path \( m \) is used to send data for flow \( f \). The MBS selects paths for each flow with a given probability (mixed strategy) [31]. We denote \( u^{m}_{f} = u_{f} (z^{m}_{f}, x^{m}_{f}) \) as a utility function of flow \( f \) when using path \( m \). The vector \( z^{m}_{f} \) denotes the flow-split vector excluding path \( m \). The MBS can choose more than one path to deliver data, from SP1, the utility gain of flow \( f \) is

\[
u_{f} = \sum_{m} u^{m}_{f} = -N_{f}.
\]

To exploit the historical information, the MBS determines a flow-split vector for each flow \( f \) from \( Z_{f} \) based on the PMF from the previous stage \( t = 1 \), i.e.,

\[
\pi_{f}(t-1) = (\pi^{1}_{f}(t-1), \ldots, \pi^{Z_{f}}_{f}(t-1)).
\]

Here, we define \( \Phi_{f}(t) = (\Phi^{1}_{f}(t), \ldots, \Phi^{m}_{f}(t), \ldots, \Phi^{Z_{f}}_{f}(t)) \) as a regret vector of determining flow-split vector for flow \( f \). The MBS selects the flow-split vector with highest regret in which the mixed-strategy probability is given as

\[
\pi^{m}_{f}(t) = \frac{[\Phi^{m}_{f}(t)]^{+}}{\sum_{m' \in Z_{f}} [\Phi^{m'}_{f}(t)]^{+}}.
\]

Let \( \hat{\Phi}_{f}(t) = (\hat{\Phi}^{1}_{f}(t), \ldots, \hat{\Phi}^{m}_{f}(t), \ldots, \hat{\Phi}^{Z_{f}}_{f}(t)) \) be the estimated regret vector of flow \( f \). Basically, with the goal of maximizing the cumulative reward in SP1, the MBS (agent) has to discover the possible paths (action set) in order to find the best paths (distribution of actions with higher pay-off) in the long run [31]. If the MBS spends much time on discovering paths (called exploration), it leads to longer convergence time. If the MBS only exploits an action (called exploitation), which gave the highest pay-off at the beginning, it may loose a chance to obtain higher reward later. Hence, balancing the trade-off between exploration and exploitation is fundamental for efficient learning. For these purpose, we have adopted the logit of Boltzmann-Gibbs (BG) kernel to efficiently learn the best paths [31, 32], \( \beta^{m}_{f}(\hat{\Phi}_{f}(t)) \), given by

\[
\beta^{m}_{f}(\hat{\Phi}_{f}(t)) = \arg \max_{\pi_{f} \in \Pi} \sum_{m \in Z_{f}} [\pi^{m}_{f}(t) \hat{\Phi}^{m}_{f}(t) - \kappa_{f} \pi^{m}_{f}(t) \ln(\pi^{m}_{f}(t))],
\]

where the trade-off factor \( \kappa_{f} \) is used to balance between exploration and exploitation [67], [32], [68]. If \( \kappa_{f} \) is small, the MBS selects \( z_{f} \) with highest payoff. For \( \kappa_{f} \rightarrow \infty \) all decisions have equal probability.

For a given set of \( \hat{\Phi}_{f}(t) \) and \( \kappa_{f} \), we solve (21) to find the probability distribution in which the solution determining the disjoint paths for each flow \( f \) is given as

\[
\beta^{m}_{f}(\hat{\Phi}_{f}(t)) = \frac{\exp \left( \frac{1}{\kappa_{f}} \left[ \Phi^{m}_{f}(t) \right]^{+} \right) \sum_{m' \in Z_{f}} \exp \left( \frac{1}{\kappa_{f}} \left[ \Phi^{m'}_{f}(t) \right]^{+} \right)}{\sum_{m \in Z_{f}} \exp \left( \frac{1}{\kappa_{f}} \left[ \Phi^{m}_{f}(t) \right]^{+} \right)}.
\]

We denote \( \hat{u}(t) \) as the estimated utility of flow \( f \) at time instant \( t \) with action \( z_{f} \), i.e., \( \hat{u}_{f}(t) = (\hat{u}^{1}_{f}(t), \ldots, \hat{u}^{m}_{f}(t), \ldots, \hat{u}^{Z_{f}}_{f}(t)) \). Upon receiving the feedback, \( \hat{u}_{f}(t) \) denotes the utility observed by flow \( f \), i.e., \( \hat{u}_{f}(t) = u_{f}(t-1) \), we propose the learning mechanism at each time instant \( t \) as follows.

