Two-Timescale Adaptive Live Video Streaming Transmission Mechanism for Vehicular Networks

Biqian Feng, Chenyuan Feng, Geyong Min, and Tony Q. S. Quek

***Abstract*—This research explores a novel adaptive live video streaming transmission strategy over vehicular networks to solve the conundrum between resource-constrained environment and user demand on high Quality-of-Experience (QoE). With an exquisite design of resource types and channel variations, we propose a two-timescale transmission mechanism which allocates the bitrate and bandwidth for each large-timescale frame based on the statistical knowledge of the channel state information (CSI) and refines the power allocation for each small-timescale slot based on instantaneous CSI. Subsequently, we formulate a QoE maximization problem under the restrictions of finite bitrates, bandwidth and power budget. To solve this problem with low computation complexity, we propose a two-stage online successive convex approximation (TOSCA)-based resource allocation algorithm. Simulation results illustrate the rationality and necessity of the proposed dual-time scale optimization, and the proposed mechanism noticeably outperforms the conventional resource allocation schemes under different power budgets and lane configurations.**

***Index Terms*—Live video transmission, vehicular networks, resource allocation, two-timescale optimization**

I. Introduction

Given the rapid rise of on-vehicle video services such as video conferencing and video surveillance, the contradiction between the highly time-varying vehicular communication environments and users' demands for a clear and seamless viewing experience has intensified [1]. Traditional video streaming distribution systems complete content acquisition, caching and transcoding in the cloud. Although this can alleviate the pressure on cloud storage, there are still serious issues such as large transmission latency, high bandwidth consumption and redundant transmission in the network edge [2]. With the advancement of vehicular edge computing (VEC) technology, the service delay and the communication overhead in the coer network can be significantly reduced by using the road side units (RSUs) to download the original bitrate version of the video from the cloud and dynamically transcoding it to the bitrate version suitable for different user channels [3].

To enhance the Quality-of-Experience (QoE) to be perceived by users in the VEC networks, some studies focus on the joint optimization of video transcoding selection and transmission resource allocation. Such an optimization problem can be formulated as a mixed-integer nonlinear programming problem with computational and wireless bandwidth resource constraints and solved with a low-complexity online scheme [4]. A joint allocation mechanism of computing resources and communication resources was also proposed in VEC-assisted video streaming delivery network [5]. However, these studies only consider the pairwise joint optimization between communication resource allocation and video bitrate version selection, and do not fully consider the dynamics of the network and vehicular users. Considering the short coherence time caused by user mobility, it is not only impractical but also uneconomical to instantly reassign all kinds of resources.

B. Feng is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China (e-mail: fengbiqian@sjtu.edu.cn).

C. Feng (*Corresponding author*) is with EURECOM, Biot 06410, France (e-mail: Chenyuan.Feng@eurecom.fr).

G. Min is with the Department of Computer Science, University of Exeter, Exeter, EX4 4QF, U.K. (e-mail: g.min@exeter.ac.uk).

T. Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

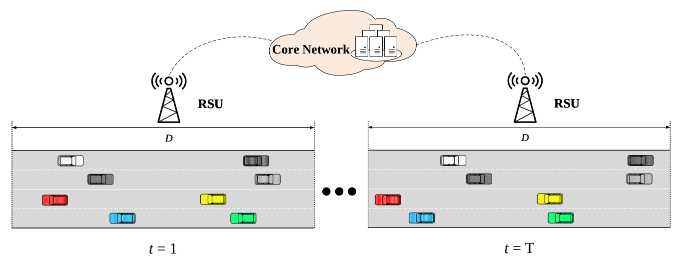

Fig. 1. A typical vehicular communication network supporting high-data-rate video live streaming services.

Inspired by this problem, we propose an adaptive video streaming transmission strategy that adapts power, bandwidth, and video bitrate versions to the communication environment at different time scales. Our contribution can be summarized as follows: i) To improve the transmission efficiency with mobile vehicular users, we formulate a two-timescale transmission decision-making problems to configure the bitrate and bandwidth allocation on large-timescale frames and adjust the transmitting power for each small-timescale slot based on CSI; ii) To reduce the computational complexity of real-time joint optimization problem, we design a two-stage online successive convex approximation (TOSCA)-based resource allocation algorithm; iii) We conducted extensive experiments and the results demonstrate the effectiveness and robustness of our proposed strategy and algorithm.

