IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 18, 2020, accepted March 8, 2020, date of publication March 19, 2020, date of current version March 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982068

mmWave-NOMA-Based Low-Latency and High-Reliable Communications for Enhancement of V2X Services

BICHENG WANG, (Student Member, IEEE), RUOQI SHI, FANWEI SHI, AND JIANLING HU, (Member, IEEE)
School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China

Corresponding author: Jianling Hu (jlhu@suda.edu.cn)

This work was supported by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJKY19_2285.

ABSTRACT The enhanced vehicle-to-everything (eV2X) systems have stricter constraints of latency and reliability than before. To meet the low latency and high reliability (LLHR) demands, the investigation of semi-persistent scheduling (SPS) is of great significance. For the first time, we integrate millimeter wave (mmWave) and non-orthogonal multiple access (NOMA) into SPS strategy. The proposed strategy includes the beam division and user clustering process, the power allocation process, and the vehicular user equipment (VUE) – resource block (RB) matching process. Firstly, the joint-optimization problem in terms of user scheduling indicators and power allocation factors is formulated to minimize the SPS period. This problem is then proved to be non-deterministic polynomial hard (NP-hard) by being reduced into a vertex coloring problem. To solve it with polynomial complexity, we divide it into two subproblems and calculate them by iteration. For one subproblem, the LLHR power control algorithm is proposed to solve the non-convex reliability-optimization problem, which can provide an evaluation indicator for the user scheduling. For the other subproblem, the beam division and VUE clustering algorithm is designed to reduce the complexity of the VUE-RB matching. After that, the matching problem with peer effects is solved by the proposed union-based matching algorithm. The analytical results on stability, convergence and computational complexity are presented, and simulation results show that the proposed SPS strategy outperforms the existing schemes.

INDEX TERMS Enhanced V2X, NOMA, mmWave, hybrid precoding, semi-persistent scheduling.

I. INTRODUCTION

Intelligent transport system (ITS) has been one of the highly concerned transmission systems in the fifth generation (5G) enabled artificial intelligence (AI) era, since it can provide security, transport efficiency, and energy conservation [1]. To meet the new requirements of 5G ITS, the 3rd generation partnership project (3GPP) has proposed an enhanced vehicle-to-everything (eV2X) network [2], [3], which contains two communication modes, i.e., PC5 interface for sidelink (SL) [4] and Uu interface for downlink/uplink (DL/UL) [5]. Particularly, the novel use cases in eV2X networks have more critical low latency and high reliability (LLHR) constraints. For instance, the advanced driving services require vehicular user equipment (VUE) to implement mobile AI [6], where the Uu interface is utilized to transmit periodic packets.

Such LLHR use cases are generally time-triggered rather than accidental [7], which is different from the event-triggered services in conventional cellular networks. It has been proved that semi-persistent scheduling (SPS) performs better for time-triggered services than dynamic scheduling (DS), since it can greatly reduce the signaling overhead [8]. Moreover, the centralized SPS utilizes global channel state information (CSI) to allocate resource blocks (RBs) for VUEs [9]. Hence, it has higher spectrum utilization and is better for collision avoiding compared with the distributed SPS.[1] However, eV2X networks also require higher data rate, e.g. Gbps, as introduced in [6], [10]. Limited by spectrum

The associate editor coordinating the review of this manuscript and approving it for publication was Tariq Umer.

[1]For distributed SPS, several VUEs are allocated with the same RB since only partial CSI can be obtained. At this point, collision avoidance algorithm will be utilized, but it may cause an indeterminate increase of latency.

resource, it is difficult for sub-6 GHz to guarantee the latency of all VUEs under LLHR constraints.

A. RELATED WORKS

To achieve higher data rate, millimeter waves (mmWave), e.g. 30 GHz and 63 GHz, are considered in eV2X networks [11], [12]. The wider spectrum bandwidth makes mmWave able to serve more VUEs than sub-6 GHz. Moreover, the shorter wavelengths are beneficial for the deployment of large-scale antenna systems, which can improve the beamforming gain [13]. However, the higher carrier-center frequency (CCF) also leads to the increase of interference and pass loss [14]. Hence, several improved beamforming schemes were designed for mmWave systems. In [15], a latency-optimal beam sweeping and alignment scheme was proposed for both the transmitter side and the receiver side in mmWave V2X systems, while in [6], an online learning algorithm in the cloud was proposed for beamforming with the ability of environmental awareness. Typically, in order to achieve a balance between radio frequency (RF) chain overhead and transmission performance, hybrid analog and digital precoding schemes are proposed for mmWave systems, including fully connected hybrid beamforming and sub-connected hybrid beamforming [16]. Specifically, a fully connected hybrid precoding structure consisting of a low-cost high dimensional analog precoder and a high-performance low dimensional digital precoder was proposed in [17], where each RF chain is connected to all antennas.

Benefit from the non-orthogonal nature in frequency domain, power-domain non-orthogonal multiple access (NOMA) has been considered to solve the problem of carrier frequency offset (CFO) and frequent handovers of mmWave cells caused by vehicle mobility [18], [19]. It is worth mentioning that the pairing-based NOMA has been shown to achieve the spectral efficiency which is about 30% higher than orthogonal multiple access (OMA) under a given bandwidth [20]. This advantage facilitates the transmission of large data packets in a limited RB. Moreover, power domain multiplexing by NOMA can be an effective complement to space division multiplexing by MIMO [21]. Recently, some MIMO-NOMA schemes in cellular networks have been investigated. A low-complexity singular value decomposition (SVD) beamforming scheme based on multi-user CSI was proposed for 5G MIMO-NOMA in [22]; a joint precoding and dynamic power control scheme was proposed for mmWave-NOMA systems with lens antenna arrays in [23]; user selection and power allocation for mmWave-NOMA networks were investigated in [24] and [25]. However, the above solutions [22]–[25] were designed to maximize the sum of users' achievable rates. Different from them, the key performance indicator (KPI) of eV2X systems is considered to be latency rather than system throughput [12], [26]. Additionally, considering the advantages of high spectrum utilization, MIMO-NOMA has been widely studied for conventional cellular services in [27]–[29], e.g., a detailed analysis of channel capacity in mmWave massive MIMO-NOMA systems was given in [27]. Additionally, in [30], a precoding scheme was proposed to increase the difference of channel gain between user 1 and user 2 in a given NOMA pair, and two schemes of power allocation were proposed with consideration of user 1's transmission outage probability and data transmission rate, respectively. However, the schemes in [27]–[30] are not suitable for the LLHR services, i.e., the ability to improve latency and reliability may be limited. Furthermore, it is difficult for such schemes to manage interference effectively in an eV2X system with dense vehicle distribution. For instance, although the system throughput can be improved with the proposed mmWave-NOMA scheme in [28], the weak users have a much lower transmission rate than the strong users, which leads to a risk of reliability. Therefore, the LLHR scenarios in eV2X systems need a novel design of mmWave-NOMA schemes.

B. MOTIVATION AND CONTRIBUTIONS

The motivation of this paper is to improve latency and reliability of the eV2X systems by utilizing MIMO-NOMA. Considering the KPI, eV2X systems prefer to utilize the least number of RBs to serve all VUEs in coverage under the constraints of quality of service (QoS), which is different from the event-triggered services. This view was also reflected in [7], where a centralized SPS strategy is designed for machine-type communications (MTC) to minimize the scheduled bandwidth. Specially, for eV2X systems, the minimization of SPS period is more feasible and reasonable, since the users have a uniform constraint of latency and reliability [26]. Additionally, to maximize the packet reception rate (PRR), a NOMA-based hybrid centralized/distributed SPS strategy was proposed in [9], which was designed for broadcast communication.

In this paper, a centralized SPS scheme based on mmWave massive MIMO and NOMA is proposed to meet the LLHR demands of eV2X systems, which includes hybrid precoding, user scheduling, and resource allocation. Particularly, the proposed strategy optimizes the SPS period under the constraints of reliability by searching and calculating two parameter sets, i.e., power allocation factors and vehicle scheduling indicators. The main contributions are summarized as follows:

- A novel centralized SPS strategy is proposed for the LLHR eV2X systems. The purpose of this strategy is to reduce the SPS period, which is the main component of scheduling latency [26], as well as to improve the lower bound of transmission reliability of VUEs that share the same RB in each time slot. To achieve this purpose, we introduce mmWave massive MIMO-NOMA to extend the freedom of each VUE, which makes the VUEs covered in eV2X systems finish the data transmission with a smaller number of RBs under QoS constraints.
- To enable the mmWave massive MIMO-NOMA-based eV2X SPS strategy, the joint-optimization problem in terms of user scheduling indicators and power allocation

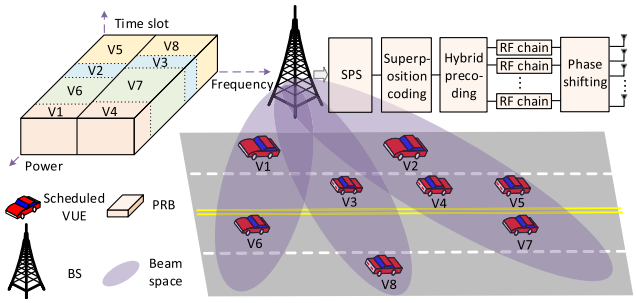


FIGURE 1. System model of cellular eV2X transmission.

factors is first formulated. This problem is proved to be non-deterministic polynomial (NP) hard, and then, three implementation processes are developed, including the power allocation process, the beam division and user clustering process, and the VUE-RB matching process.

- The stability, convergence and computational complexity in terms of the mmWave-NOMA-based SPS strategy designed in this paper are presented. For comparison, five typical schemes as well as the proposed one are evaluated by simulations. It is shown that the proposed scheme improves the SPS period compared with the existing schemes.

C. ORGANIZATION

The rest of the paper is organized as follows. In Section II, the system model and channel model of cellular V2X systems is described, and the formulation of scheduling problem is given. In Section III, we design the mmWave-NOMA-based SPS strategy and analyze stability, convergency, and complexity of the proposed scheme. Specifically, the LLHR power control (LLHR-PC) algorithm, the beam division and VUE clustering (BD&VC) algorithm and the union-based VUE-RB matching algorithm are proposed, respectively. Finally, simulation results are presented in Section IV and conclusions are drawn in Section V.

Notation: $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $||\mathbf{A}||$ denote the transpose, conjugate transpose, inversion and 2-norm of matrix $\mathbf{A}$, respectively. $\lfloor\cdot\rfloor$ denotes Gaussian rounding. $|\mathcal{N}|$ denotes the number of elements in $\mathcal{N}$. $\beta \sim \mathcal{CN}(u, \sigma^2)$ denotes that β obeys the complex Gaussian distribution with mean u and covariance σ^2.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, a cellular eV2X system model is first presented. Then, we propose and formulate the mmWave-NOMA scheme in order to reduce the latency caused by the SPS period.