**Learning procedure**: The estimates of the utility, regret, and probability distribution functions are performed, and are updated for all actions per path \( m \) as follows:

\[
\begin{align*}
\hat{u}^{m}_{f}(t) &= \hat{u}^{m}_{f}(t-1) + \gamma^{1}(t) \{ \pi^{m}_{f}(t) - \hat{\pi}^{m}_{f}(t-1) \}, \\
\Phi^{m}_{f}(t) &= \Phi^{m}_{f}(t-1) + \gamma^{2}(t) \{ \hat{u}^{m}_{f}(t) - \hat{u}^{m}_{f}(t) - \hat{\pi}^{m}_{f}(t-1) \}, \\
\pi^{m}_{f}(t) &= \pi^{m}_{f}(t-1) + \gamma^{3}(t) \{ \beta^{m}_{f}(\hat{\Phi}_{f}(t)) - \pi^{m}_{f}(t-1) \},
\end{align*}
\]

Here, \( \gamma^{1}(t), \gamma^{2}(t), \) and \( \gamma^{3}(t) \) are the learning rates (please see Section V for more details and convergence proof). Based on the probability distribution as per (23), the MBS determines the flow-split vector for each flow \( f \). The learning-aided PS is performed in a long-term period to ensure that the paths do not suddenly change, and thus, the SCBSSs have sufficient time to deliver data. For instance, at the beginning of the large time scale, the best paths are selected, and will be used for the rest of these large scale time slots as shown in Fig. 2.

B. Rate Allocation

Consider \( r_{(i,j)} = \log(1 + p^{f}_{(i,j)} | g_{(i,j)}(h) |^{2}) \) as the transmission rate, where the effective channel gain \( 4 \) for mmWave channels can be modeled as \( | g_{(i,j)}(h) |^{2} = \frac{10}{10 + | h_{i,j}(t) |^{2}} \) [19], [8]. Here, \( g_{(i,j)}(h) \) and \( f^{\text{max}} \) denote the normalized channel gain and the maximum interference, respectively. Denoting the left hand side (LHS) of (12) and (13) as \( D_{f}^{i} \) for simplicity, the optimal values of flow control \( x \) and transmit power \( p \) in the subproblem 3 (SP3) are found by minimizing

\[
\begin{align*}
\min_{x,p} & \sum_{f=1}^{F} -Y_{f} x_{f}, \quad \text{subject to} \quad 1 + p^{f}_{(i,j)} | g_{(i,j)}(h) |^{2} \geq e^{x_{f}}, \forall f \in \mathcal{F}, i = 0, \quad \text{(24a)} \\
1 + p^{f}_{(i,j)} | g_{(i,j)}(h) |^{2} &\geq e^{D_{f}^{i}}, \forall i : B, \quad \text{(24b)} \\
\sum_{f \in \mathcal{F}} p^{f}_{(i,j)} &\leq p^{\text{max}}_{i}, \forall i \in B, \forall f \in \mathcal{F}. \quad \text{(24c)}
\end{align*}
\]

\( 4 \)The effective channel gain captures the path loss, channel variations, and interference penalty (Here, the impact of interference is considered small due to highly directional beamforming and high pathloss for interfered signals at mmWave frequency band, and thus a multi-hop directional transmission can be operated at dense mmWave networks [35], [36], [38], [41], [37]).
The constraint (24c) is non-convex, motivated by the low-complexity of SCA method, we solve (24) by replacing (24c) with its proper convex approximation [69], [8], [49]. Since it is very hard to find the convex approximation of (24c) [33], [70], we introduce the slack variable \( y \) to transform (24c) into equivalent constraints, which having a proper bound satisfying the conditions in [33, Property A] as

\[
\frac{2 + p_f(t, f, l)^I |g_f(t, f, l)|^2}{2} \geq \sqrt{y^2 + \frac{p_f(t, f, l)^I |g_f(t, f, l)|^2}{2}}, \quad (25)
\]

and

\[
\frac{1 + \rho_f(t, f, l)^I |g_f(t, f, l)|^2}{2} \geq e^{-D_f l}. \quad (26)
\]

Here, the constraint (25) holds a form of the second-order cone inequalities [70], [33], [71], while the LHS of constraint (26) is a quadratic-over-affine function which is iteratively replaced by the first order to achieve a convex approximation as follows:

\[
\frac{2 + p_f^{(i)}(t, f, l)^I |g_f^{(i)}(t, f, l)|^2}{2} \geq y^{(i)} + \frac{p_f^{(i)}(t, f, l)^I |g_f^{(i)}(t, f, l)|^2}{2}, \quad (27)
\]

Here, the superscript \( l \) denotes the \( l \)th iteration. Hence, we iteratively solve the approximated convex problem of (24) as Algorithm 1 in which the approximated problem\(^5\) is given as

\[
\min_{x, p, \pi} \sum_{f=1}^{F} -Y_f x_f \quad (28)
\]

subject to (7), (24d), (24b), (25), (27).

Finally, the information flow diagram of the learning-aided PS and RA approach.

