Fast-Adapting V2X Networks: A Dynamic Context Clustering Multi-Armed Bandits

Weiqi Chi*, Xiaoyang He†, and Manabu Tsukada*
*Graduate School of Information Science and Technology, The University of Tokyo, Japan
Email: {weiqichi, mtsukada}@g.ecc.u-tokyo.ac.jp
†Graduate School of Electronics and Communication Engineering, Sun Yat-sen University, China
Email: hexy35@mail2.sysu.edu.cn

***Abstract*—While contextual multi-armed bandit (CMAB) effectively addresses non-stationary reward for problems such as millimeter-wave (mmWave) Vehicle-to-Everything (V2X) user association (UA), it suffers from slow convergence when dealing with a large-scale context space. This paper proposes a Dynamic Context Clustering Semi-distributed Multi-Armed Bandit (DCC-SMAB) algorithm for the UA problem in mmWave vehicular networks that dynamically clusters vehicle positions as contexts, thereby accelerating convergence in reward estimation without requiring additional channel state information (CSI) compared to the CMAB framework with fixed context partitioning. The algorithm exhibits strong capability in identifying mmWave channel variations, enabling rapid adaptation to non-stationary communication environments for both newly arriving and existing vehicles in the network. Extensive simulations demonstrate over 61% and 76% in average regret reduction and network transmission rate improvement compared to benchmarks using Upper Confidence Bound (UCB) policy, with robustness validated across 40 trials under varying network topologies.**

***Index Terms*—Contextual multi-armed bandit, user association, mmWave, vehicle-to-everything communication**

I. Introduction

Autonomous vehicles are required to exchange massive sensor information to address visibility constraints in dense traffic scenarios [1], especially for emerging applications such as cooperative driving and collaborative perception, which demand frequent sharing of large amounts of data. The high-frequency millimeter-wave (mmWave) provides a wide spectrum to support the high transmission rate in vehicle-to-everything (V2X) communication, which is beyond the capabilities of conventional communication systems. Nevertheless, significant technical challenges remain in mmWave vehicular networks, where frequent blockage results in limited coverage and severe channel quality fluctuations [2], leading to unstable connectivity. Furthermore, the mmWave user association (UA) becomes challenging due to the need to jointly optimize coupled objectives such as transmission rate and load balancing under these dynamic propagation conditions [3].

Multi-armed bandit (MAB) frameworks are promising for UA problems due to their ability to balance exploration and exploitation in real-time learning without requiring offline training. Previous study [4] utilizes MAB to address UA problems in static networks, while [5] additionally considers base station (BS) load balancing and users' mobility. Contextual MAB (CMAB) extends MAB by incorporating context, with the main goal of learning to map from the context space to reward distribution [6]. The CMAB framework is used to model the non-stationary reward distributions in mmWave vehicular networks [7], where the context space is divided into predefined hypercubes. Though in [8], the authors turn the context into a Hilbert space to linearize the mapping, convergence performance degrades substantially as network scale and context space grow. Both [9] and [10] present clustering approaches for CMAB with stationary reward distributions, limiting their applicability in mmWave V2X networks where reward varies with channel conditions. To conclude, though mmWave communication offers low latency suitable for advanced V2X applications, its non-stationary channel characteristics pose significant challenges to the UA problem. This necessitates the development of UA algorithms with fast convergence, real-time decision-making capabilities, and strong adaptability to dynamic mmWave vehicular environments.

In this paper, we propose a dynamic context clustering semi-distributed MAB (DCC-SMAB) algorithm for UA in mmWave vehicular networks. The key contributions are fourfold: 1) We propose a dynamic clustering approach that integrates similar context-reward patterns into a dynamic number of clusters as knowledge inputs for subsequent learning processes. Unlike fixed context partitioning that ignores similarities across predefined partitions, this approach reduces redundant explorations among vehicles with similar context, thereby accelerating the learning rate. 2) The proposed algorithm periodically updates clusters to track reward distribution shifts caused by mmWave channel variations, enabling rapid adaptation to non-stationary communication environments. 3) The semi-distributed mechanism in DCC-SMAB further accelerates convergence by enabling vehicles leverage pre-learned cluster knowledge when their contexts match existing clusters. This eliminates cold-start exploration for new arrivals and reduces adaptation time for vehicles whose contexts have shifted. 4) DCC-SMAB requires no prior channel state information (CSI) by estimating channel quality through vehicle positions as contexts, making it practical for highly dynamic mmWave vehicular networks where CSI acquisition is costly and often outdated.