Notations: We adopt x, $\mathbf{x}$, and $\mathbf{X}$ to denote a scalar, vector, and matrix, respectively. Superscript T stands for the transpose. $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|$ denote the l_1-norm and the l_2-norm of the vector $\mathbf{x}$, respectively. $\Pi_{\mathcal{B}}(\mathbf{x})$ denotes the projection of $\mathbf{x}$ onto the set $\mathcal{B}$.

II. System Model

As shown in Fig. 1, we consider a typical vehicular communication network that supports live video streaming transmission at high data rates. In this network, the RSU can only simultaneously serve V vehicular users (VUs) within its communication range, in diameter D, through the V2I links.

A. Live Video Streaming Transmission

In a realistic scenario, an RSU can transcode each live video segment into S different bitrate versions. Let $\mathcal{B} \triangleq \{b(1), b(2), \cdots, b(S)\}$ denote the set of all bitrate versions in a descending order, that is, $b(1) > b(2) > \cdots > b(S)$. Let $b_{v,t} \in \mathcal{B}$ represent the transmitting bitrate version selected by VU v at time slot t. Intuitively, as more power and bandwidth resources are allocated and the achievable data rate increases, the RSU prefers to select a higher bitrate to provide a higher definition video to the target VU.

Let $w_{v,t}$ and $p_{v,t}$ denote the bandwidth and the transmission power allocated to VU v by the RSU at time slot t, respectively. Then, the achievable data rate of the VU v at time slot t can be expressed as [7]:

$$R_{v,t} = w_{v,t} \log_2 \left(1 + \frac{p_{v,t} g_{v,t}}{w_{v,t} N_0}\right), \tag{1}$$

where $g_{v,t}$ and N_0 represent the channel gain between the RSU and VU v and the noise power spectral density, respectively. Considering that the RSU in real life has the limit of maximum transmit power $p_{\max}$ and maximum available bandwidth $w_{\max}$ at one certain time slot t, we have the following constrains:

$$\begin{aligned} \|\mathbf{p}_t\|_1 &\le p_{\max}, \qquad \forall t, \\ \|\mathbf{w}_t\|_1 &\le w_{\max}, \qquad \forall t, \end{aligned} \tag{2}$$

where $\mathbf{p}_t \triangleq (p_{1,t}, \cdots, p_{V,t})^T$ and $\mathbf{w}_t \triangleq (w_{1,t}, \cdots, w_{V,t})^T$.

B. Two-Timescale Operation Model

According to Clarke's model for the Doppler spectrum [8], [9], the short coherence time of vehicular communication channels alongside the sluggishness of resource allocation strategies, will result in a scenario where certain vehicles face resource wastage while others face resource scarcity. To overcome this problem, we propose a two-timescale adaptive live video streaming transmission mechanism which allocates the bitrate and bandwidth for each large-timescale frame based on the statistical knowledge of the CSI and updates the power allocation every small-timescale time slot based on available instantaneous CSI. The reasons for matching the allocation of different kinds of resources with different time scales are as follows: First, the frequent changes in bitrate between consecutive segments will inevitably lead to a degradation of user viewing experience, possibly even causing carsickness, and the frequent changes in bandwidth will increase the communication overhead; Secondly, real-time bitrate configuration leads to a notable increase in computational cost.

More specifically, we assume one large-timescale frame consists of T short-timescale slots, let $t = 1, 2, \cdots, T$ denote the index of short-timescale slot. During one large-timescale frame, the bandwidth allocated to the user and the selected video bitrate version remain unchanged, that is,

$$\begin{aligned} b_v &= b_{v,1} = \cdots = b_{v,T}, \\ w_v &= w_{v,1} = \cdots = w_{v,T}. \end{aligned} \tag{3}$$

For the sake of simplicity, we stack all bitrates, bandwidths, powers, and channel gains in a vector or matrix, denoted as $\mathbf{b} \in \mathbb{R}_+^{V\times 1}$, $\mathbf{w} \in \mathbb{R}_+^{V\times 1}$, $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_T) \in \mathbb{R}_+^{V\times T}$, and $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_T) \in \mathbb{R}_+^{V\times T}$, respectively.