Algorithm 1 Iterative RA

Initialization: set \( l = 0 \) and generate initial points \( y^{(0)} \).
repeat
    Solve (28) with \( y^{(i)} \) to get the optimal value \( y^{(i)\ast} \).
    Update \( y^{(i+1)} := y^{(i)\ast} \); \( l := l + 1 \).
until Convergence

PS and RA approach is shown in Fig. 2, where the RA is executed in a short-term period. Note that the PS and RA are both done at the MBS, in this work we assume that the information is shared among the base stations by using the X2 interface. As opposed to a brute-force approach yielding the global optimal solution, the proposed iterative solution that uses time scale separation remarkably reduces the search time and computational complexity, while obtaining an efficient suboptimal solution.

V. PERFORMANCE ANALYSIS

In this section, we provide a comprehensive performance analysis of our proposed Lyapunov optimization based framework. We show that there exists an \( O(1/v) \), \( O(v) \] utility-queue backlog trade-off, where \( v \) is the Lyapunov control parameter [30]. Next, we present the conditions that ensure that the proposed learning-based PS converges with probability one. Finally, a convergence analysis and a complexity computation of the SOCP based approximation method for RA sub-problem are studied.

A. Queue and Utility Performance

We scrutinize the performance analysis of our proposed algorithm and prove that the queues are stable as per the following theorem.

Theorem 1. [Optimality] Assume that all queues are initially empty. For arbitrary arrival rates, the PS and RA are chosen to satisfy (18) and the rate regime. For a given constant \( \chi > 0 \), the network utility maximization with any \( v > 0 \) provides the following utility performance with \( \chi \) - approximation

\[
U_0 \geq U_0^* - \frac{\Psi + \chi}{v},
\]

where \( U_0^* \) is the optimal network utility over the rate regime.

Proof: We first prove the queues are bounded. Let \( x_f \) denote the largest right derivative of \( U(\tilde{x}_f) \), the Lyapunov framework can guarantee the following strong stability of the virtual queues and the network queues as follows

\[
Q_f(t) \leq v x_f + a_{\text{max}} f, \quad (29)
\]

\[
Y_f(t) \leq v x_f + a_{\text{max}} f. \quad (30)
\]

Here, we first prove the bound of the virtual queues, and then the bound of the network queues are proved similarly. Suppose that all queues are initially empty at \( t = 1 \), this clearly holds for \( t = 1 \). Suppose these inequalities hold for some \( t > 1 \), we need to show that it also holds for \( t + 1 \).

From (15), if \( Y_f(t) \geq v x_f \) then \( Y_f(t+1) \leq v x_f + a_{\text{max}} f \) and the bound holds for \( t + 1 \) due to the rate constraint \( x_f(t) \leq a_{\text{max}} f \). Else, if \( Y_f(t) \geq v x_f \), since the value of auxiliary variables is determined by maximized \( \sum_{f=1}^{F} Y_f(t) \) \( \varphi_f(t) - \nu U_0(\varphi(t)) \varphi(t) \) is then forced to be zero. From (15), \( Y_f(t+1) \) is bounded by...
and choosing
By taking expectations of both sides of the above inequality
are bounded, for a given
\( \chi \)
\( \phi \)
By taking the sum over \( \tau \)
\( U \)
function, the auxiliary variable is chosen to satisfy
as
\( t \)
completes the proof of the
\( \Delta(t) \)
Similarly, we can prove that the network queue (29) is stable with network queues in (5) and (6).
We have established the network bounds, we are now going to show the utility bound. Since our solution of (8) is to minimize the Lyapunov drift and the objective function every time slot \( t \), we have the following inequality given all existing
\( \Xi(t) \)
for all \( t \),
\[
\Delta(t) - \nu \mathbb{E}[U_0(\phi(t))|\Xi(t)] \leq \Psi + \chi - \nu U_0^* - \mathbb{E}[U_0(\phi(t))|\Xi(t)].
\]
By using the definition of the Lyapunov drift and taking an expectation, obtaining
\[
\mathbb{E}[L(\Xi(t))] \leq Ct.
\]
As the definition of the Lyapunov function \( L(\Xi(t)) \), \( \forall i \in B \) we have
\[
\mathbb{E}[O_i(t)^2], \mathbb{E}[Y(t)^2] \leq 2Ct.
\]
Dividing both sides by \( t^2 \), and taking the square roots shows for all \( t > 0 \) and \( \forall i \in B \):
\[
\frac{\mathbb{E}[O_i(t)]}{t}, \frac{\mathbb{E}[Y(t)]}{t} \leq \sqrt{\frac{2C}{t}}.
\]
As \( t \to \infty \), taking the limit, we prove the queues are stable.