The rest of this paper is organized as follows. Section II introduces the system model. Section III presents the proposed DCC-SMAB algorithm in detail with regret analysis. Section IV and V provide numerical results and a conclusion.

II. System Model

We investigate a heterogeneous network (HetNet) in mmWave vehicular communication where a sub-6 GHz macro base station (MBS) overlays multiple mmWave sub-BSs (SBSs). The MBS functions as both an access point and a center unit for learning tasks, while SBSs enable advanced mmWave V2X applications. The rest of this section presents the vehicle mobility model and the channel model. Followed by the optimization problem formulation.

A. Mobility and Channel Model

Consider a finite time horizon $T \in \mathbb{N}$ with discrete time steps $t = 1, \ldots, T$. At each time step, vehicles arrive in the network following a Poisson distribution. Vehicles update their states at each time period. We consider a set of SBSs denoted as $\mathcal{J} = \{1, \ldots, j, \ldots, |\mathcal{J}|\}$, where $|\mathcal{J}|$ represents the total number of SBSs deployed for mmWave vehicular network operation. For each time step, let $\mathcal{V}^t = \{1, \ldots, k, \ldots, |\mathcal{V}^t|\}$ represent the vehicle set with $|\mathcal{V}^t|$ vehicles. Each SBS is equipped with a massive antenna array with M_j elements, enabling serving multiple vehicles. Each vehicle k is equipped with antenna arrays for both sub-6 GHz and mmWave frequency bands. For the sub-6 GHz frequency band, we adopt the widely used Gaussian MIMO channel model. For the mmWave frequency band, where one vehicle is constrained to associate with only one SBS, the channel gain between the SBS j and the vehicle k is given as:

$$y_{k,j}(t) = \mathbf{h}_{k,j}(t)\mathbf{w}_{k,j}, \tag{1}$$

where $\mathbf{h}^t_{k,j}$ and $\mathbf{w}_{k,j}$ denote the channel matrix and beamforming vector from vehicle k to SBS j, respectively. We focus on large-scale fading characteristics, including path loss and shadowing induced by buildings in the urban scenario. Since orthogonal frequency allocation is adopted across SBSs, inter-cell interference is negligible, and only intra-cell interference needs to be considered. Specifically, when vehicle i simultaneously communicates with SBS j at time t, it introduces interference to the ongoing transmission toward vehicle k, characterized by the interference channel coefficient $\tilde{y}_{i,j}(t)$. The total interference at SBS j while serving vehicle k is therefore aggregated over all vehicles that simultaneously communicate with SBS j:

$$I_{k,j}(t) = \sum_{i \in (\mathcal{V}^t \setminus k)} P_v \left|\tilde{y}_{i,j}(t)\right|^2 \mathcal{I}^t_{i,j} + N_o W, \tag{2}$$

where the W is the bandwidth of SBS j, P_v is the transmission power of vehicles, N_o is the noise power density and $\mathcal{I}_{i,j}(t) = 1$ indicates i is associating with SBS j, otherwise $\mathcal{I}_{i,j}(t) = 0$.

B. Optimization Problem

According to (1) and (2), the instantaneous transmission rate between vehicle k to SBS j:

$$R^t_{k,j} = W \log_2\left(1 + \frac{P_v |y_{k,j}(t)|^2}{I_{k,j}(t)}\right), \tag{3}$$

In each period t, vehicle k is associated with one SBS. Let $\boldsymbol{\eta}^t$ indicate the vector of the selected SBS's index for vehicle set $\mathcal{V}^t$. The association between vehicles and SBSs can be defined as:

$$\boldsymbol{\eta}^t \triangleq [\eta^t_1, \ldots, \eta^t_k, \ldots, \eta^t_{|\mathcal{V}^t|}]. \tag{4}$$

The user association optimization problem aims to maximize the sum rate of the vehicular network by finding the optimal association vector:

$$\begin{aligned} \max_{\boldsymbol{\eta}} \quad & r(\boldsymbol{\eta}^t) = \sum_{k \in \mathcal{V}^t} R^t_{k,\eta_k} \\ \text{s.t.} \quad & \sum_{j=1}^{|\mathcal{J}|} \mathcal{I}^t_{k,j} = 1, \quad \forall k \in \mathcal{V}^t. \end{aligned} \tag{5}$$

The subject constraint in (5) shows that each vehicle could only connect to one SBS while SBS can serve multiple vehicles at time step t. The optimization problem presented in (5) exhibits NP-hardness characteristics attributed to the combination of non-convex nonlinear constraints and integer variables [11]. While exhaustive search can identify optimal associations with complete CSI knowledge, it becomes impractical in highly dynamic vehicular networks due to the high computational complexity and the accuracy of CSI estimation.

