Resource Allocation for Secure URLLC in Mission-Critical IoT Scenarios

Hong Ren, *Member, IEEE*, Cunhua Pan, *Member, IEEE*, Yansha Deng, *Member, IEEE*, Maged Elkashlan, *Member, IEEE*, and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*— Ultra-reliable low latency communication (URLLC) is one of three primary use cases in the fifth-generation (5G) networks, and its research is still in its infancy due to its stringent and conflicting requirements in terms of extremely high reliability and low latency. To reduce latency, the channel blocklength for packet transmission is finite, which incurs transmission rate degradation and higher decoding error probability. In this case, conventional resource allocation based on Shannon capacity achieved with infinite blocklength codes is not optimal. Security is another critical issue in mission-critical internet of things (IoT) communications, and physical-layer security is a promising technique that can ensure the confidentiality for wireless communications as no additional channel uses are needed for the key exchange as in the conventional upper-layer cryptography method. This paper is the first work to study the resource allocation for a secure mission-critical IoT communication system with URLLC. Specifically, we adopt the security capacity formula under finite blocklength and consider two optimization problems: weighted throughput maximization problem and total transmit power minimization problem. Each optimization problem is non-convex and challenging to solve, and we develop efficient methods to solve each optimization problem. Simulation results confirm the fast convergence speed of our proposed algorithm and demonstrate the performance advantages over the existing benchmark algorithms.**

***Index Terms*— URLLC, secure communications, short packet transmission, mission-critical applications, industrial 4.0.**

I. Introduction

THE fifth-generation (5G) networks are expected to support three main use cases: enhanced mobile broadband (eMBB), massive machine type communication (mMTC), and ultra-reliable low latency communication (URLLC) [1]. Significant advancement has been achieved in the last decade for the use case of eMBB characterized by high throughput and data rate. Some typical techniques include massive multiple-input multiple-output (MIMO) and mmWave communications. For the use case of mMTC, 5G networks aim to provide massive connectivity to tens of billions of low-cost small-size machine-type devices such as smart glasses, smart thermometers, wireless sensors, etc. Some access protocols such as random access and grant-free access are shown to be effective in mMTC. However, the realization of URLLC is more challenging than eMBB and mMTC due to the fact that URLLC targets at two stringent quality-of-service (QoS) requirements in term of extremely high reliability (e.g., $1-10^{-9}$) and ultra-low latency (e.g., 1 ms), which are conflicting with each other. Specifically, to achieve high reliability, long codeword with redundancy is required, which increases the latency. On the other hand, short packet/codeword is mandated to achieve low latency, which lowers the reliability performance. The research of URLLC is still in its infancy and is main target of Release 17. In addition, URLLC is closely relevant to mission-critical internet of things (IoT) applications with emphasis on high reliability and low latency, such as autonomous factory manufacture, remote surgery, unmanned aerial vehicles (UAVs) control, and vehicular communication networks.

The primary feature associated with URLLC compared with conventional human-to-human communications is its short packet transmission feature, which is adopted to guarantee ultra-low latency. In this case, the law of large numbers is not valid and Shannon capacity cannot be applied to characterize the system capacity. Knowing that short blocklength is adopted in URLLC, the decoding error probability will not approach zero even when signal-to-noise ratio (SNR) is arbitrarily high. If Shannon capacity expression is directly applied for transmission design, the reliability and latency will be underestimated, and the QoS cannot be guaranteed. In [2], the authors first derived the approximate expression of the data rate for a point-to-point AWGN channel under the case of finite channel blocklength, which is a function of the SNR, channel blocklength, and decoding error probability. Recently, this information-theoretical result has been adopted to design the resource allocation in various communication systems [3]–[9]. Specifically, a wireless-powered IoT network with short packet communication was studied in [3], where the transmission time and packet error rate of each device was jointly optimized to maximize the effective-throughput that accounts for both the transmission efficiency and reliability performance. Joint optimization of the channel blocklength and UAV location was investigated in [4] to minimize the decoding error probability. The authors in [5] derived the average achievable data rate

Manuscript received February 17, 2020; revised April 25, 2020 and May 29, 2020; accepted May 29, 2020. Date of publication June 3, 2020; date of current version September 16, 2020. This work was supported by the U.K. Engineering and the Physical Sciences Research Council under Grant EP/N029666/1. The associate editor coordinating the review of this article and approving it for publication was D. B. Da Costa. *(Corresponding author: Cunhua Pan.)*

Hong Ren, Cunhua Pan, Maged Elkashlan, and Arumugam Nallanathan are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: h.ren@qmul.ac.uk; c.pan@qmul.ac.uk; maged.elkashlan@qmul.ac.uk; a.nallanathan@qmul.ac.uk).

Yansha Deng is with the Department of Informatics, King's College London, London WC2R 2LS, U.K. (e-mail: yansha.deng@kcl.ac.uk).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.2999628

from the control information delivery from the ground control station to UAV under a three-dimensional channel model. In [6], the resource allocation for non-orthogonal multiple access (NOMA) short-packet communications was studied for a simple two-user scenario. A cross-layer framework was established in [7] for short packet transmission in mobile edge computing IoT networks. A resource allocation problem for a factory automation scenario was optimized in [8] to minimize the decoding error probability for four transmission schemes, e.g., orthogonal multiple access, NOMA, relay-assisted, and cooperative NOMA schemes. In [9], the authors proposed to adopt the compelling massive MIMO technique to support the transmission for massive amount of devices in industrial IoT networks, in which channel hardening effect can be exploited to reduce the computational and operational latency.

On the other hand, due to the broadcast nature of wireless communications, IoT applications such as industrial robots are particularly vulnerable to security threats (e.g. critical control information leakage or malicious attack). Conventionally, the security is enhanced through cryptography at the upper layers of the communication system. However, the secret key exchange and management is complicated and needs additional channel uses to accomplish these protocols. In URLLC, channel blocklength is limited, and the cryptography method may not be applicable in URLLC applications. On the other hand, physical layer security, which exploits the nature of wireless channels, is more favourable for URLLC as the complicated key exchange procedure is unnecessary. Recently, physical-layer security has been extensively studied in the existing literature [10]–[14]. In particular, a comprehensive review of physical-layer security was conducted in [10], which described secure transmission strategies for various transmission systems. The authors in [11] presented a review on the physical-layer security for the machine-type communications (MTC), where the MTC devices have limited hardware and limited energy storage. A novel unmanned aerial vehicle (UAV) flight trajectory optimization was studied in [12] to maximize the minimum secrecy rate of ground terminals. In [13], a secure mobile edge computing (MEC) system was proposed where ground users can offload computing tasks to the legitimate UAV in the presence of multiple eavesdropping UAVs. A novel secure transmission scheme was developed in [14] to fight against the eavesdropping in downlink multi-input single-output non-orthogonal multiple access (NOMA) networks. However, infinite blocklength is assumed in these papers, and the security capacity is defined as the highest coding rate that there always exists a pair of channel encoder and decoder such that both the decoding error probability at the legitimate receiver and the information leakage to the eavesdropper can be made arbitrarily small when the channel blocklength is sufficiently large.

Unfortunately, the security capacity formula based on the infinite blocklength assumption is not applicable for secure URLLC applications, where short channel code/blocklength is adopted to reduce the latency. There are only a limited number of contributions studying the secrecy rate under finite blocklength [15], [16]. Most recently, the approximate security capacity formula under finite blocklength has been derived in [15], which is more complicated than the simple point-to-point communication system in [2]. Based on this information-theoretical result, the authors in [16] analyzed the performance of secure short-packet communications in a mission-critical IoT system with an eavesdropper. However, the resource allocation based on this result has not yet been studied.

Against the above background, the resource allocation problem for a secure mission-critical IoT system under short packet communications is studied in this paper. Specifically, the contributions are summarized as follows:

1) We first consider the weighted sum throughput (WST) maximization problem by jointly optimizing the bandwidth unit and power allocation, while guaranteeing the total power and bandwidth constraints. This optimization problem is challenging to solve due to the following reasons. First, this problem involves the discrete variables associated with the number of bandwidth unit allocation. Second, the optimization variables are coupled in the objective function. Hence, this problem is a non-convex mixed-integer programming. To handle this problem, we develop an efficient iterative algorithm based on the principles of block coordinate descent (BCD) along with the successive convex approximation (SCA) method to solve this problem. Both the convergence and complexity analysis are provided. Greedy search method is adopted to convert the continuous variables into discrete ones.
2) We then jointly optimize the power and channel bandwidth unit allocation to minimize the total transmit power (TTP) for a mission-critical IoT system under short packet communications, while guaranteeing the minimum security capacity of each device and the total channel blocklength. The optimization problem is a non-convex and mixed-integer programming problem, and NP-hard to solve. We first express the power for each device as a function of channel blocklength, and relax the discrete constraint for the channel blocklength to continuous variables. Then, a sufficient condition is proposed when the channel blocklength allocation problem is a convex optimization problem. This condition holds for typical URLLC application scenarios. At last, greedy method is used to convert the continuous solutions to discrete solutions.
3) Finally, simulation results confirm the performance advantage of the proposed algorithm over the benchmark solutions such as the conventional long packet transmission scheme, which verifies the importance of adopting the security capacity formula under finite channel blocklength in the system design.

The rest of this paper is organized as follows. System model and problem formulation are provided in Section II. Weighted sum throughput maximization problem is solved in Section III, while total transmit power minimization problem is considered in Section IV. Simulation results along with related discussions are shown in Section V. Finally, conclusions of this paper are drawn in Section VI.

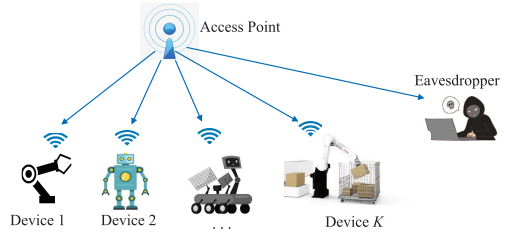

Fig. 1. Illustration of a secure mission-critical IoT communication system.

II. System Model and Problem Formulation

A. System Model

We consider a downlink mission-critical IoT communication system as depicted in Fig. 1, in which an access point (AP) transmits confidential control signals to K wireless connected devices (e.g. actuators, robots, and automated guided vehicle (AGV)). Meanwhile, there is an eavesdropper that aims to intercept the critical control signals transmitted by the AP. The AP, K devices, and the eavesdropper are assumed to be equipped with a single antenna. Due to the low-latency transmission, it is not feasible to allocate different time slots to all devices. Instead, we assume that all the devices are allocated with orthogonal frequency bands while transmitting over the same time duration, denoted as T.

In practical systems, the frequency band is divided into multiple basic bandwidth units with bandwidth B_0. Each device is assumed to operate in different frequency bands and the total frequency bandwidth allocated to the kth device is denoted as $B_k = n_k B_0$, where n_k denotes the number of bandwidth units allocated to the kth device. We assume that the total bandwidth allocated to all the devices should be no larger than channel coherence bandwidth W_c. It is assumed that W_c is divisible by B_0, which can be expressed as $W_c = n_{\max} B_0$. Therefore, we have

$$\sum\nolimits_{k=1}^{K} n_k \le n_{\max}. \tag{1}$$

Then, the number of channel uses allocated for the kth device is given by $B_k T$. In URLLC, the transmission duration T is extremely small, which is shorter than the channel coherence time. Hence, the channels from the AP to K devices and the eavesdropper stay constant over each transmission. The channels from the AP to the kth device and the eavesdropper are denoted as $h_k^d \in \mathbb{C}$ and $h^e \in \mathbb{C}$, respectively. The received signal to noise ratio (SNR) of the kth device is given by

$$\gamma_k^d = \frac{p_k g_k^d}{n_k}, \tag{2}$$

where p_k is the transmit power for the kth device, and $g_k^d = |h_k^d|^2/(\sigma_{d,k}^2 B_0)$ with $\sigma_{d,k}^2$ denoting the noise power spectrum density at the kth device. It is assumed that the eavesdropper can access all the frequency bands occupied by the devices. Thus, when the eavesdropper attempts to eavesdrop the kth device's information, the received SNR at the eavesdropper is given by

$$\gamma_k^e = \frac{p_k g^e}{n_k}, \tag{3}$$

where $g^e = |h^e|^2/\left(\sigma_e^2 B_0\right)$ and σ_e^2 is the noise power spectrum density at the eavesdropper. In this paper, we assume all the channel state information (CSI) is available at the transmitter. This is a strict assumption, and the solutions in this paper can serve as the performance benchmark for the existing more practical case when only imperfect or partial CSI is available at the transmitter.