B. Learning Convergence Conditions
Due to the space limitation, the complete convergence conditions can be found in [32]. Here, we briefly establish the convergence conditions to the \( \omega \)-coarse correlated equilibrium for the reinforcement learning based algorithm, where \( \omega \) is a very small positive value [72]. The complete proof was studied in [32], [68], the learning rates \( \eta(t) \), \( \eta_2(t) \), and \( \eta_3(t) \) are chosen to satisfy the convergence conditions as follows:
\[
\begin{align*}
\lim_{t \to \infty} \sum_{\tau=0}^t \eta(t) \tau &= +\infty, \\
\lim_{t \to \infty} \sum_{\tau=0}^t \eta_2(t) \tau &= +\infty, \\
\lim_{t \to \infty} \sum_{\tau=0}^t \eta_3(t) \tau &= +\infty,
\end{align*}
\]
\[
\begin{align*}
\lim_{t \to \infty} \sum_{\tau=0}^t \eta(t) \tau &= +\infty, \\
\lim_{t \to \infty} \sum_{\tau=0}^t \eta_2(t) \tau &= +\infty, \\
\lim_{t \to \infty} \sum_{\tau=0}^t \eta_3(t) \tau &= +\infty,
\end{align*}
\]
\[
\begin{align*}
\lim_{t \to \infty} \eta(t) \tau &= 0, \\
\lim_{t \to \infty} \eta_2(t) \tau &= 0.
\end{align*}
\]

C. Convergence Analysis of SOCP based Algorithm 1
We establish a convergence result for Algorithm 1 based on the SOCP approach. By using the SOCP approach, we have approximated the original non-convex problem (24) by a strongly convex problem (28). We briefly describe the convergence for the sake of completeness since it was studied in [33]. We assume that the Algorithm 1 obtains the solution of problem (28) at iteration \( l + 1 \) th. The updating rule in Algorithm 1 ensures that the optimal values \( y(t) \) at iteration \( l \) satisfy all constraints in (28) and are feasible to the optimization problem at iteration \( l + 1 \). Therefore, the objective obtained in the \( l + 1 \)st iteration is less than or equal to that in the in the \( l \)th iteration, since we minimize the linear function. In other words, Algorithm 1 yields a non-increasing sequence. Due to the transmit power constraints and rate constraints, the objective is bounded, and thus Algorithm 1 converges to some local optimal solution of (28). Moreover, Algorithm 1 produces a sequence of points that are feasible for the original problem (24) and this solution satisfies the Karush–Kuhn–Tucker (KKT) condition of the original problem (24) as discussed in [33].
VI. Numerical Results

In this section Monte Carlo simulations are carried out in order to evaluate the system performance of our proposed algorithm. To solve Algorithm 1, we use YALMIP toolbox to model the optimization problem with MOSEK as internal solver [73]. For simulations, we assume that there are two flows from the MBS to two UEs, while the number of available paths for each flow is four [27]. The MBS selects two paths from four most popular paths\(^6\). Each path contains two relays, the total number of SCBSs is 8, and the one-hop distance is varying from 50 to 100 meters. The maximum transmit power of MBS and each SC are 43 dBm and 30 dBm, respectively, and the SC antenna gain is 5 dB. The number of antennas \(N_b\) at each BS is set to 8 and 64 for small and large antenna arrays, respectively. The number of antennas \(N_c\) at UE is set to 2 and 16, for small and large antenna arrays, respectively. The number of RF chains at BS \(N_b\) and UE \(N_c\) are set to 8 and 2, respectively.

For simulations purposes, the general channel model for arbitrary antenna arrays is used. In particular, the estimate channel matrix \(\hat{H}_{i,j} \in \mathbb{C}^{N_i \times N_j}\) of the channel matrix \(H_{i,j} \in \mathbb{C}^{N_i \times N_j}\) between the transmitter \(i\) and the receiver \(j\) can be modeled as [57], [74]

\[
\hat{H}_{i,j} = \sqrt{N_i \times N_j} \Theta_{i,j}^{1/2} \left(1 - \tau_j^2 W_{i,j} + \tau_j W_{i,j} \right),
\]

where \(W_{i,j} = \left[w_{i,j}^1, \ldots, w_{i,j}^{N_i}, \ldots, w_{i,j}^{N_j} \right] \in \mathbb{C}^{N_i \times N_j}\) is the small-scale fading channel matrix, which is independent and identically distributed (i.i.d.) with zero mean and variance \(\frac{\kappa}{N_i \times N_j}\) in which \(w_{i,j} \sim \mathcal{CN}(0,1)\) is the small-scale fading channel vector between the transmitter antenna array and the \(n_{i,j}\) antenna of receiver \(j\). Here, \(\tau_j \in [0, 1]\) reflects the estimation accuracy for receiver \(j\), if \(\tau_j = 0\), then \(\hat{H}_{i,j} = H_{i,j}\), the perfect channel state information is assumed at the transmitters [75]. \(\hat{H}_{i,j} \in \mathbb{C}^{N_i \times N_j}\) is the estimated noise, also modeled as a realization of the circularly symmetric complex Gaussian distribution matrix with zero mean and variance of \(\frac{\kappa}{N_i \times N_j}\) [8], [57]. Moreover, \(\Theta_{i,j} \in \mathbb{C}^{N_i \times N_j}\) depicts the antenna spatial correlation matrix that accounts for the path loss and shadow fading, such that \(\text{Rank}(\Theta_{i,j}) = N_i\).