C. Problem Formulation

To evaluate the performance of live video streaming in vehicular communication networks, we adopt three widely-used performance metrics, namely bit rate, rate instability index and achievable data rate. Hence, the QoE is defined as follows:

$$\begin{aligned} \psi(\mathbf{b}, \mathbf{w}, \mathbf{P}, \mathbf{G}) \triangleq \frac{1}{V} \sum_{v=1}^{V} \Bigg(& \omega_1 \frac{b_v}{b(1)} - \omega_2 \left(\frac{b_{v,0} - b_v}{b(1)}\right)^2 \\ & - \omega_3 \frac{1}{T} \sum_{t=1}^{T} \left(\frac{R_{v,t}}{b_v} - 1\right)^2 \Bigg), \end{aligned} \tag{4}$$

where $b_{v,0}$ represents the bitrate selected in the last large-timescale frame, ω_1, ω_2 and ω_3 are non-negative weights used to balance these factors.

In this paper, our objective is to maximize QoE by jointly optimizing the transmission bitrate, power allocation, and bandwidth allocation. Accordingly, the optimization problem is formulated as

$$\max_{\mathbf{b} \in \mathcal{B}, \mathbf{w} \ge \mathbf{0}} \ \mathbb{E}\left\{\max_{\mathbf{P} \ge \mathbf{0}} \psi(\mathbf{b}, \mathbf{w}, \mathbf{P}, \mathbf{G})\right\}, \tag{5a}$$

$$\text{s.t.} \quad \|\mathbf{p}_t\|_1 \le p_{\max}, \quad \forall t, \tag{5b}$$

$$\|\mathbf{w}\|_1 \le w_{\max}. \tag{5c}$$

It is arduous to solve Problem (5) due to *Challenge I:* the objective function (5a) is non-convex with respect to $\mathbf{w}$ and $\mathbf{P}$; and *Challenge II:* the discrete variable b_v implies that the problem is NP-hard. General speaking, there is no standard optimal solution for such a two timescale optimization problem. In the following section, we will propose an efficient solution based on two-stage online successive convex approximation (TOSCA) algorithm.

III. Design of Adaptive Video Streaming Transmission Mechanism

A. Overall Description

We first relax discrete bitrates to continuous variables to address *Challenge II*. Subsequently, we adopt the TOSCA algorithm [10] to solve the approximate non-convex optimization problem efficiently to address *Challenge I*. Finally, we map the relaxed bitrates into the given bitrate set $\mathcal{B}$ and refine the bandwidth allocation using the TOSCA algorithm again.

1) Relaxation Stage: Let $\tilde{\mathcal{B}} \triangleq [b(1), b(S)]$ be the relaxed version of $\mathcal{B}$ and the initial problem (5) is transformed to the following approximated problem with convex constraints:

$$\max_{\mathbf{b} \in \tilde{\mathcal{B}}, \mathbf{w} \ge \mathbf{0}} \ \mathbb{E}\left\{\max_{\mathbf{P} \ge \mathbf{0}} \psi(\mathbf{b}, \mathbf{w}, \mathbf{P}, \mathbf{G})\right\}, \tag{6a}$$

$$\text{s.t.} \quad \|\mathbf{p}_t\|_1 \le p_{\max}, \quad \forall t, \tag{6b}$$

$$\|\mathbf{w}\|_1 \le w_{\max}. \tag{6c}$$

Using primal decomposition, the problem can be decomposed into a large-timescale optimization sub-problem and a family of small-timescale optimization sub-problems, which are described as follows:

i) Small-timescale optimization problem: Given the channel gain $\mathbf{g}_t$ at time slot t, we can find the optimal $\mathbf{p}_t^*$ by solving the following small-timescale optimization problem:

$$\mathbf{p}_t^* \triangleq \arg\min_{\mathbf{p}_t \geq \mathbf{0}} \sum_{v=1}^{V} \left(\frac{R_{v,t}}{b_v} - 1 \right)^2, \tag{7a}$$

$$\text{s.t.} \quad \|\mathbf{p}_t\|_1 \leq p_{\max}. \tag{7b}$$

ii) Large-timescale optimization problem: Given the mapping $\mathbf{P}^*(\mathbf{G}) \triangleq \{\mathbf{p}_t^*(\mathbf{g}_t)\}_{t=1}^T$, the bitrate and bandwidth can be optimized as follows:

$$(\tilde{\mathbf{b}}^*, \tilde{\mathbf{w}}^*) \triangleq \max_{\mathbf{b}\in\tilde{\mathcal{B}}, \mathbf{w}\geq\mathbf{0}} \mathbb{E}\left\{\psi(\mathbf{b}, \mathbf{w}, \mathbf{P}^*(\mathbf{G}), \mathbf{G})\right\}, \tag{8a}$$

$$\text{s.t.} \quad \|\mathbf{w}\|_1 \leq w_{\max}. \tag{8b}$$

2) Projection Stage: The relaxed bitrates are mapped into the given bitrate set, i.e.,

$$\mathbf{b}^* \triangleq \Pi_{\mathcal{B}}(\tilde{\mathbf{b}}^*) = \min_{\mathbf{b}\in\mathcal{B}} \|\mathbf{b} - \tilde{\mathbf{b}}^*\|. \tag{9}$$

