Wireless Communications and Mobile Computing

# Recent Advances in 5G Technologies: New Radio Access and Networking

Lead Guest Editor: Shao-Yu Lien
Guest Editors: Chih-Cheng Tseng, Ingrid Moerman, and Leonardo Badia

# Recent Advances in 5G Technologies: New Radio Access and Networking

Wireless Communications and Mobile Computing

# Recent Advances in 5G Technologies: New Radio Access and Networking

Lead Guest Editor: Shao-Yu Lien
Guest Editors: Chih-Cheng Tseng, Ingrid Moerman, and Leonardo Badia

# Editorial Board

# Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 8202048, 2 pages
https://doi.org/10.1155/2019/8202048

*Editorial*

# Recent Advances in 5G Technologies: New Radio Access and Networking

**Shao-Yu Lien**,[1] **Chih-Cheng Tseng**,[2] **Ingrid Moerman**,[3] **and Leonardo Badia**[4]

[1]*Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi 62102, Taiwan*
[2]*Department of Electrical Engineering, National Ilan University, Ilan 26041, Taiwan*
[3]*Department of Information Technology, Ghent University, 9000 Ghent, Belgium*
[4]*Department of Information Engineering, University of Padova, 2-35122 Padova, Italy*

Correspondence should be addressed to Shao-Yu Lien; shaoyulien@gmail.com

Received 16 January 2019; Accepted 19 January 2019; Published 10 February 2019



In Sept. 2015, the International Telecommunication Union-Radiocommunications Standardization Sector (ITU-R) has released the service recommendations of the fifth generation (5G) mobile networks known as International Mobile Telecommunications 2020 (IMT-2020). Instead of solely boosting the data rates, like the past evaluations from IMT-2000 to IMT-Advanced, an IMT-2020 system shall support three categories of wireless scenarios, including enhanced mobile broadband (eMBB), ultra-reliability and low latency communication (URLLC), and massive machine-type communication (mMTC), to sustain the 20 Gbps peak data rate, 100 Mbps user experienced data rate, 10 Mbps/$m^2$ area traffic capacity, $10^6$ devices/$km^2$ connection density, 1 ms latency, and 500 km/hr mobility. To compete for being an IMT-2020 system, 3GPP consequently launched the normative works of "New Radio (NR)" in Release 15 and Release 16. In Jun. 2018, Phase I normative work of NR (i.e., Release 15) has completed, and Phase II (i.e., Release 16) has subsequently begun. The feature technologies in NR thus include communications using millimeter/centimeter wave carriers (spectrum above 6 GHz), nonorthogonal multiple access (NOMA), advanced vehicle-to-everything (V2X), directional transmission/reception, software-defined network (SDN), etc.

To service the urgent needs in normative works of NR, this special issue thus aims at bringing together the state-of-the-art innovations, research activities (both in academia and industry), and the corresponding standardization impacts of NR, so as to comprehend the inspirations, requirements, implementation, and the promising technical options to boost, practice, and deploy the NR.

In the paper titled "Exploiting Impacts of Intercell Interference on SWIPT-Assisted Non-Orthogonal Multiple Access," the influence of intercell interference (ICI) on the system outage behavior with important derived results in the proposed model of simultaneous wireless information and power transfer (SWIPT) together with NOMA using the amplify-and-forward protocol is examined. The authors further derive the closed-form expression of coverage probability for two NOMA users as a function of the signal-to-interference-plus-noise ratio (SINR) and investigate the average outage probability by considering impacts of the reasonable number of participating ICI.

In the paper titled "MC-GiV2V: Multichannel Allocation in mmWave-Based Vehicular Ad Hoc Networks," a Giga-V2V (GiV2V) network is proposed, in which vehicles query and deliver high quality video and sensor data of smart and self-driving cars using mmWave communications instead of current dedicated short-range communications (DSRC). Vehicles probably form a grid topology along lanes of a road, which leads to align mmWave beams of the vehicles and cause mutual interference among them. As channel diversity can resolve effectively the interference between mmWave beams, several heuristic algorithms for channel assignment of each beam in the GiV2V networks are also proposed.

In the paper titled "Micro Operator Design Pattern in 5G SDN/NFV Network," the authors discuss the deployment of Micro Operator ($\mu$O) to reduce network latency in response

to the low-latency applications for future 5G edge computing environment. The authors consequently address the design pattern of 5G micro operator and propose a Decision Tree Based Flow Redirection (DTBFR) mechanism to redirect the traffic flows to neighbor service nodes. The proposed DTBFR mechanism thus allows different $\mu$Os to share network resources and speed up the development of edge computing in the future.

In the paper titled "Energy-Efficient Uplink Resource Units Scheduling for Ultra-Reliable Communications in NB-IoT Networks," the issue of how to guarantee the reliable communication and satisfy the quality of service (QoS) while minimizing the energy consumption for IoT devices is studied. The authors model the problem as an optimization problem and prove it to be NP-complete and then propose an energy-efficient, ultra-reliable, and low-complexity scheme. Extensive simulation is also conducted to show that the provided scheme can serve more devices with guaranteed QoS while saving their energy effectively.

In the paper titled "Genetic Algorithm-Based Beam Refinement for Initial Access in Millimeter Wave Mobile Networks," the initial access issue in 5G networks operating at carrier frequencies is investigated. The authors extend the proposed genetic algorithm- (GA-) based beam refinement scheme to include beamforming at both the transmitter and the receiver and compare the performance with alternative approaches in the millimeter wave multiuser multiple-input-multiple-output (MU-MIMO) networks. The effect of different parameters such as the number of transmit antennas/users/per-user receive antennas, beamforming resolutions, and hardware impairments on the system performance employing different beam refinement algorithms is investigated and shows that the proposed GA-based approach performs well in delay-constrained networks with multi-antenna users.

In the paper titled "RF Driven 5G System Design for Centimeter Waves," the authors describe their experiences in developing a centimeter waves mobile broadband concept satisfying future capacity requirements. The first step in the process is the radio channel measurement campaign and statistical modeling. Then the link level design is performed tightly together with the radio frequency (RF) implementation requirements to allow as large scalability of the air interface as possible. The authors started the concept development at 10 GHz frequency band and during the project World Radiocommunication Conference 2015 selected somewhat higher frequencies as new candidates for 5G. The main learning is to gain insight of interdependencies of different phenomena and find feasible combinations of techniques and parameter combinations that might actually work both in practice and in theory.

## Conflicts of Interest

This is to confirm that there are no conflicts of interest.

*Shao-Yu Lien*
*Chih-Cheng Tseng*
*Ingrid Moerman*
*Leonardo Badia*

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2525492, 12 pages
https://doi.org/10.1155/2018/2525492

*Research Article*

# Exploiting Impacts of Intercell Interference on SWIPT-Assisted Non-Orthogonal Multiple Access

**Thanh-Luan Nguyen**[1] **and Dinh-Thuan Do**[2]

[1]*Faculty of Electrical and Electronics Engineering, Bach Khoa University, Ho Chi Minh City, Vietnam*
[2]*Wireless Communications Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Correspondence should be addressed to Dinh-Thuan Do; dodinhthuan@tdt.edu.vn

Received 11 February 2018; Accepted 6 November 2018; Published 25 November 2018

Guest Editor: Shao-Yu Lien



In this paper, we examine the influence of intercell interference (ICI) on the system outage behavior with important derived results in the proposed model of simultaneous wireless information and power transfer (SWIPT) together with the nonorthogonal multiple access (NOMA) using the amplify-and-forward protocol. We derive the closed-form expression of coverage probability for two NOMA users as a function of the signal-to-interference-plus-noise ratio (SINR). To fully take into account the effect of ICI, we adopt more practical parameters to evaluate the optimal power splitting coefficient regarding energy harvesting system performance analysis. Furthermore, to consider a more practical scenario, based on the fact that the number of ICI sources can affect wireless powered relays, we investigate the average outage probability by considering impacts of the reasonable number of participating ICI.

## 1. Introduction

In recent years, as a significant technique in forthcoming 5G transmission, nonorthogonal multiple access (NOMA) has more attraction due to its opportunity to improve substantial spectrum efficiency (SE) [1]. Considering advantages of power domain allocation to gain simultaneous access for diversified data streams to network, NOMA is evaluated as having better performance than the conventional orthogonal multiple access (OMA) [2]. Concentrating on the SE improvement, many earlier works regarding NOMA scheme are primarily focused. For example, single-input, single-output (SISO) schemes are studied for deployment in cooperative NOMA [3, 4]. In principle, by distributing multiple users into the different power domains, signal intending for the NOMA scheme at the source superimposes the symbol data to serve multiple users at destination. At the receiver, in order to detect multiplexed users' information, successive interference cancellation (SIC) is acquired to eliminate the interference term. In addition to spectrum employment, other metrics need be considered, i.e., user fairness evaluating to multiple users in NOMA. Unlike traditional water-filling power allocation, NOMA users with better channel qualities are allocated less power while more power is allocated to users with poor channel qualities to highlight an enhanced trade-off between model throughput and user fairness. As a result, the same frequency and spreading codes are deployed in a lot of user equipment simultaneously; however they are distinguished by different power levels. Such principle results in improved spectral efficiency and ensured user fairness. The multiple-input-multiple-output (MIMO) NOMA designs are investigated in terms of the ergodic capacity maximization and an optimal power allocation strategy was then proposed as in [5]. The outage performance and the ergodic achievable rate are metrics to consider system performance of cellular downlink NOMA communications. Capacity is another metric to consider performance of NOMA as in [6]. Recently, stochastic geometry networks have been employed in NOMA to exhibit the physical layer security as in [7]. As an extension of [7], single-antenna and multiple-antenna stochastic geometry networks were investigated in two proposed schemes to increase the secrecy performance [8].

In [9], the popular metrics including optimal designs of decoding order, transmission rates, and power allocated to each user are studied in a new design of NOMA under secrecy considerations.

Regarding the idea of combining NOMA with wireless powered networks, NOMA can be deployed with the wireless power transfer; i.e., the simultaneous wireless information and power transfer (SWIPT) technique was presented together with relaying network and its outage performance is considered in [10]. As a promising technology, to prolong the lifetime of energy-constrained users, wireless power transfer is facilitated in wireless relaying networks to forward a signal to far users [11–13]. To permit energy-limited users scavenge energy and information from the transmitted radio frequency (RF) signals, two famous policies, namely, time switching (TS) and power splitting (PS) receiver architecture, are introduced as in [14]. As a scenario of green NOMA, a SWIPT-assisted cooperative NOMA (SWIPT-CNOMA) network is studied as a combination between the NOMA and SWIPT scheme [15], and such novel scheme can be employed to wireless sensor or cellular networks. In such topology, it is possible to avoid lifetime limitation of the energy-constrained NOMA user that acts as a relay which can be able to harvest energy from the received signals. The main benefit of the SWIPT NOMA can be appreciated in the scenario where the relay employs that harvested energy rather than itself to forward the signal to the far NOMA users. In addition, to serve the near NOMA users with strong channel conditions, a PS protocol was employed in the SWIPT-CNOMA system [15]. By revealing the received information in the PS protocol, energy is first harvested and then used to help the signal forwarded to serve the far NOMA users. However, a shortcoming of the PS protocol [15] can be raised in which the relay always retains a quiet slot for some target data rate requirements. Considering the locations of NOMA users, three NOMA user selection policies are performed, including random near NOMA user and random far NOMA user (RNRF) selection, nearest near NOMA user and nearest far NOMA user (NNNF) selection, and nearest near NOMA user and farthest far NOMA user (NNFF) selection. In order to obtain the better outage performance, the authors in [16] considered a best-near best-far NOMA user collection scheme.

However, there are few works that focus on interference effects on the NOMA system. Regarding the influences of cochannel interference (CCI) on the system performance, many outcomes in the literature showed that the aggressive frequency reuse is considered as main reason for CCI in the conventional cellular relaying systems [17–19]. However, FD schemes are more vulnerable to CCI due to the higher frequency reuse, and such characterization exhibits comparison between full-duplex (FD) transmission mode and traditional half-duplex (HD) transmission mode. In addition, in a multicell FD relay base stations meet influences with a much higher CCI due to adjacent cells compared with its HD counterparts [20–22]. As an alternative method, MIMO FD relaying transmission was introduced to examine the harmful effect of CCI on the system performance and such system can be applied in theoretical and practical applications. In [20], to evaluate the average spectral efficiency, a stochastic geometry is conveyed to consider the situation in which the base stations and user equipment operate in small cell network FD mode with FD mode (i.e., the dedicated antennas for transmission and reception are equipped in such FD nodes). In [21, 22], the authors studied the outage probability of a decode-and-forward FD relay with single-antenna nodes in FD relaying subject to CCI. It is worth noting that all these studies are restricted by employing a single antenna in each node of relaying networks. Therefore, deploying multiple receive and transmit antennas at the FD relay is a powerful scheme to eliminate both the CCI and loop interference channels at the FD relay and hence reliability and capacity are achieved.

However, to the best of our knowledge, the impact of ICI on the performance of CNOMA with capability of energy harvesting has yet to be fully pressed. Such analysis motivates us to find impacts of ICI on the SWIPT-assisted NOMA.

The primary contributions of our paper are summarized as follows:

(i) Different from the system model presented in [10, 14] where the relay only harvests energy from one source (i.e., base station), by deploying ICI as an extra resource to feed energy for relay, we propose a new SWIPT NOMA protocol under impacts of ICI to enhance the level of harvested energy to prolong the lifetime of NOMA systems

(ii) We derive closed-form expressions for the outage probability at NOMA users, when considering the interference channel schemes

(iii) The optimal power splitting factor is derived to obtain the maximum coverage probability at each user in SWIPT NOMA

The rest of the paper is structured as follows. In Section 2, the system model for studying cooperative SWIPT NOMA is introduced. In Section 3, new analytical expressions are derived for the outage probability when the proposed scheme is used. The optimal power splitting factor for SWIPT can be shown in Section 4. Numerical results are offered in Section 5, which is followed by the conclusion in Section 6.

## 2. System Model

Consider a SWIPT NOMA network with the help of an energy-constrained relay for transmission to two representative NOMA users considering impacts of external interferers to the relay. In particular, Figure 1 illustrates the proposed system model, where one source node (s), i.e., the main base station, one EH-assisted relay (r), and two users are considered. The relay utilizes the amplify-and-forward (AF) relaying protocol to transmit the superimposed signals from the source to each user and the power splitting (PS) protocol is adopted for energy harvesting. The channel between node $x$ and node $y$ is assumed to experience Rayleigh fading where $\Omega_{xy} = \mathbb{E}[|h_{xy}|^2]$ is the average channel

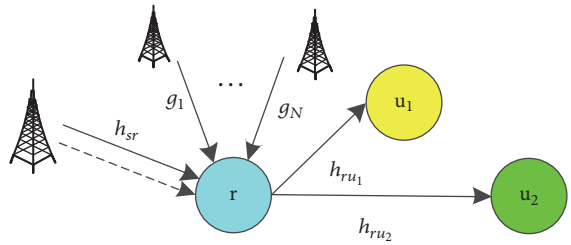



FIGURE 1: System model.

power gain. Each node is equipped with a single half-duplex antenna.

Assume that the relay is affected by $N$ external interferers. The channel coefficient between the $n$-th interferer and the relay is modeled as Nakagami-m fading with shape factor $m_n$ and variance of $\Omega_n = \mathbb{E}[|g_n|^2]$, i.e., $g_n \sim \Gamma(m_n, \Omega_n/m_n)$. All given channels are assumed to be independent and not necessarily identically distributed. The transmission between the source and two users is divided into two phases. In the first phase, the source broadcasts a superimposed mixture of signals, i.e., $x_s = \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2$, where $x_i$, $i \in \{1, 2\}$, is the intended signal for user $u_i$, $\alpha_1 + \alpha_2 = 1$.

Assuming that the relay utilizes the power splitting (PS) protocol for energy harvesting [11], thus the received signal at the relay node is given by

$$y_r = \sqrt{(1-\beta)P_s}x_s h_{sr} + \sum_{n=1}^{N}\sqrt{(1-\beta)P_n}g_n x_n + n_r, \quad (1)$$

where $\beta \in (0, 1)$ is the power splitting ratio, $P_s$ is the transmit power of the source, $x_n$ and $P_n$ are signal and the transmit power of the $n^{th}$ interferer, respectively, and $n_r$ is the zero mean additive white Gaussian noise (AWGN) at the relay with variance of $\sigma^2$. Specifically, the harvested energy at the relay node is obtained as

$$E_h = \eta\beta\left(P_s|h_{sr}|^2 + \sum_{n=1}^{N}P_n|g_n|^2\right)\frac{T}{2}, \quad (2)$$

where $\eta \in [0, 1]$ is a coefficient representing the efficiency of the harvesting circuitry and $T$ is the block time for the transmission from the source node to both users. Therefore, the transmit power at the relay is obtained as

$$P_r = \frac{E_h}{T/2} = \eta\beta\left(P_s|h_{sr}|^2 + \sum_{n=1}^{N}P_n|g_n|^2\right). \quad (3)$$

In the second phase, the relay multiplies the received signal with an amplifying gain G and then forwards the amplified signal to both users. Hence, the received signal at user$i$ is given by

$$\begin{aligned} y_i &= \sqrt{P_r}y_r h_{ru_i}G + n_i \\ &= \sqrt{P_r}h_{ru_i}\left(\sqrt{(1-\beta)P_s}x_s h_{sr}\right)G \\ &\quad + \sqrt{P_r}h_{ru_i}\left(\sum_{n=1}^{N}\sqrt{(1-\beta)P_n}g_n x_n\right)G \\ &\quad + \sqrt{P_r}h_{ru_i}n_r G + n_i, \end{aligned} \quad (4)$$

where $n_i$ is the zero mean additive white Gaussian noise (AWGN) at user $u_i$ with variance of $\sigma^2$. The amplifying factor due to the AF protocol at the relay is given by

$$G = \frac{1}{\sqrt{(1-\beta)P_s|h_{sr}|^2 + (1-\beta)\sum_{n=1}^{N}P_n|g_n|^2 + \sigma^2}}. \quad (5)$$

*Remark 1.* Without loss of generality, we assume that user $u_1$ has better channel quality than user $u_2$; i.e., $|h_{ru_2}|^2 \le |h_{ru_1}|^2$. Hence, due to NOMA, user 1 is encouraged to priorly decode $x_2$ before detecting its own signal by applying successive interference cancellation (SIC), while user 2 can directly detect $x_2$.

Due to Remark 1, the instantaneous signal-to-interference-plus-noise ratio (SINR) at user 1 to detect user 2's signal is given by

$$\gamma_{1\to 2} = \frac{\alpha_2 P_s|h_{sr}|^2}{\alpha_1 P_s|h_{sr}|^2 + \sum_{n=1}^{N}P_n|g_n|^2 + \left(\sigma^2/P_r G^2\right)/(1-\beta)\left|h_{ru_1}\right|^2 + \sigma^2/(1-\beta)}. \quad (6)$$

**Lemma 2.** *The term $\sigma^2/P_r G^2$ can be approximated in the high signal-to-noise ratio (SNR) region, in which the source transmits with a relatively large power, as*

$$\frac{\sigma^2}{P_r G^2} \simeq \frac{(1-\beta)\sigma^2}{\eta\beta}. \quad (7)$$

*Proof.* By substituting (3) and (5) into $\sigma^2/P_r G^2$ we have

$$\begin{aligned} \frac{\sigma^2}{P_r G^2} &= \frac{\left((1-\beta)P_s|h_{sr}|^2 + (1-\beta)\sum_{n=1}^{N}P_n|g_n|^2\right)\sigma^2}{\eta\beta\left(\mathscr{P}_s|h_{sr}|^2 + \sum_{n=1}^{N}P_n|g_n|^2\right)} \\ &\quad + \frac{\sigma^4}{\eta\beta\left(P_s|h_{sr}|^2 + \sum_{n=1}^{N}P_n|g_n|^2\right)}. \end{aligned} \quad (8)$$

Note that, in the high SNR region, the second term, i.e., $\sigma^4/\eta\beta(P_s|h_{sr}|^2 + \sum_{n=1}^{N} P_n|g_n|^2)$, can be neglected. Hence, after ignoring this part, we can simply achieve the result of Lemma 2.

Therefore, the approximated instantaneous $\gamma_{1\longrightarrow 2}$ is obtained by

$$\gamma_{1\longrightarrow 2} \approx \frac{\alpha_2\psi_{sr}}{\alpha_1\psi_{sr} + \psi_I + 1/\psi_{ru_1} + 1/(1-\beta)}, \tag{9}$$

where $\psi_{sr} \triangleq P_s|h_{sr}|^2/\sigma^2$, $\psi_I \triangleq \sum_{n=1}^{N} P_n|g_n|^2/\sigma^2$, and $\psi_{ru_1} \triangleq \eta\beta|h_{ru_1}|^2$. If $\gamma_{1\longrightarrow 2} > 2^{2R_2} - 1$, where $R_2$ (bits/s/Hz) is the target data rate for $u_2$, user 1 can successfully carry out the successive interference cancellation (SIC) technique to detect user 2's signal and remove this signal from $y_1$. Assuming perfect SIC, the approximated instantaneous SINR for user 1 to detect its own signal is given by

$$\gamma_1 \approx \frac{\alpha_1\psi_{sr}}{\psi_I + 1/\psi_{ru_1} + 1/(1-\beta)}. \tag{10}$$

Note that if $\gamma_1 > 2^{2R_1} - 1$, where $R_1$ (bits/s/Hz) is the target data rate for $u_1$, user 1 can successfully decode its own message. Similarly, the SINR at user 2 to detect its signal is obtained as

$$\gamma_2 = \frac{\alpha_2 P_s |h_{sr}|^2}{\alpha_1 P_s |h_{sr}|^2 + \sum_{n=1}^{N} P_n |g_n|^2 + (\sigma^2/P_r G^2)/(1-\beta)\,|h_{ru_2}|^2 + \sigma^2/(1-\beta)}. \tag{11}$$

Define $\psi_{ru_2} \triangleq \eta\beta|h_{ru_2}|^2$; the approximated $\gamma_2$ achieved by adopting Lemma 2 is given by

$$\gamma_2 \approx \frac{\alpha_2\psi_{sr}}{\alpha_1\psi_{sr} + \psi_I + 1/\psi_{ru_2} + 1/(1-\beta)}. \tag{12}$$

□

## 3. Performance Analysis

*3.1. Preliminaries.* In this section, some preliminary results are presented in terms of the solutions of some PDF and/or CDF calculation, which will be frequently invoked in the analysis. Let $Y = \sum_{n=1}^{N} X_n$ be the sum of $N$ independent and not necessarily identically distributed gamma random variables (RVs), i.e., $X_n \sim \Gamma(m_n, \Omega_n/m_n)$, where $\Omega_n = \mathbb{E}[X_n]$.

**Theorem 3.** *The approximated probability density function (PDF) of $Y$ follows the gamma distribution, where $m_I$ is the shape factor and $\Omega_I$ is the variance, which is given by [23–25]*

$$f_Y^{[1]}(\gamma) \approx \frac{1}{\Gamma(m_I)\,\mu_I^{m_I}}\gamma^{m_I-1}\exp\left(-\frac{\gamma}{\mu_I}\right), \quad \gamma \geq 0, \tag{13}$$

*where $\mu_I = \Omega_I/m_I$, $\Omega_I = \sum_{n=1}^{N}\Omega_n$. The shape parameter, $m_I$, is computed by using moment-base estimators, which is given by*

$$m_I = \frac{\Omega_I^2}{\mathbb{E}[\Phi^2] - \Omega_I^2}, \tag{14}$$

*where the first term in the denominator, $\mathbb{E}[\Phi^2]$, can be obtained with the help of*

$$\mathbb{E}\left[\Phi^{k_0}\right] = \sum_{\{k_n\}_{n=1}^{N-1}}^{k_0} \mathbb{E}\left[|X_N|^{2k_{N-1}}\right] \cdot \prod_{t=1}^{N}\binom{k_{t-1}}{k_t}\mathbb{E}\left[|X_{N-1}|^{2(k_{t-1}-k_t)}\right], \tag{15}$$

*in which*

$$\mathbb{E}\left[|X_n|^k\right] = \frac{\Gamma(m_n + k/2)}{\Gamma(m_n)}\left(\frac{\Omega_n}{m_n}\right)^{k/2}, \tag{16}$$

*where $\sum_{\{k_n\}_{n=1}^{N-1}}^{k_0}$ is a short-hand representation of $\sum_{k_1=0}^{k_0}\sum_{k_2=0}^{k_1}\cdots\sum_{k_{N-1}=0}^{k_{N-1}}\cdots$ and $\binom{n}{k}$ denotes the binomial coefficient. In addition, when $m_n$ is a nonnegative integer, the exact PDF and CDF of $Y$ are obtained through Theorem 4 as below.*

**Theorem 4.** *Without loss of generality, we assume that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_N$, where $\mu_n = \Omega_n/m_n (n = 1 \ldots N)$. Therefore, the exact probability density function (PDF) of $Y$, where $m_n \in \mathbb{Z}^+$, is given by*

$$f_Y^{[2]}(\gamma) = \sum_{p=1}^{P}\sum_{q=1}^{m_{\langle p\rangle}} \frac{R_{p,q}(\mathcal{S})}{(q-1)!\mu_{\langle p\rangle}^q}\gamma^{q-1}\exp\left(-\frac{\gamma}{\mu_{\langle p\rangle}}\right), \quad \gamma \geq 0, \tag{17}$$

*in which $\mathcal{S} = \{\mu_n\}_{n=1}^{N}$ is a vector containing all $\mu_n$'s. Without loss of generality, assume that vector $\mathcal{S}$ contains $P$ distinct elements; each element is denoted by $\mu_{\langle p\rangle}$, where $p = 1, 2, \ldots, P$. Intuitively, $m_{\langle p\rangle}$ is the sum of all $m_n$'s ($n \geq 1$), in which the subscript is the index of the $n$-th element of $\mathcal{S}$, i.e., $\mu_n$, that is identical to $\mu_{\langle p\rangle}$. Specifically, the value of $m_{\langle p\rangle}$ is determined through Figure 2 and the below description.*

In Figure 2, $\tau_p$ is the number of elements identical to $\mu_{\langle p\rangle}$ and $s_p \triangleq \sum_{k=1}^{p}\tau_k$ and $s_{p-1} + 1$ denote the indices of the last

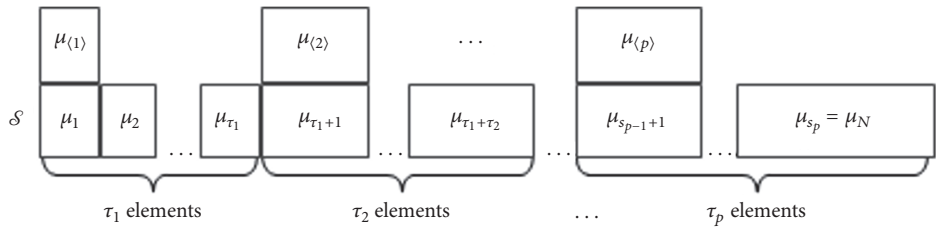



FIGURE 2: A realization between $\mu_{\langle p\rangle}$ and the elements in vector $\mathcal{S}$.

and the first element identical to $\mu_{\langle p\rangle}$, respectively. Therefore, $m_{\langle p\rangle} = \sum_{k=s_{p-1}+1}^{s_p} m_k$. Subsequently, $R_{p,q}(\mathcal{S})$ is obtained by

$$R_{p,q}(\mathcal{S}) = \frac{1}{\left(m_{\langle p\rangle}-q\right)!\mu_{\langle p\rangle}^{m_{\langle p\rangle}-q}} \times \left(\frac{d^{m_{\langle p\rangle}-q}}{dv^{m_{\langle p\rangle}-q}} \prod_{k=1,k\neq p}^{P} \frac{1}{\left(1+\mu_{\langle k\rangle}v\right)^{m_{\langle k\rangle}}}\right)\Bigg|_{v=-1/\mu_{\langle p\rangle}}. \tag{18}$$

*Proof.* See Appendix A. □

The closed-form expression of (18) is given in Appendix B in order to find the values of $R_{p,q}(\mathcal{S})$ without requiring any differentiation manipulation. In addition, Remark 5 is given in order to provide some insights of Theorem 4.

*Remark 5.* Theorem 4 is priorly defined for $m_n \in \mathbb{Z}^+$; however in case of $m_n \in \mathbb{R}^+$ one can still utilize Theorem 4 to find the sum of $N$ i.n.i.d. gamma RVs as long as it satisfies $m_{\langle p\rangle} \in \mathbb{Z}^+$. This statement can be understood by observing the agreement between the simulation curves and analytical curves in Section 5. Further, it is noticed from Appendix A that (A.3) can also extend to the case of $m_n \in \mathbb{R}^+$ without any modification. However, in order to adopt partial fraction decomposition (PFD) to achieve the closed-form expression for $f_Y^{[2]}(\gamma)$ in the third equality of (A.4), $m_{\langle p\rangle} \in \mathbb{Z}^+$ must be a positive integer, which can be easily achieved by assuming $m_n \in \mathbb{Z}^+$ is also a positive integer.

Furthermore, in order to support the analysis in Section 3.2 we introduce Theorem 6 as below.

**Theorem 6.** *Let* $\mathrm{I}_Y(\kappa) \triangleq \mathbb{E}_Y[e^{-\kappa Y}]$*; its closed-form expression is obtained as*

$$\mathrm{I}_{\psi_I}(\kappa) = \int_0^\infty \mathrm{e}^{-\kappa\gamma} f_Y^{[t]}(\gamma)\, d\gamma \tag{19}$$

$$= \begin{cases} \left(1+\mu_I\kappa\right)^{-m_I} & (t=1) \\ \sum_{p=1}^{P}\sum_{q=1}^{m_{\langle p\rangle}} \dfrac{R_{p,q}(\mathcal{S})}{\left(1+\mu_{\langle p\rangle}\kappa\right)^q} & (t=2). \end{cases} \tag{20}$$

*Proof.* Using the given PDF of $\psi_I$ in (13) for $t = 1$ and (17) for $t = 2$, with the help of [26, Eq. 3.351.3], we can easily achieve the above results.

In addition, when the interferences are statistically independent and identically distributed (i.i.d.), i.e., $\mu_1 = \mu_2 = \ldots = \mu_N = \mu$, then $P = 1$ and $m_{\langle p\rangle} = \sum_{n=1}^{N} m_n \triangleq m$; $\mathrm{I}_Y(\kappa)$ becomes

$$\mathrm{I}_{\psi_I}(\kappa) = \left(1+\mu\kappa\right)^{-m}. \tag{21}$$

When $N = 0$, i.e., $Y = 0$, it immediately follows that $\mathrm{I}_{\psi_I}(\kappa) = 1$. □

*3.2. Performance Analysis on Outage Probability at Each User.* In this section, the analytical result for the coverage probability of user $u_i$, denoted by $\overline{\mathcal{O}}_i$, is given in closed form. Hence, the outage probability at user $u_i$ is $\mathcal{O}_i = 1 - \overline{\mathcal{O}}_i$. It is worth noticing that the results in this section are obtained when $\alpha_2 > \tau_2\alpha_1$ and if $\alpha_2 < \tau_2\alpha_1$ the coverage probability of each user becomes zero. Subsequently, the coverage probability of user 1 is defined as the probability that this user successfully decodes both $x_2$ and $x_1$ which is mathematically given by

$$\overline{\mathcal{O}}_1 = \Pr\left\{\gamma_{1\longrightarrow 2} > \tau_2, \gamma_1 > \tau_1\right\}, \tag{22}$$

where $\tau_2 \triangleq 2^{2R_2} - 1$ and $\tau_1 \triangleq 2^{2R_1} - 1$. However, deriving the exact $\overline{\mathcal{O}}_1$ in closed form is not tractable due to the complexity of (6); thus we then derive the approximated coverage probability in the high SNR region by substituting (9) and (10) into (22). Note that this assumption is widely used in literature. Subsequently, this probability is approximately obtained in closed form as

$$\overline{\mathcal{O}}_1 \simeq \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_1}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_1}{\mu_{sr}}\frac{1}{1-\beta}\right) \times \sum_{k=1}^{3}(-1)^{k-1}\, 2\sqrt{\frac{\mathrm{T}_1}{\mu_{sr}\mu_k}}K_1\left(2\sqrt{\frac{\mathrm{T}_1}{\mu_{sr}\mu_k}}\right), \tag{23}$$

where $\mu_{sr} = P_s\Omega_{sr}/\sigma^2$, $\mu_{ru_1} = \eta\beta\Omega_{ru_1}$, $\mu_{ru_2} = \eta\beta\Omega_{ru_2}$, $\mathrm{T}_1 = \max(\tau_2/(\alpha_2 - \tau_2\alpha_1), \tau_1/\alpha_1)$, $\mu_1 = \mu_{ru_1}$, $\mu_2 = \mu_{ru_1}\mu_{ru_2}/\mu_{ru_2} + \mu_{ru_1}$, and $\mu_3 = \mu_{ru_2}$, $\mathrm{I}_{\psi_I}(\kappa)$ is obtained in Theorem 6 and Theorem 4 by setting $\Omega_I = \sum_{n=1}^{N} P_n\Omega_n/\sigma^2$ for $(t = 1)$ and $m_I = \sum_{n=1}^{N} P_n\Omega_n/\sigma^2$, $K_\nu(z)$ is the $\nu^{th}$ order modified Bessel function of the second kind [26], and $\alpha_2 > \tau_2\alpha_1$.

*Proof.* See Appendix C. □

Hence, the coverage probability of user 2 is defined as the probability that this user successfully decodes its own message, $x_2$, and can be represented mathematically by

$$\overline{\mathcal{O}}_2 = \Pr\left\{\gamma_2 > \tau_2\right\}. \tag{24}$$

Subsequently, the closed-form expression for the approximated $\overline{\mathcal{O}}_2$ is given by

$$\begin{aligned}\overline{\mathcal{O}}_2 \simeq{}& \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_2}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_2}{\mu_{sr}}\frac{1}{1-\beta}\right)\\ &\cdot 2\sqrt{\frac{\mathrm{T}_2}{\mu_{sr}\mu_2}}K_1\left(2\sqrt{\frac{\mathrm{T}_2}{\mu_{sr}\mu_2}}\right),\end{aligned} \tag{25}$$

where $\mathrm{T}_2 = \tau_2/\alpha_2 - \tau_2\alpha_1$ and $\alpha_2 > \tau_2\alpha_1$.

*Proof.* Substituting (12) into (22), we then obtain

$$\begin{aligned}\overline{\mathcal{O}}_2 &\simeq \Pr\left\{\psi_{sr} > \mathrm{T}_2\left(\psi_I + \frac{1}{\psi_{ru_2}} + \frac{1}{1-\beta}\right)\right\}\\ &\simeq \exp\left(-\frac{\mathrm{T}_2}{\mu_{sr}}\frac{1}{1-\beta}\right)\int_0^\infty \exp\left(-\frac{\mathrm{T}_2}{\mu_{sr}}x\right) f_{\psi_I}^{[t]}(x)\,dx\\ &\quad\times\int_0^\infty \exp\left(-\frac{\mathrm{T}_2}{\mu_{sr}}\frac{1}{y}\right) f_{\psi_{ru_2}}(y)\,dy.\end{aligned} \tag{26}$$

In addition, the first integral is obtained by adopting (20) when $\kappa = \mathrm{T}_2/\mu_{sr}$. Recall that $|h_{ru_2}|^2 \le |h_{ru_1}|^2$ due to Remark 1; thus $\psi_{ru_2} \le \psi_{ru_1}$. Using order statistic in [27], the PDF of the ordered $\psi_{ru_2}$ is given by

$$f_{\psi_{ru_2}}(y) = \frac{1}{\mu_2}\exp\left(-\frac{y}{\mu_2}\right), \quad (y \ge 0). \tag{27}$$

Therefore, the closed-form expression for the second integral is obtained by using (27) and [26, Eq. 3.471.9]. Hence, the proof is done. □

## 4. Optimal Power Splitting Factor Problem

In this section, the maximum coverage probability at a single user is obtained by adjusting the power splitting factor to maximize the SNRs/SINRs given in Section 3. Specifically, the optimal power splitting factor to achieve the maximum SNRs/SINRs at user $u_i$ is given by

$$\beta_i^{\mathrm{op}} = \arg\max_{0<\beta<1}\left(\varphi_i\right), \tag{28}$$

where $\varphi_1 \triangleq \min\{\gamma_{1\longrightarrow 2}, \gamma_1\}$ and $\varphi_2 \triangleq \gamma_2$. Since $\beta$ only appears in the denominator of $\varphi_i$ we can rewrite (28) as

$$\beta_i^{\mathrm{op}} = \arg\min_{0<\beta<1}\left(\psi_i(\beta) \triangleq \frac{1}{\eta\beta\left|h_{ru_i}\right|^2} + \frac{1}{1-\beta}\right). \tag{29}$$

**Lemma 7.** *The function $\psi_i(\beta)$ is convex.*

*Proof.* Since $1/\eta\beta|h_{ru_i}|^2$ and $1/1-\beta$ are convex functions of $\beta \in (0,1)$ and the sum of two convex functions is also convex, thus $\psi_i(\beta)$ is convex.

From Lemma 7, it is clear that (30) has a unique solution for $\beta_i^{\mathrm{op}}$ which is obtained by solving $(\partial/\partial\beta)\psi_i(\beta) = 0$. Subsequently, after some simple algebraic manipulations, the optimal energy harvesting factor is given as below:

$$\beta_i^{\mathrm{op}} = \frac{1}{1+\sqrt{\eta\left|h_{ru_i}\right|^2}}. \tag{30}$$

Substituting (30) into (9), (10), and (12), the optimal coverage probability of user $u_i$ is given approximately as

$$\begin{aligned}\overline{\mathcal{O}}_i^{op} \simeq{}& \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_i}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}}\right)\sum_{k=i}^{4-i}(-1)^{k+\bar{i}-1}\\ &\cdot\left\{\times\omega_{i,k}K_1\left(\omega_{i,k}\right) - \sqrt{\pi\frac{\mathrm{T}_i}{\mu_{sr}\eta}}\omega_{i,k}\exp\left(-\omega_{i,k}\right)\right\},\end{aligned} \tag{31}$$

where $\omega_{i,k} = 2\sqrt{\mathrm{T}_i/\mu_{sr}\eta\Omega_k}$, $\Omega_1 = \Omega_{ru_1}$, $\Omega_2 = \Omega_{ru_1}\Omega_{ru_2}/\Omega_{ru_1} + \Omega_{ru_2}$, $\Omega_3 = \Omega_{ru_2}$, and $\bar{i} = 2/i$. □

*Proof.* See Appendix D. □

Note that the relay can only choose to tune $\beta$ to achieve the best performance for a single user in a specific block time, T; thus the other user may not reach maximum coverage probability during that block time. In this case, the coverage probability of the other user is given approximately by

$$\begin{aligned}\overline{\mathcal{O}}_{\bar{i}} \simeq{}& \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\\ &\cdot\sum_{k_{\bar{i}}=\bar{i}}^{4-\bar{i}}\sum_{k_i=i}^{4-i}(-1)^{k_{\bar{i}}+i-1}(-1)^{k_i+\bar{i}-1}\\ &\times\left[\left(1-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\sqrt{\frac{\pi}{\eta\Omega_{k_i}}}\right)\omega_{\bar{i},k_{\bar{i}}}K_1\left(\omega_{\bar{i},k_{\bar{i}}}\right)\right.\\ &\left.-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}\Omega_{k_{\bar{i}}}}\sqrt{\frac{\pi\Omega_{k_i}}{\eta}}K_0\left(\omega_{\bar{i},k_{\bar{i}}}\right)\right],\end{aligned} \tag{32}$$

where $\omega_{\bar{i},k_{\bar{i}}} = 2\sqrt{\mathrm{T}_{\bar{i}}/\mu_{sr}\eta\Omega_{k_{\bar{i}}}}$.

*Proof.* See Appendix E. □

## 5. Numerical Results

In this section, in terms of the outage probability of both users in the cooperative SWIPT NOMA network, we present representative numerical results to demonstrate the performance assessments. In the considered SWIPT NOMA network under impacts of ICI, we set the energy conversion efficiency of SWIPT as $\eta = 1$. We define SNR $\triangleq P_s\Omega_{sr}/\sigma^2$

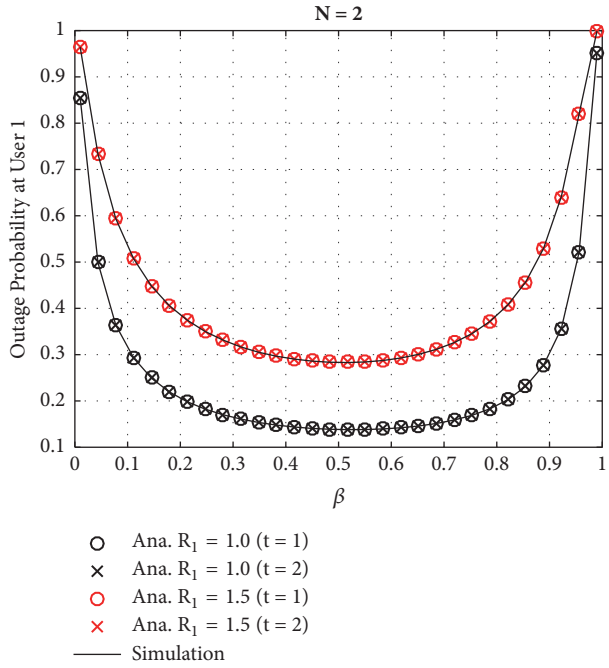



Figure 3: Outage probability at user 1 vs. the energy harvesting ratio $\beta$, where SNR = 30 dB and INR = −10 dB.

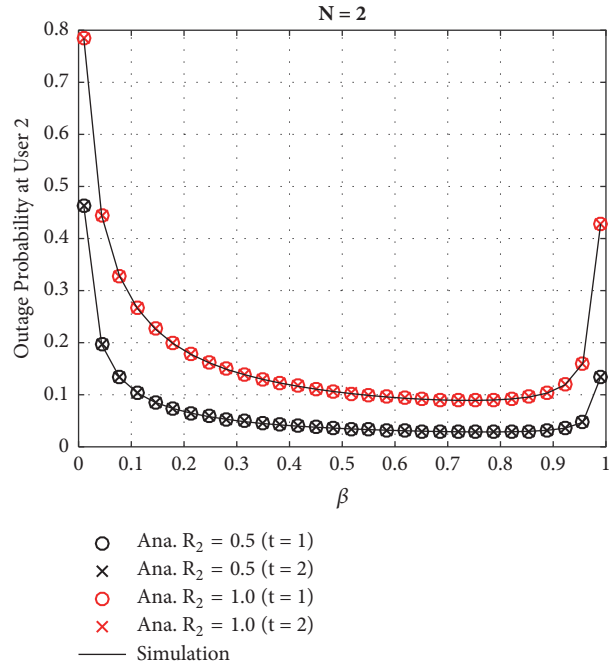



Figure 4: Outage probability at user 2 vs. the energy harvesting ratio $\beta$, where SNR = 30 dB and INR = −10 dB.

as the average SNR at the transmitter (source); assume that $P_n = P_I$ and define INR $\triangleq \sum_{n=1}^{N} P_I \Omega_n / \sigma^2$ as the average interference-to-noise ratio. $\Omega_{sr} = 1$, $\Omega_{ru_1} = 1.5$, and $\Omega_{ru_2} = 0.5$. From Figures 3–8, the number of external interferers is set to 2 ($N = 2$) with $m_1 = 1.2$, $m_2 = 1.8$, $\Omega_1 = 1.2$, and $\Omega_2 = 1.8$; the power allocation coefficients are $\alpha_1 = 0.1$ and $\alpha_2 = 0.9$. The below results are obtained by averaging 500,000 separated simulations under a MATLAB environment.

Firstly, we consider the results in Figures 3 and 4. The curves denoted by (O) can be obtained from (23) with $I_{\psi_I}(\kappa)$ in (20) achieved at $t = 1$ whereas the curves denoted by (X) can be obtained with $I_{\psi_I}(\kappa)$ obtained at $t = 2$ (20). As a clear observation, the simulation lines obtained via (22) and (24), respectively, strictly match with the analytical lines for both $t = 1$ and $t = 2$, which confirms the accuracy of our derivation. Furthermore, Figure 3 shows the outage probability for user 1, as increasing $R_1$ to serve this user with higher data rate or higher quality of service and outage performance will be the worse case. In such result, the black curve is $R_1 = 1.0$ bits/s/Hz while the red curve is $R_1 = 1.5$ bits/s/Hz. One can observe that a lower outage probability is achieved by changing approximate $\beta = 0.6 \longrightarrow 0.8$. The figures also demonstrate that as $\beta$ is very high or very low, a higher outage event can occur due to the low harvested energy at the relay where it makes impacts on the end-to-end SNR of the system. Similarly, outage probability (not optimal) is shown for user 2 in Figure 4. It can be observed that as $R_2$ is increased then outage performance will be worse; in this case, we set $R_2 = 0.5$ bits/s/Hz as the black curves and the red ones corresponding with $R_2 = 1.0$ bits/s/Hz.

In Figure 5, the outage probabilities for user 1 are achieved by varying INR levels and are shown as functions of the

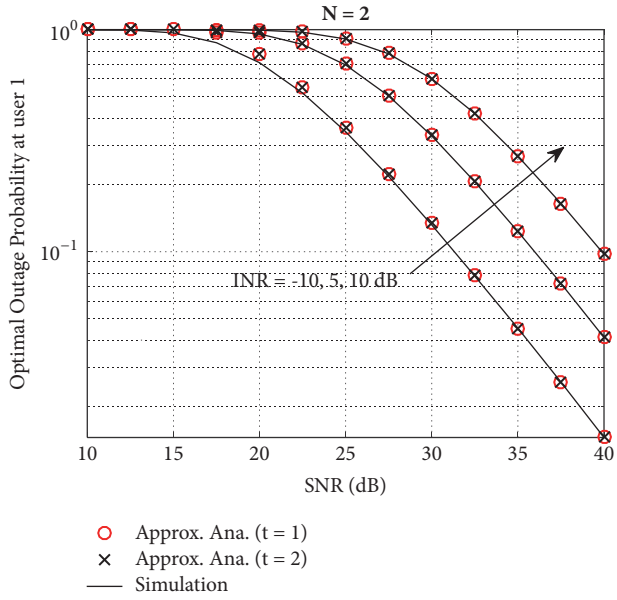



Figure 5: Optimal outage probability at user 1 vs. the average SNR.

average SNR. The simulation curves are obtained by adopting (22) with $\beta$ defined in (28) while the analytical curves are obtained from (31) with $i = 1$. As can be seen from the figure, NOMA with an INR level equal to −10 dB outperforms the other remaining scenarios, since it can ensure that the outage is achievable by all the users as controlling impacts of ICI.

Figure 6 demonstrates the outage performance for user 2 versus SNR with different INR levels. Similarly, the simulation curves can be obtained via (24) with the optimal $\beta$ in

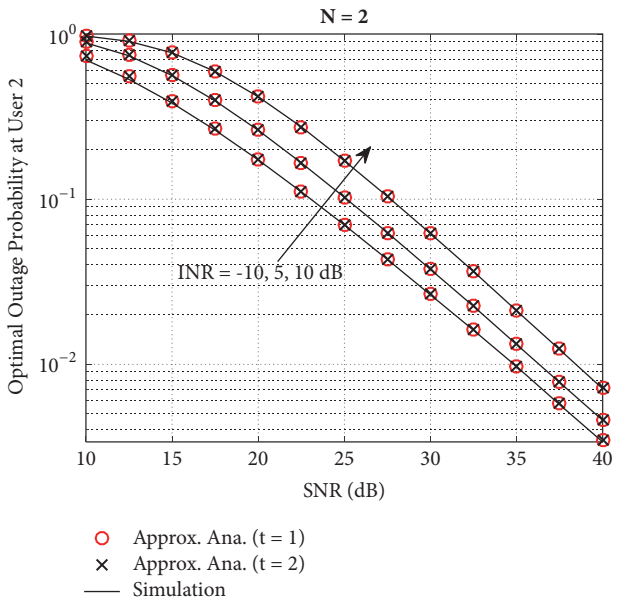



FIGURE 6: Optimal outage probability at user 2 vs. the average SNR.

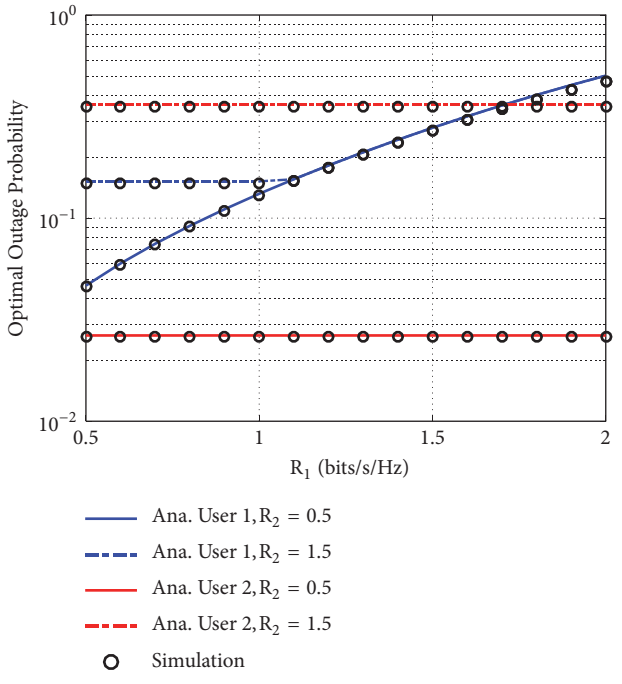



FIGURE 7: Optimal outage probability vs. user 1's target data rate.