We generate the spatial correlation matrix as \(\Theta_{i,j} = PL_{i,j} \Theta_{i,j}\) with \(\text{Rank}(\Theta_{i,j}) = R_i\), and the normalized spatial correlation matrix with \(\text{Tr}(\Theta_{i,j}) = N_i\) [74]. The mmWave path loss \(PL_{i,j}\) is modeled as a distance-based path loss for urban environments at 28 GHz with 1 GHz system bandwidth [76], [77], which may exist as a line-of-sight (LOS), non-LOS (NLOS), or blockage states. We adopt the mmWave channel model used in the system level simulation in [76], given by

\[
PL(d) = Pr(d)PL_{\text{LOS}}(d) + (1 - Pr(d))PL_{\text{NLOS}}(d),
\]

where \(PL_{\text{LOS}}(d)\) and \(PL_{\text{NLOS}}(d)\) are the distance-based path loss for LOS and NLOS states at distance \(d\), respectively [76]. Here, \(Pr(d)\) denotes a boolean random variable that is 1 with some probability. For the general blockage channel model,

\[72\)

\[61\)

\[^{6}\]As studied in [27], it suffices for a flow to maintain at least two paths provided that it repeatedly selects new paths at random and replaces if the latter provides higher throughput.

\[^{7}\]The analog end-to-end delay is defined as the sum of the average one-hop delay of all hops.

\[^{8}\]A simulation source code can be found in [78], which consists of a set of simple functions that allows to learn the path/route and allocate the transmit power in our paper.

<table>
<thead>
<tr>
<th>Path loss model</th>
<th>Values in dB</th>
<th>Bandwidth (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS @ 28 GHz</td>
<td>61.4 + 20 log(d)</td>
<td>1</td>
</tr>
<tr>
<td>NLOS @ 28 GHz</td>
<td>72 + 29.2 log(d)</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter settings</th>
<th>Path loss model [76],[77]</th>
<th>Values in dB</th>
<th>Bandwidth (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS @ 28 GHz</td>
<td>61.4 + 20 log(d)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NLOS @ 28 GHz</td>
<td>72 + 29.2 log(d)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

The LOS probability is defined as \(\exp(-0.006d)\), then the NLOS probability is \(1 - \exp(-0.006d)\) [76], [77]. For the analog beamforming, the side lobe gain \(\Gamma\) is set to \(\frac{1}{3}\), and the beamwidths at the transmitter and receiver are set to \(\frac{\pi}{4}\) and \(\frac{\pi}{3}\) radians, respectively.

We assume that the traffic flow is divided equally into two sub-flows, the arrival rate for each sub-flow is varying from 2 to 5 Gbps for small antenna array case. The maximum delay requirement \(\beta\) and the target reliability probability \(\epsilon\) are set to be 10 ms and 5\%, respectively [11]. For the learning algorithm, the Boltzmann temperature (trade-off factor) \(\eta_f\) is set to 5, while the learning rates \(\eta_f^1(t), \eta_f^2(t), \text{and} \eta_f^3(t)\) are set to \(\frac{1}{(t+1)^{\eta_f}}, \frac{1}{(t+1)^{\eta_f^{2/3}}}, \text{and} \frac{1}{(t+1)^{\eta_f}}, \text{respectively}\) [68], [12].

The parameter settings\(^7\) are summarized in Table II. To that end, we would like to notice that our work contains some main features: (i) NUM [30], [40], (ii) dynamic path selection learning [32], and (iii) URLLC-aware rate allocation [11]. We consider the following baselines: Baseline 1 employs features (i) and (ii), whereas Baseline 2 applies features (i) and (iii), finally Baseline 3 considers only feature (i). We benchmark our work and these baselines to assess the impact of the dynamic path selections and of the URLLC-constrained rate allocation, which has not been addressed in the literature in the context of mmWave communications. In addition, Single hop scheme considers that the MBS delivers data to UEs over one single hop at long distance in which the probability of LOS communication is low, and then the blockage needs to be taken into account [76].