III. Dynamic Context Clustering Semi-distributed MAB Algorithm

As indicated in [12], the mmWave vehicular channel exhibits extremely short channel coherence time. This not only makes frequent channel estimation costly, but also renders the optimization problem a "one-shot" optimization. By leveraging the CMAB framework, vehicles can predict the reward based on vehicles' contexts, obviating the need for continuous CSI acquisition and substantially alleviating computational complexity. Critically, the proposed DCC-SMAB algorithm employs a dynamic hierarchical context clustering approach that identifies and exploits similarities across context boundaries, in stark contrast to fixed context partitioning schemes that treat predefined partitions independently. Furthermore, the periodic cluster updates track reward distribution shifts in non-stationary mmWave environments, while the semi-distributed architecture enables vehicles to exploit pre-learned cluster knowledge, eliminating cold-start delays. In this section, we first introduce the system architecture, followed by a detailed description of the dynamic context clustering mechanism and regret analysis.

A. DCC-SMAB Architecture

The proposed algorithm architecture is illustrated in Fig. 1, which operates through three interconnected stages: *distributed learning*, *dynamic hierarchical context clustering*, and *semi-distributed learning*. In the distributed learning stage, each vehicle observes its current position as context and selects an SBS based on dedicated policies independently, such as Upper Confidence Bound (UCB) or Thompson Sampling (TS). Upon completing the association and receiving an instantaneous reward, the vehicle updates its reward estimation vector across all arms. This estimation vector, along with the vehicle's current context, is recorded as a tuple of historical samples. These accumulated samples are periodically uploaded to the MBS via sub-6 GHz V2X links, providing the necessary sample data for centralized coordination.

The framework then proceeds to the dynamic context clustering stage, where the MBS executes a hierarchical clustering algorithm to identify similar context-reward patterns. In contrast to fixed partitioning with rigid boundaries, this approach dynamically clusters similar context-reward patterns into a flexible number of groups, enabling more effective learning. Each cluster is characterized by a context centroid and corresponding expected rewards for all SBSs. As historical samples from various vehicles are uploaded, the cluster knowledge is increased and enhanced by identifying new clusters or merging the close clusters in MBS, enabling the algorithm to effectively track reward distribution shifts caused by mmWave channel variations and non-stationary environments.

Finally, in the semi-distributed learning stage, vehicles query the MBS to determine whether to inherit pre-learned cluster knowledge before executing their local policies. By identifying the cluster closest to a vehicle's current context, the MBS facilitates knowledge inheritance to optimize the learning process. This mechanism specifically addresses two critical challenges: first, it eliminates cold-start issues for newly arriving vehicles by allowing them to bypass initial exploration; second, it enables rapid adaptation for existing vehicles when their contexts shift significantly. By leveraging shared learning experiences across the network, this semi-distributed approach substantially reduces exploration overhead and accelerates convergence compared to purely distributed learning.

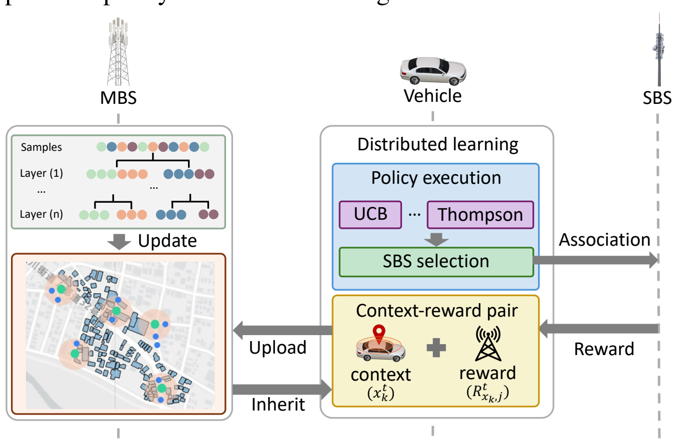

Fig. 1. System Architecture Overview

B. Multi-Agent-CMAB Framework

The Multi-Agent-CMAB (MA-CMAB) framework extends the conventional CMAB by allowing multiple agents to make simultaneous decisions, facilitating cooperative learning in complex wireless networking environments [13]. In mmWave vehicular networks, each vehicle $k \in \mathcal{V}^t$ at time step t acts as an agent, while the set of available SBSs $j \in \mathcal{J}$ represents the arms. At each time step, vehicle k identifies its current position as the context x_k^t within the context space $\mathcal{X}$. Based on this context, the vehicle follows a policy to select an SBS $\eta_{x_k}^t \in \mathcal{J}$ and subsequently receives an instantaneous reward $R^t_{x_k,j}$. To effectively balance exploration and exploitation, we mainly introduce the UCB policy [14] employed in this paper. Specifically, each vehicle calculates the UCB value $u^t_{x_k,j}$ for each SBS j under context x_k^t as follows:

$$u^t_{x_k,j} = \Gamma^{t-1}_{x_k,j} + \sqrt{\frac{2\ln t}{n^{t-1}_{x_k,j}}}, \tag{6}$$

where $n^t_{x_k,j}$ denotes the cumulative number of associations with SBS j under context $\mathbf{x}_k$, and $\Gamma^t_{x_k,j} = \mathbb{E}[R^t_{k,j}]$ represents the expected reward vector. The second term in (6) serves as the confidence bound, which incentivizes the exploration of less-visited SBSs. Consequently, the vehicle k associates with the SBS that maximizes the upper confidence bound:

$$\eta^t_{x_k} = \arg\max_{j\in\mathcal{J}} u^t_{x_k,j}. \tag{7}$$

Algorithm 1 presents the complete UCB-based UA process. The proposed DCC-SMAB algorithm leverages this MA-CMAB framework to facilitate collaborative learning by aggregating distributed knowledge learned from multiple vehicles in MBS and knowledge inheritance for vehicles when necessary. This allows vehicles to exploit pre-learned knowledge rather than learning distributedly, effectively eliminating cold-start delays and accelerating convergence in non-stationary mmWave environments.

Algorithm 1: CMAB UA using UCB

```
Input: Γ^{t-1}_{x_k,j}, n^{t-1}_{x_k,j} for all j ∈ J
Identify vehicle context x_k^t;
for j ∈ J do
    Calculate UCB value u^t_{x_k,j} according to (6);
end
Select SBS η^t_{x_k} using (7);
Receive instantaneous reward R^t_{x_k,η_{x_k}};
Update number of pulls: n^t_{x_k,η_{x_k}} = n^{t-1}_{x_k,η_{x_k}} + 1;
Update expected reward:
  Γ^t_{x_k,η_{x_k}} = (Γ^{t-1}_{x_k,η_{x_k}} · n^{t-1}_{x_k,η_{x_k}} + R^t_{x_k,η_{x_k}}) / n^t_{x_k,η_{x_k}};
Output: η^t_{x_k}, n^t_{x_k,η_{x_k}}, Γ^t_{x_k,η_{x_k}}
```

C. Dynamic Hierarchical Clustering in MBS

The MBS maintains cluster knowledge by periodically aggregating historical samples uploaded from all participating vehicles and executing a hierarchical clustering algorithm. Specifically, the proposed method performs hierarchical partitioning on the sample pool, treating each element of the reward estimation vector as a distinct hierarchical layer. In this cascaded approach, samples are clustered through successive layers, where similarity is rigorously evaluated based on a specific vector dimension at each layer to progressively refine cluster granularity. Consequently, the resulting clusters represent distinct context-reward patterns synthesized from the collective experience of multiple vehicles. By aggregating samples that exhibit similarity across all hierarchical reward layers, the MBS extracts a generalized mapping between context and reward distribution, which serves as the pre-learned cluster knowledge for subsequent training phases.

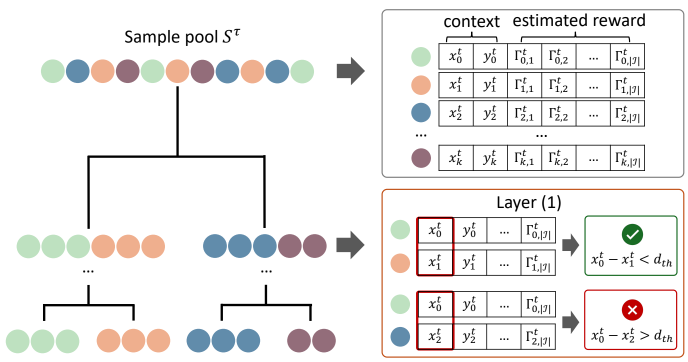

Fig. 2. Illustration of the hierarchical clustering algorithm.

