Latency-sensitive Service Delivery with UAV-Assisted 5G Networks

Shashi Raj Pandey, Student Member, IEEE, Kitae Kim, Madyan Alsenwi, Yan Kyaw Tun, Zhu Han, Fellow, IEEE, Choong Seon Hong, Senior Member, IEEE

Abstract

In this letter, a novel framework to deliver critical spread out URLLC services deploying unmanned aerial vehicles (UAVs) in an out-of-coverage area is developed. To this end, the resource optimization problem, i.e., resource block (RB) and power allocation, are studied for UAV-assisted 5G networks to meet the objective of jointly maximizing the average sum-rate and minimizing the transmit power of UAV while satisfying the URLLC requirements. To cope with the sporadic URLLC traffic problem, an efficient online URLLC traffic prediction model based on Gaussian Process Regression (GPR) is proposed to derive optimal URLLC scheduling and transmit power strategy. The formulated problem is revealed as a mixed-integer nonlinear programming (MINLP), which is solved following the introduced successive minimization algorithm. Finally, simulation results are provided to show the efficacy of our proposed solution approach.

Index Terms

Unmanned aerial vehicles (UAVs), 5G NR, URLLC, gaussian process regression (GPR).

I. INTRODUCTION

Recently, unmanned aerial vehicles (UAVs)-assisted cellular networks are deployed as a promising alternative to handle out-of-coverage issues. These UAVs work as a flying base station (BS) that offers computing and communication facilities for Internet of Things (IoT) devices and applications, e.g., disaster and rescue, autonomous control, military operations, and smart farming [1], [2]. However, UAVs are energy-constrained, which needs to delicately allocate its...
available resources while concurrently serving as a local multi-access edge computing (MEC) infrastructure. Besides, to fully reap the benefits of UAVs, it is essential to have its efficient integration with the 5G New Radio (NR) standards to deliver latency-sensitive data packets.

5G NR standardization has been a significant paradigm shift to realize full-fledged communication networks for the next-generation applications. The Third Generation Partnership Project (3GPP) Release 15 5G-NR [3] supports connectivity for massive device densities, high data rate, and ultra-reliable low-latency communication (URLLC) services. The key feature of URLLC supports a high level of reliability and low latency services, i.e., the latency of less than 1ms while guaranteeing packet error rates (PER) in the order-of-five \(10^{-5}\) [3]. This mandates immediate transmission of critical URLLC packets over the short transmission time interval (sTTI) [3] to meet the stringent latency-reliability requirements. Hence, recent studies focus on resource block (RB) allocation problem to deliver URLLC services [4]–[8]. However, the solution to overcome the challenges due to the dynamic nature of URLLC traffic is non-trivial. Inline with [6], several approaches have been recently proposed to optimize resource allocation [4], [7], considering a general arrival process to capture dynamic URLLC traffic and perform RB allocation. Authors in [8] proposed a model-free approach to guarantee end-to-end latency and end-to-end reliability imposing latency constraints in the optimization problem. However, the authors in [4], [6]–[8] consider a typical network infrastructure with a fixed BS, having no energy restrictions, and the mobile users deployed randomly under its coverage area. Furthermore, the challenges of integrating 5G features on energy-constrained UAV systems, particularly serving out-of-coverage users, are still overlooked in recent studies [1], [2], [9].

In this work, we first leverage the benefits of UAVs to ensure latency-sensitive data transmission and offer 5G services in an out-of-coverage area with unmanned aerial systems. In particular, we adopt a Gaussian Process Regression (GPR) [10] approach to capture the network dynamics and predict the URLLC traffic online for executing an efficient resource allocation (i.e., RBs and power) strategy of the UAV. GPR, a flexible and robust active learning approach that shows merit in tackling parametric models’ issues [10], allows UAVs to predict latency-sensitive data packets before serving the remote IoT devices. In summary, the main contribution of this paper is a novel framework to deliver critical URLLC services in an out-of-coverage area deploying UAVs. In particular, we develop a practical integration of 5G features with UAV networks and

\[\text{Resource block (RB) is the smallest unit of bandwidth resources defined in Long Term Evolution (LTE)}\] [3].
leverage GPR to appropriately characterize and predict the dynamic URLLC traffic online. We later fuse this prediction information to optimize the radio resources, i.e., RBs and transmit power, efficiently at the UAV. Then, we formulate a joint optimization problem for maximizing the average sum-rate while minimizing the transmit power of UAV with the constraints to satisfy stringent URLLC requirements. The formulated problem is revealed as an MINLP, which is NP-hard due to the binary constraint. Hence, we relax the binary constraints and decompose the proposed problem into two sub-problems, which are solved using a low-complexity near-optimal successive minimization algorithm. Simulation results show that the proposed approach achieves a performance gain of up to 25% as compared with the baseline for satisfying the reliability constraints.