A. Small Antenna Array System

We first evaluate the network performance under the small antenna array setting, i.e., \(N_f = 8, N_j = 2\). In Fig. 3, we report the average one-hop delay\(^8\) versus the mean arrival rates \(\bar{\mu}\). As we increase \(\bar{\mu}\), baselines 3, 2, and 1 violate the latency constraints at \(\bar{\mu} = 3.5, 4.5, \text{and} 5\) Gbps, respectively. While the average delay of our proposed algorithm is gradually increased with \(\bar{\mu}\), but under the warming level, \(\beta = 10\) ms. The reason is that the delay requirement is satisfied via the equivalent instantaneous rate by our proposed algorithm as per (12) and (13), while the baselines 1 and 3 use the traditional utility-delay trade-off approach without considering the latency constraint, and the baseline 2 considers the random PS mechanism only. The benefit of applying the learning path algorithm is that selecting the path with high payoff and less congestion, results in small latency. Let us now take a look at \(\bar{\mu} = 4.5\) Gbps, the average one-hop delay of baseline 1
with learning outperforms baselines 2 and 3, whereas our proposed scheme reduces latency by 50.64\%, 81.32\% and 92.9\% as compared to baselines 1, 2, and 3, respectively. When $\bar{\mu} = 5$ Gbps, the average delay of all baselines increases dramatically, violating the delay requirement of 10 ms, while our proposed scheme is robust to the latency requirement.

In Fig. 4, we report the tail distribution (complementary cumulative distribution function (CCDF)) of latency to showcase how often the system achieves a delay greater than the target delay levels [79] as $\bar{\mu} = 4.5$ Gbps, $\epsilon = 5\%$, $\beta = 10$ ms. In contrast to the average delay, the tail distribution is an important metric to reflect the URLLC characteristic. For instance, at $\bar{\mu} = 4.5$ Gbps, by imposing the probabilistic latency constraint, our proposed approach ensures reliable communication with better guaranteed probability, i.e., Pr(delay > 10 ms) < $10^{-6}$.

In contrast, baseline 1 with learning violates the latency constraint with high probability, where Pr(delay > 10 ms) = 0.08 and Pr(delay > 25 ms) < $10^{-6}$, while the performance of baselines 2 and 3 gets worse. For instance, as shown in Fig. 4, baselines 2 and 3 obtain Pr(delay > 10 ms) > 0.12 and Pr(delay > 10 ms) > 0.24, respectively. For throughput comparison, we observe that for $\bar{\mu} = 4.5$ Gbps, our proposed algorithm is able to deliver 4.4874 Gbps of average network throughput per each sub-flow, while the baselines 1, 2, and 3 deliver 4.4759, 4.4682, and 4.3866 Gbps, respectively. Here, the Single hop scheme only delivers 3.55 Gbps due to the high path loss, causing large latency.

Note that in this work we mainly focus on the low latency scale, i.e., 1 – 10 ms, the target achievable rate for all schemes is very high and close to each other. Hence, we report the average MBS queue length instead of the average achievable rate. Generally speaking, as per (5), the average achievable rate can be extracted from the average MBS queue length and the mean arrival rate, i.e., $\bar{x}_f = \bar{\mu} - \bar{Q}_f$. In Fig 5, we plot the average queue length of the MBS as a function of mean arrival rates. As we increase the mean arrival rate from 2 to 5 Gbps, the average MBS queue length of our proposed algorithm is increased from 0.01 Gb to 0.04 Gb, which means that the average delay at the MBS is increased from 5 ms to 8 ms, which meet the latency constraint (8b). In contrast, the average queue length of the baselines is increased up to 16 ms, which violates the latency constraint (8b).

In Fig. 6, we report the tail distribution of the one-hop latency (logarithmic scale) versus the guaranteed probability $\epsilon$ as $\beta = 10$ ms, $\kappa = 5$, and $\bar{\mu} = 4.5$ Gbps. By varying $\epsilon$ from 0.05 to 0.15, the system is allowed to achieve a delay greater than the target latency with higher probability. As can be seen in Fig. 6, the probability that the system achieves a latency greater than 4 ms increases from less than 1 % to 8 % when increasing $\epsilon$ from 0.05 to 0.15. This indicates the trade-off between reliability and latency, if we loose the reliability requirement, latency is higher.
In this work, we have applied the Boltzmann-Gibbs technique to capture the trade-off between exploration and exploitation as per (21). We run the simulations for different values of $\kappa \in \{2, 5, 10, 20\}$ for all flow $f$. As expected, with a small value of $\kappa$, the MBS decides to use the paths with highest payoff, selected at the beginning (a small probability of exploration). In this case, the algorithm converges faster, but lacks exploration, the MBS will not try other paths, which may exploit path diversity; as shown in Fig. 11, small value of $\kappa$, results in higher delay in the long run. By increasing $\kappa$, the MBS exploits the network environment with higher probability. The benefits of exploration are to utilize the path diversity, improving the performance, i.e., low latency and reducing congestion at the BSs. As shown in Fig. 11, average latency is decreased with $\kappa$, and a large value of $\kappa$ incurs
Next, we plot the convergence of the iterative algorithm as a function of the number of hops as shown in Fig. 10. Here, we provide the distribution of the number of iterations of the SOCP-based algorithm in which the convergence criteria stops running with an accuracy of $10^{-2}$. With increasing the number of hops, the number of constraints and variables is increased, and thus the number of iterations required by the algorithm for convergence is higher. Intuitively, our proposed algorithm only needs few iteration to converge at each time slot $t$ as shown in Fig. 10. For example, for three hop transmission, the probability that the number of iterations takes a value less than or equal to 7 is 90%.

