Knowledge Base Aware Semantic Communication in Vehicular Networks

Le Xia*, Yao Sun*, Dusit Niyato†, Kairong Ma*, Jiawen Kang‡, and Muhammad Ali Imran*

*James Watt School of Engineering, University of Glasgow, Glasgow, UK
†School of Computer Science and Engineering, Nanyang Technological University, Singapore
‡School of Automation, Guangdong University of Technology, Guangdong, China
Email: Yao.Sun@glasgow.ac.uk

Abstract—Semantic communication (SemCom) has recently been considered a promising solution for the inevitable crisis of scarce communication resources. This trend stimulates us to explore the potential of applying SemCom to vehicular networks, which normally consume a tremendous amount of resources to achieve stringent requirements on high reliability and low latency. Unfortunately, the unique background knowledge matching mechanism in SemCom makes it challenging to realize efficient vehicle-to-vehicle service provisioning for multiple users at the same time. To this end, this paper identifies and jointly addresses two fundamental problems of knowledge base construction (KBC) and vehicle service pairing (VSP) inherently existing in SemCom-enabled vehicular networks. Concretely, we first derive the knowledge matching based queuing latency specific for semantic data packets, and then formulate a latency-minimization problem subject to several KBC and VSP related reliability constraints. Afterward, a SemCom-empowered Service Supplying Solution (S^4) is proposed along with the theoretical analysis of its optimality guarantee. Simulation results demonstrate the superiority of S^4 in terms of average queuing latency, semantic data packet throughput, and user knowledge preference satisfaction compared with two different benchmarks.

I. INTRODUCTION

Semantic communication (SemCom) beyond the conventional Shannon paradigm has recently been recognized as a promising remedy for communication resource saving and transmission reliability promotion [1]–[6]. Such a trend, therefore, inspires us to investigate the potential of exploiting SemCom in vehicular networks to deal with the inevitable scarcity of available communication resources while enhancing information interaction efficiency, especially when considering large-scale vehicle-to-vehicle (V2V) communications.

Different from bit-oriented traditional communication, the accurate delivery of semantics becomes the cornerstone of SemCom [1]. Taking a single SemCom-enabled V2V link as an example, a sender vehicle user (VUE) first leverages background knowledge relevant to source messages to filter out irrelevant content and extract core semantic features that only require fewer bits for transmission [2]. Once the receiver VUE has the same knowledge as the sender, its local semantic interpreters are capable of accurately restoring the original meanings from the received bits, even with intolerable bit errors in data dissemination [3]. Consequently, SemCom can significantly alleviate the resource scarcity problem, while ensuring sufficient transmission efficiency along with ultra-

low semantic errors for each V2V transmission. Apart from these advancements, it should be noted that the background knowledge equivalence is of paramount importance to eliminate semantic ambiguity, which has led to a key concept of *knowledge base* (KB) in the realm of SemCom [4]–[6]. Specifically, a single KB is deemed a small information entity that contains background knowledge corresponding to only one particular application domain (e.g., music or sports, etc.) [4]. Hence, holding some common KBs becomes the necessary condition to perform SemCom between two VUEs in accordance with the knowledge equivalence principle.

Recently, there have been some technical works on SemCom from the link level [3] to the networking level [5] under cellular network architectures. Nevertheless, to the best of our knowledge, no research pertinent to service provisioning has been conducted in *SemCom-enabled vehicular networks* (SCVNs), which should be rather challenging and involves two unique yet fundamental problems. To be concrete, due to the varying practical KB sizes, personal KB preferences, and limited vehicular storage capacities, the first problem is how to devise an optimal *knowledge base construction* (KBC) policy not only proactively but also collaboratively for all VUEs to construct their respective appropriate KBs for better service provisioning. In the meantime, when considering different types of KBs equipped on numerous VUEs and unstable wireless link quality, it can be challenging to well solve the service provisioning-driven VUE pairing problem to meet the knowledge matching restriction, thus shaping the second problem namely *vehicle service pairing* (VSP).

To that end, in this paper, we propose a novel SemCom-empowered Service Supplying Solution (S^4) in SCVN. First, we theoretically derive the KB matching based queuing latency for semantic data packets received at each possible VUE pair. Afterward, a joint latency-minimization problem is mathematically formulated subject to several KBC and VSP-related reliability constraints and other practical system limitations. To address this problem with low complexity, we then develop an efficient solution S^4 and theoretically prove its optimality. Particularly, a primal-dual problem transformation method is exploited in S^4, followed by a two-stage method dedicated to solving multiple subproblems. Given the dual variable in each iteration, the first stage is to obtain the optimal KBC sub-policy for each potential VUE pair, whereby the second stage
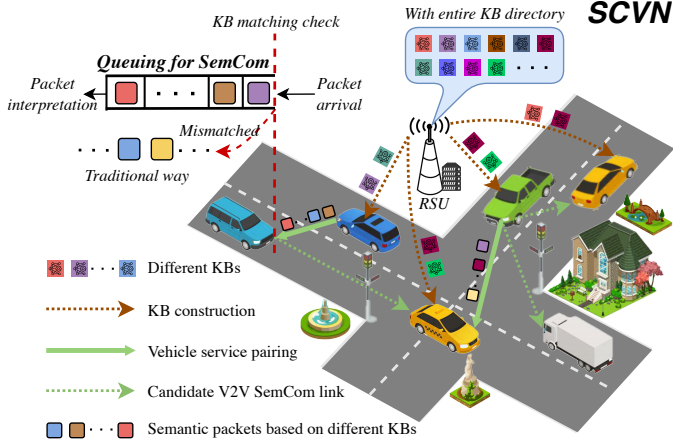
Fig. 1. The SCVN scenario and the knowledge base aware queuing model for semantic data packets transmitted between VUEs.