To our best knowledge, this is the first work that adopts GPR for performing dynamic URLLC traffic prediction and resource optimization to guarantee maximum average sum-rate and minimum transmit power, jointly, in a UAV-assisted 5G network.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

In this work, as depicted in Fig. 1, we consider a wireless network system where a single UAV is deployed in an out-of-coverage area to provide wireless communication services to a
set of URLLC users $\mathcal{U}$ of $|\mathcal{U}| = U$ at time slot in the set $\mathcal{T}$ of $|\mathcal{T}| = T$. We fixed the location of UAV at an altitude $H$, and the horizontal coordinates $(x, y)$; thus, and the position of UAV is $\mathbf{c} = [x, y, H]$. Similarly, the location of URLLC user is $\mathbf{o}_u = [x_u, y_u], \forall u \in \mathcal{U}$. The system available total bandwidth is divided into a set of RBs $\mathcal{B}$ of $|\mathcal{B}| = B$. We considered line-of-sight (LOS) link is available, and the orthogonal frequency division multiple access (OFDMA) scheme is adopted to share radio resources amongst the URLLC users. Let $a_{ub}(t) \in \{0, 1\}$ be the RB assignment variable at time slot $t$ defined as

$$a_{ub}(t) = \begin{cases} 1, & \text{if user } u \text{ is assigned to RB } b \text{ at time slot } t, \\ 0, & \text{otherwise}. \end{cases}$$

Then, following the free-space path loss model [1], the channel gain between users $u$ assigned to RB $b$ and UAV at time slot $t$ is

$$\gamma_{ub}(t) = \gamma_0 d_u^{-\theta} = \frac{\gamma_0}{(H^2 + ||\mathbf{c} - \mathbf{o}_u||^2)^{\theta/2}}, \forall u \in \mathcal{U}, \forall b \in \mathcal{B}, \forall t \in \mathcal{T},$$

where, respectively, $d_u$ is the distance between the UAV and user $u$ defined as $d_u = \sqrt{H^2 + ||\mathbf{c} - \mathbf{o}_u||^2}$, $\theta$ is the path loss exponent, and $\gamma_0$ is the channel gain at the reference distance $d_0 = 1$ m. Therefore, the achievable downlink data rate over RB $b$ for user $u \in \mathcal{U}$ at time slot $t$ is

$$r_{ub}(t) = \omega \log_2 \left(1 + \frac{p_{ub}(t) \gamma_{ub}(t)}{n_0}\right), \forall u \in \mathcal{U}, \forall b \in \mathcal{B}, \forall t \in \mathcal{T},$$

where $\omega$ is the bandwidth of each RB, $n_0$ is the additive white Gaussian noise power, $p_{ub}(t)$ is the transmit power of the UAV over RB $b$ at time slot $t$, respectively.

Let $L_u(t)$ denotes the random URLLC traffic arrivals at time slot $t$ of user $u$. Therefore, the system reliability constraint can be defined as

$$\Pr \left[ \sum_{b=1}^{B} r_{ub}(t) \leq \beta L_u(t) \right] \leq \epsilon, u \in \mathcal{U}, \forall t \in \mathcal{T},$$

where $\beta$ denotes the URLLC packet size, and $\epsilon$ is a small outage threshold value. Then, by using the Markov’s Inequality, we can rewrite the reliability constraint in (4) as a linear constraint as follows:

$$\Pr \left[ \sum_{b=1}^{B} r_{ub}(t) \leq \beta E[L_u(t)] \middle| \right] \leq \frac{\beta E[L_u(t)]}{\sum_{b=1}^{B} r_{ub}(t)}, \forall t \in \mathcal{T}. $$

$^2$We will use the term “users” to denote URLLC users henceforth.
B. Problem Formulation