D. Impact of the Learning Temperature

In addition to the previous discussion on the impact of the trade-off parameter on the convergence, in Fig. 11, we report the average one-hop latency versus the learning trade-off parameter $\kappa$ as $\varepsilon = 5\%$, $\beta = 10$ ms, and $\bar{\mu} = 3.5$ Gbps. It can be observed that at small $\kappa$, slowly increasing $\kappa$ the MBS is allowed to explore other paths to get higher gain in the long run. Hence, the average one-hop latency gradually reduces with small increased $\kappa$. However, when $\kappa$ is very large, four paths are determined uniformly for two flows, which becomes random PS. For instance, when $\kappa = 50$, the average delay is much higher. Hence, it can be observed that the average delay is a convex function of $\kappa$ in which there exists an optimal value for $\kappa$.

VII. CONCLUSION

In this paper, the authors proposed a multi-hop multi-path scheduling to support reliable communication by incorporating the probabilistic latency constraint and traffic splitting techniques in 5G mmWave networks. In particular, the problem was modeled as a network utility maximization subject to a bounded latency with a guaranteed reliability probability, and network stability. The authors employed massive MIMO and mmWave communication techniques to further improve the DL transmission of a multi-hop self-backhauled small cells. By leveraging stochastic optimization, the problem was decoupled into PS and RA, which are solved by applying the reinforcement learning and successive convex approximation methods, respectively. A comprehensive performance analysis of our proposed algorithm was mathematically provided. Numerical results shown that our proposed framework reduces latency by 50.64% and 92.9% as compared to the baselines with and without learning, respectively.

ACKNOWLEDGMENT

The authors would like to acknowledge colleagues Chen-Feng Liu, Cristina Perfecto, Mohammed Elbamby, Kien Giang Nguyen, Satya Joshi, Quang-Doanh Vu, Sumudu Samarakoon,
and Jhong Park at CWC, University of Oulu for helpful discussions on the paper. Further, the authors thank the editor and all anonymous reviewers for their constructive comments. They have been very helpful in the revision of this paper and allowed us to improve the technical contents and presentation quality.

REFERENCES


Trung Kien Vu received the B.Eng. degree from the School of Electronics and Telecommunications, Hanon University of Science and Technology, Vietnam, in 2012, and the M.Sc. degree in electrical engineering from the School of Electrical Engineering, University of Ulsan, South Korea, in 2014. From 2015 to 2018, he pursued the D.Sc. degree with the Centre for Wireless Communications (CWC), University of Oulu, Finland. He was a Visitor at Tsinghua University, Beijing, China from Dec. 2018 to Mar. 2019. His research interests focus on applications of stochastic optimization and deep reinforcement learning for 5G wireless networks and beyond. Kien Vu received the 2016 European Wireless Best Paper Award and was a recipient of the Nokia Foundation grant, the Finnish Foundation grant for Technology Promotion, and the 2014 Brain Korea 21 Plus (BK21+) Scholarship.
Mehdi Bennis (’S’07–’AM’08–SM’15) received the M.Sc. degree from the École Polytechnique Fédérale de Lausanne, Switzerland, and the Eurecom Institute, France, in 2002, and the Ph.D. degree in electrical engineering in 2009 focused on spectrum sharing for future mobile cellular systems. He is currently an Associate Professor with the University of Oulu, Finland, and an Academy of Finland Research Fellow. He has authored or co-authored over 100 research papers in international conferences, journals, book chapters, and patents. His main research interests are in radio resource management, heterogeneous networks, game theory, and machine learning. He has given numerous tutorials at the IEEE flagship conferences. He was also a recipient of several prizes, including the 2015 Fred W. Ellersick Prize from the IEEE Communications Society (ComSoc), the 2016 IEEE COMSOC Best Tutorial Prize, the 2017 EURASIP Best Paper Award for the *Journal on Wireless Communications and Networking*, and recently the Best Paper Award at the European Conference on Networks and Communications 2017.

