

RESEARCH

Open Access



# Security-aware energy-efficient design for mobile edge computing network operating with finite blocklength codes

Chenhao Shi<sup>1</sup>, Yulin Hu<sup>1,2\*</sup> , Yao Zhu<sup>1,2\*†</sup> and Anke Schmeink<sup>2</sup>

<sup>†</sup>Yulin Hu and Yao Zhu have contributed equally to this work.

\*Correspondence:  
yulin.hu@whu.edu.cn;  
yao.zhu@isek.rwth-aachen.de

<sup>1</sup> School of Electronic Information, Wuhan University, Wuhan 430000, Hubei Province, China

<sup>2</sup> Chair of Information Theory and Data Analytics, RWTH Aachen University, 52068 Aachen, Germany

## Abstract

Energy efficiency and physical-layer security are crucial considerations in the advancement of mobile edge computing systems. This paper addresses the trade-off between secure-reliability and energy consumption in finite blocklength (FBL) communications. Specifically, we examine a three-node scenario involving a user, a legitimate edge computing server, and an eavesdropper, where the user offloads sensitive data to the edge server while facing potential eavesdropping threats. We propose an optimization framework aimed at minimizing energy consumption while ensuring secure-reliability by decomposing the problem into manageable subproblems. By demonstrating the convexity of the objective function concerning the variables, we establish the existence of an optimal parameter selection for the problem. This implies that practical optimization of parameters can significantly enhance system performance. Our numerical results demonstrate that the application of FBL regime and retransmission mechanism can effectively reduce the energy consumption of the system while ensuring secure-reliability. For the quantitative analyses, the retransmission mechanism is 33.1% better than no retransmission, and the FBL regime is 13.1% better than infinite blocklength (IBL) coding.

**Keywords:** Edge computing, Finite blocklength regime, Retransmission, Physical layer security

## 1 Introduction

With the growing concern about information security in normalise the Industrial Internet, user data protection has become a key issue in mobile applications. In particular, the collected data may contain many sensitive information of users with the development of terminal devices. In traditional cloud computing applications, user data is transmitted over long distances to remote cloud centers [1] which results in a high leakage probability. Besides, cloud servers are easier targets for financially motivated attacks. To reduce the leakage risk and delay, Mobile Edge Computing (MEC) has garnered increasing attention within the academic community in recent years [2]. MEC, unlike cloud computing, places computational resources closer to applications, reducing latency and conserving bandwidth. This proximity offers faster response times and enhances security.

MEC's smaller architecture enables private use, like family clouds, minimizing data leakage risk. However, other devices in the network can be potential eavesdroppers.

To solve these problems, physical layer security (PLS) is introduced which is a security approach that focuses on safeguarding wireless network transmissions from eavesdroppers by exploiting the characteristics of the physical layer of communication [1, 3]. Differ from traditional cryptography methods that rely on complex mathematical algorithms for securing data, PLS takes advantage of the inherent properties of wireless communication channels to provide security. In particular, PLS can adapt dynamically to changing channel conditions. It can adjust its security measures based on the quality of the wireless link, ensuring that security is maintained even in challenging environments [3, 4]. This flexibility is valuable in MEC scenarios where network conditions may vary.

On the other hand, in future, ultra-reliable and low-latency communications (URLLC) will play a critical role in those delay-sensitive applications, such as Virtual Reality (VR) and Autonomous Driving [5]. URLLC refers to a communication technology for critical application scenarios, designed to provide extremely high reliability and very low transmission latency. Given the common focus of these studies on delay-sensitive and task-sensitive, transmissions with URLLC are performed over finite blocklength (FBL) codes [6]. Unlike traditional Shannon theory, Finite Blocklength refers to the technique of transmitting data with finite coding, which also results in transmissions that are no longer arbitrarily reliable. Therefore, in order to ensure URLLC, the length of the encoding is important.

The demand for the application of physical layer security (PLS) in ultra-reliable low-latency communication (URLLC) scenarios is growing steadily [7]. In practice, PLS in URLLC involves deploying advanced encryption techniques at the physical layer to secure communication channels against eavesdropping. This ensures the confidentiality of sensitive data in critical applications, such as industrial automation and healthcare, meeting stringent latency and reliability demands [7].

### 1.1 Related work and discussion

The adoption of PLS in MEC is mainly achieved through the introduction of emerging technologies. [8] introduces a novel deep-learning (DL)-based physical (PHY) layer authentication scheme in MEC networks. [9] explores physical layer security in multi-access edge computing (MEC) systems employing non-orthogonal multiple access (NOMA), amidst potential eavesdropping threats. [10] proposes an RIS-assisted secure Mobile Edge Computing (MEC) network for enhancing task offloading security in IoT devices. [11] presents a multi-access mobile write-in scheme by considering the energy consumption, communication security and latency of unmanned vehicles in combination. These works greatly expand the application scenarios of PLS in MEC networks, but the challenges posed by URLLC for the future are not well addressed.

On the other hand, finite blocklength coding in MEC networks greatly expands the application scenarios of URLLC. A bound on the transmission rate is proposed in [12]. Based on this, [13] optimizes system throughput in URLLC by jointly optimizing block and pilot lengths, considering latency and block error probability constraints. By utilizing finite blocklength transmission [14] propose a reconfigurable intelligent surface (RIS)-assisted rate-splitting multiple access to support URLLC.

[15] investigate the scheme to maximize the transmission rate for a finite blocklength while minimizing the channel blocklength. In addition, issues regarding the allocation of resources (e.g., transmit power, bandwidth, frequency, etc.) have also been widely discussed [16–18]. These works provide a sufficient theoretical basis for the implementation of URLLC in the FBL regime.

Furthermore, noting that applying PLS to URLLC scenarios will greatly enhance the communication quality of MEC networks, there are many efforts working in this direction. [19] focuses on the trade-off between network throughput and reliability, security. Considering the presence of randomly distributed eavesdroppers, [20] explores physical layer security in a large-scale wireless sensor network (WSN) with random multiple access. [21] introduces a novel method for secure and covert broadcast communication in multi-user downlink ultra-high reliability and low latency communications (URLLC). Utilizing artificial noise (AN) for the first time, the approach enhances physical layer security (PLS) while considering covertness constraints amidst a multi-antenna malicious warden (Willie) and eavesdropper (Eve). [22] proposes schemes to tackle proactive eavesdropping in short packet communications, enhancing wireless information surveillance. These works provide excellent PLS schemes based on the traditional Shannon coding theorem, but in URLLC scenarios, FBL codes perform much better. In fact, the finite blocklength regime can take care of both URLLC and PLS, specifically, we can trade off reliability for overall security by choosing an appropriate transmission blocklength, a theory proposed in [6]. Utilizing FBL regime, [23] propose an antenna transmit power minimization scheme for PLS in URLLC scenarios employing intelligent reflecting surface (IRS) technology. Furthermore, [24] jointly optimize blocklength and pilot signal length to achieve secure URLLC in Industrial Internet.

However, due to the introduction of FBL codes, transmissions are no longer arbitrarily reliable, which leads to communication failures and even data loss. For this reason, Hybrid Automatic Repeat reQuest (HARQ) needs to be introduced to improve the system reliability [25, 26]. HARQ provides a more flexible transmission scheme for the system by providing an automatic retransmission request mechanism. [27, 28] utilize HARQ, which provides flexible allocation options to improve resource utilization. Specifically, [27] provides a group-based preallocation scheme combined with HARQ to effectively reduce the average user latency. [28] utilizes Non-Orthogonal Multiple Access (NOMA) FBL codes and HARQ techniques to ensure reliability while reducing average user latency.

However, it must be mentioned that the introduction of the retransmission mechanism will in turn increase the risk of information leakage. Therefore, it is important to rationally allocate resources in order to balance security and reliability. Moreover, the problems of utilizing HARQ to deploy PLS to UPLLC scenarios have not been resolved. In addition, the extra energy consumption due to retransmission should also be taken into account. In particular, the trade-off between energy consumption and communication security based on finite blocklength and retransmission mechanism should be highlighted. This manifests itself in the expectation of smaller blocklength and fewer retransmissions in terms of energy consumption, while security requires larger blocklength and more retransmissions.

## 1.2 Motivation and contribution

Motivated by above observations, in this paper, we adopt the MEC networks with retransmission capabilities that support task offloading with the goal of minimizing the expected total energy consumption of legitimate users. In addition, we measure the secure-reliability of a system in terms of leakage-failure probability (LFP), which indicates that the message may not have been decoded by the legitimate recipient or may have been successfully decoded by the eavesdropper. Since the non-convexity of the problem makes optimization difficult, we provide an optimization framework to reformulating the original problem as a mixed integer convex problem. Our contribution in this work can be summarized as follows

- Reveal the trade-off between the overall security of the system and the energy consumption of legal transmission under the retransmission mechanism, which affirms the plausibility of our optimization problem.
- A framework is designed to describe and analyze the trade-off between energy consumption and security by formulating this trade-off as a convex optimization problem.
- To solve the optimization problem, the original problem is transformed into a mixed-integer convex problem by decomposing the subproblems, and the problem is optimized by determining the blocklength and maximum number of retransmissions to be transmitted.
- Through numerical simulations, we validate our analytical results and evaluate the performance of the proposed method, while demonstrating energy consumption versus overall LFP with channel gain, transmit power, and data size. Our results show that retransmission is 33.1% more energy efficient than no retransmission and FBL coding is 13% more energy efficient than IBL coding.

The rest of this paper is organized as follows. In Sects. 2 and 3, we discuss the system model and characterize the metrics. Then, in Sect. 4, we give an optimization framework and basic proofs aimed at minimizing the overall LFP and transmission energy consumption. Next Sect. 5 evaluates the proposed design, while Sect. 6 summarizes the entire work. In addition, Sects. 7 and 8 give the lists and statements for the article.

## 2 System model

We utilize a standard communication system comprising three nodes. In this scenario, the user equipment, represented by Alice, tries to transmit confidential information to the MEC server, known as Bob. However, an eavesdropper, denoted as Eve, is actively trying to intercept and decipher their messages. Alice sends sensitive information to Bob within a fixed duration  $T$ , i.e., both communication and computation between Alice and Bob must be completed in  $T$ . Due to secure-reliability requirements, Bob's total decoding error probability of Bob should satisfy  $\epsilon_{b,tot} \leq \epsilon_{b,max}$ , while Eve's total decoding error probability of Eve need to satisfy  $\epsilon_{e,tot} \geq \epsilon_{e,min}$ . Note that to ensure optimal system performance, these constraints should be satisfied by each transmission.