A. SCENARIO DESCRIPTION

As is shown in Fig. 1, we consider a single-cell mmWave-NOMA-supported cellular V2X system. The base station (BS) is equipped with M_{BS} transmitting antennas and G_{BS} RF chains. To be simple, each VUE is assumed to be equipped with single antenna [23], [29]. Each RF chain is connected to all antennas to form the fully connected structure [16], [17], which can achieve optimal hybrid precoding performance.

The VUEs in different beams can be multiplexed in spatial domain, while the VUEs in the same beam can be multiplexed in power domain, since strong channel correlation is required for NOMA to superimpose clusters' signals [28]. The VUE set of the n-th cluster in the g-th beam is expressed as $\mathcal{N}_{g,n}$, which can include multiple NOMA VUEs $V^i_{g,n}$, $i \geq 1$ [31]. The pairing-based NOMA is generally considered [32], [33], if each VUE has strict QoS constraints. $|\mathcal{N}_{g,n}|$ is utilized to denote the number of VUEs in $\mathcal{N}_{g,n}$. We have $|\mathcal{N}_{g,n}| \in N_+$ and $\mathcal{N}_{g,n} \cap \mathcal{N}_{g',n'} = \Phi$ for $|g - g'| + |n - n'| \neq 0$. An adaptive NOMA/OMA clustering is considered in this system, since the number of VUEs in each beam is uncertain. Specifically, NOMA is adopted when $|\mathcal{N}_{g,n}| \geq 2$, while OMA is adopted when $|\mathcal{N}_{g,n}| = 1$. Without loss of generality, assume that the channel gain of $V^i_{g,n}$ is stronger than $V^{i+1}_{g,n}$ when $|\mathcal{N}_{g,n}| \geq 2$. The NOMA/OMA cluster set of the g-th beam is expressed as $\mathcal{N}_g$. Hence, the number of VUEs and the number of clusters are $U = \sum_{g=1}^{G_D} \sum_{n=1}^{|\mathcal{N}_g|} |\mathcal{N}_{g,n}|$ and $|\mathcal{N}| = \sum_{g=1}^{G_D} |\mathcal{N}_g|$ respectively, where G_D is the number of formed beams. We have $G_D \leq G_{BS}$.

Considering the strong multiplexing capability of NOMA and hybrid precoding, the bandwidth of RB is defined as the aggregate system bandwidth ω [12], and the width of time slot is τ_{RB}, i.e., the number of equivalent RBs after aggregation is equal to the number of time slots [9], [32]. One SPS period requires K time slots to finish the transmission of packets under the QoS constraints. Hence, the SPS period is $K\tau_{RB}$.

B. CHANNEL MODEL

We adopt the channel model widely used in mmWave systems [23]–[25], [29]. The channel vector of $V^i_{g,n}$ is

$$\mathbf{h}^i_{g,n} = \sqrt{M_{BS}} \sum_{l=1}^{L} \beta^i_{g,n,l} \mathbf{a}\left(\phi^i_{g,n}\right), \tag{1}$$

where L is the number of paths from BS to each VUE. Without loss of generality, the first path is defined as line-of-sight (LoS), while the others are non-line-of-sight (NLoS). $\beta^i_{g,n,l}$ represents the complex Gaussian gain of $V^i_{g,n}$ in the l-th path, and $\beta^i_{g,n,l} \sim \mathcal{CN}(0, \ell^i_{g,n,l})$. $\ell^i_{g,n,l}$ represents the average path loss from BS to $V^i_{g,n}$. Considering the mobility of VUEs, the actual path loss is $\ell^i_{g,n,l} = |\mathbf{d}^i_{g,n} + \mathbf{v}^i_{g,n} t_w|^{-\eta_l}$, where $\mathbf{d}^i_{g,n}$ and $\mathbf{v}^i_{g,n}$ are the displacement vector and relative velocity between BS and $V^i_{g,n}$, respectively. t_w is the waiting interval between the time point when the data is ready to transmit and the time point when the transmission starts, and the maximum waiting interval can be $t_w = K\tau_{RB}$. η_l is the path loss exponent of the l-th path.

$\mathbf{a}\left(\phi^i_{g,n}\right)$ represents the array steering vector, and $\phi^i_{g,n}$ represents the azimuth angle of departure (AoD), $\phi^i_{g,n} \in [0, 2\pi)$. Particularly, the uniform linear array (ULA) is adopted, and

$\mathbf{a}\left(\phi_{g,n}^{i}\right)$ can be expressed as

$$\mathbf{a}\left(\phi_{g,n}^{i}\right)=M_{BS}{}^{-\frac{1}{2}}[e^{\frac{j2\pi\mathbf{m}f_c d\sin(\phi_{g,n}^{i})}{c}}]^{T}, \tag{2}$$

where f_c, c, d and $\mathbf{m}=\{0,1,\ldots,M_{BS}-1\}$ are respectively the CCF, the speed of light, the antenna spacing, and the antenna set. Note that a value of $\sin(\phi_{g,n}^{i})$ may correspond to two different values of AoD, e.g. 0.25π and 0.75π. Therefore, VUEs with strong channel correlation may be far apart in spatial position.

C. TRANSMITTING AND RECEIVING

According to (1), the channels of the transmitting antennas in a mmWave system are not spatially orthogonal. Thus, the channel matrix with the size of $M_{BS}\times M_{BS}$ generated by (1) cannot be full rank. At this point, the maximum number of data streams that can be distinguished in spatial domain is often much smaller than the number of antennas, even if the fully digital beamforming is adopted. Therefore, considering the cost and overhead, an analog precoding is first utilized to obtain the equivalent channel matrix that is low dimensional but low correlation, and then, a low-dimensional digital precoding is utilized to improve the level of interference suppression, especially for the services with high reliability constraints.

Considering the analog precoder with B bits quantization precision, the minimum step of phase shifter is $2^{1-B}\pi$. Combined with the channel feature of ULA, the analog precoding vector satisfies

$$\mathbf{w}_{g}^{a}\in\mathcal{W}^{a}=\left[\mathbf{a}\left(2^{1-B}\pi\mathbf{b}\right)\right], \tag{3}$$

where $\mathbf{b}=\left\{0,1,\ldots,2^{B}-1\right\}$ is unit phase shift, and $\mathbf{w}_{g}^{a}$ is the analog precoding vector of the g-th beam. Hence, the analog precoding matrix can be expressed as

$$\mathbf{W}^{a}=\left[\mathbf{w}_{1}^{a},\mathbf{w}_{2}^{a},\ldots,\mathbf{w}_{G_D}^{a}\right]. \tag{4}$$

Then, the digital precoding vector $\mathbf{w}_{g,n}^{i}$ of $V_{g,n}^{i}$ is calculated. Typically, in a conventional NOMA cellular system, $\mathbf{w}_{g,n}^{i}$ is often calculated by the CSI of the strong user (SU) which has the highest channel gain in the cluster [28], [32]. Such a SU-based scheme facilitates the optimization of system throughput, since SU has the best channel quality in its cluster. In *Section IV*, we simulate the above-mentioned "SU-based Hybrid Precoding" as one of the baseline schemes. Considering that each type of time-triggered services in eV2X systems has the same reliability constraint and a similar size of packet, a fair zero-forcing digital precoding method is adopted and can be expressed as

$$\tilde{\mathbf{W}}^{d}=(\mathbf{W}^{a})^{H}\mathbf{H}\Big(\mathbf{H}^{H}\mathbf{W}^{a}(\mathbf{W}^{a})^{H}\mathbf{H}\Big)^{-1}, \tag{5}$$

where $\mathbf{H}=\left[\mathbf{h}_{1,1}^{1},\ldots,\mathbf{h}_{2,1}^{1},\ldots,\mathbf{h}_{G_D,|\mathcal{N}_{G_D}|}^{|\mathcal{N}_{G_D},|\mathcal{N}_{G_D}||}\right]$ represents the channel matrix containing the channel vectors of all VUEs and $\tilde{\mathbf{W}}^{d}=\left[\tilde{\mathbf{w}}_{1,1}^{1},\ldots,\tilde{\mathbf{w}}_{2,1}^{1},\ldots,\tilde{\mathbf{w}}_{G_D,|\mathcal{N}_{G_D}|}^{|\mathcal{N}_{G_D},|\mathcal{N}_{G_D}||}\right]$ represents the digital precoding matrix.

After normalization, $\mathbf{w}_{g,n}^{i}$ can be expressed as

$$\mathbf{w}_{g,n}^{i}=\tilde{\mathbf{w}}_{g,n}^{i}/\left\|\tilde{\mathbf{w}}_{g,n}^{i}\right\|. \tag{6}$$

To facilitate the expression of signal, we define a scheduling indicator $\theta_{g,n}^{k}$ for the VUEs in $\mathcal{N}_{g,n}$, where

$$\theta_{g,n}^{k}=\begin{cases}1, & \text{if } \mathcal{N}_{g,n} \text{ is scheduled in the k - th RB,}\\ 0, & \text{otherwise.}\end{cases} \tag{7}$$

Then, the received signal of $V_{g,n}^{i}$ in the k-th RB can be expressed as

$$y_{g,n}^{i(k)}=\sqrt{P_t}\bar{\mathbf{h}}_{g,n}^{i}\sum_{x=1}^{G_D}\sum_{y=1}^{|\mathcal{N}_x|}\sum_{z=1}^{|\mathcal{N}_{x,y}|}\mathbf{w}_{x,y}^{z}\sqrt{\theta_{x,y}^{k}\alpha_{x,y}^{z}}x_{x,y}^{z}+v_{g,n}^{i}, \tag{8}$$

where $\bar{\mathbf{h}}_{g,n}^{i}=(\mathbf{h}_{g,n}^{i})^{H}\mathbf{W}^{a}$ is the equivalent channel vector after analog precoding. P_t represents the transmit power, and $v_{g,n}^{i}$ represents the additive noise with a power spectral density of σ_v. $\alpha_{g,n}^{i}$ is the power allocation factor, which satisfies

$$\sum_{g=1}^{G_D}\sum_{n=1}^{|\mathcal{N}_g|}\theta_{g,n}^{k}\sum_{i=1}^{|\mathcal{N}_{g,n}|}\alpha_{g,n}^{i}\le 1,\quad \forall k\in\{1,2,\ldots K\}. \tag{9}$$

Minimum mean square error - successive interference cancellation (MMSE-SIC) is adopted to decode the received signal [20], where the mean square error is given by

$$e_{g,n}^{i}=E\left[\left\|\rho_{g,n}^{i}y_{g,n}^{i(k)}-\theta_{g,n}^{k}x_{g,n}^{i}\right\|^{2}\right]. \tag{10}$$

$\rho_{g,n}^{i}$ represents the channel equalization coefficient (CEC) of $V_{g,n}^{i}$, and we assume $E\left[\left\|x_{g,n}^{i}\right\|^{2}\right]=1$.

D. SPS PROBLEM FORMULATION

The spectral efficiency for $V_{g,n}^{i}$ to transmit a packet of $L_{g,n}^{i}$ bytes is given by

$$\bar{R}_{g,n}^{i}=8\ln 2\cdot\frac{L_{g,n}^{i}}{\tau_{RB}\omega}, \tag{11}$$

where the nature unit is utilized as the unit of data to simplify subsequent calculations, and thus, the unit of $\bar{R}_{g,n}^{i}$ is $nats/(s\cdot Hz)$.