B. Achievable Secrecy Data Rate Under Finite Blocklength

It is well-known that when the number of channel uses is sufficiently large and the transmission data rate is lower than the secrecy capacity, we can always find a channel coding scheme such that both the decoding error probability and information leakage can be made as small as possible. In URLLC, the transmission blocklength (or the number of channel uses) is finite to guarantee low latency. However, short blocklength transmission suffers from a non-zero decoding error probability and non-negligible information leakage.

Based on [15], for a given channel blocklength $N_k = B_k T$, to guarantee a maximum decoding error probability of ϵ_k at the kth device, and a secrecy constraint on the information leakage of δ_k, a lower bound on maximum secrecy communication rate (bit/s/Hz) can be approximated by:

$$r_k = C_k - \sqrt{\frac{V_k^d}{N_k}}\frac{Q^{-1}(\epsilon_k)}{\ln 2} - \sqrt{\frac{V_k^e}{N_k}}\frac{Q^{-1}(\delta_k)}{\ln 2}, \tag{4}$$

where $C_k = \log_2(1+\gamma_k^d) - \log_2(1+\gamma_k^e)$ denotes the maximum secrecy capacity that can be achieved under infinite channel blocklength, $V_k^x = 1-(1+\gamma_k^x)^{-2}$, $x \in \{d, e\}$ is the channel dispersion which characterizes the random variability of a channel with respect to a deterministic channel with the same capacity [17], and $Q^{-1}(\cdot)$ is the inverse of the Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. A necessary condition to ensure a positive data rate is that $\gamma_k^d > \gamma_k^e$ [16], which is equivalent to $g_k^d > g^e$ according to (2) and (3). The total number of bits (or throughput) that can be transmitted for each transmission for the kth device is given by

$$R_k = n_k B_0 T \left(C_k - \sqrt{\frac{V_k^d}{n_k B_0 T}}\frac{Q^{-1}(\epsilon_k)}{\ln 2} - \sqrt{\frac{V_k^e}{n_k B_0 T}}\frac{Q^{-1}(\delta_k)}{\ln 2} \right). \tag{5}$$

In the following two sections, we aim to jointly optimize the number of bandwidth units and the power allocation to maximize the weighted sum throughput (WST) and minimize the total transmit power (TTP), respectively.

III. Weighted Sum Throughput Maximization

In this section, we aim to maximize the WST of all devices through jointly optimizing the number of bandwidth units and the power allocation. Specifically, we first provide the problem formulation. Then, one efficient algorithm is proposed to solve the optimization problem.

A. Problem Formulation

For the case when the AP places more emphasis on the amount of information transmitted, we aim to jointly optimize the power allocation and the bandwidth unit allocation to maximize the WST of all devices while guaranteeing the total power constraint at the AP, and the total number of available bandwidth units. Thus, the optimization problem can be formulated as follows:

$$
\begin{aligned}
(\mathbf{P1}):\ \max_{\boldsymbol{p},\boldsymbol{n}}\ & \sum\nolimits_{k=1}^{K}\omega_k R_k && \text{(6a)}\\
\text{s.t. }& \sum\nolimits_{k=1}^{K} p_k \le P_{\max}, && \text{(6b)}\\
& \sum\nolimits_{k=1}^{K} n_k \le n_{\max}, && \text{(6c)}\\
& n_k \in \mathbb{N}^+, \quad \forall k=1,\cdots,K, && \text{(6d)}\\
& p_k \ge 0, \quad \forall k=1,\cdots,K, && \text{(6e)}
\end{aligned}
$$

where $\boldsymbol{p}=\{p_1,\cdots,p_K\}$, $\boldsymbol{n}=\{n_1,\cdots,n_K\}$, ω_k is a positive weight factor used to ensure the fairness among the devices and $\mathbb{N}^+$ denotes the non-negative integer set. Inequalities (6b) and (6c) correspond to the total power constraint and total bandwidth constraint, respectively. Constraint (6d) means that the integer constraint for the number of bandwidth units. Note that $R_k=0$ when $p_k=0$, which ensures the non-negative value of R_k in the optimal solution.

Problem $(\mathbf{P1})$ is a mixed integer programming problem due to the non-negative integer constraints on $\boldsymbol{n}$. To make it tractable, we relax the integer $\boldsymbol{n}$ to continuous variables, and then convert the continuous solutions into integer ones. Therefore, Problem $(\mathbf{P1})$ is relaxed as follows:

$$
\begin{aligned}
(\mathbf{P2}):\ \max_{\boldsymbol{p},\boldsymbol{n}}\ & \sum\nolimits_{k=1}^{K}\omega_k R_k && \text{(7a)}\\
\text{s.t. }& (6b),(6c),(6e), && \text{(7b)}\\
& n_k \ge 0, \quad \forall k=1,\cdots,K. && \text{(7c)}
\end{aligned}
$$

However, Problem **(P2)** is still difficult to solve since $\boldsymbol{p}$ and $\boldsymbol{n}$ are coupled together. To circumvent this difficulty, we adopt the block coordinate descent (BCD) to decouple these optimization variables. In particular, we optimize one set of variables while keeping the others fixed, and vice versa. Then, each subproblem is solved in an iterative manner. Specifically, Problem **(P2)** is decoupled into two subproblems as

$$
\begin{aligned}
(\mathbf{P2-1}):\ \max_{\boldsymbol{p}}\ & \sum\nolimits_{k=1}^{K}\omega_k R_k(p_k) && \text{s.t. } (6b),(6e)\\
(\mathbf{P2-2}):\ \max_{\boldsymbol{n}}\ & \sum\nolimits_{k=1}^{K}\omega_k R_k(n_k) && \text{s.t. } (6c),(7c)
\end{aligned}
$$

where Problem $(\mathbf{P2-1})$ corresponds to the optimization of power allocation $\boldsymbol{p}$ with a given number of bandwidth units $\boldsymbol{n}$, while Problem $(\mathbf{P2-2})$ is the optimization of the number of bandwidth units $\boldsymbol{n}$ with given $\boldsymbol{p}$. Each subproblem will be solved in the following subsections.

B. The Solution of (P2-1)

In this subsection, we aim to solve the power allocation of Problem $(\mathbf{P2-1})$ with given $\boldsymbol{n}$. To this end, we first define $\bar{g}_k^d \triangleq \frac{g_k^d}{n_k}$, $\bar{g}_k^e \triangleq \frac{g^e}{n_k}$, $L_k^d = \frac{Q^{-1}(\epsilon_k)\sqrt{N_k}}{\ln 2}$, and $L_k^e = \frac{Q^{-1}(\delta_k)\sqrt{N_k}}{\ln 2}$. Then, $R_k(p_k)$ can be rewritten as

$$
\begin{aligned}
R_k(p_k) = & \underbrace{N_k \log_2\left(1+p_k\bar{g}_k^d\right) - N_k\log_2\left(1+p_k\bar{g}_k^e\right)}_{f_k(p_k)}\\
& \underbrace{-\left(\sqrt{V_k^d}L_k^d+\sqrt{V_k^e}L_k^e\right)}_{y_k(p_k)}. \qquad \text{(9)}
\end{aligned}
$$

Since $g_k^d > g^e$, we have $\bar{g}_k^d > \bar{g}_k^e$. Before solving Problem $(\mathbf{P2-1})$, we first analyze the concave-convex property of $R_k(p_k)$ with respect to (w.r.t.) p_k. To this end, we first show that both functions $f_k(p_k)$ and $y_k(p_k)$ are concave w.r.t. p_k as proved in the following lemma.

Lemma 1: $f_k(p_k)$ and $y_k(p_k)$ are concave w.r.t. p_k, and thus $R_k(p_k)$ is the difference of two concave functions $f_k(p_k)$ and $y_k(p_k)$.

Proof: Please refer to Appendix A. ■

According to Lemma 1, both $f_k(p_k)$ and $y_k(p_k)$ are concave w.r.t. p_k. Hence, for given $\boldsymbol{n}$, the objective function of Problem $(\mathbf{P2-1})$ is a difference of two concave functions, and thus is a non-concave function w.r.t. p_k. As a reslut, Problem $(\mathbf{P2-1})$ is a non-convex optimization problem, and the globally optimal solution is difficult to find. However, Problem $(\mathbf{P2-1})$ belongs to a class of difference of convex (DC) problems [18], where the objective is to maximize a difference of two concave functions. This type of optimization problem can be efficiently solved by using the successive convex approximation (SCA) method, which solves the DC problem in an iterative manner.

Denote the solution of $\boldsymbol{p}$ in the $(i-1)$-th iteration as $\boldsymbol{p}^{(i-1)}$. By exploiting the concavity of $y_k(p_k)$ and Jensen's inequality, we have

$$
y_k(p_k) \le y_k\left(p_k^{(i-1)}\right) + \beta_k(p_k^{(i-1)})\left(p_k - p_k^{(i-1)}\right), \quad \text{(10)}
$$

where $\beta_k(p_k^{(i-1)})$ is the first-order derivative of $y(p_k)$ at $p_k^{(i-1)}$, and is given by

$$
\begin{aligned}
&\beta_k(p_k^{(i-1)})\\
&= \frac{\left(1+p_k^{(i-1)}\bar{g}_k^d\right)^{-3}\bar{g}_k^d L_k^d}{\left(1-\left(1+p_k^{(i-1)}\bar{g}_k^d\right)^{-2}\right)^{\frac{1}{2}}} + \frac{\left(1+p_k^{(i-1)}\bar{g}_k^e\right)^{-3}\bar{g}_k^e L_k^e}{\left(1-\left(1+p_k^{(i-1)}\bar{g}_k^e\right)^{-2}\right)^{\frac{1}{2}}},\\
&> 0. \qquad \text{(11)}
\end{aligned}
$$

By replacing $y_k(p_k)$ with the right hand side (RHS) of (11), we obtain the optimization problem to be solved in the ith iteration, which is given by:

$$
\begin{aligned}
(\mathbf{P2-1-a}):\ \max_{\boldsymbol{p}}\ & \sum_{k=1}^{K}\left(\omega_k f_k(p_k) - \omega_k\beta_k(p_k^{(i-1)})p_k\right) && \text{(12a)}\\
\text{s.t. }& (6b),(6e), && \text{(12b)}
\end{aligned}
$$

where the constant values have been omitted in the objective function. Note that the objective function of Problem $(\mathbf{P2-1-a})$ is a concave function and its constraints are affine functions of $\boldsymbol{p}$. Then, Problem $(\mathbf{P2-1-a})$ is a

convex optimization problem. The optimal solution of Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ can be found in the following theorem.