The system adopts time-slotted model, dividing time into frames of length  $T$ . Within each frame, there is a communication phase to upload the data from Alice and a computation phase to execute the task at Bob. In the communication phase, each transmission is performed with a blocklength of  $m$  and a fixed amount of data  $d_{\text{bits}}$ . Assume that the channel between Alice and Bob (Eve) undergoes quasi-static fading, meaning that the channel state information remains constant within a frame but varies between frames. Denote the channel gain of the link by  $z$ , assuming that the server is aware of the gain. Then, we can provide Alice's channel parameters to Bob and Eve, respectively. Start with the signal-to-noise ratio (SNR):

$$\gamma_b = \frac{\phi_b z_b P_{\text{Alice}}}{\sigma_b^2}, \quad (1)$$

$$\gamma_e = \frac{\phi_e z_e P_{\text{Alice}}}{\sigma_e^2}, \quad (2)$$

where  $P_{\text{Alice}}$  implies the transmit power of the Alice,  $\phi$  is the channel path-loss which is related to the distance between nodes, i.e.,  $r_{\text{Bob}}$  and  $r_{\text{Eve}}$ .  $\sigma^2$  is the noise power.

Due to the application of finite blocklength codes for transmission, errors may occur during transmission. And because of the strict latency constraint, Bob sends Alice a Negative Acknowledgment (NACK) with a fixed small data volume of  $d_{\text{NK}}$  bits within a fixed duration  $t_{\text{NK}}$ , and the power of transmitting NACKs is represented by  $P_s$ . The probability in unsuccessfully decoding NACK at Alice is represented by  $\nu$ . After Alice successfully decodes the NACK, it resends the data until Bob either successfully decodes data packet or reaches the maximum allowed number of (re)transmissions. We denote the duration of a single transmission by  $t$ , and the time length of a symbol is represented by  $T_s$ . Therefore, transmission blocklength  $m = \frac{t}{T_s}$ , meanwhile, the total time duration of the transmission phase is  $(n + 1)t$ , while  $n=0$  represent initial transmission.

Utilizing Dynamic Frequency and Voltage Scaling (DVFS) technology [29, 30], Bob (edge sever) can flexibly adjust the CPU operating frequency  $f$  in each frame to the requirements of the mission at hand. Computation time  $t_c$  is determined by the CPU frequency  $f$ , i.e.,  $t_c = \frac{c}{f}$  with  $0 \leq f \leq f_{\text{max}}$ , where  $c$  represents computation cycles for the task while  $f_{\text{max}}$  is the maximum available CPU frequency (Fig. 1).

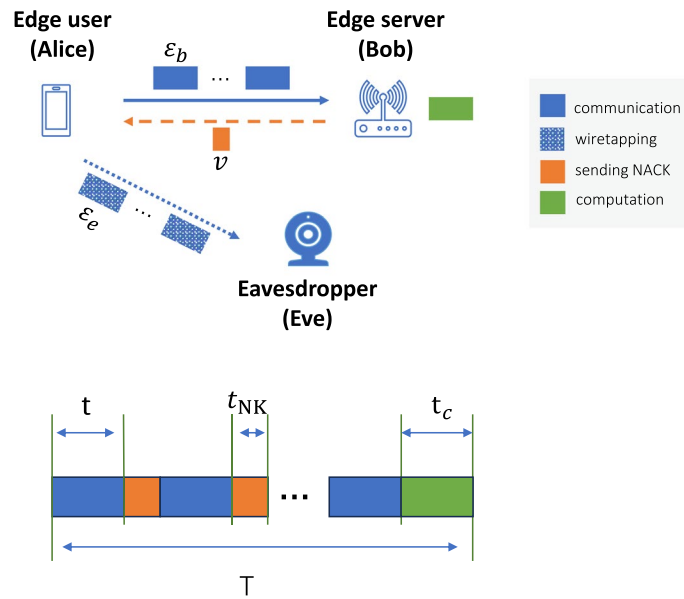
### 3 Characterizations on the leakage failure probability and the total energy consumption

#### 3.1 Characterization of END-to-END leakage-failure probability

##### 3.1.1 End-to-end error probability of decoding in FBL regime

During each communication phase (when a message arrives), Alice transmits data packets of a fixed size  $d$  through a finite blocklength code with a length of  $m$ . Then, denote the corresponding coding rate as  $r = \frac{d}{m}$ . Following the FBL regime in [12], the error probability of the  $n^{\text{th}}$  (re)transmission is given as follows:

$$\epsilon = P(r, \gamma, m) \approx Q\left(\sqrt{\frac{m}{V(\gamma)}}(C(\gamma) - r) \ln 2\right), \quad (3)$$



**Fig. 1** System Model ( where transmissions between Bob and Alice are called legitimate transmissions, and Eve makes eavesdropping attempts (initial transmission and retransmission) on each transmission until the eavesdropping is successful or the maximum number of retransmissions is reached)

where  $C(\gamma) = \log_2(1 + \gamma)$  represents the Shannon Capacity in [12]. Denote channel dispersion by  $V$ , which satisfies  $V(\gamma) = 1 - (1 + \gamma)^{-2}$  in a complex AWGN channel. To characterize security, we adopt the metric from [31], i.e., the security-reliability of a single transmission is evaluated by LFP. We denote the leak-failure event by  $Y$ , i.e., Bob fails to decode or Eve succeeds in decoding. Denote Bob’s successful decoding by  $X_b$  and Eve’s successful decoding by  $X_e$  independently of each other. Thus, we can denote the LFP by  $\epsilon_{LF} = P(Y = 1) = P(X_b = 0 \cup X_e = 1)$ . Due to the application of FBL, both Bob and Eve’s decoding could be incorrect. According to (3), the probability of decoding error for Bob and Eve can be expressed, respectively, as follows.

$$\epsilon_b = P(X_b = 0) = Q\left(\sqrt{\frac{m}{V(\gamma_b)}}(C(\gamma_b) - r) \ln 2\right), \tag{4}$$

$$\epsilon_e = P(X_e = 0) = Q\left(\sqrt{\frac{m}{V(\gamma_e)}}(C(\gamma_e) - r) \ln 2\right). \tag{5}$$

Hence, according to [6], we can denote the LFP by:

$$\begin{aligned} \epsilon_{LF} &= P(Y = 1) = P(X_b = 0 \cup X_e = 1) \\ &= 1 - (1 - \epsilon_b)\epsilon_e = \epsilon_b\epsilon_e + (1 - \epsilon_e). \end{aligned} \tag{6}$$

Note that LFP takes both security and reliability into account simultaneously, so we characterize reliability constraints and security constraints uniformly with LFP,

$\epsilon_{LFP,tot}$ .

### 3.1.2 Overall error probability of Bob and Eve in communication phase

To begin with, we model the total error probability between Bob and Eve, which also affects the expected energy consumption of the system.

$N=0$ : Retransmission is not allowed. We have

$$\epsilon_{b,tot} = \epsilon_b. \quad (7)$$

$N \geq 1$ : Error occurs when the Bob fails to decode in the  $n^{th}$  (re)transmission, i.e., Bob decodes  $(n-1)^{th}$  message mistakenly or Alice decodes NACK wrongly. According to [31], we denote  $P_v(n)$  as the decoding error probability of the  $n^{th}$  (re)transmission due to unsuccessfully decoding NACK at the Alice and let  $P_{\epsilon_b}$  denote the decoding error probability of the  $n^{th}$  (re)transmission due to the failure of decoding at the Bob. Within  $N$  maximal allowed retransmission attempts, we can give the Alice-to-Bob decoding error probability:

$$\epsilon_{b,tot} = \sum_{n=1}^N \epsilon_b^n (1-\nu)^{n-1} \nu + \epsilon_b^{N+1} (1-\nu)^N. \quad (8)$$

Secondly, we delve into characterizing the total decoding error probability of Eve. Here, we make the assumption that Eve cannot interfere with the retransmission mechanism by sending fake NACKs. However, it is crucial to note that events where Alice fails to decode the NACK or the transmission itself fails also influence Eve's overall decoding error probability. In other words,  $\epsilon_{e,tot}$  is contingent upon both  $\epsilon_b$  and  $\nu$  in each transmission attempt. Consequently,  $\epsilon_{e,tot}$  can be expressed as follows:

$$\epsilon_{e,tot} = \epsilon_e^{N+1} \epsilon_b^N (1-\nu)^N + \sum_{n=1}^{N-1} \left( \epsilon_e^{N-n} \epsilon_b^n (1-\nu)^n \nu \right). \quad (9)$$

Finally, because LFP is a metric that characterizes the overall secure-reliability of the system, we can calculate  $\epsilon_{LFP,tot}$  based on (6):

$$\epsilon_{LFP,tot} = \epsilon_{b,tot} \epsilon_{e,tot} + (1 - \epsilon_{e,tot}). \quad (10)$$

### 3.1.3 Total energy consumption between Bob and Alice

In practice, what we are primarily concerned with is the energy consumption associated with legitimate transmissions, specifically those between Bob and Alice, as opposed to those involving Eve. The energy consumption between Bob and Alice comprises three components: Alice's energy consumption for transmitting packets, denoted as  $E_t$ ; the energy consumed by Bob to transmit NACK to Alice, denoted as  $E_k$ ; and the energy consumption of Bob for computation after successful decoding, denoted as  $E_c$ . Next, we characterize the energy expectations of these three components.

**(a) Energy for transmitting data packet at Alice:** Alice's expected energy consumption for transmitting a data packet is influenced by the number of (re)transmissions and the decoding error probability of NACK. We calculate the initial transmission ground energy consumption as  $E_{t,0} = tP_{Alice} + E_{d,B}$ , where  $E_{d,B}$  represents Bob's decoding energy consumption (typically a small data quantity). Note that the initial transmission always occurs regardless of  $N$ . Then, the decoding error probability of the  $n^{th}$  transmission, along with the

decoding error probability of the NACK, determines the expected energy consumption of the  $(n + 1)^{\text{th}}$  transmission. Therefore, we can provide:

$$\begin{aligned}\bar{E}_t &= E_{t,0} + \epsilon_b(1 - \nu)E_{t,0} + \dots + \epsilon_b^N(1 - \nu)^N E_{t,0} \\ &= \sum_{n=0}^N \epsilon_b^n(1 - \nu)^n E_{t,0}.\end{aligned}\quad (11)$$

**(b) Energy consumption for delivering NACKs at Bob:** Bob's energy consumption for sending NACKs back to Alice is  $E_{k,0} = t_{\text{NK}}P_{\text{Bob}} + E_{d,A}$ , where  $E_{d,A}$  represents the constant energy consumption of Alice to receive NACKs. However, in the initial transmission,  $E_k = 0$ . i.e., there are no NACKs sent at this time. NACKs are only sent when a decoding error occurs at Bob. Then, we can characterize the expected energy consumption for sending a single NACK as  $\epsilon_b E_{k,0}$ ; hence, the energy cost of the  $n^{\text{th}}$  NACK is  $\epsilon_b^{n+1}(1 - \nu)^n E_{k,0}$ . Therefore, the overall expected energy consumption for transmitting NACKs can be calculated as follows:

$$\begin{aligned}\bar{E}_k &= \epsilon_b E_{k,0} + \epsilon_b^2(1 - \nu)E_{k,0} + \dots + \epsilon_b^{N+1}(1 - \nu)^N E_{k,0} \\ &= \sum_{n=0}^N \epsilon_b^{n+1}(1 - \nu)^n E_{k,0}.\end{aligned}\quad (12)$$

**(c) Energy cost for computation:** The nonlinear energy model in [29] is introduced, given by the following equation:

$$E_c = \kappa c f^2 = \kappa c^3 t_c^{-2}, \quad (13)$$

where  $\kappa$  is a permanent constant which is related to the architecture of the hardware. Recall that the duration for data transmission is  $(n + 1)t$ . Combined with the duration for sending NACKs  $nt_{\text{NK}}$ , the duration of the computation phase is  $t_c^{(n)} \leq T - (n + 1)t - nt_{\text{NK}}$ , since  $E_c$  is monotonically increasing with  $t_c$ . The minimum energy consumption requires that the equation always holds. Therefore, we can determine the  $n^{\text{th}}$  energy consumption for computing:

$$E_c^{(n)} = \kappa c^3 \frac{1}{(T - (n + 1)t - nt_{\text{NK}})}, \quad (14)$$

$n=0$  represents the calculation after the initial transmission, whose probability is  $1 - \epsilon_b$ . The  $n^{\text{th}}$  retransmission occurs when the  $n$  attempts fails and the  $n^{\text{th}}$  NACK is correctly decoded, i.e., the probability of success of the  $n^{\text{th}}$  retransmission is  $\epsilon_b^n(1 - \nu)^n(1 - \epsilon_b)$ . In order to meet the demand for ultra-reliability, we have  $\epsilon_b \ll 1$  and  $\nu \ll 1$ . Then, the expected energy of computation is

$$\begin{aligned}\bar{E}_c &= (1 - \epsilon_b)E_c^{(0)} + \epsilon_b(1 - \nu)(1 - \epsilon_b)E_c^{(1)} + \dots + \epsilon_b^N + (1 - \nu)^N(1 - \epsilon_b)E_c^{(N)} \\ &= E_c^{(0)} - \epsilon_b^{N+1}(1 - \nu)^N E_c^{(N)} + \sum_{n=1}^N \epsilon_b^n(1 - \nu)^{n-1} \left( (1 - \nu)E_c^n - E_c^{(n-1)} \right) \\ &\approx E_c^{(0)} + \sum_{n=1}^N \epsilon_b^n(1 - \nu)^{n-1} \left( E_c^{(n)} - E_c^{(n-1)} \right).\end{aligned}\quad (15)$$



Recall that we are considering a URLLC scenario, which requires high reliability of the transmission. Therefore decoding satisfies  $\nu \ll 1$  and  $\epsilon_b \ll 1$ , which makes the final approximate equation holds. Hence, the expected total energy consumption for legitimate transmission is the sum of the above three parts by  $\bar{E}_{b,\text{tot}} = \bar{E}_c + \bar{E}_k + \bar{E}_t$ .

So far, we characterize both the total energy cost of legal transmission (between Bob and Alice) and the total LFP (stands for overall secure-reliability).

#### 4 Optimal framework design

Our goal is to minimize the expected energy consumption of the legitimate transmission within the system. Next, we design a framework with the legitimate transmission energy as the optimization objective. Because of the symmetry of the problem, this framework also applies when the overall LFP is used as the optimization objective.

##### 4.1 Problem formulation

In the system, we aim to minimize the expected overall energy consumption by optimizing the transmission blocklength  $m$  and the retransmission number  $N$ , while ensuring that the overall secure-reliability satisfies  $\epsilon_{\text{LFP,tot}} \leq \epsilon_{\text{max}}$ . Hence, we have:

$$\min_{m,N} \bar{E}_{b,\text{tot}} \quad (16)$$

$$s.t. \quad \epsilon_{\text{LFP,tot}} \leq \epsilon_{\text{max}}, \quad (16a)$$

$$t_c^{(n)} + (n+1)t + nt_{\text{NK}} = T, \quad (16b)$$

$$\frac{c}{f_{\text{max}}} + (N+1)t - Nt_{\text{NK}} \leq T, \quad (16c)$$

$$N \in \mathbb{Z}. \quad (16d)$$

Here, constraint (16a) indicates the secure-reliability constraints of the system.<sup>1</sup>

##### 4.2 Optimal solution

To solve the optimization problem (16), we first split the original problem into  $N_{\text{max}}$  subproblems, where  $N_{\text{max}}$  is the upper limit of feasible domains of  $N$ . Next, we conclude  $N_{\text{max}}$ , which restricts the total number of subproblems. Furthermore, we formulate the subproblem and transform the original problem into a solvable convex problem.

###### 4.2.1 Original problem decomposition

Recall that  $N \leq N_{\text{max}}$  is a positive integer, there are  $N_{\text{max}}$  possible outcomes associated with the frame retransmission event. Once given  $N \in [0, 1, 2, \dots, N_{\text{max}}]$ , we have subproblem as follows:

<sup>1</sup> For practical systems, it may be necessary to have separate constraints for reliability and safety, i.e.,  $\epsilon_{b,\text{tot}} \leq \epsilon_{b,\text{threshold}}$  and  $\epsilon_{e,\text{tot}} \geq \epsilon_{e,\text{threshold}}$ .

$$\min_{m, N} \bar{E}_{b, \text{tot}} \quad (17)$$

$$s.t. \quad (16a), (16b), (16c), 16d. \quad (17a)$$

#### 4.2.2 Upper bound of $N_{\max}$

The maximum number of retransmissions is limited by the computational power of Bob (edge server). Particularly, with analyzing the constraints (16b) and (16c), an upper bound for  $N_{\max}$  can be given as

$$N_{\max} \leq \lfloor \frac{T - \frac{c}{f_{\max}} - t}{t + t_{\text{NK}}} \rfloor, \quad (18)$$

where  $\lfloor \dots \rfloor$  is for round down.

#### 4.2.3 Optimal solution for subproblem

When determining a maximum number of retransmissions  $N$ , we can solve the subproblem with the following lemma.

**Lemma 1** *The expected total energy consumption for legitimate transmission  $\bar{E}_{b, \text{tot}}$  is convex in  $m$ .*

**Proof** See Appendix A. □

Here, the objective function of problem (17) is a convex function of blocklength, and our next task is to prove the convexity or linearity of the constraints in  $m$ . Furthermore, we prove that the constraint (16a) is convex in  $m$  by the following three lemma:

**Lemma 2** *The total decoding error probability of Bob  $\epsilon_{b, \text{tot}}$  is convex in the blocklength of a single (re)transmission  $m$ .*

**Proof** See Appendix B. □

**Lemma 3** *The total error probability of Eve  $\epsilon_{e, \text{tot}}$  is concave in the blocklength of a single (re)transmission  $m$ .*

**Proof** See Appendix C. □

**Lemma 4** *The total LFP  $\epsilon_{\text{LFP}, \text{tot}}$  is convex in the blocklength of a single (re)transmission  $m$ .*

**Proof** See Appendix D. □

Moreover, (16b) and (16c) are linear in  $m$ . Thus, we prove that the constraints are either convex or linear in  $m$ .

#### 4.2.4 Problem reformulation

Based on Lemmas 1, 4 in the previous section and the upper bounds of  $N_{\max}$ , we formulate the original problem as

$$\min_{m, N} \bar{E}_{b, \text{tot}} \quad (19)$$

$$s.t. \quad (16a), (16b), (16c), (16d), \quad (19a)$$

$$N_{\max} \leq \left\lfloor \frac{T - \frac{c}{f_{\max}} - t}{t + t_{\text{NK}}} \right\rfloor. \quad (19b)$$

Combining Lemma 1, 2, 3, 4 and the proofs, the objective function is convex in  $m$ . Moreover, all constraints are either convex in  $m$  or linear. Based on these, the problem we studied also became the mixed convex integer problem, which can be solved utilizing mixed integer convex optimization methods [32].

## 5 Numerical evaluation

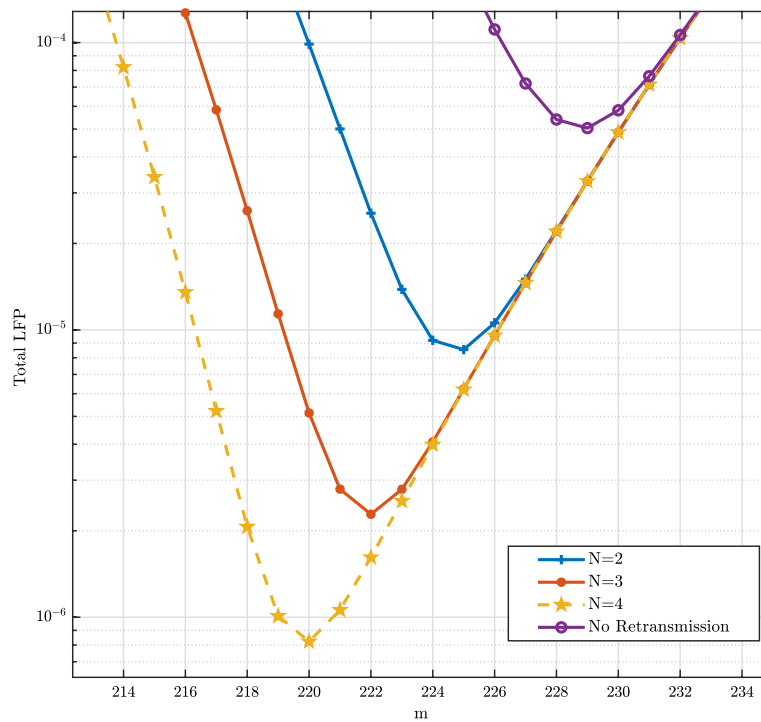
### 5.1 Parameter setup

This section sets the model parameters to verify the analytical results through numerical simulations and study the performance of the system. First of all, the parameters are set as follows: The amount of data sent by Alice is set to  $d = 320$  bits. Considering the difference in channel quality, the distance between the legitimate receiver (Bob) and the marginal user (Alice) is considered as  $r_{\text{Bob}} = 130$  m, while the eavesdropper (Eve) is  $r_{\text{Eve}} = 160$  m away from Alice. According to [33], we can give the path-loss model as  $\text{PL} = 17.0 + 40.0 \log_{10}(x)$ . Moreover, the span of each frame is given as  $T = 60$  ms, while each symbol length is  $T_s = 0.04$  ms. What is more, the noise of both is also given by  $\sigma_b^2 = -167$  dBm,  $\sigma_e^2 = -164$  dBm. Next, the transmission channel bandwidth is considered as  $B = 5$  MHz, and  $t_{\text{NK}} = 3$  ms for sending NACK. Additionally, we let the upper limit of CPU frequency  $f_{\max} = 3.5$  GHz and CPU total workload  $c = 20$  Mcycles,  $\kappa$  is hardware constants, we let it as  $\kappa = 10^{-11}$ . Finally, we constrain the problem to an ultra-safe scenario, i.e., we set  $\epsilon_{\max} = 0.001$  as the maximal allowed total error probability,  $\epsilon_{e, \min} = 0.999$  for the bound of decoding error probability about Eve, and denote  $\nu = 0.00001$  as the Alice's probability of an error in decoding NACK. The values of the parameters are listed in Table 3.

### 5.2 Overall LFP of the system

To start with, we evaluate the convexity of the leakage-failure probability (LFP) with respect to blocklength  $m$  in Fig. 2, which is one of the key analytical findings in Lemma 4. In particular, we plot LFP against  $m$  with different setups of retransmission numbers  $N = 2, 3, 4$  and no retransmission.

Clearly, the convexity can be observed in every setups within the feasible set, which confirms the results in Lemma 4. However, different retransmission numbers result in different minimum of LFP. It implies the importance of jointly design of retransmission



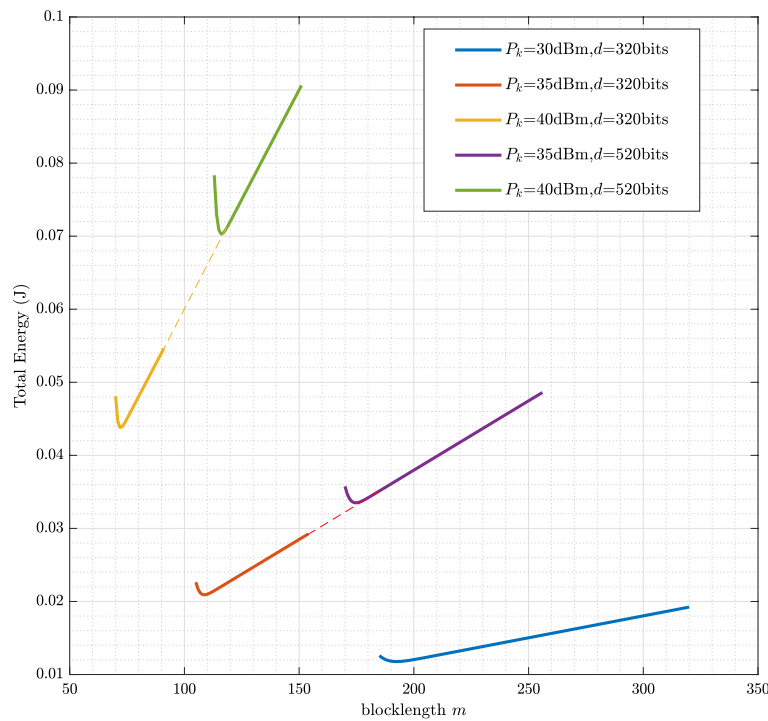
**Fig. 2** Total Probability of LFP vs Blocklength  $m$  for different retransmission attempts

number and blocklength for the system. Interestingly, LFP increases and convergences eventually with the increase of  $m$  regardless of  $N$ . This is because that the decoding error probability of Bob and Eve both monotonically decreasing in  $m$ . When blocklength is sufficiently long, the initial transmission is sufficiently reliable for Bob, but also likely to be leaked to Eve, while the rest of transmissions play insignificant role in LFP. Moreover, it can be observed that whether it is a long blocklength or a short blocklength, there will be a inflection point, which can be explained that when the blocklength is short, although the leakage probability is small, the probability of Bob decoding failure will be relatively large, resulting in a short blocklength while the LFP is high, and this situation improves as the blocklength increases. However, when the blocklength is very long, the probability of Eve decoding success becomes larger, resulting in a significant increase in total LFP. This convexity suggests that the accuracy of blocklength  $m$  selection greatly affects system performance. Furthermore, by comparing retransmission with no retransmission, we observed that the retransmission mechanism is of great assistance in improving the transmission errors caused by the FBL regime.

### 5.3 Total energy consumption between legitimate edge server (Bob) and user (Alice)

To assess the system performance comprehensively, we present a series of curves depicting the variation of key parameters as follows:

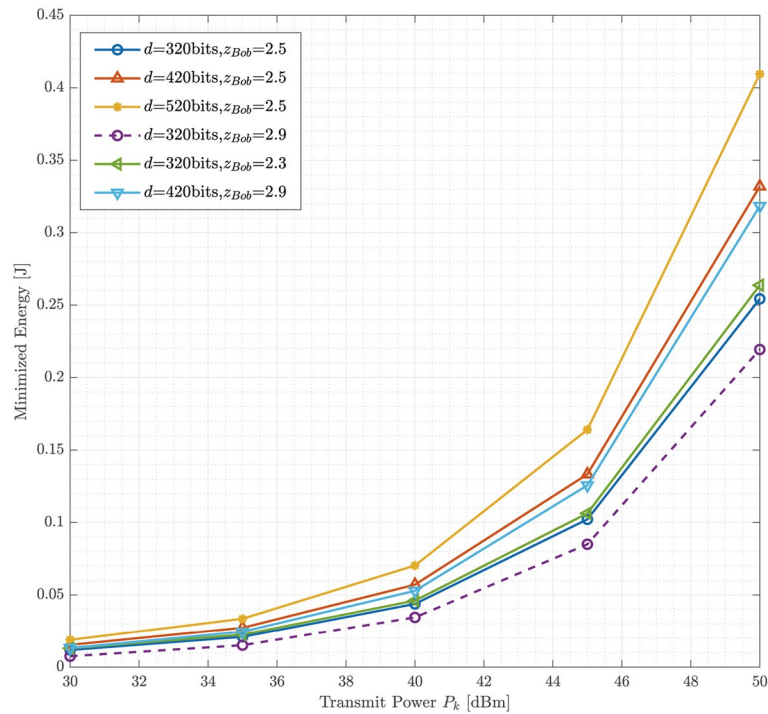
First of all, Fig. 3 displays total energy cost versus  $m$  with variant transmit power and data packet size. The curves illustrate that  $\bar{E}_{b,tot}$  is convex in  $m$ , which is consistent with Lemma 1. From the figure, we observe that transmit power and data size significantly affect energy consumption, higher transmit power and larger data packet lead to greater



**Fig. 3**  $\bar{E}_{b,tot}$  vs Blocklength  $m$  with Different Transmit Power  $P_k$  and Data Packet Size  $d$  (The feasible domains of the individual images are shorter, but when filling in the infeasible regions, we can find a correlation between them, because even if the LFP is large, the mathematics can still work out the corresponding energy consumption)

energy cost. Furthermore, it is intriguing to note that increased transmit power enhances the sharpness of the convexity curves. Specifically, higher transmit power increases the likelihood of successful decoding by Eve, bringing the leakage-failure probability (LFP) closer to its upper limits. This, in turn, tightens the constraints, ultimately leading to a sharper convexity. In essence, enforcing strict secure-reliability constraints enhances the convexity’s sharpness.

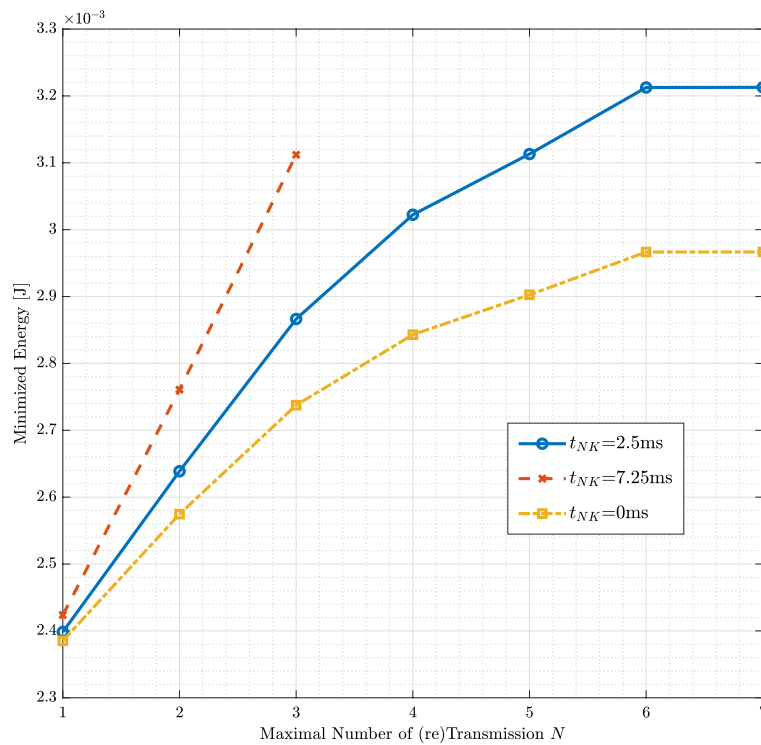
Subsequently, we demonstrate the optimal total energy consumption  $\bar{E}_{b,tot}^*$  versus transmit power  $P_k$  with varying sizes of data packets and channel gains. In Fig. 4, we observe that the optimal energy consumption  $\bar{E}_{b,tot}$  increments with the transmit power  $P_k$ , regardless of the other parameter settings which is due to the fact that transmit power is positively associated with  $\bar{E}_{b,tot}$ . In addition, larger data sizes and smaller channel gains for Bob tend to increase optimal energy consumption. This occurs because a greater channel gain enables the system to transmit information with fewer retransmission attempts, while smaller data sizes allow for transmission in shorter blocklength, consequently reducing transmission times. Therefore, to optimize energy consumption, it is preferable to choose smaller data sizes and larger channel gains. Furthermore, the escalation of transmit power accentuates the gap between channel gain and data size. This observation implies that enhancing channel gain and reducing data volume will have limited impact on optimizing energy consumption, particularly when the transmit power is constrained.



**Fig. 4** Optimal  $\bar{E}_{b,tot}^*$  vs Transmit Power with variant data size  $d$  and channel gain of Bob  $z_{Bob}$

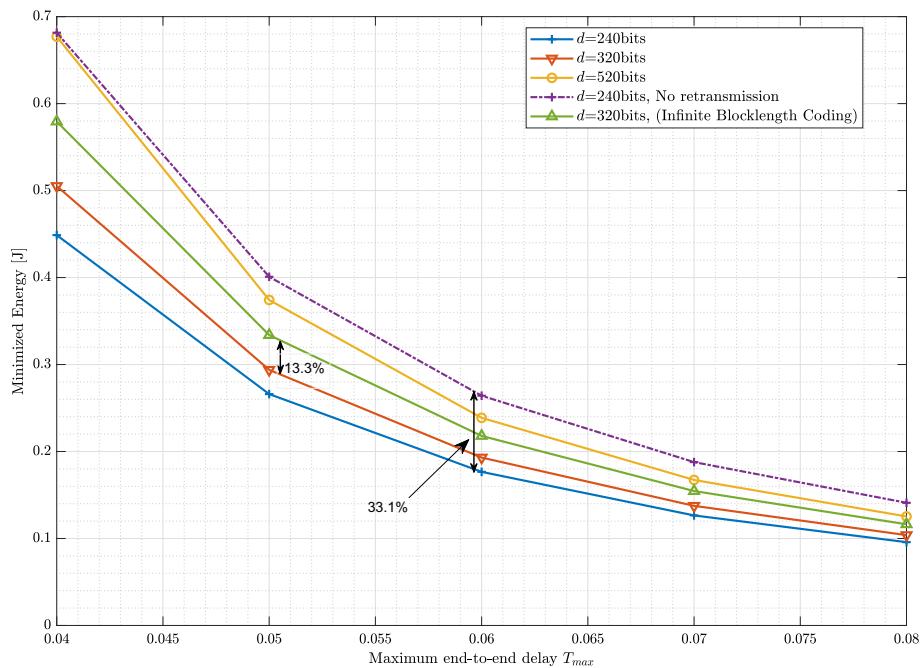
To assess the effect of the retransmission mechanism within the system design, we present the optimal energy cost  $\bar{E}_{b,tot}^*$  against the maximum number of retransmissions  $N$ , considering various NACK-transmission duration  $t_{NK}$ . This comparison is visualized in Fig. 5. Since each retransmission imposes additional energy cost on the system, the system tends to favor fewer retransmissions  $N$  when  $t_{NK}$  is large. From our simulation, we can observe that when the number of retransmissions  $N > 3$ , there is no feasible solution for problem (19). On the other hand, if  $t_{NK}$  is small, the smaller the number of retransmission attempts, the better the system performance will be, as it will result in lower energy consumption per retransmission. The bottom curve demonstrates the extreme case of  $t_{NK} = 0$  ms, which means that NACKs are sent instantaneously. Due to the practical limitations of the CPU’s maximum operating frequency  $f_{max}$ , the growth rate of energy consumption slows down significantly when the number of retransmissions is large because the upper limit has been reached. Furthermore, if  $t_{NK}$  is too high, the system will not be able to furnish computation services within the maximum allowable delay  $T_{max}$  while fulfilling the other constraints. As a result, we observe that the number of feasible retransmissions  $N$  is small when  $t_{NK}$  is large.

Moreover, we are interested in understanding the influence of the maximum allowable delay  $T_{max}$  on the total energy consumption of the system. The curves are depicted in Fig. 6. When considering various data sizes, we observe a decrease in minimum energy consumption as  $T_{max}$  decreases, aligning with the corollary of Equation (14). Specifically, strict delay constraints compel the system to adopt longer blocklength to diminish the number of retransmissions. Conversely, loose delay



**Fig. 5** Optimal  $\bar{E}_{b,tot}^*$  vs Maximal allowable (re)transmission attempts  $N$  with various transmission duration of NACK  $t_{NK}$

constraints permit more retransmission attempts with shorter blocklength and higher CPU operating frequency. Moreover, due to the limitation of CPU computing capacity, larger data sizes may exceed the system’s ability to complete data transmission and calculations within  $T_{max}$ . Consequently, for larger data sizes, the feasible domain in  $T_{max}$  contracts. Particularly noteworthy is the significant reduction in the feasible domain for  $d = 520$  bits. Furthermore, we have also investigated the efficacy of the no retransmission mechanism alongside infinite blocklength codes within the URLLC scenario. Our curves indicate a superior performance of the retransmission mechanism over no retransmission, particularly in scenarios with stringent delay constraints. On the other hand, we compare the difference between infinite blocklength (IBL) coding and FBL coding. Since the optimal coding rate in the IBL regime is the Shannon capacity  $C$ , the optimal blocklength for transmission is  $m = \frac{d}{C}$ . According to (3), the error probability of transmission is  $Q(0)=0.5$ , which is much higher than the URLLC requirement, thus more retransmission is needed. As a result, the benefits of our design in terms of reduced energy consumption and improved overall secure-reliability are realized. Moreover, for quantitative analysis, we further demonstrate the optimization percentage for different designs, i.e.,  $\frac{E_{b,tot,IBL} - E_{b,tot,FBL}}{E_{b,tot,IBL}} \times 100\%$  and  $\frac{E_{b,tot,No-Retrans} - E_{b,tot,Retrans}}{E_{b,tot,No-Retrans}} \times 100\%$ . We observe 33.1% improvement in retransmission over no retransmission energy consumption and 13% improvement in FBL coding over IBL for the same frame length and data size.

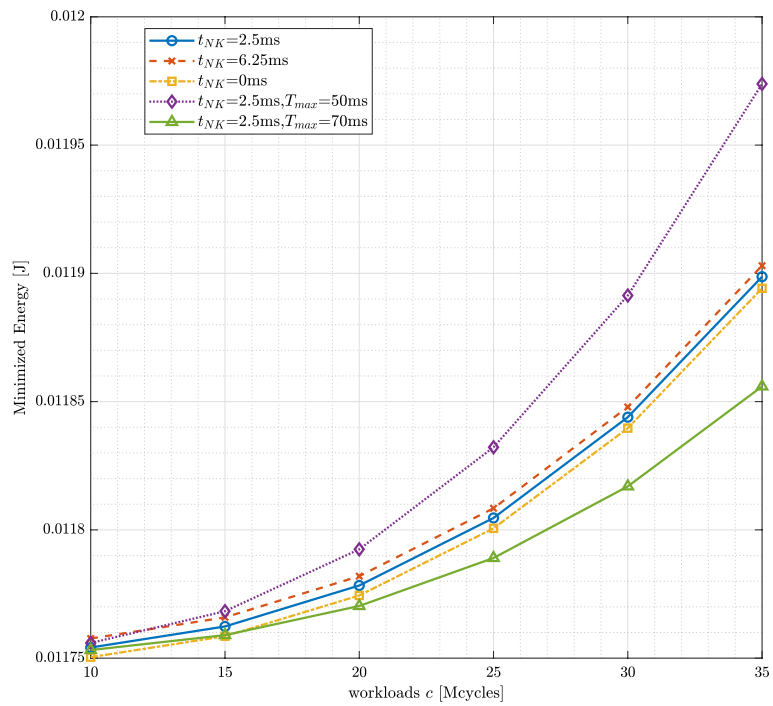


**Fig. 6** Optimal  $\bar{E}_{b,tot}^*$  vs Maximum end-to-end delay  $T_{max}$  within different data size and mechanisms (We observe 33.1% improvement in retransmission over no retransmission energy consumption and 13% improvement in FBL coding over IBL for the same frame length and data amount.)

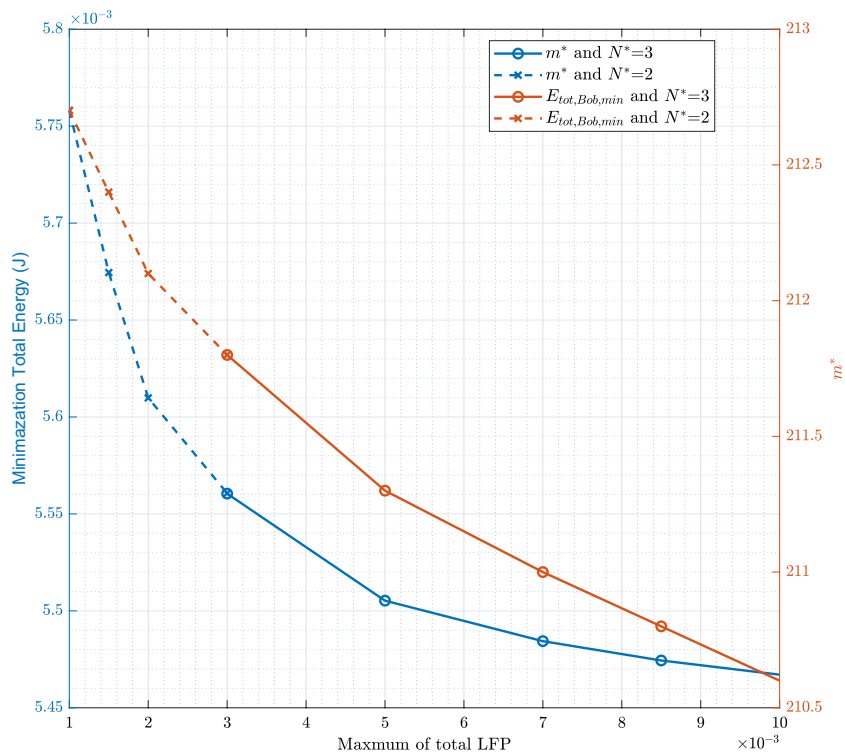
Then, we demonstrate the optimal total energy cost  $E_{b,tot}^*$  versus the CPU workload  $c$  and the NACK- transmission duration  $t_{NK}$ . In Fig. 7, an increase in workload  $c$  also increases the energy cost, which is because  $E_c$  is incremental in  $c$ . In addition, we observe that  $T_{max}$  significantly affects the energy consumption, with larger  $T_{max}$  leading to smaller energy consumption, which is consistent with our previous findings in Fig. 6. However, there is little difference between various setups of  $t_{NK}$  and  $T_{max}$  if the value of  $c$  is small. This implies that when the workload is very small, increasing the NACK time and frame length does little to optimize energy cost. This means that the computational capacity of MEC servers is crucial for the optimization of energy consumption.

Finally, we demonstrate the relationship between energy consumption minimization and the optimal blocklength and overall LFP  $\epsilon_{LFP,tot,max}$  in Fig. 8. The curves illustrate that stricter security constraints necessitate longer transmission blocklength (i.e., large transmission duration), resulting in higher energy consumption. Additionally, the figure depicts the optimal allowed retransmission attempts  $N^*$ . The dotted line corresponds to  $N^* = 2$ , while the solid line represents  $N^* = 3$ . This allows us to interpret the curves as follows: when the target LFP is high, the system tends to prioritize shorter transmission duration, thereby reducing both computational and communication energy consumption. Consequently, it utilizes more retransmission attempts  $N$  to compensate for LFP induced by a shorter blocklength  $m$ . Moreover, with the strict target LFP, the system requires longer transmission duration to guarantee security-reliability. However, for computational energy consideration, Bob’s computation time cannot be too short. As a result, the system needs to reduce retransmission attempts  $N$  for minimizing total energy consumption while larger blocklength  $m$  for secure-reliability.





**Fig. 7** Optimal  $\bar{E}_{b,tot}^*$  versus Workloads  $c$  with different duration of transmitting NACKs  $t_{NK}$  ( $T_{max}=60$  ms for the other three)



**Fig. 8** Optimal  $\bar{E}_{b,tot}^*$  and Optimal Blocklength  $m^*$  vs Different  $\epsilon_{tot,max}$  within  $N^*$

## 6 Conclusions

In this paper, we analyze the optimization of energy consumption and secure-reliability with finite blocklength and retransmission mechanism. To demonstrate the trade-off between total energy consumption for legal transmission and security-reliability, we utilize a three-node edge network model, i.e., Mobile Edge User, Edge Server and Eavesdropper, where the transfer of information between the user and the edge server is legal. Based on this, we formulate the optimization problem and then transform it into a mixed-integer convex problem by means of decomposing subproblems, followed by providing optimization framework. In particular, we discuss energy consumption as optimization objective functions and finally verify the reasonableness of our analysis through numerical simulations and show the trade-off between security and energy consumption, along with the impact of channel gain, data size, and transmit power on the system performance. The results of the simulation show that the retransmission mechanism in the FBL regime effectively improves secure-reliability while reducing energy consumption by 33.1% compared to no retransmission. Compared to the IBL regime, the FBL regime performs 13% better in terms of energy efficiency within URLLC service.

## 7 List

Then, we demonstrate all notations and their meanings. Here,  $i$  is index for Bob and Eve (Tables 1, 2, 3).

**Table 1** List of Abbreviations

Abbreviation	Full Form
MEC	Mobile Edge Computing
LFP	Leakage-Failure Probability
NACK	Negative Acknowledgment
FBL	Finite Blocklength
URLLC	Ultra-Reliable and Low-Latency Communication
VR	Virtual Reality
PLS	Physical Layer Security
WSN	Wireless Sensor Network
AN	Artificial Noise
IRS	Intelligent Reflecting Surface
HARQ	Hybrid Automatic Repeat reQuest
NOMA	Non-Orthogonal Multiple Access

**Table 2** Notations

$T$	Frame length
$T_{\max}$	Maximum frame length
$d$	Data packet size
$m$	Length of a blocklength
$z_i$	Gain of the channel
$\gamma_i$	Signal-to-noise ratio of the channel between Alice and Bob/Eve
$\phi_i$	Channel path-loss between Alice and Bob/Eve
$P_{\text{Alice}}$	The transmit power of the Alice
$\sigma_i^2$	The noise power of the channel between Alice and Bob/Eve
$d_{\text{NK}}$	The data size of NACK
$\nu$	The probability of Alice incorrectly decoding a NACK
$\epsilon_i$	Probability of decoding error for a single transmission from Bob/Eve to Alice
$\epsilon_{i,\text{tot}}$	The total probability of decoding errors in Alice's transmission to Bob/Eve
$r_i$	The distance between Bob/Eve and Alice
$t_{\text{NK}}$	The time length for NACK-transmission
$P_s$	The power of transmitting NACKs
$T_s$	The time length of a symbol
$t$	Duration of a single transmission
$n$	Index of the number of transmissions
$t_c$	Computation time of Bob
$f$	CPU frequency of Bob
$f_{\max}$	The maximum available CPU frequency
$c$	Workload of the CPU
$\kappa$	Hardware constant of the CPU
$r$	Coding rate
$C(\gamma_i)$	Shannon capacity of the channel between Bob/Eve and Alice
$V(\gamma_i)$	Channel dispersion of the channel between Bob/Eve and Alice
$\epsilon_{\text{LFP,tot}}$	The overall Leakage-failure probability of the system
$\bar{E}_t$	Energy consumption for transmitting data at Alice
$\bar{E}_k$	Energy consumption for transmitting NACKs at Alice
$\bar{E}_c$	Energy consumption for computation of Bob
$E_k^n$	Energy consumption for $n^{\text{th}}$ computation of Bob
$N$	Number of retransmissions in a frame
$N_{\max}$	Maximum number of retransmissions for system
$N$	Maximum number of retransmissions
$E_{d,B}$	Bob's decoding energy cost
$E_{d,A}$	Alice's decoding energy cost
$\epsilon_{\max}$	System secure-reliability constraint

**Table 3** Model parameters

Parameter	Value
Data sent by Alice ( $d$ )	320 bits
Distance to Bob ( $r_{\text{Bob}}$ )	130 m
Distance to Eve ( $r_{\text{Eve}}$ )	160 m
Path-loss model	$PL = 17.0 + 40.0 \log_{10}(x)$
Frame span ( $T$ )	60 ms
Symbol length ( $T_s$ )	0.04 ms
Noise of Bob ( $\sigma_b^2$ )	-167 dBm
Noise of Eve ( $\sigma_e^2$ )	-164 dBm
Transmission channel bandwidth ( $B$ )	5 MHz
Time for sending NACK ( $t_{\text{NACK}}$ )	3 ms
CPU frequency upper limit ( $f_{\text{max}}$ )	3.5 GHz
CPU total workload ( $c$ )	20 Mcycles
Hardware constant ( $\kappa$ )	$10^{-11}$
Maximal allowed total error probability ( $\epsilon_{\text{max}}$ )	0.001
Decoding error probability bound for Eve ( $\epsilon_{e,\text{min}}$ )	0.999
Alice's probability of error in decoding NACK ( $\nu$ )	0.00001

**Proof of Lemma 1**

**Proof** Since  $\bar{E}_{b,\text{tot}}$  consists of three parts. i.e.,  $\bar{E}_{b,\text{tot}} = \bar{E}_c + \bar{E}_k + \bar{E}_t$ . By the properties of convex functions, if each of the three parts is convex in  $m$ , then their sum is also convex in  $m$ . Next, we will give the proof the convexity in  $m$  from three parts.

Firstly, for  $\bar{E}_t$  we have

$$\frac{\partial^2 \bar{E}_t}{\partial t^2} = \frac{P_t}{T_s} \left( (1 - \nu)A + \sum_{n=2}^N (1 - \nu)^n n(n - 1) \epsilon (n - 2) \left( \frac{\partial \epsilon_b}{\partial m} \right)^2 t + n \epsilon (1 - n)^n A \right), \tag{A1}$$

here we define  $A = \frac{\partial^2 \epsilon_b}{\partial m^2} + 2 \frac{\partial \epsilon_b}{\partial m} \geq 0$ . Recall that  $V \leq 1$  and  $m = \frac{t}{T_s}$ , we have

$$\begin{aligned} A &= \frac{1}{T_s} \left( \frac{\partial^2 \epsilon_b}{\partial m^2} + 2 \frac{\partial \epsilon_b}{\partial m} \right) \\ &= \sqrt{\frac{m}{V}} \left( \frac{(C - k/m)(C + k/m)^2}{4Vm^2} - \frac{3C + \frac{k}{m}}{4m^2} \right) \geq \frac{B}{m^3}, \end{aligned} \tag{A2}$$

where  $B = C^3 m^3 + (C^2 k - 3C)m^2 - (Ck^2 - 3k)m - k^3$  is a third degree polynomial whose largest root  $m = \frac{k}{C}$ . Since  $\epsilon_b \leq \epsilon_{b,\text{max}} \ll 1, C > \frac{k}{m}$  holds, we can give

$$\frac{\partial B}{\partial m} = 2C^2 km - Ck^2 + 3k + 3Cm(C^2 m - 3) \geq 2Ck^2 - Ck^2 + 3k + 3Cm(C^2 - 3) \geq 0. \tag{A3}$$

This suggests that  $B \geq 0$  holds on feasible domains. Hence,  $A \geq 0$  also holds according to (A2). Based on this,  $\frac{\partial^2 \bar{E}_t}{\partial m^2} \geq 0$ , i.e.,  $\bar{E}_t$  is convex in single communication transmission duration  $t$ . Because  $m = \frac{t}{T}$ ,  $\bar{E}_t$  is convex in  $m$ .

Next,  $\bar{E}_k$  meets the condition

$$\frac{\partial^2 \bar{E}_k}{\partial m^2} = \frac{\partial^2 \epsilon_b}{\partial m^2} E_k + \sum_{n=1}^N n(n-1) \epsilon_b^{n-2} \left( \frac{\partial^2 \epsilon_b}{\partial t^2} \right)^2 + n \epsilon_b^{n-1} + \frac{\partial^2 \epsilon_b}{\partial m^2}. \tag{A4}$$

Because  $\frac{\partial^2 \epsilon_b}{\partial m^2} \geq 0$ , we obtain  $\frac{\partial^2 \bar{E}_k}{\partial m^2} \geq 0$ , which indicates that  $\bar{E}_k$  is convex in blocklength  $m$ .

Finally, the convexity of  $\bar{E}_c$  in blocklength  $m$  is studied. We have

$$\frac{\partial^2 \bar{E}_c}{\partial m^2} = \frac{\partial^2 E_c^{(0)}}{\partial m^2} + \sum_{n=1}^N \left[ \left( n(n-1) \epsilon_b^{n-2} \left( \frac{\partial \epsilon_b}{\partial m} \right)^2 + n \epsilon_b^{n-1} \frac{\partial^2 \epsilon_b}{\partial m^2} \right) D_1^n - 2n \epsilon_b^{n-1} \frac{\partial \epsilon_b}{\partial m} D_2^n + \epsilon_b^n D_3^n \right], \tag{A5}$$

where  $D_1^{(n)} = E_c^{(n)} - E_c^{(n-1)}$ ,  $D_2^{(n)} = \frac{\partial E_c^n}{\partial m} - \frac{\partial E_c^{(n-1)}}{\partial m}$ ,  $D_3^{(n)} = \frac{\partial E_c^n}{\partial m} - \frac{\partial^2 E_c^{(n-1)}}{\partial m^2}$ , we have  $\frac{\partial \epsilon_b}{\partial m} < 0$ ,  $\frac{\partial^2 \epsilon_b}{\partial m^2} \geq 0$ . In addition,  $\frac{\partial^2 E_c^{(0)}}{\partial m^2} = 6(T-t)^4 \geq 0$ .

$$\begin{aligned} D_1^{(n)} &= \frac{1}{(T - (n+1)t - nt_{\text{NK}})^2} - \frac{1}{(T - nt - (n-1)t_{\text{NK}})^2} \\ &\geq \frac{1}{(T - nt - (n-1)t_{\text{NK}})^2} - \frac{1}{(T - nt - (n-1)t_{\text{NK}})^2} = 0. \end{aligned} \tag{A6}$$

Similarly, with exploiting  $n+1 \geq n$  to execute the inequality chain, we can proof that  $D_2^{(n)} \geq 0$ ,  $D_3^{(n)} \geq 0$ , also hold. So we proved that  $\frac{\partial^2 \bar{E}_c}{\partial m^2} \geq 0$ .

So far, we give the proof of the convexity of  $\bar{E}_t$ ,  $\bar{E}_k$  and  $\bar{E}_c$  in  $m$ , thus  $\bar{E}_{\text{tot}} = \bar{E}_t + \bar{E}_k + \bar{E}_c$  is also convex in  $m$ , i.e., objective function is convex in  $m$ .  $\square$

### Proof of Lemma 2

**Proof** Let's start with convexity of  $\epsilon_b$  in blocklength  $m$ .

Define  $\omega(m) = \sqrt{\frac{m}{V(\gamma)}} (C(\gamma) - d/m) \ln 2$  to facilitate subsequent proof. In addition, we have  $Q(\omega) = \int_{\omega}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ , which is monotonically decreasing in  $\omega$ , i.e.,  $\frac{\partial \epsilon_b}{\partial \omega} \leq 0$ . Based on [30] we can obtain

$$\begin{aligned} \frac{\partial \omega}{\partial m} &= \frac{\ln 2}{2} m^{-\frac{1}{2}} V^{-\frac{1}{2}} (C_b + m^{-1}d) \geq 0, \\ \frac{\partial^2 \omega}{\partial m^2} &= -\frac{\ln 2}{4} m^{-\frac{3}{2}} V^{-\frac{1}{2}} (C_{\text{textb}} + 3m^{-1}d) \leq 0, \\ \frac{\partial^2 \epsilon_b}{\partial \omega^2} &= \frac{\omega}{\sqrt{2\pi}} e^{-\frac{\omega^2}{2}} \geq 0. \end{aligned} \tag{B7}$$

Hence, we proof that  $\omega$  is concave in blocklength  $m$ .

Then, the convexity in  $m$  can be directly given by the following equation:

$$\frac{\partial \epsilon_b}{\partial m} = \frac{\partial \epsilon_b}{\partial \omega} \frac{\partial \omega}{\partial m} = \frac{\ln 2}{2} m^{-\frac{1}{2}} V^{-\frac{1}{2}} (C + m^{-1}d) \leq 0. \tag{B8}$$

Recall that  $\epsilon_b \leq \epsilon_{b,\max}$ , which indicates  $C_b \geq \frac{d}{m} = r$ . The convexity of  $\epsilon_b$  in  $m$  can be given as

$$\frac{\partial^2 \epsilon_b}{\partial m^2} = \frac{\partial^2 \epsilon_b}{\partial \omega_b^2} \left( \frac{\partial \omega_b}{\partial m} \right)^2 + \frac{\partial \epsilon_b}{\partial m} \frac{\partial^2 \omega_b}{\partial m^2} \geq 0. \tag{B9}$$

Furthermore, we can show that  $\epsilon_{b,\text{tot}}$  is convex in  $m$ . For  $N=0$ , we have  $\frac{\partial^2 \epsilon_{b,\text{tot}}}{\partial m^2} = \frac{\partial^2 \epsilon_b}{\partial m^2}$ . As for  $N \leq 1$ , we have

$$\frac{\partial^2 \epsilon_{b,\text{tot}}}{\partial m^2} = \frac{\partial^2 \epsilon_b}{\partial m^2} \nu + \sum_{n=2}^N n \left( (n-1) \epsilon_b^{(n-2)} \left( \frac{\partial \epsilon_b}{\partial m} \right)^2 + \epsilon_b^{(n-1)} \frac{\partial^2 \epsilon_b}{\partial m^2} \right) + N(N+1) \epsilon_b^{N-1} (1-\nu)^N \frac{\partial^2 \epsilon_b}{\partial m^2} \geq 0, \tag{B10}$$

i.e.,  $\epsilon_{b,\text{tot}}$  is convex in blocklength  $m$ . □

### Proof of Lemma 3

**Proof**  $\epsilon_e \geq \epsilon_{e,\min}$  implies  $C_e \leq \frac{d}{m} = r$ , its second derivative is given by

$$\begin{aligned} \frac{\partial^2 \epsilon_e}{\partial m^2} &= \frac{\partial^2 \epsilon_e}{\partial \omega_e^2} \left( \frac{\partial \omega_e}{\partial m} \right)^2 + \frac{\partial \epsilon_e}{\partial m} \frac{\partial^2 \omega_e}{\partial m^2} \\ &\stackrel{\frac{e^{-\omega_e^2/2}}{\sqrt{2\pi}} \geq 0}{\rightarrow} \frac{1}{4\sqrt{V_e m^3}} \left( (C_e + r)^2 \omega_e + (C_e + 3r) \right), \end{aligned} \tag{C11}$$

because of  $V_e = V_{\gamma_e} = 1 - (1 + \gamma_e)^{-2} \leq 1$ , the original formula can be reduced to

$$\begin{aligned} &\leq \frac{1}{4\sqrt{V_e m^3}} \left( (C_e + r)^2 \sqrt{m} (C_e - r) + (C_e + 3r) \right) \\ &= \frac{\sqrt{m}}{4\sqrt{V_e m^3}} \left[ -r^3 - C_e r^2 + \left( \frac{3}{\sqrt{m}} + C_e^2 \right) r + \frac{C_e}{\sqrt{m}} + C_e^3 \right]. \end{aligned} \tag{C12}$$

Define a function as  $h(r) = -r^3 - C_e r^2 + \left( \frac{3}{\sqrt{m}} + C_e^2 \right) r + \frac{C_e}{\sqrt{m}} + C_e^3$  is a monotonically decreasing function with respect to  $r \geq 0$  by

$$\frac{\partial h(r)}{\partial r} = \left( \frac{3}{\sqrt{m}} + C_e^2 \right) - 2C_e r - 3r^2. \tag{C13}$$

This is a quadratic polynomial function that possesses the open-down property with real roots. We have  $(2C_e)^2 + 4 \cdot 3 \left( \frac{3}{\sqrt{m^3}} + C_e \right) \geq 0$ , but its axis of symmetry is  $r_{\text{sym}} = -\frac{C_e}{3} \leq 0$ , recall that  $0 \leq C_e \leq r_e$ , then we have  $h(r) \leq h(0) \leq 0$ , which indicates

$$\frac{\partial^2 \epsilon_e}{\partial m^2} = \frac{1}{4\sqrt{V_e m^3}} h(r) \leq 0, \tag{C14}$$

moreover  $\frac{\partial \epsilon_e}{\partial m} = \frac{\partial \epsilon_e}{\partial \omega_e} \frac{\partial \omega_e}{\partial m} \geq 0$ .  $v$  is a very small value. We can approximate that:

$$\frac{\partial^2 \epsilon_{e,tot}}{\partial m^2} = (1 - \epsilon_b) \sum_{n=1}^N \epsilon_e^n \epsilon_b^{n-1} + \epsilon_e^{N+1} \epsilon_b^N. \tag{C15}$$

Hence, we have

$$\begin{aligned} \frac{\partial^2 \epsilon_{e,tot}}{\partial m^2} &= \frac{\partial^2 \epsilon_e}{\partial m^2} \left[ 1 + \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^{n-1} - \epsilon_b - \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^n \right] \\ &+ \frac{\partial^2 \epsilon_b}{\partial m^2} \left[ \sum_{n=2}^N (n-1)(n-2) \epsilon_e^n \epsilon_b^{n-3} - \epsilon_e - \sum_{n=2}^N n(n-1) \epsilon_e^n \epsilon_b^{n-2} \right] \\ &+ \frac{\partial \epsilon_b}{\partial m} \frac{\partial \epsilon_e}{\partial m} \left[ \sum_{n=2}^N 2n(n-1) \epsilon_e^{n-1} \epsilon_b^{n-2} - 2 - \sum_{n=2}^N 2n^2 \epsilon_e^{n-1} \epsilon_b^{n-1} \right] + \frac{\partial^2 \epsilon_e}{\partial m^2} [N(N+1) \epsilon_e^{N-1} \epsilon_b^N] \\ &+ \frac{\partial \epsilon_e}{\partial m} \frac{\partial \epsilon_b}{\partial m} [2N(N+1) \epsilon_e^N \epsilon_b^{N-1}] + \frac{\partial^2 \epsilon_b}{\partial m^2} [N(N-1) \epsilon_e^{N+1} \epsilon_b^{N-2}]. \end{aligned} \tag{C16}$$

Firstly, because  $0 \leq \epsilon_b \leq \epsilon_{max} \leq 0.1$ , so  $A^{(1)} = \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^{n-1} - \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^n = \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} (\epsilon_b^{n-1} - \epsilon_b^n) \geq 0$ , and  $\frac{\partial^2 \epsilon_e}{\partial m^2} \leq 0$ , hence  $\frac{\partial^2 \epsilon_e}{\partial m^2} \left( 1 + \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^{n-1} - \epsilon_b - \sum_{n=2}^N n(n-1) \epsilon_e^{n-2} \epsilon_b^n \right) \leq 0$ .

Secondly, we denote

$$\begin{aligned} B^{(1)} &= \sum_{n=2}^N (n-1)(n-2) \epsilon_e^n \epsilon_b^{n-3} \\ B^{(2)} &= \sum_{n=2}^N n(n-1) \epsilon_e^n \epsilon_b^{n-2}, \end{aligned} \tag{C17}$$

because  $R = \frac{B^{(1)}}{B^{(2)}} = \frac{(n-2)\epsilon_b}{n} \leq 1$ , so we have

$$\sum_{n=2}^N (n-1)(n-2) \epsilon_e^n \epsilon_b^{n-3} - \sum_{n=2}^N n(n-1) \epsilon_e^n \epsilon_b^{n-2} \leq 0, \tag{C18}$$

moreover,  $B^{(3)} = N(N-1) \epsilon_e^{N+1} \epsilon_b^{N-2} - \epsilon_e$ , since  $N(N-1) \epsilon_e^{N+1} \epsilon_b^{N-2} \leq N(N-1) \epsilon_e^N \epsilon_b^{N-2} \leq 1$ , we can find that  $\epsilon_e (N(N-1) \epsilon_e^N \epsilon_b^{N-2} - 1) = B^{(3)} \leq 0$ . According to  $B^{(3)}$  and  $R$ , we have

$$\frac{\partial^2 \epsilon_b}{\partial m^2} \left[ \sum_{n=2}^N (n-1)(n-2) \epsilon_e^n \epsilon_b^{n-3} - \epsilon_e - \sum_{n=2}^N n(n-1) \epsilon_e^n \epsilon_b^{n-2} + N(N-1) \epsilon_e^{N+1} \epsilon_b^{N-2} \right] \leq 0. \tag{C19}$$

Furthermore, we denote that

$$\begin{aligned}
 D^{(1)} &= \sum_{n=2}^N 2n(n-1)\epsilon_e^{n-1}\epsilon_b^{n-2} - 2 - \sum_{n=2}^N 2n^2\epsilon_e^{n-1}\epsilon_b^{n-1} \\
 &= 2 \sum_{n=1}^N n(n-1)\epsilon_e^{n-1}\epsilon_b^{n-2} - n^2\epsilon_e^{n-1}\epsilon_b^{n-1} \\
 &= 2 \sum_{n=1}^N n(n-1)\epsilon_e^{n-1}\epsilon_b^{n-2} \left( (\epsilon_b - 1)n^2 - n \right).
 \end{aligned} \tag{C20}$$

Additionally,  $(\epsilon_b - 1)n^2 - n$  is a quadratic function of the downward direction of an opening that does not intersect with the real axis and  $2n(\epsilon_b - 1) - 1 \leq 0$ .  $D^{(1)} \leq D_0^{(1)} \leq 0$ , so we have,

$$\frac{\partial \epsilon_b}{\partial m} \frac{\partial \epsilon_e}{\partial m} \left[ \sum_{n=2}^N 2n(n-1)\epsilon_e^{n-1}\epsilon_b^{n-2} - 2 - \sum_{n=2}^N 2n^2\epsilon_e^{n-1}\epsilon_b^{n-1} \right] \leq 0. \tag{C21}$$

□

So far, we have proved that  $\frac{\partial^2 \epsilon_{e,tot}}{\partial m^2} \leq 0$ , i.e.,  $\epsilon_{e,tot}$  is concave in  $m$ .

#### Proof of Lemma 4

**Proof** According to (10), we can conclude  $\epsilon_{LFP}$  is convex in  $m$  by showing

$$\frac{\partial^2 \epsilon_{LF}}{\partial m^2} = \frac{\partial^2 \epsilon_b}{\partial m^2} \epsilon_e + 2 \frac{\partial \epsilon_b}{\partial m} \frac{\partial \epsilon_e}{\partial m} + (\epsilon_b - 1) \frac{\partial^2 \epsilon_e}{\partial m^2} \geq 0. \tag{D22}$$

Given our previous proof, we have

$$\frac{\partial^2 \epsilon_{LF,tot}}{\partial m^2} = \frac{\partial^2 \epsilon_{b,tot}}{\partial m^2} \epsilon_{e,tot} + 2 \frac{\partial \epsilon_{b,tot}}{\partial m} \frac{\partial \epsilon_{e,tot}}{\partial m} + (\epsilon_{b,tot} - 1) \frac{\partial^2 \epsilon_{e,tot}}{\partial m^2} \geq 0, \tag{D23}$$

among them  $\frac{\partial \epsilon_{e,tot}}{\partial m} \frac{\partial \epsilon_{b,tot}}{\partial m} \geq 0$  holds. Hence, the total LFP  $\epsilon_{LFP,tot}$  is convex in the block-length of a single (re)transmission  $m$ . □

#### Acknowledgements

Not applicable.

#### Author contributions

S was a contributor to the article; YZ, YH and Anke guided the article to completion.

#### Funding

The work of C. Shi and Y. Hu is supported by NSFC Grant with No. 62101389 and the National Key RD Program of China with No. 2023YFE0206600, and the Fundamental Research Funds for the Central Universities (2042024kf1006). The work of Y. Zhu and A. Schmeink is supported by Federal Ministry of Education and Research (BMBF) Germany in the program of "Souverän. Digital. Vernetzt." Joint Project 6 G-ANNA with project identification number 16KISK097, and in part by the German Research Council (DFG) through the basic research project under Grant DFG SCHM 2643/17.

#### Availability of data and materials

Not applicable.



## Declarations

### Competing interests

The authors declare that they have no conflict of interest.

Received: 20 November 2023 Accepted: 13 August 2024

Published online: 03 September 2024

## References

- Z. Xiang, W. Yang, G. Pan, Y. Cai, Y. Song, Physical layer security in cognitive radio inspired noma network. *IEEE J. Sel. Topics Signal Process.* **13**(3), 700–714 (2019). <https://doi.org/10.1109/JSTSP.2019.2902103>
- B. Radouane, G. Lyamine, K. Ahmed, B. Kamel, Scalable mobile computing: From cloud computing to mobile edge computing. 2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS) pp. 1–6 (2022). <https://doi.org/10.1109/NISS55057.2022.10085600>
- M. Mitev, A. Chorti, H.V. Poor, G.P. Fettweis, What physical layer security can do for 6g security. *IEEE Open J. Veh. Technol.* **4**, 375–388 (2023). <https://doi.org/10.1109/OJVT.2023.3245071>
- M.S.J. Solajija, H. Salman, H. Arslan, Towards a unified framework for physical layer security in 5g and beyond networks. *IEEE Open J. Veh. Technol.* **3**, 321–343 (2022). <https://doi.org/10.1109/OJVT.2022.3183218>
- R. Raman, R. Singh, Z. Gupta, S. Verma, A. Rajput, S.M. Parikh, Wireless communication with extreme reliability and low latency: tail, risk and scale. In: 2022 5th International conference on contemporary computing and informatics (IC3I) pp. 1699–1703 (2022). <https://doi.org/10.1109/IC3I56241.2022.10073096>
- Y. Zhu, X. Yuan, Y. Hu, R.F. Schaefer, A. Schmeink, Trade reliability for security: leakage-failure probability minimization for machine-type communications in URLLC. *IEEE J. Sel. Areas Commun.* **41**(7), 2123–2137 (2023). <https://doi.org/10.1109/JSAC.2023.3280960>
- R. Chen, C. Li, S. Yan, R. Malaney, J. Yuan, Physical layer security for ultra-reliable and low-latency communications. *IEEE Wirel. Commun.* **26**(5), 6–11 (2019). <https://doi.org/10.1109/MWC.001.1900051>
- R.F. Liao, H. Wen, J. Wu, F. Pan, A. Xu, H. Song, F. Xie, Y. Jiang, M. Cao, Security enhancement for mobile edge computing through physical layer authentication. *IEEE Access* **7**, 116390–116401 (2019). <https://doi.org/10.1109/ACCESS.2019.2934122>
- T.X. Zheng, X. Chen, Y. Wen, N. Zhang, D.W.K. Ng, N. Al-Dhahir, Secure offloading in noma-enabled multi-access edge computing networks. *IEEE Trans. Commun.* pp. 1–1 (2023). <https://doi.org/10.1109/TCOMM.2023.3342242>
- S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, V.C.M. Leung, Reconfigurable intelligent surface-assisted secure mobile edge computing networks. *IEEE Trans. Veh. Technol.* **71**(6), 6647–6660 (2022). <https://doi.org/10.1109/TVT.2022.3162044>
- J. Qin, J. Liu, Multi-access edge offloading based on physical layer security in c-v2x system. *IEEE Trans. Veh. Technol.* **71**(7), 6912–6923 (2022). <https://doi.org/10.1109/TVT.2022.3164896>
- Y. Polyanskiy, H.V. Poor, S. Verdú, Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **56**(5), 2307–2359 (2010). <https://doi.org/10.1109/TIT.2010.2043769>
- J. Cao, X. Zhu, Y. Jiang, Y. Liu, F.C. Zheng, Joint block length and pilot length optimization for URLLC in the finite block length regime. In: 2019 IEEE Global Communications Conference (GLOBECOM) pp. 1–6 (2019). <https://doi.org/10.1109/GLOBECOM38437.2019.9013958>
- S. Pala, M. Katwe, K. Singh, B. Clerckx, C.P. Li, Spectral-efficient ris-aided rsma URLLC: toward mobile broadband reliable low latency communication (mbrllc) system. *IEEE Trans. Wirel. Commun.*, pp. 1–1 (2023). <https://doi.org/10.1109/TWC.2023.3309028>
- R. Hashemi, S. Ali, N.H. Mahmood, M. Latva-Aho, Joint sum rate and blocklength optimization in ris-aided short packet URLLC systems. *IEEE Commun. Lett.* **26**(8), 1838–1842 (2022). <https://doi.org/10.1109/LCOMM.2022.3180396>
- W.J. Ryu, S.Y. Shin, Power allocation for URLLC using finite blocklength regime in downlink noma systems. In: 2019 International Conference on Information and Communication Technology Convergence (ICTC) pp. 770–773 (2019). <https://doi.org/10.1109/ICTC46691.2019.8939713>
- Q. Peng, H. Ren, C. Pan, N. Liu, M. El-kashlan, Resource allocation for uplink cell-free massive mimo enabled URLLC in a smart factory. *IEEE Trans. Commun.* **71**(1), 553–568 (2023). <https://doi.org/10.1109/TCOMM.2022.3224502>
- Y. Wu, D. Qiao, H. Qian, Efficient bandwidth allocation for URLLC in frequency-selective fading channels. In: GLOBECOM 2020—2020 IEEE global communications conference, pp. 1–6 (2020). <https://doi.org/10.1109/GLOBECOM42002.2020.9322582>
- K. Yu, J. Yu, A. Dong, Cooperative communication and mobility for securing URLLC of future wireless networks. *IEEE Trans. Veh. Technol.* **71**(5), 5331–5342 (2022). <https://doi.org/10.1109/TVT.2022.3151063>
- T.X. Zheng, X. Chen, C. Wang, K.K. Wong, J. Yuan, Physical layer security in large-scale random multiple access wireless sensor networks: a stochastic geometry approach. *IEEE Trans. Commun.* **70**(6), 4038–4051 (2022). <https://doi.org/10.1109/TCOMM.2022.3167047>
- C. Wang, Z. Li, H. Zhang, D.W.K. Ng, N. Al-Dhahir, Achieving covertness and security in broadcast channels with finite blocklength. *IEEE Trans. Wirel. Commun.* **21**(9), 7624–7640 (2022). <https://doi.org/10.1109/TWC.2022.3160051>
- D. Xu, H. Zhu, Proactive eavesdropping via jamming over short packet suspicious communications with finite blocklength. *IEEE Trans. Commun.* **70**(11), 7505–7519 (2022). <https://doi.org/10.1109/TCOMM.2022.3208621>
- M. Naseri-Tehrani, S. Farahmand, Resource allocation for irs-enabled secure multiuser multi-carrier downlink URLLC systems. In: 2022 IEEE 23rd international workshop on signal processing advances in wireless communication (SPAWC) pp. 1–5 (2022). <https://doi.org/10.1109/SPAWC51304.2022.9833996>

24. A. Pradhan, S. Das, M. Jalil Piran, Blocklength optimization and power allocation for energy-efficient and secure URLLC in industrial iot. *IEEE Internet Things J.* **11**(6), 9420–9431 (2024). <https://doi.org/10.1109/JIOT.2023.3324379>
25. F. Nadeem, Y. Li, B. Vucetic, M. Shirvanimoghaddam, Analysis and optimization of harq for URLLC. In: 2021 IEEE Globecom workshops (GC Wkshps), pp. 1–6 (2021). <https://doi.org/10.1109/GCWkshps52748.2021.9682028>
26. J. Östman, R. Devassy, G.C. Ferrante, G. Durisi, Low-latency short-packet transmissions: fixed length or harq? In: 2018 IEEE Globecom workshops (GC Wkshps), pp. 1–6 (2018). <https://doi.org/10.1109/GLOCOMW.2018.8644397>
27. R. Santos, D. Castanheira, A. Silva, A. Gameiro, Multi-user ir-harq latency and resource optimization for URLLC. *IEEE Access* **11**, 129994–130009 (2023). <https://doi.org/10.1109/ACCESS.2023.3334256>
28. M.M. Ebrahimi, K. Khamforoosh, M. Amini, A. Sheikahmadi, H. Khamfroush, Adaptive-persistent nonorthogonal random access scheme for URLL massive IoT networks. *IEEE Syst. J.* **17**(1), 1660–1671 (2023). <https://doi.org/10.1109/JSYST.2022.3190132>
29. Y. Mao, J. Zhang, K.B. Letaief, Dynamic computation offloading for mobile-edge computing with energy harvesting devices. *IEEE J. Sel. Areas Commun.* **34**(12), 3590–3605 (2016). <https://doi.org/10.1109/JSAC.2016.2611964>
30. Y. Hu, Y. Zhu, M.C. Gursoy, A. Schmeink, Swipt-enabled relaying in iot networks operating with finite blocklength codes. *IEEE J. Sel. Areas Commun.* **37**(1), 74–88 (2019). <https://doi.org/10.1109/JSAC.2018.2872361>
31. Y. Zhu, Y. Hu, A. Schmeink, J. Gross, Energy minimization of mobile edge computing networks with finite retransmissions in the finite blocklength regime. In: 2019 IEEE 20th international workshop on signal processing advances in wireless communications (SPAWC), pp. 1–5 (2019). <https://doi.org/10.1109/SPAWC.2019.8815391>
32. M. Lubin, E. Yamangil, R. Bent, J.P. Vielma, Polyhedral approximation in mixed-integer convex optimization. *Math. Program.* **172**(1–2), 139–168 (2017). <https://doi.org/10.1007/s10107-017-1191-y>
33. Y. Corre, J. Stephan, Y. Lostanlen, Indoor-to-outdoor path-loss models for femtocell predictions. In: 2011 IEEE 22nd international symposium on personal, indoor and mobile radio communications, pp. 824–828 (2011). <https://doi.org/10.1109/PIMRC.2011.6140082>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.