As depicted in Fig. 2, at each cluster update round τ, the MBS aggregates all uploaded samples from vehicles into a sample pool, denoted as $\mathbf{S}^{\tau}$(Distinct colors of samples denote different context-reward patterns for clarity). Each sample consists of the vehicle's current context and the expected reward vector $[\mathbf{x}_k^t, \mathbf{\Gamma}_{x_k}^t]$. The hierarchical clustering algorithm leverages the fundamental principles of the K-Dimensional Tree (KD-Tree) [15] framework. Specifically, we define a threshold vector $\boldsymbol{\theta} = [d_{th}, r_{th}]$, where d_{th} specifies the maximum spatial distance and r_{th} defines the maximum reward vector difference permitted within each hierarchical layer to determine sample similarity. After similar samples are aggregated into one cluster, the MBS derives the cluster centroid by computing the mean of all samples within cluster q, denoted as $\mathbf{c}_q^{\tau} = [\bar{\mathbf{x}}_q^{\tau}, \bar{\mathbf{\Gamma}}_q^{\tau}]$, where $\bar{\mathbf{x}}_q^{\tau}$ and $\bar{\mathbf{\Gamma}}_q^{\tau}$ represent the average position and average reward vector of the samples in cluster q, respectively.

Following the hierarchical clustering process, the MBS initially generates a set of pseudo-clusters for cluster round τ, denoted as $\tilde{\mathcal{C}}^{\tau} = \{\tilde{\mathbf{c}}_1^{\tau}, \ldots, \tilde{\mathbf{c}}_m^{\tau}, \ldots, \tilde{\mathbf{c}}_M^{\tau}\}$, which represent new clusters that contain the context-reward patterns extracted from the current sample pool $\mathbf{S}^{\tau}$. The MBS subsequently executes a merging procedure that integrates these newly formed pseudo-clusters with the existing cluster set. Specifically, the fusion of clusters is determined exclusively by the context proximity between their centroids. This logic is based on the observation that vehicles at close geographical positions exhibit similar channel characteristics. Through this dynamic adjustment, the MBS maintains an updated set of cluster centroids $\mathcal{C}^{\tau} = \{\mathbf{c}_1^{\tau}, \ldots, \mathbf{c}_q^{\tau}, \ldots, \mathbf{c}_Q^{\tau}\}$ as the pre-learned cluster knowledge, where the total number of clusters Q adapts evolutionarily to the varying vehicular network environment. Algorithm 2 details how samples and clusters are handled and updated in the MBS at cluster update round τ

During the semi-distributed learning stage, a vehicle initiates an inquiry to the MBS under two specific conditions: 1) newly arriving vehicles inherit from the cluster knowledge if their contexts (i.e., positions) fall within existing clusters, thereby bypassing initial exploration; 2) existing vehicles request updates when their context shifts significantly from the previous learning period. Upon inquiry, the MBS assigns the corresponding cluster knowledge to the vehicle if its current context is sufficiently close to a cluster centroid. This hierarchical clustering architecture accelerates convergence by enabling collective learning across similar contexts, exploiting low-dimensional position-based representations, and adapting to reward distribution shifts through periodic cluster updates.

Algorithm 2: Cluster Update in MBS

Input: $\boldsymbol{\theta}$, $\mathcal{C}^{\tau-1}$, $\mathcal{V}^t$
for *each* $k \in \mathcal{V}^t$ **do**
  Vehicle upload samples to sample pool $\mathbf{S}^{\tau}$, $\forall k \in \mathcal{V}^t$;
  Clear vehicle sample storage ;
end
Pesudo-clusters: $\tilde{\mathcal{C}}^{\tau} \leftarrow$Hierarchical clustering $(\mathbf{S}^{\tau}, \boldsymbol{\theta})$;
Centroids merging:
for *each* $\mathbf{c}_n^{\tau-1} \in \mathcal{C}^{\tau-1}, \tilde{\mathbf{c}}_m^{\tau} \in \tilde{\mathcal{C}}^{\tau}$ **do**
  $\mathbf{c}_q^{\tau} \leftarrow \begin{cases} \text{mean}(\mathbf{c}_n^{\tau-1}, \tilde{\mathbf{c}}_m^{\tau}) & \text{if mergeable} \\ \tilde{\mathbf{c}}_m^{\tau} & \text{otherwise} \end{cases}$
  Central update: $\mathcal{C}^{\tau} \leftarrow \mathbf{c}_q^{\tau}$;
end
Output: Updated cluster knowledge $\mathcal{C}^{\tau}$

Finally, the complete DCC-SMAB algorithm is presented in Algorithm 3, which incorporates the three primary processes described previously within the MA-MAB framework.

Algorithm 3: DCC-SMAB

Input: t, τ, $\mathcal{C}^{(0)}$, $\mathcal{V}^t$, $\mathcal{J}$, $\boldsymbol{\theta}$
for $t \in T$ **do**
  if $t = \tau$ **then**
    Execute cluster update (Algorithm 2);
  end
  for *each* $k \in \mathcal{V}^t$ **do**
    Identify vehicle context: $\mathbf{x}_k^t$;
    Position difference: $\Delta d = \|\mathbf{x}_k^t - \mathbf{x}_k^{t-1}\|$;
    if $k \notin \mathcal{V}^{t-1}$ *OR* $\Delta d > d_{th}$ **then**
      Inherit cluster knowledge;
        $q^* \leftarrow \arg\min_q \|\mathbf{x}_k^t - \bar{\mathbf{x}}_q^{\tau}\|$, $\mathbf{c}_{q^*}^{\tau} \in \mathcal{C}^{\tau}$;
        $\mathbf{\Gamma}_k^t \leftarrow \bar{\mathbf{\Gamma}}_{q^*}^{\tau}$;
    end
    Execute Algorithm 1: $\eta_{x_k}^t$, $\Gamma_{x_k, \eta_{x_k}}^t$;
    Update $\mathbf{\Gamma}_{x_k}^t$ and store current sample: $[\mathbf{x}_k^t, \mathbf{\Gamma}_{x_k}^t]$;
  end
end
Output: System association vector $\boldsymbol{\eta}$

D. Regret Analysis

In this work, we adopt the notion of expected regret [16], defined as the expectation of the difference between the rewards of the optimal sequence of actions and those chosen by the algorithm. For vehicle k, the expected regret up to step T is:

$$\mathbb{E}[\mathcal{R}_k] = \mathbb{E}\left[\max_{\eta_k^t \in \mathcal{J}} \sum_{t=1}^{T} R_{k,\eta_k^t}^t - \sum_{t=1}^{T} R_{k,\eta_k^t}^t\right]. \tag{8}$$

For the computation of the expected regret in the following simulation, the optimal action is determined via an offline oracle that identifies the SBS providing the highest instantaneous communication rate for vehicle k at each time step t, based on the complete knowledge of channel realizations and interference levels. Our simulated result in Section IV shows the fast convergence compared with benchmark methods.

IV. Numerical Results

We evaluate the proposed DCC-SMAB algorithm in a mmWave HetNet scenario using large-scale simulations over a 1.9×1.9 km urban area. The simulation setup approximates real-world deployment by incorporating actual base station locations from OpenCelliD [17], realistic vehicular traffic generated by SUMO [18] following the recommended evaluation scenario in [19], and urban road topology from OpenStreetMap [20] for the Hachioji district in Tokyo, Japan. The mmWave channel is modeled using the Clustered Delay Line (CDL) model with ray tracing [21] to simulate channel fluctuations caused by static building obstacles. Key simulation parameters are listed in TABLE I. The proposed DCC-SMAB is compared against stochastic MAB and CMAB benchmarks, with both UCB and TS policies evaluated.

TABLE I
Simulation Parameters and Settings

Category	Parameter	Value
Network Setup	Bandwidth	50 MHz
	SBS carrier frequency	28 GHz
	SBS Antenna array	4×4
	Transmit power	$15 \sim 35$ dBm
	Noise power spectral density	-174 dBm/Hz
Traffic Parameters	Number of SBSs	12
	Height of SBSs	5 m
	Simulation area	1.9×1.9 km
	Vehicle Dimensions	$5 \times 2 \times 0.75$ m (20%)
		$5 \times 2 \times 1.6$ m (50%)
		$13 \times 2.6 \times 3$ m (30%)

Fig. 3 illustrates the average regret convergence, defined as the average cumulative expected regret across all vehicles. At 35 dBm transmit power, DCC-SMAB demonstrates stable and rapid convergence, while baseline methods exhibit fluctuations and fail to converge within the given simulation time. The proposed algorithm achieves significant performance improvements of 65.58% and 61.31% over MAB and CMAB with UCB policy, and 49.03% and 47.07% over MAB and CMAB with TS policy, respectively.

The CMAB approach fails to converge within the limited learning duration due to the large context space. Moreover, it exhibits significant fluctuations, with a sharp regret increase after time step 30, coinciding with new vehicle arrivals in

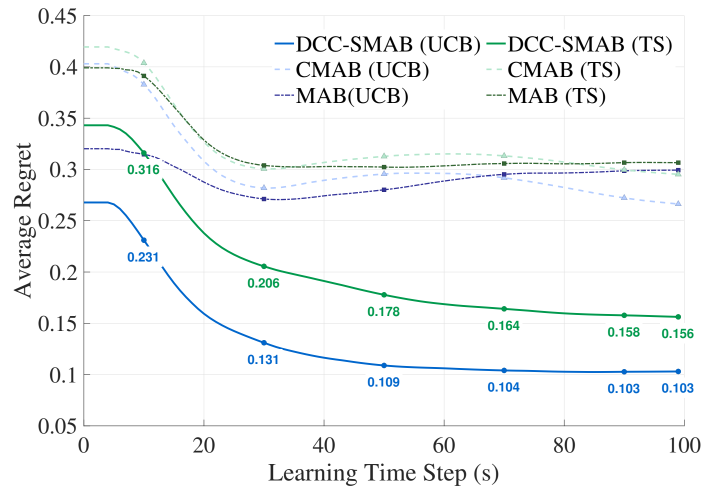

Fig. 3. Average regret comparison across three methods

previously unexplored context partitions. Without prior knowledge of these novel contexts, vehicles must restart exploration from scratch. While the stochastic MAB method avoids context space explosion, it cannot capture the non-stationary mmWave channel conditions during vehicle mobility, resulting in persistent fluctuations throughout the learning process. In contrast, the proposed DCC-SMAB method addresses both limitations through semi-distributed learning with dynamic context clustering. This approach effectively reduces the effective dimensionality of the context space while capturing communication channel dynamics. Furthermore, the centralized knowledge aggregation enables newly arriving vehicles to inherit relevant historical information, allowing them to rapidly assess network conditions and make informed association decisions without cold-start exploration.

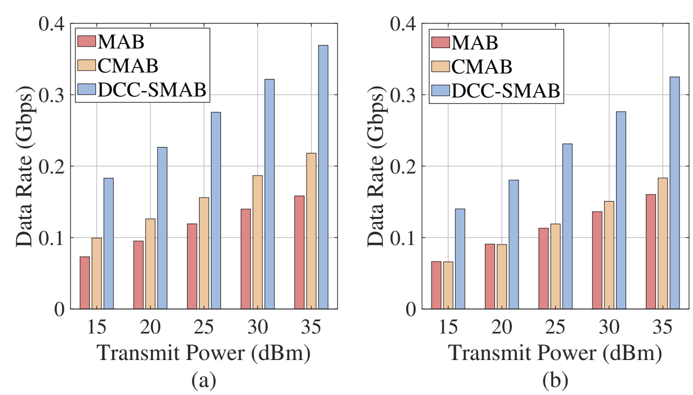

Fig. 4. Transmission rate comparison across three methods. (a) UCB policy (b) TS policy

Fig. 4 illustrates the network transmission rate under various vehicle transmit powers using both UCB and TS policies. The proposed DCC-SMAB algorithm demonstrates substantial performance gains across all power levels. Specifically, DCC-SMAB achieves average improvements of 136.57% and 76.40% over CMAB and MAB with the UCB policy, and 103.82% and

93.33% over the corresponding benchmarks with the TS policy. As shown in both Fig. 3 and Fig. 4, UCB consistently outperforms TS in this vehicular scenario. This performance gap is primarily due to UCB's deterministic exploration strategy, which is more effective at handling environmental dynamics than the stochastic approach of TS. Furthermore, the frequent arrival and departure of vehicles provide insufficient temporal samples for TS to achieve accurate posterior distribution estimation, leading to slower convergence compared to the proposed scheme.

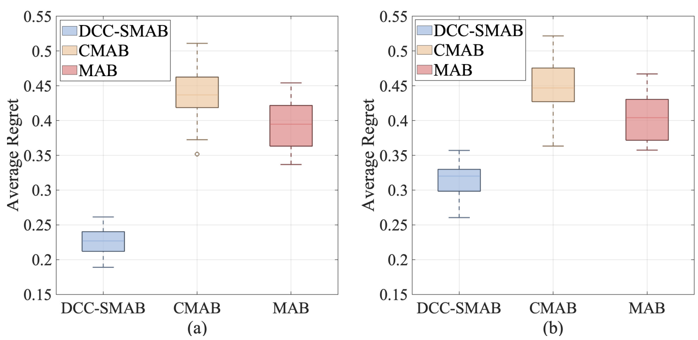

Fig. 5. Algorithm robustness validation. (a) UCB policy (b) TS policy

Fig. 5 presents a box plot analysis to evaluate algorithm robustness across 40 trials with different SBS location configurations. DCC-SMAB achieves a significantly lower median average regret compared to both MAB and CMAB benchmarks, regardless of the policy used, which validates the effectiveness of centralized knowledge sharing. Moreover, the proposed method exhibits a notably narrower interquartile range, indicating higher performance stability across varying network topologies and vehicle densities. By leveraging the pre-learned knowledge organized in hierarchical layers, DCC-SMAB effectively mitigates the performance variance induced by environmental fluctuations and channel condition changes, ensuring consistent and reliable behavior in diverse vehicular network deployments.

V. Conclusion

This paper proposed the DCC-SMAB algorithm for user association in mmWave vehicular networks. By identifying contextual similarities, the algorithm dynamically formed clusters that provided centralized knowledge for newly arriving vehicles and adapted to context shifts of existing ones. Simulation results demonstrated fast and stable convergence without prior CSI knowledge, yielding substantial improvements over benchmarks in both convergence rate and network transmission rate. Robustness analysis across 40 independent trials further confirmed consistent performance under diverse BS deployments and varying network conditions. Future work will focus on formalizing computational and signaling complexity to ensure scalability in ultra-dense deployments, and on expanding discussions of limitations — particularly in scenarios where highly heterogeneous channels at similar spatial locations may render position-based context insufficient for distinguishing reward distributions.

Acknowledgment

This work was supported by JST SPRING, Grant Number JPMJSP2108.

References

[1] V. Va *et al.*, "Millimeter wave vehicular communications: A survey," *Foundations and Trends® in Networking*, vol. 10, no. 1, pp. 1–113, 2016. [Online]. Available: http://dx.doi.org/10.1561/1300000054

[2] C. G. Ruiz *et al.*, "Analysis of blocking in mmWave cellular systems: Application to relay positioning," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1329–1342, 2021.

[3] M. L. Attiah *et al.*, "A survey of mmWave user association mechanisms and spectrum sharing approaches: An overview, open issues and challenges, future research trends," *Wireless Networks*, vol. 26, no. 4, pp. 2487–2514, May 2020.

[4] S. Maghsudi *et al.*, "Distributed user association in energy harvesting small cell networks: A probabilistic model," *CoRR*, vol. abs/1601.07795, 2016. [Online]. Available: http://arxiv.org/abs/1601.07795

[5] A. Alizadeh *et al.*, "Reinforcement learning for user association and handover in mmWave-enabled networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9712–9728, 2022.

[6] S. C. Hoi *et al.*, "Online learning: A comprehensive survey," *Neurocomputing*, vol. 459, pp. 249–289, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221006706

[7] G. H. Sim *et al.*, "An online context-aware machine learning algorithm for 5G mmWave vehicular communications," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2487–2500, 2018.

[8] X. He *et al.*, "Learning-based user association for mmWave vehicular networks with kernelized contextual bandits," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*. Milan, Italy: IEEE, March 2025, pp. 1–6.

[9] Y. Ban *et al.*, "Local clustering in contextual multi-armed bandits," *CoRR*, vol. abs/2103.00063, 2021. [Online]. Available: https://arxiv.org/abs/2103.00063

[10] C. Yan *et al.*, "Dynamic clustering based contextual combinatorial multi-armed bandit for online recommendation," *Knowledge-Based Systems*, vol. 257, p. 109927, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705122010206

[11] Z. Mlika *et al.*, "User association under SINR constraints in HetNets: Upper bound and NP-hardness," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1672–1675, 2018.

[12] V. Va *et al.*, "The impact of beamwidth on temporal channel variation in vehicular channels and ITS implications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5014–5029, 2017.

[13] S. Maghsudi *et al.*, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, 2016.

[14] A. Slivkins, "Introduction to multi-armed bandits," *CoRR*, vol. abs/1904.07272, 2019.

[15] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, p. 509–517, Sep. 1975. [Online]. Available: https://doi.org/10.1145/361002.361007

[16] S. Bubeck *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," 2012. [Online]. Available: https://arxiv.org/abs/1204.5721

[17] "OpenCellID – the world's largest open database of cell towers," https://opencellid.org, accessed: 2024.

[18] P. A. Lopez *et al.*, "Microscopic traffic simulation using SUMO," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018.

[19] 3GPP, "Study on evaluation methodology of new vehicle-to-everything (V2X) use cases for LTE and NR," 3rd Generation Partnership Project (3GPP), Technical Report TR 37.885, June 2019, release 15.

[20] OpenStreetMap Contributors, "Planet dump retrieved from https://planet.osm.org," https://www.openstreetmap.org, 2017.

[21] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901, June 2025, release 19.