In order to satisfy the stringent latency requirements of URLLC traffic (5), we need to consider the limited resource capacity, i.e., transmit power and RBs, at the UAV. Therefore, we formulate our optimization problem to jointly maximize the average sum-rate and minimize the transmit power of the UAV, while ensuring the URLLC constraints, as follows:

\[
P: \text{maximize } \frac{1}{T} \sum_{t=1}^{T} \left( \sum_{u=1}^{U} \sum_{b=1}^{B} r_{u}^{b}(t) - \zeta \sum_{u=1}^{U} \sum_{b=1}^{B} p_{u}^{b}(t) \right) \tag{6a}
\]

subject to

\[
\sum_{b=1}^{B} r_{u}^{b}(t) \geq \frac{\beta \mathbb{E}[L_{u}(t)]}{\epsilon}, \quad \forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \tag{6b}
\]

\[
\sum_{u=1}^{U} a_{u}^{b}(t) \leq 1, \quad \forall b \in \mathcal{B}, \forall t \in \mathcal{T}, \tag{6c}
\]

\[
a_{u}^{b}(t) \in \{0, 1\}, \quad \forall u \in \mathcal{U}, b \in \mathcal{B}, \forall t \in \mathcal{T}, \tag{6d}
\]

\[
\sum_{u=1}^{U} \sum_{b=1}^{B} a_{u}^{b}(t)p_{u}^{b}(t) \leq P_{\text{max}}, \forall t \in \mathcal{T}, \tag{6e}
\]

\[
0 \leq p_{u}^{b}(t) \leq P_{\text{max}}, \forall u \in \mathcal{U}, \forall b \in \mathcal{B}, \forall t \in \mathcal{T}, \tag{6f}
\]

where \(a\) and \(p\) are, respectively, the vectors of resource allocation and transmit power, \(\zeta > 0\) is a scaling constant, (6b) is the URLLC reliability constraint, and constraints (6c) and (6d) ensure one RB is assigned to only one user at most. Constraint (6e) and (6f) defines the total transmit power of UAV over all RBs is bounded by the system power budget \(P_{\text{max}}\). Besides, \(a\) and \(p\) are the vectors of resource allocation and transmit power.

III. Proposed Solution Approach

The formulated problem in (6) is an MINLP, which may require exponential-complexity to solve. To solve (6) efficiently and suboptimally, we decompose it into two sub-problems: (i) RB allocation, and (ii) transmit power allocation.

A. RB Allocation Problem for a Given Power Allocation

For a given power allocation, we can relax the binary constraint (6c) and recast the integer programming problem (6) as a RB allocation problem. Then, the fractional solution is rounded
to get a solution to the original integer problem. Hence, we pose (6) as

\[ \text{P1: maximize } \sum_{t=1}^{T} \sum_{u=1}^{U} \sum_{b=1}^{B} a_{u}^{b}(t) \]

subject to

\[ \sum_{b=1}^{B} r_{u}^{b}(t) \geq \frac{\beta \mathbb{E}[L_{u}(t)]}{\epsilon}, \quad \forall u \in U, \forall t \in T, \quad (7b) \]

\[ \sum_{u=1}^{U} a_{u}^{b}(t) \leq 1, \quad \forall b \in B, \forall t \in T, \quad (7c) \]

\[ a_{u}^{b}(t) \in [0, 1], \quad \forall u \in U, b \in B, \forall t \in T. \quad (7d) \]

The above problem is a maximization problem with concave objective and linear constraints; hence, a convex optimization problem which can be solved efficiently using ECOS solver in the CVXPY toolkit.

B. Transmit Power Allocation Problem Given RB Allocation

For a given RB allocation, we can recast the integer programming problem (6) as a transmit power allocation problem as

\[ \text{P2: maximize } \sum_{t=1}^{T} \left( \sum_{u=1}^{U} \sum_{b=1}^{B} r_{u}^{b}(t) - \zeta \sum_{u=1}^{U} \sum_{b=1}^{B} p_{u}^{b}(t) \right) \]

subject to

\[ \sum_{b=1}^{B} r_{u}^{b}(t) \geq \frac{\beta \mathbb{E}[L_{u}(t)]}{\epsilon}, \quad \forall u \in U, \forall t \in T, \quad (8b) \]

\[ \sum_{u=1}^{U} \sum_{b=1}^{B} a_{u}^{b}(t)p_{u}^{b}(t) \leq P_{\text{max}}, \forall t \in T, \quad (8c) \]

\[ 0 \leq p_{u}^{b}(t) \leq P_{\text{max}}, \forall u \in U, \forall b \in B, \forall t \in T. \quad (8d) \]

For any given RB allocation, the above problem is a convex optimization problem which can be solved efficiently by the UAV.