3) Refinement Stage: For a given fixed bitrate vector $\mathbf{b}^*$, the problem reduces to

$$\max_{\mathbf{w}\geq\mathbf{0}} \mathbb{E}\left\{\max_{\mathbf{P}\geq\mathbf{0}} \psi(\mathbf{b}^*, \mathbf{w}, \mathbf{P}, \mathbf{G})\right\}, \tag{10a}$$

$$\text{s.t.} \quad \|\mathbf{p}_t\|_1 \leq p_{\max}, \quad \forall t, \tag{10b}$$

$$\|\mathbf{w}\|_1 \leq w_{\max}, \tag{10c}$$

which can be solved by the same strategies as in Relaxation Stage. Therefore, we then only focus on solving the optimization problems (7) and (8) in Relaxation Stage.

B. Short-Term Optimization Problem

At time slot t, the RSU firstly acquires the channel gain vector $\mathbf{g}_t$, and aims to adjust the power allocation $\mathbf{p}_t$ according to the selected bitrate version $\mathbf{b}$, and allocated bandwidth $\mathbf{w}$.

1) Optimal-Solution Algorithm Design: The stationary point of the short-term power allocation $\mathbf{p}_t$ can be obtained when the condition in the following proposition is satisfied.

Proposition 1. *The optimal $\mathbf{p}_t^*$ satisfies*

$$R_{v,t} \leq b_v, \quad \forall v.$$

In particular, the equality holds if the power budget satisfies

$$p_{\max} \geq \sum_v \frac{w_v N_0}{g_{v,t}} \left(2^{\frac{b_v}{w_v}} - 1 \right),$$

Proof. Consider the partial Lagrangian function:

$$\mathcal{L}(\mathbf{p}_t, \lambda) = \sum_{v=1}^{V} \left(\frac{R_{v,t}}{b_v} - 1 \right)^2 + \lambda \left(\|\mathbf{p}_t\|_1 - p_{\max} \right), \tag{11}$$

The optimal solution $\mathbf{p}_t^*$ must satisfy

$$\frac{\partial \mathcal{L}(\mathbf{p}_t, \lambda)}{\partial p_{v,t}} = \frac{2\left(\frac{R_{v,t}}{b_v} - 1\right) w_v g_{v,t}}{(w_v N_0 + p_{v,t} g_{v,t}) b_v \ln 2} + \lambda = 0. \tag{12}$$

We can conclude that $R_{v,t} \leq b_v$ must hold since all parameters $\omega_{v,t}, g_{v,t}, p_{v,t}, \lambda$ are non-negative. In particular, if the total power budget satisfies

$$p_{\max} \geq \sum_v \frac{w_v N_0}{g_{v,t}} \left(2^{\frac{b_v}{w_v}} - 1 \right), \tag{13}$$

the RSU can allocate $p_{v,t} = \frac{w_v N_0}{g_{v,t}} \left(2^{\frac{b_v}{w_v}} - 1 \right)$ to VU v such that $R_{v,t} = b_v$.

□

Based on Proposition 1, power allocation optimization is only required when

$$p_{\max} < \sum_v \frac{w_v N_0}{g_{v,t}} \left(2^{\frac{b_v}{w_v}} - 1 \right). \tag{14}$$

To solve it, we introduce an auxiliary optimization variables $\mathbf{z} = [z_1, z_2, \cdots, z_V]^T$ and the initial problem is equivalent to

$$\min_{\mathbf{p}_t \geq \mathbf{0}, \mathbf{z}} \sum_{v=1}^{V} z_v^2, \tag{15a}$$

$$\text{s.t.} \quad \frac{R_{v,t}}{b_v} - 1 \geq z_v, \quad \forall v, \tag{15b}$$

$$\|\mathbf{p}_t\|_1 \leq p_{\max}, \tag{15c}$$

which can be solved by CVX, and the computational complexity is $\mathcal{O}(V^{3.5})$. Although it yields the optimal solution, the high computational complexity renders its execution challenging, especially in scenarios with high-speed VUs and extremely short coherence time.

2) Low-Complexity Algorithm Design: As demonstrated in Proposition 1, the basic idea of power allocation strategy is to reduce the objective function in (15a) by allocating more power to VUs with superior channel gain. Simultaneously, it should ensure fairness by prohibiting any VU from unnecessarily holding an excessive amount of power.

Inspired by this idea, we resort to the classic water-filling algorithm to reduce the complexity. Specifically, there are three primary steps in the allocation procedure. Step 1: The remaining power p is allocated to the set of the remaining VUs $\mathcal{V}$ according to Eq. (2) reported in [11], which can be expressed as:

$$p_{v,t} = \left[\frac{w_v}{\lambda} - \frac{w_v N_0}{g_{v,t}} \right]^+, \tag{16}$$

where $\lambda \geq 0$ is the optimal dual variable for the transmit power constraint in (15c). It can be obtained via one dimensional search techniques, e.g., bisection method. ii) Step 2: The achievable data rate for each VU is assessed. In order to avoid resource waste and violate Proposition 1 in the event that the achievable data rate of a VU exceeds the target bitrate, we employ a bisection method to incrementally reduce the allocated power until the achievable data rate matches the target bitrate. Step 3: Re-distribute the remaining power to the VUs that haven't yet received sufficient power by repeating Steps 1 and 2 until all the power is completely assigned.

C. Long-Term Optimization Problem

The large-timescale optimization is executed once at the beginning of each large-timescale frame to improve the QoE, which can be solved recursively by the TOSCA algorithm. Specifically, at iteration (k), a new channel gain matrix $\mathbf{G}^{(k)}$ is randomly drawn from the known distribution according the future positions of the vehicles. Given $(\mathbf{b}^{(k)}, \mathbf{w}^{(k)})$ and $\mathbf{G}^{(k)}$, the intermediate variable $\mathbf{P}^{(k)}$ is calculated by solving the small-timescale optimization problem in the aforementioned subsection. Thus, we can solve the following problem:

$$(\bar{\mathbf{b}}^{(k)}, \bar{\mathbf{w}}^{(k)}) \triangleq \max_{\mathbf{b}\in\tilde{\mathcal{B}}, \mathbf{w}\geq\mathbf{0}} \quad \bar{\psi}^{(k)}(\mathbf{b}, \mathbf{w}, \mathbf{P}^{(k)}, \mathbf{G}^{(k)}), \tag{17a}$$

$$\text{s.t.} \quad \|\mathbf{w}\|_1 \leq w_{\max}, \tag{17b}$$

with the surrogate function $\bar{\psi}^{(k)}(\mathbf{b}, \mathbf{w}, \mathbf{P}^{(k)}, \mathbf{G}^{(k)})$ defined as

$$\begin{aligned} &\bar{\psi}^{(k)}(\mathbf{b}, \mathbf{w}, \mathbf{P}^{(k)}, \mathbf{G}^{(k)}) \\ &= \mathrm{tr}((\mathbf{b}-\mathbf{b}^{(k)})^T(\rho^{(k)}\nabla_{\mathbf{b}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_b^{(k)})) \\ &+ \mathrm{tr}((\mathbf{w}-\mathbf{w}^{(k)})^T(\rho^{(k)}\nabla_{\mathbf{w}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_w^{(k)})) \\ &+ \tau_b\|\mathbf{b}-\mathbf{b}^{(k)}\|^2 + \tau_w\|\mathbf{w}-\mathbf{w}^{(k)}\|^2, \end{aligned} \tag{18}$$

where $\tau_b > 0$ and $\tau_w > 0$ are constant to ensure strong convexity; $\mathbf{f}_b^{(k)}$ and $\mathbf{f}_w^{(k)}$ are the accumulation vectors that can be updated recursively as follows:

$$\begin{aligned} \mathbf{f}_b^{(k)} &= (1-\rho^{(k)})\mathbf{f}_b^{(k-1)} + \rho^{(k)}\nabla_{\mathbf{b}}\bar{\psi}^{(k)}, \\ \mathbf{f}_W^{(k)} &= (1-\rho^{(k)})\mathbf{f}_W^{(k-1)} + \rho^{(k)}\nabla_{\mathbf{W}}\bar{\psi}^{(k)}, \end{aligned} \tag{19}$$

with $\rho^{(k)} \in (0,1]$ being a sequence to be properly chosen [10, Assumption 5]: $\rho^{(k)} \to 0$, $\frac{1}{\rho^{(k)}} \leq O(k^{\beta})$ for $\beta \in (0,1)$, and $\sum_k(\rho^{(k)})^2 < \infty$.

Proposition 2. *The optimal solution to Problem in (17) can be decomposed into two independent sub-problems w.r.t.* $\mathbf{b}$ *and* $\mathbf{w}$*, which are, respectively, given by*

$$\bar{\mathbf{b}}^{(k)} = \Pi_{\tilde{\mathcal{B}}}\left[\mathbf{b}^{(k)} - \frac{1}{2\tau_b}(\rho^{(k)}\nabla_{\mathbf{b}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_b^{(k)})\right], \tag{20a}$$

$$\bar{\mathbf{w}}^{(k)} = \left[\mathbf{w}^{(k)} - \frac{1}{2\tau_w}(\rho^{(k)}\nabla_{\mathbf{w}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_w^{(k)} + \lambda\mathbf{1})\right]^+, \tag{20b}$$

where $\lambda \geq 0$ *is the optimal dual variable for the transmit power constraint in (17b). It can be obtained via one dimensional search techniques, e.g., bisection method.*

Proof. i) Optimize $\mathbf{b}$*:* The objective function is quadratic w.r.t. each element of $\mathbf{b}$ and the constraint is in a convex domain. Therefore, The optimal solution can be derived by projecting the stationary point of the objective onto the convex domain, which leads to the desired Eq. (20a).

ii) Optimize $\mathbf{w}$*:* The partial Lagrangian function is given by

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \lambda) = &\mathrm{tr}((\rho^{(k)}\nabla_{\mathbf{w}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_w^{(k)})^T(\mathbf{w}-\mathbf{w}^{(k)})) \\ &+ \tau_w\|\mathbf{w}-\mathbf{w}^{(k)}\|^2 + \lambda\left(\|\mathbf{w}\|_1 - w_{\max}\right). \end{aligned} \tag{21}$$

Algorithm 1 The TOSCA-based adaptive live video streaming transmission mechanism

---

Initialization: $\{\gamma^{(k)}\}, k=0, \mathbf{b}^{(0)} \in \mathcal{B}, \mathbf{w}^{(0)} \geq \mathbf{0}$.
Relaxation Stage: Repeat the following **S1-S5**:
S1: Obtain the random channel gain matrix $\mathbf{G}^{(k)}$.
S2: At each time slot k, if $p_{\max}$ satisfies Eq. (13), set $p_{v,t} = \frac{w_v N_0}{g_{v,t}}\left(2^{\frac{b_v}{w_v}} - 1\right)$.
Otherwise, compute $\mathbf{p}_t^{(k)}$ by the following steps:
- **S2.1:** Compute the set of remaining VUs $\mathcal{V}$ and the total remaining power p.
- **S2.2:** Compute power allocation according to Eq. (16).
- **S2.3:** Assess the achievable rate for each VU in $\mathcal{V}$. If $R_{v,t} \geq b_v$, reduce the power such that $R_{v,t} = b_v$ and remove VU v from $\mathcal{V}$. Otherwise, $p_{v,t} = 0$.
- **S2.4:** If all VUs satisfies $R_{v,t} < b_v$, compute power allocation according to Eq. (16) and STOP. Otherwise, $\tau = \tau + 1$ and go to **S2.1**.

S3: Update $\mathbf{f}_b^{(k)}$ and $\mathbf{f}_w^{(k)}$ according to Eq. (19).
S4: Compute $\bar{\mathbf{b}}^{(k)}$ and $\bar{\mathbf{w}}^{(k)}$ according to Eq. (20).
S5: Update $\mathbf{b}^{(k+1)}$ and $\mathbf{w}^{(k+1)}$ according to Eq. (24).
S6: $k = k+1$ and go to **S1**.
Projection Stage:
S7: Project the relaxed bitrate $\mathbf{b}^* \triangleq \Pi_{\mathcal{B}}$.
Refinement Stage:
S8: Execute the similar steps as **S1-S6** with setting $\mathbf{f}_b^{(k)} = \mathbf{0}$ and $\bar{\mathbf{b}}^{(k)} = \mathbf{b}^{(k)}$.

The dual function is given by $g(\lambda) = \inf_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \lambda)$. Since $\mathcal{L}(\mathbf{w}, \lambda)$ is a convex function w.r.t $\mathbf{w}$, we can find the optimal matrix $\mathbf{w}$ from the following optimality condition:

$$\begin{aligned} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}, \lambda) = &\rho^{(k)}\nabla_{\mathbf{w}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_w^{(k)} \\ &+ 2\tau_w(\mathbf{w}-\mathbf{w}^{(k)}) + \lambda\mathbf{1} = \mathbf{0}, \end{aligned} \tag{22}$$

which yields

$$\mathbf{w} = \mathbf{w}^{(k)} - \frac{1}{2\tau_w}(\rho^{(k)}\nabla_{\mathbf{w}}\bar{\psi}^{(k)} + (1-\rho^{(k)})\mathbf{f}_w^{(k)} + \lambda\mathbf{1}). \tag{23}$$

Then, $\mathbf{w}$ is projected onto the non-negative domain, which leads to the desired Eq. (20b). □

Finally, $(\mathbf{b}, \mathbf{w})$ is updated according to

$$\begin{aligned} \mathbf{b}^{(k+1)} &= (1-\gamma^{(k)})\mathbf{b}^{(k)} + \gamma^{(k)}\bar{\mathbf{b}}^{(k)}, \\ \mathbf{w}^{(k+1)} &= (1-\gamma^{(k)})\mathbf{w}^{(k)} + \gamma^{(k)}\bar{\mathbf{w}}^{(k)}, \end{aligned} \tag{24}$$

where $\gamma^{(k)} \in (0,1]$ is a step size sequence satisfying $\gamma^{(k)} \to 0$, $\sum_k \gamma^{(k)} = \infty$, $\sum_k(\gamma^{(k)})^2 < \infty$, and $\lim_{k\to\infty}\gamma^{(k)}/\rho^{(k)} = 0$ [10, Assumption 5].

The above procedure is summarized in Algorithm 1.

D. Convergence and Complexity Analysis

Convergence analysis: Firstly, according to [10, Thm. 2], the objective value is convergent in both relaxation and refinement stages. Secondly, the projection has a closed-form solution. Therefore, our proposed algorithm is convergent.

Complexity analysis: In both relaxation and refinement stages, the complexity of updating $\mathbf{P}$ at small timescale is $\mathcal{O}(VT)$. The complexity of updating $\mathbf{b}$ and $\mathbf{w}$ mainly depends on computing the gradients $\nabla_{\mathbf{b}}\bar{\psi}$ and $\nabla_{\mathbf{w}}\bar{\psi}$, which is $\mathcal{O}(V)$. In the projection stage, the complexity of projection is $\mathcal{O}(V)$.

IV. Numerical Results

To numerically evaluate the performance of the proposed algorithm, the coherence time τ is set as 10 ms, the number of time slots within each large-scale frame T is 500. In live video transmission services, the commonly-used available bitrates include 5.8 Mbps, 3 Mbps, 1.75 Mbps, and 0.75 Mbps[1]. The available bandwidth $w_{\max}$ is 10 MHz. The coverage radius of the RSU is 200 m. The speed of the vehicles ranges from 100 to 120 km/h. Considering the safety distance between vehicles, we assume that there are two moving VUs in each lane within the coverage radius of the RSU.

The proposed algorithm are compared with the following baselines: i) Baseline 1: Bandwidth and bitrate are optimized at each large-timescale frame, and the power is uniformly allocated; ii) Baseline 2: Bandwidth and bitrate are optimized based on a fixed uniform power allocation at the beginning of each large-timescale frame, and the power is optimized at each small-timescale slot; iii) Baseline 3: Bandwidth is uniformly allocated, the selected bitrate is consistent with the last large-timescale frame, and the power is optimized at each small-timescale slot.

Figs. 2-3 illustrates the QoE for different configurations of transmission power at the RSU and the number of lanes on the road. Firstly, across various configurations, the proposed design routinely beats three baselines in terms of QoE. Secondly, the uniform power allocation strategy in baseline 1 makes some vehicles experience resource wastage, while others face resource scarcity, which is exacerbated by the increase of transmission power. It demonstrate the significance and necessity of dynamically adjusting power for improving the QoE. The fact that baseline 2 considerably enhances baseline 1's performance by optimizing the power allocation in each time slot further supports this conclusion. Baseline 2 has a reduced complexity because bitrate, bandwidth, and power allocation are not optimized jointly in large-timescale frames. This suggests that baseline 2 can be utilized as a sub-optimal method in situations when computation complexity is low or decision-latency is minimal. Thirdly, the performance gap between baseline 3 and the proposed algorithm demonstrates the need for dynamic bandwidth and bitrate optimization. In conclusion, the mismatches in bitrate, bandwidth and power allocation will cause resource waste and deteriorate the QoE.