Theorem 1: The optimal solution of Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ is given by

$$p_k^\star(\lambda) = \left[\frac{-\left(\bar{g}_k^d+\bar{g}_k^e\right)+\sqrt{\left(\bar{g}_k^d+\bar{g}_k^e\right)^2-4\bar{g}_k^d\bar{g}_k^e\left(1-\eta_k^{(i-1)}(\lambda)\right)}}{2\bar{g}_k^d\bar{g}_k^e}\right]^+, \ \forall k \tag{13}$$

where $[x]^+$ is equal to $\max\{x,0\}$ and $\eta_k^{(i-1)}(\lambda)$ is given by

$$\eta_k^{(i-1)}(\lambda) = \frac{N_k}{\ln 2}\frac{\omega_k(\bar{g}_k^d-\bar{g}_k^e)}{\omega_k\beta_k(p_k^{(i-1)})+\lambda}, \forall k. \tag{14}$$

If $\sum_{k=1}^K p_k^\star(0) \le P_{\max}$, then $\lambda = 0$. Otherwise, λ is the root of the following equation:

$$\sum\nolimits_{k=1}^K p_k^\star(\lambda) - P_{\max} = 0. \tag{15}$$

Proof: Please refer to Appendix B. ∎

If $\sum_{k=1}^K p_k^\star(0) > P_{\max}$, we need to find a λ to satisfy Equation (15). For the case of $p_k^\star(\lambda) > 0$, by taking the first-order derivative of $p_k^\star(\lambda)$ w.r.t. λ, we have $\frac{\partial p_k^\star(\lambda)}{\partial\lambda} < 0$.

C. The Solution of (P2-2)

In this subsection, our aim is to solve Problem (P2-2) by optimizing the number of bandwidth units with given power allocation. For simplicity, we first define $\tilde{N}_0 = B_0T, \tilde{g}_k^d = p_kg_k^d, \tilde{g}_k^e = p_kg^e$, $\tilde{L}_k^d = \sqrt{\tilde{N}_0}\frac{Q^{-1}(\epsilon_k)}{\ln 2}$, and $\tilde{L}_k^e = \sqrt{\tilde{N}_0}\frac{Q^{-1}(\sigma_k)}{\ln 2}$. Since $g_k^d > g^e$, we have $\tilde{g}_k^d > \tilde{g}_k^e$. Then, $R_k(n_k)$ can be rewritten as

$$R_k(n_k) = \underbrace{\tilde{N}_0n_k\log_2\left(1+\frac{\tilde{g}_k^d}{n_k}\right)-\tilde{N}_0n_k\log_2\left(1+\frac{\tilde{g}_k^e}{n_k}\right)}_{F_k(n_k)} - \underbrace{\left(\sqrt{z_k^d(n_k)}\tilde{L}_k^d+\sqrt{z_k^e(n_k)}\tilde{L}_k^e\right)}_{G_k(n_k)}, \tag{16}$$

where $z_k^x(n_k) = n_k - \frac{n_k^3}{(n_k+\bar{g}_k^x)^2}$, $x \in \{d,e\}$. Before solving Problem (P2-2), we first analyze the concave-convex property of $R_k(n_k)$. In particular, the following lemma shows that $F_k(n_k)$ and $G_k(n_k)$ are concave functions w.r.t. n_k.

Lemma 2: $F_k(n_k)$ and $G_k(n_k)$ are concave w.r.t. n_k, and thus $R_k(n_k)$ is the difference of two concave functions $F_k(n_k)$ and $G_k(n_k)$.

Proof: Please refer to Appendix C. ∎

Then, similar to the optimization of power allocation, we adopt the SCA method to solve Problem (P2-2). By denoting the solution $\boldsymbol{n}$ in the $(j-1)$-th iteration as $\boldsymbol{n}^{(j-1)}$ and using Lemma 2 and Jensen's inequality, we have

$$G_k(n_k) \le G_k(n_k^{(j-1)}) + \alpha_k(n_k^{(j-1)})\left(n_k - n_k^{(j-1)}\right), \tag{17}$$

where $\alpha_k(n_k^{(j-1)})$ is the first-order derivative of $G_k(n_k)$ w.r.t. n_k at $n_k = n_k^{(j-1)}$ and is given by

$$\alpha_k(n_k^{(j-1)}) = \frac{\tilde{L}_k^d\left(3\left(\tilde{g}_k^d\right)^2n_k^{(j-1)}+\left(\tilde{g}_k^d\right)^3\right)}{2\sqrt{z_k^d(n_k^{(j-1)})}\left(n_k^{(j-1)}+\tilde{g}_k^d\right)^3} + \frac{\tilde{L}_k^e\left(3(\tilde{g}_k^e)^2n_k^{(j-1)}+(\tilde{g}_k^e)^3\right)}{2\sqrt{z_k^e(n_k^{(j-1)})}\left(n_k^{(j-1)}+\tilde{g}_k^e\right)^3}. \tag{18}$$

By replacing $\alpha_k(n_k^{(j-1)})$ with the RHS of (18), the subproblem to be solved in the jth iteration is given by

$$(\mathbf{P2}-\mathbf{2}-\mathbf{a}): \max_{\boldsymbol{n}} \sum_{k=1}^K\left(\omega_kF_k(n_k)-\omega_k\alpha_k(n_k^{(j-1)})n_k\right) \tag{19a}$$

$$\text{s.t. } (6c),(7c), \tag{19b}$$

where the constant terms are omitted in the objective function.

Note that the objective function of Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$ is a concave function of n_k and the constraints are affine functions. Hence, Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$ is a convex optimization problem. In the following, we provide a low-complexity algorithm to obtain the globally optimal solution by using the Lagrangian dual decomposition method [19]. Since Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$ is a convex optimization problem and the slater's condition is satisfied,[1] the dual gap is zero and the original problem can be solved by solving its dual problem. Specifically, we introduce the non-negative Lagrange multiplier $\mu \ge 0$ corresponding to the constraint of the total number of bandwidth units, and the partial Lagrange function of Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$ is given by

$$\mathcal{L}(\boldsymbol{n},\mu) = \sum\nolimits_{k=1}^K\left(\omega_kF_k(n_k)-\omega_k\alpha_k(n_k^{(j-1)})n_k\right) -\mu\left(\sum\nolimits_{k=1}^K n_k - n_{\max}\right). \tag{20}$$

The dual function can be obtained by solving the following optimization problem:

$$Y(\mu) \triangleq \max_{n_k\ge 0,\forall k}\sum\nolimits_{k=1}^K \mathcal{L}_k(n_k,\mu)+\mu n_{\max}, \tag{21}$$

where $\mathcal{L}_k(n_k,\mu)$ is given by

$$\mathcal{L}_k(n_k,\mu) = \omega_kF_k(n_k)-\omega_k\alpha_k(n_k^{(j-1)})n_k-\mu n_k. \tag{22}$$

Then, the dual problem is given by

$$\min_{\mu\ge 0} Y(\mu). \tag{23}$$

To solve the dual problem (23), we need to first obtain the expression of dual function $Y(\mu)$, which needs to solve Problem (21) with given μ. For given μ, Problem (21) can be decoupled into K independent optimization problems, and the optimization problem associated with the kth device is given by

$$\max_{n_k\ge 0}\ \mathcal{L}_k(n_k,\mu)=\omega_kF_k(n_k)-\omega_k\alpha_k(n_k^{(j-1)})n_k-\mu n_k. \tag{24}$$

[1]There exist strictly feasible $\boldsymbol{n}$ such as $\sum_{k=1}^K n_k < n_{\max}$.

The first-order derivative of $\mathcal{L}_k(n_k,\mu)$ w.r.t. n_k is derived as

$$\frac{\partial \mathcal{L}_k(n_k,\mu)}{\partial n_k} = \omega_k \tilde{N}_0 \log_2\left(\frac{n_k+\tilde{g}_k^d}{n_k+\tilde{g}_k^e}\right) + \frac{\omega_k \tilde{N}_0\left(\tilde{g}_k^e-\tilde{g}_k^d\right) n_k}{\ln 2\left(n_k+\tilde{g}_k^d\right)\left(n_k+\tilde{g}_k^e\right)} - \omega_k \alpha_k\left(n_k^{(j-1)}\right) - \mu. \quad (25)$$

Since $\mathcal{L}_k(n_k,\mu)$ is a concave function, the optimal solution of Problem (24) can be derived as follows:

- If $\frac{\partial \mathcal{L}_k(n_k,\mu)}{\partial n_k}\Big|_{n_k=0} \leq 0$, the optimal n_k for given μ is given by $n_k^\star(\mu)=0$;
- If $\frac{\partial \mathcal{L}_k(n_k,\mu)}{\partial n_k}\Big|_{n_k=0} > 0$, the optimal n_k should satisfy the equation $\frac{\partial \mathcal{L}_k(n_k,\mu)}{\partial n_k} = 0$, and its root is denoted as $n_k^\star(\mu)$. As $\mathcal{L}_k(n_k,\mu)$ is a concave function, $\frac{\partial \mathcal{L}_k(n_k,\mu)}{\partial n_k}$ is a decreasing function w.r.t. n_k. Thus, $n_k^\star(\mu)$ can be obtained by the bisection search method when μ is given.

Next, we turn to solve the dual problem by finding the optimal μ. The optimal value of μ should satisfy the complementary slackness condition for the constraint (6c):

$$\mu\left(\sum\nolimits_{k=1}^{K} n_k^\star(\mu) - n_{\max}\right) = 0. \quad (26)$$

From (26), if $\sum_{k=1}^K n_k^\star(0) \leq n_{\max}$ holds, the optimal solution is given by $n_k(0), \forall k$; Otherwise, the optimal μ should satisfy $W(\mu) = \sum_{k=1}^K n_k^\star(\mu) = n_{\max}$. In contrast to the power allocation solution in (13), the bandwidth unit allocation $n_k(\mu)$ cannot be expressed in an explicit function of μ. As a result, the monotonicity of $n_k(\mu)$ w.r.t. μ cannot be proved by checking the sign of the first-order derivative. To deal with this issue, we have the following lemma:

Lemma 3: $W(\mu)$ is a monotonically decreasing function w.r.t. μ.

Proof: Please refer to Appendix D. ∎

Based on Lemma 3, the optimal μ can be obtained by using the bisection search method.

D. Algorithm Analysis

1) Algorithm Description: Based on the above analysis, we summarize the proposed BCD algorithm in Algorithm 1, where $R(\boldsymbol{n},\boldsymbol{p})$ is the weighted throughput defined as $R(\boldsymbol{n},\boldsymbol{p}) = \sum_{k=1}^K \omega_k R_k(\boldsymbol{n},\boldsymbol{p})$. This algorithm is a two-layer iterative algorithm, where the inner layer is the iteration of the SCA algorithm to solve Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ and Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$, and the outer layer is the BCD algorithm to solve Problem **(P2)**. In Line 7 of Algorithm 1, $\boldsymbol{p}^\star$ denotes the optimal solution obtained by the inner layer to solve Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$, and $\boldsymbol{n}^\star$ in Line 12 corresponds to the inner layer to solve Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$.

2) Convergence Analysis: In the tth outer iteration, the SCA algorithm is adopted to solve Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ to find the power allocation solution. Based on the property of the SCA algorithm [20], the SCA algorithm is guaranteed to converge. Then, we have $R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t)}\right) \geq R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t-1)}\right)$. Afterwards, the SCA algorithm is used to find the channel bandwidth unit solution. We then have $R\left(\boldsymbol{n}^{(t)},\boldsymbol{p}^{(t)}\right) \geq R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t)}\right)$. Hence, we have $R\left(\boldsymbol{n}^{(t)},\boldsymbol{p}^{(t)}\right) \geq R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t-1)}\right)$, which shows the solutions obtained by the BCD algorithm are monotonically increasing. In addition, due to the power and total bandwidth limits, there exists an upper bound on the total throughput. Hence, the BCD algorithm is guaranteed to converge.