For brevity, $\varsigma_{g,n}^{i}$ is utilized to describe the intra-beam interference, the inter-beam interference and the additive noise of $V_{g,n}^{i}$. According to (8), the interference plus noise can be expressed as

$$\varsigma_{g,n}^{i}=P_t\sum_{x\neq g}\sum_{y=1}^{|\mathcal{N}_x|}\theta_{x,y}^{k}\sum_{z=1}^{|\mathcal{N}_{x,y}|}\left\|\mathbf{h}_{g,n}^{i}\mathbf{w}_{x,y}^{z}\right\|^{2}\alpha_{x,y}^{z}+P_t\theta_{g,n}^{k}\sum_{z=1}^{i-1}\left\|\mathbf{h}_{g,n}^{z}\mathbf{w}_{g,n}^{i}\right\|^{2}\alpha_{g,n}^{z}+\sigma_v^{2}. \tag{12}$$

The signal to interference plus noise ratio (SINR) of $V_{g,n}^i$ is given by

$$\gamma_{g,n}^i = \frac{P_t \left\| \bar{\mathbf{h}}_{g,n}^i \mathbf{w}_{g,n}^i \right\|^2 \theta_{g,n}^k \alpha_{g,n}^i}{\varsigma_{g,n}^i}. \tag{13}$$

Then, the achievable spectral efficiency of $V_{g,n}^i$ can be expressed as

$$R_{g,n}^i = \ln(1 + \gamma_{g,n}^i). \tag{14}$$

Since latency is set as the objective function of the optimization problem, the QoS constraints are mostly concerned with reliability. In this paper, we utilize PRR to numerically describe the transmission reliability [12], which can be calculated by a logical function defined as follows,

$$\delta_{g,n}^i = \frac{1}{1 + e^{-\mu(R_{g,n}^i - \bar{R}_{g,n}^i)}} \geq \delta_{th}. \tag{15}$$

μ is the slope parameter [1], and δ_{th} represents the threshold of reliability in one SPS period.

The VUE set scheduled in the k-th time slot is described as Γ_k. The number of VUEs in Γ_k is $|\Gamma_k| = \sum_{g=1}^{G_D} \sum_{n=1}^{|\mathcal{N}_g|} \sum_{i=1}^{|\mathcal{N}_{g,n}|} \theta_{g,n}^k$, where $\left\{\theta_{g,n}^k\right\}$ should meet the QoS constraints of all VUEs by the optimization of $\left\{\alpha_{g,n}^i\right\}$ in k-th time slot. In the worst case, we have $|\Gamma_k| = 1$. Define function $f(K)$ to represent the number of VUEs scheduled in the first K time slots, where $f(K) = \sum_{k=1}^{K} |\Gamma_k|$. Let $f(K) = U$. The minimum SPS period can be expressed as the inverse function of $f(K)$, i.e., $f^{-1}(U)$. Then, the SPS period optimization problem in terms of power control and time-frequency resource allocation can then be formulated as follows,

$$\min_{\left\{\theta_{g,n}^k\right\}, \left\{\alpha_{g,n}^i\right\}} K = f^{-1}(U), \tag{16}$$

$$s.t.\ (9), (15),$$

$$\alpha_{g,n}^i \geq 0, \forall 1 \leq g \leq G_D, 1 \leq n \leq |\mathcal{N}_g|, 1 \leq i \leq |\mathcal{N}_{g,n}|, \tag{16a}$$

$$\sum_{k=1}^{K} \theta_{g,n}^k = 1, \forall 1 \leq g \leq G_D, 1 \leq n \leq |\mathcal{N}_g|, \tag{16b}$$

$$\sum_{n=1}^{|\mathcal{N}_g|} \theta_{g,n}^k \leq 1, \forall 1 \leq g \leq G_D, 1 \leq k \leq K. \tag{16c}$$

The constraints of transmit power and QoS are described in constraints (9), (16a) and (15). Specifically, the power allocation factors are nonnegative, and the sum of power in each time slot does not exceed P_t. Constraint (16b) represents that each cluster in each beam is scheduled once and only once in one SPS period, and thus, $\sum_{g=1}^{G_D} \sum_{n=1}^{|\mathcal{N}_g|} \sum_{k=1}^{K} \theta_{g,n}^k = |\mathcal{N}|$. The constraint (16c) represents that no more than one of the VUEs in the same beam is scheduled in the same time slot, since the clusters in the same beam cannot have strong channel correlation.

The scheduling and resource allocation problem of SPS described in problem (16) is more complex than DS, since more than one time slot is considered in SPS. Moreover, the multiplexing in power domain and spatial domain increases the complexity. Particularly, such a problem is proved to be NP-hard, where the details are given in Appendix A.

To obtain the polynomial complexity, problem (16) can be split into the power control process and the VUE-RB matching process, and an iterative algorithm is designed to solve the joint optimization problem of the parameter sets $\{\theta_{g,n}^k\}$ and $\{\alpha_{g,n}^i\}$.

III. MmWave-NOMA-BASED SPS STRATEGY

In this section, a novel SPS strategy is proposed for eV2X systems based on mmWave and NOMA. For the power control subproblem, we solve the optimization problem of $\{\alpha_{g,n}^i\}$ under QoS constraints. For the VUE-RB matching subproblem, we propose a low-complexity clustering algorithm based on the player union idea in matching theory. After that, a cluster-RB matching problem is formed, and a LLHR matching algorithm is designed to solve this problem. Finally, the analysis of stability, convergence, and complexity is given.

A. OPTIMAL LLHR POWER CONTROL ALGORITHM

The optimal power allocation process is performed after the VUE-RB matching process, since the power allocation depends on the result of matching. Considering a candidate SPS period can be determined after the VUE-RB matching process, the main purpose of power allocation is to find the lower bound of reliability for each time slot instead of latency optimization.

According to the above discussion, the premise of calculating $\{\alpha_{g,n}^i\}$ is that $\{\theta_{g,n}^k\}$ is determined. Hence, a temporary period $K\tau_{RB}$ is also known. The power control process needs to judge whether the candidate matching $\{\theta_{g,n}^k\}$ can satisfy constraints (9) and (16a) in each time slot and constraint (15) for each VUE. For the k-th time slot, this sub-problem can be described as

$$\max_{\{\alpha_{g,n}^i\}} \min_{V_{g,n}^i \in \Gamma_k} \delta_{g,n}^i, \tag{17}$$

$$s.t.\ \alpha_{g,n}^i \geq 0, \tag{17a}$$

$$\sum_{g=1}^{G_D} \sum_{n=1}^{|\mathcal{N}_g|} \sum_{i=1}^{|\mathcal{N}_{g,n}|} \theta_{g,n}^k \alpha_{g,n}^i \leq 1. \tag{17b}$$

The power constraints in the k-th time slot are shown in (17a) and (17b). Problem (17) can get the max-min reliability result $\{\alpha_{g,n}^i\}$ with a given $\{\theta_{g,n}^k\}$. Specifically, if this result $\{\alpha_{g,n}^i\}$ satisfies constraint (15), it can be one of the feasible solutions to problem (16). Then, we can try to schedule another more suitable cluster in the current time slot to shorten the SPS period, until constraint (15) cannot be satisfied.

According to (10), the mean square error is generated when the received signal of $V_{g,n}^i$ is decoded, which can be expressed

as

$$e^i_{g,n} = \left\| \rho^i_{g,n} \sqrt{P_t \theta^k_{g,n} \alpha^i_{g,n}} \bar{\mathbf{h}}^i_{g,n} \mathbf{w}^i_{g,n} - \theta^k_{g,n} \right\|^2 + \left\| \rho^i_{g,n} \right\|^2 \varsigma^i_{g,n}. \tag{18}$$

We have $\theta^k_{g,n} = 1$ when $V^i_{g,n}$ needs to be decoded. By deriving $e^i_{g,n}$ with respect to $\rho^i_{g,n}$, we can get the optimal CEC when the minimum mean square error is obtained. The optimal CEC is given by

$$\rho^i_{g,n} = 1 - \varsigma^i_{g,n} \left(\sqrt{P_t \theta^k_{g,n} \alpha^i_{g,n}} \left\| \bar{\mathbf{h}}^i_{g,n} \mathbf{w}^d_{g,n} \right\| + \varsigma^i_{g,n} \right)^{-1}. \tag{19}$$

At this point, the mean square error can be formulated as follows,

$$\min_{\rho^i_{g,n}} e^i_{g,n} = \varsigma^i_{g,n} \left(\sqrt{P_t \theta^k_{g,n} \alpha^i_{g,n}} \left\| \bar{\mathbf{h}}^i_{g,n} \mathbf{w}^d_{g,n} \right\| + \varsigma^i_{g,n} \right)^{-1}. \tag{20}$$

According to (14) and (20), we have

$$R^i_{g,n} = -\min_{\rho^i_{g,n}} \ln e^i_{g,n}. \tag{21}$$

Substituting it into (15), the transmission reliability can be expressed as

$$\delta^i_{g,n} = \max_{\rho^i_{g,n}} \left(1 + e^{\mu \bar{R}^i_{g,n}} (e^i_{g,n})^{\mu}\right)^{-1}. \tag{22}$$

Hence, the objective function of problem (17) can be converted into the following form,

$$\min_{\left\{\alpha^i_{g,n}\right\}} \max_{V^i_{g,n} \in \Gamma_k} \min_{\rho^i_{g,n}} \left(1 + e^{\mu \bar{R}^i_{g,n}} (e^i_{g,n})^{\mu}\right)^{-1}, \tag{23}$$

which is still non-convex with respect to $\{\alpha^i_{g,n}\}$. Hence, we then convert it into a convex problem. According to (22) and (15), we have

$$\max_{\rho^i_{g,n}} \left(1 + e^{\mu \bar{R}^i_{g,n}} (e^i_{g,n})^{\mu}\right)^{-1} \geq \delta_{th}. \tag{24}$$

According to (20), the optimal CEC can be uniquely determined by $\{\alpha^i_{g,n}\}$ after $\{\theta^k_{g,n}\}$ is given. Therefore, constraint (15) can be formulated as

$$\varsigma^i_{g,n} - [e^{-\mu \bar{R}^i_{g,n}} (\tilde{\delta}^{-1}_{th} - 1)]^{\frac{1}{\mu}} * \left(\sqrt{P_t \theta^k_{g,n} \alpha^i_{g,n}} \left\| \bar{\mathbf{h}}^i_{g,n} \mathbf{w}^i_{g,n} \right\| + \varsigma^i_{g,n} \right) \leq 0, \tag{25}$$

where $\tilde{\delta}_{th}$ is the PRR threshold of the current round in LLHR-PC algorithm. Then, problem (17) can be transferred into the following form,

$$\min_{\{\alpha^i_{g,n}\}} \sum_{g=1}^{G_D} \sum_{n=1}^{|\mathcal{N}_g|} \sum_{i=1}^{|\mathcal{N}_{g,n}|} \theta^k_{g,n} \alpha^i_{g,n}, \quad s.t. \quad (17a), (25). \tag{26}$$

Both (17a) and the objective function of problem (26) are linear, and (25) is a unitary quadratic inequality. Therefore, problem (26) is convex.