(28) while the analytical curves are obtained from (31) with $i = 2$. One can observe that the proposed method achieves the lowest outage since it has the lowest impacts on the system among three scenarios. The figure also demonstrates the existence of the outage ceilings in the low SNR region (i.e., SNR less than 10 dB). This is due to the fact that the system probability is approaching an outage event and such outage is determined only by the SNR while other parameters do not affect outage. It is worth noting that increasing SNR can improve the outage; however, for the case INR $= -10$ dB,

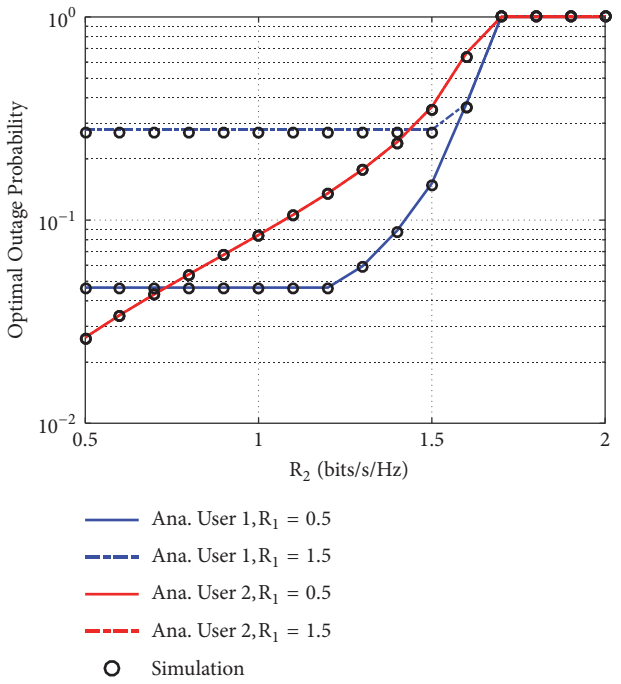



FIGURE 8: Optimal outage probability vs. user 2's target data rate.

the throughput is the lowest among three scenarios. This is because, in the latter case, the lower impact for information processing at user 2, and hence an outage, will be improved. Therefore, we see that it is important to select appropriate ICI when designing practical NOMA downlink transmission systems.

Figures 7 and 8 illustrate the optimal outage probability of both users versus the target rates, i.e., $R_1$ and $R_2$, in two different time blocks. The red and blue solid lines are both obtained from (31) when $i = 1$ (ensuring QoS at user 1) and $i = 2$ (ensuring QoS at user 2), respectively, while the simulation results for user 1 and user 2 are obtained by substituting $\beta = (1 + \sqrt{\eta|h_{ru_1}|^2})^{-1}$ into (22) and $\beta = (1 + \sqrt{\eta|h_{ru_2}|^2})^{-1}$ into (24), respectively. It is worth noting that, in the same signal block, the relay only satisfies QoS criteria for one user only; however we intend to compare the optimal outage performance each user can achieve. Firstly, one can observe that the outage probabilities of user 2 and user 1 can increase as $R_2$ and $R_1$ increase, respectively. The reason is that increasing $R_2$ and $R_1$ can lead to the higher threshold of decoding and therefore can result in more outage. More specifically, the outage probability at user 2 increases as $R_2$ increases but it remains constant varying $R_1$ as depicted in Figure 8, where the solid line overlaps the dashed line. The reason is that the performance of this user only depends on the data rate of $x_2$ but not that of $x_1$. In addition, the outage probability at user 1 depends highly on $T_1$; thus if $T_1 = \tau_1/\alpha_1$ increasing $R_2$ will not affect the outage probability at this user resulting in a segment of straight blue lines in Figure 8 and a portion of the blue curves in Figure 7, where the solid curve

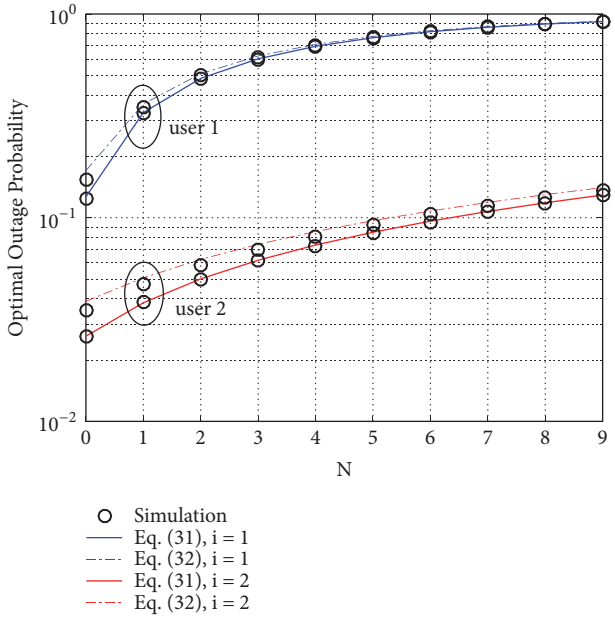



FIGURE 9: Optimal outage probability vs. number of external interferers.

overlaps the dashed curve. However, the system performance can be decreased in case of $T_1 = \tau_2/\alpha_2 - \tau_2\alpha_1$.

Figure 9 presents the optimal outage probability at both users as considering on increasing number of interferers. The solid curves are obtained through (31) representing the optimal outage probability of user $i$ ($i = 1, 2$) whereas the dashed curves are achieved via (32) denoting the outage probability of user $i$ when the relay focuses on optimizing the QoS of the other user, i.e., user $\bar{i}$. In such experiment, $I_{\psi_I}(\kappa)$ can be extracted from (21) as $m_1 = \ldots = m_N = 1$ and $\Omega_1 = \ldots = \Omega_N = 1$ (i.i.d. scenario). It is confirmed that if we try to optimize the outage probability for a user then another user cannot achieve maximum performance. This situation is equivalent to the case of requirement for maximum QoS for the selected user (i.e., user 1) resulting in system performance of the remaining user (i.e., user 2) being dropped.

## 6. Conclusions

In this paper, the application of simultaneously wireless information and power transfer (SWIPT) in nonorthogonal multiple access (NOMA) under the influence of external interferers has been investigated. In particular, we derived the closed-form expression of outage probability in the SWIPT NOMA protocol. In order to provide a complete framework, we illustrate system performance through simulation results used to address the impacts of the target rates of the near and far users and the number of interference sources and turn in evaluating the outage performance of the proposed protocol. In terms of outage probability, new analytical results have been derived to conclude the system efficiency. Such numerical results have been demonstrated to corroborate our analysis. We conclude that, by carefully choosing the parameters of the network regarding external interferers, acceptable system performance can be ensured by applying the SWIPT NOMA protocol in practical networks.

## Appendix

## A. Proof of Theorem 4

The characteristic function of $Y$ is given by

$$\begin{aligned}\Phi_Y(j\omega) &= \mathbb{E}\left[e^{j\omega Y}\right] = \mathbb{E}\left[e^{j\omega\sum_{n=1}^{N}X_n}\right] \\ &= \prod_{n=1}^{N}\int_0^{\infty} e^{j\omega\varphi} f_{X_n}(\varphi)\, d\varphi \\ &= \prod_{n=1}^{N}\frac{1}{(1-j\omega\mu_n)^{m_n}},\end{aligned} \tag{A.1}$$

where $j = \sqrt{-1}$ and $\gamma_n = P_n|h_n|^2$. The third equality is obtained by using the PDF of $\gamma_n$ which is given by

$$f_{\gamma_n}(\varphi_n) = \frac{1}{\Gamma(m_n)\,\mu_n^{m_n}}\varphi_n^{m_n-1}\exp\left(-\frac{\varphi_n}{\mu_n}\right). \tag{A.2}$$

In addition, $\Phi_Y(j\omega)$ can be rewritten in terms of P distinct elements of $\mu_n$ in descending order as

$$\Phi_Y(j\omega) = \prod_{p=1}^{P}\frac{1}{\left(1-j\omega\mu_{\langle p\rangle}\right)^{m_{\langle p\rangle}}}. \tag{A.3}$$

The PDF of $Y$ is obtained through the characteristic function in (A.3) as

$$\begin{aligned}f_Y^{[2]}(\gamma) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega x}\Phi_Y(j\omega)\, d\omega \\ &= \frac{1}{2\pi}\prod_{p=1}^{P}\int_{-\infty}^{\infty}\frac{e^{j\omega\gamma}}{\left(1+j\omega\mu_{\langle p\rangle}\right)^{m_{\langle p\rangle}}}d\omega \\ &\overset{(a)}{=\!=} \frac{1}{2\pi}\sum_{p=1}^{P}\sum_{q=1}^{m_{\langle p\rangle}}R_{p,q}(\mathcal{S})\int_{-\infty}^{\infty}\frac{e^{j\omega\gamma}}{\left(1+j\omega\mu_{\langle p\rangle}\right)^{q}}d\omega \\ &\overset{(b)}{=\!=} \sum_{p=1}^{P}\sum_{q=1}^{m_{\langle p\rangle}}\frac{R_{p,q}(\mathcal{S})}{(q-1)!\mu_{\langle p\rangle}^{q}}\gamma^{q-1}\exp\left(-\frac{\gamma}{\mu_{\langle p\rangle}}\right),\end{aligned} \tag{A.4}$$

where (a) is obtained by using partial fraction decomposition of (A.3) [26] and (b) is achieved with the help of [26, Eq. 3.382.6]. In addition, when $P = 1$, i.e., $\mu_1 = \mu_2 = \ldots \mu_N$, $R_{p,q}(\mathcal{S})$ becomes

$$R_{1,q}(\mathcal{S}) = \begin{cases}0, & q = 1, 2, \ldots, m-1 \\ 1, & q = m.\end{cases} \tag{A.5}$$

The proof leads to the same result as in Theorem 4.

## B.

Define $\delta_p(\upsilon) \triangleq (1+\upsilon\mu_{\langle p\rangle})^{-m_{\langle p\rangle}}$; the n-th derivative of $\delta_p(\upsilon)$ is given by

$$\delta_p^{(n)}(\upsilon) = (-1)^n \frac{\left(m_{\langle p\rangle}+n-1\right)!}{\left(m_{\langle p\rangle}-1\right)!} \frac{\mu_{\langle p\rangle}^n}{\left(1+\upsilon\mu_{\langle p\rangle}\right)^{m_{\langle p\rangle}+n}}. \quad \text{(B.1)}$$

*Proof.* In case of $n = 0$, it is obvious that $\delta_p^{(0)}(\upsilon) = \delta_p(\upsilon)$, which satisfies (B.1). In another case, assuming that $n = k$ holds, it can be proved that $n = k+1$ holds for $k \in \mathbb{N}$. It is noted that the $(k+1)$-*th* derivative of $\delta_p(\upsilon)$ is obtained as

$$\begin{aligned}\delta_p^{(k+1)}(\upsilon) &= \left[\delta_p^{(k)}(\upsilon)\right]^{(1)} \\ &= (-1)^{k+1} \frac{\left(m_{\langle p\rangle}-1+k+1\right)!}{\left(m_{\langle p\rangle}-1\right)!} \frac{\mu_{\langle p\rangle}^{k+1}}{\left(1+\upsilon\mu_{\langle p\rangle}\right)^{m_{\langle p\rangle}+k+1}}.\end{aligned} \quad \text{(B.2)}$$

Therefore, using the induction hypothesis, (B.1) holds for all $k \in \mathbb{N}$.

Define $\mathscr{L}_p(\upsilon) \triangleq \prod_{k=1,k\neq p}^{P} \delta_k(\upsilon)$; then according to Leibniz's rule [26], the *n-th* derivative of $\mathscr{L}(\upsilon)$ is given as

$$\begin{aligned}\mathscr{L}_p^{(n)}(\upsilon) &= \sum_{\{n_k\}_{k=1,k\neq p}^{P}}^{n} \prod_{k=1,k\neq p}^{P} \binom{n_{k-1}}{n_k} \delta_k^{(n_k)}(\upsilon) \\ &= \sum_{n_1=0}^{n} \binom{n}{n_1} \cdots \sum_{\substack{n_k=0 \\ k\neq p}}^{n_{k-1}} \binom{n_{k-1}}{n_k} \\ &\quad \cdots \sum_{n_P=0}^{n_{P-1}} \binom{n_{P-1}}{n_P} \prod_{\substack{k=1 \\ k\neq p}}^{P} \delta_k^{(n_k)}(\upsilon).\end{aligned} \quad \text{(B.3)}$$

Substituting (B.3) into (18) $R_{p,q}(\mathcal{S})$ is rewritten for $P \geq 1$ as

$$R_{p,q}(\mathcal{S}) = \frac{\mathscr{L}_p^{(m_{\langle p\rangle}-q)}\left(-1/\mu_{\langle p\rangle}\right)}{\left(m_{\langle p\rangle}-q\right)!\mu_{\langle p\rangle}^{m_{\langle p\rangle}-q}}. \quad \text{(B.4)}$$

□

## C. Proof of (22)

Substituting (9) and (10) into (22), after some algebraic steps, we can obtain

$$\begin{aligned}\overline{\mathcal{O}}_1 &= \Pr\{\gamma_{1\longrightarrow 2} > \tau_2, \gamma_1 > \tau_1\} = \Pr\left\{\psi_{sr}\right. \\ &> \frac{\tau_2}{\alpha_2 - \tau_2\alpha_1}\left(\psi_I + \frac{1}{\psi_{ru_1}} + \frac{1}{1-\beta}\right), \psi_{sr} \quad \text{(C.1)} \\ &\left. > \frac{\tau_1}{\alpha_1}\left(\psi_I + \frac{1}{\psi_{ru_1}} + \frac{1}{1-\beta}\right)\right\} \\ &= \Pr\left\{\psi_{sr} > \mathrm{T}_1\left(\psi_I + \frac{1}{\psi_{ru_1}} + \frac{1}{1-\beta}\right)\right\}. \quad \text{(C.2)}\end{aligned}$$

The CDF of $\psi_{sr}$ and the PDF of the ordered $\psi_{ru_1}$ [28] are given by

$$F_{\psi_{sr}}(\gamma) = 1 - \exp\left(-\frac{\gamma}{\mu_{SR}}\right), \quad (\gamma \geq 0) \quad \text{(C.3)}$$

$$f_{\psi_{ru_1}}(\gamma) = \sum_{k=1}^{3} \frac{(-1)^{k-1}}{\mu_k} \exp\left(-\frac{\gamma}{\mu_k}\right), \quad (\gamma \geq 0). \quad \text{(C.4)}$$

Substituting (C.2) and (C.3) into (C.1) we obtain

$$\begin{aligned}&\overline{\mathcal{O}}_1 \\ &= \exp\left(-\frac{\mathrm{T}_1}{\mu_{sr}}\frac{1}{1-\beta}\right)\int_0^\infty \exp\left(-\frac{\mathrm{T}_1}{\mu_{sr}}\gamma\right) f_{\psi_I}(\gamma)\, d\gamma \quad \text{(C.5)} \\ &\quad \times \sum_{k=1}^{3} \frac{(-1)^{k-1}}{\mu_k} \int_0^\infty \exp\left(-\frac{\mathrm{T}_1}{\mu_{sr}}\frac{1}{\gamma} - \frac{\gamma}{\mu_k}\right) d\gamma,\end{aligned}$$

where the first integral is obtained by adopting Proposition 2 and the second integral is obtained with the help of [26, Eq. 3.471.9]. Therefore, (C.5) immediately follows (23).

## D. Proof of (31)

First, let us rewrite the PDF of the ordered channels, $|h_i|^2$ $(i = 1, 2)$, as

$$f_{|h_i|^2}(x) = \sum_{k=i}^{4-i} \frac{(-1)^{k+\bar{i}-1}}{\Omega_k} \exp\left(-\frac{x}{\Omega_k}\right). \quad \text{(D.1)}$$

From (C.2) and (26), we can see the similarities of the two equations, and, therefore, it is possible to achieve a unified equation to determine the coverage probability at user $u_i$ which is given by

$$\overline{\mathcal{O}}_i = \Pr\left\{\psi_{sr} > \mathrm{T}_i\left(\psi_I + \frac{1}{\psi_{ru_i}} + \frac{1}{1-\beta}\right)\right\}. \quad \text{(D.2)}$$

Substituting (30) into (D.2), using similar steps in Appendix A, the above equality is rewritten as follows:

$$\overline{\mathcal{O}}_i = 1 - \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_i}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}}\right)\times\sum_{k=i}^{4-i}\frac{(-1)^{k+\bar{i}-1}}{\Omega_k} \cdot\int_0^{\infty}\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}\eta}\frac{2}{\sqrt{y}}-\frac{\mathrm{T}_i}{\mu_{sr}\eta}\frac{1}{y}-\frac{1}{\Omega_k}y\right)dy. \tag{D.3}$$

The above equation cannot be expressed in closed form; however it is noticed that when $x \longrightarrow 0$ the exponential function is approximated as

$$\exp(-x) \simeq 1 - x. \tag{D.4}$$

Applying (D.4) into (D.3), $\overline{\mathcal{O}}_i$ is approximated as

$$\overline{\mathcal{O}}_i \simeq 1 - \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_i}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}}\right)\sum_{k=i}^{4-i}\frac{(-1)^{k+\bar{i}-1}}{\Omega_k} \times\left[\int_0^{\infty}\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}\eta}\frac{1}{y}-\frac{1}{\Omega_k}y\right)dy - 2\frac{\mathrm{T}_i}{\mu_{sr}\eta}\int_0^{\infty}\frac{1}{\sqrt{y}}\exp\left(-\frac{\mathrm{T}_i}{\mu_{sr}\eta}\frac{1}{y}-\frac{1}{\Omega_k}y\right)dy\right]. \tag{D.5}$$

One can utilize [26, Eq. 3.471.9] and [26, Eq. 3.471.15] to solve the first and second integral, respectively. Therefore, after some algebraic manipulations we then achieve (31).

## E. Proof of (32)

From (D.2) and (30) the coverage probability of the other user is given by

$$\overline{\mathcal{O}}_{\bar{i}} = \Pr\left\{\psi_{sr} > \mathrm{T}_{\bar{i}}\left(\psi_I + \frac{1+\sqrt{\eta\left|h_{ru_i}\right|^2}}{\eta\left|h_{ru_{\bar{i}}}\right|^2} + \frac{1+\sqrt{\eta\left|h_{ru_i}\right|^2}}{\sqrt{\eta\left|h_{ru_i}\right|^2}}\right)\right\}. \tag{E.1}$$

Substituting (D.1) and adopting Theorem 6 into (E.1), we then have

$$\overline{\mathcal{O}}_{\bar{i}} = \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\times\int_0^{\infty}\left\{\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\frac{1}{\eta y}\right) \cdot f_{|h_{ru_{\bar{i}}}|^2}(y)\times\int_0^{\infty}\exp\left[-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}\sqrt{\eta}}\left(\frac{\sqrt{x}}{y}+\frac{1}{\sqrt{x}}\right)\right] \cdot f_{|h_{ru_i}|^2}(x)\,dx\right\}dy. \tag{E.2}$$

Adopting (D.1) and (D.4) into (E.2), $\overline{\mathcal{O}}_{\bar{i}}$ is obtained approximately as

$$\overline{\mathcal{O}}_{\bar{i}} \simeq \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\sum_{k_{\bar{i}}=\bar{i}}^{4-\bar{i}}\sum_{k_i=i}^{4-i}\left\{\frac{(-1)^{k_{\bar{i}}+i-1}}{\Omega_{k_{\bar{i}}}}\times\frac{(-1)^{k_i+\bar{i}-1}}{\Omega_{k_i}} \cdot\int_0^{\infty}\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\frac{1}{\eta y}\right)\exp\left(-\frac{y}{\Omega_{k_{\bar{i}}}}\right)\times\left[\Omega_{k_i}-\int_0^{\infty}\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}\sqrt{\eta}}\frac{1}{\sqrt{x}}\exp\left(-\frac{x}{\Omega_{k_i}}\right)dx-\int_0^{\infty}\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}\sqrt{\eta}}\frac{\sqrt{x}}{y}\exp\left(-\frac{x}{\Omega_{k_i}}\right)dx\right]dy\right\}. \tag{E.3}$$

By applying [26, Eq. 3.371] to solve the second and third integral, we then achieve

$$\overline{\mathcal{O}}_{\bar{i}} \simeq \mathrm{I}_{\psi_I}\left(\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right)\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\right) \cdot\sum_{k_{\bar{i}}=\bar{i}}^{4-\bar{i}}\sum_{k_i=i}^{4-i}\left\{\frac{(-1)^{k_{\bar{i}}+i-1}}{\Omega_{k_{\bar{i}}}}\times(-1)^{k_i+\bar{i}-1}\int_0^{\infty}\left[\exp\left(-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}\eta}\frac{1}{y}-\frac{y}{\Omega_{k_{\bar{i}}}}\right)\times\left(1-\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\sqrt{\frac{\pi}{\eta\Omega_{k_i}}}-\frac{1}{2}\frac{\mathrm{T}_{\bar{i}}}{\mu_{sr}}\sqrt{\frac{\pi\Omega_{k_i}}{\eta}}\frac{1}{y}\right)dy\right]\right\}, \tag{E.4}$$

where the integral part in (E.4) can be evaluated in closed form by adopting [[26], Eq. 3.471.9]. Hence, after some manipulations, we then get (32).

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References


[1] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proceedings of the 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC*, pp. 611–615, September 2013.

[2] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, 2014.

[3] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Communications Letters*, vol. 18, no. 2, pp. 313–316, 2014.

[4] Z. Ding, M. Peng, and H. V. Poor, "Cooperative Non-Orthogonal Multiple Access in 5G Systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, 2015.

[5] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "On the Ergodic Capacity of MIMO NOMA Systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, 2015.

[6] J.-B. Kim and I.-H. Lee, "Capacity Analysis of Cooperative Relaying Systems Using Non-Orthogonal Multiple Access," *IEEE Communications Letters*, vol. 19, no. 11, pp. 1949–1952, 2015.

[7] Z. Qin, Y. Liu, Z. Ding, Y. Gao, and M. Elkashlan, "Physical layer security for 5G non-orthogonal multiple access in large-scale networks," in *Proceedings of the ICC 2016 - 2016 IEEE International Conference on Communications*, pp. 1–6, Kuala Lumpur, Malaysia, May 2016.

[8] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the Physical Layer Security of Non-Orthogonal Multiple Access in Large-Scale Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1656–1672, 2017.

[9] B. He, A. Liu, N. Yang, and V. K. N. Lau, "On the Design of Secure Non-Orthogonal Multiple Access Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2196–2206, 2017.

[10] X.-X. Nguyen and D.-T. Do, "Maximum harvested energy policy in full-duplex relaying networks with SWIPT," *International Journal of Communication Systems*, vol. 30, no. 17, 2017.

[11] N. T. Luan and D.-T. Do, "A new look at AF two-way relaying networks: energy harvesting architecture and impact of co-channel interference," *Annals of Telecommunications-Annales des Télécommunications*, vol. 72, no. 11-12, pp. 669–678, 2017.

[12] D.-T. Do and C.-B. Le, "Application of NOMA in Wireless System with Wireless Power Transfer Scheme: Outage and Ergodic Capacity Performance Analysis," *Sensors*, vol. 18, no. 10, p. 3501, 2018.

[13] D.-T. Do, H.-S. Nguyen, M. Voznak, and T.-S. Nguyen, "Wireless powered relaying networks under imperfect channel state information: System performance and optimal policy for instantaneous rate," *Radioengineering*, vol. 26, no. 3, pp. 869–877, 2017.

[14] D.-T. Do, "Power switching protocol for two-way relaying network under hardware impairments," *Radioengineering*, vol. 24, no. 3, pp. 765–771, 2015.

[15] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative Non-orthogonal Multiple Access with Simultaneous Wireless Information and Power Transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 938–953, 2016.

[16] N. T. Do, D. B. Da Costa, T. Q. Duong, and B. An, "A BNBF User Selection Scheme for NOMA-Based Cooperative Relaying Systems with SWIPT," *IEEE Communications Letters*, vol. 21, no. 3, pp. 664–667, 2017.

[17] D.-T. Do and H.-S. Nguyen, "A tractable approach to analyzing the energy-aware two-way relaying networks in the presence of co-channel interference," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, 2016.

[18] S. S. Ikki and S. Aissa, "Performance analysis of two-way amplify-and-forward relaying in the presence of co-channel interferences," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 933–939, 2012.

[19] I. Trigui, S. Affes, and A. Stéphenne, "Ergodic capacity of two-hop multiple antenna AF systems with co-channel interference," *IEEE Wireless Communications Letters*, vol. 4, no. 1, pp. 26–29, 2015.

[20] H. Alves, C. de Lima, P. Nardelli, R. Souza, and M. Latva-aho, "On the Average Spectral Efficiency of Interference-Limited Full-Duplex Networks," in *Proceedings of the 9th International Conference on Cognitive Radio Oriented Wireless Networks*, 554, 550 pages, Oulu, Finland, June 2014.

[21] H. Alves, R. D. Souza, D. B. da Costa, and M. Latva-aho, "Full-Duplex Relaying Systems Subject to Co-Channel Interference and Noise in Nakagami-m Fading," in *Proceedings of the 2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Glasgow, United Kingdom, May 2015.

[22] G. Sharma, P. K. Sharma, and P. Garg, "Performance analysis of full duplex relaying in multicell environment," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2501–2505, Delhi, India, September 2014.

[23] D. Benevides da Costa, . Haiyang Ding, and . Jianhua Ge, "Interference-Limited Relaying Transmissions in Dual-Hop Cooperative Networks over Nakagami-m Fading," *IEEE Communications Letters*, vol. 15, no. 5, pp. 503–505, 2011.

[24] D. B. da Costa and M. D. Yacoub, "Outage performance of two hop AF relaying systems with co-channel interferers over Nakagami-m fading," *IEEE Communications Letters*, vol. 15, no. 9, pp. 980–982, 2011.

[25] D. B. Da Costa, H. Ding, M. D. Yacoub, and J. Ge, "Two-way relaying in interference-limited AF cooperative networks over nakagami-m fading," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3766–3771, 2012.

[26] A. Jeffrey and D. Zwillinger, *Table of integrals, series, and products*, Academic press, 2007.

[27] H. A. David and H. N. Nagaraja, *Order Statistics*, John Wiley, New York, NY, USA, 3rd edition, 2003.

[28] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "The Impact of Power Allocation on Cooperative Non-orthogonal Multiple Access Networks with SWIPT," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4332–4343, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2753025, 15 pages
https://doi.org/10.1155/2018/2753025

*Research Article*

# MC-GiV2V: Multichannel Allocation in mmWave-Based Vehicular Ad Hoc Networks

**Wooseong Kim**

*Department of Computer Engineering, Gachon University, 1342 Seongnam-si, Gyeonggi, Republic of Korea*

Correspondence should be addressed to Wooseong Kim; wooseong@gachon.ac.kr

Received 26 February 2018; Revised 2 June 2018; Accepted 10 June 2018; Published 17 July 2018

Academic Editor: Ingrid Moerman



During last several years, mobile communications using mmWave spectrum have been intensively researched for 5G wireless networks. Now the mmWave wireless technologies are evolved into direct device-to-device communications for a single or multihop communication via Giga-bit links. Vehicular ad hoc networks (VANETs) are one of the most attractive areas to apply the direct mmWave communications. In this paper, we propose a Giga-V2V (GiV2V) network, in which vehicles query and deliver high quality video and sensor data of smart and self-driving cars using mmWave communications instead of current dedicated short-range communications (DSRC). In the GiV2V networks, vehicles probably form a grid topology along lanes of a road, which leads to align mmWave beams of the vehicles and cause mutual interference among them. As channel diversity can resolve effectively the interference between mmWave beams, we propose several heuristic algorithms for channel assignment of each beam in the GiV2V networks. We investigate the proposed algorithms using simulation and compare performance with well-known metaheuristic algorithms for this NP-Hard problem.

## 1. Introduction

5G wireless technology opens a new era of Giga-bit rate data communications using mmWave spectrums for high quality and real-time multimedia data. Many companies and universities built testbeds for measurement study of the mmWave communications and made efforts to demonstrate feasibility of beam forming and tracking developed for mobile communications. The 5G mobile communications are now being standardized in ITU [1], 5GPPP [2], 3GPP [3], and so forth and ready to commercialize. The mmWave communications are also developed for local area communications (e.g., WPAN and WLAN) using 60 GHz unlicensed bands such as IEEE 802.15.3 Task Group 3c (TG3c) [4] and IEEE 802.11ad [5].

Due to severe penetration loss and reflection from short wavelength, mmWave communications are almost feasible only in Line-of-Sight (LoS) environment. When a mmWave link between a sender and receiver is blocked (i.e., non-LoS), relay operation is necessary; in 802.11ad WLAN, the mobile station can access to an access point (AP) via a relay station. Such Device-to-Device (D2D) direct communications using the mmWave spectrum attract attentions to support Giga-bit data rate in proximity services and offloading in cellular networks.

For dissemination of safety messages over roads, V2X communications (e.g., vehicle-to-vehicle, infrastructure, or pedestrian) have been researched and developed intensively during the last decade. At the end, auto companies recently release solutions based on the IEEE 802.11p/WAVE standard which satisfy requirements of safety messages (e.g., low-latency delivery less than 100 *ms*) and support infotainment communications up to 6-27 Mbps using separate service channels. However, a future smart car capable of autonomous driving demands much higher data rate and low latency for vehicle control technology, which relies on large amount of data from near or medium range radars and camera sensors of neighboring vehicles. In particular, higher resolution visual data like Ultra High Definition (UHD) video can enable precise vehicle control; for example, if using 2 M pixel camera instead of 0.3 M in the lane keeping system, curvature recognition accuracy on the front road increases from 30 to 50 *m*, which leads to more safe and fuel efficient driving.

In this paper, Giga-bit vehicle-to-vehicle communication (GiGaV2V or GiV2V) using the mmWave is proposed to support aforementioned high quality multimedia data. Research on the GiV2V has not been conducted popularly and not matured yet to the best of our knowledge. The GiV2V can improve network throughput because of spatial frequency reuse by directional antennas that are typically used to compensate high path loss of the mmWave. However, the spatial division may not occur constructively since vehicles are mostly aligned along lanes of roads and form a grid topology where mmWave beams are also aligned and cause mutual interference. Directivity of the directional antenna increases not only the antenna gain and signal to noise ratio (SNR), but also interferences to other nodes. To mitigate the interference, we propose multichannel- (MC-) GiV2V, a channel diversity scheme in GiV2V networks. Here we introduce several multichannel allocation algorithms with many available channels in the mmWave spectrum; for example, IEEE 802.11ad has 6 channels of each 2 GHz bandwidth.

Our proposed algorithms are distributed, centralized greedy and hybrid algorithms. The distributed algorithm searches a local optimal allocation within an interference region and the greedy algorithm assigns channels based on global information (i.e., SINR of all receive nodes). The hybrid approach is a mixed algorithm of the above two algorithms. Details of algorithms are explained in Section 6. According to simulation results in Section 7, the hybrid approach shows best throughput among them since it probably searches a globally optimal allocation with well-distributed initial conditions. Furthermore, three well-known metaheuristic algorithms are investigated for comparison study with our proposed algorithms.

## 2. Related Works

Directional antenna was exhaustively exploited for a MAC protocol in multihop ad hoc networks. Most of those researches assume 2.4 or 5 GHz Wi-Fi, but similar challenges also exist in mmWave-based WLANs. Ko et al. [6] first propose a modified 802.11 Distributed Coordination Function (DCF) for directional antennas, which maintains directional channel availability based on the GPS information. Takai et al. [7] use Angle of Arrivals (AOAs) of Request to Send (RTS) and Clear To Send (CTS) instead of GPS information. Choudhury et al. [8] propose a basic directional MAC (DMAC) which includes Directional Network Allocation Vector (DNAV) and listens incoming packets omnidirectionally to trace their Direction of Arrival (DoA). Kolar et al. [9] introduce a greed queuing to solve a Head of Line (HoL) problem with the DNAV table of beam directions. Ramanathan et al. [10] suggest different backoff algorithms for different events such as busy channel and missing CTS or ACK and also tight power control scheme. In order to solve a hidden terminal problem in the DMAC, Circular Directional RTS (CDR) [11], CRCM [12], and DtD-MAC [13] conduct sequential RTS and CTS transmissions to all directions, which deals with deafness and directional hidden terminals from unheard nodes or asymmetric antenna gain. But the circular transmissions suffer from control overhead and excessive delay according to number of sectors. Gossain et al. [14] propose simultaneous circular RTS/CTS to reduce the delay with Diametrically Opposite Direction (DOD) which removes duplicate transmissions of RTSs and CTSs in the overlapped area. Furthermore, Deafness Avoidance and Collision Avoidance (DMAC-DACA) [15] and DMAC with Deafness Avoidance (DMAC/DA) [16] reduce circular transmission overhead by DNAV reservation and beam direction information, with which nodes can determine spatial diversity and schedule pending transmitters. Singh et al. [17] propose a Memory-guided DMAC (MDMAC), as a fully distributed MAC protocol, which enables approximate TDM scheduling for wireless meshes using the memory about transmission success or fail.

Recently, mmWave communications emerge as one of the key 5G technologies. Its feasibility has been explored by many universities and companies. Rappaport et al. [18] perform measurement campaign in New York City on the 28 GHz, 38 GHz, and 73 GHz bands [19, 20] and establish a channel model of the mmWave communications. The mmWave links are considered not only for access links of mobile devices, but also for backhaul links that can constitute wireless mesh networks [21, 22]. 3GPP [3] completes standards of a new radio (i.e., mmWave communications) and now moves into device-to-device communications for the mmWave which can be applied to legacy proximity services in cellular networks.

The mmWave for WLANs has already been explored for home and mobile appliances at indoor environment. Also, several standards using 60 GHz unlicensed bands were released such as IEEE 802.15.3 Task Group 3c (TG3c) [4] and IEEE 802.11ad [5], which specify physical and MAC protocols (Carrier Sensing Multiple Access/Collision Avoidance (CSMA/CA), Time Division Multiple Access (TDMA), etc.). Also, those WPAN/WLAN standards define relay operation that can be utilized for multihop ad hoc networks. For instance, an access point (AP) arranges service periods (SPs) for Directional Multi-Giga-bit (DMG) mobile stations (STAs), when the AP receives a request of Relay DMG STA (RDS) search from a STA for NLOS environment. During the SPs, a source and destination STA exchange the packets with candidate RDSs nearby. Then, the source STA asks several RDSs with good channel (i.e., high SNR) to report channel condition to both source and destination STAs. Finally, the source STA selects a best RDS that has highest SNR in both links.

In [23], demanded rate-based coordination of the directional or omnidirectional transmissions is proposed with allocation of time slots for spatial and time reuse of frequency in mmWave WLANs. Sing et al. [24] propose a multihop MAC protocol for indoor mmWave environment where diffraction and blockage highly occur due to fixed or moving obstacles (e.g., people and furniture). They develop a diffraction model to estimate link connectivity and decide multihop relays. From simulation, it is proved that proposed approach improves network throughput with low overhead rather than an AP-based single hop communication. Reference [25] describes tactical scenarios using mmWave links for a secure channel in military ad hoc networks and relay operation in NLOS environment. In [26], a CSMA/CA-like MAC protocol

for directional mmWave is proposed. Chen et al. [27] propose a spatial reuse strategy with directional antennas in IEEE 802.11 ad networks. Son et al. [28] propose a Frame based Directive MAC protocol (FDMAC) which is a centralized scheduling algorithm for the Pico-Net Controller (PNC) based on greedy coloring providing multiple concurrent transmissions. Thornburg et al. [29] analyze throughput of ad hoc networks using mmWave communications. Authors establish a 2D-PPP model of nodes and obstacles deployment and evaluate performance of one- or two-way communications in terms of SINR and coverage with simulation. In [30], Park et al. propose a Multiband Directional Neighbor Discovery (MDND) for self-organization of ad hoc networks, which utilizes dual radios with different bands and antenna types, a 2.4 GHz band with an omnidirectional antenna and 60 GHz band with a directional antenna. Reference [31] proposes a stochastic model of vehicular communications at highway for mmWave communications, where mmWave-based road side units are deployed for infrastructure to vehicle communications with high data rate rather than vehicle-to-vehicle communications. Blockage probability according to vehicle density and speed is shown from the model. In [32], authors show design and implementation of a long-range and broadband aerial communication system with directional antennas (ACDA), which enables unmanned aerial vehicle (UAV) to extend communication range, increase throughput, and reduce interference. In the testbed, the ACDA achieves 48 Mbps throughput at a distance of 300 m and 2 Mbps at 5000 m, promising long-distance Wi-Fi aerial communication. Reference [33] proposes joint optimization to select relay and link to get around obstacle and reduce delivery latency in 60 GHz mmWave networks and develops a less complex algorithm by decomposing the problem into subproblems. In [34], research results on propagation characteristics for V2V channels, particularly shadowing effects induced by obstructing vehicles between transmitter (Tx) and receiver (Rx), are introduced. In [35], measurement campaign is conducted in the mmWave band for the 12 most common railway materials; influence of typical objects to the mmWave propagation channel is analyzed for railway scenarios with various configurations. Reference [36] proposes an IEEE 802.11ad-based radar for long-range radar (LRR) applications at the 60 GHz unlicensed band, which enables a joint waveform for automotive radar and mmWave vehicle-to-vehicle communications reusing hardware.

## 3. GiV2V Network Architecture

Figure 1 depicts an architecture of GiV2V networks, in which vehicles deployed at an intersection form mmWave beams toward neighbor vehicles to exchange safety and infotainment data. Also, they can share own storage or processing power to maintain floating data and process those data. For instance, video clips captured in the intersection area are held and analyzed by vehicles or road side units (RSUs) for object or event detection [37]. In order to query and deliver the floating data in this vehicular cloud, the Information Centric Networking (ICN) mechanism can be used [37–39].

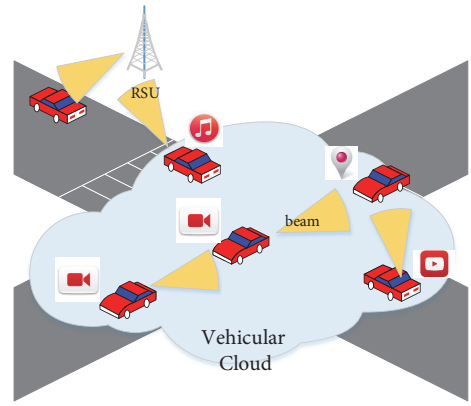



Figure 1: GiV2V network architecture.

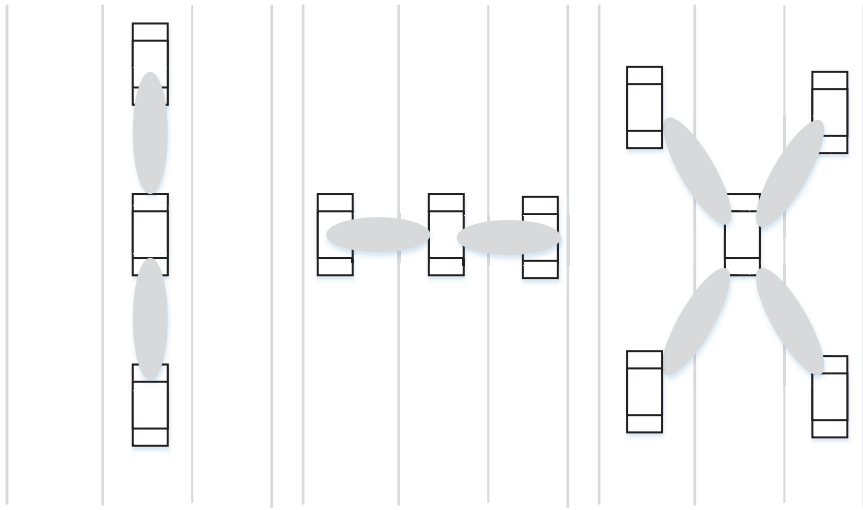
Due to road structure, neighbor vehicles are located on limited positions, which are mostly front, back, and side directions as shown in Figure 2. First, a simple convoy model of Figure 2(a) is a typical traffic pattern at roads and appropriate to create a vehicle flow (i.e., vehicle train) for autonomous driving of smart cars. In this model, transmission direction is also limited, forward or backward, which can cause considerable interference among vehicles without transmission power control. However, de- and acceleration of vehicle speed lead to varying distance between vehicles, so the power control probably makes vehicle connectivity unstable. Second, a vehicle searches vehicles in next lanes with side beams to couple partitioned networks along the lanes as shown in Figure 2(b). This scenario can cause more interference than the convoy model due to small lane width. The beam directions are more various according to road shapes (e.g., curve and intersection), road width (e.g., multilane highway), and vehicle speed. Accordingly, vehicles can be located on front side or rear side in next lanes as shown in Figure 2(c). Such diagonal beams diverse beam directions like a random topology that has lower interference than a grid topology of Figures 2(a) and 2(b). However, the grid topology has advantages in connection establishment with small efforts to sweep beam directions compared to the random topology. As a consequence, most of the beams in GiV2V communications belong to scenarios in Figure 2, and considerable interference can exist due to limited beam directions.

## 4. Directional Antenna

*4.1. Directivity Model.* In this paper, a beamforming model is expressed by a sectorized directional antenna following ITU-R reference [40] that covers 400 to 70 GHz spectrum as below. The radiation intensity at azimuthal $F(\phi)$ and elevation plane $F(\theta)$ is modelled by two different radiation intensity functions: rectangular and exponential sectoral radiation.

Directivity of omnidirectional and sectoral antennas is

$$D = \frac{U_M}{U_0}, \tag{1}$$



(a) Front-rear beam (b) Side beam (c) Front-rear side beam

Figure 2: GiV2V topology and beam pattern at road.

where $D$ is directivity (i.e., gain) and $U_0$ is radiation intensity of an isotropic source.

$$P_t = \int_0^{2\pi} \int_0^{\pi} F(\theta) F(\phi) \sin(\theta) \, d\theta d\phi, \tag{2}$$

then the omnidirectional power is $U_0 = 1/4\pi P_t$.

Here those two different radiation intensity functions in the azimuthal plane can be considered while the elevation plane is assumed to be an exponential function. In rectangular sectoral radiation, the azimuthal power intensity is derived as

$$F(\phi) = \begin{cases} 0 & \text{if } \dfrac{\phi_s}{2} - |\phi| \geq 0 \\ 1 & \text{if } \dfrac{\phi_s}{2} - |\phi| < 0 \end{cases} \tag{3}$$

And elevation power is

$$F(\theta) = e^{-a\theta^2}, \quad a = -\ln(0.5) \frac{4}{\theta_{bw}}, \tag{4}$$

where $\theta_{bw}$ is beamwidth.

Accordingly, the omnidirectional intensity is calculated approximately as

$$U_0 = \frac{\phi_s \theta_{bw}}{4\pi} \sqrt{\frac{\pi}{2.773}} e^{-\theta^2/11.09} \tag{5}$$

The directivity $D_r$ of the rectangular radiation model is

$$D_r = \frac{38750}{\phi_s \theta_{bw}} e^{\theta_{bw}^2/36400}, \tag{6}$$

when $U_M$ is 1.

In the exponential function for sectoral radiation, the azimuthal function $F(\phi)$ is replaced by the following exponential function:

$$F(\phi) = e^{b\phi^2}, \quad b = \ln(0.5) \frac{4}{\phi_s^2} \tag{7}$$

Here the gain $D_e$ of the exponential radiation model is

$$D_e = \frac{36400}{\phi_s \theta_{bw}} e^{\theta_{bw}^2/36400} \tag{8}$$

Side lobes are smaller than the main lobes with the front-to-back ratio (FBR) (i.e., ratio of front-side lobes) denoted by $\gamma$ ($0 < \gamma \leq 1$, 1 for omnidirection). Accordingly, the gains of a main lobe and side lobes are $G_m = (1 - \gamma) D(\theta_{bw})$ and $\gamma G_m$, respectively. Table 1 shows gains of main and side lobes with varying $\gamma$ values.

For our experiment that appeared in Section 7, 30, 60, and 90 degrees of beam width are used; the directivity antenna gains are 16.8, 8.4, and 5.6 dBi, respectively. While the widths of those beams are in a linear scale, gains increase exponentially as shown in Figure 3.

*4.2. Coverage in GiV2V Networks.* Figure 4 shows an example of communication range in GiV2V networks. The coverage is varying according to beam directions of neighbor vehicles in contrast to a coverage using omnidirectional antenna. In the figure, 10 vehicles exit near the transmitter $V_t$ but only 4 vehicles, from $V_{r1}$ to $V_{r4}$, have connection to the transmitter. Supposing that the transmitter vehicle, $V_t$, forms a beam shadowed among 4 sectors (i.e., 90 degrees), the receiver vehicle $V_{r1}$ located on the transmission sector is reachable even with its different beam direction in the $d_2$ coverage. However, other $V_{r2}$ and $V_{r3}$ that are not on the first quarter sector but within the $d_2$ must create beams toward the transmitter $V_t$ for connections. In the $d_3$ coverage, only the $V_{r4}$ has a connection to the $V_t$ since both the transmitter and receiver have to beam to each other.

Table 2 describes beamforming gain and corresponding reachable radio ranges. When antenna gain from bore sight of a main lobe is denoted as $G$ and gain from other directions is $g$ as a side lobe, vehicles in $d_3$ should have beamforming to each other and in $d_2$, one of a sender and receiver has to make a beam to a peer node at least.

Table 1: Gain of main and side lobes.

| | Directivity (Gain) | $\gamma$=0.2 | $\gamma$=0.5 | $\gamma$=0.7 |
|---|---|---|---|---|
| 90 | 5.613824327 | 1.122765 | 2.806912 | 3.929677 |
| 60 | 8.42073649 | 1.684147 | 4.210368 | 5.894516 |
| 45 | 11.22764865 | 2.24553 | 5.613824 | 7.859354 |
| 30 | 16.84147298 | 3.368295 | 8.420736 | 11.78903 |
| 15 | 33.68294596 | 6.736589 | 16.84147 | 23.57806 |
| 10 | 50.52441894 | 10.10488 | 25.26221 | 35.36709 |

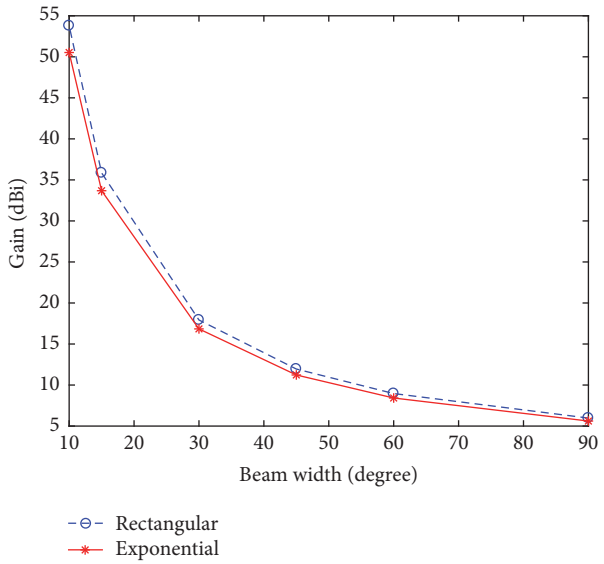



Figure 3: Directivity (gain) of rectangular and exponential radiation models.

## 5. mmWave Channel Propagation Model

The mmWave pathloss model at 60 GHz was established for LoS environment based on measurement study [41].

$$L_d\,(dB) = A + 20\log_{10}\left(f_{MHz}\right) + 10\alpha\log_{10}\left(d\right), \quad (9)$$

where the $A$ is 32.5 dB and no shadow factor. $d$ is a distance between a transmitter and receiver ($km$) and $\alpha$ is a pathloss exponent of LoS (e.g., 2).

In outdoor GiV2V communication, additional attenuation from vapour water ($L_{vap}$), oxygen ($L_{O_2}$), and rain ($L_R$) is considered as below. Total pathloss can be $PL(d) = L_{(f,d)} + L_a$.

$$L_a\,(dB) = d\left(L_{vap} + L_{O_2} + L_R\right). \quad (10)$$

Those atmosphere parameters ($dB/km$) for further loss are assumed constant for relatively short communication period in this study [42, 43]. From simplicity of the constant $L_a$ during the short communication period, the path loss is only determined by the distance, $d$, at a given operational frequency, $f$ (e.g., 60 GHz).

$$P_r = P_t - PL\,(d) - NF + G_R + G_T - IL - CB, \quad (11)$$

where $P_r$ and $P_t$ are transmission and receive power and NF and $N_{th}$ are noise floor and thermal noise. Maximum antenna gain, $G_T$, $G_R$ of a transmitter $T$ and receiver $R$ is assumed to be the same (i.e., same antenna array). $IL$ is implementation loss like from cables $CB$.

Table 2: Beam direction and radio range.

| Gain | Range |
|---|---|
| $G_t$ $G_r$ | d3 |
| $G_t$ $g_r$ | d2 |
| $g_t$ $G_r$ | d2 |
| $g_t$ $g_r$ | d1 |

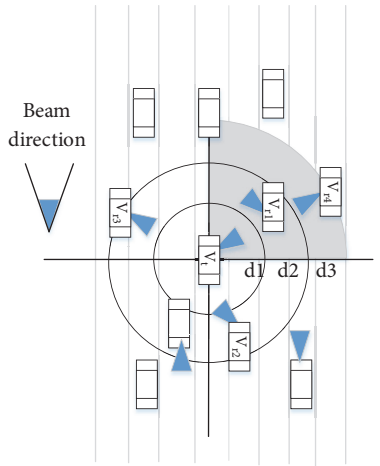


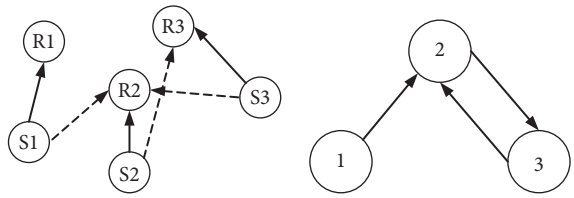

Figure 4: Varying coverages in GiV2V networks.

In the LoS environment, the radio range can be derived by the following outage probability with a required SNR of a target Modulation Coding Scheme (MCS).

$$P\left(P_r \geq T\right) = P\left(PL\left(d\right) \leq P_t + G_R + G_T - T - G_n\right), \quad (12)$$

where $PL(d)$ is pathloss of distance $d$, $T$ is sensitivity for the required MCS level, and $G_n = NF + IL + CB$. For instance, minimum $T$ is -78 or -68 dBm for control signals and data with lowest MCS, $\pi/2$-BPSK, respectively.

In the above equation, maximum coverage $d$ can be calculated by $PL^{-1}(P_t + G_R + G_T - T - G_n)$. Accordingly, the effective range is decided by only antenna gain of transmission and reception (i.e., beamforming factor) as shown in Table 2 while other parameters are assumed to be constant; in this study,




(a) Network topology (b) Interference relation

Figure 5: GiV2V topology and interference.

no transmission power control is achieved between vehicles. From (12), maximum range, $d$, can be expressed as follows.

$$d = \left( \frac{P_t G_R G_T}{T G_n} \right)^{1/\alpha} \tag{13}$$

Consequently, the coverage is exponentially increasing with beamforming gain according to the (13).

## 6. Multichannel Beamforming for GiV2V Networks

The GiV2V enables multiple vehicles to create mmWave links to near vehicles with directional beams for concurrent Giga-bit communications. Those beams can be aligned or diversified according to network topologies and interference also occurs by the beam patterns.

Figure 5(a) illustrates an example of interfering 3 pairs of communication vehicles with directional beams, where each vertex indicates a vehicle node, sender *Sn*, and receiver *Rn*, and receive interference is different to each pair because of adjacent beam directions; *R*2 receives interference from both transmitters *S*1 and *S*3 denoted as dashed lines, while *R*3 has interference only from *S*2 and there is no interference for *R*1. The interference among nodes can be expressed by a directed graph as in Figure 5(b) in which each vertex indicates a pair of communication nodes, i.e., a link or beam.

In the interference graph of the GiV2V networks, multiple channels can be assigned to the vertices for collision and interference avoidance. For example, only 2 channels can remove the mutual interference completely in Figure 5(b): CH1 for vertexes 1 and 3 and CH2 for vertex 2. This channel assignment for each communication pair is a coloring problem of the interference graph, which is a combinatorial problem known as a NP-Hard. Accordingly, we propose several algorithms that can be realized in centralized or distributed manners and compare their performance through simulation.

*6.1. System Model.* We build a system model to design and analyze our algorithms. We define variables according to the directed interference graph in Figure 5(b). Each communication link is denoted as a vertex and directional interference as an edge; there are $i$ vertices and $ij$ directional edges. Symbols of the our system model are described in Table 3.

Table 3: System model parameters.

| Symbol | Description |
|---|---|
| N | A set of communication links |
| C | A set of available channels |
| G | Directional antenna gain |
| P | Power of transmission |
| B | Bandwidth |
| $r_i$ | Data rate at link $i$ |
| L | Path loss |
| $d$ | Distance between vehicles |

The channel assignment for each link $i \in N$ can be expressed by $x_i^c$. If the link $i$ is tuned to the channel $c$,

$$x_i^c = \begin{cases} 1, & i \text{ on } c \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Data rate of a communication link, $i$, on a channel $c$, $r_i^c$, can be calculated approximately by Shannon and Friis transmission equation as follows:

$$r_i^c = B_c \log_2 \left( 1 + \frac{P_i G_i \Lambda L_i x_i^c}{\sigma^2 + \sum_{j:j \in H_i} P_k G_j \Lambda L_j x_j^c} \right), \tag{15}$$

where $\Lambda = (\lambda/4\pi)^2$ and $L_i = d_i^\alpha$. The $\alpha$ is an exponent for free-space pathloss and $d$ is distance of a communication link or an edge of Figure 5(b). Other parameters are denoted in Table 3. Afterwards the distance term is only used for the interference edge rather than the other. $H_i$ is a hyperarc that consists of set of incoming edges at the vertex $i$ in the interference graph. For an example of Figure 5(b), $H_1 = \{\}$, $H_2 = \{1, 3\}$, and $H_3 = \{2\}$. System bandwidth $B_c$ can be 2.1 GHz for each channel according to IEEE 802.11ad.

In this study, our objective is maximizing sum of utility of each communication link.

$$\max \sum_i U\left(r_i^c\right) \tag{16}$$

The utility function $U$ is defined as follows.

$$U\left(r_i^c\right) = \gamma + \kappa \sum_{i \in N} \sum_{c \in C} x_i^c r_i^c, \tag{17}$$

where $\gamma$ is minimum data rate for a pair of communication nodes on a channel $c$ and $\kappa$ is a small value like 1e-3 for max-min fairness among links [43].

*6.2. Random Channel Assignment.* Each link is the same as a vertex. To clarify, each link (a vertex in the interference graph) chooses a channel randomly in a distributed manner, which is mostly simple and powerful compared to complicated channel assignment algorithms. In addition, vehicle nodes can use position information (i.e., GPS) for random seeds to diversify channel selection within an interference region.

*6.3. Distributed Channel Assignment.* As the same distributed approach, neighboring nodes can exchange channel selection information to avoid collisions rather than the random selection, which enables nodes to select a minimum used channel within interference region. In this section, we introduce a simple Distributed Channel Assignment (DCA) algorithm.

The achievable rate of each link is varying by the channel assignment in the system model, which is intractable. Accordingly, the objective is redefined as a local general assignment problem to minimize maximum aggregated gain within interference region from (15). For the simplicity, fixed transmission power and constant parameters are omitted. This local solution from the redefined problem does not guarantee to find a global optimum, but it is valuable for notable throughput and easily realized in the distributed architecture.

$$\text{minimize} \quad \max_{c} \sum_{i \in H} w_i x_i^c \tag{18}$$

$$\text{subject to} \quad w_i = G_i d_i^{-\alpha} \tag{19}$$

$$\sum_{c \in C} x_i^c = 1, \quad i \in N \tag{20}$$

$$x_i^c \in \{0, 1\}, \quad i \in N, \ c \in C, \tag{21}$$

where the interference weight $w_i$ is decided by distance to an interferer and beam direction. The hyperarch $H$ is one of interference regions in a whole network. Equation (18) can be reformulated into an equivalent epigraph form and solved by Lagrangian relaxation as follows.

$$\text{minimize} \quad t \tag{22}$$

$$\text{subject to} \quad \sum_{i \in H} w_i x_i^c \le t \tag{23}$$

$$\text{Eq. (20)} - (21) \tag{24}$$

Partial Lagrangian can be derived for relaxation by dualizing first constraint (23),

$$\begin{aligned} L(t, x, \boldsymbol{\lambda}) &= t + \sum_{c \in C} \lambda_c \left( \sum_{i \in H} w_i x_i^c - t \right) \\ &= t \left( 1 - \sum_{c \in C} \lambda_c \right) + \sum_{i \in H} \sum_{c \in C} \lambda_c w_i x_i^c, \end{aligned} \tag{25}$$

where $\lambda$ is nonnegative Lagrange multiplier for the first inequality constraint (23). Accordingly, we can have a dual function $g(\lambda) = L(t, x, \boldsymbol{\lambda})$ by minimizing above partial Lagrangian with regard of $x$ and $t$, where the dual function can have $-\infty$ by $t$ if $\sum_{c \in C} \lambda_c \neq 1$. Therefore, we can redefine the dual function with constraint for $\lambda$.

$$\text{maximize} \quad g(\boldsymbol{\lambda}) \tag{26}$$

$$\text{subject to} \quad \sum_{c \in C} \lambda_c = 1 \tag{27}$$

$$\lambda_c \ge 0 \tag{28}$$

```
1: procedure VehicleChannelSelectFunc
2:     Initialization: λ_c(t) for all c, and estimate w_i
3:     loop:
4:     Select c ⟵ arg min_c λ_c w_i according to Eq.(29)
5:     Broadcast the selected channel c, x_i^c
6:     Update λ_c(t + 1) according to Eq.(30).
7:     if λ_c(t) = λ_c(t + 1) then
8:         stop;
9:     Broadcast λ_c(t)
10:    goto loop.
```

Algorithm 1: Distributed channel selection algorithm.

In this dual problem, the analytical solution for the combinatorial problem can be derived in a closed form expression. For optimal $x^*$,

$$x_i^{c*} = \begin{cases} 1, & c = c* \\ 0, & \text{otherwise}, \end{cases} \tag{29}$$

where the optimal channel ($c^* = \arg\min_c \lambda_c w_i$ ) from (25) can be easily found, which is one minimizing sum of weight within the interference range by choosing minimum prices, $\boldsymbol{\lambda}$. Accordingly, the derivation of the $\mathbf{x}$ only takes $O(N)$ linear time.

The dual function is convex although it is not differentiable. Therefore optimal $\lambda$ value in (26) can be acquired using a subgradient method. A following projected subgradient method updates the $\lambda$ value by given channel allocation, $\mathbf{x}$, with which the algorithm newly assigns channels to nodes to converge into an optimum.

$$\lambda_c(t + 1) = \lambda_c(t) + \alpha_t \sum_{i \in H} w_i x_i^{c*}, \tag{30}$$

where $t$ indicates an algorithm iteration and $\alpha_t$ is a $t$th step size ($\alpha_t > 0$).

Above the closed form solution can be realized in distributed manner as shown in Algorithm 1.

*6.4. Interference-Aware Channel Assignment.* In contrast to the local solution introduced in Algorithm 1, we propose a greedy SINR-based Channel Assignment (SCA) algorithm in a centralized architecture, where vehicles report chosen channels to a controller and then the controller calculates optimal allocation for the objective. As shown in (17) and (18), our objective is minimizing maximum aggregated interference. Thus, the controller chooses a link with maximum aggregated interference and assigns a separate channel first as a greedy manner. Algorithm 2 explains a procedure of the algorithm to assign a channel to each link based on degree of interference. First, the algorithm selects a pair of vehicles that suffer from highest interference and assigns a channel that brings maximum throughput enhancement in overall networks. The algorithm continues until no more link can have gain after changing own channel.

1: **procedure** ControllerGreedyScheduleFunc
2: Initialization: $c = 1$ and $x_i^1 = 1$
3: Receive $\sum_{j \in H_i} w_j$ from $i$
4: MAX = $\sum_i \sum_{j \in H_i} w_j$
5: *loop*:
6: $i \longleftarrow \arg\max_i \sum_{j \in H_i} w_j$
7: $c^* \longleftarrow \arg\min_c \sum_{j \in H_i} w_j x_j^c$
8: **if** MAX > $\sum_i \sum_{j \in H_i} w_i$) **then**
9: Update $c = c^*$, $x_e^c = 0$ and $x_e^{c^*} = 1$
10: MAX = $\sum_i \sum_{j \in H_i} w_j$)
11: **goto** *loop*.
12: **else**
13: **stop**;

Algorithm 2: Greedy channel selection algorithm.

*6.5. Hybrid Channel Assignment.* Previously proposed SCA algorithm assumes no initial channel distribution; all vehicles have the same channel (e.g., channel 1) at the beginning. However, it is not realistic and able to lead the algorithm to reach local maxima. Therefore, we propose an initialized SCA (InitSCA) algorithm as a hybrid approach that utilizes the DCA and SCA in sequence; the DCA allows vehicles to find a local optimum and then a controller of the SCA adjusts channel allocation based on SINR of each pair of communication vehicles to enhance total network throughput.

## 7. Experiments

In this section, we evaluate channel allocation algorithms introduced in Section 6 using simulations. In GiV2V networks, vehicles communicate with one-hop neighbors regardless of a final destination at a certain time. Previous studies about a directional MAC show that a HoL problem can be solved by a neighbor location table and scheduling order [9, 44]. From this, we assume that a vehicle has a communication peer nearby and establish a beam toward the peer node.

We consider an exemplary scenario at highway that has a simple vehicle traffic pattern with varying vehicle densities and uses wireless local area (WLAN) networking for V2V communications. This GiV2V network topology at the highway follows 2D-PPP that is shown similar with a PPP model along multiple lanes in terms of average throughput [45]. For simulation parameters, the highway space is 20 x 400 $m$ and vehicle density is varying from 3 to 11e-3. Transmission power of each node is configured as 20 dBm in the WLAN and the channel model is applied as shown in Section 5.

Average distance $\mathbb{E}(D)$ between vehicles can be derived by the PPP model as follows. 2D-Poisson distribution with density $\lambda$ is

$$\mathbb{P}\left(\Phi\left(A\right) = n\right) = e^{-\lambda|A|} \frac{\left(\lambda\left|A\right|\right)^n}{n!} \tag{31}$$

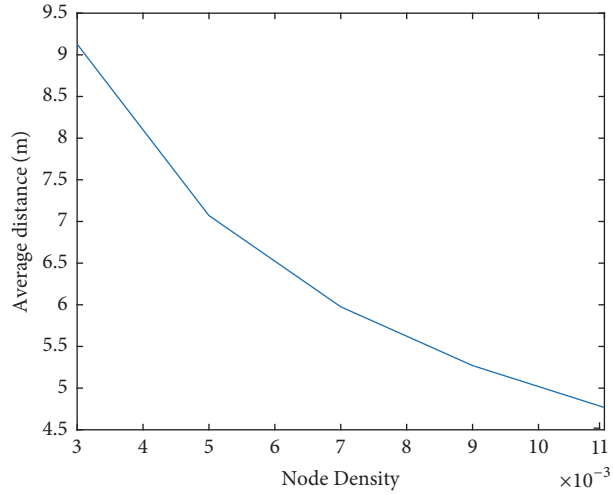



Figure 6: Average distance by vehicle density.

Then the CCDF of the distance of the nearest point of the process is the same as probability of empty set in circle area of a radius, $r$, $\mathbb{P}(\Phi(\pi r^2) = 0)$.

$$\mathbb{P}\left(D > r\right) = e^{-\lambda \pi r^2} \tag{32}$$

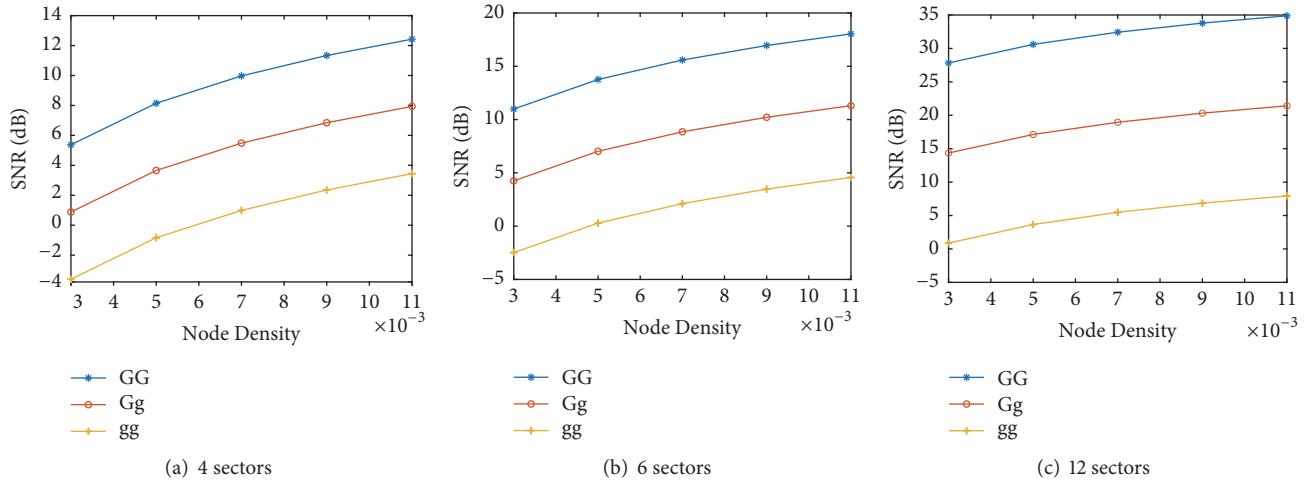
Hence average distance can be calculated as

$$\mathbb{E}\left(D\right) = \int_0^\infty f\left(r\right) r dr = \frac{1}{2\sqrt{\lambda}}, \tag{33}$$

where the PDF $f(r)$ of the nearest point is given from the CCDF as below.

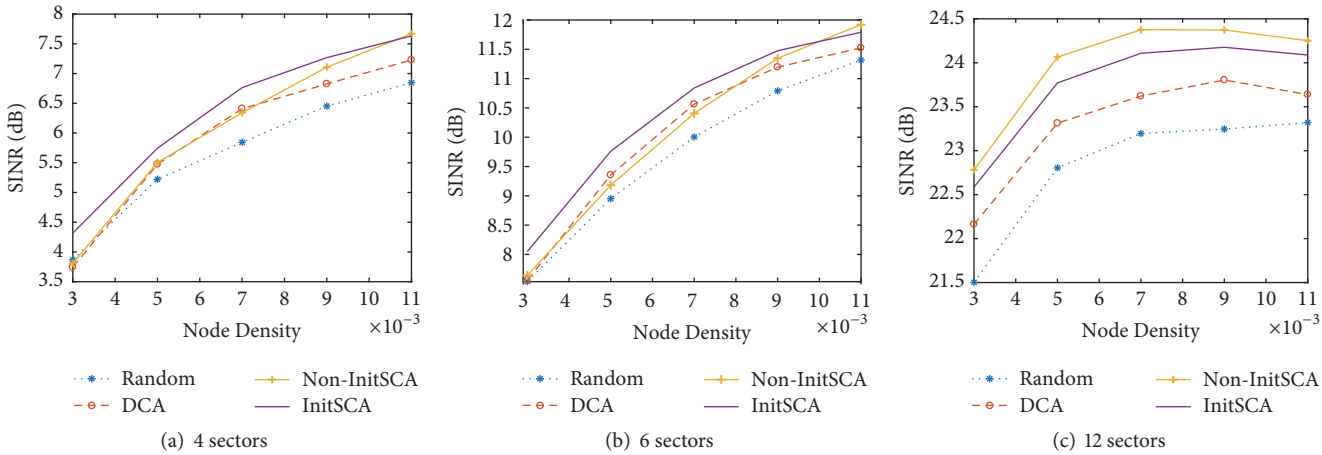
$$f\left(r\right) = 2\pi r \lambda e^{-\lambda \pi r^2} \tag{34}$$

Figure 6 shows the average distance with varying vehicle density. Density 3e-3 shows average 9 $m$ distance to the nearest node while density 11e-3 shows less than 5 $m$. According to this average distance, SNR of a link to the nearest node is plotted in Figure 7 based on the channel model discussed in Section 5. The SNR increases as the vehicle density grows because path loss reduces by decreasing link distance to a neighbor node. Furthermore, the SNR is varying by directivity gain as shown in Figure 3 (also refer to Table 1). The SNR increases as the number of sectors decreases; in 4 sectors, the SNR is around 6 dB while it is much higher in 6 and 12 sectors, about 10 and 27 dB, respectively, in Figure 7. Also, beam alignment affects the SNR; in particular, a narrower beam causes more difference in SNR. For instance, there is only a 4 dB gap between $Gg$ and $gg$ in 4 sectors, but 15 dB in 12 sectors (beam alignments $Gg$ and $gg$ described at Table 2). In this study, the $Gg$ or $gg$ case is only for interference since we assume that all communication pairs tune to each other. Here we can conclude that the GiV2V can suffer from higher interference by the aligned network topologies, in which a node can receive $GG$ interference in addition to the $Gg$ and $gg$.




(a) 4 sectors (b) 6 sectors (c) 12 sectors

Figure 7: Average SNR of the shortest GiV2V link.




(a) 4 sectors (b) 6 sectors (c) 12 sectors

Figure 8: Average SINR with 3 channels.

*7.1. Comparison of Proposed Algorithms.* We compare our 4 channel assignment algorithms in a PPP highway model with varying density of vehicles. To obtain average throughput, total 200 runs with random topologies are conducted. Each vehicle chooses nearest another randomly as a receiver or relay and transmits data. Transmission power is 20 dBm, and topology change due to vehicle mobility is not considered.

Figure 8 shows average SINR of transmission links at different node densities when applying each algorithm with 3 channels. Performance shows a similar pattern regardless of number of sectors. The random allocation achieves the lowest throughput while the InitSCA mostly outperforms others. Noninitialized SCA (later called just SCA) and DCA approaches are comparable. In particular, the SCA shows better throughput in 12 sectors than the InitSCA because of increasing interference from narrow beams.

Here we observe that the node density and antenna directivity affect SINR with different degree of interference. For instance, node density 3e-3 has about 6 dB for SNR in Figure 7(a) and 4 dB as SINR in Figure 8(a) with interference in case of the random assignment, while node density 11e-3 has almost 12 dB SNR but only 7 dB for SINR. In other words, the higher node density 11e-3 suffers more interference, 2 versus 5 dB reduction. For the directivity, Figure 7(c) shows more or less 30 dB SNR in node density 3e-3 and SINR of the same density is 20 dB in case of the random algorithm in Figure 8(c). Thus, almost 10 dB reduction occurs due to the interference in 12 sectors, which is much higher than the case of 4 sectors, 2 dB reduction because nodes can cause stronger interference and interference coverage expansion with high directivity. The interference becomes severe as the density increases; in 12 sectors, node density 11e-3 shows that 35 dB SNR is degraded to only 23 dB SINR in case of the random algorithm.

Figure 9 shows experiment results with 6 channels (e.g., IEEE 802.11ad channels), which allows more degree of freedom for interference avoidance. Compared to 3 channel results, low density allows comparable throughput among all algorithms because of orthogonality in space and channel divisions. However, as density grows, gaps among algorithms

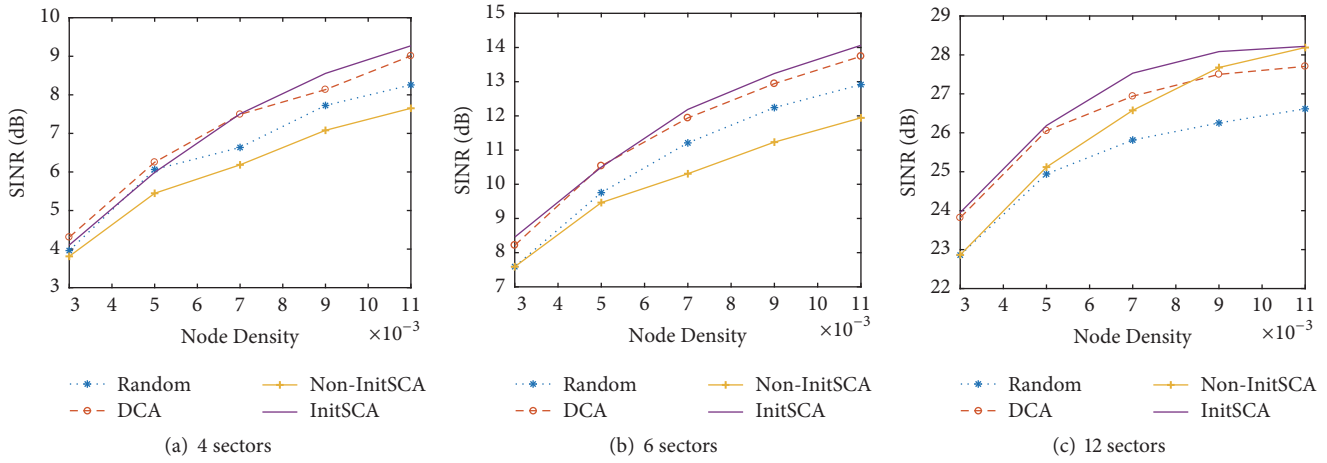



Figure 9: Average SINR with 6 channels.

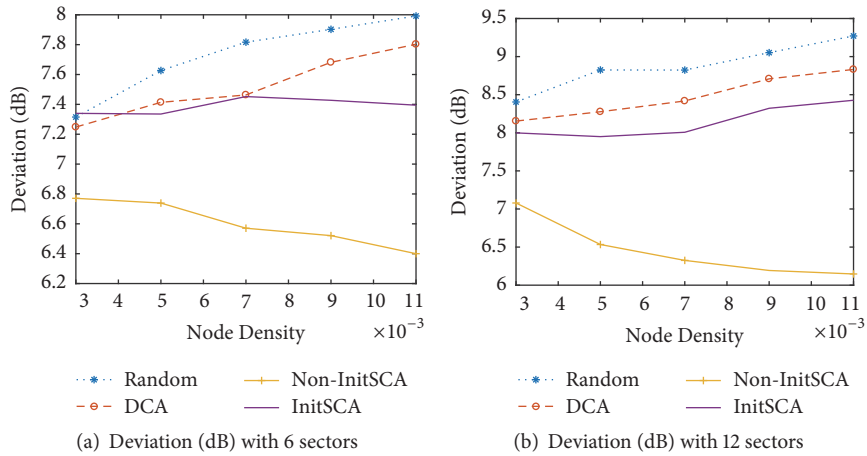
increase; the SCA achieves lower throughput than others including the random algorithm. From this, the SCA algorithm is assumed to stop at a local maximum while the random or DCA mechanism searches a solution closer to a global optimum. However, the SCA shows bit better performance in 12 sectors than the random and DCA distributed algorithms since the high interference avoidance may reduce overall average SINR.

Figures 10 and 11 show standard deviation and 0 dB probability (i.e., a level of lowest MCS) of SINR of results in Figure 9. Standard deviation indicates difference of SINR among transmission links. Thus, the max-min-like SCA reduces the difference effectively because it assigns a channel first to a link that has worst SINR compared to others. The InitSCA shows lower deviation than the DCA by adjusting channel assignment for some links with low SINR. In low density, the deviation is affected mostly by length of each link while the deviation is determined by interference in high density. Figures 10(a) and 10(b) show that the deviation of the SCA and InitSCA is consistent or decreasing while the deviation of others increases; increasing interference from node density is effectively managed in case of the SCA and InitSCA compared to others. Furthermore, the probability that is less than 0-dB shows clearly max-min fairness achievement of the SCA in Figures 11(a) and 11(b). Similar with the deviation, the SCA has the lowest probability of 0-dB SINR among algorithms and shows robustness with increasing density. In 12 sectors, the 0-dB probability of the SCA is almost zero because of high spatial diversity; other algorithms' probabilities are also quite lower than ones of 6 sectors. Note that the probability decreases as the density grows because of path loss reduction. As a result, multihop relays can be effective rather than direct communications in dense GiV2V networks.

Figure 12 depicts throughput with reduced transmission power, 10 dBm instead of 20 dBm. A total of 6 channels are assigned to nodes. Compared to results in Figure 9(a), SINR decreases, for an example of the 4 sectors, from 9 to 2 dB in high density due to Tx power reduction as shown in Figure 12(a). In low density, SINR shows less than 0 dB, i.e., -4 dB. Also, SINR difference among algorithms decreases, which implies that throughput difference from varying channel diversity gain becomes reduced as interference effect reduces. However, the throughput gap increases again as the interference increases due to high beam directivity in Figure 12(c). Transmission power control can reduce interference but together with receive signal strength that mainly affects the throughput rather than the interference that can be avoidable by spatial or channel diversity.

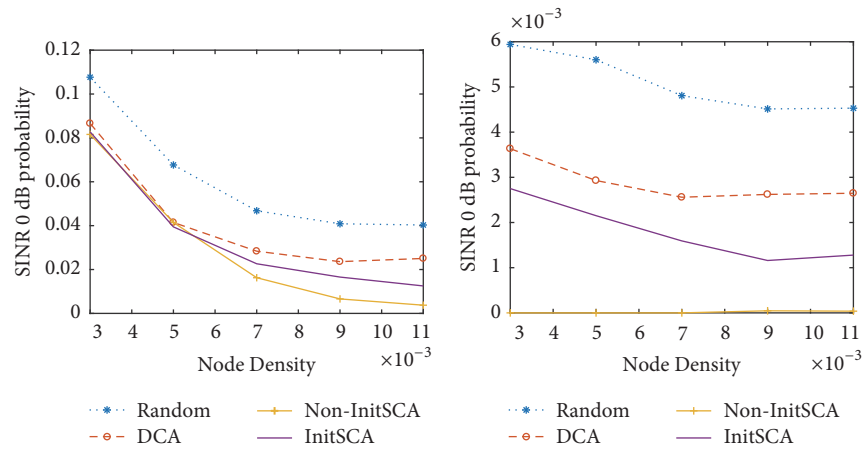
*7.2. Comparison of Global Optimization Algorithms.* According to experiment results in Section 7.1, the InitSCA is the most effective algorithm among proposed algorithms across all densities and sectors. In this section, we compare the InitSCA performance with three well-known metaheuristic algorithms for seeking a global optimum, which are popularly used for large scale optimization and NP-Hard problems. For this comparison, we apply the same parameters used in the simulation of Section 7.1. Actual global optimum can be found by exhaustive search algorithms like brand and bound, but its average value (e.g., from 200 runs) requires much computation time due to the complexity of experiment scenario.

Here we introduce briefly those three algorithms: Simulated Annealing (SA) [46], Generic Algorithm (GA) [47], and Particle Swarm Optimization (PSO) [48]. First, the SA follows a physical annealing analogy, in which heated particles in a liquid state are cooled down slowly to reach thermal equilibrium. In the algorithm, the particles are converged into lowest energy level by probability that moves to a new state, which decreases to zero along the cooling temperature. Here final energy level is determined by cooling speed; slow cooling leads to acquire a global optimum but delay might not be negligible for real-time systems. In our system, we limit number of iterations as 1e4. Second, the GA follows a process of natural DNA evolution, which is based on DNA operations such as mutation, crossover, and selection.



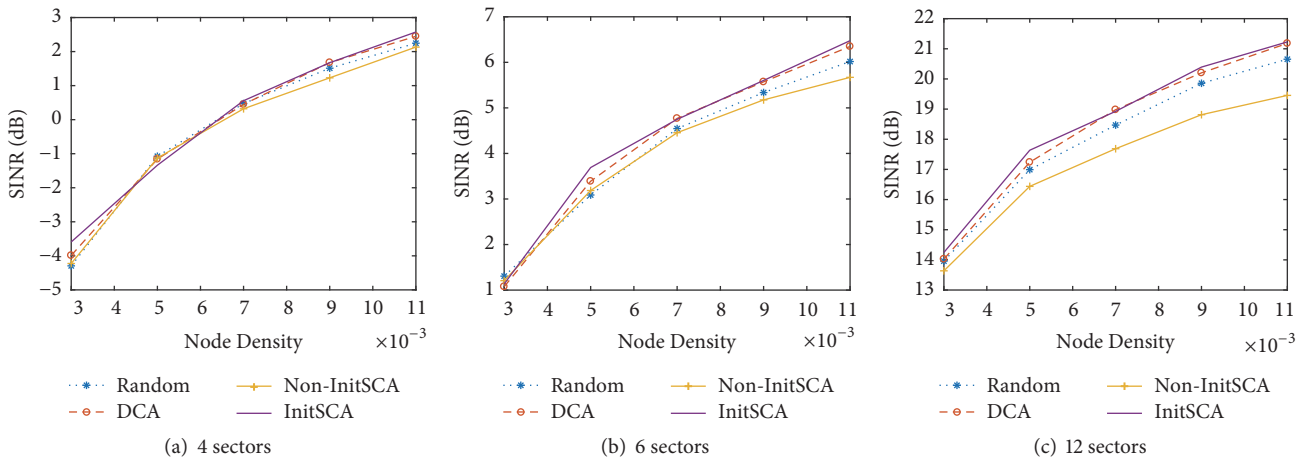

(a) Deviation (dB) with 6 sectors (b) Deviation (dB) with 12 sectors

Figure 10: Standard deviation with 6 channels.




(a) SINR 0 dB probability with 6 sectors (b) SINR 0 dB probability with 12 sectors

Figure 11: Zero SINR probability with 6 channels.




(a) 4 sectors (b) 6 sectors (c) 12 sectors

Figure 12: Average SINR with Tx power, 10 dBm.

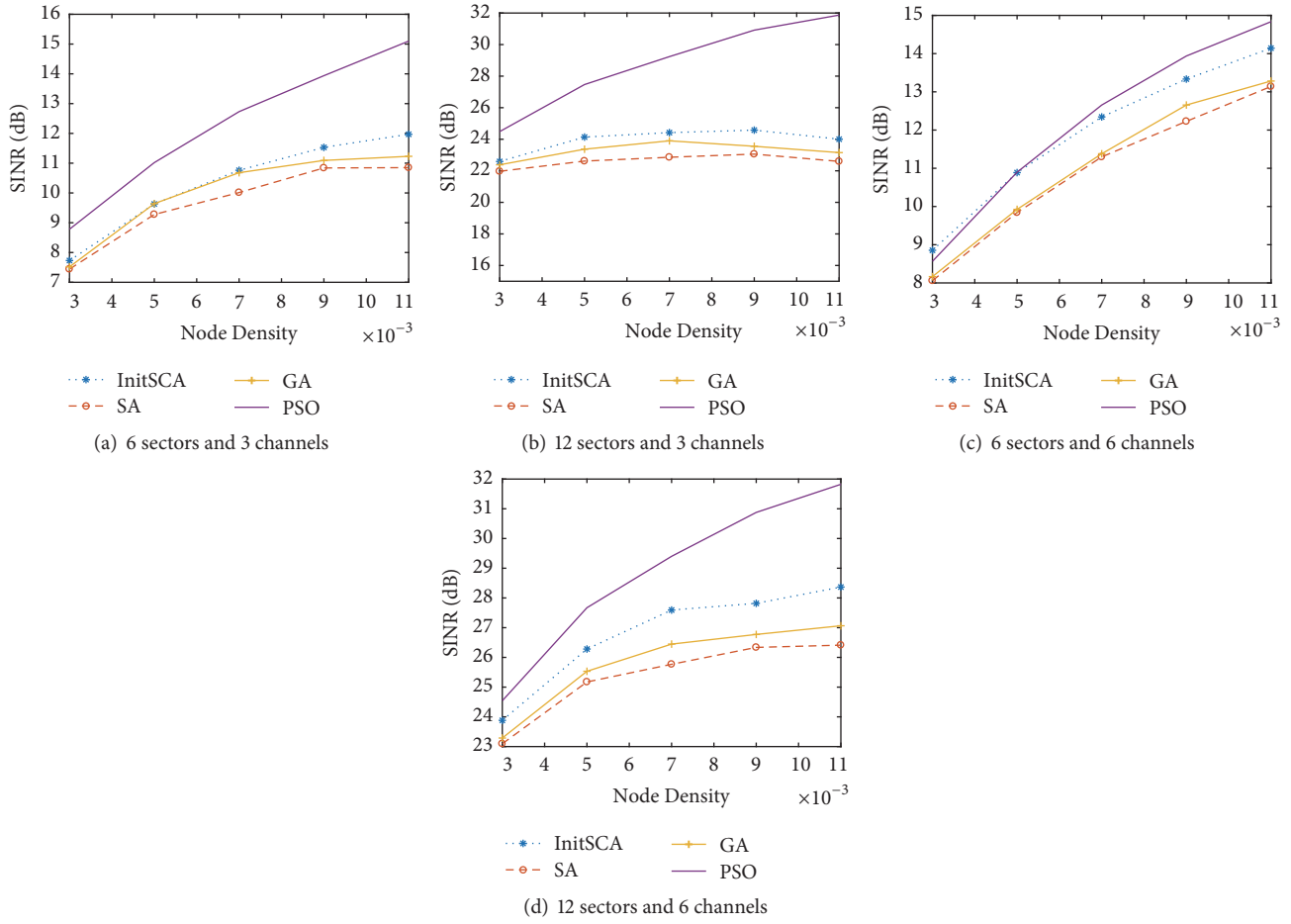



Figure 13: Average SINR of global optimization algorithms.

Candidate solutions are evolved and selected by such bio-inspired operations until finding a global optimum. Third, PSO generates similarly swarm of particles as candidate random solutions, in which particles search for better solutions in the search space according to swarm's best known position and find optima by updating generations. Compared to the GA, PSO is easy to implement with a few parameters to adjust for simple formulae.

Figure 13 shows average SINR of algorithms with varying channels and sectors. In all cases, PSO outperforms other algorithms, especially for 12 sectors. InitSCA is comparable with the GA or SA in case of 3 channels regardless of sectors. With 6 channels, the InitSCA shows slightly better throughput, about 1 dB than the GA and SA. Consequently, the GA and SA probably do not reach a global optimum although they are good solutions. Optimization of SA and GA such as cooling speed and crossover strategy in order to find the optimum is left for our future works.

Metaheuristic algorithms require conventionally considerable search time which is inappropriate for real-time applications like wireless communications. In the GiV2V networks, channels should be reassigned according to topology change; the GiV2V can have a new topology in a couple of seconds considering vehicle speed. Time complexity of our algorithm is shown as $O(n)$ in Section 6.5, but other metaheuristic algorithms have different complexity with a population size (refer to [49] for each complexity). Table 4 and Figure 14 show average elapsed time for the global algorithms at each node density. The initialized SCA takes only several ten milliseconds for all density cases, but SA and GA consume about 5 to 35 seconds; time increases drastically according to the node density. Due to this long delay, those two algorithms are hard to apply to the GiV2V networks. PSO delay, less than 2 seconds at highest density, seems competitive considering its performance shown in our simulation.

## 8. Conclusion

In this study, we propose a new VANET architecture, GiV2V, using mmWave spectrum and investigate its performance with simulation. Beamforming for mmWave links can increase receive signal quality and overcome high propagation loss of the mmWave. However, it can also cause higher interference in beam-aligned ad hoc networks, especially in high node density. In this study, we propose a simple distributed algorithm and centralized greedy algorithm based

Table 4: Elapsed time of global algorithms.

| | 3e-3 | 5e-3 | 7e-3 | 9e-3 | 11e-3 |
|---|---|---|---|---|---|
| InitSCA | 0.006994 | 0.014559 | 0.024938 | 0.041612 | 0.067441 |
| SA | 4.525196 | 8.652447 | 14.37165 | 22.30713 | 32.49279 |
| GA | 1.811456 | 4.840629 | 13.46396 | 24.10401 | 37.83527 |
| PSO | 0.385824 | 0.625597 | 0.92461 | 1.237252 | 1.665746 |

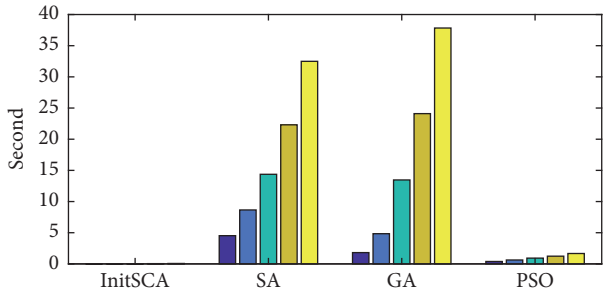



Figure 14: Average computation time of algorithms with increasing density (5 $\lambda$ from 3e-3 to 11e-3).

on SINR. Although the centralized algorithm outperforms distributed one, the distributed algorithm is still competitive at high degree of freedom from many channels and less complicated to implement. The centralized greedy algorithm shows comparable throughput with several metaheuristic algorithms while its complexity is lower and appropriate for real-time GiV2V systems. We will experiment further to evaluate the proposed algorithms under vehicle mobility simulation and look for optimum values using branch and bound algorithm to compare as future works.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (no. NRF-2017R1C1B1006607).

## References

[1] "ITU," https://www.itu.int/.

[2] "5GPPP," https://5g-ppp.eu/.

[3] "3GPP," http://www.3gpp.org/.

[4] "IEEE, "IEEE 802.15.3 Working Group, Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs), IEEE Unapproved Draft Std P802.15.3c/D10," bibtex:ieee80215c.

[5] *Part11: Wireless LAN Medium Acess Control (MAC) and Physical Layer (PHY) Specifications C Amendment 5: Enhancements for Very High Throughput in the 60 GHz Band*, 2010, IEEE P802.11ad/D1.0.

[6] Y.-B. Ko, V. Shankarkumar, and N. H. Vaidya, "Medium access control protocols using directional antennas in ad hoc networks," in *Proceedings of the INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, 2000.

[7] M. Takai, J. Martin, A. Ren, and R. Bagrodia, "Directional virtual carrier sensing for directional antennas in mobile ad hoc networks," in *Proceedings of the MOBIHOC 2002: PROCEEDINGS OF The Third ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 183–193, Switzerland, June 2002.

[8] R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, "Using directional antennas for medium access control in ad hoc networks," in *Proceedings of The Eight Annual International Conference on Mobile Computing and Networking*, pp. 59–70, USA, September 2002.

[9] V. Kolar, S. Tilak, and N. Abu-Ghazaleh, "Avoiding head of line blocking in directional antenna [MAC protocol]," in *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks. LCN 2004*, pp. 385–392, Tampa, FL, USA.

[10] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad hoc networking with directional antennas: A complete system solution," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 496–506, 2005.

[11] T. Korakis, G. Jakllari, and L. Tassiulas, "CDR-MAC: A protocol for full exploitation of directional antennas in ad hoc wireless networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 2, pp. 145–155, 2008.

[12] G. Jakllari, I. Broustis, T. Korakis, S. V. Krishnamurthy, and L. Tassiulas, "Handling asymmetry in gain in directional antenna equipped ad hoc networks," in *Proceedings of the IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 2, pp. 1284–1288, IEEE, Berlin, Germany, September 2005.

[13] E. Shihab, L. Cai, and J. Pan, "A distributed asynchronous directional-to-directional MAC protocol for wireless ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5124–5134, 2009.

[14] H. Gossain, C. Cordeiro, and D. P. Agrawal, "MDA: An efficient directional MAC scheme for wireless ad hoc networks," in *Proceedings of the GLOBECOM'05: IEEE Global Telecommunications Conference, 2005*, pp. 3633–3637, USA, December 2005.

[15] Y. Li and A. M. Safwat, "DMAC-DACA: Enabling efficient medium access for wireless ad hoc networks with directional

antennas," in *Proceedings of the 2006 1st International Symposium on Wireless Pervasive Computing*, pp. 1–5, Thailand, January 2006.

[16] M. Takata, M. Bandai, and T. Watanabe, "A MAC protocol with directional antennas for deafness avoidance in ad hoc networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference, GLOBECOM 2007*, pp. 620–625, USA, November 2007.

[17] S. Singh, R. Mudumbai, and U. Madhow, "Distributed coordination with deaf neighbors: efficient medium access for 60 GHz mesh networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, March 2010.

[18] T. Rappaport, S. Sun, R. Mayzus et al., "Millimeter wave mobile communications for 5G cellular: it will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[19] S. Sun, G. R. Maccartney, M. K. Samimi, S. Nie, and T. S. Rappaport, "Millimeter wave multi-beam antenna combining for 5G cellular link improvement in New York City," in *Proceedings of the 1st IEEE International Conference on Communications (ICC '14)*, pp. 5468–5473, IEEE, Sydney, Australia, June 2014.

[20] G. R. Maccartney Jr. and T. S. Rappaport, "73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City," in *Proceedings of the 1st IEEE International Conference on Communications (ICC '14)*, pp. 4862–4867, Sydney, Australia, June 2014.

[21] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: from fixed access to cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 168–178, 2015.

[22] C. Dehos, J. González, A. de Domenico, D. Ktéenas, and L. Dussopt, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 88–95, 2014.

[23] X. An and R. Hekmat, "Directional MAC protocol for millimeter wave based wireless personal area networks," in *Proceedings of the IEEE 67th Vehicular Technology Conference-Spring (VTC '08)*, pp. 1636–1640, May 2008.

[24] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. Rodwell, "Blockage and directivity in 60 GHz wireless personal area networks: from cross-layer model to multihop MAC design," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1400–1413, 2009.

[25] S. L. Cotton, W. G. Scanlon, and B. K. Madahar, "Millimeter-wave soldier-to-soldier communications for covert battlefield operations," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 72–81, 2009.

[26] M. X. Gong, R. Stacey, D. Akhmetov, and S. Mao, "A directional CSMA/CA protocol for mmWave wireless PANs," in *Proceedings of the IEEE Wireless Communications and Networking Conference 2010, WCNC 2010*, Australia, April 2010.

[27] Q. Chen, X. Peng, J. Yang, and F. Chin, "Spatial reuse strategy in mmWave WPANs with directional antennas," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 5392–5397, Anaheim, Calif, USA, December 2012.

[28] I. K. Son, S. Mao, M. X. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 2149–2157, March 2012.

[29] A. Thornburg, T. Bai, and J. Heath, "Performance analysis of outdoor mmWave ad hoc networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 4065–4079, 2016.

[30] H. Park, Y. Kim, T. Song, and S. Pack, "Multiband directional neighbor discovery in self-organized mmWave Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1143–1155, 2015.

[31] A. Tassi, M. Egan, R. J. Piechocki, and A. Nix, "Modeling and Design of Millimeter-Wave Networks for Highway Vehicular Communication," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10676–10691, 2017.

[32] J. Chen, J. Xie, Y. Gu et al., "Long-Range and Broadband Aerial Communication Using Directional Antennas (ACDA): Design and Implementation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10793–10805, 2017.

[33] Z. He, S. Mao, S. Kompella, and A. Swami, "On Link Scheduling in Dual-Hop 60-GHz mmWave Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11180–11192, 2017.

[34] Q. Wang, D. W. Matolak, and B. Ai, "Shadowing Characterization for 5 GHz Vehicle-to-Vehicle Channels," *IEEE Transactions on Vehicular Technology*, 2017.

[35] D. He, B. Ai, K. Guan et al., "Influence of Typical Railway Objects in mmWave Propagation Channel," *IEEE Transactions on Vehicular Technology*, 2017.

[36] P. Kumari, J. Choi, N. Gonzalez Prelcic, and R. W. Heath, "IEEE 802.11ad-based Radar: An Approach to Joint Vehicular Communication-Radar System," *IEEE Transactions on Vehicular Technology*, 2017.

[37] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds," in *Proceedings of the IEEE World Forum on Internet of Things (WF-IoT '14)*, pp. 241–246, March 2014.

[38] E. Lee, E.-K. Lee, M. Gerla, and S. Y. Oh, "Vehicular cloud networking: architecture and design principles," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 148–155, 2014.

[39] W. Kim, "Beyond LTE-advance for information centric networking," *Computer Standards & Interfaces*, vol. 49, pp. 59–66, 2017.

[40] "F.1336: Reference radiation patterns of omnidirectional, sectoral and other antennas for the fixed and mobile service for use in sharing studies in the frequency range from 400 mhz to about 70 ghz".

[41] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Khoryaev, and A. Lomayev, "Experimental investigations of 60 GHz WLAN systems in office environment," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1488–1499, 2009.

[42] E. Zola, A. J. Kassler, and W. Kim, "Joint user association and energy aware routing for green small cell mmWave backhaul networks," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference, WCNC 2017*, USA, March 2017.

[43] W. Kim, "Dual Connectivity in Heterogeneous Small Cell Networks with mmWave Backhauls," *Mobile Information Systems*, vol. 2016, 2016.

[44] O. Bazan and M. Jaseemuddin, "An opportunistic directional MAC protocol for multihop wireless networks with switched beam directional antennas," in *Proceedings of the IEEE International Conference on Communications, ICC 2008*, pp. 2775–2779, China, May 2008.

[45] M. J. Farooq, H. Elsawy, and M.-S. Alouini, "A Stochastic Geometry Model for Multi-Hop Highway Vehicular Communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2276–2291, 2016.

[46] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm," *ACM Transactions on Mathematical Software*, vol. 13, no. 3, pp. 262–280, 1987.

[47] W. K. Lai and G. G. Coghill, "Channel Assignment through Evolutionary Optimization," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 1, pp. 91–96, 1996.

[48] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*, pp. 760–766, Springer US, Boston, MA, USA, 2011.

[49] P. S. Oliveto, J. He, and X. Yao, "Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results," *International Journal of Automation and Computing*, vol. 4, no. 3, pp. 281–293, 2007.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3471610, 14 pages
https://doi.org/10.1155/2018/3471610

*Research Article*

# Micro Operator Design Pattern in 5G SDN/NFV Network

**Chia-Wei Tseng**[1], **Yu-Kai Huang**[1], **Fan-Hsun Tseng**[2], **Yao-Tsung Yang**[1], **Chien-Chang Liu**[1] **and Li-Der Chou**[1]

[1]*Department of Computer Science and Information Engineering, National Central University, Taoyoan 32001, Taiwan*
[2]*Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei 10610, Taiwan*

Correspondence should be addressed to Li-Der Chou; cld@csie.ncu.edu.tw

Received 27 February 2018; Accepted 30 May 2018; Published 10 July 2018

Academic Editor: Shao-Yu Lien



The trend of 5G mobile networks is increasing with the number of users and the transmission rate. Many operators are turning to small cell and indoor coverage of telecom network service. With the emerging Software Defined Networking and Network Function Virtualization technologies, Internet Service Provider is able to deploy their networks more flexibly and dynamically. In addition to the change of the wireless mobile network deployment model, it also drives the development trend of the Micro Operator ($\mu$O). Telecom operators can provide regional network services through public buildings, shopping malls, or industrial sites. In addition, localized network services are provided and bandwidth consumption is reduced. The distributed architecture of $\mu$O tackles computing requirements for applications, data, and services from cloud data center to edge network devices or to the micro data center of $\mu$O. The service model of $\mu$O is capable of reducing network latency in response to the low-latency applications for future 5G edge computing environment. This paper addresses the design pattern of 5G micro operator and proposes a Decision Tree Based Flow Redirection (DTBFR) mechanism to redirect the traffic flows to neighbor service nodes. The DTBFR mechanism allows different $\mu$Os to share network resources and speed up the development of edge computing in the future.

## 1. Introduction

Many new operators have chosen to turn to small cells and micro telecommunication network service that covers indoor areas in the wake of the coming of the 5G era, the changes of habits of users when going online, and the increasing demands for applications. Telecom operators can provide services via a variety of access networks such as 3G, 4G, 5G, and even Wi-Fi [1–3]. With the emerging Software Defined Networking (SDN)/Network Function Virtualization (NFV) technologies, Internet Service Provider (ISP) is able to deploy their networks more flexibly and dynamically [4–6]. ISPs are now capable of providing localized services by means of public buildings, shopping malls, or industrial facilities, and this has not only changed the model of distributing and establishing mobile network but also gave birth to the idea of the micro operator ($\mu$O)[7–9]. Factors such as scarcity of spectrum resources, the effective employment of bandwidth, the scale of the mobile market, and consumers' options for the diversity of services conspire to produce the inception of Mobile Virtual Network Operator (MVNO), which in turn brings more diverse applications and services to the telecommunication market [10]. MVNO can refer to any telecommunication operators who provide wireless communication services to consumers and do not have their own wireless network infrastructures. In order to expand the coverage of their business, Mobile Networker Operators (MNO) might choose to cooperate with other MNOs by renting the bandwidth and time of use in their domains to rapidly develop the business of MVNO. MVNO provides brand new opportunities for development for the matured mobile communication market and spurs the telecommunication market to move toward a service-oriented competition. As for 5G mobile network operators, development of small cells and micro telecommunication network service that covers indoor areas is now being initiated to satisfy users' demand for indoor coverage, maturing the concept of $\mu$O [11–13].

Figure 1 shows the $\mu$O network architecture. $\mu$O is relatively small in scale and holds limited resources to provide particular and necessary services to a certain number of

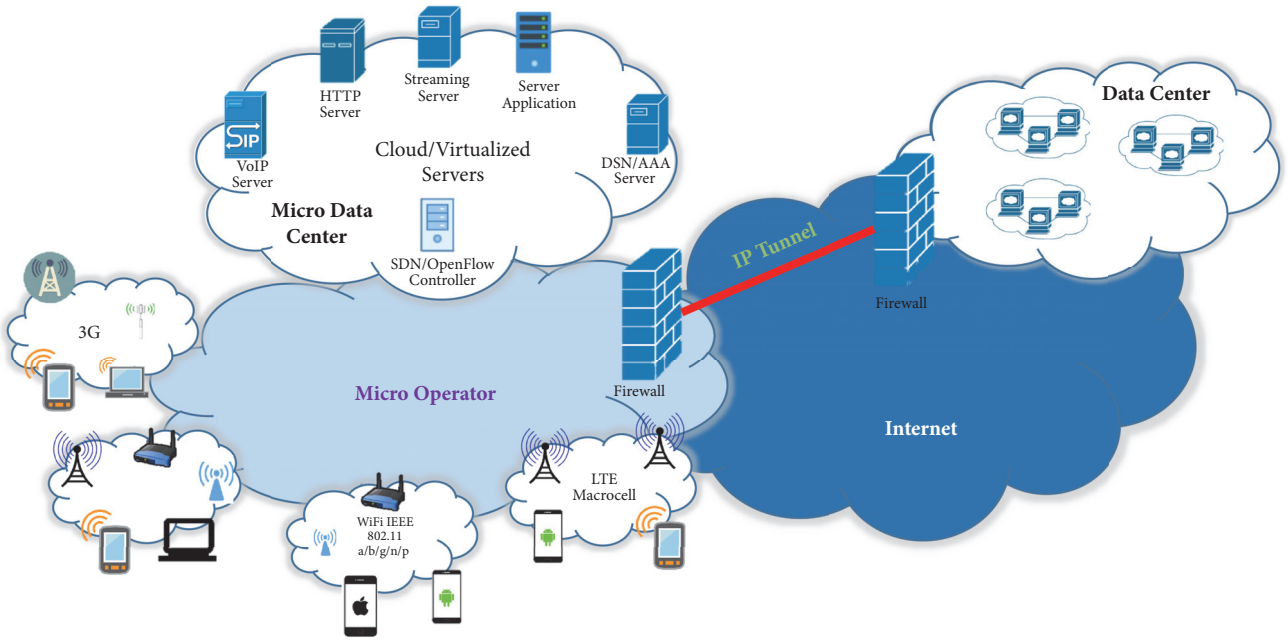



Figure 1: Micro operator network architecture.

users. The $\mu$O is originated from Oulu University in Finland that consists of three features [14]. (1) Mobile devices access local network specific service (2) spatially confined area with specific service (3) dependent on appropriate available network resources. Facilities like hospitals, schools, large conference, sports venues, shopping malls, hypermarkets, and factories can use local area network service to satisfy users' needs for all sorts of application services.

Restricted by resources and space, $\mu$O has the character of regionalization service that allows it to provide, under circumstances where hardware infrastructures are scarce and the resources are limited, different regional services according to different domains, making it possible for users of mobile devices to gain access to services in the nearby regions. Apart from reducing the consumption of bandwidth resources by offering nearby network services, this kind of localization service can also transfer applications, data, and the computing of services from nodes in the data center on cloud to edge nodes in logical LAN to be processed and realize the developing environment for edge computing [15, 16], reduce network latency, and satisfy 5G demands for amelioration of latency.

The purpose of this paper is to discuss the $\mu$O design pattern in 5G SDN/NFV network and establishes an experimental environment to evaluate the decision tree based traffic flow redirection mechanism. To realize the customized experiment environment, the network functions virtualization is taken into consideration in our research. Our contributions are listed as follows:

(1) A $\mu$O design pattern is constructed based on SDN and NFV that combines network slicing and tunneling technologies to build a network infrastructure for $\mu$Os.

(2) A DTBFR mechanism for $\mu$O is proposed and the decision tree theory is utilized to serve as the reference in the determining the SDN traffic flows direction.

(3) We established a $\mu$O environment for validation of the DTBFR mechanism and $\mu$Os communication experiments.

The rest of the paper is organized as follows: in Section 2, the background and related works are addressed. Section 3 describes the $\mu$O design pattern and DTBFR mechanism. Section 4 presents the experiment results. The last section concludes this paper.

## 2. Background and Related Works

The development of 5G application services will be largely on the Internet of Things (IoT) and encourages the new communication market to move toward a more vertical-subdivision one [17, 18]. The result will be the formation of different emerging application service scenarios and more diverse demands for networks. However, as far as telecommunication is concerned, despite the fact that there is now a globally agreed requirement on IoT constructed by 5G on characters such as transmitting speed, capacity, coverage, and security, there is still room for the business model of 5G to improve itself. 5G mobile broadband network puts the accent on small cells/base stations, how to amplify indoor coverage, providing faster services for users, and reducing the latency in network transmission, all of which pose a great challenge for telecommunication operators [19–21]. Furthermore, the appearance of MVNOs has brought new opportunities for development for MNOs who have not yet gained licenses to operate mobile communication business.

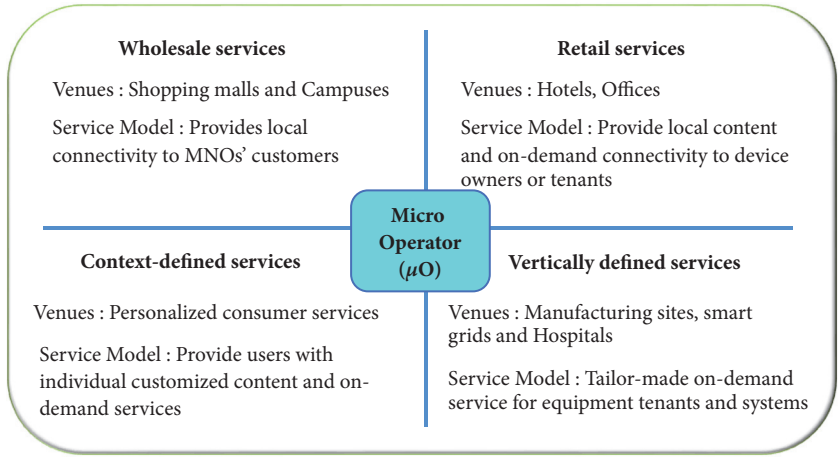



Figure 2: Micro operator service model [24].

MVNOs use MNOs' spectrum and network to provide customized mobile communication services, the mobile virtual private network for enterprises, or scores of other micro markets that operators still have not stretched their services to such as less lucrative or regional local emerging markets. A trend of regionalization has been observed among small cell/base stations and the regional services do not processes the characters of public area network, industrial application network, and a hybrid network. $\mu$O is an emerging network business model that has started developing under this backdrop [22]. Reference [14] illustrates the relation between $\mu$O and MNO and scrutinizes the use cases of $\mu$O-MNO to set the foundation for the possible business model for $\mu$O. Reference [23] proposes a mechanism in regard to $\mu$O's shared spectrum access communication and shares the infrastructure of mono-network/physical network via virtual technology to fully utilize the valuable bandwidth resource.

As shown in Figure 2, the service model of $\mu$O can be divided into four quadrants:

(1) The first quadrant is closed, controlled, independent, and general service that often takes places in office. The users in this region can gain access to the services provided by the region.

(2) The second quadrant is shared, transparent, and general service that often takes places in shopping malls and schools. Such regions are often packed with people and a great deal of data is required as well. A micro data center might be set in the region for users to quickly gain access to the services.

(3) The third quadrant is shared, transparent, and special service that often takes place in scenarios like telematics network and customized services. Such services are likely to be used by end-users. The communication among end-users requires special network protocols.

(4) The fourth quadrant is closed, controlled, independent, and special service that might take place in hospitals, automated factories, and smart grids. Hospitals provide irreplaceable medical services. Users will have to connect to such regions before gaining access to the services.

Due to the increasing demand of mobile device users accessing Internet services, operators need to dynamically adjust and combine to meet the requirements of different applications in order to improve network performance. The service architecture of 5G $\mu$O must also be developed in conjunction with different technologies such as SDN and NFV [25, 26]. SDN abstracts the network architecture by decoupling the network control and forwarding functions enabling the network control to become directly programmable and the underlying infrastructure to be abstracted for applications and network services [27]. By separating the control plane from the data forwarding plane and virtualizing all of the connections, administrator can remove the hard-wired barriers of networks and quickly change structures to suit their needs. In addition to the SDN, NFV is a core structural change in the way telecommunication infrastructure gets deployed. The goal of NFV is to enable service providers to reduce costs and faster service delivery. The requirements and open standards that underpin NFV are being developed by the European Telecommunications Standards Institute (ETSI) [28].

Since 5G technology is now undergoing rapid development, the conventional network infrastructure is not able to meet 5G's diverse demands and $\mu$Os' development of service model in the future. Therefore, the international standards organization and equipment manufacturers now set out to promote the technology of network slicing which allows ISP to use SDN and NFV technologies to realize network virtual slicing, the division of several different service scenarios, and the provision of customized network service [29–31]. The 5G white paper proposed by Next Generation Mobile Networks (NGMN) that expounds 5G's network infrastructure in the future had included the concept of network slicing technology in order to support particular connection forms and use particular methods to deal with the communication services of control/user-plane [32]. To satisfy the requirements of 5G network's flexible configuration, NOKIA had come up with the programmable 5G network infrastructure where network slices of several independent virtual subnetworks

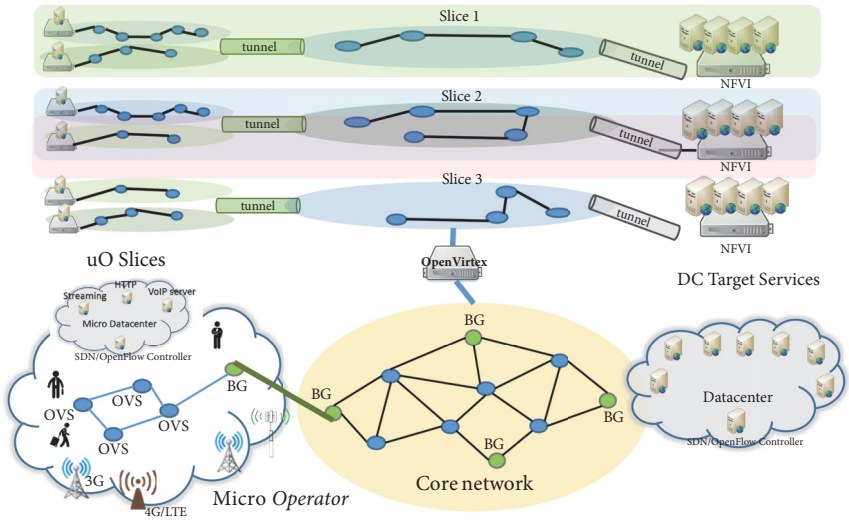



Figure 3: A $\mu$O design pattern with network slicing.

are constructed in the same basic infrastructure to serve the purpose of the specific application [33]. Network slicing technology can divide network resources into multiple independent virtual networks. Not only can it satisfy the diverse demands of small cells/base stations and $\mu$O but it can also work as the crucial technology for the development of 5G network communication. As for virtualization tools, in order to realize the function of network slicing, [34] proposes the concept of Network Virtualization Layer–FlowVisor that divides a physical network into several virtual networks that each possesses their own controllers and enables personal virtual network to be built in the same physical network equipment to achieve the goal of resource allocation. Reference [35] proposes a platform of network virtualization called OpenVirtex (OVX) that is capable of realizing network virtualization for multiple tenants. OVX can map a user's network topology to a physical topology according to the user's needs and accomplish the network-connecting process. The major difference between OVX and FlowVisor is that users of OVX do not have to worry about the basic physical topology and can define their won virtual network topology. OVX will assist its users to accomplish the mapping of virtual network topology to physical network topology.

The conventional offline benchmark planning is no longer suitable to satisfy 5G $\mu$Os' efficient utilization of resources and demands of sharing and the full utilization of computing resources to respond to the complicated situations of different demands to determine the direction of the traffic flow. Machine learning is on the list of the most potential candidates [36]. It provides a much stronger and complicated decision-making ability. Decision tree is one of the most popular machine learning algorithms used all along. Decision tree, on the other hand, is highly efficient supervised learning model that is ideal for the prediction of classification and regression data type such as ID3 and C4.5 [37, 38]. C4.5 is the refined version of feature selection in ID3 algorithm that can be used to process continual class label and the problem of missing data in training data. It can also correct the problem of overfitting of the decision tree by means of pruning. As for the application of route choosing and traffic flow identification in regard of mobile edge network, [39] uses the characters of SDN and combines them with decision tree algorithm to categorize traffic flow, allowing the controller to choose an access point from the local network that is suitable to be connected to a mobile device according to the level of congestion of the backhaul route. Reference[40] proposes a model called AMPS that combines SDN and machine learning. The model is capable of categorizing the differences of each traffic flow by means of learning packet identification and then chooses the best route of transmission for the traffic flow to achieve the automation of the optimal selection of routing path.

## 3. 5G $\mu$O Design Pattern and DTBFR Mechanism

The interaction between SDN and NFV enables the 5G network to build a system infrastructure in an abstract manner and further increase the flexibility of the network, allows the vertical system to be divided into multiple sliced constructor blocks, and builds a network that is connectable, programmed, and virtual. The operator can advance to use network slicing technology to virtualize networks and, as the aforesaid paragraphs, flexibly divide a physical network into multiple independent and segregated $\mu$O networks according to different usage scenarios. As for the design of the infrastructure, this paper uses SDN and NFV technologies as the base and combines network slicing and tunneling technologies to build a network infrastructure for $\mu$Os. The infrastructure allows users of different $\mu$Os to be concatenated via tunneling technology and then realizes the rapid connection of network to effectively enhance the interconnectivity of networks.

The proposed $\mu$O design pattern is shown in Figure 3. Network slicing technology realizes the logical partition of $\mu$Os' networks via OpenVirtex. The connection between the

core network and the $\mu$O is carried out by the tunnel built by SDN border gateway (BG). The Internet cloud data center and the micro data center are constructed by NFV infrastructure defined by ETSI to save the cost of equipment investment. The mission of SDN controller is to use OpenFlow to construct a passageway between BG at the edge of the core network and BG at the edge of the $\mu$O's network. Once the SDN controller begins executing the matching and concatenation between network passageways, the $\mu$O's network can build a connection via the tunnel and the main core network. The $\mu$O can continue finishing the construction of virtual network via OpenVirtex's virtual slicing technology, allowing users of the $\mu$O to gain access to the data on the nearby micro data center. The users can also, via tunnels, connect to the cloud data center on the Internet to access the network service of particular application. The proposed architecture pattern combines network slicing and tunneling to realize the communication model for $\mu$Os and further integrates bandwidth controlling technology to be applied to the communication bandwidth application of the $\mu$O's network. This will increase the utilization ratio of network resources and the efficiency of traffic flow and result in a better quality of experience (QoE) for the network users.

In response to the demand of $\mu$O's network resource distribution that allows users to gain access of nearby network resources, the paper proposes a DTBFR mechanism for $\mu$O and uses decision tree theory to serve as the reference in the determining the SDN traffic flows direction. Decision tree algorithms have been used for solving predictive analytics problems in the past few years. Decision tree is a supervised machine learning model with simple process intuition and high execution efficiency. Decision tree algorithm is mainly used to conduct systematic result and integration of dataset and to find the special class and label relation during the decision-making process. Compared with other ML models, execution speed is a major advantage. To quantify uncertain information and dataset, the paper uses entropy as a method to gauge uncertainty and level of chaos in (1). Entropy is a measure of the impurity in a collection of training examples. The entropy increases with the increase in uncertainty or randomness and decreases with a decrease in uncertainty or randomness. The production of information comes with uncertainty and entropy can gauge it according to its probability of occurrence [41]. With higher probability, the chaos is more likely to occur and the uncertainty is low, while on the contrary, the uncertainty is high.

$$Entropy\,(s) = \sum_{i=1}^{c} P_i \log_2 P_i \tag{1}$$

Decision trees algorithm uses feature selection to guide the decision of the most useful attributes. Different feature selection criteria result in different types of decision trees. ID3 uses entropy and information gain to construct a decision tree. Information gain is used to decide which feature to split on at each step in building the tree. Information gain measures the expected reduction in entropy by partitioning the examples according to an attribute. Calculations of ID3 information gain are shown in (2). The information gains criteria to measure the strength of the association between an attribute and class. Symbol S is the target class and Symbol A represents the class label. C4.5 is an extension of ID3 which is a similar tree generation algorithm. C4.5 uses gain ratio by splitting the training sets based on its test attributes. The gain ratio takes number and size of branches into account when choosing an attribute. It corrects the information gain by taking the intrinsic information of a split into account. The split information can be defined as shown in (3). The gain ration can be defined as shown in (4). The flowchart of decision tree construction is shown in Figure 4.

$$Gain\,(\boldsymbol{S},\boldsymbol{A}) = Entropy\,(\boldsymbol{S}) - \sum_{j=1}^{v} \frac{|S_j|}{|S|} Entropy\left(S_j\right) \tag{2}$$

$$SplitInfo\,(S, A) = -\sum_{i=1}^{c} \frac{|s_i|}{|S|} \log_2 \left( \frac{|S_i|}{|S|} \right) \tag{3}$$

$$GainRatio\,(S, A) = \frac{Gain\,(S, A)}{SpliteInfo\,(S, A)} \tag{4}$$

The dataset used by the decision tree consists of both class labels and decision result. Training data collects user application traffic primarily through SDN controller. The dataset includes user's network area, server's CPU/memory, and traffic oriented result. The data storage format is represented as [*user_area, s1_cpu, s2_cpu, s3_cpu, s1_mem, s2_mem, s3_mem, result*]. For example, the dataset $[2, 49.0, 35.6, 6.0, 29.0, 13.0, 36.3, 2]$ corresponds to the class label [*$\mu$O, s0_cpu, s0_memory, s1_cpu, s1_mem, s2_cpu, s2_mem, result*]. Training data is a certain percentage of an overall dataset along with a testing set. The more complete training data makes classification easier. The duplicate values and divorced values must be removed first from the dataset. Duplicate values refer the same values for each class label in the dataset. Divorced values are calculated from the average value of the class label. When the value of the class label is greater than the average of ±5, the dataset will be removed from dataset.

After completing training on data collection, we can start to establish the node of the decision tree. Calculate the *Gain(S, A)*, *SplitInfo(S, A)*, and *Gain_ratio(S, A)* from the class labels of the dataset. Pick the best *Gain_ratio (S, A)* from the list and save it to BestFeature as the root of the decision tree. The remaining unselected class labels are added to Unselected_labels. The execution of the recursion begins when the decision tree node is established. Pruning is a technique in machine learning that reduces the size of decision trees by pruning the tree based on the statistical confidence estimates. When the decision tree is created, load the training data and start traverse tree node from root to leaf. We adopted the pessimistic pruning strategy proposed used in C4.5 to avoid the need of pruning set and continuity correction to error rate at each node.

In this paper, the calculation of continuous class labels and discrete class labels is different. Discrete data is the type of data that has clear spaces between values while continuous

Table 1: Example of continuous class labels.

| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CPU | 85.4 | 90.3 | 78.4 | 80.5 | 80.5 | 75.2 | 66.0 | 90.3 | 75.2 | 80.5 | 75.8 | 90.3 | 75.2 | 90.3 |
| S | No | No | Yes | No | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes |

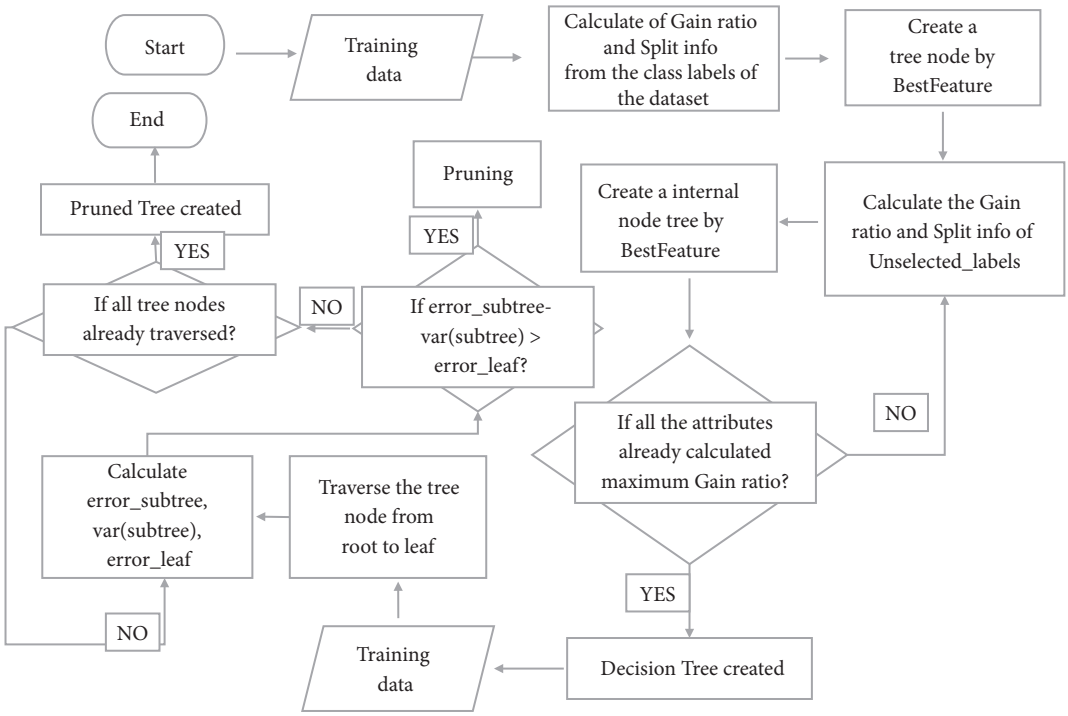



Figure 4: The decision tree construction flowchart.

data is data that falls on a continuous sequence. Table 1 shows the example of continuous class label. The second column shows the servers' CPU usage in $\mu$O. The third column S represents the output of the decision. In this example we start with number 66 to calculate the gain ratio. The dataset greater than 66.0 and the dataset less than or equal to 66.0 use (1) to compute the entropy value as 0 and 0.96, respectively. Then, the results are calculated by (2) and (3). Dividing the *Gain (S, 66.0)* by the (3) yields the gain ratio value of 0.129. Table 2 shows the calculation results and the class label CPU 66.0 gets the maximum gain ratio value in this example. Based on the above calculation, the best value is chosen in each round to serve as the node of the tree until the decision tree is established.

Based on the results of Table 2, the decision tree is constructed as shown in Figure 5. In order to correct the overlearning problem caused by the impure information in the decision tree, we adopted the pessimistic error pruning strategy to recursively estimate the error rate of the sample nodes covered by each internal node. Pessimistic error pruning is based on error estimates derived from the training data. If the internal node's error rate is lower than the estimated value, it will be replaced by the largest number of leaf nodes in the subtree.

Assume that an internal node of the tree covers N samples with E errors; then the error rate of the leaf nodes is given by (5). This penalty factor is 0.5. Pessimistic error pruning adds a constant to the training error of a subtree by assuming that each leaf automatically classifies a certain fraction of an instance incorrectly. This fraction is taken to be 1/2 divided by the total number of instances covered by the leaf and is called a continuity correction in statistics.

$$E\left(leaf\_error\_rate\right) = \frac{(E + 0.5)}{N} \tag{5}$$

$$E\left(subtree\_error\_rate\right) = \frac{\left(E + 0.5 * leaf\right)}{N} \tag{6}$$

After pruning, the internal node becomes a leaf node and the false positive number also need to add a penalty factor. The error rate of subtree is given by (6). Equations (7) and (8) show the false positive number and the standard deviation of

Table 2: Calculation result of the continuous class labels.

| Value | 66.0 | | 75.2 | | 75.8 | | 78.4 | | 80.5 | | 85.4 | | 90.3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| interval | ≦ | > | ≦ | > | ≦ | > | ≦ | > | ≦ | > | ≦ | > | ≦ | > |
| Yes | 1 | 8 | 3 | 6 | 4 | 5 | 5 | 4 | 7 | 2 | 7 | 2 | 9 | 0 |
| No | 0 | 5 | 1 | 4 | 1 | 4 | 1 | 4 | 2 | 3 | 3 | 2 | 5 | 0 |
| Entropy | 0 | 0.96 | 0.81 | 0.97 | 0.72 | 0.99 | 0.65 | 1 | 0.76 | 0.97 | 0.88 | 1 | 065 | 0 |
| Info(S,A) | 0.891 | | 0.92 | | 0.89 | | 0.85 | | 0.838 | | 0.91 | | 0.651 | |
| Gain | 0.048 | | 0.015 | | 0.045 | | 0.09 | | 0.102 | | 0.02 | | 0 | |
| Split Info | 0.371 | | 0.863 | | 0.94 | | 0.98 | | 0.94 | | 0.863 | | 0.651 | |
| Gain ratio | 0.129 | | 0.017 | | 0.047 | | 0.091 | | 0.108 | | 0.023 | | 0 | |

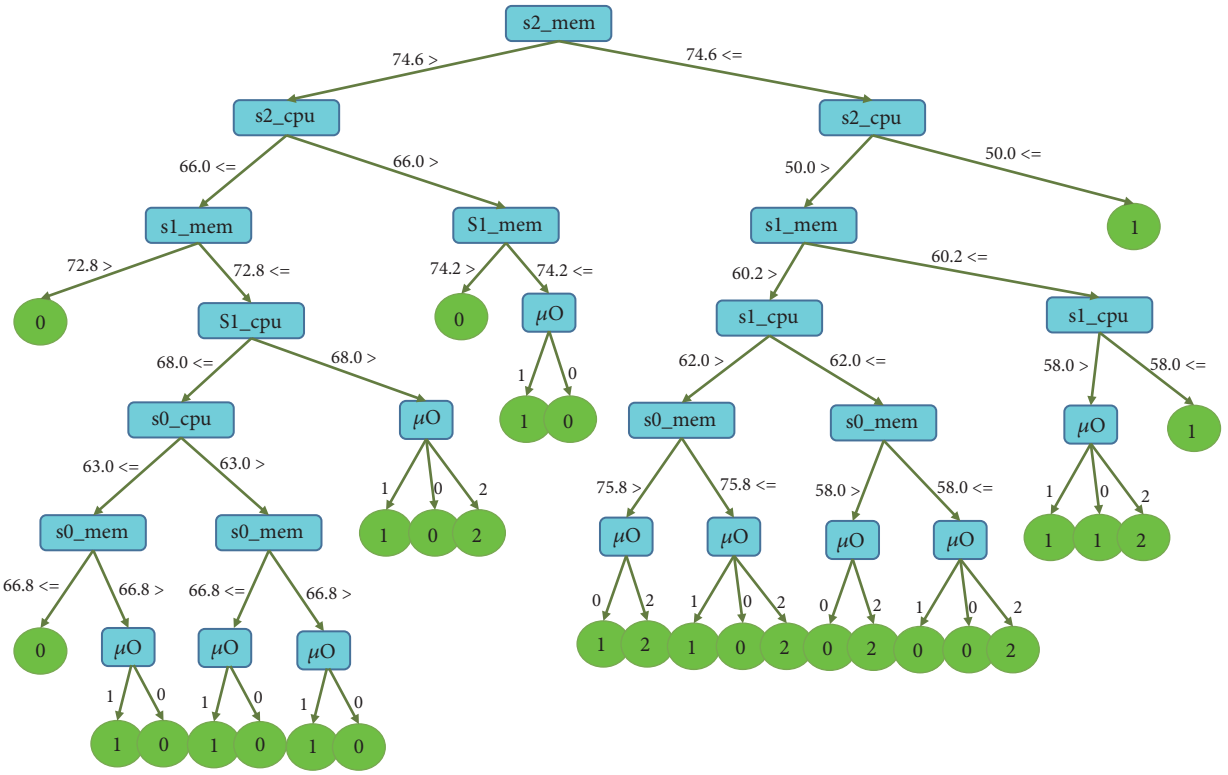



Figure 5: An example of established decision tree.

the subtree. The false positive number of leaf node is shown in (9).

$$E\left(subtree\_error\_count\right) = N * e \tag{7}$$

$$var\left(subtree\_error\_count\right) = \sqrt{N * e * (1 - e)} \tag{8}$$

$$E\left(leaf\_error\_count\right) = N * e \tag{9}$$

$$E\left(sub\_error\_count\right) - var\left(subtree\_error\_count\right) > E\left(leaf\_error\_count\right) \tag{10}$$

As shown in (10), if the number of misclassified subtrees minus the standard error of the error rate is still greater than the number of false positives corresponding to the leaf node; then the pruning will occur and the subtree will be removed from the tree. Until all nodes in all decision trees have been visited, the pruning process is completed. The pruning process is finished when all the nodes in all the decision trees are visited. Figure 6 shows the pruned decision tree.

The proposed DTBFR mechanism utilizes the SDN technology to collect the network traffic information. The calculation results of decision tree algorithm can provide instructions for SDN controller to guide user traffic flows to the services in the nearby 5G $\mu$O micro data center. The DTBFR mechanism combined with 5G SDN/NFV programmable network architecture can improve service efficiency and save available bandwidth resources.

## 4. Experiment Results

The purpose of this experiment is to verify $\mu$O's communication infrastructure pattern during operation. The experiment

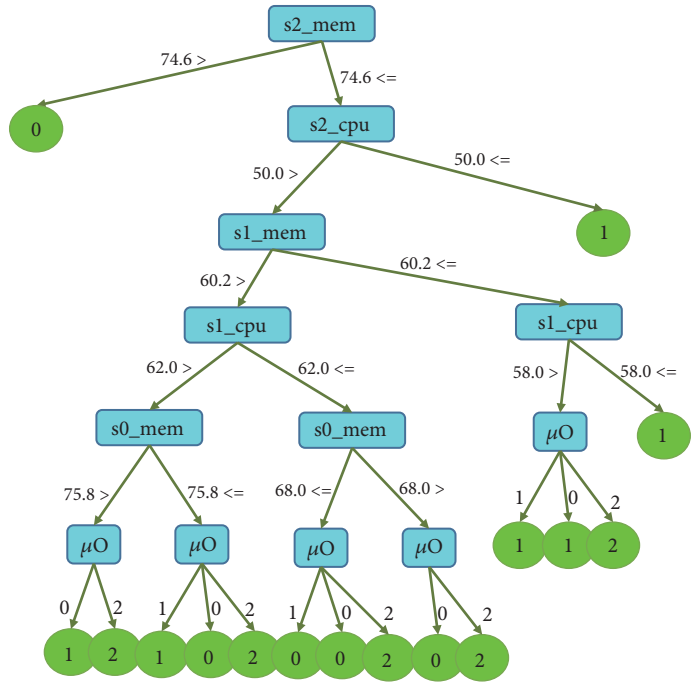



Figure 6: The pruned decision tree.

condition consists of multiple local area networks controlled by the $\mu$O and central tunnel controllers. In our evaluation, we simulated a network with tree-like topology in mininet and measured the overhead caused by network view division [42]. Mininet is used to simulate $\mu$O's network, and tunnels can be built among different $\mu$O areas for quick exchange of information. The experiment comprises two hypervisors, one OpenVirtex, and multiple SDN controllers shown in Figure 7. The topology is constructed by the mininet in VM in the hypervisor and the $\mu$O areas are concatenated by tunneling technology. The OpenVirtex functions as the system management system. The IT staff can construct network slicing for the system via network interface and the back-end server of the network interface will analyse the operating information before interpreting it into OpenVirtex to produce commands for network slicing. The webpage uses socket and slice manager to transmit information. After the slice manager receives the information, it will continue to use OpenVirtex's API to build network slicing for the manager.

Network slicing has a wide range of applications according to the cohorts of users, the types of application services, or different demands for resources. The paper uses regionalization service as the benchmark for network slicing and connects homogeneous regionalization service servers with tunneling technology to form a complete network slicing of regionalization service. A C4.5 decision tree model is built based on the mechanism proposed by Ross Quinlan to serve as the reference for determination of 5G $\mu$O's traffic flow direction and allows the users' traffic flow to be directed to a closer $\mu$O access service to reduce the network latency in packet transmission.

In order to verify $\mu$O 0 and $\mu$O 1's virtual communication service infrastructures, the experiment uses hypervisor to stimulate the scenario of adjacent $\mu$O communication. The infrastructure of the experiment consists of one Xen hypervisor and two Ubuntu 16.04 virtual machines. Mininet is used to virtualize S1 for VM1, add GRE interface, and set the remote IP as 10.10.10.126, which is the IP of VM2. VM2 has the same setting with the only difference that its remote IP is set as 10.10.10.116, which is the IP of VM1. Both VM1 and VM2's control plane are set as OpenVirtex and OVX will simultaneously finish logical network slicing when the slicing is built. As shown in Figure 8, the TCP bandwidth test is conducted by Iperf in mono-hypervisor. Iperf is a network benchmarking tool which is used to measure the throughput of the network carrying UDP and TCP data [43]. The average bandwidth of GRE tunnels is 1.65 Gbps and the average bandwidth of VxLAN tunnels is 1.63 Gbps. The results of the experiment are shown in Figure 9. As shown in Figure 10, logical network slicing infrastructure is built by the two hypervisors. Hypervisor 1 and Hypervisor 2 are separated physically. The results of the experiment are shown in Figure 11. In this case, the average bandwidth of GRE tunnels is 915 Mbps and the average bandwidth of VxLAN tunnels is 908 Mbps.

Judging from the results, we can see that even though tunnels between the two VMs in mono-hypervisor are successfully built, the packets are unable to actually leave hypervisor's network card but exchange memory instead. This is because both VMs are located in the same hypervisor. In the two hypervisors scenario, tunnels are successfully built as well and the packets actually reach each other's hypervisor network cards to carry out the exchange. This is attributed

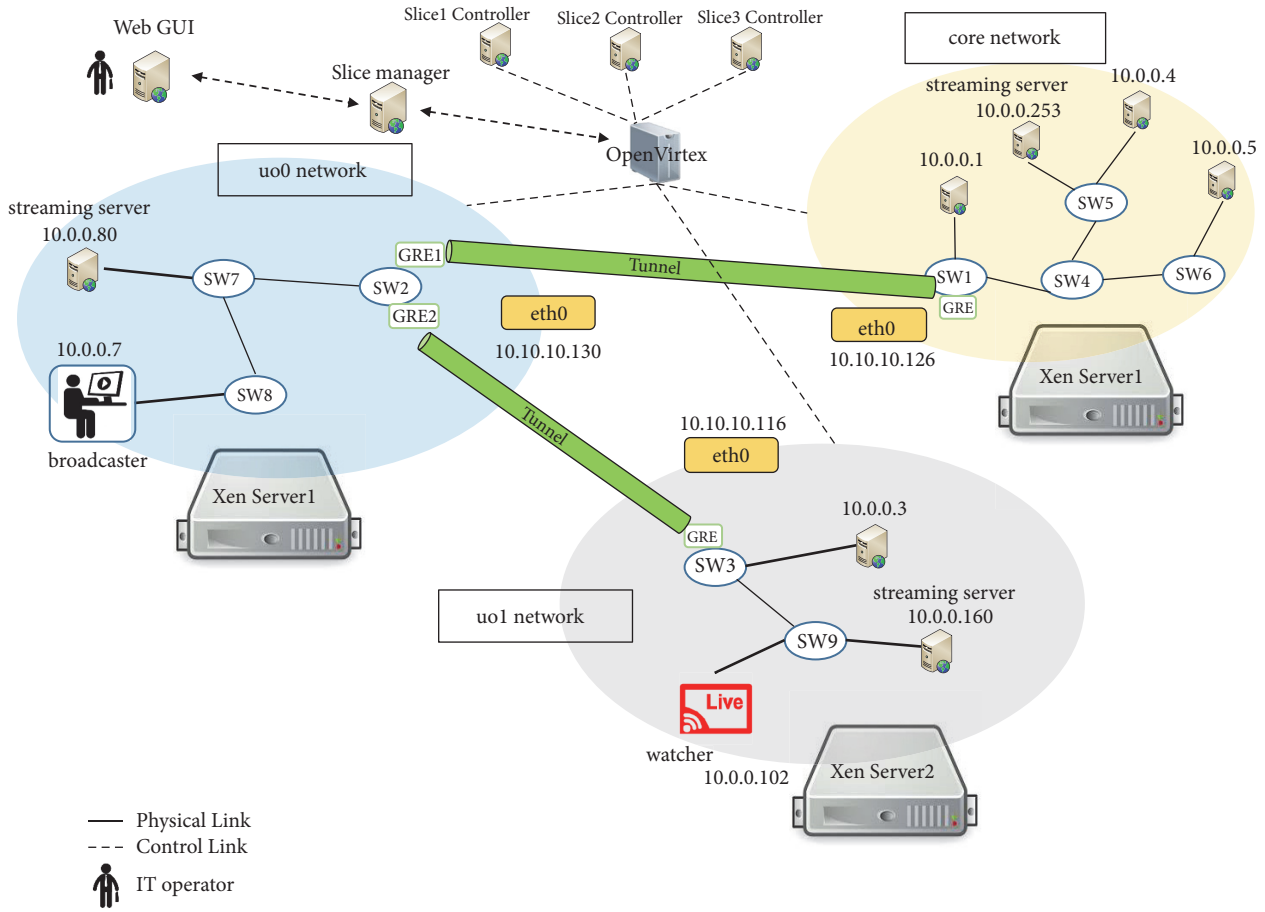



Figure 7: The experiment environment.

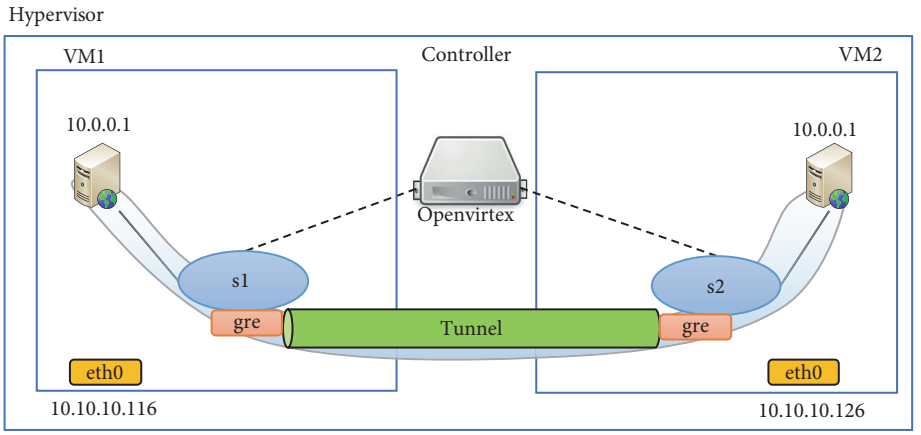



Figure 8: Average bandwidth test in a signal hypervisor scenario.

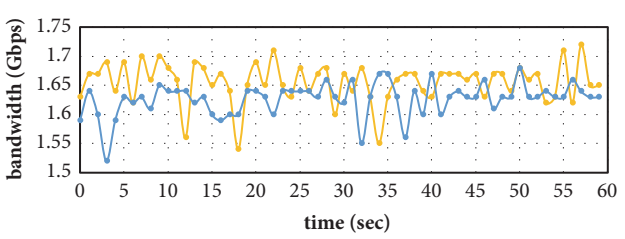



Figure 9: The experiment result in a signal hypervisor.

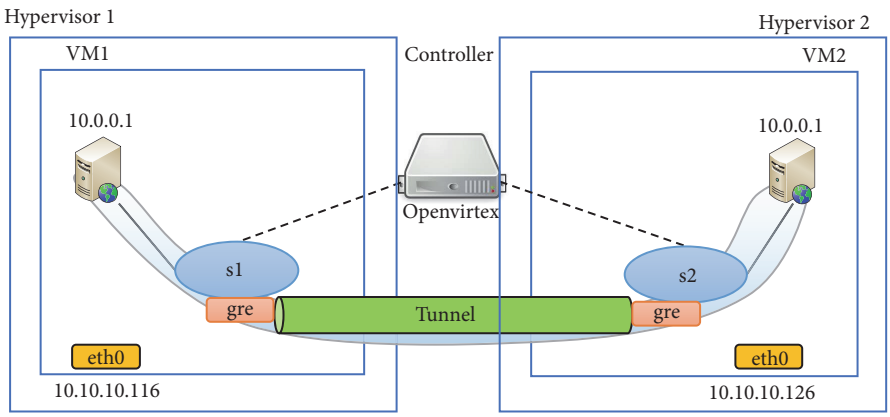



Figure 10: Average bandwidth test in two hypervisors scenario.

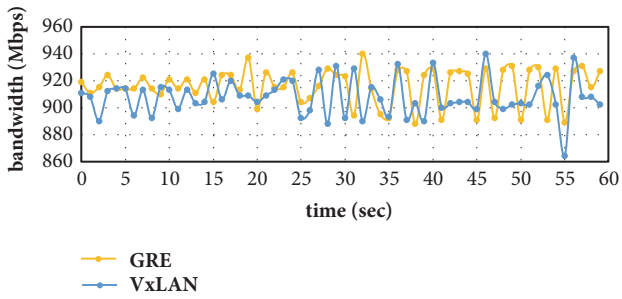



Figure 11: The experiment result in cross-hypervisors.

to that fact that a bridge of communication has been built by the tunnels and the packets are repackaged. It is therefore determined that the two hypervisors scenario requires less bandwidth than mono-hypervisor.

As for the scenario of nonadjacent $\mu$O communication, we replace the experiment infrastructure with two-tier mininet to simulate the whole environment and construct a virtual experiment infrastructure in different domains. As shown in Figure 12, in the first tier, we use the mininet to construct topology in different domains and connect domain 1 and domain 2 via the OpenVSwitch (OVS) of the mininet. Domain 1 and Domain 2 can also construct their own topology via the mininet. The topology of the mininet in the first tier uses OpenVirtex to conduct the network slicing for tenants. The slices are composed of tunnel and nontunnel, and the slices are subject to the management of different controllers. For example, the tunnel slice is subject to tunnel controller and so forth. The tunnel network slicing uses the exclusive network slices built by the tunneling technology. The experiment uses two mainstream tunneling technologies, GRE and VxLAN, to conduct the comparison. The experiment result is shown in Figure 13. The experiment sets the bandwidth of the link in the topology at 100Mbytes and the major purpose of the experiment is to test the transmission efficiency of the two tunneling technologies. GRE tunnel package belongs to the third tier routing; VxLAN uses UDP to pack the packet. The average bandwidth of GRE is around 73.7 Mbytes and the average bandwidth of VxLAN is around 34.8 Mbytes. The results clearly suggest that VxLAN's packet transmission efficiency is not even half of that of GRE's. The cause behind this is because of the different ways of packet transmission. UDP is by itself an unreliable way of transmission and the packet is even more likely to be lost with the OpenVirtex serving as the intermediary tier. This might explain the poor performance in bandwidth.

To evaluate the efficiency and the performance, the DTBFR mechanism proposed in this paper is compared with the Minimize Loading First (MLF) mechanism. The MLF is a greedy algorithm that can be used to find the server with the lowest load and direct the traffic to the server via SDN controller.

In order to make users of $\mu$O gain access to the nearby micro data center or link to the Internet data center for access to particular application of network services via tunnels, three indicators are available to determine the quality of the decision-making of traffic flow redirection. $\mu$O_num represents the number of accessed services in $\mu$O. Near_$\mu$O_num represents the number of times when the traffic flow has to access services in nearby $\mu$O because the servers in $\mu$O are all busy. Datacenter_num represents the number of times recorded when the traffic flow has to access remote data center because nearby $\mu$Os are all busy. The results of the experiment are shown in Figure 14. The value of $\mu$O_num of the proposed DTBFR mechanism is apparently higher than that of MLF mechanism. This means that DTBFR mechanism is able to prioritize the user's traffic flow according to the servers in $\mu$O's area. If the server is busy, DTBFR mechanism will consider accessing services in nearby $\mu$O and, because the number of HTTP requests varies, (11) is used to calculate the redirection ratio.

$$\text{Ratio} = \frac{(\mu\text{O\_num/near\_}\mu\text{O\_num/Datacenter\_num})}{\text{Total number of HTTP requests}} \quad (11)$$

Under the same experiment conditions, the decision-making of traffic flow direction of DTBFR mechanism is as follows: 68.1% of the traffic flow will be directed to the $\mu$O in the current area to access services; 28.7% of the traffic flow will be directed to the nearby $\mu$O to access services; only 3.2% of the traffic will go to the data center to access services.

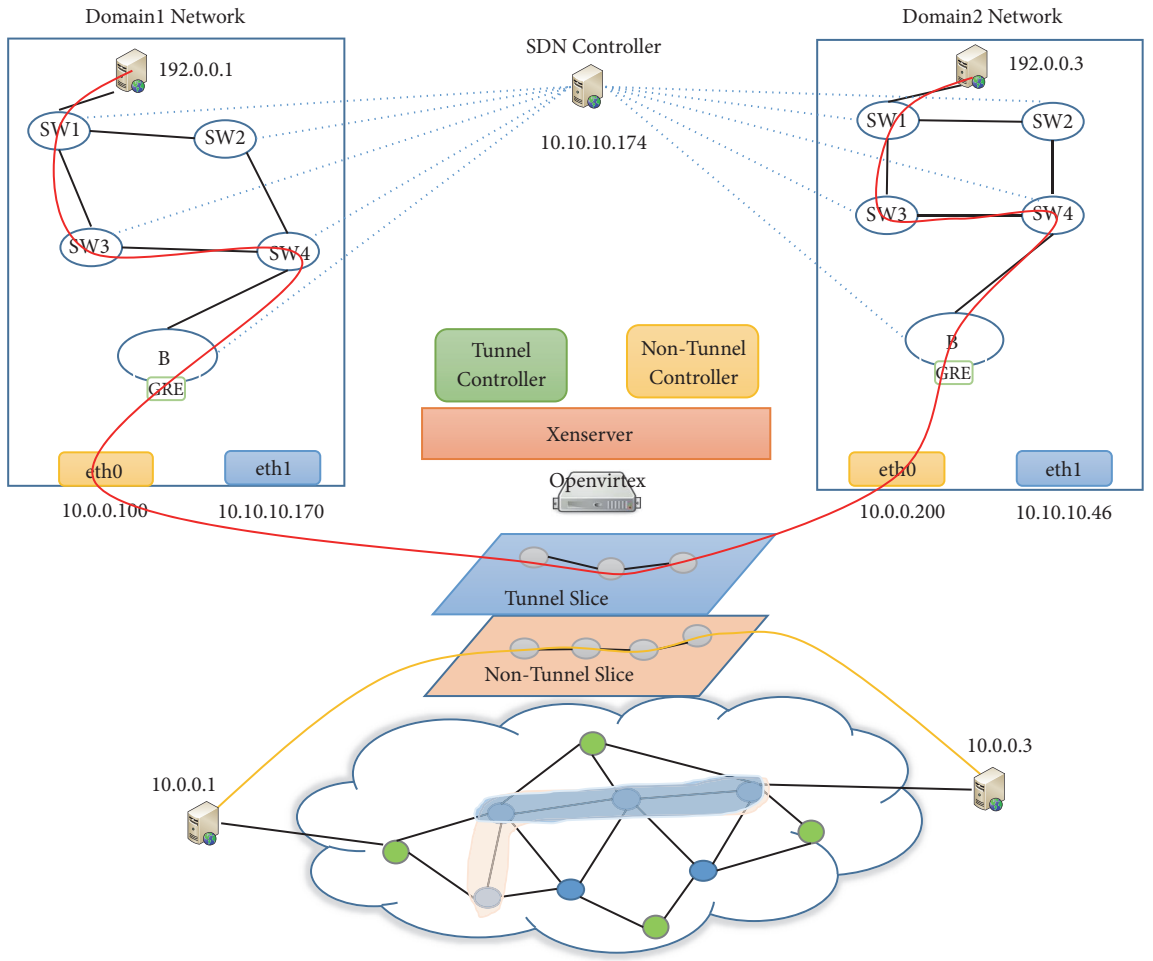



Figure 12: Nonadjacent $\mu$O communication scenario.

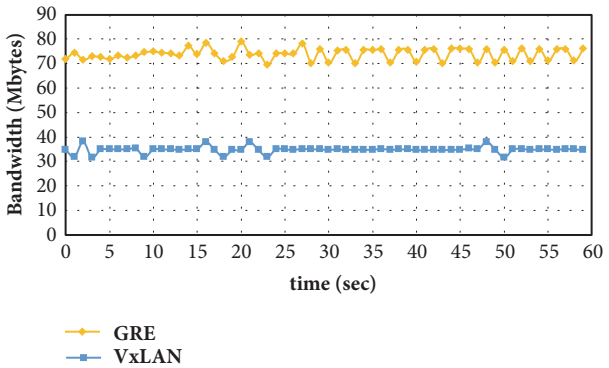



Figure 13: The comparison of GRE and VxLAN experiment result in nonadjacent $\mu$O communication scenario.

MLF algorithm is a directing method based on server load which directs 29.8% of the traffic flow to the $\mu$O in the current area to access services, 32.7% of the traffic flow to the nearby $\mu$O to access services, and 37.5% of the traffic flow to the Internet data center to access services. The $\mu$O's accessed services indicator in the area shows that DTBFR mechanism takes up 68.1% while MLF algorithm takes up only 29.8%. The experiment verifies that DTBFR mechanism's performance on traffic redirection is relatively good.

In summary, the design pattern of 5G micro operator combines network slicing and tunneling technologies to realize the communication model for $\mu$O, although network slicing holds much promise for 5G, but not without its share of hurdles. The traditional network architecture will need to be redesigned to enable network slicing. Interoperability should be tested in order to ensure network slicing works as expected in the 5G network. In terms of tunneling, the inefficient packet overhead may have a negative effect on the performance of the network. For the purpose of micro operator to providing regional services, the DTBFR mechanism allows different $\mu$Os to share network resources and improves data processing by directing the users' traffic flow to a closer $\mu$O for reducing the network transmission latency, building the foundations of ultrareliable and low-latency communications (URLLC) in 5G. Conclusively, the $\mu$O design pattern provided in this paper helps to reduce the computing burden of the traditional cloud network architecture, improves the load capacity of the cloud network and the servers, and is also able to be used as development basis for

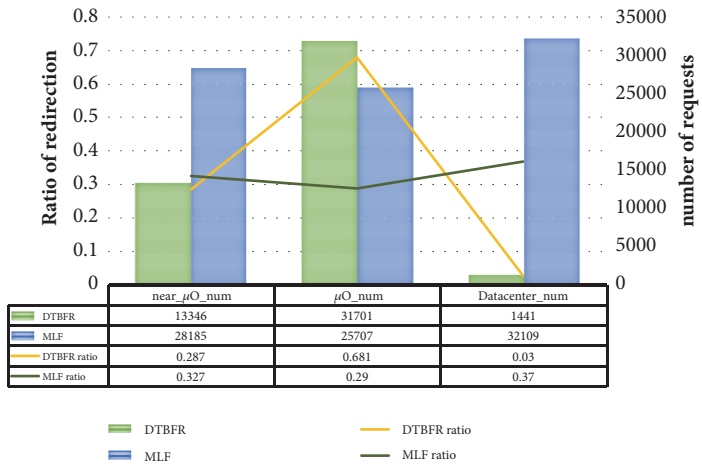



Figure 14: Comparison with experiment results of DTBFR and MLF.

integration and innovation application of the Cloudy-Edge Computing achieved by 5G.

## 5. Conclusions

This paper uses SDN and NFV technologies as the base and combines network slicing and tunneling technologies to come up with a network infrastructure pattern for $\mu$Os. The pattern allows users of different $\mu$Os to be concatenated via tunneling technology and then realizes the rapid connection of network to effectively enhance the interconnectivity of networks. In order to meet the demand of regional service of micro operator, this paper proposes DTBFR mechanism, which uses decision tree theory as the basis of SDN-based traffic decision-making. Under the same experiment conditions, the decision-making of traffic flow direction of DTBFR mechanism is as follows: 68.1% of the traffic flow will be directed to the $\mu$O in the current area to access services; 28.7% of the traffic flow will be directed to the nearby $\mu$O to access services; only 3.2% of the traffic will go to the Internet cloud data center to access services. The features deployed by the regional service of the micro operator can effectively reduce the burden of the data center on the Internet and accelerate the development of the edge computing service in the future 5G network.

Future work is needed to create an edge computing service model for the wide-spread adoption of the $\mu$O design pattern, especially the integration of the edge computing technologies into the 5G $\mu$O deployment path by improving management efficiency, and provides service on-demand application for micro operator customers. The resource estimation, tasking scheduling, and lightweight virtual network function configuration techniques are the primary research directions.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The work described in this paper was supported in part by the Ministry of Science and Technology of the Republic of China (Project nos. 104-2221-E-008 -039, 105-2221-E-008-071, 107-2623-E-008-002, and 107-2636-E-003-001).

## References

[1] X. Huang, Y. Li, S. Tang, and Q. Chen, "Coexistence of cognitive small cell and WiFi system: a traffic balancing dual-access resource allocation scheme," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 4092681, 17 pages, 2018.

[2] A. Raschellà, F. Bouhafs, G. C. Deepak, and M. Mackay, "QoS aware radio access technology selection framework in heterogeneous networks using SDN," *Journal of Communications and Networks*, vol. 19, no. 6, pp. 577–586, 2017.

[3] H. Beyranvand, M. Levesque, M. Maier, J. A. Salehi, C. Verikoukis, and D. Tipper, "Toward 5G: FiWi enhanced LTE-A hetnets with reliable low-latency fiber backhaul sharing and WiFi offloading," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 690–707, 2017.

[4] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN-Key technology enablers for 5G networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2468–2478, 2017.

[5] C. W. Tseng, P. H. Lai, B. S. Huang, L. D. Chou, and M. C. Wu, "NFV deployment strategies in SDN network," *International Journal of High Performance Computing and Networking (IJHPCN)*, 2018, Available: http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijhpcn.

[6] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, "Wireless network virtualization with SDN and C-RAN for 5G networks: requirements, opportunities, and challenges," *IEEE Access*, vol. 5, pp. 19099–19115, 2017.

[7] uO5G, "Micro operator concept for boosting local service delivery in 5G," Available: http://www.oulu.fi/uo5g/.

[8] M. Matinmikko, M. Latva-aho, P. Ahokangas, S. Yrjölä, and T. Koivumäki, "Micro operators to boost local service delivery in 5G," *Wireless Personal Communications*, vol. 95, no. 1, pp. 69–82, 2017.

[9] A. Prasad, Z. Li, S. Holtmanns, and M. A. Uusitalo, "5G micro-operator networks — A key enabler for new verticals and markets," in *Proceedings of the 2017 25th Telecommunication Forum (TELFOR)*, pp. 1–4, Belgrade, Serbia, November 2017.

[10] J. S. Walia, H. Hammainen, and M. Matinmikko, "5G Micro-operators for the future campus: A techno-economic study," in *Proceedings of the 2017 Internet of Things - Business Models, Users, and Networks*, pp. 1–8, Copenhagen, Denmark, November 2017.

[11] M. Matinmikko-Blue and M. Latva-aho, "Micro operators accelerating 5G deployment," in *Proceedings of the 2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pp. 1–5, Peradeniya, Sri Lanka, December 2017.

[12] K. B. Manosha, M. Matinmikko-Blue, and M. Latva-aho, "Framework for spectrum authorization elements and its application to 5G micro-operators," in *Proceedings of the 2017 Internet of Things - Business Models, Users, and Networks*, pp. 1–8, Copenhagen, Denmark, November 2017.

[13] T. Sanguanpuak, S. Guruacharya, E. Hossain, N. Rajatheva, and M. Latva-aho, "On spectrum sharing among micro-operators in 5G," in *Proceedings of the 2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–6, Oulu, Finland, June 2017.

[14] P. Ahokangas, S. Moqaddamerad, and M. Matinmikko, "Future micro operators business models in 5G," *The Business and Management Review*, vol. 7, no. 5, pp. 143–149, 2016.

[15] P. Mach and Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[16] G. Li, J. Song, J. Wu, and J. Wang, "Method of resource estimation based on QoS in edge computing," *Wireless Communications and Mobile Computing*, vol. 2018, 9 pages, 2018.

[17] M. A. Lema, A. Laya, T. Mahmoodi et al., "Business case and technology analysis for 5G low latency applications," *IEEE Access*, pp. 5917–5935, 2017.

[18] P. Kiss, A. Reale, C. J. Ferrari, and Z. Istenes, "Deployment of IoT applications on 5G edge," in *Proceedings of the 2018 IEEE International Conference on Future IoT Technologies (Future IoT)*, pp. 1–9, Eger, Hungary, January 2018.

[19] M. Matinmikko, A. Roivainen, M. Latva-aho, and K. Hiltunen, "Interference study of micro licensing for 5g micro operator small cell deployments," in *Cognitive Radio Oriented Wireless Networks*, vol. 228 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 264–275, Springer International Publishing, Switzerland, 2018.

[20] L. Falconetti, R. Karaki, and S. Corroy, "Practical energy-aware cell association for small cell deployment," *Wireless Communications and Mobile Computing*, vol. 16, no. 16, pp. 2436–2448, 2016.

[21] T. Zahid, X. Hei, We. Cheng, A. Ahmad, and P. Maruf, "On the tradeoff between performance and programmability for software defined WiFi networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 1083575, pp. 1–12, 2018.

[22] P. Ahokangas, M. Matinmikko, S. Yrjölä, and I. Atkova, "Disruptive revenue models for future micro operator driven mobile business ecosystem," in *Proceedings of The 24th Nordic Academy of Management Conference (NFF)*, pp. 23–25, Bodo, Norway, 2017.

[23] M. G. Kibria, G. P. Villardi, K. Nguyen, W.-S. Liao, K. Ishizu, and F. Kojima, "Shared spectrum access communications: a neutral host micro operator approach," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1741–1753, 2017.

[24] "Micro operators for vertical specific service delivery in 5G," Available: http://www.5gsummit.org/berlin/docs/slides/Matti-Latva-Aho.pdf/.

[25] A. Basta, A. Blenk, K. Hoffmann, H. J. Morper, M. Hoffmann, and W. Kellerer, "Towards a cost optimal design for a 5G mobile core network based on SDN and NFV," *IEEE Transactions on Network and Service Management*, vol. 14, no. 4, pp. 1061–1075, 2017.

[26] C. Bouras, A. Kollia, and A. Papazois, "SDN and NFV in 5G: Advancements and challenges," in *Proceedings of the 20th Conference on Innovations in Clouds, Internet and Networks, ICIN 2017*, pp. 107–111, France, March 2017.

[27] ONF, "SDN architecture," Available: https://www.opennetworking.org/images/stories/downloads/sdnresources/technical-reports/TRSDNARCH1.006062014.pdf.

[28] ETSI, "Network Functions Virtualization (NFV), architectural framework," ETSI Ind. Specification Group, Sophia Antipolis Cedex, France, 2014.

[29] Q. Jia, R. Xie, T. Huang, J. Liu, and Y. Liu, "Efficient caching resource allocation for network slicing in 5G core network," *IET Communications*, vol. 11, no. 18, pp. 2792–2799, 2017.

[30] X. Li, M. Samaka, H. A. Chan et al., "Network slicing for 5G: challenges and opportunities," *IEEE Internet Computing*, vol. 21, no. 5, pp. 20–27, 2017.

[31] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 32–39, 2016.

[32] NGMN Alliance, "5G White Paper," Available: https://www.ngmn.org/5g-white-paper.html.

[33] NOKIA, "Dynamic end-to-end network slicing for 5G White Paper," Available: http://www.hit.bme.hu/~jakab/edu/litr/5G/NOKIA_dynamic_network_slicing_WP.pdf.

[34] S. Min, S. Kim, J. Lee, B. Kim, W. Hong, and J. Kong, "Implementation of an OpenFlow network virtualization for multi-controller environment," in *Proceedings of the 14th International Conference on Advanced Communication Technology, ICACT 2012*, pp. 589–592, Republic of Korea, February 2012.

[35] A. Al-Shabibi, M. De Leenheer, M. Gerola et al., "OpenVirteX: Make your virtual SDNs programmable," in *Proceedings of the 3rd ACM SIGCOMM 2014 Workshop on Hot Topics in Software Defined Networking, HotSDN 2014*, pp. 25–30, Chicago, IL, USA, August 2014.

[36] E. Alpaydin, *Introduction to Machine Learning*, The MIT Press, Cambridge, Mass, London, England, 2nd edition, 2009.

[37] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree ID3 and C4.5," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, 2014, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications 2014.

[38] R. C. Barros, M. P. Basgalupp, A. C. P. L. F. De Carvalho, and A. A. Freitas, "A survey of evolutionary algorithms for decision-tree induction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 3, pp. 291–312, 2012.

[39] D. Lee and C. S. Hong, "Access point selection algorithm for providing optimal AP in SDN-based wireless network," in *Proceedings of the 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pp. 362–365, Seoul, South Korea, September 2017.

[40] S. T. V. Pasca, S. S. P. Kodali, and K. Kataoka, "AMPS: Application aware multipath flow routing using machine learning in SDN," in *Proceedings of the Twenty-third National Conference on Communications (NCC)*, pp. 1–6, Chennai, India, 2017.

[41] "Entropy," Available: https://en.wikipedia.org/wiki/Entropy_(information_theory).

[42] "Mininet," Available: http://mininet.org/.

[43] "iPerf," Available: https://iperf.fr/.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4079017, 17 pages
https://doi.org/10.1155/2018/4079017

*Research Article*

# Energy-Efficient Uplink Resource Units Scheduling for Ultra-Reliable Communications in NB-IoT Networks

**Jia-Ming Liang**[1,2] **Kun-Ru Wu**[3] **Jen-Jee Chen**[4] **Pei-Yi Liu**[3] **and Yu-Chee Tseng**[3,5]

[1]*Department of Computer Science and Information Engineering, Chang Gung University, Kweishan, Taoyuan 33378, Taiwan*
[2]*Department of General Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan*
[3]*Department of Computer Science, National Chiao Tung University, HsinChu 30010, Taiwan*
[4]*Department of Electrical Engineering, National University of Tainan, Tainan, Taiwan*
[5]*Research Center for Information Technology Innovation, Academia Sinica, Taipei 11574, Taiwan*

Correspondence should be addressed to Jen-Jee Chen; jjchen@mail.nutn.edu.tw

Received 1 March 2018; Revised 2 May 2018; Accepted 6 May 2018; Published 2 July 2018

Academic Editor: Shao-Yu Lien



For 5G wireless communications, the *3GPP Narrowband Internet of Things (NB-IoT)* is one of the most promising technologies, which provides multiple types of *resource unit (RU)* with a special repetition mechanism to improve the scheduling flexibility and enhance the coverage and transmission reliability. Besides, NB-IoT supports different operation modes to reuse the spectrum of LTE and GSM, which can make use of bandwidth more efficiently. The IoT application grows rapidly; however, those massive IoT devices need to operate for a very long time. Thus, the energy consumption becomes a critical issue. Therefore, NB-IoT provides discontinuous reception operation to save devices' energy. But, how to further reduce the transmission energy while ensuring the required ultra-reliability is still an open issue. In this paper, we study how to guarantee the reliable communication and satisfy the *quality of service (QoS)* while minimizing the energy consumption for IoT devices. We first model the problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient, ultra-reliable, and low-complexity scheme, which consists of two phases. The first phase tries to optimize the default transmit configurations of devices which incur the lowest energy consumption and satisfy the QoS requirement. The second phase leverages a weighting strategy to balance the emergency and inflexibility for determining the scheduling order to ensure the delay constraint while maintaining energy efficiency. Extensive simulation results show that our scheme can serve more devices with guaranteed QoS while saving their energy effectively.

## 1. Introduction

The *Internet of Things (IoT)* is one of the key applications and technologies in the *fifth-generation (5G)* communications. Since IoT is widely used for remote monitoring and reporting, such as smart building, smart transportation, smart grid, e-health, and/or factory automation, it makes our life more convenient and makes industry more efficient. Therefore, the *3rd generation partnership project (3GPP)* develops a new technology, *Narrowband Internet of Things (NB-IoT)* [1, 2], as the communication standard for IoT, which is featured by low cost, low energy consumption, low complexity, and low throughput. Besides, the NB-IoT supports massive connectivity and enhances the benefit of spectrum reuse. Specifically, it supports multiple types of *resource units (RU)* with specific repetitions for data transmission to improve the scheduling flexibility and enhance communication reliability. In addition, NB-IoT also provides three-operation modes to flexibly reuse the spectrum of LTE and GSM, which can achieve higher spectrum utilization and reduce the extra deployment cost for the operators.

On the other hand, due to the inherent behaviors of IoT applications, such as remote monitoring and reporting, IoT devices need to operate for a very long time [3]. Thus, energy consumption becomes a key issue. Currently, NB-IoT provides discontinuous reception to save devices' energy based on wake-up and sleep operation. However, how to further decrease their transmission energy during wake-up period is an open problem. In addition, the reliability of transmission is also a key issue in QoS for uplink transmission

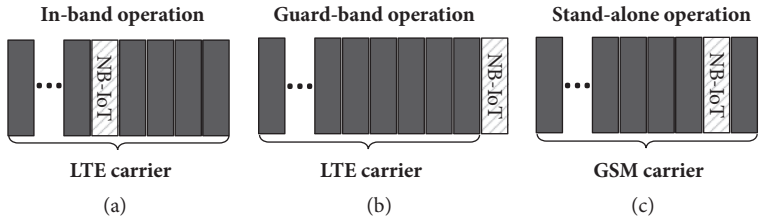



Figure 1: Three-operation modes of NB-IoT.

especially for mission critical applications (e.g., e-health, intelligent transport), voice applications, and the high timing precision factory automation [4–6]. Thus, NB-IoT provides the maximum repetition time up to 128 for a scheduling resource allocation to the ultra-reliability issue. But, how to optimize the repetition operation to achieve high reliability while reducing the waste of resource and energy is still an open issue.

In this paper, we study how to ensure the high transmission reliability to guarantee the strict QoS for devices based on the RU scheduling and repetition determination while minimizing their energy consumption in NB-IoT networks. We first model the problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient and ultra-reliable heuristic, which consists of two phases. The first phase tries to select the primary parameters which conduct the lowest energy consumption and ensure QoS requirements for uplink transmission. The second phase applies a weighting strategy to determine the precise scheduling order of uplink requests based on the scheduling emergency and inflexibility. In addition, it also adjusts the corresponding results appropriately to satisfy the strict delay constraint if needed while considering energy efficiency. Extensive simulation results show that our scheme can enlarge the number of serving devices with guaranteed QoS and decrease the packet drop ratio while saving energy.

The rest of this paper is organized as follows. Related work is discussed in Section 2. Preliminaries are given in Section 3. Section 4 presents our scheme. Simulation results are shown in Section 5. Section 6 concludes the paper.

## 2. Related Work

In the literature, the studies [7, 9–11] give an overview of NB-IoT and conclude that NB-IoT can enhance bandwidth efficiency and increase network coverage. Reference [12] proposes a new procedure for cell search and initial synchronization in NB-IoT which can speed up the access operation for the devices with low SNR. In [13], it proposes a new channel equalization algorithm to optimize the sampling rate of devices when NB-IoT and LTE share the same spectrum. However, they neglect the QoS and reliability of transmissions. The research [14] leverages the *modulation and coding scheme (MCS)* and repetition number to enhance the QoS satisfaction and transmission latency. However, it does not leverage different types of RUs; thus, it will reduce the service coverage of NB-IoT and cannot allocate resource flexibly and effectively. In [15], the authors develop a new detection mechanism for random access procedure to enhance the coverage and access efficiency of NB-IoT. However, it does not consider the transmission reliability and energy efficiency. The study [16] proposes a transmission scheme without connection setup to reduce connectivity latency. But, it may cause extra energy consumption. Reference [17] develops a detection scheme based on maximum likelihood to detect timing acquisition with low delay while reducing energy consumption. However, the QoS satisfaction and the transmission reliability are ignored in this paper. In [18], the authors develop a prediction method to allocate resource in advance based on the occurrence and delay time of uplink transmission. Although it can accelerate the transmission procedure, it may decrease the scheduling flexibility and resource efficiency. In [19], it leverages *Nonorthogonal Multiple Access (NOMA)* to allocate common subcarriers to multiple devices and thus to enhance the spectrum efficiency. However, it does not discuss how to ensure the energy efficiency and transmission reliability, which are the key issues in NB-IoT.

Based on the above observation, it motivates us to address the issue of considering both transmission reliability and energy efficiency by scheduling multitypes of RUs with optimal repetition in NB-IoT networks.

## 3. Preliminary

In this section, we first give an overview of the operation modes of NB-IoT. Then, we introduce the resource unit and the repetition mechanism used in NB-IoT. Finally, we formally define our resource allocation problem and show it to be NP-complete.

*3.1. NB-IoT Operation Modes.* In NB-IoT, all devices connect with the centralized base station (also called the *Evolved Node B, eNB*). In order to enhance the spectrum utilization and reduce the cost of operators, NB-IoT provides three-operation modes for devices to access the eNB by reusing the existing spectrum of LTE and GSM [20–22]:

*3.1.1. Inband.* Using the bandwidth of one *resource block (RB)* inside the LTE carrier as the access spectrum is shown in Figure 1(a).

*3.1.2. Guard-Band.* Using the bandwidth of one RB in the guard-band of LTE carrier as the access spectrum is shown in Figure 1(b).

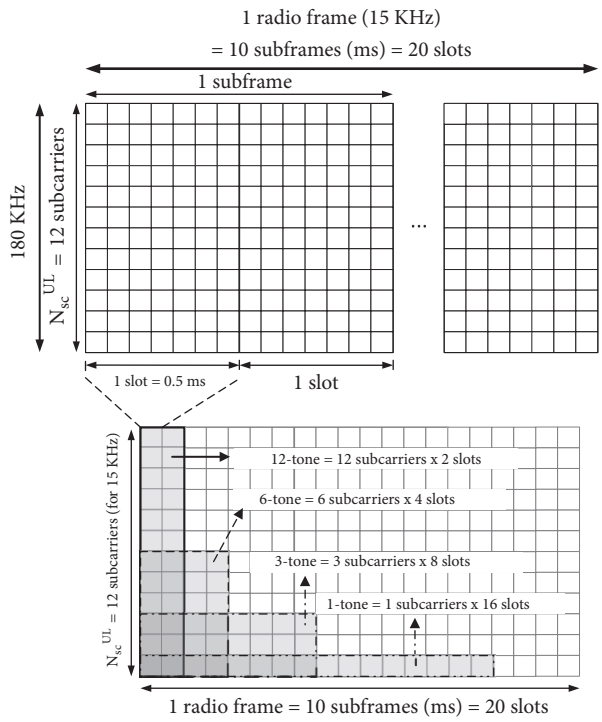



Figure 2: Multiple types of RUs.

Table 1: The types of resource units (RUs) supported in NB-IoT.

| Subcarrier spacing | Number of tones (subcarriers) | Classification | Number of slots |
|---|---|---|---|
| 15 KHz | 1 | single-tone | 16 |
| 15 KHz | 3 | multi-tone | 8 |
| | 6 | | 4 |
| | 12 | | 2 |
| 3.75 KHz | 1 | single-tone | 16 |

*3.1.3. Stand-Alone.* Using the bandwidth of a reframed GSM carrier as the access spectrum is shown in Figure 1(c).

*3.2. Resource Unit (RU).* In NB-IoT, the resource is divided into *frames*, where each frame consists of 10 *subframes*. The length of a subframe is 1 ms, which is further divided into two *slots*. For the uplink transmission, data is transmitted through *Narrowband Uplink Shared Channel (NPUSCH)*. The *resource unit (RU)* is the basic transmission resource unit allocated in the bandwidth of 180 KHz. The transmission data can be carried by one or multiple RUs depending on the request size and MCS. Specifically, NB-IoT supports multiple types of RUs based on the subcarrier spacing as shown in Table 1. Since the subcarrier spacing of 15 KHz is mandatory in the standard, we focus on it in this paper [1]. For the subcarrier spacing of 15 KHz, there are 4 types of RUs that are classified as *single-tone* (1-tone) or *multitone* (3-tone, 6-tone, and 12-tone). Each type of RU is with a specific number of subcarriers and time slots, as shown in Figure 2.

*3.3. Repetition Mechanism.* In NB-IoT, one of the key features is the repetition mechanism, which is designed to enhance the reliability of transmission and enlarge the network coverage. According to the NB-IoT standard, the transmission RUs of each device $i$ (also called *user equipment* $i$ ($\mathrm{UE}_i$) in the following) can associate with a specific number of repetition $N_i^{\mathrm{rep}} = 2^l$, where $l \in \{0 \cdots 7\}$. Thus, each UE can ensure the transmission reliability based on the individual physical status, such as channel quality, path loss, bit-error-rate (BER), and transmission power. Figure 3 shows an example of 5 UEs with different repetition numbers. Note that the transmission finish time of each UE should not violate its delay constraint.

*3.4. Downlink Control Information (DCI).* *Downlink Control Information (DCI)* is the control message in *Narrowband Physical Downlink Control Channel (NPDCCH)*, which is responsible for describing the scheduling results for both downlink and uplink transmissions. Each DCI is with the length of 1 ms. When the eNB completes the RU scheduling, each scheduling result will be carried by one DCI to inform the corresponding UE about its uplink transmission with the designate RU type, subcarrier set, allocation time, and the number of repetitions. Table 2 shows the main parameters

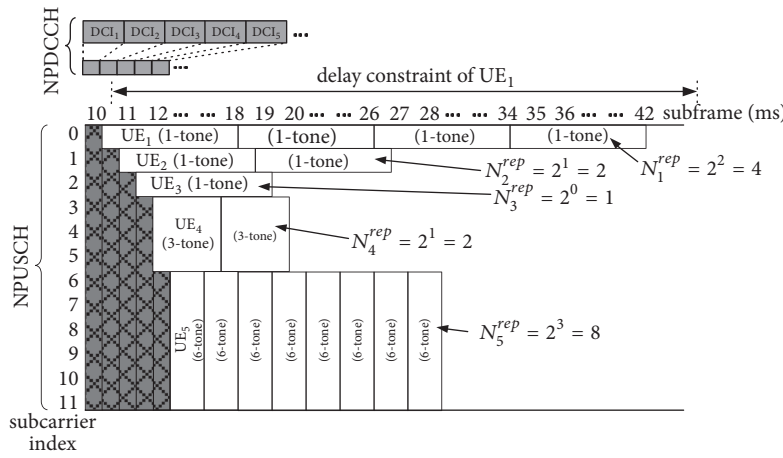



Figure 3: The example of repetitive transmissions.

Table 2: Main parameters of DCI (format N0).

| Parameter | Value |
|---|---|
| subcarrier indication ($I_i^{\mathrm{sc}}$) | 0~63 |
| resource assignment ($N_i^{\mathrm{RU}}$) | 0~7 |
| modulation and coding scheme ($\mathrm{MCS}_i$) | 0~10 |
| repetition number ($N_i^{\mathrm{rep}}$) | $2^l$, $l \in \{0 \cdots 7\}$ |

Table 3: Subcarrier indication and the corresponding subcarrier sets.

| Subcarrier indication ($I_i^{\mathrm{sc}}$) | Set of Allocated subcarriers ($S_i^{\mathrm{sc}}$) |
|---|---|
| 0–11 | $I_i^{\mathrm{sc}}$ |
| 12–15 | $3(I_i^{\mathrm{sc}} - 12) + \{0, 1, 2\}$ |
| 16–17 | $6(I_i^{\mathrm{sc}} - 16) + \{0, 1, 2, 3, 4, 5\}$ |
| 18 | $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ |
| 19–63 | reserved |

in DCI (format N0), which is for *uplink grant* and *scheduling* in NPUSCH. Specifically, the *subcarrier indication* ($I_i^{\mathrm{sc}}$) describes the RU type and the corresponding *subcarrier set* to locate RUs. The *resource assignment* ($N_i^{\mathrm{RU}}$) represents the number of allocated continuous RUs for this transmission schedule excluding repetition. The *modulation and coding scheme* ($\mathrm{MCS}_i$) means which MCS is applied on this RU transmission. Note that NB-IoT supports 11 types of modulation and coding schemes for uplink, which depend on the bit-error-rate and received signal-to-noise ratio (this will be clear later on). The *repetition number* ($N_i^{\mathrm{rep}}$) represents the number of repetitions for the scheduled RUs. So, the total amount of RUs assigned to $\mathrm{UE}_i$ is $N_i^{\mathrm{RU}} \times N_i^{\mathrm{rep}}$.

Specifically, subcarrier indication ($I_i^{\mathrm{sc}} \in \{0\sim63\}$) is used for the description of RU types and their subcarrier set, as shown in Table 3. When the subcarrier spacing is 15 KHz, $I_i^{\mathrm{sc}} \in \{0\sim11\}$ represents the fact that the RU type is single-tone and locates at the subcarrier set of $S_i^{\mathrm{sc}} = I_i^{\mathrm{sc}}$. Thus, it has 12 possible locations. When $I_i^{\mathrm{sc}} \in \{12\sim15\}$, the RU type is 3-tone and locates at $S_i^{\mathrm{sc}} = 3(I_i^{\mathrm{sc}} - 12) + \{0, 1, 2\}$, which has 4 possible locations, i.e., $S_i^{\mathrm{sc}} \in \{\{0, 1, 2\}, \{3, 4, 5\}, \{6, 7, 8\}, \{9, 10, 11\}\}$. When $I_i^{\mathrm{sc}} \in \{16\sim17\}$, it indicates the RU type of 6-tone, which has 2 possible locations, i.e., $S_i^{\mathrm{sc}} \in \{\{0, 1, 2, 3, 4, 5\}, \{6, 7, 8, 9, 10, 11\}\}$. Finally, when $I_i^{\mathrm{sc}} = 18$, the RU type is 12-tone, which has a unique location, i.e., $S_i^{\mathrm{sc}} = \{\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}\}$ and thus $|S_i^{\mathrm{sc}}| = 1$.

*3.5. Problem Definition.* In this paper, we consider an NB-IoT network with a base station (eNB) serving $N$ UEs. Each $\mathrm{UE}_i$, $i = 1 \cdots N$, has an uplink request with data size $D_i \geq 0$ (bits), required reliability $R_i \in [0, 1]$, and strict delay constraint $d_i$ (ms). To guarantee QoS, assume that the arrival time of the $\mathrm{UE}_i$'s request is at $T_i^{\mathrm{req}}$th (ms) and then the data must be uploaded to the base station before the delay deadline $T_i^{\mathrm{req}} + d_i$. For each $\mathrm{UE}_i$, the transmit power is denoted as $P_i$ (mW) which is constrained by the maximum transmit power $P_i^{\max}$; i.e.,

$$0 \leq P_i \leq P_i^{\max}. \tag{1}$$

When scheduling, each $\mathrm{UE}_i$ has to be assigned one type of RUs; $N_i^{\mathrm{sc}} \in \{1, 3, 6, 12\}$; according to the designate subcarrier indication $I_i^{\mathrm{sc}} \in \{0\sim18\}$; i.e.,

$$N_i^{\mathrm{sc}} = \begin{cases} 1, & \text{if } 0 \leq I_i^{\mathrm{sc}} \leq 11 \\ 3, & \text{if } 12 \leq I_i^{\mathrm{sc}} \leq 15 \\ 6, & \text{if } 16 \leq I_i^{\mathrm{sc}} \leq 17 \\ 12, & \text{if } I_i^{\mathrm{sc}} = 18. \end{cases} \tag{2}$$

For each $\mathrm{UE}_i$'s RUs, the amount of data that $\mathrm{UE}_i$ can carry depends on the modulation and coding scheme $\mathrm{MCS}_i \in \{0 \cdots 10\}$. Specifically, the bit-error-rate of the data received by the base station relies on the *received signal-to-noise ratio* $\mathrm{SNR}_{\mathrm{dB}}(i)$; i.e.,

$$\begin{aligned} \mathrm{SNR}_{\mathrm{dB}}(i) &= 10 \log_{10} \left( \frac{\widetilde{P}(P_i)/N_i^{\mathrm{sc}}}{BN_0 + I} \right) \\ &\geq \mathrm{SNR}_{\mathrm{dB}}^{\mathrm{Req}}(\mathrm{MCS}_i, \mathrm{BER}_i), \end{aligned} \tag{3}$$

where $\widetilde{\mathbf{P}}(P_i) = G_i G_{\text{eNB}} P_i / L(i, \text{eNB})$ is the received power at base station and $G_i$, $G_{\text{eNB}}$, and $L(i, \text{eNB})$ are the transmitter gain, receiver gain, and the path loss between $\text{UE}_i$ and the eNB, respectively; $B$ is the subcarrier bandwidth; i.e., 15 KHz, $N_0$ is the noise power and $I$ is the interference perceived at the eNB. Note that

$\text{SNR}_{\text{dB}}^{\text{Req}}(\text{MCS}_i, \text{BER}_i)$ is the SNR threshold to apply $\text{MCS}_i$ with the measured *bit-error-rate* ($\text{BER}_i$).

According to Table 1, the number of required RUs ($N_i^{\text{RU}}$) for each $\text{UE}_i$ is

$$N_i^{\text{RU}} = \begin{cases} \left\lceil \dfrac{D_i}{r(\text{MCS}_i) \times 16} \right\rceil, & \text{if } N_i^{\text{sc}} = 1 \\ \left\lceil \dfrac{D_i}{r(\text{MCS}_i) \times 24} \right\rceil, & \text{otherwise,} \end{cases} \tag{4}$$

where $r(\text{MCS}_i)$ is the data rate of $\text{MCS}_i$ (bits per subcarrier $\times$ slot). To guarantee the transmission reliability $R_i$, we have to leverage the number of repetitions $N_i^{\text{rep}}$ and the successful probability of data transmission $P_i^s$; i.e.,

$$1 - \left(1 - P_i^s\right)^{N_i^{\text{rep}}} \geq R_i, \tag{5}$$

where $P_i^s = (1 - \text{BER}_i)^{D_i}$ is the successful probability [23, 24] if data $D_i$ is transmitted one time and $1 - (1 - P_i^s)^{N_i^{\text{rep}}}$ is the successful probability after $N_i^{\text{rep}}$ repetitions. Thus, to ensure the reliability requirement $R_i$ of $D_i$, Equation (5) is the necessary requirement.

Note that the scheduling results will be carried by the DCI message, which is scheduled at $T_i^{\text{DCI}}$ (subframe) for each $\text{UE}_i$. Thus, it has to satisfy the delay deadline of $\text{UE}_i$; i.e.,

$$T_i^{\text{DCI}} + \left( N_i^{\text{RU}} \times \frac{N_i^{\text{slot}}}{2} \times N_i^{\text{rep}} \right) \leq T_i^{\text{req}} + d_i, \tag{6}$$

where $N_i^{\text{slot}}$ is the number of slots of single RU (two slots constitutes one ms), which depends on the RU type; i.e.,

$$N_i^{\text{slot}} = \begin{cases} 16, & \text{if } N_i^{\text{sc}} = 1 \\ 8, & \text{if } N_i^{\text{sc}} = 3 \\ 4, & \text{if } N_i^{\text{sc}} = 6 \\ 2, & \text{if } N_i^{\text{sc}} = 12. \end{cases} \tag{7}$$

Now, we consider the current scheduling subframe is $T^s$ (ms), the feasible subcarrier set is $K$ (e.g., $|K| = 12$ if subcarrier spacing is 15 KHz), and the available earliest subframe for each subcarrier $k$ to allocate resource to devices is $S_k$, $k = 1 \cdots |K|$. Our problem asks how to optimize the uplink scheduling results for each $\text{UE}_i$, $i = 1 \cdots N$, by determining (1) the type of RUs ($N_i^{\text{sc}}$), (2) the number of RUs ($N_i^{\text{RU}}$), (3) the subcarrier set of RUs ($S_i^{\text{sc}}$), and (4) the allocation start time of RUs ($T_i^{\text{sc}}$) with (5) the number of repetitions ($N_i^{\text{rep}}$), (6) transmit power of $\text{UE}_i$ ($P_i$), (7) the modulation and coding scheme ($\text{MCS}_i$), and (8) the subframe index of DCI ($T_i^{\text{DCI}}$) to ensure that no two RUs overlap with each other and the QoS parameters including the request data size ($D_i$), delay constraint ($d_i$), and reliability of transmission ($R_i$) are satisfied while the total energy consumption of UEs, denoted as $\sum_{i=1 \cdots N} E_i$, is minimized, where

$$E_i = P_i \times \left( N_i^{\text{RU}} \times N_i^{\text{slot}} \times N_i^{\text{rep}} \right). \tag{8}$$

Therefore, our problem can be summarized as an optimization-like problem:

$$\min_{P_i, \text{MCS}_i, N_i^{\text{RU}}, I_i^{\text{sc}}, N_i^{\text{rep}}} \sum_{i=1 \cdots N} E_i, \tag{9}$$

subject to (1), (2), (3), (4) (5), (6), and (7).

Table 4 also summarizes the notations used in this paper.

**Theorem 1.** *The addressed problem is NP-complete.*

*Proof.* To simplify the proof, we consider the case of subcarrier spacing with 3.75 KHz where the UEs use the single-tone only and the modulation and coding scheme (MCS) is monotonic. So, the number of repetitions with minimal transmission power to meet required reliability of each UE is unique. Thus, the *energy cost* of an UE on each parameter list is also uniquely determined. Then, we formulate the resource allocation problem as a decision problem: *energy-efficient ultra-reliable scheduling decision (EUSD)* problem. Given a NB-IoT network and the UEs with required reliability for uplink transmission, we ask whether or not there exists a set of numbers of repetitions $S^{\text{Rep}}$ such that all UEs can conserve the total energy of $\boldsymbol{Q}$ to satisfy their uplink transmission with reliability. Then, we show that EUSD problem is NP-complete.

We first show that the EUSD problem belongs to NP. Given a problem instance and a solution containing the set of repetition numbers it can be verified whether or not the solution is valid in polynomial time. Thus, this part is proved.

We then reduce the *multiple-choice knapsack (MCK)* problem [25], which is known to be NP-complete, to the EUSD problem. Consider that there are $\boldsymbol{n}$ disjointed classes of objects, where each class $\boldsymbol{i}$ contains $\mathbf{N_i}$ objects. In each class $\boldsymbol{i}$, every object $\mathbf{x_{i,j}}$ has a profit $\mathbf{q_{i,j}}$ and a weight $\mathbf{u_{i,j}}$. Besides, there is a knapsack with capacity of $\boldsymbol{U}$. The MCK problem asks whether or not we can select exact one object from each class such that the total object weight is no larger than $\boldsymbol{U}$ and the total object profit is $\boldsymbol{Q}$.

We then construct an instance of the EUSD problem as follows. Let $\boldsymbol{n}$ be the number of UEs. Each $\text{UE}_i$ has $\mathbf{N_i}$ repetition selections to transmit data to the eNB. When $\text{UE}_i$ selects the repetition number $\mathbf{x_{i,j}}$, it will conserve energy of $\mathbf{q_{i,j}}$. Note that the conserved energy of an UE with a particular number of repetition is compared to the same UE's number with the most energy cost. Thus, the system should allocate a RU size of $\mathbf{u_{i,j}}$ to transmit $\text{UE}_i$'s data to the eNB. The total frame space is $\boldsymbol{U}$. Our goal is to let all UEs conserve energy of $\boldsymbol{Q}$ and satisfy their transmission requirement. We show that the MCK problem has a solution if and only if the EUSD problem has a solution.

Suppose that we have a solution to the EUSD problem, which is a set of repetition parameters $S^{\text{Rep}}$ with the conserved

Table 4: Summary of notations.

| notation | definition |
|---|---|
| $N_i^{\text{rep}}$ | number of repetitions of $\text{UE}_i$ |
| $I_i^{\text{sc}}$ | subcarrier indication of $\text{UE}_i$ |
| $N_i^{\text{RU}}$ | number of RUs of $\text{UE}_i$ |
| $\text{MCS}_i$ | modulation and coding scheme of $\text{UE}_i$ |
| $S_i^{\text{sc}}$ | set of allocated subcarriers of $\text{UE}_i$ |
| $D_i$ | uplink request of $\text{UE}_i$ (bits) |
| $R_i$ | required transmission reliability of $\text{UE}_i$ |
| $d_i$ | delay constraint of $\text{UE}_i$ (ms) |
| $T_i^{\text{req}}$ | data arrival time of $\text{UE}_i$ (ms) |
| $P_i$ | transmit power of $\text{UE}_i$ |
| $P_i^{\max}$ | maximum transmit power of $\text{UE}_i$ |
| $N_i^{\text{sc}}$ | RU type of $\text{UE}_i$ |
| $\text{SNR}_{\text{dB}}(i)$ | received signal-to-noise ratio of $\text{UE}_i$ (dB) |
| $\widetilde{P}$ | received power |
| $G_i$ | transmitter gain of $\text{UE}_i$ |
| $G_{\text{eNB}}$ | receiver gain of $\text{UE}_i$ |
| $L(i, \text{eNB})$ | path loss between $\text{UE}_i$ and the eNB |
| $B$ | subcarrier bandwidth of the NB-IoT (Hz) |
| $N_o$ | noise power |
| $I$ | interference perceived at the eNB |
| $\text{BER}_i$ | bit-error-rate of $\text{UE}_i$ |
| $\text{SNR}_{\text{dB}}^{\text{Req}}(\cdot)$ | SNR threshold (dB) |
| $r(\text{MCS}_i)$ | data rate of $\text{MCS}_i$ (bits per subcarrier × slot) |
| $P_i^s$ | successful probability of data transmission of $\text{UE}_i$ |
| $T_i^{\text{DCI}}$ | DCI subframe index of $\text{UE}_i$ (ms) |
| $T^s$ | index of current scheduling subframe (ms) |
| $K$ | feasible subcarrier set |
| $S_k$ | index of earliest available subframe of subcarrier $k$ |
| $T_i^{\text{sc}}$ | allocation start time of RU of $\text{UE}_i$ (ms) |
| $N_i^{\text{slot}}$ | number of slots for a single RU of $\text{UE}_i$ |
| **notations of the proposed scheme** | **definition** |
| $A_i$ | feasible setting pairs of RU type and MCS of $\text{UE}_i$ |
| $\text{Score}_i$ | score value of $\text{UE}_i$ |
| $W_1, W_2$ | weighting factors |
| $Em_i$ | urgent level of $\text{UE}_i$'s request |
| $T_j^R$ | remaining time of $\text{UE}_j$'s request from the scheduling subframe $T^s$ (ms) to the delay deadline |
| $\text{Waste}(i, S_i^{\text{sc}})$ | potential waste of $\text{UE}_i$ with its allocated subcarrier set |
| $\widehat{S}_k$ | earliest available allocation subframe (ms) of RUs for subcarrier $k$ |
| $S_i^{\text{sc}^*}$ | best subcarrier of $\text{UE}_i$ |
| $C_i^{\alpha,\beta}$ | cost ratio of $\text{UE}_i$ |
| $\Psi_{N^{\text{sc}}}()$ | number of choices of RU types |

energy of RUs. Each UE can choose exact one repetition number and we need to assign each number to each UE to satisfy their transmission reliability. The total size of RUs cannot exceed the frame space $\boldsymbol{U}$ and the conserved energy of all UEs is $\boldsymbol{Q}$. By viewing the possible number of repetitions of an UE as a class of objects and the frame as the knapsack, the repetition numbers in $S^{\text{Rep}}$ all constitute a solution to the MCK problem. This proves the *if* part.

Conversely, let $\mathbf{x_{1,\alpha 1}}, \mathbf{x_{2,\alpha 2}}, \ldots$ be a solution to the MCK problem. Then, for each $\text{UE}_i$, we select a repetition number such that $\text{UE}_i$ conserves energy of $\mathbf{q_{i,\alpha i}}$ and the size of allocated RUs to transmit $\text{UE}_i$'s data to the eNB is $\mathbf{u_{i,\alpha i}}$. In this way, the conserved energy of all UEs will be $\boldsymbol{Q}$ and the overall RU size is no larger than $\boldsymbol{U}$. This constitutes a solution to the EUSD problem, thus proving the *only if* part. □

## 4. The Proposed Scheme

Since the EUSD problem is NP-complete, finding the optimal solution is impractical due to the time complexity. Thus, we

propose a low-complexity, energy-efficient, and high-reliable scheme to tackle this problem. This scheme consists of two phases. The first phase exploits the strategy of "*minimal energy cost*" to determine the scheduling parameters of UEs. This scheme first quantifies the consumed energy for each UE and then chooses the one with minimal energy cost and reserves the corresponding parameters as the default transmit setting while satisfying the required reliability. The second phase determines the scheduling order based on the "*score function*" which considers the emergency level of requests and inflexibility of the scheduling transmission. Then, it determines the best resource location of RUs of each UE based on the "*potential resource waste*" function to enhance the resource utilization. Finally, if the predetermined results violate an UE's delay requirements on scheduling, the scheme will calculate a "*cost ratio*" to adaptively adjust the RU assignments. The details of the scheme are described as follows.

*4.1. Phase 1. Minimal Energy Cost.* The goal of the first phase is to determine the default parameters for each UE, including the type of RUs ($N_i^{\mathrm{sc}}$), the number of RUs ($N_i^{\mathrm{RU}}$), the best number of repetitions ($N_i^{\mathrm{rep}}$), and transmit parameters ($\mathrm{MCS}_i$ and $P_i$), to guarantee QoS and the transmission reliability. These operations are described as follows.

*Step 1.* For each $\mathrm{UE}_i$, $i = 1 \cdots N$, we first calculate the required number of RUs $N_i^{\mathrm{RU}}$ according to (4) based on the available RU types and MCS selections. Specifically, the required transmit time to carry the amount of data $D_i$ cannot be greater than the delay requirements. These results are collected as the feasible setting pairs of RU type and MCS setting for each $\mathrm{UE}_i$, denoted as set $A_i$; i.e.,

$$A_i = \left\{ \left( \mathrm{N}_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j} \right) \mid N_{i,j}^{\mathrm{RU}} \times \frac{N_{i,j}^{\mathrm{slot}}}{2} \le d_i,\ N_{i,j}^{\mathrm{sc}} \in \{1, 3, 6, 12\}, \mathrm{MCS}_{i,j} \in \{0 \cdots 10\} \right\}, \tag{10}$$

where $j$ is the index of feasible setting pair of RU type and MCS for each $\mathrm{UE}_i$ and $N_{i,j}^{\mathrm{slot}}$ is the number of required slots when the RU type is $N_{i,j}^{\mathrm{sc}}$, which can be obtained by (7). Note that $N_{i,j}^{\mathrm{slot}}$ is divided by 2 because two slots constitute 1 ms, which is the unit of delay constraint $d_i$.

*Step 2.* For each $\mathrm{UE}_i$, $i = 1 \cdots N$, consider the feasible RU type and MCS setting pair $(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}) \in A_i$; we calculate the allowed repetition numbers $N_{i,j}^{\mathrm{rep}}$ in which each $N_{i,j,k}^{\mathrm{rep}} \in N_{i,j}^{\mathrm{rep}}$ can make $\mathrm{UE}_i$ not only satisfy the required reliability $R_i$ but also ensure the corresponding transmission power $P_{i,j,k}$ in the feasible ranges; i.e.,

$$N_{i,j}^{\mathrm{rep}} = \left\{ N_{i,j,k}^{\mathrm{rep}} \mid 1 - \left(1 - P_{i,j,k}^{s}\right)^{N_{i,j,k}^{\mathrm{rep}}} \ge R_i, N_{i,j,k}^{\mathrm{rep}} \in \left\{ 2^l \mid l \in \{0 \cdots 7\} \right\}, N_{i,j}^{\mathrm{RU}} \times \frac{N_{i,j}^{\mathrm{slot}}}{2} \times N_{i,j,k}^{\mathrm{rep}} \le d_i, 0 \le P\left(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, \mathrm{BER}_{i,j,k}\right) \le P_i^{\max} \right\}, \tag{11}$$

where

$$\mathrm{BER}_{i,j,k} = 1 - \left(1 - \left(1 - R_i\right)^{1/N_{i,j,k}^{\mathrm{rep}}}\right)^{1/D_i} \tag{12}$$

is derived from (5) and $P(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, \mathrm{BER}_{i,j,k})$ is a function which returns the minimum transmit power for the RU type $N_{i,j}^{\mathrm{sc}}$, MCS setting $\mathrm{MCS}_{i,j}$, and target bit-error-rate $\mathrm{BER}_{i,j,k}$; i.e.,

$$P\left(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, \mathrm{BER}_{i,j,k}\right) = 10^{\mathrm{SNR}_{\mathrm{dB}}^{\mathrm{Req}}(\mathrm{MCS}_{i,j}, \mathrm{BER}_{i,j,k})/10} \times \frac{\left(BN_0 + I\right) \cdot L\left(i, \mathrm{eNB}\right) \cdot N_{i,j}^{\mathrm{sc}}}{G_i G_{\mathrm{eNB}}}, \tag{13}$$

which can be derived from (3).

After that we have all the feasible RU type and MCS setting pairs with each of their allowed repetition numbers $N_{i,j}^{\mathrm{rep}}$.

*Step 3.* Based on the results of Steps 1 and 2, we calculate the most energy-saving repetition number $N_{i,j}^{\mathrm{rep}^*}$ for each feasible combination pair $(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}) \in A_i$, where

$$N_{i,j}^{\mathrm{rep}^*} = \arg \min_{N_{i,j,k}^{\mathrm{rep}} \in N_{i,j}^{\mathrm{rep}}} E\left(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, N_{i,j,k}^{\mathrm{rep}}\right),$$
$$E\left(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, N_{i,j,k}^{\mathrm{rep}}\right) = P\left(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, \mathrm{BER}_{i,j,k}\right) \times N_{i,j}^{\mathrm{RU}} \times \frac{N_{i,j}^{\mathrm{slot}}}{2} \times N_{i,j,k}^{\mathrm{rep}}. \tag{14}$$

Then, reform $A_i$ as a set of triplets $(N_{i,j}^{\mathrm{sc}}, \mathrm{MCS}_{i,j}, N_{i,j}^{\mathrm{rep}^*})$. Each triplet in $A_i$ is a feasible configuration of RU type, MCS setting, and repetition number.

*Step 4.* Then, we choose the best triplet of $(\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}})$ from $A_i$ as the default parameter of $\mathrm{UE}_i$, which incurs the minimum energy consumption by

$$\left(\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}}\right) = \arg \min_{(\mathbf{N_{i,j}^{sc}}, \mathbf{MCS_{i,j}}, \mathbf{N_{i,j}^{rep^*}}) \in A_i} \left\{ \mathbf{E}\left(\mathbf{N_{i,j}^{sc}}, \mathbf{MCS_{i,j}}, \mathbf{N_{i,j}^{rep^*}}\right) \right\}. \tag{15}$$

Through the above steps, we can determine the best RU type $\mathbf{N_i^{sc^*}}$, MCS setting $\mathbf{MCS_i^*}$, and repetition number $\mathbf{N_i^{rep^*}}$ that can incur the least energy consumption and meet the reliability requirement $\mathbf{R_i}$ of each $\mathrm{UE}_i$.

*4.2. Phase 2. Weighting Based Flexible Scheduling.* The goal of the second phase is to optimize the scheduling results of requests from UEs, including the subcarrier set of RUs ($S_i^{\mathrm{sc}}$) and the start time of RUs ($T_i^{\mathrm{sc}}$). In addition, if needed, it can adaptively adjust the transmission parameters of UEs to ensure the delay constraint and enhance spectrum utilization. The detailed steps are depicted as follows.

*Step 1.* We first define a *score function* to evaluate the emergency and inflexibility for each $\mathrm{UE}_i$ with uplink transmission request, i.e.,

$$\mathrm{Score}_i = W_1 \times Em_i + W_2 \times \widetilde{\mathrm{Inf}_i}, \tag{16}$$

where $W_1 \in [0,1]$ and $W_2 \in [0,1]$ are the weighting factors of the degrees of the emergency and inflexibility, respectively, that satisfy $W_1 + W_2 = 1$. Note that $Em_i$ is the urgent level of $\mathrm{UE}_i$'s request compared to others; i.e.,

$$Em_i = \frac{\max_j \left\{T_j^R\right\} - T_i^R}{\max_j \left\{T_j^R\right\} - \min_j \left\{T_j^R\right\}}, \tag{17}$$

where $T_j^R$ is the remaining time from the scheduling subframe $T^S$ (or current subframe) to the delay deadline $T_j^{\mathrm{req}} + d_j$ of $\mathrm{UE}_j$; i.e.,

$$T_i^R = \max\left(\left(T_i^{\mathrm{req}} + d_i\right) - T^s, 0\right). \tag{18}$$

$\widetilde{\mathrm{Inf}_i}$ is the number of RU types that $\mathrm{UE}_i$ can choose, which is defined by

$$\widetilde{\mathrm{Inf}_i} = \frac{\mathrm{Inf}_i}{\max_j \left\{\mathrm{Inf}_j\right\}}, \tag{19}$$

where

$$\mathrm{Inf}_i = \begin{cases} 4, & \text{if } \Psi_{N^{\mathrm{sc}}}(A_i) = 1 \\ 3, & \text{if } \Psi_{N^{\mathrm{sc}}}(A_i) = 2 \\ 2, & \text{if } \Psi_{N^{\mathrm{sc}}}(A_i) = 3 \\ 1, & \text{if } \Psi_{N^{\mathrm{sc}}}(A_i) > 3 \end{cases} \tag{20}$$

and $\Psi_{N^{\mathrm{sc}}}(A_i)$ is the number of choices of RU types for the feasible setting pair $A_i$. That means if $\mathrm{UE}_i$ has fewer choices, its inflexibility is higher and needs to be scheduled earlier.

Now, for each $\mathrm{UE}_i$, $i = 1 \cdots N$, we calculate its $\mathrm{Score}_i$ and sort them in descending order. For the UEs without any request, define its score as $-\infty$. Without loss of generality, we use *List*$\boldsymbol{L}$ to represent the sorted sequence of the UEs.

*Step 2.* Before determining the subcarrier set of RUs, we first define a function $\mathrm{Waste}(i, S_i^{\mathrm{sc}})$ to reflect the potential waste of resource if $\mathrm{UE}_i$'s RUs are allocated at subcarrier set $S_i^{\mathrm{sc}}$; i.e.,

$$\begin{aligned} &\mathrm{Waste}\left(i, S_i^{\mathrm{sc}}\right) \\ &= \sum_{k' \in K - S_i^{\mathrm{sc}}} \left( \left( \max_{k \in S_i^{\mathrm{sc}}} \left\{\widehat{S}_k\right\} + \left( N_i^{\mathrm{RU}} \times \frac{N_i^{\mathrm{slot}}}{2} \times N_i^{\mathrm{rep}} \right) \right) \right. \\ &\left. - \widehat{S}_{k'} \right)^+ + \sum_{k \in S_i^{\mathrm{sc}}} \left( \max_{k \in S_i^{\mathrm{sc}}} \left\{\widehat{S}_k\right\} - \widehat{S}_k \right), \end{aligned} \tag{21}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$ outputs the value larger than or equal to 0; $\max_{k \in S_i^{\mathrm{sc}}}\{\widehat{S}_k\}$ means the earliest available resource allocation start time of RUs if the subcarrier set is $S_i^{\mathrm{sc}}$, where $\widehat{S}_k = \max\{S_k, T_i^{\mathrm{DCI}} + 1\}$ is to ensure allocating RU after DCI. Note that (21) sums up the unused resource space before the resource allocation finish time of $\mathrm{UE}_i$ if $\mathrm{UE}_i$'s RUs are allocated at $S_i^{\mathrm{sc}}$.

Then, based on (21), we choose the best subcarrier set $S_i^{\mathrm{sc}^*}$ that makes $\mathrm{UE}_i$ have the minimal $\mathrm{Waste}(i, S_i^{\mathrm{sc}})$ without violating its delay deadline; i.e.,

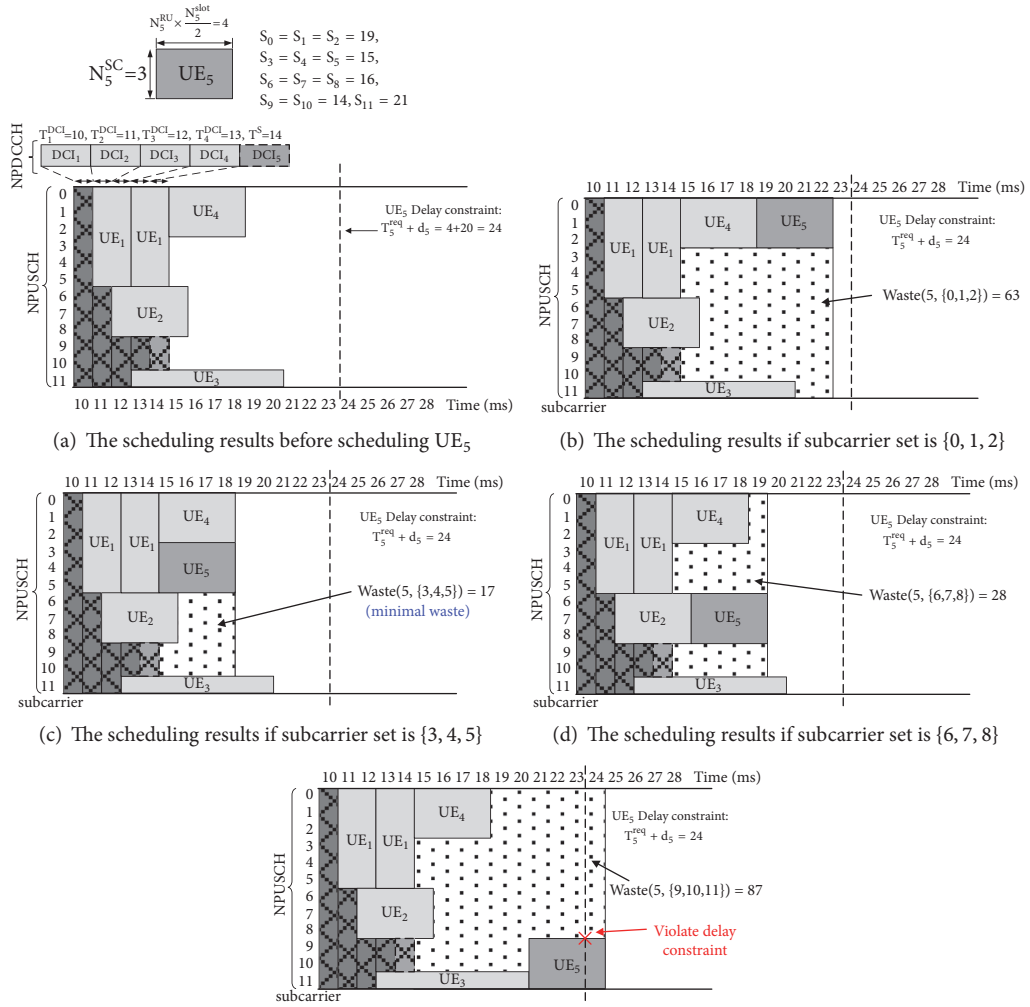
$$\begin{aligned} S_i^{\mathrm{sc}^*} = \arg \min_{S_i^{\mathrm{sc}} \subseteq \Theta(N_i^{\mathrm{sc}^*})} \left\{ \mathrm{Waste}\left(i, S_i^{\mathrm{sc}}\right) \mid \max_{k \in S_i^{\mathrm{sc}}} \left\{\widehat{S}_k\right\} \right. \\ \left. + \left( N_i^{\mathrm{RU}^*} \times \frac{N_i^{\mathrm{slot}^*}}{2} \times N_i^{\mathrm{rep}^*} \right) < \left(T_i^{\mathrm{req}} + d_i\right) \right\}, \end{aligned} \tag{22}$$

where $\Theta(N_i^{\mathrm{sc}^*})$ is the set of available subcarrier sets when default RU type $N_i^{\mathrm{sc}^*}$ is used.

If $S_i^{\mathrm{sc}^*} \neq \emptyset$, we set the subframe index of $\mathrm{DCI}_i$ by $T_i^{\mathrm{DCI}} = T^s$ and start time $T_i^{\mathrm{sc}} = \max_{k \in S_i^{\mathrm{sc}^*}}\{\widehat{S}_k\}$. Then, update the available scheduling subframe for subcarriers $k \in S_i^{\mathrm{sc}^*}$ and $k' \in \Theta(N_i^{\mathrm{sc}^*}) - S_i^{\mathrm{sc}^*}$ by $S_k = \max\{\max_{k \in S_i^{\mathrm{sc}^*}}\{\widehat{S}_k\} + (N_i^{\mathrm{RU}^*} \times (N_i^{\mathrm{slot}^*}/2) \times N_i^{\mathrm{rep}^*}), T_i^{\mathrm{DCI}} + 1\}$ and $S_{k'} = \widehat{S}_{k'}$, respectively. Finally, update $T^s = T_i^{\mathrm{DCI}} + 1$ and then remove $\mathrm{UE}_i$ from List $\boldsymbol{L}$. However, if $S_i^{\mathrm{sc}^*} = \emptyset$, it means that current transmission parameter setting is infeasible and then we check whether or not $\mathrm{UE}_i$ has other feasible triplet in $A_i$ other than the default parameter. If yes, go to Step 3 for further adjusting. If no, we remove such $\mathrm{UE}_i$ from List $\boldsymbol{L}$ and go back to Step 2 to schedule the next UE. The above steps are repeated until List $\boldsymbol{L}$ is empty and then terminate this phase.

*Step 3.* Here, we try to change the type of RUs and/or MCSs of $\mathrm{UE}_i$ by referring $A_i$ and choose the new triplet that can satisfy the delay deadline while incurring the least extra energy consumption and resource as follows.

First, we define a *cost ratio* $\mathbf{C_i^{\alpha,\beta}}$ to reflect the results of extra consumed energy over the extra required resource space when the original pair of RU type and MCS, denoted as $\boldsymbol{\alpha} = (\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}})$, changes to the new pair, denoted




(a) The scheduling results before scheduling $UE_5$

(b) The scheduling results if subcarrier set is {0, 1, 2}

(c) The scheduling results if subcarrier set is {3, 4, 5}

(d) The scheduling results if subcarrier set is {6, 7, 8}

(e) The scheduling results if subcarrier set is {9, 10, 11}

FIGURE 4: Examples to schedule $UE_5$ based on the potential resource waste.

as $\boldsymbol{\beta} = (\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}})$ for $(\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}}) \in A_i - (\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}})$; i.e.,

$$C_i^{\alpha,\beta} = \begin{cases} \dfrac{\Delta E_i^{\alpha,\beta}}{\Delta \text{Area}_i^{\alpha,\beta}}, & \text{if } \Delta \text{Area}_i^{\alpha,\beta} > 0 \\ \Delta E_i^{\alpha,\beta}, & \text{if } \Delta \text{Area}_i^{\alpha,\beta} = 0, \end{cases} \tag{23}$$

where the extra consumed energy is $\Delta E_i^{\alpha,\beta} = (E(\boldsymbol{\beta}) - E(\boldsymbol{\alpha}))^+$ and the extra resource space is $\Delta \text{Area}_i^{\alpha,\beta} = (\mathbf{N_i^{sc'}} \times T(\boldsymbol{\beta}) - \mathbf{N_i^{sc^*}} \times T(\boldsymbol{\alpha}))^+$.

Then, we choose the new pair $\boldsymbol{\beta}^*$ which incurs the minimal cost ratio; i.e.,

$$\begin{aligned} \boldsymbol{\beta}^* &= \left(\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}}\right) \\ &= \arg\min \left\{ C_i^{\alpha,\beta} \mid \boldsymbol{\beta} \in A_i - \boldsymbol{\alpha} \right\} \end{aligned} \tag{24}$$

and replace the default parameters by $\mathbf{N_i^{sc^*}} = \mathbf{N_i^{sc'}}$, $\mathbf{MCS_i^*} = \mathbf{MCS_i'}$, and $\mathbf{N_i^{rep^*}} = \mathbf{N_i^{rep'}}$ accordingly. Finally, go back to Step 2 for further allocation.

Through the above steps, we can determine each $UE_i$'s subcarrier set $S_i^{sc}$, start time $T_i^{sc}$, and the corresponding configurations of $MCS_i$, $N_i^{rep}$, and $P_i$ while ensuring the delay deadline and reducing the waste of spectrum resource and energy.

Below, we give an example in Figure 4, where there are four scheduled UEs ($UE_1$~$UE_4$) and one UE ($UE_5$) to be scheduled. The subframe indexes of DCIs for the four scheduled UEs are $T_1^{DCI} = 10$, $T_2^{DCI} = 11$, $T_3^{DCI} = 12$, and $T_4^{DCI} = 13$ (ms), separately. The current earliest subframe index for DCI is subframe 14 and scheduling subframe is $T^s = 14$ (ms). Now, we consider schedule $UE_5$, whose RU type is 3-tone, number of RUs is $N_5^{RU} = 1$, total length is $N_5^{RU} \times (N_5^{slot}/2) = 1 \times (8/2) = 4$ (ms), arrival time is $T_5^{req} = 4$ (ms), and delay constraint is $d_5 = 20$ (ms). From

Table 5: The simulation parameters [7, 8].

| Parameter | Value |
|---|---|
| maximum transmit power ($\mathbf{P_i^{max}}$) | 23 dBm |
| antenna gain of transmitter ($\mathbf{G_i}$) | −4 dBi |
| antenna gain of receiver ($\mathbf{G_{eNB}}$) | 18 dBi |
| thermal noise density ($\mathbf{N_0}$) | −174 dBm/Hz |
| path loss ($\mathbf{L(i, eNB)}$) | 120.9 + 30.76 log ($d$) dB, $d$ in Km |
| distance from the base station | 0~15 (Km) |
| number of UEs ($\boldsymbol{N}$) | 3000~30000 |
| request data size ($\mathbf{D_i}$) | 50~200 bytes |
| delay constraint ($\mathbf{d_i}$) | 50, 100, 150, 300 (ms) |
| required reliability ($\mathbf{R_i}$) | 90%~99% |

the current scheduling results, as shown in Figure 4(a), the available start time for each subcarrier is $S_0 = 19$, $S_1 = 19$, $S_2 = 19$, $S_3 = 15$, $S_4 = 15$, $S_5 = 15$, $S_6 = 16$, $S_7 = 16$, $S_8 = 16$, $S_9 = 14$, $S_{10} = 14$, and $S_{11} = 21$, respectively. Thus, the available subcarrier set for $UE_5$ is $S_5^{sc} \in \{\{0, 1, 2\}, \{3, 4, 5\}, \{6, 7, 8\}, \{9, 10, 11\}\}$. So, for the 4 possible subcarrier sets, the potential resource waste for $UE_5$ is $\text{Waste}(5, \{0, 1, 2\}) = \sum_{k'=3\sim11}(\max\{19, 19, 19\}+1\times(8/2)\times1)^+ - \widehat{S}_{k'} = 63$, $\text{Waste}(5, \{3, 4, 5\}) = \sum_{k'=0\sim2,6\sim11}(\max\{15, 15, 15\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 17$, $\text{Waste}(5, \{6, 7, 8\}) = \sum_{k'=0\sim5,9\sim11}(\max\{16, 16, 16\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 28$, and $\text{Waste}(5, \{9, 10, 11\}) = \sum_{k'=0\sim8}(\max\{15, 15, 21\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 87$, which are the doted regions in Figures 4(b), 4(c), 4(d), and 4(e), respectively. Moreover, the subcarrier set of $\{9, 10, 11\}$ in Figure 4(e) is infeasible because it violates the delay deadline of $UE_5$ (i.e., $\max_{k\in\{9,10,11\}}\{\widehat{S}_k\} + (N_5^{RU} \times (N_5^{slot}/2) \times N_5^{rep}) = 21 + 1 \times (8/2) \times 1 = 25 \geq (T_5^{req} + d_5) = 4 + 20 = 24$). Thus, based on (21), the scheme chooses the best feasible subcarrier set $S_5^{sc^*} = \{3, 4, 5\}$ and the start time of $UE_5$'s schedules RUs is $T_5^{sc} = \max_{k\in\{3,4,5\}}\{\widehat{S}_k\} = 15$ because it has the minimal value of $\text{Waste}(5, \{3, 4, 5\}) = 17$.

## 5. Simulation Results

In this section, we develop a simulator in C++ language to verify the efficiency of the proposed scheme (currently, the well-known simulator, such as ns-3 [26], has not supported NB-IoT model in terms of channels mappings and access procedures.). The parameters of the simulation are shown in Table 5. Specifically, the simulator emulates a base station to serve $N = 3000\sim30000$ UEs. Based on the model of *MAR (Mobile Autonomous Reporting)* [8], the uplink requests of UEs arrive randomly with the data size of $D_i = 50\sim200$ bytes according to the Pareto distribution with $\alpha = 2.5$. In addition, the interarrival time of requests includes 30 minutes (5%), 1 hour (15%), 2 hours (40%), and 1 day (40%), respectively.

In this simulation, we compare our scheme *(Ours)* with the standard scheme *(Spec)* [2], Narrowband Link Adaptation scheme *(NBLA)* [14], random scheduling scheme *(Random)*, and round robin scheme *(RR)*. Specifically, Spec scheme chooses the single-tone for UEs with constant repetition number for simplicity (the numbers are 1 and 2 in the simulation, denoted as Spec(1) and Spec(2)). NBLA can adjust the MCSs and repetition levels iteratively to ensure the transmission quality and delay. Random scheme schedules the UEs in a random order with a random repetition number. RR scheme schedules the UEs in round robin order with the repetition number of 1. Note that the weighting factors of our scheme is $W_1 = 0.5$ and $W_2 = 0.5$.

We consider five performance metrics: (i) *system throughput*: the total number of data bits received successfully by the eNB during the experiment period; (ii) *the number of serving UEs*: the average number of UEs that satisfy QoS and reliability; (iii) *resource consumption*: the frame space allocated to the uplink transmission over the total frame space; (iv) *packet drop rate*: the number of dropped packets due to violating delay constraint over the total number of packets; (v) *energy consumption per UE*: the average consumed energy of each served UE. Note that the scheduling interval is 30 ms and the simulator emulates for 24 hours.

*5.1. System Throughput.* First, we investigate the effects of number of request UEs on system throughput. As shown in Figure 5(a), we can see that when the number of request UEs increases, the system throughput of all scheme increases fast and then slows down due to system saturation. RR, Spec, and Random perform worse because they cannot satisfy UEs' QoS requirement and reliability. Thus, the total number of data bits received successfully by the eNB is few. Specifically, Spec(2) performs slightly better than Spec(1) because it can meet more UEs' reliability due to applying the larger repetition number. NBLA is better than the above schemes because it can adjust MCSs and repetition levels to satisfy QoS and reliability. Note that our scheme outperforms other schemes because it can flexibly adjust RU type and optimize repetition number to satisfy QoS while ensuring transmission reliability.

We also investigate the effects of distribution of request data size on system throughput. As shown in Figure 5(b), similarly, when the request data size increases, the system throughput of most schemes increases and then slows down due to system saturation. RR and Spec(1) decrease fast because their repetition number is 1 and may not satisfy the successful transmission probability due to larger request size. Our scheme is still the best because it can schedule UEs flexibly while ensuring QoS requirement and reliability.

Then, Figure 5(c) shows the impact of the distribution of delay constraints on system throughout. As can be seen, when the distribution of delay constraints increases, the system throughput of most schemes increases slowly. This is because the longer delay constraint can help more UEs to be satisfied until the frame space is exhausted. RR and Spec(1) increase slowly because they fix repetition number by 1 that would limit the successful transmission probability even when the UEs have longer delay constraint. Note that our scheme has the highest throughput because of its flexible scheduling and appropriate parameter setting.

Finally, we investigate the impact of the distribution of required reliability on system throughout. In Figure 5(d), when the distribution of reliability increases, the system throughput of most schemes decreases. The reason is that more UEs with strict reliability make all schemes harder to

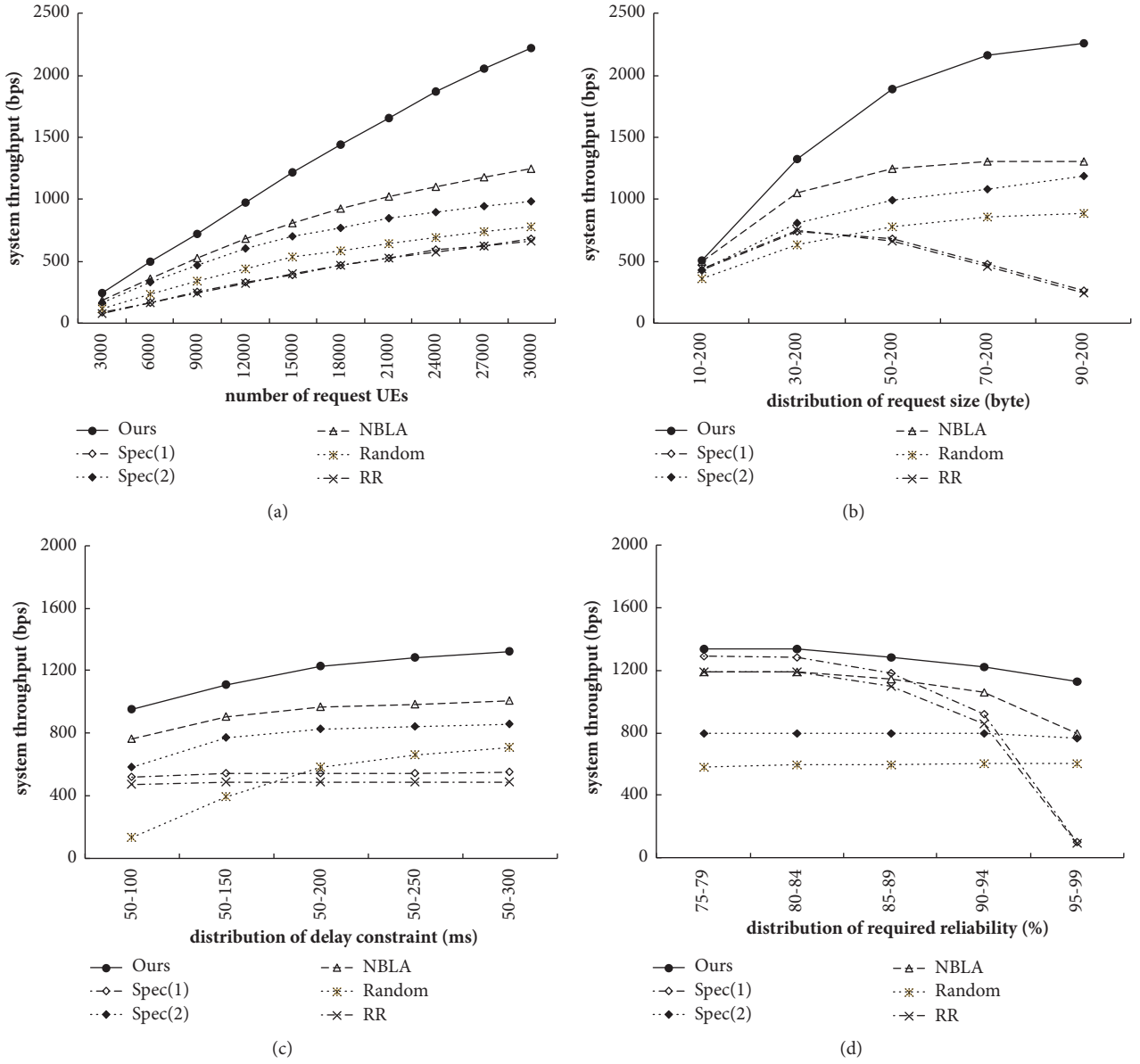



Figure 5: Comparisons on the system throughput of all schemes.

serve them due to their higher requirements. RR and Spec(1) decrease dramatically because their repetition number is 1 that could not serve most UEs with higher reliability. Note that our scheme still outperforms others.

*5.2. Number of Serving UEs.* Then, we investigate the effects of number of request UEs on number of serving UEs. As shown in Figure 6(a), similarly, Random performs the worst because it randomly schedules the UEs with a random repetition number; thus, the QoS and reliability of UEs may not be met. Spec and RR perform slightly better than Random scheme because they prefer to choose single-tone with the fixed repetition number for UEs; thus, it could potentially satisfy more UEs with small data request and lower reliability requirement. Spec(2) is better than Spec(1) because a larger repetition number can achieve higher reliability. NBLA is better than the above schemes because it can adjust the repetition levels and MCSs interactively to satisfy the transmission reliability and delay. Note that our scheme outperforms all others because our scheme can optimize the number of repetitions to satisfy the transmission reliability in phase 1 and apply the best configuration pair of RU type and MCS to ensure QoS while enhancing the spectrum utilization in phase 2.

We also investigate the effects of distribution of request data size on number of serving UEs. As shown in Figure 6(b), contrarily, when the request data size increases, the number of serving UEs of all schemes decreases because the UEs with larger request size consume more frame space. RR and Spec(1) decrease fast because their repetition number is fixed by 1 that would not satisfy the UEs with larger request size.

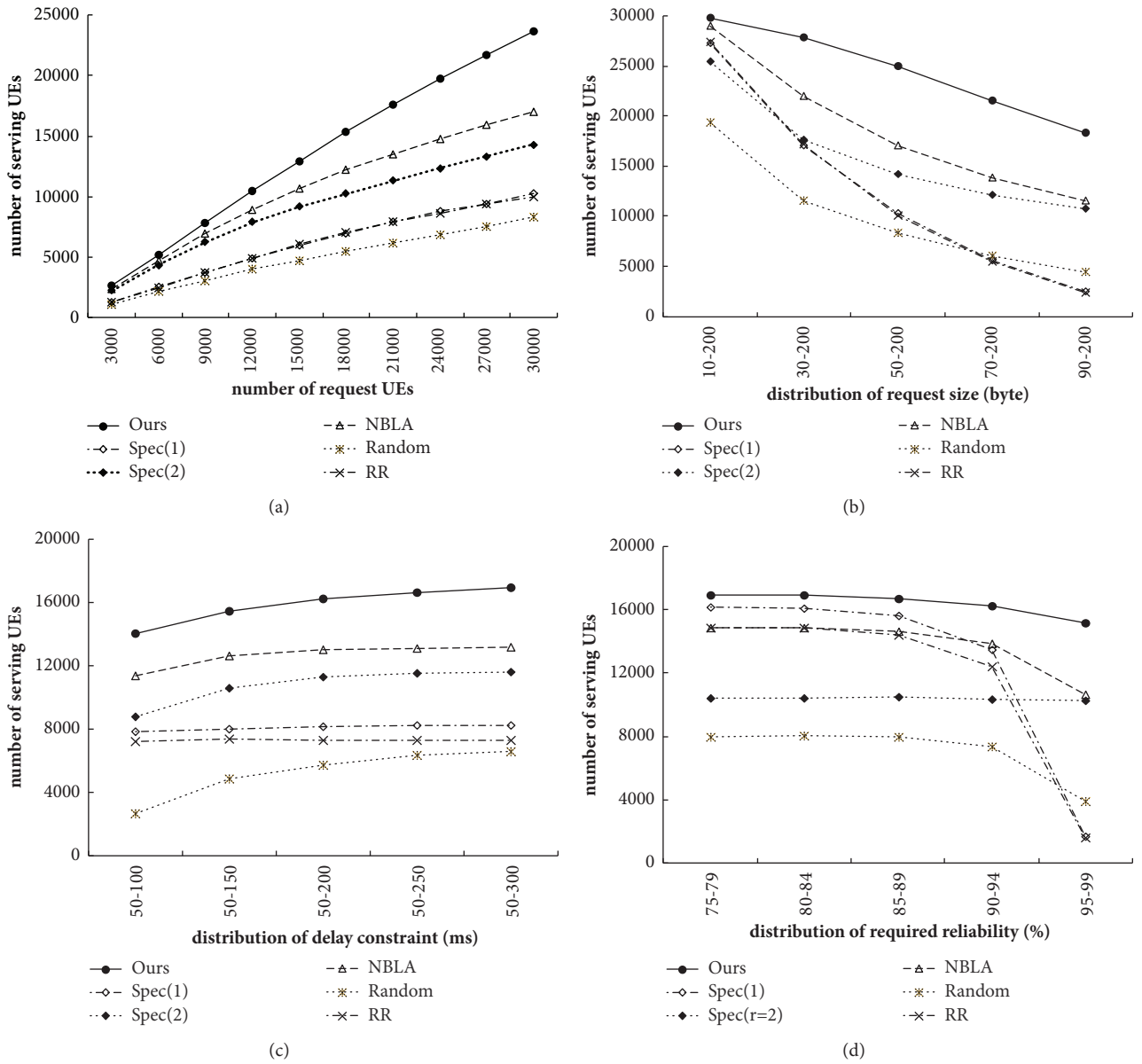



Figure 6: Comparisons on the number of serving UEs of all schemes.

Our scheme is the best because it flexibly schedules UEs to satisfy their delay requirement and reliability.

In Figure 6(c), we observe the impact of the distribution of delay constraints on number of serving UEs. As can be seen, when the distribution of delay constraints increases, the number of serving UEs of all schemes increases slowly. This is because the longer delay constraint can make more UEs tolerate allocation time until the frame space is exhausted. RR and Spec(1) increase very slowly because their fixed repetition number (i.e., 1) would limit the transmission probability although the UEs have longer delay time. Note that our scheme has the highest performance because it can well determine the scheduling setting.

Finally, we investigate the impact of the distribution of required reliability on number of serving UEs in Figure 6(d). We can see that when the distribution of reliability increases, the number of serving UEs of all schemes decreases. The reason is that more UEs with strict reliability will make schemes hardly serve them due to higher requirements. Note that RR and Spec(1) decrease dramatically because they fix the repetition number by 1 that could not serve the UEs with higher reliability requirements. Also note that our scheme still outperforms others even when the reliability requirements become higher.

*5.3. Packet Drop Rate.* Then, we investigate the effects of number of request UEs on packet drop rate. As shown in Figure 7(a), we can see that when the number of request UEs increases, the packet drop rate of all schemes increases due to network saturated gradually. Spec, Random, and RR perform

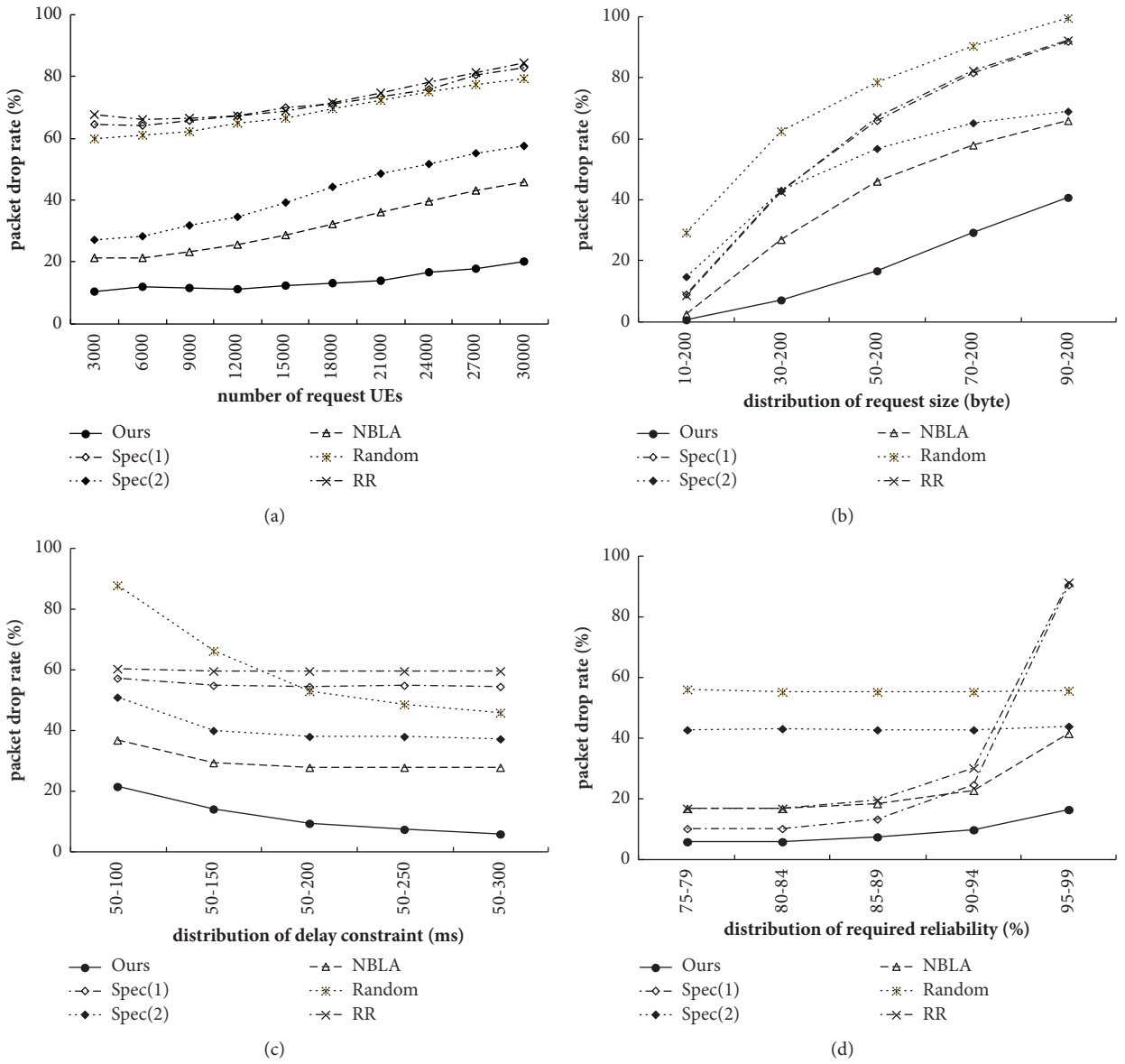



Figure 7: Comparisons on the packet drop rate of all schemes.

worse because they usually schedule UEs with single-tone so the UEs' RU length may exceed their delay constraint. NBLA performs better than above schemes because it can choose better repetition number to mitigate packets being dropped. Our scheme has the lowest packet drop rate because our scheme can flexibly select multitype RUs and minimize the number of repetitions to guarantee the QoS while avoiding packets being dropped.

We also investigate the effects of distribution of request data size on packet drop rate. As shown in Figure 7(b), similarly, when the request data size increases, the packet drop rate of all schemes increases because the larger request size consumes more resources that may exceed the frame space which makes packet being dropped. RR and Spec(1) increase fast because they could not satisfy the UEs with larger request size, which require higher transmission probability. Our scheme is the best because it can schedule UEs according to their delay and reliability requirements.

Then, in Figure 7(c), we investigate the impact of the distribution of delay constraints on packet drop rate. As can be seen, when the distribution of delay constraints increases, the packet drop rate of all schemes decreases gradually. This is because the longer delay constraint can make more packets tolerate the allocation time before delay expires. RR and Spec(1) decrease very slowly because they fix repetition number by 1 so that it would not help packets being transmitted. Note that our scheme has the lowest packet drop rate because it can well determine the scheduling results.

Finally, we investigate the impact of the distribution of required reliability on packet drop rate in Figure 7(d). As

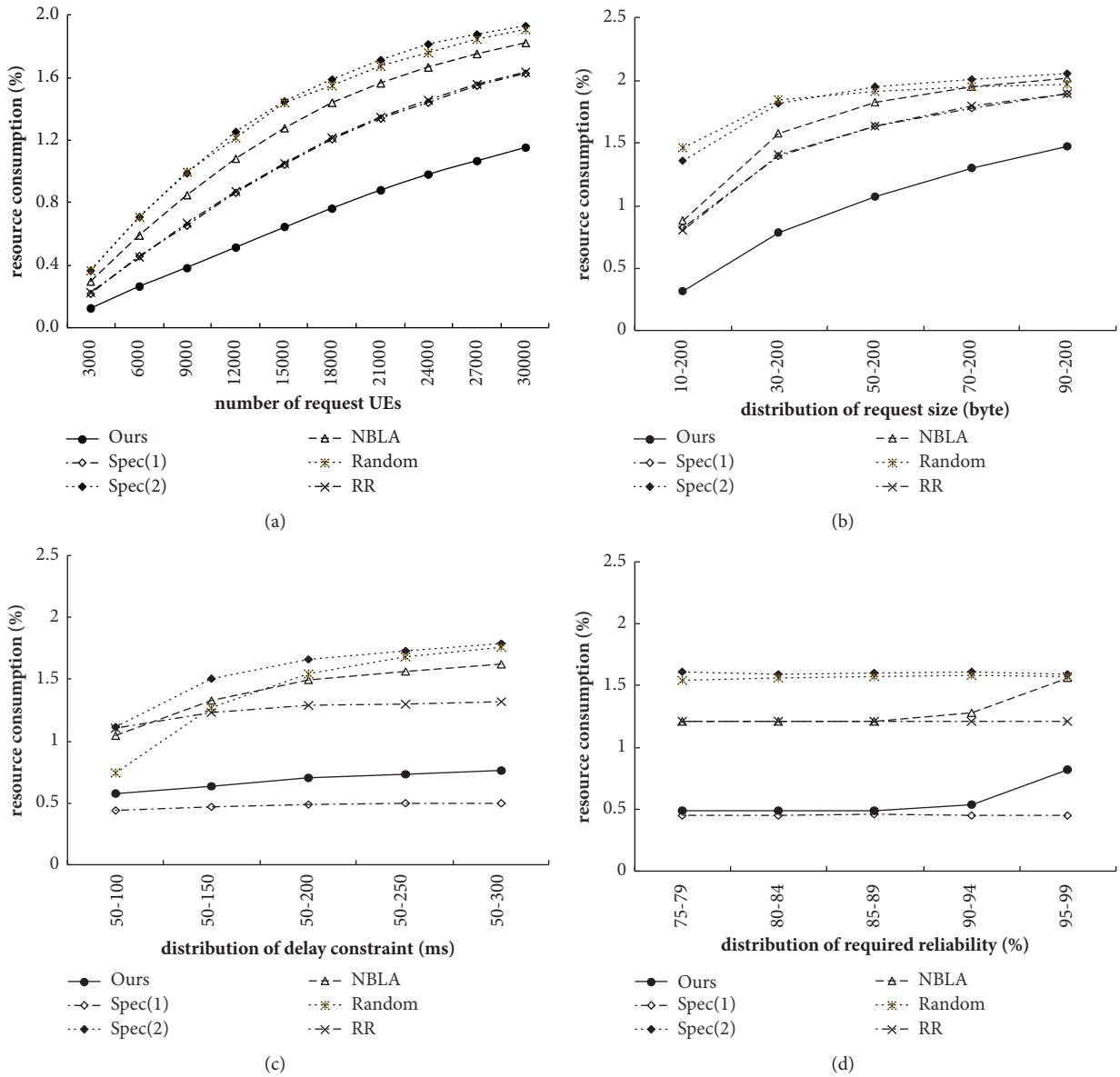



Figure 8: Comparisons on the resource consumption of all schemes.

can be seen, when the distribution of reliability increases, the packet drop rate of all schemes increases. Our scheme still outperforms others even when the reliability requirements are up to 95–99%.

*5.4. Resource Consumption.* Here, we investigate the effects of number of request UEs on resource consumption. As shown in Figure 8(a), we can see that when the number of request UEs increases, the resource consumption of all schemes increases. Spec(2), Random, and NBLA perform worst because they usually schedule the UEs with multiple number of repetitions and thus consume more resources. Spec(1) and RR have lower resource consumption because they prefer to choose the UEs with single repetition to reduce resource. Note that our scheme needs the least spectrum resource because our scheme exploits a waste function to avoid resource consumption.

Here, we also investigate the effects of distribution of request data size on resource consumption. As shown in Figure 8(b), similarly, when the request data size increases, the resource consumption of most schemes increases slowly. Spec(2), Random, and NBLA waste more resources because they usually schedule the UEs with multiple repetitions and thus consume more resources. Our scheme is still the best because it can schedule UEs flexibly while ensuring QoS requirement and reliability.

Then, in Figure 8(c), we investigate the impact of the distribution of delay constraints on resource consumption. As can be seen, when the distribution of delay constraints increases, the resource consumption of all schemes increases.

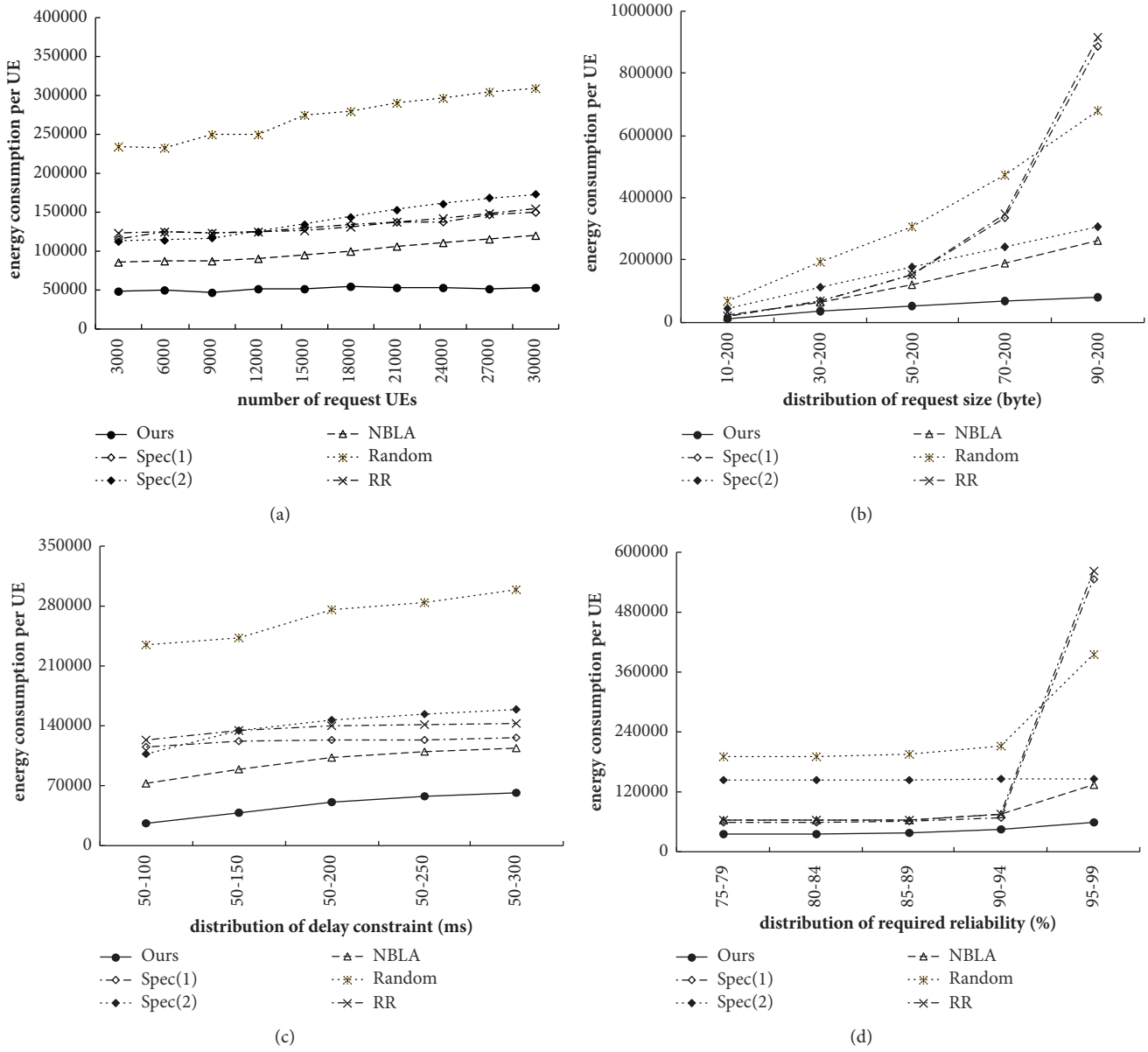



Figure 9: Comparisons on energy consumption per UE of all schemes.

This is because the longer delay constraint can make more UEs satisfied that may consume more resources. Note that although Spec(1) has the lowest resource consumption, it incurs lower system throughput, lower number of serving UEs, and higher packet drop rate, as compared to ours.

Finally, we investigate the impact of the distribution of required reliability on resource consumption, as shown in Figure 8(d). When the distribution of reliability increases, the resource consumption of most schemes increases. Our scheme increases slightly because our scheme can enlarge the repetition number with multiple tone to satisfy the UEs with higher reliability requirement.

*5.5. Energy Consumption per UE.* Finally, we investigate the effects of the number of request UEs on energy consumption per UE. As shown in Figure 9(a), we can see that the energy consumption per UE of all schemes increases when the number of request UEs increases. This is because the network is saturated and most satisfied UEs are with higher MCS, which require less resource but consume more energy. Random scheme performs the worst because it randomly chooses the number of repetitions that may potentially increase the transmission time, thus consuming more energy. Spec and RR are better than Random scheme because they only serve the UEs with small size request which consumes less energy. NBLA performs slightly better because it can determine the number of repetitions appropriately but neglects to minimize the transmission power. Note that our scheme performs the best because our scheme can choose the best scheduling parameters of RUs with least energy consumption in phase

1 and leverage the cost ratio to reduce energy consumption in phase 2, thus saving energy more efficiently.

In Figure 9(b), we also investigate the effects of distribution of request data size on energy consumption per UE. As can be seen, when the request data size increases, the energy consumption per UE of all schemes increases. Our scheme incurs the lowest energy consumption because it can determine the minimal transmit power and the corresponding RU setting to save energy.

Then, in Figure 9(c), we observe the impact of the distribution of delay constraints on energy consumption per UE. We can see that when the distribution of delay constraints increases, the energy consumption per UE of all schemes increases. Similarly, our scheme still has the lowest energy consumption because it can calculate the best transmit power and exploit the cost ratio to reduce energy consumption.

Finally, we investigate the impact of the distribution of required reliability on energy consumption per UE. Figure 9(d) shows that when the distribution of reliability increases, the energy consumption of most schemes increases. Our scheme still outperforms all other schemes because it can determine the best scheduling parameters based on the required reliability of UEs.

## 6. Conclusion

In this paper, we have addressed the problem of energy saving and reliable communications in NB-IoT networks. We first model this problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient and high-reliable scheme which has two phases. The first phase leverages the strategy of minimal energy cost to choose the default scheduling parameters with least energy consumption. The second phase exploits the weighting score function to balance the emergency and inflexibility of the request transmission and then serve the UEs with least potential resource waste. Simulation results have verified that our scheme can satisfy more UEs with ultra-reliability and QoS requirement while saving their energy.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research is cosponsored by MOST 106-2221-E-024-004-, 102-2218-E-182-008-MY3, 105-2221-E-182-051, 106-2221-E-182-015-MY3, 105-2745-8-182-001, 106-2221-E-024-004, 105-2221-E-009-100-MY3, 105-2218-E-009-029, 105-2923-E-009-001-MY2, 104-2221-E-009-113-MY3, MoE ATU Plan, Delta Electronics, ITRI, Institute for Information Industry, Academia Sinica AS-105-TP-A07, and Chang Gung Memorial Hospital, Taoyuan.

## References

[1] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Physical Channels and Modulation, v14.4.0*, pp. 1–6, 2017.

[2] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Physical Layer Procedures, v14.4.0*, pp. 1–6, 2017.

[3] J.-M. Liang, J.-J. Chen, H.-H. Cheng, and Y.-C. Tseng, "An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE-advanced networks for internet of things," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, pp. 13–22, 2013.

[4] E. Mohyeldin, "Minimum Technical Performance Requirements for IMT-2020 radio interface(s)," *ITU-R Workshop on IMT-2020 Terrestrial Radio Interfaces*, pp. 1–12, 2016.

[5] R. Ma, K. H. Teo, S. Shinjo, K. Yamanaka, and P. M. Asbeck, "A GaN PA for 4G LTE-Advanced and 5G: Meeting the telecommunication needs of various vertical sectors including automobiles, robotics, health care, factory automation, agriculture, education, and more," *IEEE Microwave Magazine*, vol. 18, no. 7, pp. 77–85, 2017.

[6] O. Teyeb, G. Wikstr, M. Stattin, T. Cheng, S. Faxér, and H. Do, "Evolving LTE to fit the 5G future," *Ericsson Technology Review*, 2017.

[7] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT system for M2M communication," in *Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, WCNC 2016*, pp. 1–5, April 2016.

[8] TR 45.820, "Cellular system support for ultra low complexity and low throughput internet of things," *v2.1.0*, pp. 1–6, 2015.

[9] Y. E. Wang, X. Lin, A. Adhikary et al., "A primer on 3GPP narrowband Internet of Things (NB-IoT)," *IEEE Communications Magazine*, vol. 15, no. 3, pp. 117–123, 2017.

[10] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. P. Koskinen, "Overview of narrowband IoT in LTE Rel-13," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, pp. 1–7, 2016.

[11] A. D. Zayas and P. Merino, "The 3GPP NB-IoT system architecture for the Internet of Things," in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 277–282, Paris, France, May 2017.

[12] N. Mangalvedhe, R. Ratasuk, and A. Ghosh, "NB-IoT deployment study for low power wide area cellular IoT," in *Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, Valencia, Spain, September 2016.

[13] L. Zhang, A. Ijaz, P. Xiao, and R. Tafazolli, "Channel Equalization and Interference Analysis for Uplink Narrowband Internet of Things (NB-IoT)," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2206–2209, 2017.

[14] C. Yu, L. Yu, Y. Wu, Y. He, and Q. Lu, "Uplink Scheduling and link adaptation for narrowband internet of things systems," *IEEE Access*, vol. 5, pp. 1724–1734, 2017.

[15] X. Lin, A. Adhikary, and Y. Eric Wang, "Random access preamble design and detection for 3GPP narrowband IoT systems," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 640–643, 2016.

[16] S. Oh and J. Shin, "An efficient small data transmission scheme in the 3GPP NB-IoT system," *IEEE Communications Letters*, vol. 21, no. 3, pp. 660–663, 2017.

[17] H. Kroll, M. Korb, B. Weber, S. Willi, and Q. Huang, "Maximum-likelihood detection for energy-efficient timing

acquisition in NB-IoT," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–5, San Francisco, CA, USA, March 2017.

[18] J. Lee and J. Lee, "Prediction-based energy saving mechanism in 3GPP NB-IoT networks," *Sensors*, vol. 17, no. 9, 2017.

[19] A. E. Mostafa, Y. Zhou, and V. W. Wong, "Connectivity maximization for narrowband IoT systems with NOMA," in *IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.

[20] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Radio Resource Control (RRC), v15.1.0*, pp. 1–786, March 2018.

[21] G. Tsoukaneri, M. Condoluci, T. Mahmoodi, M. Dohler, and M. K. Marina, "Group communications in Narrowband-IoT: Architecture, procedures, and evaluation," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 1–10, 2018.

[22] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran, "Radio resource management scheme in NB-IoT systems," *IEEE Access*, vol. 6, pp. 15051–15064, 2018.

[23] A. Rico-Alvarino, R. Lopez-Valcarce, C. Mosquera, and R. W. Heath, "FER estimation in a memoryless BSC with variable frame length and unreliable ACK/NAK feedback," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3661–3673, 2017.

[24] M. Jacobsson and C. Rohner, "Estimating packet delivery ratio for arbitrary packet sizes over wireless links," *IEEE Communications Letters*, vol. 19, no. 4, pp. 609–612, 2015.

[25] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, Springer, 2004.

[26] NS-3 Consortium, "ns-3 network simulator," 2018, https://www.nsnam.org/.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5817120, 10 pages
https://doi.org/10.1155/2018/5817120

*Research Article*

# Genetic Algorithm-Based Beam Refinement for Initial Access in Millimeter Wave Mobile Networks

**Hao Guo, Behrooz Makki, and Tommy Svensson**

*Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 41258, Sweden*

Correspondence should be addressed to Hao Guo; hao.guo@chalmers.se

Received 29 December 2017; Accepted 12 April 2018; Published 4 June 2018

Academic Editor: Shao-Yu Lien



Initial access (IA) is identified as a key challenge for the upcoming 5G mobile communication system operating at high carrier frequencies, and several techniques are currently being proposed. In this paper, we extend our previously proposed efficient genetic algorithm- (GA-) based beam refinement scheme to include beamforming at both the transmitter and the receiver and compare the performance with alternative approaches in the millimeter wave multiuser multiple-input-multiple-output (MU-MIMO) networks. Taking the millimeter wave communications characteristics and various metrics into account, we investigate the effect of different parameters such as the number of transmit antennas/users/per-user receive antennas, beamforming resolutions, and hardware impairments on the system performance employing different beam refinement algorithms. As shown, our proposed GA-based approach performs well in delay-constrained networks with multiantenna users. Compared to the considered state-of-the-art schemes, our method reaches the highest service outage-constrained end-to-end throughput with considerably less implementation complexity. Moreover, taking the users' mobility into account, our GA-based approach can remarkably reduce the beam refinement delay at low/moderate speeds when the spatial correlation is taken into account. Finally, we compare the cases of collaborative users and noncollaborative users and evaluate their difference in system performance.

## 1. Introduction

The next generation of cellular systems (5G) requires both higher data rates (in the order of 10-100 Gbps) and lower end-to-end latencies (down to 1 ms) than previous generations [1]. For this reason, it is aimed at utilizing frequency bands in the 30-300 GHz range in order to obtain sufficiently large bandwidths/data rates. Due to power limitation and high path loss at these frequencies, the coverage range is typically small so that highly directional transmissions are required for such millimeter wave (MMW) communications. On the other hand, the physical size of antennas at the MMW band is relatively small, such that large-scale beamforming can be performed in practice [2, 3]. Employing large-scale beamforming during the initial access (IA) procedure can be a good way to overcome the increased path loss experienced at higher frequencies (see Section 2 for literature review of the IA systems).

One of the most challenging tasks of IA is that the base stations (BSs) make omnidirectional cell searches with directional beams and at the receiver side the users choose their best beam direction to detect the BSs. Successful access means, e.g., that the received power or the signal-to-noise ratio (SNR) is beyond certain thresholds. After a basic connection is established, the BSs and the users can begin exchanging messages and implement a beam refinement procedure to further improve the beam directions and do additional control actions [4].

For example, the user mobility can be handled by beam refinement. With 5G, it is expected to access wireless networks not only at home or in the office, but also at moving speeds such as in a vehicle. In the moving scenario, the beam refinement process can keep tracking the beams by exploiting spacial correlations so that the computational delay can be remarkably reduced. Furthermore, for vehicular user equipment (VUEs), the system-level performance is improved if we allow a scheme using device-to-device (D2D) communications to enhance the links [5].

IA beamforming at MMW is different from the conventional one since it is hard to acquire the channel state information (CSI) at these frequencies. For this reason, codebook-based beamforming has been recently proposed

as an efficient method to reduce the dependency on CSI estimation/feedback [6, 7]. Also, several works have been presented on both physical layer and procedural algorithms of IA beamforming [8–15]. However, in these works either the algorithms are designed for special metrics, precoding/combining schemes, and channel models or the implementation complexity grows significantly by an increasing number of BSs/users. Moreover, the running delay of the algorithm has been rarely considered in the performance evaluation. On the other hand, generic machine learning-based schemes have been recently proposed which can be effectively applied for different channel models with acceptable implementation complexity [6, 7, 16–18].

In this paper, we study the effect of beam refinement on the performance of MMW networks. In our previous work, we proposed an efficient genetic algorithm- (GA-) based beamforming approach [18] which reaches almost the same performance compared to the exhaustive search with low complexity. Based on [18], the contributions of this paper are as follows. (1) We include the GA-based beam refinement at both the transmitter and the receiver side. Also, (2) we compare different machine learning-based analog beamforming approaches for the beam refinement during IA, including GA-based beamforming [18], Tabu search beamforming [16], link-by-link beamforming [17], and two-level codebook beamforming [6, 7] in large-but-finite multiuser multiple-input-multiple-output (MU-MIMO) MMW communication systems. Moreover, (3) we analyze the effect of various parameters such as the number of transmit/receive antennas, total power budget, and the power amplifier (PA) efficiency on the network performance. As opposed to the literature, we take the algorithm running delay into account. Thus, there is a trade-off between finding the optimal beamforming matrices and reducing the data transmission time slot, and the highest throughput may be achieved by few iterations. We study the system performance in terms of the end-to-end throughput with service outage constraints as well as the implementation complexity. (4) Furthermore, we evaluate and compare the performance of the considered algorithms under various mobile speed of the users. (5) Finally, we consider the case of collaborative users and compare the system performance in the cases with and without information exchanges among users.

Our results demonstrate that the running delay of the algorithms and power amplifier inefficiency affect the system performance remarkably, which should be carefully considered in the system design. Moreover, our proposed GA-based approach outperforms the considered state-of-the-art schemes, in terms of throughput, and reaches (almost) the same results as in the exhaustive search-based approach with fewer number of iterations. Furthermore, when taking the user mobility into account, the GA-based approach can remarkably reduce the algorithm running delay based on the beamforming results in the previous time slots. With collaborative users, the end-to-end throughput can be improved due to the data exchange by D2D links. Thus, the GA-based beamforming approach can be an appropriate candidate for IA in future wireless networks.

## 2. Literature Review

In this part, we present some related research work on IA. The reader familiar with the research area can skip this section and go to Sections 3–5 where we present the system model, the algorithm descriptions, and the simulation results, respectively.

Beamforming techniques at MMW bands have been considered in standard developments IEEE 802.15.3c (TG3c) [19], IEEE 802.11ad (TGad) [20], and ECMA-387 [21]. The problem formulation for IA beamforming at MMW frequencies is introduced in [8] where a fast-discovery hierarchical search method is proposed. Moreover, several design options for MMW IA are presented in [22], where the basic steps in the 3rd-Generation Partnership Project (3GPP) Long-Term Evolution (LTE) standard are used as references, and the overall delay of each design option as a function of the system overhead is evaluated. Then, [11] compares three approaches, namely, exhaustive search, two-step, and context information-based, in terms of miss-detection probability and discovery time. Another comparison work is presented in [12], where it is shown that different IA protocols have a trade-off between delay and average user-perceived throughput.

In [18], we introduce a genetic algorithm-based initial beamforming approach and evaluate the effect of the algorithm running delay on the network performance. There are also previous works using the GA-based selection approach in different communication networks. For instance, in [23] an efficient scheduling scheme is designed based on the genetic algorithm in the return-link of a multibeam satellite system. A turbo-like beamforming scheme based on the Tabu search algorithm is proposed in [16] to reduce both searching complexity and system overhead. A concurrent beamforming protocol, which we refer to as link-by-link beamforming, is presented in [17] to achieve high capacity in indoor MMW networks. Finally, for multistage beamforming, a tree-structured multilevel beamforming codebook is designed for MMW wireless backhaul systems in [6]. Also, in [7], a low-complexity multistage codebook is designed to support the IEEE 802.15.3c protocol. In [9], an exhaustive beam search method is proposed. Two beamforming schemes, namely, random-phase beamforming and directional beamforming, have been tested in [10] under the line-of-sight (LOS) channel conditions. A low-complexity beamforming scheme for initial user discovery is proposed in [13] where limited feedback-type codebooks are used. In [14], an accurate analytical framework for MMW system performance has been developed. Impact of obstacles on the cell search process is considered in [15] for the first time, and a geo-located context database is proposed to speed up the cellular attachment operations by storing and processing the information about the previous cell discovery attempts.

## 3. System Model

We consider a MU-MIMO setup with $M$ transmit antennas at a BS and $\tau$ multiantenna VUEs, each with $\beta$ antennas. As a result, there are $N = \tau \times \beta$ total antennas at the receiver side (see Figure 1). This is an extension of our work [18] with single

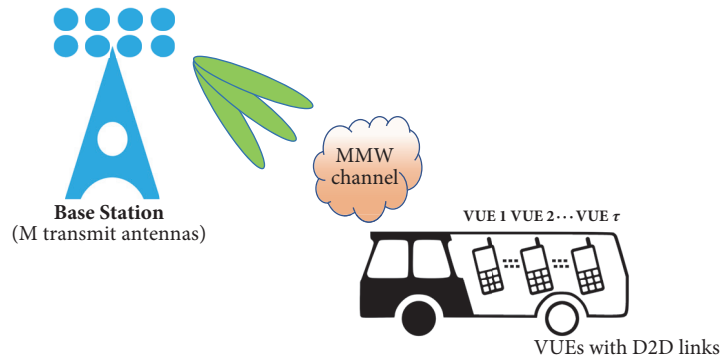



FIGURE 1: An illustration of MMW cooperative downlink communication of VUEs.

receive antennas and allows for beamforming at the receiver side. We assume that each user has perfect CSI. Also, as a more explicit model compared with [24], VUEs are allowed to exchange data with each other by using D2D links which is similar to the model in, e.g., [5]. We set $M > N$. At each time slot $t$, the aggregated received signal vector $\mathbf{y}(t)$ at time $t$ over the users after receive beamforming can be described as

$$\mathbf{y}(t) = \sqrt{\frac{P}{M}}\mathbf{U}(t)^H \mathbf{H}(t)\mathbf{V}(t)\mathbf{x}(t) + \mathbf{z}(t), \quad (1)$$

where $P$ is the total power budget, $\mathbf{H}(t) \in \mathscr{C}^{N\times M}$ is the channel matrix with the $(i, j)$th element given by $H_{i,j}(t) = d_{i,j}^{\gamma} h_{i,j}(t)$, where $d_{i,j}$ is the distance between the receiver antenna $i$ and the transmitter antenna $j$ and $\gamma$ is a path loss parameter, and $h_{i,j}(t)$ denotes the small scale fading. $\mathbf{x}(t) \in \mathscr{C}^{M\times 1}$ is the intended message signal, $\mathbf{V}(t) \in \mathscr{C}^{M\times M}$ is the precoding matrix at the BS, $\mathbf{U}(t) \in \mathscr{C}^{N\times N}$ is the aggregated combining matrix at the users' side, and $\mathbf{z}(t) \in \mathscr{C}^{N\times 1}$ denotes the independent and identically distributed (IID) Gaussian noise matrix. We assume channels remain the same during the whole algorithm running procedure. In this way, we can drop the time index $t$ in the following. In our algorithm we assume that each user can share their received signal in order to reach the optimal performance; i.e., $y_i$ is known by user $j$ with $j \neq i$. However, we also compare this user-collaborate scheme with the case that users have no collaborations; i.e., $y_i$ is not known by user $j$ with $j \neq i$.

Furthermore, the channel model $\mathbf{H}$ is described as

$$\mathbf{H} = \sqrt{\frac{k}{k+1}}\mathbf{H}_{\text{LOS}} + \sqrt{\frac{1}{k+1}}\mathbf{H}_{\text{NLOS}}, \quad (2)$$

where $\mathbf{H}_{\text{LOS}}$ and $\mathbf{H}_{\text{NLOS}}$ denote the line-of-sight and the non-line-of-sight (NLOS) components of the channel, respectively, and the NLOS component is assumed to follow a complex Gaussian distribution. Also, $k$ controls the relative strength of the LOS and the NLOS components. In (2), setting $k = 0$ represents an NLOS condition while $k \to \infty$ gives a LOS channel. We use this model because most cases in MMW systems have the LOS channel.

*3.1. Initial Beam Refinement Procedure.* Unlike a conventional beamforming procedure acquiring CSI, in MMW systems we suggest to perform codebook-based beam refinement, which means selecting a precoding matrix $\mathbf{V}$ out of a predefined codebook $\mathbf{W}_{\text{T}}$ at the BS while selecting a combining matrix $\mathbf{U}$ out of a predefined codebook $\mathbf{W}_{\text{R}}$ at the receiver side, sending test signal, and finally making decisions on transmit/receive beam patterns based on the users' feedback about their performance metrics. As the final step of IA [4], the beam refinement procedure can obtain a refined beam alignment at the cost of computational delay. The time structure for a packet transmission can be seen in Figure 2, where part of the packet period is dedicated to design appropriate beams in the IA procedure (mainly the beam refinement part) and the rest is used for data transmission. Thus, we need to find a balance between the beam design delay and the data transmission period by choosing an efficient approach.

Here, we use discrete Fourier transform- (DFT-) based codebooks [25] at both sides which are defined as

$$\mathbf{W}_{\text{T}} = \{w(m,u)\} = \left\{e^{-j2\pi(m-1)(u-1)/N_{\text{vec}}}\right\}, \quad m = 1,2,\ldots,M,\ u = 1,2,\ldots,N_{\text{vec}}, \quad (3)$$

for the BS, while

$$\mathbf{W}_{\text{R}} = \{w(n,u)\} = \left\{e^{-j2\pi(n-1)(u-1)/N_{\text{vec}}}\right\}, \quad n = 1,2,\ldots,N,\ u = 1,2,\ldots,N_{\text{vec}}, \quad (4)$$

for the users, where $N_{\text{vec}} \geq \max(M, N)$ is the number of codebook vectors. Note that since our algorithm is generic, one can apply our proposed algorithm for different kinds of codebooks.

*3.2. Performance Metrics.* The machine learning-based schemes of [6, 7, 16–18] are generic, in the sense that they can be implemented for different metrics. For the simulations, however, we consider the service outage-constrained end-to-end throughput, the complexity and the average number of required iterations as the system performance metric. In some scenarios, it may be required to serve the users with some minimum required rates; otherwise *service outage* occurs. In the $K$-th iteration round of the algorithm,

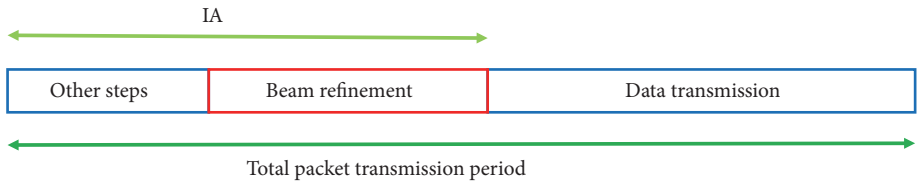



Figure 2: Schematic of a packet transmission period.

the service outage-constrained end-to-end throughput in bit-per-channel-use (bpcu) is defined as

$$R(K) = (1 - \alpha K) \sum_{i=1}^{\tau} r_i^K U\left(r_i^K, \log_2(1+\theta)\right), \quad (5)$$

where

$$r_i^K = \log_2\left(1 + \mathbf{SINR}_i^K\right), \quad (6)$$

$$U\left(r_i^K, \log_2(1+\theta)\right) = \begin{cases} 1 & r_i^K \geq \log_2(1+\theta) \\ 0 & r_i^K < \log_2(1+\theta). \end{cases} \quad (7)$$

Here, $r_i^k$ denotes the achievable rate of the user $i$ at the end of the $K$-th iteration. Also, parameter $\alpha$ is the relative delay cost for running each iteration of the algorithm which fulfills $\alpha N_{\text{it}} < 1$ with $N_{\text{it}}$ being the maximum possible number of iterations. Then, $\log_2(1+\theta)$ is the minimum per-user rate while $\theta$ represents the minimum required signal-to-interference-plus-noise ratio (SINR) of each user. Also,

$$\mathbf{SINR}_i^K = \frac{(P/M)\, g_{i,i}^K}{BN_0 + (P/M) \sum_{i \neq j}^{N} g_{i,j}^K} \quad (8)$$

is the SINR at the receiver of user $i$ in the iteration round $K$. Hence, we define the satisfied user as $\mathbf{SINR}_i^K \geq \theta$. Here, $g_{i,j}$ is the $(i,j)$-th element of the matrix $\mathbf{G}_K = |\mathbf{U}_K^H \mathbf{H} \mathbf{V}_K|^2$ which is referred to as the channel gain throughout the paper. Moreover, $B$ is the system bandwidth and $N_0$ is the power spectral density of the noise. We set $BN_0 = 1$ to simplify the system so that the power $P$ (in dB, $10\log_{10} P$) denotes the receiver side SNR as well.

The optimization problem of (5) is formulated as

$$\begin{aligned} \max_{K,\mathbf{U},\mathbf{V}} \quad & R(K) \\ \text{s.t.} \quad & \forall K \in \{1, 2, 3, \ldots, N_{\text{it}}\} \\ & \forall \mathbf{V} \subseteq \mathbf{W}_{\text{T}} \\ & \forall \mathbf{U} \subseteq \mathbf{W}_{\text{R}}. \end{aligned} \quad (9)$$

As opposed to, e.g., [17, Eq. 3], [22, Eq. 1], [26, Eq. 43], [27, Eq. 3], [28, Eq. 5] and [29, Eq. 5], we consider the algorithm running delay in the performance analysis. As seen in the following, there is a trade-off between optimizing beamforming matrices and reducing the data transmission period. In this case the optimal solution may be achieved by running the algorithms for a limited number of iterations.

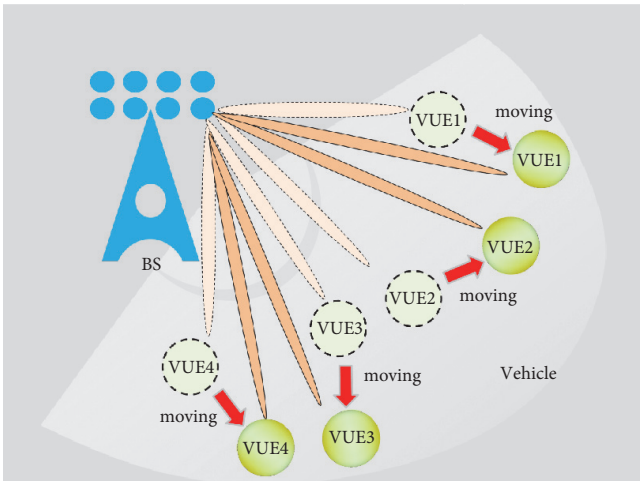



Figure 3: Mobility model, assuming that we know the moving distance for each user. The spatial correlation can be exploited by setting the queen of the previous time slot as one of the initial guesses of the next time slot.

*3.3. On the Effect of Power Amplifier Efficiency.* The efficiency of the radio-frequency high power amplifier (PA) should be taken into consideration in the multiantenna systems. Here, we consider the state-of-the-art PA efficiency model [30, Eq. 13], [31, Eq. 3]:

$$\rho_{\text{cons}} = \frac{\rho_{\max}^{\mu}}{\epsilon \times \rho_{\text{out}}^{\mu-1}} \quad (10)$$

where $\rho_{\text{cons}}$, $\rho_{\text{out}}$, $\rho_{\max}$ refer to as the consumed power, the output power, and the maximum output power of the PA, respectively. Also, $\epsilon \in [0,1]$ represents the power efficiency and $\mu \in [0,1]$ is a parameter depending on the PA class. Setting $\epsilon = 1$, $P_{\max} = \infty$ and $\mu = 0$ in (10) represents the special case (with an ideal PA).

*3.4. On the Effect of User Mobility.* Beamforming solutions for mobile users at high carrier frequencies are important in 5G wireless mobile communications. Here, we use the following mobility model to evaluate the performance of our proposed GA-based beamforming approach and compare the results with those of the considered state-of-the-art schemes. Consider Figure 3 with $\tau = 4$ multiple-antenna VUEs with data exchange D2D links. Here, we have two cases during the users' mobility.

*Case 1.* This case includes beam refinement with a random queen as initial guess (dash-line VUEs in Figure 3).

In each time slot with instantaneous channel realization $\mathbf{H} \in \mathscr{C}^{N\times M}$, do the followings:

(I) Initialization: Consider $L$, e.g., $L = 10$, sets of precoding matrices $\mathbf{V}_l$ and combining matrices $\mathbf{U}_l$, $l = 1, \ldots, L$, randomly selected from the pre-defined codebook $\mathbf{W}_{\mathrm{T}}$ and $\mathbf{W}_{\mathrm{R}}$.

(II) Selection: For each $\mathbf{V}_l$ and $\mathbf{U}_l$, evaluate the instantaneous value of the objective metric $R_l$, $l = 1, \ldots, L$, for example end-to-end throughput (5). Find the best beamforming matrix which results in the best value of the considered metric, named as the *Queen*, e.g., $\mathbf{V}_q$ and $\mathbf{U}_q$ satisfies $R(\mathbf{V}_l, \mathbf{U}_l) \leq R(\mathbf{V}_q, \mathbf{U}_q)$, $\forall l = 1, \ldots, L$ if the end-to-end throughput is the objective function.

(III) Save the Queen: $\mathbf{V}_1 \leftarrow \mathbf{V}_q$, $\mathbf{U}_1 \leftarrow \mathbf{U}_q$

(IV) Genetic operation I-Crossover: Create $S < L$, e.g., $S = 5$, beamforming matrices $\mathbf{V}_s^{\mathrm{new}}$ and $\mathbf{U}_s^{\mathrm{new}}$, $s = 1, \ldots, S$, around the Queen $\mathbf{V}_1$ and $\mathbf{U}_1$. These sets are generated by making small changes in the Queen $\mathbf{V}_q$ and $\mathbf{U}_q$.

(V) $\mathbf{V}_{s+1} \leftarrow \mathbf{V}_s^{\mathrm{new}}$, $\mathbf{U}_{s+1} \leftarrow \mathbf{U}_s^{\mathrm{new}}$, $s = 1, \ldots, S$.

(VI) Genetic operation II-Mutation: Regenerate the remaining sets $\mathbf{V}_s$ and $\mathbf{U}_s$, $s = S + 2, \ldots, L$, randomly with the same procedure as in Step (I).

(VII) Go back to Step (II) and run for $N_{\mathrm{it}}$ iterations, $N_{\mathrm{it}}$ is a fixed number decided by designer.

Return the final Queen as the beam selection rule for the current time slot.

Algorithm 1: GA-based beam refinement algorithm.

*Case 2.* This case includes beam refinement using the queen in Case 1 as initial guess (full-line VUEs in Figure 3).

By mobility we exploit the spatial correlation by setting the queen of the previous time slot as one of the initial guesses of the next time slot. For $d_{i,j}$ in (1) we assume that we know the moving speed $v$ and the time duration of mobility $\Delta t$. In this way, we can get an estimate of the user position in Case 2 in a circle whose radius is found by $v \cdot \Delta t$ with the user position at the previous time slot being the center.

## 4. Algorithm Description

In this study, we compare the performance of different IA beamforming methods as follows.

**Extended GA-based search [18]:** the algorithm starts by making $L$ possible beam selection sets at both transmitter and receiver, i.e., submatrices of each codebook. During each iteration, we choose the best set, named as the *Queen*, based on the performance metrics (for example, (5)). Next, we keep the queen and regenerate $S < L$ similar sets around the Queen by making small changes to the Queen (in the simulations, we replace 10% of the Queen columns randomly without loss of generality). Finally, the other $L - S - 1$ beamforming matrices are selected randomly to avoid the algorithm from being trapped in a local minima. Note that reducing $S$ for a given $L$ can increase the chance of being trapped. After $N_{\mathrm{it}}$ iterations (set by the designer), the queen is returned as the beam selection result in the current time slot. In this way, this is an extended version of our GA-based approach with beamforming at both the transmitter and the receiver, the basic principles of which can be found in Algorithm 1.

**Tabu search [16]:** The Tabu search approach follows the basic idea as in the GA-based scheme [16] where we choose and update the queen by iterations. The only difference is the evolution method of the queen in successive iterations. With Tabu, we use the definition of *neighborhood* in [16]: one matrix **A** is defined as another matrix **B**'s neighborhood if (1) **A** has only one different column compared with **B** or (2) the index difference between the two corresponding columns in **A** and **B** is equal to one. To make $S$ beam selection sets, we change the queen from previous round to its neighbors.

**Link-by-link search [17]:** in this strategy, the beam design of $\tau$ users is not optimized simultaneously. Instead, with a greedy approach, the beamforming solution is settled user-by-user by considering the interference from the other $\tau - 1$ links. The system performance improves in successive iterations until it converges to some (sub)optimal beamforming rules.

**Two-level search [6, 7]:** being inspired by multistage beamforming techniques, e.g., [6, 7], we design a two-level-codebook search scheme for our system. In the first level, the BS transmits messages over wider sectors using the codebook with $N_{\mathrm{vec}}/2$ columns, while in the second level it searches the optimal solution within the best such sector by steering narrower beams with an $N_{\mathrm{vec}}$-column codebook.

*4.1. On the Implementation Complexity.* To compare different methods, it is necessary that we consider the implementation complexity of each algorithm. For this reason, we derive the per-iteration complexity of different algorithms based on the fact that the product of matrices of size $N \times M$ and $M \times M$ has the complexity $\mathcal{O}(NM^2)$ in MATLAB. In this way, the per-iteration complexity for the GA-based approach is given by

$$C_{\mathrm{GA}} = L\left(2\mathcal{O}\left(N^2M\right) + \mathcal{O}\left(NM^2\right) + \mathcal{O}\left(NM\right)\right), \tag{11}$$

and $C_{\mathrm{Tabu}} = C_{\mathrm{GA}}$, $C_{\text{link-by-link}} = \tau \times C_{\mathrm{GA}}$, $C_{\text{two-level}} = 2 \times C_{\mathrm{GA}}$. $L$ is the number of beam selection sets within each iteration.

*4.2. On the Effect of User Collaboration.* In order to optimize the end-to-end system throughput (5), each user needs to share its received signal with the other users via the D2D links as mentioned in Section 2. Note that we do not consider the overhead of building up the D2D links in this work. We compare two cases regarding the user collaboration.

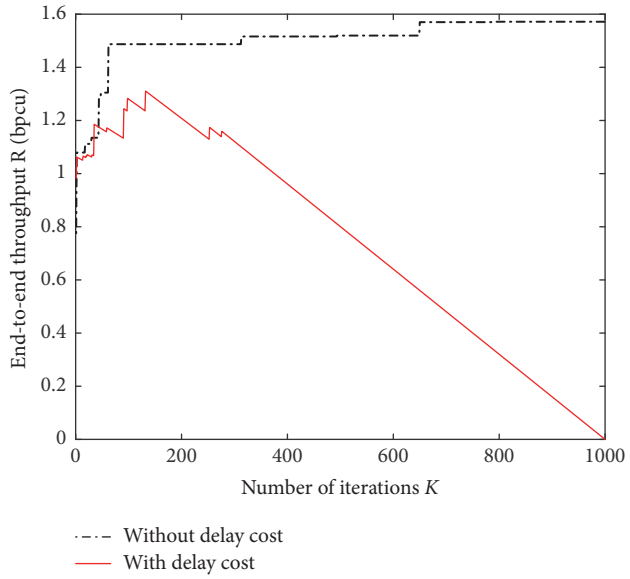



Figure 4: An example of the convergence process of the extended GA-based beamforming for systems with (subplot a) and without (subplot b) delay costs of the algorithm. $M = 32$, $\tau = 4$, $N = 12$, $P = -10$ dB, $k = 0$.

*Case 1* (collaborative users (CUs)). Each user knows the received signals of the other users and the system throughput is optimal.

*Case 2* (noncollaborative users (NCUs)). Each user only knows their own received message and the system throughput is suboptimal.

In Section 5, we evaluate the performance of the GA and the Tabu methods in these two cases and investigate the potential gains of collaboration.

## 5. Simulation Results

In the simulations, we use the channel model in (2) in the cases with $k = 0, 3$. We set $\mathbf{H}_{\text{LOS}} = \mathbf{1}_{N\times M}$ where $\mathbf{1}_{a\times b}$ refers to the normalized all-ones complex matrix. Except for Figure 4 which shows an example of the GA-based procedure, for each point in the curves the results are obtained by averaging over $10^4$ different channel realizations. In all figures, we set $N_{\text{it}} = 1000$ since it is a sufficiently large number of iterations after which no performance improvement is observed. Also, in all figures except for Figure 11, we use the normalized distance $d_{i,j} = d = 1$. Moreover, we set $L = 10$, $S = 5$ and $N_{\text{vec}} = 128$. In all figures, except for Figure 9, we use the ideal PA; i.e., set $P_{\max} = \infty$, $\mu = 0$, $\epsilon = 1$ in (10). In Figure 9 we study the effect of imperfect PAs. In Figures 4, 7, 9, and 10, we consider the service outage-constrained end-to-end throughput (5) as the performance metric with $\theta = -4$ dB. Finally, Table 1 shows the average number of required iterations in each algorithm to reach the (sub)optimal solution.

*On the convergence behavior:* Figure 4 gives an example of the GA performance in the cases with ($\alpha = 0.001$) and without costs of running the algorithm ($\alpha = 0$), respectively (see (5)). Here, example means we run our algorithm within one single channel realization. From Figure 4 we observe that very few iterations are required to reach the maximum throughput for the cases with delay cost, which is around $K = 130$. That is, considering the cost of running the algorithm, the maximum throughput is obtained by finding a suboptimal beamforming matrix and leaving the rest of the time slot for data transmission (see Figure 2). As a result, as the number of iterations increases, the cost of running the algorithm reduces the end-to-end throughput converging to zero at $K = 1/\alpha$ (see (5)). Note that the top value of the delay case is less than the other one due to the delay cost.

Table 1: Average number of required iterations $\overline{N}$ in different situations.

| $M/N$ | GA | Tabu | link-by-link | two-level |
|---|---|---|---|---|
| 32/12 | 502 | 498 | 307 | 501 |
| 32/8 | 500 | 501 | 288 | 498 |
| 32/4 | 488 | 502 | 261 | 500 |

If there is no running delay, on the other hand, the system performance improves with the number of iterations monotonically. However, the developed algorithm leads to (almost) the same performance as the exhaustive search-based scheme with very limited number of iterations. For example, with the parameter settings of Figure 4, our algorithm reaches more than 90% of the maximum achievable throughput with less than 100 iterations. On the other hand, with the parameter settings of Figure 4, exhaustive search implies testing in the order of $10^{30}$ possible beamforming matrices. Note that we cannot guarantee that the results are exactly the same with the optimal but because of the "random" part of the algorithm they become very close with large number of $N_{\text{it}}$. The trade-off between the performance and the delay cost is the concern here instead of the exact throughput value.

Finally, all considered schemes follow the same ladder-type convergence behavior as in Figure 4. This is because with the considered algorithms the system performance is not necessarily improved in each iteration and may be trapped into local minima. However, considering a couple of random solution checks in each iteration helps to avoid the local minima as the number of iterations increases.

*On the effect of service outage:* Figure 5 demonstrates the service outage-constrained end-to-end throughput (5) for different values of the required received SNR thresholds $\theta$ in (5). Also, Fig. 6 studies the service outage probability in the cases optimizing (5). Here, the results are presented for $N = 8$, $\tau = 8$, $M = 32$, $k = 0$, $N_{\text{vec}} = 128$, which means single-antenna user at the receiver side. As demonstrated in Figures 5 and 6, the service outage constraint affects the end-to-end and the per-user throughput significantly at low SNRs/severe service outage constraints. However, the effect of the service outage probability decreases as the SNR increases or $\theta$ decreases (Figures 5 and 6).

*Comparison of schemes:* in Figure 7, we compare the throughput (5) reached by the considered algorithms. It can be seen from the figure that for a broad range of SNRs the GA-based beamforming [18] leads to the best system throughput,

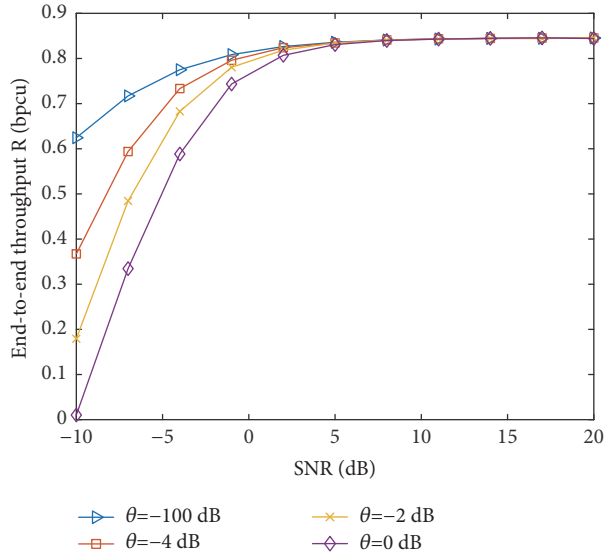



FIGURE 5: Service outage-constrained end-to-end throughput of the GA method with different $\theta$'s. $M = 32$, $\tau = 8$, $N = 8$, $k = 0$, $\alpha = 0$.

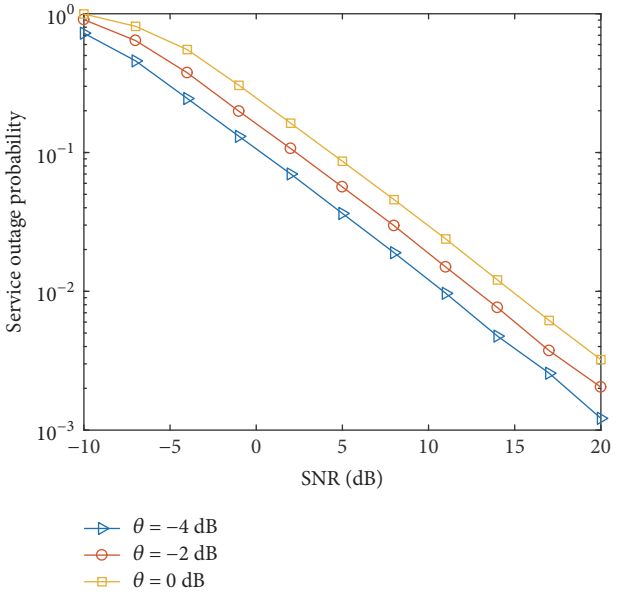



FIGURE 6: Service outage probability of the GA method with different $\theta$'s. $M = 32$, $\tau = 8$, $N = 8$, $k = 0$, $\alpha = 0$.

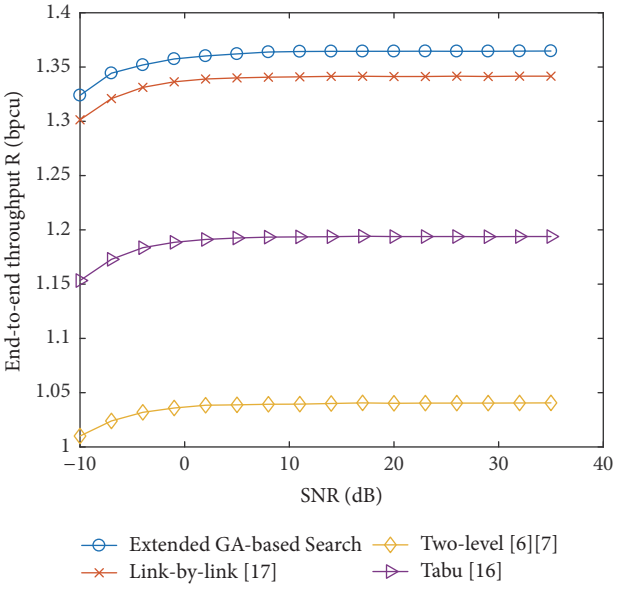



FIGURE 7: Service outage-constrained end-to-end throughput of different methods. $M = 32$, $\tau = 4$, $N = 12$, $k = 0$, $\alpha = 0$.

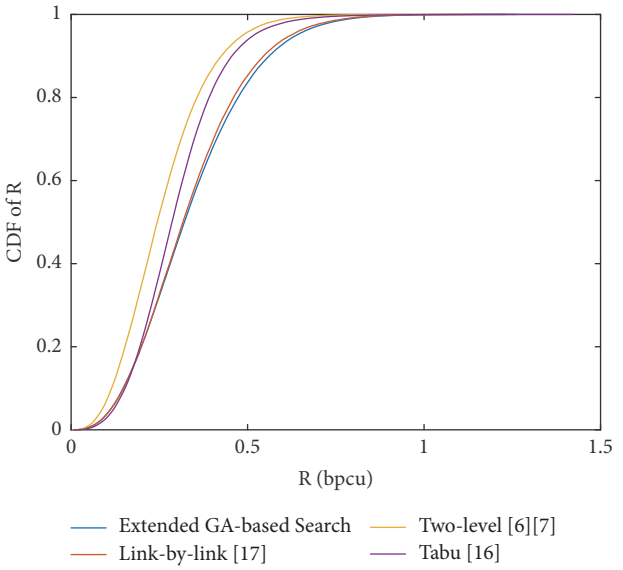



FIGURE 8: CDF of per-user throughput with different methods. $M = 32$, $\tau = 4$, $N = 12$, $k = 0$, $\alpha = 0$.

followed by the link-by-link search [17], Tabu search [16], and two-level search [6, 7].

Moreover, using the same parameter settings of Figure 7, in Figure 8 we compare the cumulative distribution function (CDF) of the per-user throughput (5) reached by the considered algorithms. From the figure we can see that the GA-based beamforming [18] leads to the best per-user throughput distribution, which means more users can be served by higher throughput, followed by the link-by-link search [17], Tabu search [16], and two-level search [6, 7].

Table 1 shows the average number of iterations $\overline{N}$ that is required in each scheme to reach a (sub)optimal solution. Here, the results are presented for $k = 0$, $M = 32$, $N = 4, 8, 12$. We can see that, in all methods, except for the link-by-link approach, the required number of iterations is almost insensitive to the number of receive antennas for the considered parameter setting of Table 1.

*On the effect of imperfect power amplifier:* Figure 9 evaluates the effect of the power amplifier on the throughput (5). We can see that the inefficiency of the PA affects the performance remarkably but this effect decreases with the SNR. This is reasonable because the effective efficiency of the PAs $\epsilon^{\text{effective}} = \epsilon(p_{\text{out}}/p_{\max})^{\mu}$ increases with SNR.

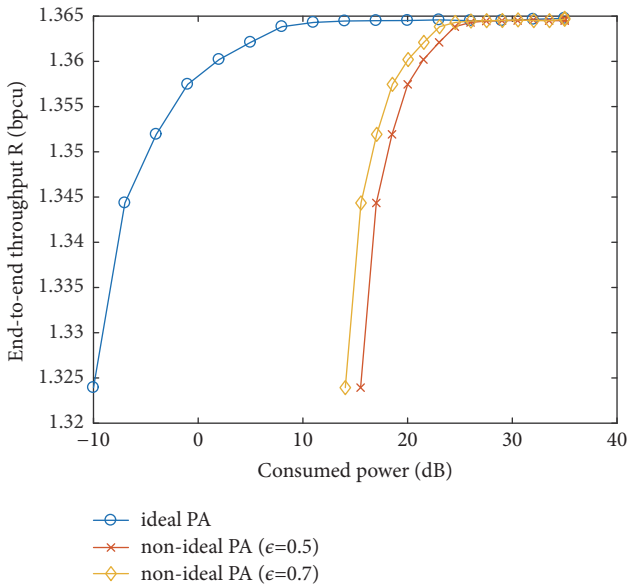



FIGURE 9: The effect of power budget and PAs efficiency on the end-to-end throughput (5). $M = 32$, $\tau = 4$, $N = 12$, $k = 0$, $\alpha = 0$.

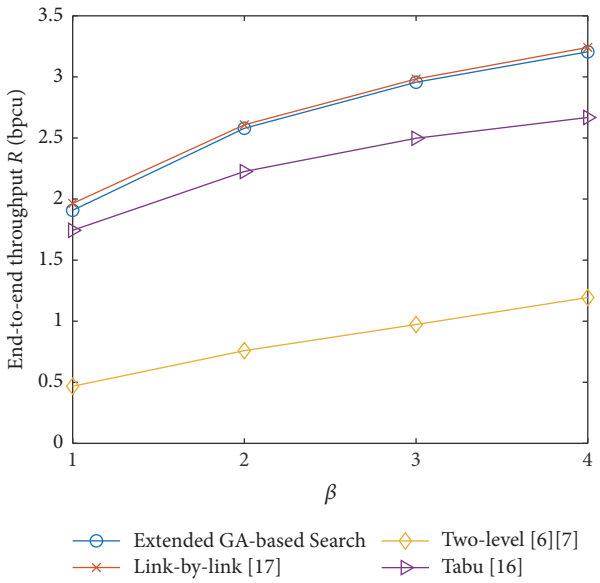



FIGURE 10: Throughput (5) with different number of receive antennas at the user side $\beta$. $M = 32$, $\tau = 4$, $\beta = 1, 2, 3, 4$, $k = 3$, $\alpha = 0$, $P = 2$ dB.

*On the effect of the number of receive antennas:* Figure 10 shows the effect of number of receive antennas per-user $\beta$ on the throughput (5). As seen in the figure, the end-to-end throughput increases with the number of per-user antennas as expected, since multiantenna techniques can improve the data rate remarkably. Moreover, the relative performance gain of the GA-based and the link-by-link scheme, compared to the other considered schemes, increases with the number of receive antennas, which is an interesting point when designing large-scale networks.

*On the effect of the user mobility:* Figure 11 shows the effect of the users' mobility on the beam refinement delay for

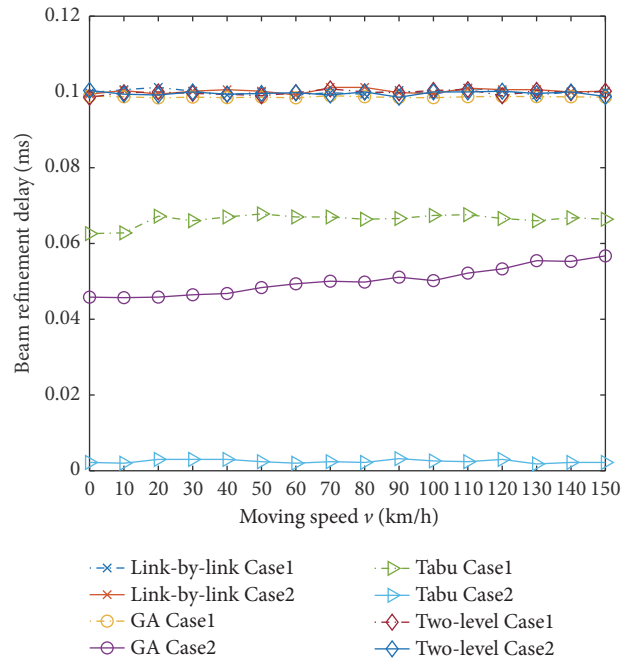



FIGURE 11: Beam refinement delay with different moving speed $v$. $M = 32$, $\tau = 4$, $\beta = 2$, $k = 0$, $\alpha = 0$, $P = 32$ dB, moving time $\Delta t = 1$ ms, $\gamma = -3.5$.

the considered algorithms. Inspired by [11], we evaluate the beam refinement delay (we assume that each iteration takes $10^{-4}$ overhead of $\Delta t$) of each algorithm in Cases 1 and 2 to check how well these algorithms are suitable for the mobile users. The algorithm running delays in Cases 1 and 2 of each method are all presented in the plot. Here, the results are presented with $M = 32$, $\tau = 4$, $\beta = 2$, $k = 0$, $\alpha = 0$, $P = 32$ dB, moving time $\Delta t = 1$ ms, $\gamma = -3.5$. As seen in the figure, both the GA-based algorithm and the Tabu-based algorithm can remarkably reduce the beam refinement delay for a broad range of users speeds, since they can use the beam refinement solution in Case 1 as the initial guess in Case 2 when the moving distance is not large. Note that Tabu search has the lowest delay in both cases since it simply changes the queen to its neighbors which takes full advantage of the spacial correlations. However, for GA-based scheme as the users speed increases the beam refinement delay increases slightly, intuitively because the spatial correlation between the positions in successive time slots decreases. Moreover, both the link-by-link search and the two-level-based search do not show noticeable performance gain.

*On the effect of collaborative users:* Figure 12 shows the effect of the users' collaboration on the end-to-end throughput for the GA and Tabu algorithms. Also, Table 2 presents the average number of required iterations for both the GA and the Tabu search in the cases with the CUs and the NCUs. Here, the results are presented with $M = 32$, $\tau = 4$, $\beta = 2$, $k = 0$, $\alpha = 0$. As seen in the figure, the performance of both the GA-based algorithm and the Tabu-based algorithm are reduced in the case of NCU. Also, these reductions decrease as the SNR increases. On the other hand, in Table 2 it can be seen that the NCU case requires

Table 2: Average number of required iterations $\overline{N}$ in different situations.

| $M/N$ | GA, CUs | GA, NCUs | Tabu, CUs | Tabu, NCUs |
|---|---|---|---|---|
| 32/12 | 502 | 1 | 498 | 1 |
| 32/8 | 500 | 1 | 501 | 1 |
| 32/4 | 488 | 1 | 502 | 1 |

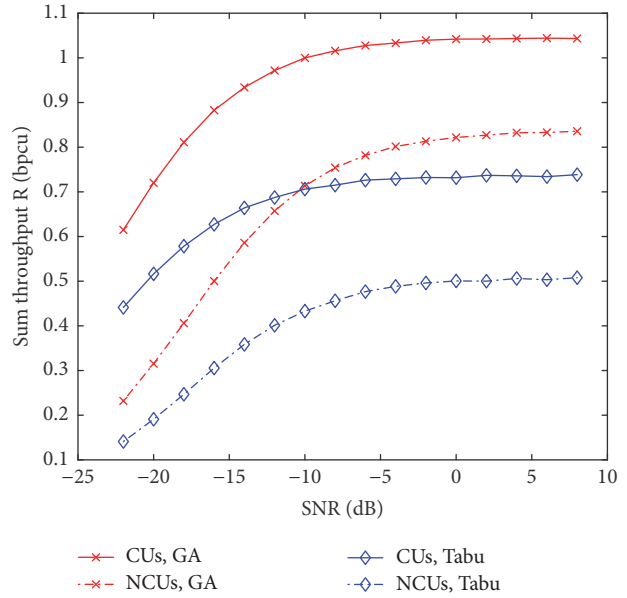



Figure 12: End-to-end throughput for different SNR with the CUs and NCUs cases for GA and Tabu. $M = 32$, $\tau = 4$, $\beta = 2$, $k = 0$, $\alpha = 0$.

much smaller iteration time compared with the CUs case for different system configurations. Only one iteration is required for the case with $\beta \leq 3$.

## 6. Conclusion

We extended our previously proposed genetic algorithm-(GA-) based beam refinement scheme to include beamforming at both the transmitter and the receiver, and we compared the performance with alternative beam refinement algorithms in an MU-MIMO system, in terms of the service outage-constrained end-to-end throughput and the implementation complexity. Particularly, our extended genetic algorithm-based scheme can reach almost the same throughput as in the exhaustive search-based approach with relatively few iterations in delay-constrained systems. Also, compared to the considered state-of-the-art schemes, our scheme leads to the highest throughput/per-user throughput and the lowest per-iteration implementation complexity, and the relative performance gain increases with the number of receive antennas. Moreover, non-ideal power amplifiers affect the system performance remarkably, which should be carefully considered during the system design. Furthermore, the GA-based approach can exploit the spatial correlation and remarkably reduce the beam refinement delay for a broad range of users speeds, which means it is an appropriate approach for mobile users. Finally, collaborative users can improve the system-level performance at the expense of computational complexity. For future work, we will investigate our proposed algorithm with more realistic parameter settings/scenarios and compare the result with other structured beamforming methods.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work is partly based on [32], and it has been partly performed in the framework of the H2020 Project 5GCAR cofunded by the EU. It has also been supported in part by VINNOVA (Swedish Government Agency for Innovation Systems) within the VINN Excellence Center ChaseOn. The authors would like to acknowledge the contributions of their colleagues.

## References

[1] M. Cudak, A. Ghosh, T. Kovarik et al., "Moving Towards Mmwave-Based Beyond-4G (B-4G) Technology," in *Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, June 2013.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[3] S. Sun, G. R. Maccartney, M. K. Samimi, S. Nie, and T. S. Rappaport, "Millimeter wave multi-beam antenna combining for 5G cellular link improvement in New York City," in *Proceedings of the 1st IEEE International Conference on Communications (ICC '14)*, pp. 5468–5473, IEEE, Sydney, Australia, June 2014.

[4] mmMagic Project D4.2, Final radio interface concepts and evaluations for mm-wave mobile communications, https://bscw.5g-mmmagic.eu/pub/bscw.cgi/d214055/mmMAGIC_D4.2.pdf.

[5] Y. Sui and T. Svensson, "Uplink enhancement of vehicular users by using D2D communications," in *Proceedings of the 2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 649–653, Atlanta, GA, USA, December 2013.

[6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proceedings of the 2011 IEEE Globecom Workshops*, pp. 253–257, Houston, TX, USA, December 2011.

[7] . Li Chen, . Ying Yang, . Xiaohui Chen, and . Weidong Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proceedings of the 2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)*, pp. 361–365, Harbin, China, August 2011.

[8] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmWave communications," in *Proceedings of the 2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 1926–1930, Pacific Grove, CA, USA, November 2014.

[9] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, 2015.

[10] Z. Abu-Shaban, H. Wymeersch, X. Zhou, G. Seco-Granados, and T. Abhayapala, "Random-phase beamforming for initial access in millimeter-wave cellular networks," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, usa, December 2016.

[11] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial Access in 5G mmWave Cellular Networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, 2016.

[12] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and J. Zhang, "On the Initial Access Design in Millimeter Wave Cellular Networks," in *Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Washington, DC, USA, December 2016.

[13] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming Tradeoffs for Initial UE Discovery in Millimeter-Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 543–559, 2016.

[14] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and C. J. Zhang, "Design and Analysis of Initial Access in Millimeter Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6409–6425, 2017.

[15] I. Filippini, V. Sciancalepore, F. Devoti, and A. Capone, "Fast Cell Discovery in mm-wave 5G Networks with Context Information," *IEEE Transactions on Mobile Computing*, 2017.

[16] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5731–5737, 2016.

[17] J. Qiao, X. Shen, J. W. Mark, and Y. He, "MAC-layer concurrent beamforming protocol for indoor millimeter-wave networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 327–338, 2015.

[18] H. Guo, B. Makki, and T. Svensson, "A genetic algorithm-based beamforming approach for delay-constrained networks," in *Proceedings of the 2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 1–7, Paris, France, May 2017.

[19] J. P. Gilb, *IEEE standards 802.15. 3cpart 15.3: wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs) amendment 2: millimeter-wave-based alternative physical layer extension [s]*, IEEE Computer Society, New York, NY, USA, 2009.

[20] C. Cordeiro et al., *IEEE P802. 11 Wireless LANs, PHY/MAC Complete Proposal Specification (IEEE 802.11-10/0433r2)*, 2010.

[21] H. Rate, "GHz PHY, MAC and PALs, Standard ECMA-387, ser," 2010, https://www.ecma-international.org/publications/files/ECMA-ST/ECMA-387.pdf.

[22] C. N. Barati, S. A. Hosseini, M. Mezzavilla et al., "Initial Access in Millimeter Wave Cellular Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7926–7940, 2016.

[23] B. Makki, T. Svensson, G. Cocco, T. De Cola, and S. Erl, "On the throughput of the return-link multi-beam satellite systems using genetic algorithm-based schedulers," in *Proceedings of the IEEE International Conference on Communications, ICC 2015*, pp. 838–843, gbr, June 2015.

[24] H. Guo, B. Makki, and T. Svensson, "A comparison of beam refinement algorithms for millimeter wave initial access," in *Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–7, Montreal, Canada, October 2017.

[25] L. Wan, X. Zhong, Y. Zheng, and S. Mei, "Adaptive codebook for limited feedback MIMO system," in *Proceedings of the 2009 IFIP International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–5, Cairo, Egypt, April 2009.

[26] J. Choi, "Beam Selection in mm-Wave Multiuser MIMO Systems Using Compressive Sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936–2947, 2015.

[27] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[28] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, "On the efficient beam-forming training for 60GHz wireless personal area networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 504–515, 2013.

[29] H.-H. Lee and Y.-C. Ko, "Low complexity codebook-based beamforming for MIMO-OFDM systems in millimeter-wave WPAN," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3607–3612, 2011.

[30] B. Makki, T. Svensson, T. Eriksson, and M.-S. Alouini, "On the Required Number of Antennas in a Point-To-Point Large-but-Finite MIMO System: Outage-Limited Scenario," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1968–1983, 2016.

[31] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input single-output capacity," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 913–919, 2014.

[32] H. Guo, *Initial Access in mm-wave 5G Mobile Communications [Master thesis]*, Chalmers University of Technology, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7852896, 9 pages
https://doi.org/10.1155/2018/7852896

*Research Article*

# RF Driven 5G System Design for Centimeter Waves

**Pekka Pirinen, Harri Pennanen, Ari Pouttu, Tommi Tuovinen, Nuutti Tervo, Petri Luoto, Antti Roivainen, Aarno Pärssinen, and Matti Latva-aho**

*Centre for Wireless Communications, P.O. Box 4500, FI-90014 University of Oulu, Finland*

Correspondence should be addressed to Pekka Pirinen; pekka.pirinen@oulu.fi

Received 22 December 2017; Accepted 12 April 2018; Published 23 May 2018

Academic Editor: Shao-Yu Lien



5G system design is a complex process due to a great variety of applications and their diverse requirements. This article describes our experiences in developing a centimeter waves mobile broadband concept satisfying future capacity requirements. The first step in the process was the radio channel measurement campaign and statistical modeling. Then the link level design was performed tightly together with the radio frequency (RF) implementation requirements to allow as large scalability of the air interface as possible. We started the concept development at 10 GHz frequency band and during the project World Radiocommunication Conference 2015 selected somewhat higher frequencies as new candidates for 5G. Thus, the main learning was to gain insight of interdependencies of different phenomena and find feasible combinations of techniques and parameter combinations that might actually work in practice, not only in theory.

## 1. Introduction

All mobile communication system generations have had a clear key application driver: 1G for analog voice, 2G for digital voice and text messaging, 3G for multimedia and Internet connectivity, and 4G for true mobile broadband [1]. 5G needs to cooperate and internetwork seamlessly with these legacy networks as long as they are in operation and provide added value to the ecosystem. Foreseen application areas for 5G are much broader than in earlier generations, and therefore, such qualities as system parameter adaptation, scalability, reconfigurability, virtualization, and self-organization become necessities for 5G.

This article overviews our approach to tackle the challenge of moving to higher center frequencies and address its implications to 5G system concept design. The starting point was to evaluate the new candidate frequencies in the centimeter wave bands that were investigated in METIS project [2–5]. Our selection at that point was 10 GHz as it was the lowest frequency and thereby it would be less risky and more predictable to go there than jump to much higher uncharted frequencies. Then we planned and completed some measurement campaigns to evaluate the propagation characteristics at 10 GHz. Also, to keep the concept development manageable with limited resources, we decided to have the main emphasis on the flexible and scalable enhanced mobile broadband communications air interface with a well-established orthogonal frequency-division multiplexing (OFDM) waveform. An integral part of the work has been a tight linkage to RF implementation issues and cross-checking the feasibility of different communications options simultaneously from the RF perspective. This kind of comprehensive view on the practical system concept development is far too often ignored and is therefore highlighted in this article.

Now, being aware of the World Radiocommunication Conference (WRC) 2015 decisions that the lowest beyond 6 GHz serious candidate frequencies have moved to above 24 GHz, the interest toward 10 GHz band is fading. However, irrespective of the actual center frequency many of the lessons learned are common and transferrable to any concept development process when utilizing previously unexplored spectral resources and planning a new ultra-scalable communications platform.

Our main contributions in this paper are as follows:

(i) Set performance targets and key metrics for 5G system design.

(ii) Provide high-level functional architecture of 5G network.

(iii) Present large-scale parameters derived from performed 10 GHz channel measurements.

(iv) Define flexible physical layer parameterization, signaling, and multiaccess structure.

(v) Elaborate RF link budget, beamforming, and practical implementation issues that are critical to 5G system design and performance at centimeter waves.

The rest of the article discusses first various 5G use cases and their respective system design objectives. Then, 5G system concept design is described from the network architecture viewpoint. The next two sections consider channel measurements and modeling issues and link level design issues. Implications of merging RF architecture design and considerations to the concept are presented next. Finally, some concluding remarks are given.

## 2. Use Cases and System Design Targets

A fully evolved 5G system needs to support diverse application areas such as *enhanced mobile broadband (eMBB)*, *massive Internet of things (MIoT)*, and *mission-critical communications (MCC)* [1]. All these use cases have distinct and partly contradictory requirements in terms of their key performance indicators, making the system concept design, as a whole, extremely complex. In most of the cases, not all of the requirements need to be simultaneously met. Thus, advanced 5G infrastructures move away from a "one architecture fits all" nature towards a "multiple architectures adapted to each service" concept. In this paper, the 5G system concept is mainly designed for eMBB, whereas the equally important use cases of MIoT and MCC have earned a fair share of attention as exemplified in [2, 6, 7]. In addition to extremely high throughputs, another main aspect of eMBB is the total system capacity. The ultra-high density of broadband user connections needs to be supported as well. New spectrum allocations, cell densification, and massive MIMO technology are seen as the key enablers to achieve these challenging goals.

The key design targets of the proposed 5G concept are presented below. These targets are the theoretical maximums that the system could support in ideal conditions.

(1) Support for scalable bandwidths up to 0.5-1 GHz in carrier frequencies around 10 and 30 GHz.

(2) Peak data rate that scales with system bandwidth, meaning tens of Gbps for 0.5-1 GHz bandwidths.

(3) Supported antenna and stream configurations:

  (a) Max. 256 transmit (TX) antennas and 16 receive (RX) antennas.

  (b) Max. 16 independent data streams.

(4) Spectrum efficiency

  (a) Max. 100 bits/s/Hz.

(5) Latency:

  (a) Control plane: < 10 ms to establish user plane.

  (b) User plane: < 1 ms from the user to server.

(6) Mobility:

  (a) Home and office, optimized for speeds < 5 km/h

  (b) Extreme mobility, speeds up to 500 km/h.

(7) Coverage:

  (a) Indoor coverage up to 30 m.

  (b) Outdoor coverage up to 300 m.

  (c) Operation > 300 m using lower cm-wave frequencies.

Theoretical values for maximum data rates are calculated with respect to the allocated spectrum in Section 5. The maximum multiple-input multiple-output (MIMO) configuration set for 5G system concept is 256 × 16 with the maximum number of 16 data streams. If this MIMO deployment can be achieved, the maximum spectrum efficiency level may be up to 100 bit/s/Hz given that 256QAM (quadrature amplitude modulation) is used. 256QAM requires error vector magnitude (EVM) levels of −33 dB, which sets rather stringent design targets for RF chain designs. The RF architecture aspects of 5G system concept are discussed in Section 6. One key target of 5G system is to define signaling structures, which enable very low delays in data transmission. The 5G systems should be able to provide 10 ms end-to-end (E2E) latency in general and 1 ms in extremely low latency use cases. The stringent latency requirements are addressed in the link level physical resource block design.

## 3. Network Architecture

The main target is to find critical solutions for a system concept suitable for a small cell deployment scenario of 5G. Furthermore, the design of 5G network architecture should be flexible. The current thinking is that there are such public spaces as stadiums and city centers wherein existing network operators shall have a role for deploying the access points and operating the networks. On the other hand, there are privately owned properties (housing blocks and shopping malls) where the network access points are privately/corporate owned or rented and the deployment and operation may be purchased from a new type of network operator. The following assumptions are used in the network level system design:

(i) Infrastructure sharing between operators allowed

(ii) Small cells and dedicated spectrum

(iii) Multioperator environment

(iv) Private networks, private access points

(v) Support for contention-based and scheduled resource usage

The generic functional architecture model for 5G system concept illustrating the basic functional entities of the control

and user plane in the device and network infrastructure parts is shown in Figure 1. The physical location of the functional entities in the infrastructure part can vary between radio nodes and more centralized units depending on the practical network deployment. The small cell radio networks are connected to common core cloud network ("EPC", Evolved Packet Core type network) with high capacity connections. Common core cloud network can serve multiple operators. The connectivity management related functionalities (for example, mobility management) are implemented in the core cloud network. The radio access network related functionalities could be distributed between the local radio network and the common core cloud network. The location of different functionalities, such as radio resource control and air interface management, may depend on the local RAN, core cloud network connection quality. In a general case, we can assume that air interface L1/L2 control is placed close to access points, while higher layers could be centrally located into the core cloud network. If very high speed connections (fiber cable) are available, then also the air interface L1/L2 control could be implemented into servers located in the core cloud network.

## 4. Radio Channel Models for System Design

Appropriate channel model is a starting point and mandatory for any system design. In the geometry-based stochastic channel model (GSCM), the propagation channel is characterized by statistical parameters obtained from the radio channel measurements. This gives a possibility of using the same framework of the model for the simulations in different frequencies and the different number or type of antennas. Due to missing characterizations of propagation channel at 10 GHz and in order to utilize GSCM promptly at 10 GHz frequency band, we carried out radio channel measurement campaigns with vector network analyzer and virtual arrays in the campus area of the University of Oulu. The measurements covered two different propagation scenarios, namely, two-story lobby and urban small cell scenarios. From the collected measurement data, complete parameterizations were derived for three-dimensional (3D) GCSM. The parameterizations are directly applicable to the 3rd Generation Partnership Project (3GPP) model [8].

The most important large-scale (LS) characteristics of the propagation channel are path loss and shadow fading. Based on our results, the path loss models are in some extent similar than in other frequency bands. However, the standard deviation of shadow fading $\sigma_{\mathrm{SF}}$ is significantly smaller due to static measured propagation environment. In addition to the path loss models and $\sigma_{\mathrm{SF}}$, the parameterizations consisted of 50 different propagation parameters. For example, LS parameters are modeled by log-normal distribution with specific mean $\mu$ and standard deviation $\sigma$ values giving higher level characterization of the propagation channel. Determined LS parameters are summarized in Table 1 and the full set of parameters is presented in [9, 10].

Although the model parameters are heavily dependent on the measured propagation environment, the following conclusions can be drawn from the determined LS and small-scale parameters:

(i) Parameters describing the delay and angular dispersion, that is, DS and angle spreads (ASs), seem to decrease in comparison to the parameters in the existing models at the frequency bands below 6 GHz due to the higher attenuation of delayed components.

(ii) When compared with lower frequency bands, specular reflection is more dominant propagation mechanism in comparison to diffuse scattering, leading to smaller cluster ASs.

Also, several research projects including industry and academia have been targeting to fulfill the requirements for designing and evaluating new channel models for the frequency bands up to 100 GHz. For instance, the initial parameterizations have been proposed for extending the quasi-deterministic radio channel generator (QuaDRiGa) over 10-80 GHz frequency band in [11]. Also, METIS project [12] addressed the challenges of the future channel modeling and recently developed a new map-based channel model up to 86 GHz as a pioneering work for 5G mobile communication system evaluations.

METIS model was intended to take into account all radio channel characteristics, which are important for any 5G mobile communications scenario. The model is based on the ray-tracing (RT) using a simplified 3D geometric description of the propagation environment. In the model, the building walls are modeled as rectangular surfaces with specific electromagnetic material properties, and the propagation paths are modeled deterministically. However, the model is not fully deterministic since the random objects representing for instance people and vehicles on the radio link are modeled stochastically. Therefore, the model can be understood as semideterministic model having significantly shorter processing time in comparison to traditional RT. Even though several properties of the model have already been successfully validated, the model still needs to be validated by additional measurements.

## 5. Link Level Design

*5.1. Physical Layer Design.* The physical layer design of the 5G system concept is based on the "OFDM signals with new numerology" approach. The signal structure has originally been designed to operate at 10 GHz band with bandwidths up to 1 GHz. Key OFDM signal parameters are selected so that synergy remains with the existing Long Term Evolution (LTE) radio implementations. Also, we shall use the same channel coding solutions as with LTE, where appropriate. However, in general the backwards compatibility with LTE is not maintained, since the 5G requirements lead to different physical layer designs when aiming to optimize the system performance.

The subcarrier spacing has been selected as 120 kHz. With subcarrier spacing of 120 kHz, the useful symbol duration becomes 8.3 $\mu$s. The cyclic prefix should be short compared to the symbol duration, but long enough to eliminate impairments in the OFDM signal detection due to propagation channel.

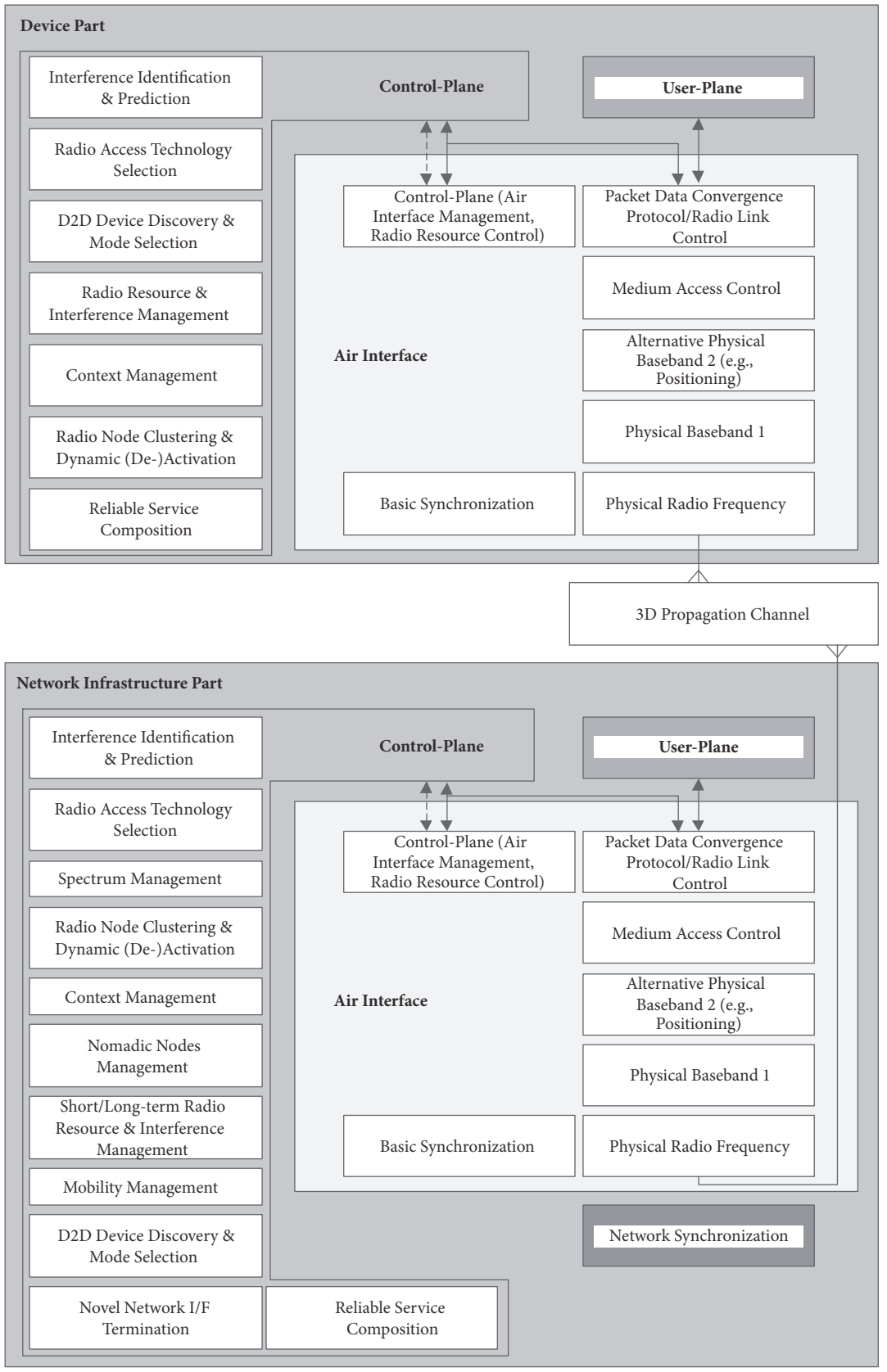



Figure 1: Functional network architecture.

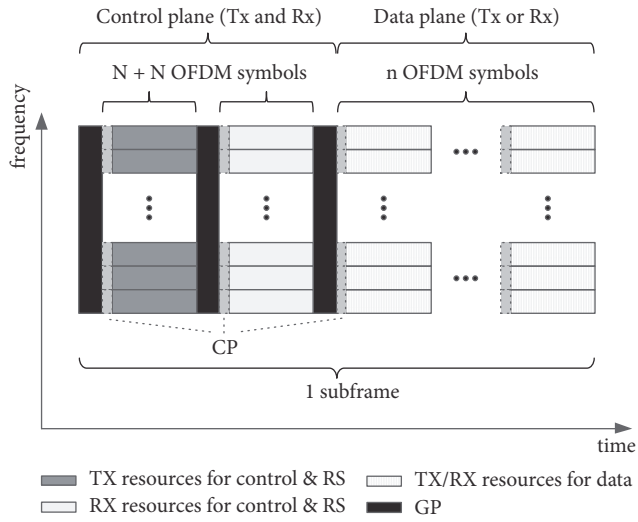



Figure 2: Physical layer subframe structure.

For the small cells (cell radius in the order of 75 m), cyclic prefix (CP) duration of 0.5 $\mu$s is large enough to prevent overlapping of OFDM symbols due to propagation delays even if no timing alignment is used in the UL transmissions. Measured delay spread values indicate that a typical delay spread at 10 GHz frequency with short transmission distances in indoor and outdoor environments is below 50 ns. For these reasons, we conclude that CP duration of 0.5 $\mu$s is enough for the 5G small cell operation scenarios. Table 2 shows alternative values for the 5G physical layer parameters, covering several (but not all possible) channel bandwidths between 31.25 MHz and 1000 MHz.

Given the above physical layer design, we have also computed the achievable data rates for several parameter sets as indicated by Table 2. The maximum data rates for a given bandwidth are achieved for 256QAM, 16 MIMO data streams, and coding rate $r = 1$, while the minimum data rates are achieved for BPSK, 1 data stream, and coding rate $r = 1/2$. We assume 90% bandwidth efficiency and rather optimistic protocol efficiency of 100%.

As the duplexing method we have chosen asymmetric dynamic time division duplexing [13]. In this particular approach, the uplink and downlink capacities may be chosen based on the traffic need within each cell, interference mitigation, and/or management requirements and user densities.

*5.2. Multiple Access Design.* The 5G subframe structure is shown in Figure 2. One subframe contains 11 OFDM symbols in time domain. In the control plane ($n = 2$ symbols), we introduce a guard period of $T_{GP} = 0.94$ $\mu$s, and the cyclic prefix is $T_{CP} = 0.5$ $\mu$s. The OFDM symbol duration is $T_{symbol} = 8.33$ $\mu$s. Therefore, for the subframe duration, we get $T_{subframe} = 3 \cdot T_{GP} + 11 \cdot (T_{symbol} + T_{CP}) = 100$ $\mu$s.

In the case where the system does not request extremely small E2E latencies (1 ms), the amount of control overhead can be reduced by concatenating multiple subframes together. In this case, the TX and RX control parts are embedded into the first subframe with 9 data symbols, while the remaining subframes (11 symbols) contain only data plane signal (either TX or RX) with first three symbols in each concatenated subframe having an extended cyclic prefix of $T_{CP,EXT} = 1.44$ $\mu$s.

We propose that the system shall have random access (contention-based) and scheduled resources. For the scheduled resources, we shall use orthogonal frequency-division multiple access (OFDMA), where the resource block (RB) size is a compromise between high granularity (to support transmission of very low amount of data) and signaling overhead. The minimum RB size is selected here as 72 resource elements, consisting of 8 consecutive subcarriers and 9 data plane symbols (note: resource element is defined as 1 subcarrier and 1 data plane symbol). This is comparable to LTE resource block size (84 resource elements) and can work both with machine-type services and mobile broadband data.

Random access resources shall be used by simple IoT devices, which are constrained by small form factor and/or battery operation. The scheduled resources can be used by more complex IoT devices and especially mobile cellular users as well as mobile broadband customers. The sharing between the resources shall be handled by the spectrum manager described earlier.

In the contention-based medium access case, the Resource Coordination functionality of layered resource management provides the template frame to a cluster of nodes allowing contention-based access inside the cluster. Control signals are transmitted in time-frequency resources separated from the data resources.

## 6. Towards 5G RF Implementation

Large antenna arrays will be one of the key enablers for 5G RF implementations for both capacity and link range. In this section, we briefly discuss multibeam link budget and RF restrictions for implementing cm-wave multiantenna transceivers (TRXs). Further in-depth discussion on the subject is available in [15].

*6.1. Link Budget.* For achieving the target data rates in practice, the link budget must address at least the following:

(i) Capacity evaluations with different modulations and waveforms

(ii) Hardware assumptions including physical dimensions, power, noise, and nonlinearity

(iii) Partitioning of signal-to-noise ratio (SNR) budget for different parts of TX and RX

(iv) Multistream transmission and adaptive beamforming

(v) Spatial channel model

Practical RF link budget consists of optimizing tens of different parameters together for setting the design targets for TRX design. Furthermore, requirements are very dependent on the target scenario including waveform assumptions, propagation environment, required physical dimensions, and user positions. Table 3 presents an example of system level RF specifications for two different frequency bands at indoor

Table 1: Derived LS parameters from the channel measurements.

| Channel model parameter | | LOS | | NLOS | |
|---|---|---|---|---|---|
| | | Two-story lobby | Urban small cell | Two-story lobby | Urban small cell |
| DS $\log_{10}([s])$ | $\mu_{DS}$ | −7.78 | −7.70 | −7.55 | −7.41 |
| | $\sigma_{DS}$ | 0.13 | 0.16 | 0.17 | 0.14 |
| KF [dB] | $\mu_{KF}$ | 8.5 | 5.1 | N/A | N/A |
| | $\sigma_{KF}$ | 3.5 | 3.2 | N/A | N/A |
| SF [dB] | $\sigma_{SF}$ | 2 | 2 | 3 | 2 |
| ASD $\log_{10}([^\circ])$ | $\mu_{ASD}$ | 0.86 | 1.08 | 1.32 | 1.24 |
| | $\sigma_{ASD}$ | 0.23 | 0.35 | 0.23 | 0.32 |
| ASA $\log_{10}([^\circ])$ | $\mu_{ASA}$ | 1.44 | 1.47 | 1.64 | 1.77 |
| | $\sigma_{ASA}$ | 0.11 | 0.20 | 0.18 | 0.08 |
| ESD $\log_{10}([^\circ])$ | $\mu_{ESD}$ | 0.91 | 0.80 | 0.54 | 0.89 |
| | $\sigma_{ESD}$ | 0.31 | 0.17 | 0.49 | 0.07 |
| ESA $\log_{10}([^\circ])$ | $\mu_{ESA}$ | 0.61 | 1.12 | 0.82 | 1.08 |
| | $\sigma_{ESA}$ | 0.17 | 0.10 | 0.29 | 0.13 |

DS = root mean square delay spread; KF = Rician K-factor; SF = shadow fading; ASD = azimuth angle spread of departure; ASA = azimuth angle spread of arrival; ESD = elevation angle spread of departure; and ESA = elevation angle spread of arrival.

Table 2: Physical layer signal parameters for 5G system concept (LTE values as a reference).

| Property | LTE | 5Gto10G | | | |
|---|---|---|---|---|---|
| Channel BW [MHz] | 20 | 31.25 | 125 | 500 | 1000 |
| Subframe length [ms] | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Sampling frequency [MHz] | 30.72 | 30.72 | 122.88 | 491.52 | 983.04 |
| FFT size | 2048 | 256 | 1024 | 4096 | 8192 |
| Subcarrier spacing [kHz] | 15 | 120 | 120 | 120 | 120 |
| Occupied subcarriers | 1201 | 234 | 938 | 3750 | 7500 |
| Guard subcarriers | 847 | 22 | 86 | 346 | 692 |
| Occupied bandwidth [MHz] | 18.015 | 28 | 113 | 450 | 900 |
| DL BW efficiency | 90% | 89.9% | 90% | 90% | 90% |
| OFDM symbols/subframe | 7 | 11 | 11 | 11 | 11 |
| Symbol duration excl. CP [$\mu$s] | 66.7 | 8.33 | 8.33 | 8.33 | 8.33 |
| CP duration [$\mu$s] | 5.2/4.69 | 0.5 | 0.5 | 0.5 | 0.5 |
| Data rate on full BW Min … Max [Mbps] | | 13 … 3,391 | 53 … 13,592 | 211 … 54,340 | 422 … 108,680 |

LOS scenario. These very abstracted requirements must then be divided further for different parts of the TRX.

In practice, adaptive modulation and coding scheme defines the minimum SNR required at the RX input. The nonidealities of TX limit the achievable SNR with respect to absolute power level. Figure 3(a) shows an example of signal-to-noise plus distortion ratio (SNDR) simulated with OFDM/256QAM waveform and commercial linear power amplifier (PA) for 10 GHz. These results are then combined with EVM values of other parts of the TX. It is clearly observed that the achievable linear power and hence data rate is easily overestimated, if only some of the TX nonidealities are taken into account. RX is treated in a similar fashion. Furthermore, the overall SNR-budget must be distributed between TX and RX. In RX, the SNR is typically limited by the noise at lower signal levels, whereas other nonidealities including phase noise of the synthesizer, analog-to-digital converter (ADC) quantization noise, I/Q mismatches, and cochannel interference limit the SNR at higher power levels.

In MIMO systems, spatial channel model is required for link budget evaluations. Multistream link budget is determined based on MIMO beam-specific path gains, where each stream is handled as an independent link [16]. The required TX power for rank 1-4 data transmissions at 10.1 GHz in indoor LOS scenario is presented in Figure 3(b). The used TX configuration is given in Table 3 including the notations: base station (BS), uniform rectangular array (URA), mobile terminal (MT), half-power beamwidth (HPBW), uniform linear array (ULA), and output power (Pout).

*6.2. Beamforming Arrays for 5G.* One of the fundamental questions in cm-wave communications is the number of antennas. The direct consequence of increasing frequency is the decreased antenna aperture. Hence, we can increase the number of antennas while maintaining the same physical area and eventually provide more beamforming gain. Multiple antennas are not only needed for increasing the data rate, but fundamentally for providing any reasonable link range. High

Table 3: Example of system level RF specifications.

| Center frequency | 10.1 GHz | 26 GHz |
|---|---|---|
| Signal bandwidth | 500 MHz | 1000 MHz |
| Array configuration at BS | 8 × 2 URA | 16 × 4 URA |
| BS HPBW (azim, elev) | (15°, 90°) | (7°, 30°) |
| Array configuration at MT | 4-element ULA | 8-element ULA |
| MT HPBW (azim, elev) | (30°, 120°) | (7°, 120°) |
| TX EVM for uncoded 256 QAM | 2.2% | 2.2% |
| PA back-off (for OFDM signal) | 9.6 dB | 9.6 dB |
| Pout Peak per PA in (BS, MT) | (10, 0.1) W | (1, 0.01) W |
| Total Noise Figure of (BS, MT) | (8, 10) dB | (8, 10) dB |

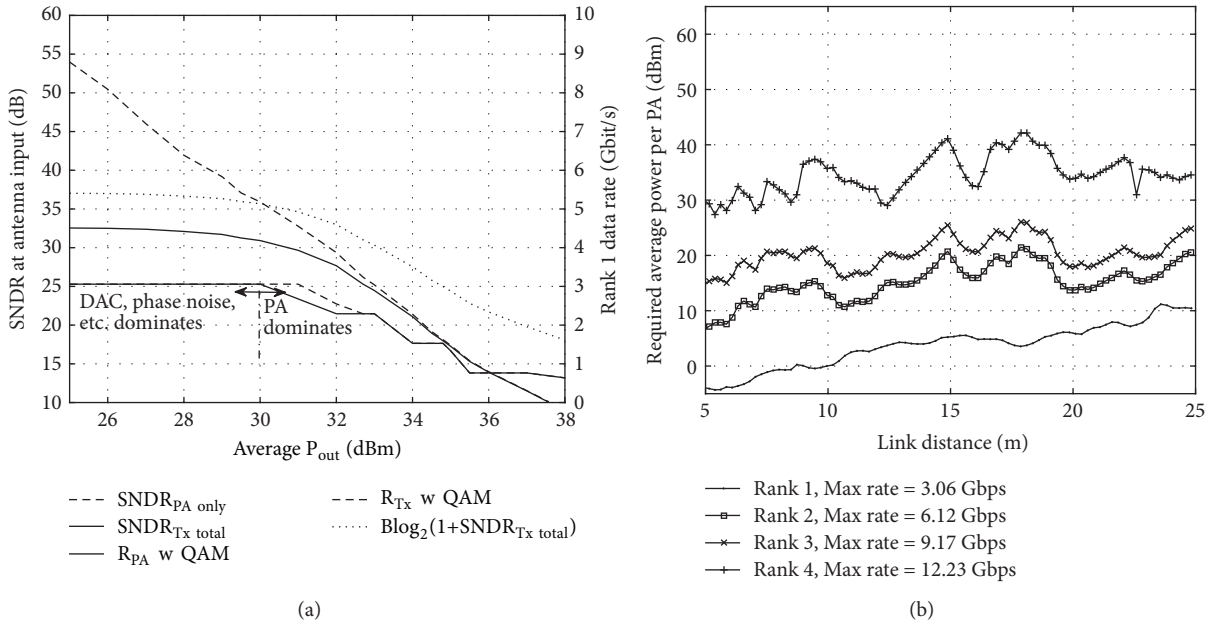



Figure 3: (a) Nonlinear PA and TX SNDR models with achievable data rates and (b) required average powers per PA for Rank N transmission with indoor LOS channel model [14] at 10.1 GHz.

directivity of the array results easily in very high effective isotropic radiated power (EIRP), which might be harmful for human tissue in case it is close to human body, limiting the maximum EIRP to be used. Decreasing antenna aperture also enables arrays for mobile terminals. However, the area of feeding network for large arrays limits the size of the antenna array in small form factors. Furthermore, because of higher circuit-level losses, the RF frontend must be embedded close to antennas to maintain the power efficiency.

In order to achieve the benefits of multiple antennas, beamforming system must be controlled adaptively. Traditionally, each antenna has individual RF chain and the control can be done in digital domain. However, because of extremely wide bandwidths, the TRX power consumption is not dominated only by analog components such as PA. Moreover, digital parallelism and wideband ADC's/digital-to-analog converters (DACs) are playing a crucial role when minimizing the power consumption. Because of these aspects, hybrid/RF beamforming is considered as de facto in cm-wave cellular systems. Adaptive RF phase and amplitude control of individual antenna elements is essential for maintaining the connection in adaptive user scenarios and minimizing the interference between data transmissions. However, the array scanning has an impact on the impedance matching of individual elements. Furthermore, the single element pattern affects the array scanning region. Hence, practical assumption of array scanning angle is in range of +/−30°.

*6.3. Practical Design Challenges.* The key challenges in cm-wave RF design are as follows:

(i) Wideband ADC/DAC dynamic range versus power consumption

(ii) Synthesizer phase noise

(iii) Linear (enough) output power with cm-wave PAs

(iv) Implementing high-efficiency PA array

(v) Physical form factors in antenna-RF integration

(vi) RF- and hybrid-beamforming array design.

The required sum power will be produced with several PA elements, resulting in decreased power per PA. Furthermore, multiple signals with different power levels in PA input give practical constraints for beam synthesis and power allocation per PA. These aspects give new tradeoffs for PA technologies although producing power at any of the options from CMOS to GaN will be a major challenge at high frequencies. The practical PA solution must be cheap, power efficient, linear, and small. However, these requirements cannot be optimized independently. Power efficient PA architectures, such as Doherty [17], are physically larger, cost more, and require linearization, which is traditionally done by digital predistortion (DPD). For RF/hybrid-beamforming arrays, conventional DPD is not possible because the waveform at each PA input cannot be controlled.

Different use scenarios set very different requirements for the RF implementation, especially in terms of required power, linearity, and beamforming. From the array scanning perspective, a convenient location of indoor BS is in the corner of the room/office area. A typical room layout indicates that it might be beneficial to have wider beams in the elevation than in the azimuth domain for beam/user separation. For outdoor BSs, high array gain for cell-edge users is a necessity that results in narrow beams. This complicates the beam scanning and tracking when serving mobile users. On the other hand, benefits of spatial filtering in network level management become apparent.

As the BS array design can be site-specific, the MT must adapt to various propagation scenarios. At 10 GHz, the practical number of elements varies between 2 and 8 with $\lambda/2$ antenna spacing in small devices. Hence, a linear array is considered to be the only practical solution, because the impact of mechanics forces the array to be designed at the bottom or top end of a device. However, other device types such as tablets and laptops may contain more antennas.

## 7. Conclusions and Way Forward

Various 5G system concept design aspects at centimeter waves were discussed. As the main use case, we selected eMBB built upon OFDM and set the key performance targets for it. One of the first objectives was to measure and characterize the radio channel in the chosen new cm-frequency. Based on the channel characteristics, link level design was carried out. It provides great adaptivity in air interface parameter setting, access schemes, and duplexing. Feasibility of the design objectives was checked in parallel with RF implementation aspects so that, for example, multiantenna beamforming, link budget, and power efficiency were be taken as integral system design elements. It was concluded that the high data rates achieved by high-order modulations and MIMO set extreme challenges for the nonlinearity, noise, and physical form factors of the RF devices. Based on this study, it is clear that RF design should be always the key driver when designing 5G solutions for centimeter waves and above that.

Millimeter-wave communications [18–22] are gaining increasing interest in 5G as the largest chunks of spectra are available at the bands above 30 GHz. Due to broader contiguous bands the link level design can be somewhat relaxed and yet very high data rates are achievable. Propagation and penetration losses tend to increase with frequency, limiting the feasible link range. However, at the same time antenna size and spacing go down enabling easier deployment of large MIMO systems with high array gain and beam directivity. Therefore, hybrid-beamforming architectures and power- and cost-efficient RF transceiver design remain in the focal role when moving from centimeter waves to millimeter-waves.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This article is mostly based on the research work in the "5G Radio Access Solutions 10 GHz and Beyond Frequency Bands (5Gto10G)" project over the years 2014-2017. Project partners Bittium, Huawei, Keysight, Nokia, and Tekes are hereby gratefully acknowledged for their support.

## References

[1] Qualcomm. Technologies, "Inc., Making 5G NR a reality Leading the technology innovations for a unified, more capable 5G air interface," *White Paper*, 2016.

[2] Mobile., "Mobile and Wireless Communication Enablers for the Twenty-twenty Information Society (METIS)," in *FP7-ICT-317669 Project*, https://www.metis2020.com.

[3] A. Osseiran, F. Boccardi, V. Braun et al., "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[4] H. Tullberg, P. Popovski, Z. Li et al., "The METIS 5G System Concept: Meeting the 5G Requirements," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 132–139, 2016.

[5] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ö. Bulakci, "METIS research advances towards the 5G mobile and wireless system definition," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–16, 2015.

[6] FANTASTIC-5G, "5GPPP Project," http://fantastic5g.com.

[7] C. Bockelmann, N. Pratas, H. Nikopour et al., "Massive machine-type communications in 5g: Physical and MAC-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59–65, 2016.

[8] "a 3rd Generation Partnership Project (3GPP)," Study on channel model for frequency spectrum above 6 GHz TR 38.900 v14.0.0", 2016.

[9] A. Roivainen, C. Ferreira Dias, N. Tervo, V. Hovinen, M. Sonkki, and M. Latva-aho, "Geometry-based stochastic channel model for two-story lobby environment at 10 GHz," *Institute of Electrical and Electronics Engineers. Transactions on Antennas and Propagation*, vol. 64, no. 9, pp. 3990–4003, 2016.

[10] A. Roivainen, P. Kyösti, C. F. Dias et al., "Parametrization and validation of geometry-based stochastic channel model for urban small cells at 10 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 7, pp. 3809–3814, 2017.

[11] "H2020-ICT-671650 Millimetre Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications (mmMAGIC) project," Measurement campaigns and initial channel models for preferred suitable frequency ranges Deliverable D2.1, 2016.

[12] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter Wave Cellular Networks: A MAC Layer Perspective," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, 2015.

[13] P. Jayasinghe, A. Tolli, and M. Latva-aho, "Bi-directional signaling strategies for dynamic TDD networks," in *Proceedings of the 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 540–544, Stockholm, Sweden, June 2015.

[14] T. Tuovinen, N. Tervo, and A. Parssinen, "RF system requirement analysis and simulation methods towards 5G radios using massive MIMO," in *Proceedings of the 2016 46th European Microwave Conference (EuMC)*, pp. 142-45, London, United Kingdom, October 2016.

[15] T. Tuovinen, N. Tervo, and A. Parssinen, "Analyzing 5G RF System Performance and Relation to Link Budget for Directive MIMO," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6636–6645, 2017.

[16] T. Tuovinen, N. Tervo, and A. Pärssinen, "Downlink Output Power Requirements with an Experimental-Based Indoor LOS/NLOS MIMO Channel Models at 10 GHz to Provide 10 Gbit/s," in *Proceedings of the 46th European Microwave Conference, EuMC 2016*, pp. 505–508, gbr, October 2016.

[17] R. S. Pengelly, "The Doherty power amplifier," in *Proceedings of the 2015 IEEE MTT-S International Microwave Symposium (IMS2015)*, pp. 1–4, Phoenix, AZ, USA, May 2015.

[18] T. Rappaport, S. Sun, R. Mayzus et al., "Millimeter wave mobile communications for 5G cellular: it will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[19] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[20] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[21] M. Xiao, S. Mumtaz, Y. Huang et al., "Millimeter Wave Communications for Future Mobile Networks (Guest Editorial), Part I," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1425–1431, 2017.

[22] L. Li, D. Wang, X. Niu et al., "mmWave communications for 5G: implementation challenges and advances," *Science China Information Sciences*, vol. 61, 2018.