Algorithm 1 BCD Algorithm for Solving Total Throughput Maximization

1 **Initialize** $\boldsymbol{n}=\boldsymbol{n}^{(0)}$, $\boldsymbol{p}=\boldsymbol{p}^{(0)}$, accuracy ε, the iteration number $t=1$ and calculate $R\left(\boldsymbol{n}^{(0)},\boldsymbol{p}^{(0)}\right)$;
2 **repeat**
3 Set $\boldsymbol{n}=\boldsymbol{n}^{(t-1)}$, $i=1$;
4 **repeat**
5 Given $\boldsymbol{p}^{(i-1)}$, calculate $\boldsymbol{p}^{(i)}$ by solving Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$, and $i \leftarrow i+1$;
6 **until** $\boldsymbol{p}$ *converges*;
7 Update $\boldsymbol{p}^{(t)}=\boldsymbol{p}^\star$;
8 Set $\boldsymbol{p}=\boldsymbol{p}^{(t)}$, $j=1$;
9 **repeat**
10 Given $\boldsymbol{n}^{(j-1)}$, calculate $\boldsymbol{n}^{(j)}$ by solving Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$, and $j \leftarrow j+1$;
11 **until** $\boldsymbol{n}$ *converges*;
12 Update $\boldsymbol{n}^{(t)}=\boldsymbol{n}^\star$ and set $t \leftarrow t+1$;
13 **until** $\left|R\left(\boldsymbol{n}^{(t)},\boldsymbol{p}^{(t)}\right)-R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t-1)}\right)\right| / R\left(\boldsymbol{n}^{(t-1)},\boldsymbol{p}^{(t-1)}\right) \leq \varepsilon$;

3) Complexity Analysis: In this part, we analyze the complexity of the BCD algorithm. Note that the main complexity in each outer layer iteration lies in the SCA algorithms to solve Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ and Problem $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$. For each inner layer of the SCA algorithm to solve $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$, the bisection search method is adopted to find λ, and its complexity is $\mathcal{O}\left(K\log_2\left(\frac{1}{\varepsilon}\right)\right)$, where ε is the accuracy. Denote I_{in} as the total number of iterations required for the convergence of the SCA algorithm. The total complexity to solve $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ in each outer layer is given by $\mathcal{O}\left(I_{\text{in}}K\log_2\left(\frac{1}{\varepsilon}\right)\right)$. By using similar analysis, the total complexity to solve $(\mathbf{P2}-\mathbf{2}-\mathbf{a})$ in each outer layer is given by $\mathcal{O}\left(J_{\text{in}}K\log_2\left(\frac{1}{\varepsilon}\right)\right)$, where J_{in} is the total number of iterations required for the convergence of the SCA algorithm. Denote the total number of iterations for the BCD algorithm to converge as N_{BCD}. The total complexity of the BCD algorithm is given by $\mathcal{O}\left(N_{\text{BCD}}\left(I_{\text{in}}+J_{\text{in}}\right)K\log_2\left(\frac{1}{\varepsilon}\right)\right)$. Hence, the BCD algorithm can converge to the locally optimal solution in polynomial time computational complexity.

E. Integer Conversion for $\boldsymbol{n}$

In general, the solution of $\boldsymbol{n}$ obtained from the BCD algorithm are positive continuous values, which may violate the integer constraints. In this part, we provide a greedy search method to convert the continuous solution into integer ones. Specifically, denote the solution of $\boldsymbol{n}$ obtained by the BCD algorithm as $\bar{\boldsymbol{n}} = \{\bar{n}_1,\cdots,\bar{n}_K\}$. The integer conversion problem is a combinatorial optimization problem, and it

requires exponential time complexity to find the globally optimal solution. In the following, we propose a low-complexity algorithm based on the greedy search method to find a suboptimal solution. Firstly, we set the initial value of the solution as $n_k^\star = \lfloor \bar{n}_k \rfloor, \forall k$, where $\lfloor \cdot \rfloor$ denotes the flooring operation. Then, there are $N_{\mathrm{Rem}} = \sum_{k=1}^{K} \bar{n}_k - \sum_{k=1}^{K} n_k^\star$ bandwidth units that are not allocated. The remaining task is to allocate these bandwidth units to the devices. The main idea of the greedy search method is that each time we allocate one bandwidth unit to the device with the highest increment of the total throughput. Denote $\boldsymbol{n}^\star = \{n_1^\star, \cdots, n_K^\star\}$ and $\tilde{\boldsymbol{n}}_k = \{n_1^\star, \cdots, n_k^\star + 1, \cdots, n_K^\star\}$. For each given $\boldsymbol{n}$, we adopt the SCA algorithm to solve Problem $(\mathbf{P2}-\mathbf{1})$, and denote the optimal value of the total throughput as $R(\boldsymbol{n})$. Then, the device index to be allocated one bandwidth unit is given by $k^* = \arg\max_{k\in\mathcal{K}} \{R(\tilde{\boldsymbol{n}}_k) - R(\boldsymbol{n}^\star)\}$, where $\mathcal{K}$ denotes the set of all devices. For the k^*th device, update $n_k^\star = n_k^\star + 1$. Repeat the above procedure until all the remaining bandwidth units are allocated, and the power allocation is updated accordingly based on the final integer solution.

IV. Total Transmit Power Minimization

In this section, each device is assumed to have its minimum throughput requirement, and our goal is to minimize the TTP by jointly optimizing the bandwidth unit and power allocation. We first provide the problem formulation and then propose one efficient algorithm to solve the problem.

A. Problem Formulation

In some application scenarios where the power consumption of the AP is of great concern, the design paradigm should be shifted to the energy efficient design by minimizing the power consumption. Specifically, we aim to jointly optimize the bandwidth unit and power allocation to minimize the TTP while guaranteeing each device's minimum throughput requirement and the budget of the total available bandwidth units. Mathematically, this optimization problem is formulated as follows:

$$
\begin{aligned}
(\mathbf{P3}):\ \min_{\boldsymbol{p},\boldsymbol{n}}\ & \sum\nolimits_{k=1}^{K} p_k && (27a)\\
\text{s.t. } & R_k \geq D_k^{\min}, \quad \forall k, && (27b)\\
& (6c), (6d), (6e), && (27c)
\end{aligned}
$$

where $D_k^{\min}$ is the minimum throughput requirement of the kth device. In the following, we always assume that the problem is feasible.

Problem $(\mathbf{P3})$ can be readily known as a mixed-integer programming problem due to the integer constraint on the number of bandwidth units, which is NP-hard to solve. We notice that the objective function of Problem $(\mathbf{P3})$ is not related to the number of bandwidth units and only depends on the power allocation. Hence, the BCD algorithm that alternately optimizes the bandwidth unit and power allocation is not applicable. In the following, we assume that the problem is feasible and we propose one low-complexity algorithm to solve this problem.

B. Approximation Method

The complicated expression of data rate R_k makes Problem (27) difficult to solve. To make it tractable, we approximate V_k^x as one, i.e., $V_k^x \approx 1$, where $x \in \{d, e\}$. The approximation is very accurate when SNR rate γ_k^x is very high, $\gamma_k^x \gg 1$. This approximation has been widely adopted in the current literature [21], [22]. Define $\tilde{h}_k^d = T|h_k|^2 \big/ \sigma_{d,k}^2$ and $\tilde{h}^e = T\|\mathbf{h}_e\|_2^2 \big/ \sigma_e^2$ with $\tilde{h}_k^d > \tilde{h}^e$, the achievable data rate can be approximated as

$$
\begin{aligned}
R_k \approx \tilde{R}_k = N_k \left(\log_2\left(1 + \frac{p_k \tilde{h}_k^d}{N_k}\right) - \log_2\left(1 + \frac{p_k \tilde{h}^e}{N_k}\right)\right)\\
- N_k \left(\sqrt{\frac{1}{N_k}} \frac{Q^{-1}(\epsilon_k)}{\ln 2} - \sqrt{\frac{1}{N_k}} \frac{Q^{-1}(\delta_k)}{\ln 2}\right). \quad (28)
\end{aligned}
$$

Since $V_k^x < 1$, $\tilde{R}_k$ is actually a lower bound of the original data rate R_k. Hence, if $\tilde{R}_k \geq D_k^{\min}$, then $R_k \geq D_k^{\min}$ always holds.

By substituting (28) into Problem $(\mathbf{P3})$, we can now optimize the channel blocklength allocation $\boldsymbol{N}(\boldsymbol{N} = \{N_1, \cdots, N_K\})$ and the power allocation $\boldsymbol{p}$, which is formulated as

$$
\begin{aligned}
(\mathbf{P4}):\ \min_{\boldsymbol{p},\boldsymbol{N}}\ & \sum\nolimits_{k=1}^{K} p_k && (29a)\\
\text{s.t. } & \tilde{R}_k \geq D_k^{\min}, \quad \forall k, && (29b)\\
& \sum\nolimits_{k=1}^{K} N_k \leq W_c T, && (29c)\\
& p_k \geq 0, \quad \forall k && (29d)\\
& N_k \in \{B_0 T, 2B_0 T, \ldots, n_{\max} BT\}, \quad \forall k, && (29e)
\end{aligned}
$$

where W_c is the coherence channel bandwidth. Due to the discrete constraint on N_k, Problem $(\mathbf{P4})$ is difficult to solve. To solve this problem, we first remove this constraint and relax it to continuous values, which is given by

$$
\begin{aligned}
(\mathbf{P4}-\mathbf{a}):\ \min_{\boldsymbol{p},\boldsymbol{N}}\ & \sum\nolimits_{k=1}^{K} p_k && (30a)\\
\text{s.t. } & (29b), (29c), (29d), && (30b)\\
& N_k \geq 0. && (30c)
\end{aligned}
$$

When Problem (**P4-a**) is solved, we convert the continuous N_ks into discrete values.

We first solve Problem (**P4-a**). Obviously, for any given channel blocklength allocation N_k, $\tilde{R}_k$ is a monotonically increasing function of p_k. Hence, inequality (29b) holds with equality at the optimal point. Then, the power allocation can be expressed as a function of N_k:

$$
p_k(N_k) = -\frac{N_k}{\tilde{h}^e} + \frac{c_k N_k}{d_k - e^{\frac{a_k}{N_k} + \frac{b_k}{\sqrt{N_k}}}}, \quad (31)
$$

where $a_k = D_k^{\min} \ln 2, b_k = Q^{-1}(\epsilon_k) + Q^{-1}(\delta_k)$, $c_k = \left(\tilde{h}_k^d - \tilde{h}^e\right) \Big/ \left(\tilde{h}^e\right)^2$, and $d_k = \tilde{h}_k^d \big/ \tilde{h}^e$. To guarantee that N_k is positive, by using $p_k(N_k) > 0$, we can obtain the lower bound of N_k as

$$
N_k \geq \left(\frac{b_k + \sqrt{b_k^2 + 4a_k \ln d_k}}{2 \ln d_k} \right)^2 \triangleq N_k^{\mathrm{lb}}. \quad (32)
$$

By substituting (31) into Problem (**P4-a**) and considering the lower bound of N_k, we have

$$(\mathbf{P4-b}): \min_{\boldsymbol{N}} \sum\nolimits_{k=1}^{K} p_k(N_k) \tag{33a}$$
$$\text{s.t. } (29c), (32). \tag{33b}$$

Then, in the following theorem, we provide a sufficient condition when $p_k(N_k)$ is a convex function.