We focus on the power allocation in the k-th time slot, where we have $\Gamma_k = \left\{ V^i_{g,n} | \theta^k_{g,n} = 1, V^i_{g,n} \in \mathcal{N}_{g,n} \right\}$. To simplify the expression, the index of VUEs can be redefined as $V_m \triangleq V^i_{g,n} \in \Gamma_k$. Then, problem (26) can be described as follows,

$$\min_{\{\alpha_m\}} \sum_{m=1}^{|\Gamma_k|} \alpha_m, \tag{27}$$

$$s.t. \ \alpha_m \geq 0, \tag{27a}$$

$$(1 - A_m) \varsigma_m - B_m \sqrt{\alpha_m} \leq 0, \tag{27b}$$

where the constant $A_m = \left[e^{-\mu \bar{R}_m} \left(\tilde{\delta}^{-1}_{th} - 1 \right) \right]^{\frac{1}{\mu}}$, and the constant $B_m = A_m \sqrt{P_t} \left\| \bar{\mathbf{h}}_m \mathbf{w}_m \right\|$.

The Lagrangian function of problem (27) can be written by

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{m=1}^{|\Gamma_k|} \left[\alpha_m + \lambda_m (1 - A_m) \varsigma_m - \lambda_m B_m \sqrt{\alpha_m} \right], \tag{28}$$

where $\boldsymbol{\alpha} = \left(\alpha_1, \alpha_2, \ldots, \alpha_{|\Gamma_k|} \right)^T$. According to the Karush-Kuhn-Tucker (KKT) conditions, the optimal solution should satisfy

$$\frac{\partial \mathcal{L}}{\partial \alpha_m} = 1 + P_t (1 - A_m) \sum_{i \in \Theta_m} \lambda_i \left\| \bar{\mathbf{h}}_m \mathbf{w}_m \right\|^2 - \frac{\lambda_m B_m}{2 \sqrt{\alpha_m}} = 0, \tag{29}$$

where set $\Theta_m = \Gamma_k / \{V_j | V_j$ and V_m in the same beam, $j \geq m\}$.

When $\lambda_m = 0$, we have

$$1 + P_t (1 - A_m) \sum_{i \in \Theta_m} \lambda_i \left\| \bar{\mathbf{h}}_m \mathbf{w}_m \right\|^2 = 0. \tag{30}$$

Considering $0.5 < \delta_{th} < 1$, we have $A_m > 0$, $B_m > 0$, and $1 - A_m > 0$. Hence, it can be obtained that $1 + P_t (1 - A_m) \sum_{i \in \Theta_m} \lambda_i \left\| \bar{\mathbf{h}}_m \mathbf{w}_m \right\|^2 > 0$, which is contradictory to (30). Therefore, $\forall \lambda_m > 0$.

The solution of α_m can be obtained by

$$\alpha_m = \left(\frac{\lambda_m B_m}{2 + 2 (1 - A_m) P_t \sum_{i \in \Theta_m} \lambda_i \left\| \bar{\mathbf{h}}_m \mathbf{w}_m \right\|^2} \right)^2. \tag{31}$$

To further obtain α_m, the gradient based method can be utilized to update the value of λ_m [34], [35]. The primal variables can be obtained by assuming a certain value of the power difference in each NOMA cluster. In each iteration, the new value of the Lagrangian variables can be calculated by

$$\lambda_m(t+1) = \lambda_m(t) - \theta(t) \left[(1 - A_m) \varsigma_m(t) - B_m \sqrt{\alpha_m(t)} \right], \tag{32}$$

where $\theta(t)$ is a dynamically chosen step sequence as introduced in [35].

In LLHR-PC algorithm, the optimal power allocation solution $\{\alpha^i_{g,n}\}$ is calculated for the k-th time slot. First, the boundary of $\tilde{\delta}_{th}$ is determined. The ideal PRR is $\delta_{th} \to 1$. Moreover, if the solution to problem (27) cannot satisfy constraint (27b) when $\tilde{\delta}_{th} = \delta_{th}$, the current matching $\{\theta^k_{g,n}\}$ is proved to be infeasible. Otherwise, we have $\tilde{\delta}_{th} \in [\delta_{th}, 1)$. In each round, we utilize a temporary PRR threshold, i.e. $\tilde{\delta}_{th}$, to calculate $\{\alpha^i_{g,n}\}$, until the search space of $\tilde{\delta}_{th}$ is limited to ε_δ. The details are described in **Algorithm 1**.

Algorithm 1 Proposed LLHR-PC Algorithm

Input: $\{\theta^k_{g,n}\}$ of the k-th time slot.
Output: $\{\alpha^i_{g,n}\}$; the lower bound of PRR $\tilde{\delta}_{th} = \delta_{LB}$.

```
Init. δ_LB = δ_th, δ_UB = 1, δ̃_th = δ_LB;
Solve problem (27);
if Σ_{g=1}^{G_D} Σ_{n=1}^{|N_g|} Σ_{i=1}^{|N_{g,n}|} θ^k_{g,n} α^i_{g,n} > 1 or (27) is infeasible then
    {θ^k_{g,n}} is infeasible;
else
    while δ_UB − δ_LB ≥ ε_δ do
        δ̃_th = (δ_DB + δ_UB) /2 and solve problem (27);
        if Σ_{g=1}^{G_D} Σ_{n=1}^{|N_g|} Σ_{i=1}^{|N_{g,n}|} θ^k_{g,n} α^i_{g,n} ≤ 1 then
            δ_LB = δ̃_th, and record {α^i_{g,n}};
        else
            δ_UB = δ̃_th;
        end if
    end while
end if
```

Remark 1: The calculation of $\boldsymbol{\alpha}$ according to (31) has linear complexity, and the calculation of Lagrangian variables according to (32) with proper choice of step $\theta(t)$ has iteration complexity. Furthermore, $\tilde{\delta}_{th}$ is updated by bisection method. Therefore, the complexity of **Algorithm 1** is $\mathcal{O}\left(\frac{|\Gamma_k|}{\sqrt{\kappa}} \log \frac{1-\delta_{th}}{\varepsilon_\delta}\right)$, where κ is the accuracy of Lagrangian variables.

B. MATCHING PROBLEM FORMULATION

LLHR-PC algorithm shows that $\{\alpha^i_{g,n}\}$ can be determined by $\{\theta^k_{g,n}\}$. However, the matching result $\left\{\theta^k_{g,n}\right\}$ will also be affected by $\{\alpha^i_{g,n}\}$. Thus, we form and analyze the following VUE-RB matching problem to calculate both $\{\theta^k_{g,n}\}$ and $\{\alpha^i_{g,n}\}$.

Use two disjoint sets of V and F to respectively represent U VUEs and K RBs in an eV2X system, where $V = \{v_1, v_2, \ldots, v_U\}$, $F = \{f_1, f_2, \ldots, f_K\}$. The matching problem with respect to V and F can be described as follows.

Definition 1: A matching Ψ is a mapping from the set $V \cup F$ to the set $V \cup F$, where
1) $\Psi(v_u) \in F \cup \{v_u\}$;
2) $\Psi(f_k) \subset V \cup \{f_k\}$;
3) $\Psi(v_u) = f_k \Leftrightarrow v_u \in \Psi(f_k)$.

Condition 1) implies that each VUE is scheduled in no more than one RB, i.e., $|\Psi(v_u)| = 1$ and $|\Psi(v_u) \cap F| \leq 1$, while condition 2) implies that each RB can serve several VUEs. Condition 3) indicates that the mapping between v_u and f_k is symmetrical. Specially, $\Psi(j) = \{j\}$ happens when there is no other matching for element j. Note that if $\exists j \in V \cup F$ such that $\Psi(j) = \{j\}$, the current matching Ψ will be infeasible.

When $\Psi(v_u) = \Psi(v'_u) = f_k$ and $v_u, v'_u \in V$, the co-channel interference will occur between v_u and v'_u, since they are multiplexed in the same RB. Particularly, the co-channel interference can be either intra-beam or inter-beam. Thus, compared with the existing matching problem [8], Ψ suffers from more complex peer effects. To solve this problem, a preference relationship is formed for both f_k and v_k, which can help players to search for other players according to their interests. Specifically, the preference $\mathcal{R}$ for f_k and v_k is defined as follows:

Definition 2: Both f_k and v_k have strict[2] preferences. For two sets of VUEs $V^i, V^j \subseteq V$, $\mathcal{R}\left(V^i f_k V^j\right)$ represents that f_k prefers V^i than V^j; for two RBs $f_i, f_j \in F$, $\mathcal{R}\left(f_i v_u f_j\right)$ represents that v_u prefers f_i than f_j.

Remark 2: For f_k, $\mathcal{R}$ is determined by the PRR threshold $\tilde{\delta}_{th}(f_k, \Psi(f_k))$ of the k-th RB with respect to the current matching $\Psi(f_k)$. Specifically, substituting $(f_k, \Psi(f_k))$ into LLHR-PC algorithm, $\tilde{\delta}_{th}$ of the k-th RB can be calculated. For v_u, $\mathcal{R}$ is determined by the achievable PRR of v_u in matching $(\Psi(v_u), \Psi(\Psi(v_u)))$, where $\Psi(\Psi(v_u))$ represents the player union of v_u. Hence, Definition 2 can be formulated as

$$\mathcal{R}\left(V^i f_k V^j\right) \Leftrightarrow \tilde{\delta}_{th}\left(f_k, V^i\right) > \tilde{\delta}_{th}\left(f_k, V^j\right). \tag{33}$$

$$\begin{aligned}\mathcal{R}\left(f_i v_u f_j\right) \Leftrightarrow\ & \min\{\tilde{\delta}_{th}(f_i, \Psi(f_i) \cup v_u), \tilde{\delta}_{th}(f_j, \Psi(f_j))\} \\ >\ & \min\{\tilde{\delta}_{th}(f_i, \Psi(f_i)), \tilde{\delta}_{th}(f_j, \Psi(f_j) \cup v_u)\}.\end{aligned} \tag{34}$$

The preference is transitive, and thus, a complete preference list can be formed by (33) and (34). With $\mathcal{R}$, the eV2X system can be expressed as $(V \cup F, \mathcal{R})$, where a feasible Ψ should satisfy that for $\forall v_u \in V$ and $\forall f_k \in F$, $\mathcal{R}(\Psi(v_u) v_u \{v_u\})$ and $\mathcal{R}(\Psi(f_k) f_k \{f_k\})$ are always established.[3] Additionally, according to *Remark 2*, v_u can find an ordered selection set (SSet), i.e., feasible union, satisfying its QoS constraints, which is defined as follows:

Definition 3: For a given VUE $v_u \in V^i$ and $V^i \subseteq V$, $S\left(v_u, V^i\right)$ represents the SSet of v_u in V^i, where $S\left(v_u, V^i\right) \subseteq V^i$. v_u satisfies the QoS constraints in $S\left(v_u, V^i\right)$. For $\forall V^j \subseteq V^i$ and $V^j \neq S\left(v_u, V^i\right)$, $\mathcal{R}\left(V^j \Psi(v_u) S\left(v_u, V^i\right)\right)$ does not hold. Similarly, for a given $f_k \in F$ and $V^i \subseteq V$, $S\left(f_k, V^i\right)$ represents the SSet of f_k in V^i. The QoS constraints are satisfied when $\Psi(f_k) = S\left(f_k, V^i\right)$. For $\forall V^j \subseteq V^i$ and $V^j \neq S\left(f_k, V^i\right)$, $\mathcal{R}\left(V^j f_k S\left(f_k, V^i\right)\right)$ does not hold.