Mérouane Debbah (’S’01–’AM’03–M’04–SM’08–F’15) received the M.Sc. and Ph.D. degrees from the École Normale Supérieure Paris-Saclay, France. In 1996, he joined the École Normale Supérieure Paris-Saclay. He was with Motorola Labs, Saclay, France, from 1999 to 2002, and also with the Vienna Research Center for Telecommunications, Vienna, Austria, until 2003. From 2003 to 2007, he was an Assistant Professor with the Mobile Communications Department, Institut Eurecom, Sophia Antipolis, France. From 2007 to 2014, he was the Director of the Alcatel-Lucent Chair on Flexible Radio. Since 2007, he has been a Full Professor with CentraleSupélec, Gif-sur-Yvette, France. Since 2014, he has been a Vice-President of the Huawei France Research Center and the Director of the Mathematical and Algorithmic Sciences Lab. He has managed 8 EU projects and more than 24 national and international projects. His research interests lie in fundamental mathematics, algorithms, statistics, information, and communication sciences research. He is an IEEE Fellow, a WWRF Fellow, and a Membre émérite SEE. He was a recipient of the ERC Grant MORE (Advanced Mathematical/Tools for Complex Network Engineering) from 2012 to 2017. He was a recipient of the Mario Boella Award in 2005, the IEEE Glavieux Prize Award in 2011, and the Qualcomm Innovation Prize Award in 2012. He received 19 best paper awards, among which the 2007 IEEE GLOBECOM Best Paper Award, the Wi-Opt 2009 Best Paper Award, the 2010 Newcom+ Best Paper Award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC Best Paper Award, the 2015 ICC Best Paper Award, the 2015 IEEE Communications Society Leonard G. Abraham Prize, the 2015 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications Society Best Tutorial Paper Award, the 2016 European Wireless Best Paper Award, the 2017 EURASIP Best Paper Award, the 2018 IEEE Marconi Prize Paper Award, and the Valuetools 2007, Valuetools 2008, CrownCom 2009, Valuetools 2012, SAM 2014, and 2017 IEEE Sweden VTCOM-IT Joint Chapter best student paper awards. He is an Associate Editor-in-Chief of the journal Random Matrix: Theory and Applications. He was an Associate Area Editor and Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2011 to 2013 and from 2013 to 2014, respectively.

Matti Latva-aho received the M.Sc., Lic.Tech. and Dr. Tech (Hons.) degrees in Electrical Engineering from the University of Oulu, Finland in 1992, 1996 and 1998, respectively. From 1992 to 1993, he was a Research Engineer at Nokia Mobile Phones, Oulu, Finland after which he joined Centre for Wireless Communications (CWC) at the University of Oulu. Prof. Latva-aho was Director of CWC during the years 1998–2006 and Head of Department for Communication Engineering until August 2014. Currently he serves as Academy of Finland Professor in 2017–2022 and is Director for 6Genesis - Finnish Wireless Flagship for 2018–2026. His research interests are related to mobile broadband communication systems and currently his group focuses on 5G and beyond systems research. Prof. Latva-aho has published 350+ conference or journal papers in the field of wireless communications. He received Nokia Foundation Award in 2015 for his achievements in mobile communications research.

Mériam Bennis (S’01–AM’08–SM’15) received the M.Sc. and Ph.D. degrees from the École Normale Supérieure Paris-Saclay, France. In 1996, she joined the École Normale Supérieure Paris-Saclay. She was with Motorola Labs, Saclay, France, from 1999 to 2002, and also with the Vienna Research Center for Telecommunications, Vienna, Austria, until 2003. From 2003 to 2007, she was an Assistant Professor with the Mobile Communications Department, Institut Eurecom, Sophia Antipolis, France. From 2007 to 2014, she was the Director of the Alcatel-Lucent Chair on Flexible Radio. Since 2007, she has been a Full Professor with CentraleSupélec, Gif-sur-Yvette, France. Since 2014, she has been a Vice-President of the Huawei France Research Center and the Director of the Mathematical and Algorithmic Sciences Lab. She has managed 8 EU projects and more than 24 national and international projects. Her research interests lie in fundamental mathematics, algorithms, statistics, information, and communication sciences research. She is an IEEE Fellow, a WWRF Fellow, and a Membre émérite SEE. She was a recipient of the ERC Grant MORE (Advanced Mathematical/Tools for Complex Network Engineering) from 2012 to 2017. She was a recipient of the Mario Boella Award in 2005, the IEEE Glavieux Prize Award in 2011, and the Qualcomm Innovation Prize Award in 2012. She received 19 best paper awards, among which the 2007 IEEE GLOBECOM Best Paper Award, the Wi-Opt 2009 Best Paper Award, the 2010 Newcom+ Best Paper Award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC Best Paper Award, the 2015 ICC Best Paper Award, the 2015 IEEE Communications Society Leonard G. Abraham Prize, the 2015 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications Society Best Tutorial Paper Award, the 2016 European Wireless Best Paper Award, the 2017 EURASIP Best Paper Award, the 2018 IEEE Marconi Prize Paper Award, and the Valuetools 2007, Valuetools 2008, CrownCom 2009, Valuetools 2012, SAM 2014, and 2017 IEEE Sweden VTCOM-IT Joint Chapter best student paper awards. She is an Associate Editor-in-Chief of the journal Random Matrix: Theory and Applications. She was an Associate Area Editor and Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2011 to 2013 and from 2013 to 2014, respectively.