is to finalize the optimal solutions of KBC and VSP for all VUEs. Finally, numerical results demonstrate the performance superiority of S^4 in terms of average queuing latency, semantic data packet throughput, and user knowledge preference satisfaction compared with two different benchmarks.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Considering an SCVN scenario as shown in Fig. 1, the total of V VUEs are randomly distributed within the coverage of a single roadside unit (RSU), and each VUE $i \in \mathcal{V} = \{1, 2, \dots, V\}$ is capable of providing SemCom-empowered services to others. Let $\gamma_{i,j}$ denote the signal-to-interference-plus-noise ratio (SINR) experienced by the V2V link between VUE i and VUE j. With this, the set of communication neighbors of VUE i is defined as $\mathcal{V}_i = \{j | j \in \mathcal{V}, j \neq i, \gamma_{i,j} \geqslant \gamma_0\}$, $i \in \mathcal{V}$, where γ_0 is a prescribed SINR threshold.

A. Semantic Service Provisioning Model

Due to the unique mechanism of semantic interpretation, the acquisition of necessary background knowledge is inevitable for all SemCom-enabled transceivers. In this work, assuming that each VUE i proactively downloads and constructs its required KBs from the RSU under its finite KB storage capacity C_i. Meanwhile, suppose that there is a KB library $\mathcal{K}$ with a total of N differing KBs in the considered SCVN, and each requires a unique storage size s_n, $n \in \mathcal{K} = \{1, 2, \dots, N\}$. Furthermore, we define a binary KBC indicator as

$$\alpha_i^n = \begin{cases} 1, & \text{if KB } n \text{ is constructed at VUE } i; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

It is worth mentioning that the same KB cannot be constructed repeatedly at one VUE to promote the storage efficiency.

Besides, note that different VUEs may have different preferences for these KBs corresponding to their desired services. Without loss of generality, we assume that the KB popularity at each VUE meets Zipf distribution [7].[1] Hence,

[1]Other distributions with known probabilities can also be adopted without changing the remaining modeling and solution.

the probability of VUE i requesting its desired KB n-based services (generating the corresponding semantic data packets) is $p_i^n = (r_i^n)^{-\xi_i} / \sum_{e \in \mathcal{K}} e^{-\xi_i}, \forall (i, n) \in \mathcal{V} \times \mathcal{K}$, where ξ_i ($\xi_i \geqslant 0$) is the skewness of the Zipf distribution, and r_i^n is the popularity rank of KB n at VUE i.[2] Based on p_i^n, we specially develop a KBC-related metric η_i below, namely *knowledge preference satisfaction*, to measure the satisfaction degree of VUE i constructing its interested KBs as

$$\eta_i = \sum_{n \in \mathcal{K}} \alpha_i^n p_i^n. \quad (2)$$

It is further required that $\eta_i \geqslant \eta_0$, where η_0 is the unified minimum threshold that needs to be achieved at each VUE.

Moreover, considering that each VUE can be paired with only one (another) VUE at a time. Let $\beta_{i \looparrowright j}$ denote the binary VSP indicator for a VUE i-VUE j pair (suppose that VUE i is the sender and VUE j is the receiver), where

$$\beta_{i \looparrowright j} = \begin{cases} 1, & \text{if VUE } i \text{ is associated with VUE } j; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Here, we use the notation $\looparrowright$ as an auxiliary illustration to specify the roles of sender and receiver in each VUE pair.

B. Knowledge Base Aware Queuing Model

As depicted in Fig. 1, the knowledge matching based semantic packet queuing delay is employed as the latency metric of SemCom, to characterize the average sojourn time of semantic data packets in the receiver VUE's queue buffer (following a first-come first-serve rule). Note that not all semantic packets arriving at the receiver VUE are always allowed to enter its queue, as some of them may mismatch the KBs currently held, rendering these packets uninterpretable. To avoid pointless queuing, these mismatched packets may have to choose the traditional communication way for information interaction, which will not be discussed in depth here.