V. Conclusion

In this paper, we proposed a two-timescale adaptive live video streaming transmission mechanism, where the bitrate and bandwidth are allocated for each large-timescale frame based on the statistical CSI and the power allocation is updated every small-timescale. To solve the two-timescale optimization problem efficient, we employed the TOSCA algorithm to decoupled the QoE maximization problem into large- and short-timescale problems. Simulation results confirmed the effectiveness of the proposed mechanism across various power budgets and lane configurations.

[1] https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2

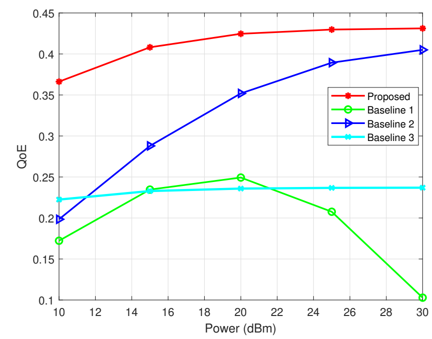

Fig. 2. Evaluation of QoE under different transmit power.

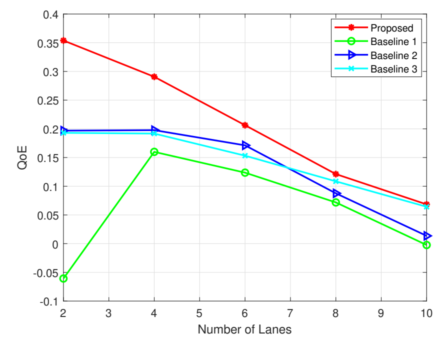

Fig. 3. Evaluation of QoE under different number of lanes.

References

[1] L. Cui, D. Su, S. Yang, Z. Wang, and Z. Ming, "TCLiVi: Transmission control in live video streaming based on deep reinforcement learning," *IEEE Trans. Multimedia*, vol. 23, pp. 651-663, 2021.

[2] X. Jiang, F. R. Yu, T. Song, and V. C. M. Leung, "Resource allocation of video streaming over vehicular networks: A survey, some research issues and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 5955-5975, Jul. 2022.

[3] K. Wu, X. Cao, P. Yang, Z. Yu, D. O. Wu, and T. Q. S. Quek, "QoE-driven video transmission: Energy-efficient multi-UAV network optimization," *IEEE Trans. Network Science Engineering*, vol. 11, no. 1, pp. 366-379, Jan.-Feb. 2024.

[4] S. Wang, S. Bi, and Y. -J. A. Zhang, "Adaptive wireless video streaming: Joint transcoding and transmission resource allocation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3208-3221, 2022

[5] Y. Guo, F. R. Yu, J. An, K. Yang, C. Yu, and V. C. M. Leung, "Adaptive bitrate streaming in wireless networks with transcoding at network edge using deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3879-3892, 2020.

[6] B. Feng, C. Feng, D. Feng, Y. Wu, and X.-G. Xia, "Proactive content caching scheme in urban vehicular network," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 4165–4180, May 2023.

[7] O. Semiari, W. Saad, and M. Bennis, "Joint millimeter wave and microwave resources allocation in cellular networks with dual-mode base stations," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4802–4816, Jul. 2017.

[8] T. Feng and T. R. Field, "Statistical analysis of mobile radio reception: an extension of clarke's model," *IEEE Trans. Commun.*, vol. 56, no. 12, pp. 2007-2012, Dec. 2008.

[9] V. Va, J. Choi, and R. W. H. Jr., "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5014–5029, Nov. 2016.

[10] A. Liu, V. K. N. Lau, and M.-J. Zhao, "Online successive convex approximation for two-stage stochastic nonconvex optimization," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5941–5955, Nov. 2018.

[11] N. Kalpana and M. Z. A. Khan, "Fast computation of generalized waterfilling problems," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1884–1887, Nov. 2015.