Theorem 2: By defining $\rho_k = -\frac{12a_k+b_k^2}{3}$ and $\kappa_k = -\frac{2b_k^4+36a_kb_k^2+108a_k^2}{27b_k}$, the function $p_k(N_k)$ is a monotonically decreasing and convex function w.r.t. N_k when the following condition holds:

$$\sqrt{N_k} \leq 2\sqrt{-\frac{\rho_k}{3}}\cosh\left(\frac{1}{3}\text{arcosh}\left(\frac{3\kappa_k}{2\rho_k}\sqrt{\frac{-3}{\rho_k}}\right)\right)+\frac{b_k}{3}. \tag{34}$$

Proof: Please see Appendix E. ■

Fortunately, the RHS of (34) only depends on the long-term system parameters such as $D_k^{\min}$, ϵ_k, and δ_k, which is not related to the relative channel gains. For a typical URLLC communication system, the number of transmission bits for each device is around 100 bits (i.e., $D_k^{\min} = 100$), the decoding error probability ϵ_k is about 10^{-9}, the information leakage δ_k is roughly 10^{-2} [16]. Then, the value of the RHS of (34) can be calculated as 23.9. Hence, when $N_k \leq 572$, the inequality in (34) holds. For a typical system, the channel coherence bandwidth is around 0.5 MHz, and the transmission delay requirement is 1 ms. Hence, the total number of channel uses is 500, which should be allocated among all devices. Then, the number of channel uses allocated to each device is much smaller than the value of 500. As a result, for practical communication systems, the inequality in (34) holds and thus $p_k(N_k)$ is a convex function.

Since constraints (29c) and (32) are affine functions, Problem $(\mathbf{P4-b})$ is a convex problem, which can be solved by using Lagrangian dual decomposition method. We first introduce a positive Lagrange multiplier ς associated with constraint (29c), the partial Lagrangian function of Problem $(\mathbf{P4-b})$ is given by

$$\mathcal{L}(\boldsymbol{N},\varsigma) = \sum\nolimits_{k=1}^{K} p_k(N_k) + \varsigma\left(\sum\nolimits_{k=1}^{K} N_k - W_cT\right). \tag{35}$$

We first need to obtain the optimal $\boldsymbol{N}$ by minimizing $\mathcal{L}(\boldsymbol{N},\varsigma)$ over $\boldsymbol{N}$ for a given ς:

$$\min_{\boldsymbol{N}} \mathcal{L}(\boldsymbol{N},\varsigma). \tag{36}$$

We denote the optimal N_k for given ς as $N_k^\star(\varsigma)$. For given ς, $\mathcal{L}(\boldsymbol{N},\varsigma)$ is a convex function, and thus the optimal $N_k^\star(\varsigma)$ can be obtained as follows:

- If $\frac{\partial\mathcal{L}(\boldsymbol{N},\varsigma)}{\partial N_k}|_{N_k=N_k^{\text{lb}}} \geq 0$, the optimal N_k is given by $N_k^\star(\varsigma) = N_k^{\text{lb}}$;
- If $\frac{\partial\mathcal{L}(\boldsymbol{N},\varsigma)}{\partial N_k}|_{N_k=N_k^{\text{lb}}} < 0$, $N_k^\star(\varsigma)$ is the solution to the equation $\frac{\partial\mathcal{L}(\boldsymbol{N},\varsigma)}{\partial N_k} = 0$, which can be obtained by the bisection search method.

Once obtaining the optimal $N_k^\star(\varsigma)$, we can obtain the sum of all channel uses defined as

$$F(\varsigma) \triangleq \sum\nolimits_{k=1}^{K} N_k^\star(\varsigma). \tag{37}$$

We need to solve the following equation to find the optimal dual variable ς:

$$\varsigma\left(F(\varsigma) - W_cT\right) = 0. \tag{38}$$

If $F(0) \leq W_cT$, then the optimal ς is equal to zero. Otherwise, we need to solve the equation $F(\varsigma) = W_cT$. By using a similar method as in Lemma 3, we can prove that $F(\varsigma)$ is a monotonically decreasing function of ς. Hence, the bisection search method can be used to find the solution of equation $F(\varsigma) = W_cT$.

Denote the solution obtained from Problem $(\mathbf{P4-b})$ as $\bar{\boldsymbol{N}} = \{\bar{N}_1, \cdots, \bar{N}_K\}$. Obviously, the solution $\bar{\boldsymbol{N}}$ obtained by using the above the Lagrangian dual decomposition method do not satisfy the discrete constraint in (29e). Hence, we need to transfer $\bar{\boldsymbol{N}}$ to satisfy its discrete constraint. As mentioned before, this kind of problem is a combinatorial optimization problem, which is NP-hard to solve. We again adopt the greedy search method to solve this problem. Denote the solution of $\boldsymbol{N}$ that satisfies the discrete constraint as $\boldsymbol{N}^\star = \{N_1^\star, \cdots, N_K^\star\}$. Specifically, we first initialize the solution of $\boldsymbol{N}^\star$ as $N_k^\star = \left\lfloor \frac{\bar{N}_k}{B_0T} \right\rfloor \cdot B_0T, \forall k$. There are other channel uses that have not been allocated, the number of which is given by $\left(n_{\max} - \sum_{k=1}^{K}\left\lfloor \frac{\bar{N}_k}{B_0T} \right\rfloor\right) \cdot B_0T$. As proved in Theorem 2, $p_k(N_k)$ is a monotonically decreasing function of N_k. Hence, we can assign the unallocated channel uses to additionally reduce the power consumption. We allocate one channel use to the device with the largest decrement of $p_k(N_k)$, i.e., $k^\star = \arg\max_{k\in\mathcal{K}}\{p_k(N_k^\star) - p_k(N_k^\star + B_0T)\}$. For the $k^\star$th device, we allocate one bandwidth unit to it and update $N_k^\star = N_k^\star + B_0T$. Repeat this procedure until $\sum_{k=1}^{K} N_k^\star = W_cT$.

V. Simulation Results

In this section, we provide simulation results to evaluate the performance of our proposed algorithms. Unless otherwise specified, the adopted simulation parameters are given as follows: bandwidth of channel unit of $B_0 = 1$ KHz, noise power spectrum density of -173 dBm/Hz, number of devices of $K = 4$, $\epsilon_k = 10^{-9}, \forall k$, $\delta_k = 10^{-2}, \forall k$, time duration of $T = 1$ ms, channel coherence bandwidth of $W_c = 0.5$ MHz. The channel path loss is modeled as $PL = 35.3 + 37.6\log_{10}l$ (dB) [23], where l (m) is the distance between the devices/eavesdropper and the AP. The distance between the eavesdropper and the AP is set as $l_e = 180$ (m).

A. Weighted Sum Throughput Maximization

In this subsection, we provide simulation results to evaluate the performance of the BCD algorithm in Algorithm 1 for the WST maximization problem. The distances between the AP and the devices are assumed to be randomly generated within 100 m $\sim$ 120 m, and the following results are obtained by averaging over 200 device location generations.

In Fig. 2, we illustrate the convergence behavior of the BCD algorithm for various number of devices. It is observed from this figure that the BCD algorithm converges rapidly for all considered values of K, and roughly ten iterations are

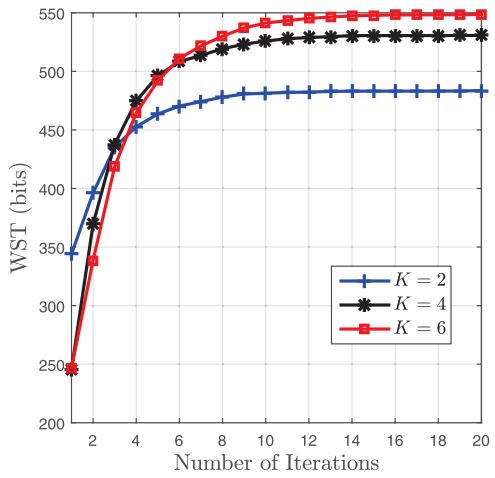

Fig. 2. Convergence behavior of the BCD algorithm.

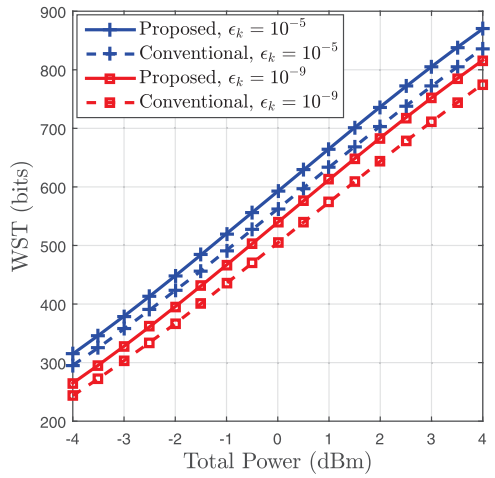

Fig. 3. WST versus the total power limit.

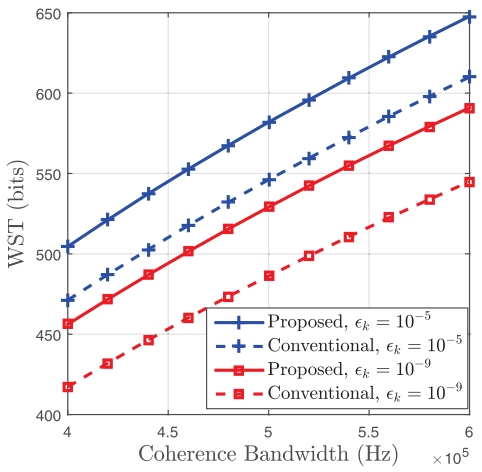

Fig. 4. WST versus the channel coherence bandwidth.

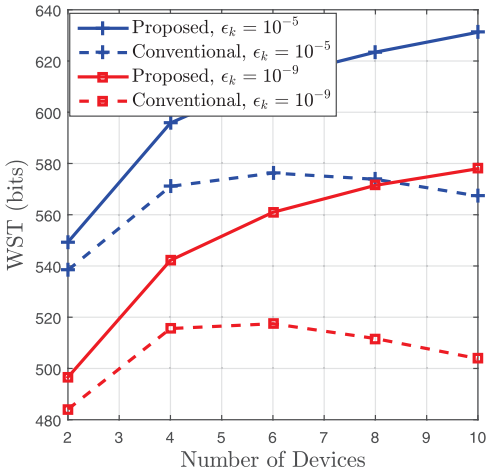

Fig. 5. WST versus the number of devices.

sufficient for the convergence of the BCD algorithm. Fig. 2 also shows that larger number of devices leads to slower convergence speed. The reason is that larger number of devices corresponds to more optimization variables to be optimized and require more iterations.

Next, we compare the performance of the proposed BCD algorithm with the conventional long packet transmission, where the penalty terms in (4) are not considered and the throughput of the kth device is given by

$$R_k = n_k B_0 T \left(\log_2(1+\gamma_k^d) - \log_2(1+\gamma_k^e)\right). \quad (39)$$

The BCD algorithm can be directly applied by setting some parameters to zero. This algorithm is labeled as 'Conventional'. Then, the solution obtained is applied in calculating the throughput in (5) by considering the penalty terms. This means that the solution obtained from 'Conventional' is used for obtaining solutions, while the throughput under finite blocklength is used for performance evaluation.

Fig. 3 shows the WST versus the total power limit for various decoding error probabilities ϵ at the devices. As expected, the WST of each algorithm increases with an increase of the maximum available transmit power as higher transmit power will bring higher value of SNR. The proposed BCD algorithm is observed to outperform the conventional long packet transmission scheme, and the performance gap increases as the transmit power limit becomes larger. This may be due to the fact that larger transmit power corresponds to a higher value of SNR, and thus V_k^x will approach one. Then, the impact of the penalty terms will increase, which is not considered in the conventional long packet transmission scheme. We can also find from this figure that a lower value of the decoding error probability requirement brings a lower WST. This is because when δ_k is large, the value of $Q^{-1}(\delta_k)$ increases, thus leading to larger values of the penalty terms.