$S\left(v_u, V^i\right)$ can reflect the expected union of v_u, while other VUEs' constraints are not guaranteed.

[2]For a given player, any two preference values are assumed to be unequal.
[3]We called it individual rationality. The size of RB, i.e. ω or τ_{RB}, should be adjusted if there is an element in Ψ that is not individual rationality.

Each VUE prefers its SSet over the set of all VUEs. Particularly, such SSets should be stable, where a stable SSet is defined as follows:

Definition 4: For VUE set $V^i \subseteq V$, if $S\left(V^i, V\right) = \bigcup_{v_u \in V^i} S\left(v_u, V\right) = V^i$, i.e. the SSet of V^i is itself, we call V^i a stable SSet.

Remark 3: A feasible Ψ satisfies that $\forall f_k, f_{k'} \in F$, $\Psi\left(f_k\right) = S\left(V^i, V\right) = V^i$, $\Psi\left(f_{k'}\right) = S\left(V^{i'}, V\right) = V^{i'}$ and $V^i \cap V^{i'} = \varnothing$, where $k \neq k'$.

Moreover, there may be alternative relationship between some VUEs. For instance, if v_u and v'_u are in the same beam, v'_u may replace the position of v_u in one SSet. Such relationship can be expressed as follows:

Definition 5: Given $v_u, v'_u \in V^i$, if $v_u \in S\left(f_k, V^i\right)$, we always have $v_u \in S\left(f_k, V^i/\{v'_u\}\right)$, and if $v'_u \in S\left(f_k, V^i\right)$, we always have $v'_u \in S\left(f_k, V^i/\{v_u\}\right)$. v'_u and v_u are called alternative.

Intuitively, the alternative can exclude a part of players, thereby reducing the search space of Ψ. Furthermore, it can also be utilized to improve the current matching Ψ.

C. UNION-BASED VUE-RB MATCHING ALGORITHM

In $(V \cup F, \mathcal{R})$ with unknown $\{\alpha^i_{g,n}\}$, the channel correlation difference and the channel gain difference among VUEs can be obtained by UL. This information helps VUEs to form different unions, and thus, V can be divided into several clusters. There are two benefits to this approach. For one thing, the unions facilitate the analysis of peer effects caused by intra-beam and inter-beam interference. For another, the resource allocation for unions is more efficient than the allocation for VUE individuals. The basic condition for forming a union is that each VUE in the union can successfully decode its desired signal under the QoS constraints.

Based on the above discussion, we propose a union-based VUE-RB matching algorithm, which includes two phases, i.e., the clustering phase and the matching phase.

In the clustering phase, VUEs are divided into a series of minimum unions based on CSI. To improve the beamforming gain, the beam division and VUE clustering process is performed according to the beam space constructed by the analog precoder. Since the analog precoder has a fixed phase shifter, it may result in a decrease of precoding performance for analog precoding to be selected based on the result of clustering. Therefore, a novel division algorithm is designed for each VUE to select the most appropriate analog precoding vector from the matrix in (4).

In BD&VC algorithm, each VUE is sequentially allocated to the optimal analog beam according to the channel correlation. Particularly, one vector $\mathbf{w}^a_g$ may not match with any VUE, especially when the analog phase shifter has a high precision. Considering the overhead of the digital precoder, such $\mathbf{w}^a_g$ is deleted by $\mathbf{W}_a$ to decrease the number of beams G_D. Note that the quantization precision of the analog precoder is assumed to be $B = \left\lfloor \log_2(G_{BS}) \right\rfloor$ to ensure that $G_D \leq G_{BS}$ can always be satisfied [23]. Then, to make the VUEs in the same NOMA cluster have a certain difference of channel gain [28], [32], the VUEs in each beam is sorted according to the channel gain. Specially, the VUEs with the low channel gain may be allocated in an OMA cluster. Finally, both $\mathbf{w}^a_g$ and $\mathbf{w}^i_{g,n}$ of each VUE are determined. The details of the proposed algorithm are described in **Algorithm 2**.

Algorithm 2 Proposed BD&VC Algorithm

Input: $\mathcal{W}^a$ in (3); VUE set V; channel vectors $\mathbf{h}_u$, $u \in \{1, 2, \ldots, U\}$; maximum number of VUEs in each cluster $|\mathcal{N}|_{\mathrm{MAX}}$.

Output: Cluster set $\mathcal{N}$; number of effective beams G_D; analog precoding matrix $\mathbf{W}^a$; digital precoding vector set $\{\mathbf{w}^i_{g,n}\}$.

1: Init. VUE set in each beam $\Pi = \left\{\Pi_1, \Pi_2, \ldots, \Pi_{G_{BS}}\right\}$, where $\Pi_x = \varnothing$ for $x \in \{1, 2, \ldots, G_{BS}\}$;
2: **for all** $u \in \{1, 2, \ldots, U\}$ **do**
3: $\quad g = \arg \max_{x \in G_{BS}} \left\{ \frac{\left\|\mathbf{h}_u^H \mathbf{w}^a_x\right\|}{\|\mathbf{h}_u\| \left\|\mathbf{w}^a_x\right\|} \right\}$;
4: $\quad \Pi_g = \Pi_g \cup v_u$;
5: **end for**
6: $\mathbf{G} = \{x | \Pi_x \neq \emptyset, 1 \leq x \leq G_{BS}\}$;
7: $G_D = |\mathbf{G}|$ and $\mathbf{W}^a = \bigcup_{g \in \mathbf{G}} \mathbf{w}^a_g$;
8: **for all** $g \in \mathbf{G}$ **do**
9: $\quad \mathbf{H} = \left\{ \left\| \mathbf{h}_x^H \mathbf{W}^a_g \right\|^2 | v_x \in \Pi_g \right\}$;
10: $\quad [\sim, \Lambda] = sort(\mathbf{H}, 'descend')$;
11: $\quad V^i_{g,n} = \Lambda\left((n-1)|\mathcal{N}|_{\mathrm{MAX}} + i\right)$, where $1 \leq n \leq \left\lfloor \frac{|\Pi_g|}{|\mathcal{N}|_{\mathrm{MAX}}} \right\rfloor + 1$, $1 \leq i \leq |\mathcal{N}|_{\mathrm{MAX}}$, and $(n-1)|\mathcal{N}|_{\mathrm{MAX}} + i \leq |\Pi_g|$;
12: $\quad$ Calculate $\mathbf{w}^i_{g,n}$ according to (6);
13: **end for**

Remark 4: BD&VC algorithm has the polynomial complexity. Specifically, the maximum complexity is $\mathcal{O}\left(G_{BS} U\right)$ from step 2 to step 5, while the maximum complexity is $\mathcal{O}(\sum_{g=1}^{G_{BS}} |\Pi_g| \log |\Pi_g|) \leq \mathcal{O}(G_{BS} U \log U)$ from step 8 to step 13. Therefore, the complexity of **Algorithm 2** is $\mathcal{O}\left(G_{BS} U \log(U)\right)$.

Cluster-RB matching has the compatible definitions and properties with VUE-RB matching. Specifically, for the definition of Ψ, the VUE set $V = \{v_1, v_2, \ldots, v_U\}$ can be represented by the cluster set $\mathcal{N} = \{\mathcal{N}_{1,1}, \mathcal{N}_{1,2}, \ldots, \mathcal{N}_{g,n}, \ldots\}$. Moreover, the properties of Ψ can be re-described as follows:

Proposition 1: The union $\mathcal{N}_{g,n}$ in Ψ satisfies the properties of cluster alternative.

Proof: See Appendix B.

Proposition 2: The union $\mathcal{N}_{g,n}$ in Ψ satisfies the properties of cluster common preference.

Proof: See Appendix C.

Algorithm 3 describes the union-based VUE-RB matching algorithm consisting of clustering phase and matching phase. In clustering phase, a series of minimal unions are formed by BD&VC algorithm, where the prior knowledge for matching, i.e., the properties of alternative and

preference, can be obtained. In matching phase, the largest stable unions in each RB are formed. Specifically, we first find the beam g that has the most unscheduled clusters, and then, find the cluster $\mathcal{N}_{g,n}$ that has the highest channel gain, since such $\mathcal{N}_{g,n}$ has the strictest matching condition according to constraint (16c) and has the strongest multiplexing ability among the clusters in $\mathcal{N}_g$. After that, the properties of alternative and preference are utilized for k-th RB f_k to update the matching $\Psi(f_k)$, until $(f_k, \Psi(f_k))$ cannot be affected by any unscheduled or scheduled cluster. At this point, $\Psi(f_k)$ is regarded as the current largest stable union of f_k. Additionally, in step 22, both $(f_k, \Psi(f_k))$ and $(f_{ind}, \Psi(f_{ind}))$ will be updated if there exists $\left(\mathcal{N}_{g,n}, \mathcal{N}_{g,n'}\right)$ such that $\min\{\tilde{\delta}_{th}^{k}(f_k, \Psi'(f_k)), \tilde{\delta}_{th}^{ind}(f_{ind}, \Psi'(f_{ind}))\} > \min(\tilde{\delta}_{th}^{k}, \tilde{\delta}_{th}^{ind})$, where $\Psi'(f_k) = \Psi(f_k) / \left\{\mathcal{N}_{g,n}\right\} \cup \left\{\mathcal{N}_{g,n'}\right\}$, $\Psi'(f_{ind}) = \Psi(f_{ind}) / \left\{\mathcal{N}_{g,n'}\right\} \cup \left\{\mathcal{N}_{g,n}\right\}$. Finally, the matching phase terminates when all VUEs are scheduled in one SPS period.

Algorithm 3 Union-Based VUE-RB Matching Algorithm

Input: The input in **Algorithm 2**; PRR threshold δ_{th}.
Output: SPS period $K\tau_{RB}$; $\{\theta_{g,n}^{k}\}$; $\{\alpha_{g,n}^{i}\}$.

1: Init. $\Psi = \varnothing$; $k = 0$;
2: *Phase 1. Clustering Phase*:
3: Run BD&VC algorithm;
4: *Phase 2. Matching Phase*:
5: **repeat**
6: $c = 1$ and $k = k + 1$;
7: $g = \arg \max_{l \in \{1,2,\ldots,G_D\}} \left(|\mathcal{N}_l| - \sum_{i=1}^{k} \sum_{j=1}^{|\mathcal{N}_l|} \theta_{l,j}^{i}\right)$
8: $n = \arg \max_{j \in \mathcal{N}_g} \left\|\mathbf{h}_{g,j}^{1}\right\|^2$
9: $\Psi(f_k) = \Psi(f_k) \cup \mathcal{N}_{g,n}$ and $\Psi\left(\mathcal{N}_{g,n}\right) = f_k$;
10: **repeat**
11: Form $\mathcal{R}$ for $\left\{\mathcal{N}_{g,n} | \Psi\left(\mathcal{N}_{g,n}\right) == \mathcal{N}_{g,n}, \mathcal{N}_{g,n} \in \mathcal{N}\right\}$ to find the most preferred $\mathcal{N}_{g',n'}$ for $(f_k, \Psi(f_k))$;
12: **if** $\tilde{\delta}_{th}\left(f_k, \Psi(f_k) \cup \left\{\mathcal{N}_{g',n'}\right\}\right) < \delta_{th}$ or problem (27) is infeasible **then**
13: $c = 0$;
14: **else**
15: $\Psi(f_k) = \Psi(f_k) \cup \mathcal{N}_{g',n'}$ and $\Psi\left(\mathcal{N}_{g',n'}\right) = f_k$;
16: Update $\{\theta_{g,n}^{k}\}$, then record $\left\{\alpha_{g,n}^{i}\right\}$ and $\tilde{\delta}_{th}$ for f_k;
17: **end if**
18: **until** $c == 0$
19: **if** $k > 1$ **then**
20: $[\sim, ind] = min\left\{\tilde{\delta}_{th}^{1}, \tilde{\delta}_{th}^{2}, \ldots, \tilde{\delta}_{th}^{k-1}\right\}$;
21: Find the alternative-pair set $\left\{\left(\mathcal{N}_{g,n}, \mathcal{N}_{g,n'}\right)\right\}$, where $\mathcal{N}_{g,n} \in \Psi(f_k)$, $\mathcal{N}_{g,n'} \in \Psi(f_{ind})$, $ind < k$;
22: Judge for alternative pairs according to (34);
23: **end if**
24: **until** $\forall v_u \in V$, $\Psi(v_u) \neq \{v_u\}$

D. PERFORMANCE ANALYSIS

Firstly, the concept of pairwise stable matching is introduced to prove the stability of the proposed matching algorithm.

Definition 6: If there exists a matching Ψ' that can achieve a shorter SPS period, or improve the lower bound of reliability without increasing the latency, Ψ will be blocked by Ψ'. If Ψ is not blocked by any matching between clusters and RBs that does not exist in Ψ, Ψ is called the cluster-wise stable matching.

Proposition 3: The matching obtained by the union-based VUE-RB matching algorithm is a cluster-wise stable matching.

Proof: See Appendix D.

This implies that stable SSets can be obtained by the proposed union-based matching algorithm. Particularly, whether $\mathcal{N}_{g,n}$ has been scheduled or not, $|\Psi(f_k)|$ will always be non-decreasing when $\mathcal{N}_{g,n}$ is proposed to f_k. Moreover, the number of scheduled clusters that have performed the alternative judgment will increase when there is no increase in $|\Psi(f_k)|$. Therefore, the matching process can be completed in a limited number of iterations.

For the union-based VUE-RB matching algorithm, there are two types of matching proposals. The number of proposals for unscheduled clusters is $N_1^k \leq \frac{1}{2}[(|\mathcal{N}| - \sum_{i=1}^{k-1} |\Psi(f_i)|)^2 + |\mathcal{N}| - \sum_{i=1}^{k-1} |\Psi(f_i)|]$, while the number of proposals for scheduled clusters is $N_2^k \leq |\Psi(f_k)|$. Hence, the upper bound of proposals for f_k can be expressed as

$$\begin{aligned} N_1^k + N_2^k &\leq \max_{n=1,2,\ldots,|\mathcal{N}|} \left\{\frac{(|\mathcal{N}| - n)^2 + |\mathcal{N}| - n}{2} + n\right\} \\ &\leq 0.5|\mathcal{N}|^2 - 0.5|\mathcal{N}| + 1. \end{aligned} \tag{35}$$

Two special cases of proposals are considered. For the first case, any two VUEs cannot be multiplexed in power domain, i.e., all VUEs are in different beams. Hence, we have $|\mathcal{N}| = U$ and $G_D = U$. In the worst case, only one cluster is scheduled in each RB. Hence, we have $N_1^k = |\mathcal{N}| - \sum_{i=1}^{k-1} |\Psi(f_i)| = U - (k-1) = 1 + U - k$, $N_2^k = 1$, and $K = U$. At this point, the number of proposals is given by

$$\begin{aligned} \mathcal{O}\left(\sum_{k=1}^{K} (N_1^k + N_2^k)\right) &= \mathcal{O}(0.5U^2 + 2.5U + 3) \\ &= \mathcal{O}\left(U^2\right). \end{aligned} \tag{36}$$

For the second case, any two VUEs cannot be multiplexed in spatial domain, i.e. $|\mathcal{N}| = \lceil U/|\mathcal{N}|_{MAX} \rceil$ and $G_D = 1$. Hence, for $\forall \mathcal{N}_{g,n}, \mathcal{N}_{g,n'} \in \mathcal{N}'$ and $\mathcal{N}' \subseteq \mathcal{N}$ such that $\mathcal{N}_{g,n} \in S\left(f_k, \mathcal{N}'\right)$, we have $\mathcal{N}_{g,n} \in S\left(f_k, \mathcal{N}'/\{\mathcal{N}_{g,n'}\}\right)$. According to the cluster alternative, $N_1^k = 0$, $N_2^k = 1$ and $K = U$. At this point, the number of proposals is given by

$$\mathcal{O}\left(\sum_{k=1}^{K} (N_1^k + N_2^k)\right) = \mathcal{O}(U). \tag{37}$$

Combining with the complexity of LLHR-PC algorithm and BD&VC algorithm, the complexity of **Algorithm 3** is

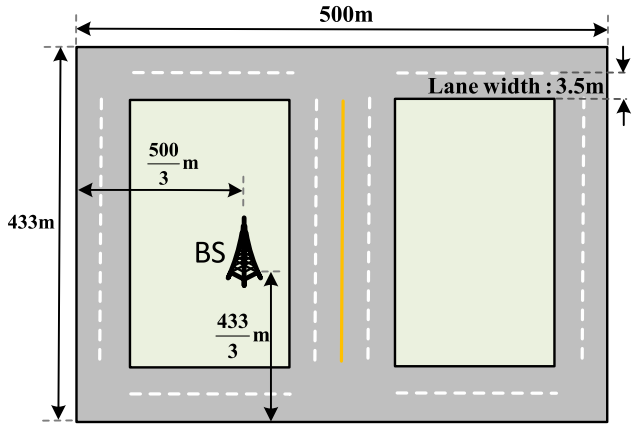


FIGURE 2. Road configuration for urban grid.

given as follows,

$$\mathcal{O}\left(G_{BS}U\log U+\sum_{k=1}^{K}(N_1^k+N_2^k)\frac{|\Gamma_k|}{\sqrt{\kappa}}\log\frac{1-\delta_{th}}{\varepsilon_\delta}\right)=\mathcal{O}(U\log U)+O(|\mathcal{N}|^3). \quad (38)$$

IV. SIMULATION RESULTS

The performance in terms of the SPS period and transmission reliability of the mmWave-NOMA-based SPS strategy proposed in this paper is evaluated by simulations. For comparison on the performance of resource allocation, multiplexing, and precoding, the following five typical schemes are considered as the benchmark methods: 1) "Greedy VUE-RB Matching", where each VUE is matched with RBs in turn until all VUEs can meet the QoS constraints in one SPS period, which is also mentioned in [8]; 2) "Hybrid Precoding (HP) OMA", where the scheduling strategy is similar to the proposed scheme, but all VUEs are orthogonal in power domain, which is also mentioned in [9], [23], [25]; 3) "Fully Analog Precoding (FAP) NOMA" [29], where only (4) is utilized for precoding; 4) "Strong user (SU) based HP NOMA" [28], [32], where the HP scheme is similar to the proposed one, while the digital precoding vectors are calculated according to the normalized channel vectors of SU; 5) "Weak user (WU) based HP NOMA", where the digital precoding vectors are calculated according to the normalized channel vectors of WU.

Specifically, the simulation parameters are described as follows. The CCF is set as $f_c = 63GHz$, and the aggregated system bandwidth is $\omega = 1GHz$ [12]. For the transmitting process, $P_t = 43dBm$, $\sigma_v = -174dbm/Hz$, $M_{BS} = 64$ and $G_{BS} = 16$. For $V_{g,n}^i$, the channel vector is generated by (1), where we assume the number of paths is $L = 3$, including the LoS component with the path loss exponent $\eta_1 = 4$ and the NLoS component with $\eta_2 = \eta_3 = 5$. $|\mathcal{N}|_{\text{MAX}}$ is set as 2. PRR is calculated by (15), where we assume: the slope parameter $\mu = 8$, the size of packets $L_{g,n}^i = 1200bytes$, short time slot $\tau_{RB} = 1/14ms$ [36], [37]. Moreover, the road configuration for unban grid discussed in this paper is presented in Fig. 2 according to 3GPP NR-V2X systems defined in [12], where the distance between a vehicle and the following vehicle in the same lane is $\left|\mathbf{v}_{g,n}^i\right| \times 2.5$ sec [8], [9].

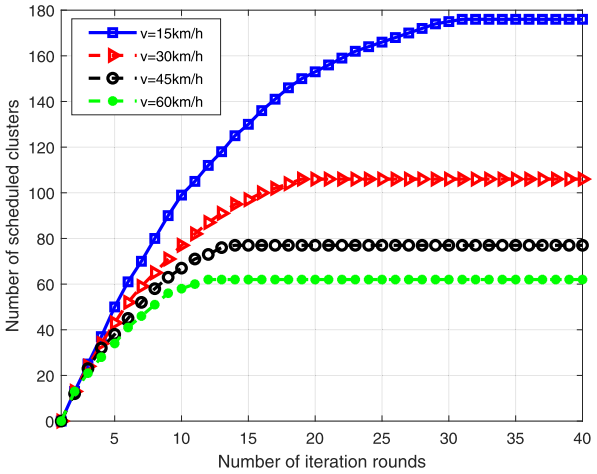


FIGURE 3. Number of scheduled clusters against number of iteration rounds with different speed of vehicles.

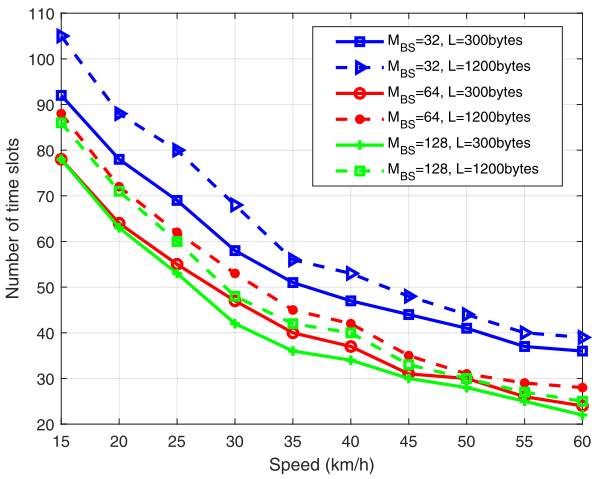


FIGURE 4. Number of time slots against speed of vehicles with different packet sizes and different number of antennas and RF chains.

Fig. 3 indicates the number of scheduled clusters in one period against the number of iteration rounds with the speed of vehicles from $15km/h$ to $60km/h$, where the PRR threshold is 99%. The density of VUE decreases when the speed of vehicles increases, which is shown in Fig. 2. Thus, the number of scheduled users decreases for a longer vehicle distance. It can be observed that the number of scheduled clusters converges to a stable value as the number of iteration rounds increases, which means that the covered VUEs are scheduled after a limited number of iterations. Hence, simulation results reflect the convergence of the union-based VUE-RB matching algorithm.

Fig. 4 indicates the number of time slots against the speed of vehicles with different sizes of packets and different numbers of antennas and RF chains, where the number of RF

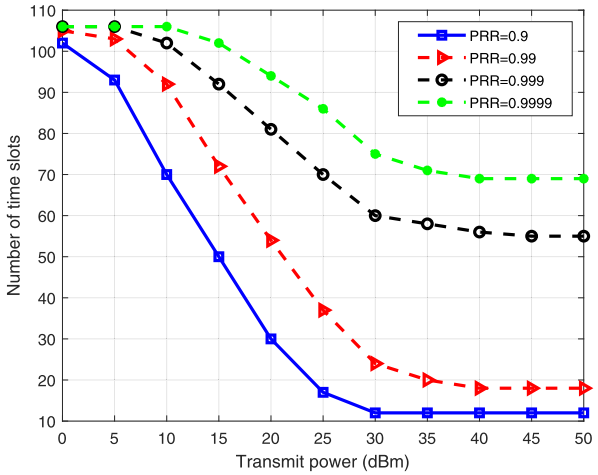


FIGURE 5. Number of time slots against transmit power with different PRR thresholds.

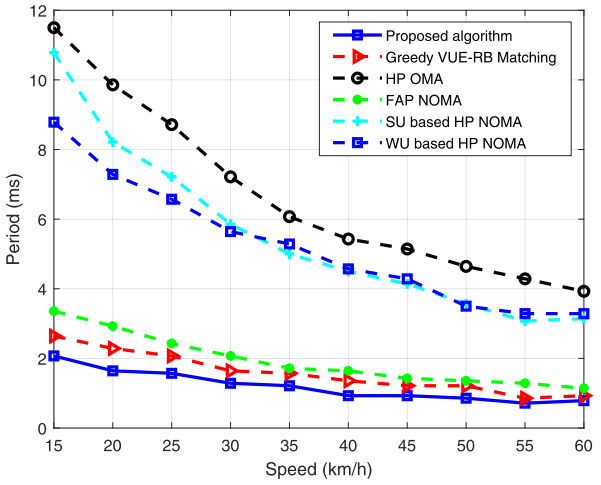


FIGURE 6. SPS period against speed of vehicles with different schemes.

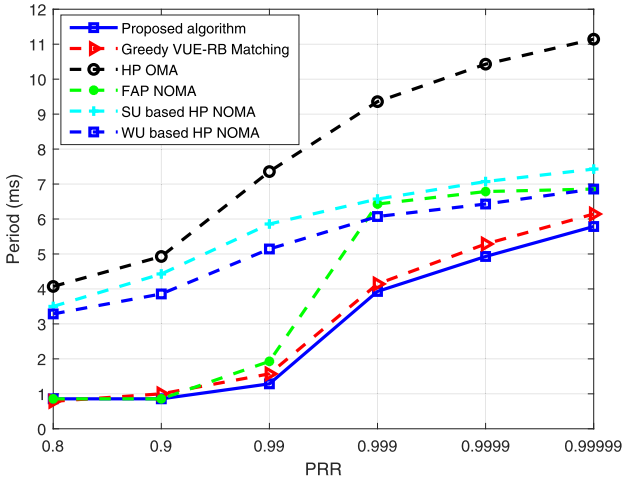


FIGURE 7. SPS period against PRR threshold with different schemes.

chains is $G_{BS} = M_{BS}/4$. It is easily observed that a smaller size of packets and a larger number of antennas and RB chains result in a decrease of the number of time slots. As the speed of vehicles increases, the difference in the number of time slots between large data packets, i.e. $L^i_{g,n} = 1200bytes$, and small data packets, i.e. $L^i_{g,n} = 300bytes$, decreases. Moreover, although there is a sharp decrease in the number of time slots when the number of antennas increases from 32 to 64, there is no significant change in the number of time slots when the number of antennas increases from 64 to 128.

Fig. 5 indicates the number of time slots against the transmit power with different PRR thresholds from 99% to 99.99%, where the speed is set as $30km/h$. We observe that the number of time slots converges to a stable value as the transmit power increases, i.e., a higher transmit power cannot result in a smaller number of time slots after the transmit power increases to approximately $43dBm$. The main reason is that a higher transmit power can reduce the additive noise, but the interference signal cannot be reduced after the power allocation factors are determined by LLHR-PC algorithm. This implies that the lower SPS period cannot be obtained by increasing the transmit power indefinitely.

Fig. 6 shows the number of time slots in one SPS period against the speed of vehicles from $15km/h$ to $60km/h$ with different schemes. For each VUE $v^i_{g,n}$, the data packet is successfully received only when the PRR satisfies $\delta^i_{g,n} \geq 99\%$. The density of VUE decreases as the speed increases, and thus, there are fewer clusters, leading to the shortening of SPS period. Moreover, it is intuitive that the proposed strategy works better than other five schemes, especially in a dense network. This implies that the proposed algorithm can improve the SPS period for any speed of vehicles.

Fig. 7 shows the SPS period, i.e. $K\tau_{RB}$, against the PRR threshold from 80% to 99.999% with different SPS schemes, where the speed is set as $30km/h$. The SPS period increases as PRR increases. For the proposed algorithm, "Greedy VUE-RB Matching" and "FAP NOMA", the eV2X systems can achieve a very short SPS period when the reliability threshold is low. However, the period of "FAP NOMA" increases rapidly as PRR increases, since the space division multiplexing ability of analog precoding is limited. Additionally, the proposed algorithm always outperforms "Greedy VUE-RB Matching", since the properties of preference are utilized.

As is shown in Fig. 6 and Fig. 7, the SPS period of the proposed SPS strategy often does not exceed $6ms$. Assuming that the speed of vehicles is $60km/h$, it can be obtained that the displacement of a vehicle in one SPS period is no more than $0.1m$. Therefore, the change in the position of a vehicle in spatial domain is very limited. According to constraint (16c), only users in different analog beams can be multiplexed in spatial domain. Hence, the directivity of mmWave has a limited impact on the space division multiplexing.

Fig. 8 shows the average PRR and the lower bound of PRR against different SPS schemes, where the PRR threshold is

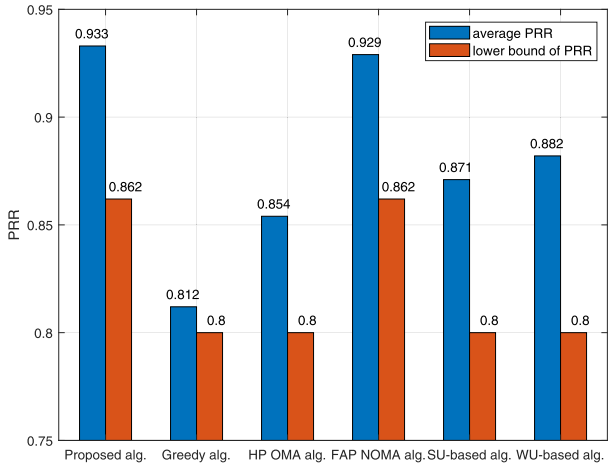


FIGURE 8. Average PRR and lower bound of PRR against different SPS schemes.

80%, and the speed is $30km/h$. Although the SPS period of "Greedy VUE-RB Matching" always closes to that of the proposed algorithm shown in Fig. 6 and Fig. 7, both the average PRR and the lower bound of PRR of the greedy algorithm are significantly lower than the proposed algorithm. We can observe that the reliability of "FAP NOMA" comes after the proposed one, since it is feasible to utilize fully analog beamforming for space division multiplexing when the reliability threshold is low, which is also shown in Fig. 7. It is also intuitive that the reliability of the proposed algorithm as well as the SPS period of this algorithm outperforms other schemes.

V. CONCLUSION

In this paper, the SPS strategy in mmWave-NOMA-based eV2X systems was studied, including hybrid precoding, user scheduling, and resource allocation. The VUE-RB matching problem with peer effects was solved by the proposed union-based two-phase matching algorithm. Specifically, in the clustering phase, BD&VC algorithm was designed to reduce the matching complexity; in the matching phase, LLHR-PC algorithm was proposed to provide evaluation indicators for scheduling. Simulation results showed that the scheduling period of eV2X systems was improved by the proposed SPS strategy compared with the conventional mmWave SPS schemes.

APPENDIXES

APPENDIX A

Proof OF THE NP-HARDNESS OF PROBLEM (16)

The reduced problem of (16), i.e., the VUE-RB matching problem, can be proved to be NP-hard, which implies the problem (16) is also NP-hard.

Given the power allocation factors $\{\alpha^i_{g,n}\}$, problem (16) can be reduced into a matching problem in terms of VUEs and RBs, where BS allocates RBs to VUEs for the minimization of SPS period. Then, the equivalence between the reduced problem and a vertex coloring problem can be established [38]. Specifically, the VUE set is viewed as the node set in the considered undirected conflicting graph, while the RB set is viewed as the color set. There exists an edge between two nodes, e.g., v and v', if they cannot be allocated to the same color, where the conditions of this case are listed as follows,

1) VUEs v and v' are in the same analog beam, i.e., corresponding to constraint (16c);

2) the reliability constraints of VUEs v and v' are not satisfied, i.e., $\min(\delta_v, \delta_{v'}) < \delta_{th}$, corresponding to constraint (15).

We aim at utilizing as few colors as possible to color all nodes such that no two adjacent vertices are with the same color. All VUEs, i.e., the dyed nodes, are promised to decode the desired signals successfully under the given conditions. Hence, the equivalence between the reduced problem and the above vertex coloring problem is established.

Since the above problem is NP-hard, the SPS problem in (16) is also NP-hard.