To preserve generality, we first suppose a Poisson data arrival process with average rate $\lambda_i^n = \lambda_i p_i^n$ for a sender VUE i to account for its local semantic packet generation based on KB n, where λ_i is the total arrival rate of all semantic packets at VUE i. Then we can obtain the effective packet arrival rate from sender VUE i to receiver VUE j as $\lambda_{i \looparrowright j}^{eff} = \sum_{n \in \mathcal{K}} \alpha_i^n \alpha_j^n \lambda_i^n$ by checking the KB matching status for all received semantic packets. Herein, denoting the ratio of the number of mismatched packets to the number of received packets in total for VUE i-VUE j pair as $\theta_{i \looparrowright j}$, namely *knowledge mismatch degree*, which is explicitly calculated by

$$\theta_{i \looparrowright j} = \frac{\sum_{n \in \mathcal{K}} \alpha_i^n (1 - \alpha_j^n) \lambda_i^n}{\sum_{n \in \mathcal{K}} \alpha_i^n \lambda_i^n}. \quad (4)$$

In parallel, let a random variable I_j^n denote the Markovian interpretation time [8] required by KB n-based packets at VUE j with mean $1/\mu_j^n$, which is determined by the computing capability of the vehicle and the type of the desired KB.

[2]The KB popularity rank of each VUE can be estimated based on its historical messaging records, which will not be discussed in this paper.

However, since multiple packets based on different KBs are allowed to queue at the same time, the interpretation time distribution for a receiver VUE should be treated as a general distribution. If further taking into account the KB popularity, we can calculate the ratio of the amount of KB n-based packets to the total packets in the VUE i-VUE j pair's queue by $\epsilon_{i \leftrightarrow j}^n = p_i^n / \sum_{f \in \mathcal{K}} \alpha_i^f \alpha_j^f p_i^f$. Keeping in mind the independence among packets based on different KBs, the interpretation time required by each packet in the queue is now phrased as $W_{i \leftrightarrow j} = \sum_{n \in \mathcal{K}} \alpha_i^n \alpha_j^n \epsilon_{i \leftrightarrow j}^n I_j^n$.

In full view of the above, the queue of each VUE pair in the SCVN fully qualifies as an M/G/1 model. According to the *Pollaczek-Khintchine formula* [9], the average queuing latency for VUE i-VUE j pair (denoted as $\delta_{i \leftrightarrow j}$) should be[3]

$$\delta_{i \leftrightarrow j} = \frac{\lambda_{i \leftrightarrow j}^{eff} \cdot \left(\mathbb{E}^2 \left[W_{i \leftrightarrow j} \right] + Var \left(W_{i \leftrightarrow j} \right) \right)}{2 \left(1 - \lambda_{i \leftrightarrow j}^{eff} \cdot \mathbb{E} \left[W_{i \leftrightarrow j} \right] \right)}. \tag{5}$$

On this basis, again leveraging the independence of I_j^n over n, we can obtain the expectation of the interpretation time for all semantic data packets in the queue by

$$\mathbb{E} \left[W_{i \leftrightarrow j} \right] = \sum_{n \in \mathcal{K}} \frac{\alpha_i^n \alpha_j^n \epsilon_{i \leftrightarrow j}^n}{\mu_j^n}, \tag{6}$$

and the variance of $W_{i \leftrightarrow j}$ is given by

$$Var \left(W_{i \leftrightarrow j} \right) = \sum_{n \in \mathcal{K}} \alpha_i^n \alpha_j^n \left(\frac{\epsilon_{i \leftrightarrow j}^n}{\mu_j^n} \right)^2. \tag{7}$$

Based on (6) and (7), $\delta_{i \leftrightarrow j}$ in (5) can be calculated explicitly.

C. Problem Formulation

According to the above discussions, we identify and formulate below a joint queuing latency-minimization problem under the KBC indicator set $\boldsymbol{\alpha} = \{\alpha_i^n | i \in \mathcal{V}, n \in \mathcal{K}\}$ and the VSP indicator set $\boldsymbol{\beta} = \{\beta_{i \leftrightarrow j} | i \in \mathcal{V}, j \in \mathcal{V}_i\}$.

$$\mathbf{P0}: \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \beta_{i \leftrightarrow j} \delta_{i \leftrightarrow j} \tag{8}$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{K}} \alpha_i^n \cdot s_n \leqslant C_i, \ \forall i \in \mathcal{V}, \tag{8a}$$

$$\eta_i \geqslant \eta_0, \ \forall i \in \mathcal{V}, \tag{8b}$$

$$\sum_{j \in \mathcal{V}_i} \beta_{i \leftrightarrow j} = 1, \ \forall i \in \mathcal{V}, \tag{8c}$$

$$\beta_{i \leftrightarrow j} = \beta_{j \leftrightarrow i}, \ \forall (i, j) \in \mathcal{V} \times \mathcal{V}_i, \tag{8d}$$

$$\sum_{j \in \mathcal{V}_i} \beta_{i \leftrightarrow j} \theta_{i \leftrightarrow j} \leqslant \theta_0, \ \forall i \in \mathcal{V}, \tag{8e}$$

$$\alpha_i^n \in \{0, 1\}, \ \forall (i, n) \in \mathcal{V} \times \mathcal{K}, \tag{8f}$$

$$\beta_{i \leftrightarrow j} \in \{0, 1\}, \ \forall (i, j) \in \mathcal{V} \times \mathcal{V}_i. \tag{8g}$$

Constraint (8a) indicates the storage limitation of VUEs, while constraint (8b) corresponds to the aforementioned knowledge

---

[3]In order to guarantee the steady-state of the queuing system, a condition of $\lambda_{i \leftrightarrow j}^{eff} \mathbb{E}[W_{i \leftrightarrow j}] < 1$ must be reached before proceeding. In this work, we assume that the packet interpretation rate is larger than the packet arrival rate to make the queuing latency finite and thus solvable.

preference satisfaction requirement for each VUE. Constraints (8c) and (8d) mathematically model the single-association requirement of VUEs. Constraint (8e) represents that the knowledge mismatch degree of each VUE pair should be over the threshold θ_0. Constraints (8f) and (8g) characterize the binary properties of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. As such, the main difficulty in solving **P0** lies on its combinatorial nature and non-convexity of its highly complicated objective function.

III. PROPOSED SEMCOM-EMPOWERED SERVICE SUPPLYING SOLUTION

In this section, we illustrate how to design our proposed solution S[4] to efficiently cope with **P0** for SCVNs. First, a Lagrange dual method is leveraged to eliminate the cross-term constraints in **P0** with a corresponding dual optimization problem transformed (referring to **D0** in Section III.A). Then given the dual variable in each iteration, we dedicatedly develop a two-stage method to determine $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the dual problem, where the optimality will be theoretically proved in Section III.B. Specifically, in the first stage, we subtly construct multiple subproblems (referring to **P1_{i,j}**), each of which aims to independently seek the optimal KBC sub-policy (with respect to only α_i^n and α_j^n) for each individual VUE pair. After solving all these subproblems, an optimal coefficient matrix is obtained for all potential VUE pairs, by which we further construct a new subproblem (referring to **P2**) in the second stage to find the optimal VSP strategy for $\boldsymbol{\beta}$.

A. Primal-Dual Problem Transformation

We first incorporate constraint (8e) into the objective function (8) by associating a Lagrange multiplier $\boldsymbol{\tau} = \{\tau_i | \tau_i \geqslant 0, i \in \mathcal{V}\}$, so that we have its associated Lagrange function as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \beta_{i \leftrightarrow j} \left(\delta_{i \leftrightarrow j} + \tau_i \theta_{i \leftrightarrow j} \right) - \theta_0 \sum_{i \in \mathcal{V}} \tau_i$$

$$\triangleq \widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \theta_0 \sum_{i \in \mathcal{V}} \tau_i, \tag{9}$$

where $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is defined for the sake of expression brevity. Then the Lagrange dual problem of **P0** is formulated by

$$\mathbf{D0}: \max_{\boldsymbol{\tau}} \quad D(\boldsymbol{\tau}) = g_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\tau}) - \theta_0 \sum_{i \in \mathcal{V}} \tau_i, \tag{10}$$

where we have

$$g_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\tau}) = \inf_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

$$\text{s.t.} \quad (8a) - (8d), (8f), (8g). \tag{11}$$

Notably, the optimality of the convex problem **D0** gives at least the best lower bound of **P0**, even if **P0** is nonconvex, according to the duality property [10].

Given the initial dual variable $\boldsymbol{\tau}$, we can solve problem (11) in the first place to find the optimal solution $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the details of which will be presented later. After that, a subgradient method is employed in charge of updating $\boldsymbol{\tau}$ to solve **D0** in an iterative fashion, and generally, the convergence of subgradient descent can be ensured with the proper stepsize [11].

B. Two-Stage Method Based on KBC and VSP

In this work, we develop a two-stage method to obtain the exactly optimal $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with a low computational complexity. In the first stage, we focus on multiple independent KB construction subproblems, each corresponding to a potential VUE pair in the SCVN. Specially, here the performances of the VUE i-VUE j pair (i.e., the sender VUE i and the receiver VUE j) and the VUE j-VUE i pair (i.e., the sender VUE j and the receiver VUE i) need to be considered together, and for ease of distinction, we refer to the two as a *VUE i, j pair*, $\forall (i,j) \in \mathcal{V} \times \mathcal{V}_i, j > i$. That way, different KBC subproblems can be solved independently, so we let

$$\omega_{i,j} = (\delta_{i \leftrightarrow j} + \tau_i \theta_{i \leftrightarrow j}) + (\delta_{j \leftrightarrow i} + \tau_j \theta_{j \leftrightarrow i}). \quad (12)$$

In this context, we now construct $U = \left(\sum_{i \in \mathcal{V}} |\mathcal{V}_i| \right)/2$ subproblems, each denoted as $\mathbf{P1_{i,j}}$ to seek the optimal KBC sub-policy only for an individual VUE i, j pair. Herein, it is worth pointing out that the optimal KBC solution to problem (11) cannot be achieved by simply combining the obtained sub-policies of these $\mathbf{P1_{i,j}}$, but these sub-policies will be used to construct the subsequent VSP subproblem to finalize the joint optimal solution of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for (11). Given the dual variable $\boldsymbol{\tau}$ in each iteration, $\mathbf{P1_{i,j}}$ becomes

$$\mathbf{P1_{i,j}}: \min_{\{\alpha_i^n\}, \{\alpha_j^n\}} \quad \omega_{i,j} \quad (13)$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{K}} \alpha_i^n \cdot s_n \leqslant C_i, \quad (13a)$$

$$\sum_{n \in \mathcal{K}} \alpha_j^n \cdot s_n \leqslant C_j, \quad (13b)$$

$$\eta_i \geqslant \eta_0, \ \eta_j \geqslant \eta_0, \quad (13c)$$

$$\alpha_i^n \in \{0,1\}, \alpha_j^n \in \{0,1\}. \quad (13d)$$

Since the problem scale of $\mathbf{P1_{i,j}}$ is quite small, here we directly employ a metaheuristic algorithm based on tabu search to quickly determine its solution [12]. Afterward, we can obtain the optimal KBC sub-policies for VUE i (denoted as $\boldsymbol{\alpha}_{i_{(j)}}^*$) and VUE j (denoted as $\boldsymbol{\alpha}_{j_{(i)}}^*$),[4] corresponding to the individual VUE i, j pair. The following proposition explicitly shows how the sub-policy of $\mathbf{P1_{i,j}}$ correlates to problem (11).

Proposition 1: Let $\boldsymbol{\alpha}^* = [\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \cdots, \boldsymbol{\alpha}_V^*]^T$ be the optimal KBC solution to the problem in (11) given the dual variable $\boldsymbol{\tau}$, where $\boldsymbol{\alpha}_i^*$ represents the optimal KBC policy of VUE i. Then we have $\forall i \in \mathcal{V}, \exists j \in \mathcal{V}_i$, such that $\boldsymbol{\alpha}_{i_{(j)}}^* = \boldsymbol{\alpha}_i^*$.

Proof: Given the optimal KBC solution $\boldsymbol{\alpha}^*$, let $\boldsymbol{\beta}^* = \left[\beta_{1 \leftrightarrow j_1^*}, \beta_{2 \leftrightarrow j_2^*}, \cdots, \beta_{V \leftrightarrow j_V^*}\right]^T$ be the corresponding optimal VSP solution to the problem in (11) under the same dual variable $\boldsymbol{\tau}$, where $\beta_{i \leftrightarrow j_i^*}$ ($\forall i \in \mathcal{V}$) indicates that VUE j_i^* is the optimal SemCom node for VUE i, i.e., $\beta_{i \leftrightarrow j_i^*} = 1$.

From $\omega_{i,j}$ defined in (12), the objective function $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (11) can be rewritten as $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i, j > i} \beta_{i \leftrightarrow j} \omega_{i,j}$, and then substituting $\boldsymbol{\beta}^*$ into

---

[4]For auxiliary illustration, we use $(\cdot)$ in the subscript to specify the VUE pair attribute (relation) for each VUE's KBC sub-policy obtained from $\mathbf{P1_{i,j}}$.

$\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can yield $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^*) = \sum_{i \in \mathcal{V}, i < j_i^*} \omega_{i, j_i^*}$, where ω_{i, j_i^*} is the term only related to VUE i, j_i^* pair.

Undoubtedly, if $\boldsymbol{\alpha}^*$ is further substituted into $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^*)$, we can straightforwardly reach the optimality of the problem in (11). Since different VUE i, j_i^* pairs are independent of each other, it means that different terms related to ω_{i, j_i^*} are independent of each other as well in $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^*)$. Therefore, achieving the optimality of $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^*)$ is equivalent to achieving the optimality of each ω_{i, j_i^*} under the given $\boldsymbol{\beta}^*$, where the optimality can be reached when $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$.

In view of the above, we know that $\boldsymbol{\alpha}_i^*$ must be the optimal solution of $\omega_{i, j_i^*}, \forall i \in \mathcal{V}$. Further combined with another fact that $\omega_{i,j}$ is the objective of $\mathbf{P1_{i,j}}, \forall (i,j) \in \mathcal{V} \times \mathcal{V}_i, j > i$ where $\boldsymbol{\alpha}_{i_{(j)}}^*$ is the corresponding optimal solution, we can be sure that the equality $\boldsymbol{\alpha}_{i_{(j)}}^* = \boldsymbol{\alpha}_i^*$ holds when $j = j_i^*$. $\blacksquare$

From Proposition 1, it is observed that the optimal KBC policy of each VUE can be found by solving a certain $\mathbf{P1_{i,j}}$. Hence, considering the single-association requirement, the optimal VSP strategy becomes the only key to finalize the optimal solution to (11). To achieve this, we first obtain the optimal coefficient matrix for $\boldsymbol{\beta}$ in (11) to account for all VSP possibilities. By calculating optimum $\omega_{i,j}$ (denoted as $\omega_{i,j}^*$) in $\mathbf{P1_{i,j}}$, the optimal coefficient matrix is formed as

$$\boldsymbol{\Omega} = \begin{bmatrix} +\infty & \omega_{1,2}^* & \omega_{1,3}^* & \cdots & \omega_{1,V}^* \\ \omega_{2,1}^* & +\infty & \omega_{2,3}^* & \cdots & \omega_{2,V}^* \\ \omega_{3,1}^* & \omega_{3,2}^* & +\infty & \cdots & \omega_{3,V}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_{V,1}^* & \omega_{V,2}^* & \omega_{V,3}^* & \cdots & +\infty \end{bmatrix}. \quad (14)$$

$\boldsymbol{\Omega}$ is a $V \times V$ symmetric matrix where $\omega_{i,j}^* = \omega_{j,i}^*$, and all elements on its main diagonal are set to $+\infty$ to indicate the fact that a vehicle cannot communicate with itself, i.e., $j \neq i$. Besides, note that some $\omega_{i,j}^*$s in $\boldsymbol{\Omega}$ also have a value $+\infty$ if VUE j is not the direct neighbor of VUE i, i.e., $j \notin \mathcal{V}_i$.

Next, we concentrate upon the optimal vehicle service pairing strategy by constructing a new subproblem in the second stage. In line with the objective $\widetilde{L}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\boldsymbol{\beta}$-related constraints in (11), the VSP subproblem is written as

$$\mathbf{P2}: \min_{\boldsymbol{\beta}} \quad \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \beta_{i \leftrightarrow j} \omega_{i,j}^* \quad (15)$$

$$\text{s.t.} \quad (8c), (8d), (8g). \quad (15a)$$

Given the dual variable, the optimal VSP strategy $\boldsymbol{\beta}$ (denoted as $\boldsymbol{\beta}^* = \left[\beta_{1 \leftrightarrow j_1^*}, \beta_{2 \leftrightarrow j_2^*}, \cdots, \beta_{V \leftrightarrow j_V^*}\right]^T$) can be directly finalized by solving $\mathbf{P2}$, where $\beta_{i \leftrightarrow j_i^*}$ ($\forall i \in \mathcal{V}$) indicates that VUE j_i^* is the optimal SemCom node for VUE i, i.e., $\beta_{i \leftrightarrow j_i^*} = 1$. To solve this problem, we first relax $\boldsymbol{\beta}$ into a continuous variable between 0 and 1 to make $\mathbf{P2}$ a linear problem, which can be efficiently solved with toolboxes such as CVXPY [13]. Afterward, we restore the binary nature of the obtained continuous solution (denoted as $\boldsymbol{\beta}^R$) by

$$\beta_{i' \leftrightarrow j'}^* = \beta_{j' \leftrightarrow i'}^* = 1, \quad (16)$$

if and only if

$$\left(i^{'}, j^{'}\right) = \arg \max_{i \in \mathcal{V}, j \in \mathcal{V}_i, j > i} \beta^R_{i \leftrightarrow j}. \qquad (17)$$

Meanwhile, for the remaining $\beta^*_{i \leftrightarrow j}$ with respect to VUE $i^{'}$ and VUE $j^{'}$, we naturally have

$$\begin{cases} \beta^*_{i^{'} \leftrightarrow j} = \beta^*_{j \leftrightarrow i^{'}} = 0, & \forall j \in \mathcal{V}_{i^{'}}, j \neq j^{'} \\ \beta^*_{j^{'} \leftrightarrow i} = \beta^*_{i \leftrightarrow j^{'}} = 0, & \forall i \in \mathcal{V}_{j^{'}}, i \neq i^{'} \end{cases}. \qquad (18)$$

Then we let $\mathcal{V} = \mathcal{V} \backslash \{i, j\}$, and repeat the above progresses until determining the optimal VSP solution for all VUEs. Finally, the obtained $\boldsymbol{\beta}^*$ is fed back to $\boldsymbol{\Omega}$ in order to further finalize the optimal KBC policy $\boldsymbol{\alpha}^*$. In the context of the solution to $\mathbf{P1_{i,j}}$, the approach to finalize $\boldsymbol{\alpha}^*$ can be stated more precisely as: for any $i \in \mathcal{V}$, we directly have $\boldsymbol{\alpha}^*_i = \boldsymbol{\alpha}^*_{i_{(j^*_i)}}$.

IV. NUMERICAL RESULTS AND DISCUSSIONS

In this section, numerical simulations are conducted to evaluate the performance of the proposed solution S^4 in SCVNs. We model a six-lane freeway (lane width: 4 m) passing through a single cell with the RSU at its center (cell radius: 500 m), where a total of 60 VUEs are dropped according to the spatial Poisson process [14]. As for the settings relevant to SemCom, a total of 12 different KBs are preset to provide VUEs with a variety of distinct services, and each of them has a storage size randomly distributed from 1 to 5 units. Correspondingly, we set a uniform KB storage capacity of 24 units for all VUEs. Besides, each VUE's preference ranking for all KBs (i.e., r^n_i) is generated independently and randomly, where assuming that their respective Zipf distributions have the same skewness 1.0. Likewise, either the average arrival rate of total semantic data packets, or the average interpretation time for packets based on the same KB n, is considered to be the same for all VUEs. Here, we fix the average total arrival rate λ_i at 100 packets/s and randomly generate the value of $1/\mu^n_i$ in a range of $5 \times 10^{-3} \sim 1 \times 10^{-2}$ s/packet with respect to different KB n. Further, the minimum knowledge preference satisfaction threshold η_0 and the maximum knowledge mismatch degree threshold θ_0 are prescribed as 0.5 and 0.1, respectively.

For comparison purposes, two different benchmarks are used in the simulations: 1) Distance-first pairing (DFP) strategy which assumes each VUE to choose its nearest unpaired VUE for V2V pairing; 2) Knowledge-first pairing (KFP) in which each VUE selects its neighboring unpaired VUE with the highest KB matching degree for V2V pairing. Among them, a personal preference-first KBC policy is considered for both benchmarks, which allows each VUE to construct KBs with the highest preferences until η_0 is satisfied, and then randomly select these unconstructed KBs until reaching respective maximum capacity.

Fig. 2 first depicts the average queuing latency performance of a VUE pair against varying numbers of VUEs, where two different KB preference skewness $\xi = 0.8$ and $\xi = 1.4$ are considered. In this figure, the latency of S^4 declines at the beginning with the number of VUEs, then remains stable beyond 70 VUEs, and it can always outperform both benchmarks
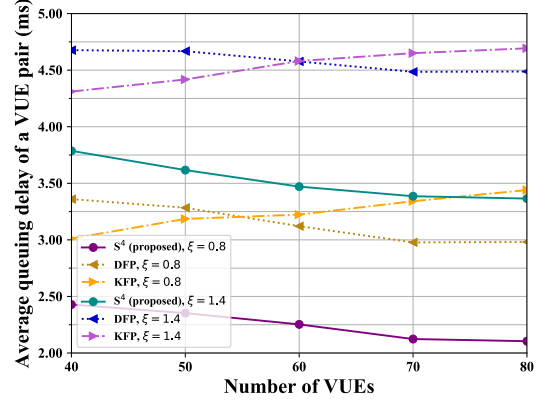

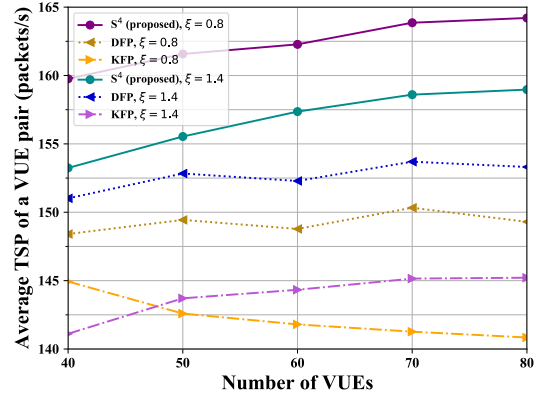Fig. 2. Average queuing latency of a VUE pair vs. varying numbers of VUEs.

Fig. 3. Average TSP of a VUE pair vs. varying numbers of VUEs.

with an average latency reduction of around 1 ms at any ξ. The rationale behind this trend is that the more neighbors each VUE can have, the better chance of getting the low queuing latency for each VUE pair, which will be eventually stabilized when reaching the respective best achievable latency with a fixed bandwidth budget. Moreover, it is also observed that a larger ξ causes a higher latency penalty, since the vast majority of VUEs' KBC is concentrated on a small number of KBs when ξ increases. Clearly, a larger ξ will make each participant more difficult to find a highly pairing VUE with the low latency under the given knowledge mismatching requirement θ_0, thus resulting in a degraded performance.

The above analysis also applies to Fig. 3, which compares all the three methods under the same settings as Fig. 2 to demonstrate the performance of the average throughput in semantic packets (TSP). Specifically, the TSP represents the total number of semantic packets that can be interpreted at a VUE pair per second, whose value is determined based on $\mathbb{E}[W_{i \leftrightarrow j}]$ in (6). Likewise, the higher TSP is obtained as the number of VUEs increases, and our S^4 is still far better than the benchmarks at any point, e.g., with an average performance gain of 14 packets/s compared with DFP and 20 packets/s with KFP at $\xi = 0.8$. Again, we see a better TSP when the KB popularity is diluted by a smaller ξ.

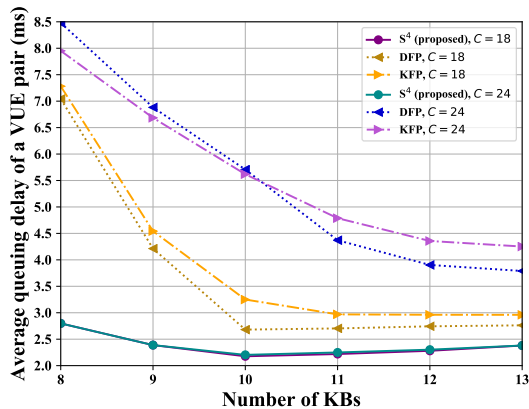Next, we explore the impact of varying number of KBs on

Fig. 4. Average queuing latency of a VUE pair vs. varying numbers of KBs.
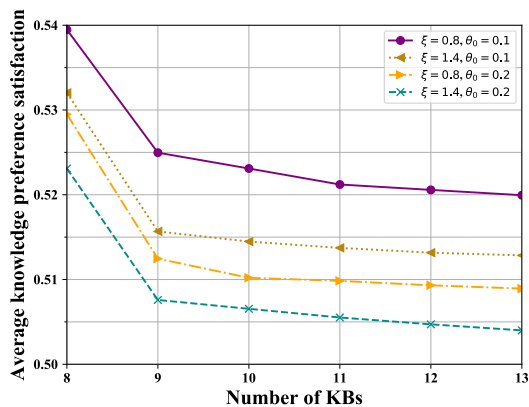

Fig. 5. Average knowledge preference satisfaction vs. varying numbers of KBs.

the average queuing latency of a VUE pair with different VUE capacities $C = 18$ and $C = 24$, as demonstrated in Fig. 4. It can be found that the latency drops fast at the beginning, and then rises slightly after exceeding 10 KBs, whereas the performance of our S^4 still surpasses the benchmarks. This trend is attributed to the fact that more KBs imply less discrepancy in VUEs' preferences for different KBs given the fixed ξ, thereby at first leading to the higher probability for two paired VUEs constructing the KBs with high interpretation rates so as to render a lower delay. However, such performance gains will saturate and even worsen when these KBs with low interpretation rates become inevitably dominant in order to meet the minimum knowledge preference satisfaction threshold η_0. Besides, it is seen that different VUE capacities have little effect on the latency of S^4, although the larger capacity makes more KBs capable of construction. This is due to the latency-minimization objective we particularly focus on in the delay-sensitive SCVN, and only the KBs with low interpretation time should be selected.

Finally, we validate the average knowledge preference satisfaction $\bar{\eta} = \frac{1}{V} \sum_{i \in \mathcal{V}} \eta_i$ reached at each VUE with varying numbers of KBs as shown in Fig. 5, where $\xi = 0.8$, $\xi = 1.4$, $\theta_0 = 0.1$, and $\theta_0 = 0.2$ are taken into account. As the number of KBs increases, a lower $\bar{\eta}$ is obtained, which is to prevent these unnecessary KBs from being constructed while satisfying η_0 to the greatest extent. For the two curves with different ξ, referring to the analysis of Fig. 2, a higher ξ indicates a more concentrated KB preference, which means some extra KBs need to be constructed to meet the maximum θ_0 requirement. Because of this, we also see a lower $\bar{\eta}$ at a higher θ_0, since a more tolerable knowledge mismatch degree is more likely to avoid the unnecessary KBC.

V. CONCLUSIONS

In this paper, we proposed a novel solution S^4 to address the SemCom-empowered service provisioning problem in the SCVN. First, the KB matching based queuing latency expression of semantic data packets was derived, and then we identified and formulated the fundamental problem of KBC and VSP to minimize the queuing latency for all VUE pairs. After the primal-dual problem transformation, a two-stage method was developed specifically to solve multiple subproblems related to KBC and VSP with low computational complexity, and the solution optimality has been theoretically proved. Numerical results verified the sufficient performance superiority of S^4 in terms of both latency and reliability by comparing it with two different benchmarks. Accordingly, this work can be served as a preliminary research of applying SemCom to next-generation vehicular networks.

REFERENCES

[1] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC: A Review of General Semantics*, pp. 261–281, 1953.
[2] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*. IEEE, 2011, pp. 110–117.
[3] L. Xia, Y. Sun, C. Liang, D. Feng, R. Cheng, Y. Yang, and M. A. Imran, "WiserVR: Semantic communication enabled wireless virtual reality delivery," *IEEE Wireless Communications*, 2022.
[4] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.
[5] L. Xia, Y. Sun, D. Niyato, X. Li, and M. A. Imran, "Wireless semantic communication: A networking perspective," *arXiv preprint arXiv:2212.14142*, 2022.
[6] L. Xia, Y. Sun, X. Li, G. Feng, and M. A. Imran, "Wireless resource management in intelligent semantic communication networks," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022, pp. 1–6.
[7] S. T. Piantadosi, "Zipfs word frequency law in natural language: A critical review and future directions," *Psychonomic Bulletin & Review*, vol. 21, no. 5, pp. 1112–1130, 2014.
[8] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 39, no. 5, pp. 489–504, 2009.
[9] S. M. Ross, *Introduction to Probability Models*. Academic Press, 2014.
[10] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[11] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.
[12] F. Glover, "Tabu search-part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
[13] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2909–2913, 2016.
[14] *3GPP: Study on LTE-Based V2X Services; (Release 14)*, document TR 36.885, V2.0.0, 3GPP, Jun. 2016.