Fig. 4 shows the WST versus the channel coherence bandwidth W_c. We observe from Fig. 4 that the WST achieved by all the schemes increase with the increase of channel coherence bandwidth. The reason is that higher channel coherence bandwidth corresponds to larger number of channel uses for each transmission, and thus brings higher throughput. In contrast to Fig. 3 where the WST logarithmically increases with the increasing transmit power, the WST linearly increases with W_c, which demonstrates the significant impact of the channel coherence bandwidth on the WST performance. It is again observed that the performance of the proposed BCD algorithm is better than the conventional long packet transmission scheme.

Fig. 5 shows the WST versus the number of devices. It is found from this figure that the WST achieved by the proposed

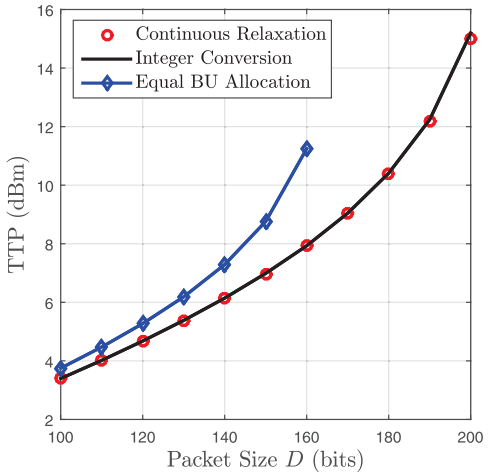

Fig. 6. TTP versus the minimum packet size requirement D.

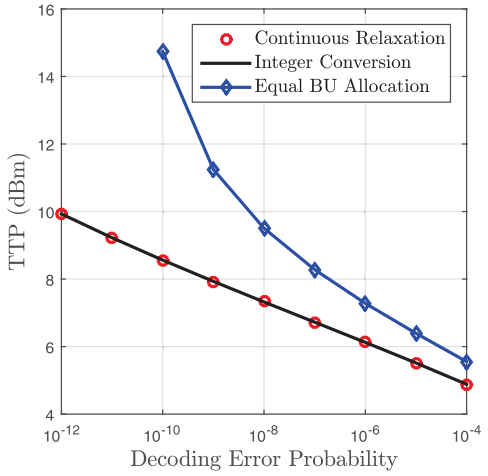

Fig. 7. TTP versus the decoding error probability of the devices.

BCD algorithm increases with the number of devices as we can employ the multiuser diversity to achieve higher performance. In contrast, the WST of the conventional long packet scheme first increases with K and then decreases with K. The main reason is that the conventional long packet scheme targets at optimizing (39) without considering the penalty incurred due to the short packet transmission. The solution that maximizes (39) may not perform well for the short packet throughput formula in (5). This again emphasizes the importance of optimizing the short packet throughput formula in URLLC applications.

B. Total Transmit Power Minimization

In this subsection, we consider the performance of the proposed method in Section IV for the TTP minimization problem. The distance between the AP and the devices are set as: $l_k = 100 + 5(k-1)$ (m), where k denotes the device index. The minimum date packet size is $D_k^{\min} = 160$ bits. Three methods are compared. The first one is the solution obtained by solving Problem $(\mathbf{P4}-\mathbf{a})$ (with legend 'Continuous Relaxation'), which is a relaxed version of the original Problem $(\mathbf{P4})$. The second one is the solution obtained by converting the continuous solution of $\bar{\mathbf{N}}$ into the discrete solution by using the greedy method (with legend 'Integer Conversion'). The final one is the solution obtained by equally allocating the channel bandwidth units to the devices, $n_k = n_{\max}/K$ (with legend 'Equal BU Allocation'), and the power allocated to each device can be obtained based on (31).

In Fig. 6, we first study the impact of packet size requirement of each device on the TTP. When packet size D ranges from 100 bits to 200 bits with the interval of 10 bits in Fig. 6, the upper bound of N_k given in Theorem 2 are respectively calculated as 573, 625, 677, 730, 783, 837, 890, 944, 998, 1054 and 1109. As $W_cT = 500$, the condition in Theorem 2 is always satisfied. As expected, the TTP required monotonically increases with D for all the methods. Both the Continuous Relaxation and the Integer Conversion methods have almost the same performance, which implies the marginal performance loss incurred by the greedy integer conversion procedure. Moreover, both these methods are shown to achieve superior performance over the naive Equal BU Allocation method, and the performance gain monotonically increases with D. When $D \geq 160$ bits, the Equal BU Allocation method becomes even infeasible, while the proposed algorithm can support the packet size up to 200 bits. The reason can be explained as follows. When the channel blocklength is equally allocated among the devices, the QoS of the device with worse channel gain cannot be satisfied even when all the available power is allocated to this device. Hence, the optimization problem associated with the Equal BU Allocation has no feasible solution. In contrast, our proposed algorithm dynamically allocated the channel blocklength bases on each device's channel condition: the device with bad channel condition can be allocated with more channel blocklength to satisfy its QoS requirement. Hence, our proposed algorithm not only achieved high WST, but also has larger feasible range. This implies the importance of optimizing the bandwidth unit allocation.

In Fig. 7, we investigate the impact of the decoding error probability requirement of the devices on the TTP. When the decoding error probability ranges from 10^{-12} to 10^{-4} with the interval shown in Fig. 6, the upper bound of N_k given in Theorem 2 are respectively calculated as 896, 890, 887, 887, 888, 890, 892, 895, and 899. As $W_cT = 500$, the condition in Theorem 2 is always satisfied. It is observed that the TTP required by all the methods decrease with ϵ_k. This can be explained as follows. According to (4), R_k is a monotonically decreasing function of ϵ_k. When ϵ_k is small, more power is required to achieve the desired throughput requirement. Again, the proposed algorithms are observed to have better performance than the Equal BU Allocation method, especially when the decoding error probability is extremely small.

The impact of the channel bandwidth on the system performance is shown in Fig. 8. For various values of W_c, the upper bounds of N_k given in Theorem 2 are the same that are all equal to 891. In addition, the lower bounds of N_k given in (32) for four devices are respectively given by 85, 95, 107, and 120. When $W_c = 1.1$ MHz, W_cT has the maximum value of 1100. Then, the maximum values of $N_k, k = 1, \cdots, 4$ are given by

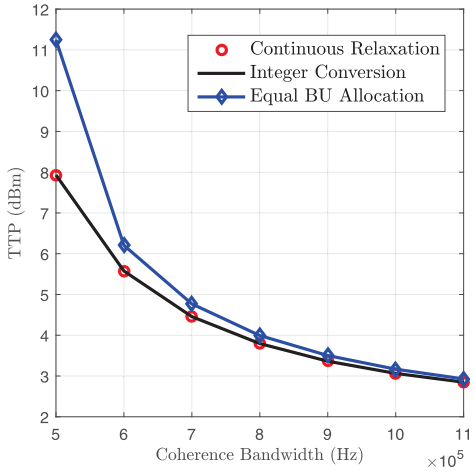

Fig. 8. Sum power versus the channel coherence bandwidth.

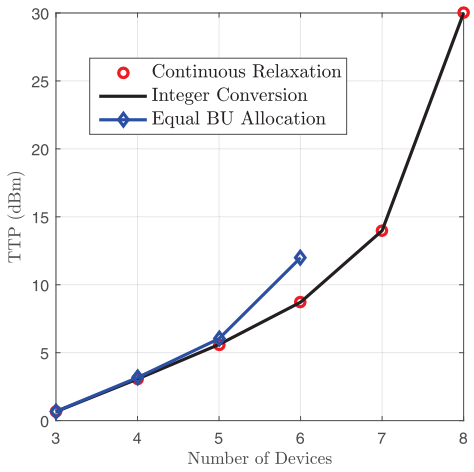

Fig. 9. Sum power versus the number of devices.

778, 788, 800, and 813, respectively. This means the maximum achievable value of N_k is smaller than upper bound of N_k given in Theorem 2. Hence, the condition in Theorem 2 is always satisfied. We can find from Fig. 8 that the TTP required by all the methods decreases with increasing channel coherence bandwidth. This is mainly due to the fact that when the channel coherence bandwidth increases, the total number of channel users increases, which can enhance throughput. It is interesting to observe that when the channel coherence bandwidth is sufficiently large, the proposed algorithms can only achieve negligible performance advantage over the Equal BU Allocation method, which implies the equal bandwidth unit allocation is nearly optimal for large channel coherence bandwidth.

In Fig. 9, we study the impact of the number of devices on the system performance where the channel coherence bandwidth is assumed to be $W_c = 1$ MHz. For various values of K, the upper bounds of N_k given in Theorem 2 are the same that are equal to 891. In addition, the lower bounds of N_k given in (32) for the case of three devices are respectively given by 85, 95, and 107. When $W_c = 1$ MHz, W_cT has the maximum value of 1000. Then, the maximum values of

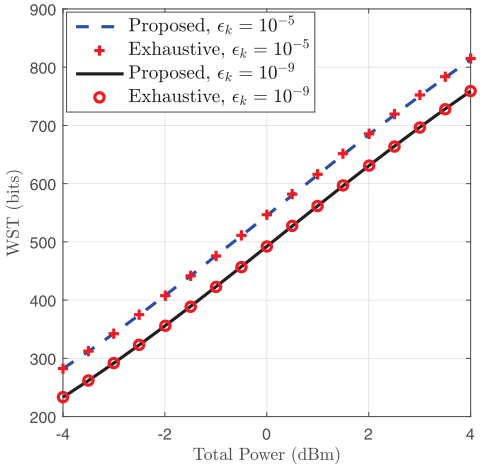

Fig. 10. WST performance comparison between exhaustive search method and the proposed method.

$N_k, k = 1, \cdots, 3$ are given by 798, 808, and 820, respectively. This means the maximum achievable value of N_k is smaller than the upper bound of N_k given in Theorem 2. Hence, the condition in Theorem 2 is always satisfied. It is observed that the sum power increases rapidly with the number of devices. This is because when the number of devices is large, the number of bandwidth units allocated to each device will decrease, which requires more power to transmit the targeted throughput.

C. Comparison With Exhaustive Search Method

Finally, we compare our proposed method with the exhaustive search method. Due to the high complexity of the exhaustive search method, we consider a small network with two devices.

First, we study the WSR performance comparison between the proposed algorithm with the exhaustive search method in Fig. 10. The other simulation parameters are set to be same as in Subsection V-A. For the exhaustive search method, we enumerate all possible bandwidth unit allocations, and the algorithm in Subsection III-B is adopted to solve the power allocation problem for each given allocation of the bandwidth units. It can be seen from Fig. 10 that our proposed algorithm can achieve almost the same performance as the exhaustive search method for the whole range of transmission power, which verifies the effectiveness of our proposed method.