APPENDIX B

Proof OF PROPOSITION 1

We have known that the union $\mathcal{N}_{g,n}$ in set $\mathcal{N}$ satisfy $\mathcal{N} = \bigcup_{g=1}^{G_D} \bigcup_{n=1}^{|\mathcal{N}_g|} \mathcal{N}_{g,n}$ and $\mathcal{N}_{g,n} \cap \mathcal{N}_{g',n'} = \varnothing$, where $|g - g'| + |n - n'| \neq 0$. Several subsets of $\mathcal{N}$ have been defined as $\mathcal{N}_g \subseteq \mathcal{N}$, where $g \in \{1, 2, \ldots, G_D\}$. According to **Algorithm 2**, unions $\mathcal{N}_{g,n}, \mathcal{N}_{g,n'} \in \mathcal{N}_g$ are in the same analog beam space, i.e., the channel vectors of $\mathcal{N}_{g,n}$ and $\mathcal{N}_{g,n'}$ are strongly correlated. According to constraint (16c), such $\mathcal{N}_{g,n}$ and $\mathcal{N}_{g,n'}$ cannot be allocated in the same RB. Hence, if there exists $\mathcal{N}^j \subseteq \mathcal{N}$ such that $\mathcal{N}_{g,n} \in S\left(f_k, \mathcal{N}^j\right)$, we have $\mathcal{N}_{g,n} \in S\left(f_k, \mathcal{N}^j / \left\{\mathcal{N}_{g,n'}\right\}\right)$. Hence, the cluster alternative is proved.

APPENDIX C

Proof OF PROPOSITION 2

For $g \neq g'$, $\mathcal{N}_{g,n}$ and $\mathcal{N}_{g',n'}$ is allowed to be multiplexed in the same RB. The key question is whether there is a uniform preference between $\mathcal{N}_{g,n}$ and $\mathcal{N}_{g',n'}$. As non-independent individuals in $\mathcal{N}_{g,n}$, the VUEs $\mathcal{N}^i_{g,n}$ must be allocated in the same RB. According to **Algorithm 1**, the VUEs in the same RB have the same $\tilde{\delta}_{th}$, i.e., the VUEs in each union have a common preference when we evaluate the current matching result $\{\theta^k_{g,n}\}$. Hence, the cluster common preference is proved.

APPENDIX D

Proof OF PROPOSITION 3

The stability of matching Ψ is proved by contradiction [39].

Assume that there exists a RB $f_k \in F$ and a cluster $\mathcal{N}_{g,n} \in \mathcal{N}$, where the clusters in the considered eV2X system are divided into two types. On the one hand, for $\mathcal{N}_{g,n}$ that has not been scheduled, we have $\mathcal{R}\left(\mathcal{N}^2 f_k \mathcal{N}^1\right)$, where $\mathcal{N}^1 = \Psi(f_k)$, $\mathcal{N}^2 = \Psi(f_k) \cup \left\{\mathcal{N}_{g,n}\right\}$. On the other hand, for $\mathcal{N}_{g,n}$ that has been scheduled, we have $\min\{\tilde{\delta}^k_{th}(f_k, \Psi'(f_k)), \tilde{\delta}^{ind}_{th}(f_{ind}, \Psi'(f_{ind}))\} > \min(\tilde{\delta}^k_{th}, \tilde{\delta}^{ind}_{th})$, where $\Psi'(f_k) = \Psi(f_k) / \left\{\mathcal{N}_{g,n}\right\} \cup \left\{\mathcal{N}_{g,n'}\right\}$, $\Psi'(f_{ind}) = \Psi(f_{ind}) / \left\{\mathcal{N}_{g,n'}\right\} \cup \left\{\mathcal{N}_{g,n}\right\}$.

Apparently, $\mathcal{N}_{g,n}$ will be successfully proposed itself to f_k in the next round of the matching phase in **Algorithm 3** to pursue lower SPS period or higher reliability. However, the matching phase of f_k has already been terminated since the algorithm has converged to Ψ. Thus, $\mathcal{N}_{g,n}$ will fail to update $(f_k, \Psi(f_k))$ regardless of whether it has been scheduling, unless $\mathcal{N}_{g,n} \notin \mathcal{N}$, which is contradictory to our assumption. Hence, *Proposition 3* is proved.

REFERENCES

[1] B. Di, L. Song, Y. Li, and G. Y. Li, "Non-orthogonal multiple access for high-reliable and low-latency V2X communications in 5G systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2383–2397, Oct. 2017.

[2] *Study Enhancement 3GPP Support for 5G V2X Services, Release 16*, document 3GPP TR 22.886, Dec. 2018.

[3] *Enhancement 3GPP Support for V2X Scenarios, Release 16*, document 3GPP TR 22.186, Dec. 2018.

[4] N. Bonjorn, F. Foukalas, F. Canellas, and P. Pop, "Cooperative resource allocation and scheduling for 5G eV2X services," *IEEE Access*, vol. 7, pp. 58212–58220, 2019.

[5] J. Lianghai, A. Weinand, B. Han, and H. D. Schotten, "Feasibility study of enabling V2X communications by LTE-Uu radio interface," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Qingdao, China, Oct. 2017, pp. 1–6.

[6] A. Asadi, S. Muller, G. H. Sim, A. Klein, and M. Hollick, "FML: Fast machine learning for 5G mmWave vehicular communications," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Honolulu, HI, USA, Apr. 2018, pp. 1961–1969.

[7] G. Karadag, R. Gul, Y. Sadi, and S. Coleri Ergen, "QoS-constrained semi-persistent scheduling of machine-type communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2737–2750, May 2019.

[8] P. Wang, B. Di, H. Zhang, K. Bian, and L. Song, "Cellular V2X communications in unlicensed spectrum: Harmonious coexistence with VANET in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5212–5224, Aug. 2018.

[9] B. Di, L. Song, Y. Li, and G. Y. Li, "NOMA-based low-latency and high-reliable broadcast communications for 5G V2X services," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[10] A. Orsino, O. Galinina, S. Andreev, O. N. C. Yilmaz, T. Tirronen, J. Torsner, and Y. Koucheryavy, "Improving initial access reliability of 5G mmWave cellular in massive V2X communications scenarios," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–7.

[11] C. K. Anjinappa and I. Guvenc, "Millimeter-wave V2X channels: Propagation statistics, beamforming, and blockage," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–6.

[12] *Study on Evaluation Methodology of New Vehicle-to-Everything (V2X) Use Cases for LTE and NR, Release 15*, document TR 37.885, 3GPP, Dec. 2018.

[13] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter wave communication in vehicular networks: Challenges and opportunities," in *Proc. 6th Int. Conf. Modern Circuits Syst. Technol. (MOCAST)*, Thessaloniki, Greece, May 2017, pp. 1–6.

[14] S. Mumtaz, J. Miquel Jornet, J. Aulin, W. H. Gerstacker, X. Dong, and B. Ai, "Terahertz communication for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5617–5625, Jul. 2017.

[15] S.-Y. Lien, Y.-C. Kuo, D.-J. Deng, H.-L. Tsai, A. Vinel, and A. Benslimane, "Latency-optimal mmWave radio access for V2X supporting next generation driving use cases," *IEEE Access*, vol. 7, pp. 6782–6795, 2019.

[16] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.

[17] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[18] D. Zhang, Y. Liu, L. Dai, A. K. Bashir, A. Nallanathan, and B. Shim, "Performance analysis of FD-NOMA-Based decentralized V2X systems," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5024–5036, Jul. 2019.

[19] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Non-orthogonal multiple access vehicular small cell networks: Architecture and solution," *IEEE Netw.*, vol. 31, no. 4, pp. 15–21, Jul./Aug. 2017.

[20] F. L. Luo and C. J. Zhang, *Signal Processing for 5G: Algorithms and Implementations*, London, U.K.: Wiley, 2016, pp. 143–166.

[21] B. Di, L. Song, Y. Li, and Z. Han, "V2X meets NOMA: Non-orthogonal multiple access for 5G-enabled vehicular networks," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 14–21, Dec. 2017.

[22] C. Chen, W. Cai, X. Cheng, L. Yang, and Y. Jin, "Low complexity beamforming and user selection schemes for 5G MIMO-NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2708–2722, Dec. 2017.

[23] B. Wang, L. Dai, X. Gao, and L. Hanzo, "Beamspace MIMO-NOMA for millimeter-wave communications using lens antenna arrays," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, USA, Sep. 2017, pp. 1–5.

[24] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "A multi-beam NOMA framework for hybrid mmWave systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–7.

[25] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "User selection and power allocation for mmWave-NOMA networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[26] K. J. Choi and K. S. Kim, "Optimal semi-persistent uplink scheduling policy for large-scale antenna systems," *IEEE Access*, vol. 5, pp. 22902–22915, 2017.

[27] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, "Capacity analysis of NOMA with mmWave massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1606–1618, Jul. 2017.

[28] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.

[29] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[30] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[31] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 611–615.

[32] Y. Alsaba, C. Y. Leow, and S. K. Abdul Rahim, "Full-duplex cooperative non-orthogonal multiple access with beamforming and energy harvesting," *IEEE Access*, vol. 6, pp. 19726–19738, 2018.

[33] Y. Chen, L. Wang, Y. Ai, B. Jiao, and L. Hanzo, "Performance analysis of NOMA-SM in Vehicle-to-Vehicle massive MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2653–2666, Dec. 2017.

[34] Z. Yang, W. Xu, and Y. Li, "Fair non-orthogonal multiple access for visible light communication downlinks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 66–69, Feb. 2017.

[35] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA, USA: Athena Scientific, 2009.

[36] *Slot Mini-Slot Numerology Alignment*, document 3GPP R1-1609504, TSG 537 RAN WG1 Meeting #86bis, Lisbon, Portugal, Oct. 2016.

[37] K. Lee, J. Kim, Y. Park, H. Wang, and D. Hong, "Latency of cellular-based V2X: Perspectives on TTI-proportional latency and TTI-independent latency," *IEEE Access*, vol. 5, pp. 15800–15809, 2017.

[38] T. Jensen and B. Toft, *Graph Coloring Problems*. New York, NY, USA: Wiley, 1995.

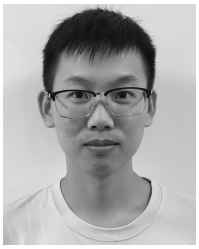
[39] P. Wang, B. Di, H. Zhang, X. Hou, and L. Song, "Cellular V2X communications in unlicensed spectrum for 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

BICHENG WANG (Student Member, IEEE) received the B.Eng. degree in communication engineering from Soochow University, Suzhou, China, in 2017, where he is currently pursuing the master's degree with the School of Electronic and Information Engineering.

His research interests include vehicular networks, non-orthogonal multiple access, and embedded systems.

RUOQI SHI received the B.Eng. degree in communication engineering from Soochow University, Suzhou, China, in 2018, where he is currently pursuing the master's degree with the School of Electronic and Information Engineering.

His research interests include wireless communications and signal processing.



FANWEI SHI is currently pursuing the B.Eng. degree with the School of Electronic and Information Engineering, Soochow University, Suzhou, China.

His research interests include wireless communications and signal processing.

JIANLING HU (Member, IEEE) received the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University, Shanghai, China, in 2000.

From 2001 to 2008, he worked as an Associate Professor at the Department of Electronic Engineering, Shanghai Jiao Tong University. Since 2008, he has been a Professor with the School of Electronic and Information Engineering, Soochow University, Suzhou, China. His research interests include wireless communications, embedded system design, and multimedia signal processing.