However, to solve (6) and (7), we first need to efficiently predict the expected random URLLC \( L_{u}(t), \forall u \) traffic load at time \( t \). A naive approach is to quantify \( L_{u}(t), \forall u \) as a random variable with some known distribution \( [6] \); however, it may result poor performance in making online scheduling decision for URLLC traffic placements. Hence, we resort to a GPR approach, which is a flexible and robust mechanism to capture the network dynamics and provide online URLLC traffic prediction with minimal errors.
Algorithm 1 Iterative solution approach for the relaxed problem

1: **Initialization:** Set $k = 0$, $\delta_1, \delta_2 > 0$, and initial solutions $(a^{(0)}(t), p^{(0)}(t))$;
2: Obtain URLLC traffic prediction $L_u(t)$ from (11);
3: repeat
   4: Compute $a^{(k+1)}(t)$ from (P1) at given $p^k(t)$;
   5: Compute $p^{(k+1)}(t)$ from (P2) at given $a^{(k+1)}(t)$;
   6: $k = k + 1$;
   7: until $\|a^{(k+1)}(t) - a^{(k)}(t)\| \leq \delta_1$, $\|p^{(k+1)}(t) - p^{(k)}(t)\| \leq \delta_2$;
8: Then, set $(a^{(k+1)}(t), p^{(k+1)}(t))$, as the desired solutions.

C. GPR-based URLLC Traffic Prediction

Our aim is to perform an online prediction for the incoming URLLC traffic gain of the next time slot $\hat{L}_u(t + 1)$ at each time slot $t$. To achieve that, we update the learning parameters over a moving window. Let $N$ be the window size, i.e., the window composed of the last $N$ time slots, in the set $\mathcal{N}$ of $|\mathcal{N}| = N$. The model parameters are trained on the data, i.e., URLLC traffic, inside the window. Then, trained parameters are used to predict the URLLC traffic load of the next time slot. In this view, for a finite data set $(t_n, L_u(t_n)), \forall n \in \mathcal{N}$, a general GPR-based prediction model \cite{10} can be modified as

$$\hat{L}_u(t + 1) = f(L_u(t)) + \varepsilon, \forall u \in \mathcal{U},$$

(9)

where $f(\cdot)$ is the regression function modeled as a Gaussian process with the mean function set to zero when there is no prior knowledge, and $\varepsilon$ is a Gaussian distribution random variable with $\sigma^2_\varepsilon$ variance and zero mean that represents the independent noise, with the kernel function $g(\cdot)$ defined as

$$g(L(t - m), L(t - n), \theta) = \exp \left( \frac{-1}{\theta_1} \sin^2 \left( \frac{\pi}{\theta_2} (L(t - m) - L(t - n)) \right) \right),$$

(10)

where $m, n \in \{0, 1, 2, \ldots, N\}$, and $\theta = [\theta_1, \theta_2]$ defines a vector of the lengths and period hyper-parameters, respectively. Accordingly, the URLLC traffic load prediction at time slot $t + 1$ is given as

$$\hat{L}_u(t + 1) = g^1(t)G^{-1}[L_u(t - N), L_u(t - N + 1, \ldots L_u(t))],$$

(11)
where $G = [g(t - m, t - n)]$, and $g(t) = [g(t, t - n)]$, $\forall m, n \in \mathcal{N}$. Moreover, the variance (uncertainty) on the predicted value is given by

$$\text{Var}(\hat{L}_u(t + 1)) = g(t, t) - g^\dagger(t)G^{-1}g(t).$$

(12)

The traffic prediction is obtained from (11) and exploring highly uncertain traffic provides more insight.

IV. Performance Evaluation

In our simulation, we consider the UAV is deployed at a fixed height in the range $H = [100, 150]$m with the coverage area of $(250 \times 250)$m$^2$. We set the number of users in the range $[5, 20]$, which are positioned randomly in the UAV’s coverage area. We consider the bandwidth of each RB $\omega = 180$ kHz. The total available transmit power is set as $P_{\text{max}} = 10$ Watts, the noise power $n_0 = -174$ dBm/Hz, and the channel gain at the reference distance is $\gamma_0 = -30$ dBm. We consider the URLLC packet size as $\beta = 32$ bytes, and the window size $N = 1000$ time slots for traffic prediction. Due the absence of real URLLC traffic datasets, we adopt the real world stock market dataset to replicate and characterize the URLLC traffic load dynamics; and hence, evaluating the performance of the proposed algorithms. Moreover, we show the performance evaluation of the proposed approach in terms of overall transmission energy that well-captures both the transmit power and average transmission rate.

Fig. (2a) shows the impact of the outage threshold $\epsilon$ for different system bandwidth configuration on the overall UAV transmission energy. In this figure, decreasing $\epsilon$ leads to high URLLC reliability. We compare the performance of the proposed approach with an intuitive baseline Maximum power, which is the worst case scenario to satisfy reliability constraints. For a given system bandwidth, the UAV increases the transmit power, so higher transmission energy improves the average sum-rate required for obtaining higher reliability (i.e., smaller $\epsilon$); however, a performance gain of up to 25% as compared with the baseline. On the other hand, we observe the UAV significantly lowers the transmit power when the available system bandwidth is high; thus, a low transmission energy. This is expected as a lower transmit power is sufficient enough to maximize the average sum-rate for URLLC users that guarantees its reliability requirements. Thus, we observe the tradeoff between the overall transmission power and the average sum-rate, as defined in (6).

3https://www.kaggle.com/szrlee/stock-time-series-20050101-to-20171231.
Fig. 2: The tradeoff between UAV transmission energy and the system bandwidth varying (a) the outage threshold $\epsilon$, and (b) the network density $U$. Fig. (2c) is the performance of URLLC traffic load prediction model for $N = 1000$.

Fig. (2b) demonstrates the impact of network density, i.e., the number of URLLC users, on the UAV transmission energy for $\epsilon = 0.1$. For a given available system bandwidth, it is shown that the UAV requires higher transmission energy to satisfy the stringent requirements of a large number of users. Moreover, with the increase in the available bandwidth, UAV can reduce its transmit energy consumption without compromising the achievable average sum-rate.

Finally, Fig. 2c captures the real trend of URLLC traffic load using a GPR prediction approach. It is observed that the prediction accuracy surpassed 99% when trained 1,000 data samples over the window of 1,000 time slots. Consequently, a better margin of performance gain is obtained while performing RB and transmit power optimization to satisfy the URLLC requirements.

V. CONCLUSIONS

In this letter, we have studied the problems of practical integration of 5G features with resource-constrained UAV to deliver URLLC services in an out-of-coverage area. In doing so, firstly, we have exploited a GPR approach to capture the real trend of URLLC traffic manipulating real word datasets. We have then formulated a joint optimization problem for maximizing the average sum-rate while minimizing the transmit power of UAV with the constraints to satisfy stringent URLLC requirements. We have revealed the formulated problem as an MINLP, challenging to solve directly using conventional optimization techniques. To tackle this issue, we have introduced the low-complexity near-optimal successive minimization algorithm. Finally,
we have presented numerical results to validate the efficiency of our proposed solution approach where our approach outperforms the other baselines.

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, “A tutorial on uavs for wireless networks: Applications, challenges, and open problems,” IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2334–2360, March 2019.

[2] Y. K. Tun, Y. M. Park, N. H. Tran, W. Saad, S. R. Pandey, and C. S. Hong, “Energy-efficient resource management in uav-assisted mobile edge computing,” IEEE Communications Letters (Early Access), Sep. 2020.

[3] UL Inter-UE Multiplexing Between eMBB and URLLC, document 3GPP TSG RAN WG1 95, Nov. 2018.

[4] A. Anand, G. De Veciana, and S. Shakkottai, “Joint scheduling of urllc and embb traffic in 5g wireless networks,” IEEE/ACM Transactions on Networking, vol. 28, no. 2, pp. 477–490, Feb. 2020.

[5] M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and low-latency wireless communication: Tail, risk, and scale,” Proceedings of the IEEE, vol. 106, no. 10, pp. 1834–1853, Sep. 2018.

[6] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, “embb-urlle resource slicing: A risk-sensitive approach,” IEEE Communications Letters, vol. 23, no. 4, pp. 740–743, Feb. 2019.

[7] Y. Huang, S. Li, C. Li, Y. T. Hou, and W. Lou, “A deep reinforcement learning-based approach to dynamic embb/urlle multiplexing in 5g nr,” IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6439–6456, Mar. 2020.

[8] A. T. Z. Kasgari and W. Saad, “Model-free ultra reliable low latency communication (urllc): A deep reinforcement learning framework,” in IEEE International Conference on Communications (ICC), Shanghai, China, May 2019.

[9] W. Khawaja, I. Guvenc, D. W. Matolak, U.-C. Fiebig, and N. Schneckenburger, “A survey of air-to-ground propagation channel modeling for unmanned aerial vehciles,” IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2361–2391, May 2019.

[10] C. K. Williams, “Prediction with gaussian processes: From linear regression to linear prediction and beyond,” in Learning in graphical models. Springer, 1998, pp. 599–621.