Then, we compare our proposed algorithm for the TTP minimization problem with the exhaustive search method in Fig. 11. The parameters are set to the same as in Subsection V-B. The condition in Theorem 2 can be verified to hold. For the exhaustive search method, we enumerate all possible bandwidth unit allocations, and we use the line search method to find the minimum transmit power for each user to satisfy the minimum throughput requirement. It can be observed from Fig. 11 that the proposed method achieves the similar performance when the channel bandwidth is small, while the performance gap increases with the channel bandwidth. However, the maximum performance

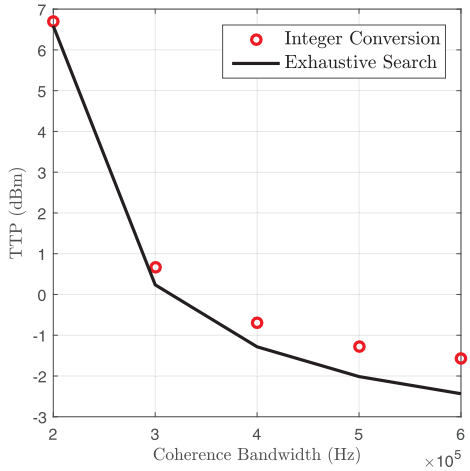

Fig. 11. TTP performance comparison between exhaustive search method and the proposed method.

gap is less than 1 dBm, which is acceptable considering the lower complexity associated with the proposed method. This means our proposed algorithm is more appealing for practical applications.

VI. Conclusion

In this paper, we studied a secure mission-critical IoT communication system under URLLC requirements, where the AP transmits safety-critical messages to the devices and there exists an eavesdropper that attempts to eavesdrop this critical message. Under this context, we considered the WST maximization problem and the TTP minimization problem through joint bandwidth unit and power allocation. For the WST maximization problem, we provided the BCD algorithm to decouple the original coupled optimization problem, and obtain its solution in an iterative manner. For the TTP minimization problem, we derived the sufficient condition when this problem is a convex problem, and we showed that most of the typical URLLC applications satisfy this condition. Low-complexity and efficient algorithms were proposed to find the globally optimal solution, and the greedy method was utilized to convert the continuous solutions into discrete solutions. Simulation results demonstrate the rapid convergence of the BCD algorithm, and performance advantages over the conventional long packet transmission scheme. For the method to solve the TTP minimization problem, simulation results validate the performance advantages in terms of power savings compared with the naive equal bandwidth unit allocation scheme.

Appendix A
Proof Lemma 1

We first prove that $f_k(p_k)$ is a concave function. The second-order derivative of $f_k(p_k)$ w.r.t. p_k is given by

$$f_k''(p_k) = \frac{N_k}{\ln 2}\frac{\bar{g}_k^e-\bar{g}_k^d}{\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)}\left(\frac{\bar{g}_k^e}{1+p_k\bar{g}_k^e}+\frac{\bar{g}_k^d}{1+p_k\bar{g}_k^d}\right) < 0, \tag{40}$$

where the last inequality holds since $\bar{g}_k^d > \bar{g}_k^e$. Hence, $f_k(p_k)$ is a concave function w.r.t. p_k. Similarly, the second-order derivative of $y_k(p_k)$ w.r.t. p_k is given by

$$y_k''(p_k) = -\frac{\left(\bar{g}_k^d\right)^2 L_k^d\left(3-2\left(1+p_k\bar{g}_k^d\right)^{-2}\right)}{\left(1-\left(1+p_k\bar{g}_k^d\right)^{-2}\right)^{\frac{3}{2}}\left(1+p_k\bar{g}_k^d\right)^4} - \frac{\left(\bar{g}_k^e\right)^2 L_k^e\left(3-2\left(1+p_k\bar{g}_k^e\right)^{-2}\right)}{\left(1-\left(1+p_k\bar{g}_k^e\right)^{-2}\right)^{\frac{3}{2}}\left(1+p_k\bar{g}_k^e\right)^4} < 0. \tag{41}$$

Hence, $y_k(p_k)$ is a concave function w.r.t. p_k. As a result, $R_k(p_k)$ is the difference of two concave functions $f_k(p_k)$ and $y_k(p_k)$, which completes the proof.

Appendix B
Proof Theorem 1

Let us define $\lambda \geq 0$ and $\boldsymbol{\mu} = \{\mu_1, \cdots, \mu_K\}$ as the non-negative dual variables associated with the total power constraint (6b) and the individual non-negative power constraint (6e), respectively. The Lagrangian function of Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ can be formulated as

$$\mathcal{L}(\boldsymbol{p},\boldsymbol{\mu},\lambda) = \sum\nolimits_{k=1}^{K}\left(\omega_k f_k\left(p_k\right) - \omega_k\beta_k(p_k^{(i-1)})p_k\right) -\lambda\left(\sum\nolimits_{k=1}^{K}p_k - P_{\max}\right) + \sum\nolimits_{k=1}^{K}\mu_k p_k. \tag{42}$$

Since Problem $(\mathbf{P2}-\mathbf{1}-\mathbf{a})$ is a convex optimization problem, the globally optimal solution satisfies the Karush-Kuhn-Tucker (KKT) conditions as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{p},\boldsymbol{\mu},\lambda)}{\partial p_k} &= \frac{N_k}{\ln 2}\frac{\omega_k(\bar{g}_k^d-\bar{g}_k^e)}{\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)} - \omega_k\beta_k(p_k^{(i-1)})\\ &\quad -\lambda+\mu_k = 0, \quad \forall k,\\ &\quad \mu_k p_k = 0, p_k \geq 0, \ \forall k, \lambda\left(\sum_{k=1}^{K}p_k - P_{\max}\right) = 0,\\ &\quad \sum_{k=1}^{K}p_k \leq P_{\max}.\end{aligned} \tag{43}$$

Note that $\mu_k, \forall k$ are slack variables in the first equation, which can be eliminated. We then have:

$$\begin{aligned}&\left(\omega_k\beta_k(p_k^{(i-1)}) + \lambda - \frac{N_k}{\ln 2}\frac{\omega_k(\bar{g}_k^d-\bar{g}_k^e)}{\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)}\right)p_k = 0, \ \forall k,\\ &\omega_k\beta_k(p_k^{(i-1)}) + \lambda \geq \frac{N_k}{\ln 2}\frac{\omega_k(\bar{g}_k^d-\bar{g}_k^e)}{\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)}, \ \forall k,\\ &\lambda\left(\sum\nolimits_{k=1}^{K}p_k - P_{\max}\right) = 0, \sum\nolimits_{k=1}^{K}p_k \leq P_{\max}, p_k \geq 0, \ \forall k.\end{aligned} \tag{44}$$

By defining $\eta_k^{(i-1)}(\lambda)$ in Theorem 1, the KKT conditions in (44) can be rewritten as

$$\begin{aligned}&\left(\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)-\eta_k^{(i-1)}(\lambda)\right)p_k=0,\quad \forall k,\\&\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)\geq\eta_k^{(i-1)}(\lambda),\quad \forall k,\\&\lambda\left(\sum\nolimits_{k=1}^{K}p_k-P_{\max}\right)=0,\sum\nolimits_{k=1}^{K}p_k\leq P_{\max},p_k\geq 0,\ \forall k.\end{aligned}\tag{45}$$

If $\eta_k^{(i-1)}(\lambda)>1$, the conditions in (45) hold only when $p_k>0$. This can be proved by using the contradiction method. Assume that $p_k=0$. Then, based on the second condition of (45), we have $1\geq\eta_k^{(i-1)}(\lambda)$, which contradicts the condition of $\eta_k^{(i-1)}(\lambda)>1$. Hence, $p_k>0$ should hold. Then, based on the first condition of (45), p_k should satisfy the following equation:

$$\left(1+p_k\bar{g}_k^d\right)\left(1+p_k\bar{g}_k^e\right)-\eta_k^{(i-1)}(\lambda)=0,\tag{46}$$

and its solution is given in (13) in Theorem 1.

On the other hand, if $\eta_k^{(i-1)}(\lambda)\leq 1$, then p_k must be equal to zero. This can also be proved by using the contradiction method. Assume that $p_k>0$. Then, based on the first condition of (45), the equation in (46) should hold, and p_k is derived as:

$$\begin{aligned}p_k&=\frac{-\left(\bar{g}_k^d+\bar{g}_k^e\right)+\sqrt{\left(\bar{g}_k^d+\bar{g}_k^e\right)^2-4\bar{g}_k^d\bar{g}_k^e\left(1-\eta_k^{(i-1)}\right)}}{2\bar{g}_k^d\bar{g}_k^e}\\&\leq\frac{-\left(\bar{g}_k^d+\bar{g}_k^e\right)+\sqrt{\left(\bar{g}_k^d+\bar{g}_k^e\right)^2}}{2\bar{g}_k^d\bar{g}_k^e}=0,\end{aligned}$$

which contradicts the assumption of $p_k>0$. Hence, p_k must be equal to zero.

Combining the above two cases, the optimal solution of p_k is given in (13) in Theorem 1. The remaining part of Theorem 1 can be readily proved by using the similar analysis for the total power constraint, details of which are omitted for simplicity.

Appendix C
Proof of Lemma 2

We first prove that $F_k(n_k)$ is a concave function w.r.t. n_k. With some manipulations, the second-order derivative of $F(n_k)$ w.r.t. n_k is given by

$$F_k''(n_k)=\frac{\tilde{N}_0\left(\tilde{g}_k^e-\tilde{g}_k^d\right)\left(\left(\tilde{g}_k^e+\tilde{g}_k^d\right)n_k+2\tilde{g}_k^e\tilde{g}_k^d\right)}{\ln 2\left(n_k+\tilde{g}_k^e\right)^2\left(n_k+\tilde{g}_k^d\right)^2}<0,\tag{47}$$

where the inequality holds since $\bar{g}_k^e<\bar{g}_k^d$. Hence, $F_k(n_k)$ is a concave function w.r.t. n_k.

Now, we start to prove that $G_k(n_k)$ is also a concave function w.r.t. n_k. With some manipulations, the second-order derivative of $G_k(n_k)$ w.r.t. n_k is given by

$$\begin{aligned}G_k''(n_k)=&\frac{2\frac{\partial^2 z_k^d(n_k)}{\partial n_k^2}z_k^d(n_k)-\left(\frac{\partial z_k^d(n_k)}{\partial n_k}\right)^2}{4z_k^d(n_k)\sqrt{z_k^d(n_k)}}\tilde{L}_k^d\\&+\frac{2\frac{\partial^2 z_k^e(n_k)}{\partial n_k^2}z_k^e(n_k)-\left(\frac{\partial z_k^e(n_k)}{\partial n_k}\right)^2}{4z_k^e(n_k)\sqrt{z_k^e(n_k)}}\tilde{L}_k^e,\end{aligned}\tag{48}$$

where $\frac{\partial^2 z_k^x(n_k)}{\partial n_k^2}$ is given by

$$\frac{\partial^2 z_k^x(n_k)}{\partial n_k^2}=-\frac{6n_k\left(\tilde{g}_k^x\right)^2}{\left(n_k+\tilde{g}_k^x\right)^4}<0,x\in\{d,e\}.\tag{49}$$

Then, combining (49) with (48), we know that $\frac{\partial^2 G_k(n_k)}{\partial n_k^2}<0$. Hence, $G_k(n_k)$ is also a concave function w.r.t. n_k, and $R_k(n_k)$ is the difference of two concave functions $F_k(n_k)$ and $G_k(n_k)$.

Appendix D
Proof of Lemma 3

We consider a pair of dual variables μ_1 and μ_2, where $\mu_1\geq\mu_2$. Denote $n_k^\star(\mu_1)$ and $n_k^\star(\mu_2)$ as the optimal solution of Problem (24) when $\mu=\mu_1$ and $\mu=\mu_2$, respectively. Since $n_k^\star(\mu_1)$ is the optimal solution of Problem (24) when $\mu=\mu_1$, we have

$$\begin{aligned}&\mathcal{L}_k(n_k^\star(\mu_1),\mu_1)\\&\quad=F_k(n_k^\star(\mu_1))-\alpha_k(n_k^{(j-1)})n_k^\star(\mu_1)-\mu_1 n_k^\star(\mu_1)\\&\quad\geq\mathcal{L}_k(n_k^\star(\mu_2),\mu_1)\\&\quad=F_k(n_k^\star(\mu_2))-\alpha_k\left(n_k^{(j-1)}\right)n_k^\star(\mu_2)-\mu_1 n_k^\star(\mu_2).\end{aligned}\tag{50}$$

Furthermore, $n_k^\star(\mu_2)$ is the optimal solution of Problem (24) when $\mu=\mu_2$, we have

$$\begin{aligned}&\mathcal{L}_k(n_k^\star(\mu_2),\mu_2)\\&\quad=F_k(n_k^\star(\mu_2))-\alpha_k(n_k^{(j-1)})n_k^\star(\mu_2)-\mu_2 n_k^\star(\mu_2)\\&\quad\geq\mathcal{L}_k(n_k^\star(\mu_1),\mu_2)\\&\quad=F_k(n_k^\star(\mu_1))-\alpha_k(n_k^{(j-1)})n_k^\star(\mu_1)-\mu_2 n_k^\star(\mu_1).\end{aligned}\tag{51}$$

By adding these two inequalities and simplifying them, we have $\left(n_k^\star(\mu_1)-n_k^\star(\mu_2)\right)(\mu_2-\mu_1)\geq 0$. Since $\mu_1>\mu_2$, we have $n_k^\star(\mu_1)<n_k^\star(\mu_2)$. Then, by summing all these K inequalities, we have $W(\mu_1)=\sum_{k=1}^K n_k^\star(\mu_1)<\sum_{k=1}^K n_k^\star(\mu_2)=W(\mu_2)$. Hence, $W(\mu)$ is a monotonically decreasing function of μ.

Appendix E
Proof of Theorem 2

We first prove its convexity. With some manipulations, the second-order derivative of $p_k(N_k)$ w.r.t. N_k is calculated as

$$p_k''(N_k)=\frac{\frac{c_k d_k}{N_k^3}e^{\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}}\Xi(N_k)+\frac{c_k}{N_k^3}e^{\frac{2a_k}{N_k}+\frac{2b_k}{\sqrt{N_k}}}\Phi(N_k)}{\left(d_k-e^{\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}}\right)^3},\tag{52}$$

where $\Xi(N_k)$ and $\Phi(N_k)$ are given by

$$\Xi(N_k) = -\frac{b_k}{4}N_k^{\frac{3}{2}} + \frac{b_k^2}{4}N_k + a_k b_k N_k^{\frac{1}{2}} + a_k^2,$$
$$\Phi(N_k) = \frac{b_k}{4}N_k^{\frac{3}{2}} + \frac{b_k^2}{4}N_k + a_k b_k N_k^{\frac{1}{2}} + a_k^2. \quad (53)$$

Since $N_k > N_k^{\mathrm{lb}}$, the denominator of (52) is larger than zero. Obviously, $\Phi(N_k)$ is larger than zero. Hence, if $\Xi(N_k) > 0$, then $p_k''(N_k) > 0$ holds and $p_k(N_k)$ is a convex function of N_k. Next, we derive the condition when $\Xi(N_k) > 0$.

Denote $t_k = N_k^{\frac{1}{2}}$. Then, $\Xi(N_k)$ can be re-expressed as

$$\Xi(t_k) = -\frac{b_k}{4}t_k^3 + \frac{b_k^2}{4}t_k^2 + a_k b_k t_k + a_k^2. \quad (54)$$

Note that $\Xi(0) = a_k^2 > 0$ and $\Xi(+\infty) = -\infty$. Since $\Xi(t_k)$ is a continuous function, there must exist at least one positive solution for the equation $\Xi(t_k) = 0$. In the following, we prove that the solution is unique.

We rewrite equation $\Xi(t_k) = 0$ as a standard cubic equation:

$$u_k t_k^3 + v_k t_k^2 + w_k t_k + z_k = 0, \quad (55)$$

where $u_k = -\frac{b_k}{4}$, $v_k = \frac{b_k^2}{4}$, $w_k = a_k b_k$, and $z_k = a_k^2$.

By dividing (55) by u_k and inserting $t_k = x_k - v_k/3u_k$, we have

$$x_k^3 + \rho_k x_k + \kappa_k = 0, \quad (56)$$

where ρ_k and κ_k are defined in Theorem 2. It can be readily verified that

$$4\rho_k^3 + 27\kappa_k^2 > 0, \quad \kappa_k > 0, \ \rho_k < 0. \quad (57)$$

As a result, there exists only one real solution for (56), which is given by

$$x_k^* = -2\sqrt{-\frac{\rho_k}{3}}\cosh\left(\frac{1}{3}\mathrm{arcosh}\left(\frac{-3\kappa_k}{2\rho_k}\sqrt{\frac{-3}{\rho_k}}\right)\right). \quad (58)$$

Thus, the unique solution of equation (55) is given by $t_k^* = x_k^* - v_k/3u_k$.

Based on the above discussion, we can conclude that when $t_k < t_k^* = x_k^* - v_k/3u_k$, $\Xi(t_k)$ is positive and $p_k(N_k)$ is a convex function.

Now we proceed to prove that $p_k(N_k)$ is a monotonically decreasing function when inequality (34) holds. The first-order derivative of $p_k(N_k)$ w.r.t. p_k is given by

$$p_k'(N_k) = -\frac{1}{\tilde{h}^e} + \frac{c_k\left(d_k - e^{\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}}\right) - c_k e^{\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}}\left(\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}\right)}{\left(d_k - e^{\frac{a_k}{N_k}+\frac{b_k}{\sqrt{N_k}}}\right)^2}. \quad (59)$$

Since $p_k(N_k)$ is a convex function when inequality (34) holds, $p_k'(N_k)$ is a monotonically increasing function. Then, we have

$$p_k'(N_k) < p_k'(\infty) = 0. \quad (60)$$

In consequence, $p_k(N_k)$ is a monotonically decreasing function when inequality (34) holds.

REFERENCES

[1] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[3] J. Chen, L. Zhang, Y.-C. Liang, X. Kang, and R. Zhang, "Resource allocation for wireless-powered IoT networks with short packet communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1447–1461, Feb. 2019.

[4] C. Pan, H. Ren, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint blocklength and location optimization for URLLC-enabled UAV relay systems," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 498–501, Mar. 2019.

[5] H. Ren, C. Pan, K. Wang, Y. Deng, M. Elkashlan, and A. Nallanathan, "Achievable data rate for URLLC-enabled UAV systems with 3-D channel model," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1587–1590, Dec. 2019.

[6] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.

[7] C. She, Y. Duan, G. Zhao, T. Q. S. Quek, Y. Li, and B. Vucetic, "Cross-layer design for mission-critical IoT in mobile edge computing systems," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9360–9374, Dec. 2019.

[8] H. Ren *et al.*, "Joint power and blocklength optimization for URLLC in a factory automation scenario," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786–1801, Mar. 2020.

[9] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020. [Online]. Available: https://arxiv.org/abs/1912.12438

[10] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1550–1573, 3rd Quart., 2014.

[11] A. Mukherjee, "Physical-layer security in the Internet of Things: Sensing and communication confidentiality under resource constraints," *Proc. IEEE*, vol. 103, no. 10, pp. 1747–1761, Oct. 2015.

[12] Z. Li, M. Chen, C. Pan, N. Huang, Z. Yang, and A. Nallanathan, "Joint trajectory and communication design for secure UAV networks," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 636–639, Apr. 2019.

[13] Y. Zhou *et al.*, "Secure communications for UAV-enabled mobile edge computing systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 376–388, Jan. 2020.

[14] Y. Cao *et al.*, "Secure transmission via beamforming optimization for NOMA networks," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 193–199, Feb. 2020.

[15] W. Yang, R. F. Schaefer, and H. V. Poor, "Wiretap channels: Nonasymptotic fundamental limits," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4069–4093, Jul. 2019.

[16] H.-M. Wang, Q. Yang, Z. Ding, and H. V. Poor, "Secure short-packet communications for mission-critical IoT applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2565–2578, May 2019.

[17] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Resource allocation for URLLC in 5G mission-critical IoT networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[18] Q. Dinh *et al.*, "Local convergence of sequential convex programming for nonconvex optimization," in *Recent Advances in Optimization and its Applications in Engineering*. Berlin, Germany: Springer, 2010, pp. 93–102.

[19] S. Boyd, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[20] C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Robust beamforming design for ultra-dense user-centric C-RAN in the face of realistic pilot contamination and limited feedback," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 780–795, Feb. 2019.

[21] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[22] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 402–415, Jan. 2019.

[23] *Further Advancements for E-UTRA Physical Layer Aspects*, E. U. T. R. Access, 3GPP document TR 36.814, 2010.

Hong Ren (Member, IEEE) received the B.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2014 and 2018, respectively. From October 2016 to January 2018, she was a Visiting Student with the School of Electronics and Computer Science, University of Southampton, U.K. She is currently a Post-Doctoral Scholar with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. Her research interests lie in the areas of communication and signal processing, including cooperative transmission, the Internet of Things, and ultra-reliability and low latency communications.

Cunhua Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively.

From 2015 to 2016, he was a Research Associate with the University of Kent, U.K. He held a post-doctoral position at the Queen Mary University of London, U.K., from 2016 and 2019, where he is currently a Lecturer. His research interests mainly include intelligent reflection surface (IRS), machine learning, UAV, the Internet of Things, and mobile edge computing. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He also serves as an Editor of IEEE WIRELESS COMMUNICATION LETTERS and IEEE ACCESS.

Yansha Deng (Member, IEEE) received the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2015. From 2015 to 2017, she was a Post-Doctoral Research Fellow with King's College London, U.K., where she is currently a Lecturer (Assistant Professor) with the Department of Informatics. Her research interests include molecular communication, machine learning, and 5G wireless networks. She was a recipient of the Best Paper Awards from ICC 2016 and Globecom 2017 as the first author. She is currently an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS, and a Senior Editor of IEEE COMMUNICATION LETTERS. She also received the Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2018. She has also served as a TPC member for many IEEE conferences, such as IEEE GLOBECOM and ICC.

Maged Elkashlan (Member, IEEE) received the Ph.D. degree in electrical engineering from The University of British Columbia, Canada, in 2006. From 2007 to 2011, he was with Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. During this time, he held visiting appointments at the University of New South Wales and the University of Technology Sydney. In 2011, he joined the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests include communication theory and statistical signal processing. He received the Best Paper Award at the IEEE International Conference on Communications (ICC) in 2016 and 2014, the International Conference on Communications and Networking in China (CHINACOM) in 2014, and the IEEE Vehicular Technology Conference (VTC-Spring) in 2013. He currently serves as an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Arumugam Nallanathan (Fellow, IEEE) has been a Professor of wireless communications and the Head of the Communication Systems Research (CSR) Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, since September 2017. He was with the Department of Informatics, King's College London, from December 2007 to August 2017, where he was a Professor of wireless communications from April 2013 to August 2017 and has been a Visiting Professor since September 2017. He was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from August 2000 to December 2007. He published nearly 500 technical papers in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, the Internet of Things (IoT), and molecular communications.

Dr. Nallanathan was a co-recipient of the Best Paper Award presented at the IEEE International Conference on Communications 2016 (ICC'2016), IEEE Global Communications Conference 2017 (GLOBECOM'2017), and IEEE Vehicular Technology Conference 2018 (VTC'2018). He received the IEEE Communications Society SPCE outstanding service award 2012 and IEEE Communications Society RCC outstanding service award 2014. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and a Technical Program Chair and a member of Technical Program Committees in numerous IEEE conferences. He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016. He was an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2006 to 2011, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2006 to 2017, and IEEE SIGNAL PROCESSING LETTERS. He is an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS.