

RESEARCH

Open Access



# Statistical QoS provisioning for MTC networks under finite blocklength

Mohammad Shehab\* , Endrit Dosti, Hirley Alves and Matti Latva-aho

## Abstract

This paper analyzes the effective capacity of delay-constrained machine-type communication (MTC) networks operating in the finite blocklength regime. First, we derive a closed-form mathematical approximation for the effective capacity in quasi-static Rayleigh fading channels. We characterize the optimum error probability to maximize the concave effective capacity function with reliability constraint and study the effect of signal-to-interference-plus-noise ratio (SINR) variations for different delay constraints. The trade-off between reliability and effective capacity maximization reveals that we can achieve higher reliability with limited sacrifice in effective capacity specially when the number of machines is small. Our analysis reveals that SINR variations have less impact on effective capacity for strict delay-constrained networks. We present an exemplary scenario for massive MTC access to analyze the interference effect proposing three methods to restore the effective capacity for a certain node which are power control, graceful degradation of delay constraint, and joint compensation. Joint compensation combines both power control and graceful degradation of delay constraint, where we perform the maximization of an objective function whose parameters are determined according to the delay and SINR priorities. Our results show that networks with stringent delay constraints favor power controlled compensation, and compensation is generally performed at higher costs for shorter packets.

**Keywords:** Effective capacity, Machine-type communication, Finite blocklength, Ultra-reliable communication

## 1 Introduction

Modern communication systems are becoming an indispensable part of our lives. Driven by the demands of users for extra services, the fifth generation (5G) of mobile communication is expected to introduce new features such as *ultra-reliable low-latency communications* (URLLC) and *massive machine-type communication* (m-MTC) [1–6]. These features may serve many yet unforeseen applications to enable the Internet of Things (IoT). IoT aims at bringing connectivity to anything that can benefit from internet connection [7]. URLLC has emerged to provide solutions for reliable and low-latency transmissions in wireless systems. The design of URLLC systems imposes strict quality of service (QoS) constraints to fulfill very low latency in the order of milliseconds with expected reliability of higher than 99.9% [1, 4]. In [8], Schulz et al. discussed the reliability requirements for different IoT applications. According to their study, latency bounds range from 1 ms

in factory automation to 100 ms in road safety. In addition, the packet loss rate constraints range from  $10^{-9}$  in printing machines to  $10^{-3}$  for traffic efficiency. Such requirements are far more stringent than the ones in the current Long-Term Evolution (LTE) standards [9].

The m-MTC refers to networks that can support a variety of connected smart devices at the same time with the same base station (BS). It obligates a certain level of connectivity to a machine via ultra-reliable communication (URC) over relatively long term ( $> 10$  ms) [1]. The number of connected devices is expected to cross the 28 billion border by 2021, where a single macro-cell may need to uphold 10,000 or more devices in the future [10, 11]. Moreover, the traffic behavior of MTC is quite different from the HTC (human-type communication), where [12]:

- MTC is coordinated (i.e, there are simultaneous access attempts from many machine reacting to the same events), while HTC is uncoordinated.
- MTC uses short as well as small number of packets.
- MTC traffic is real as well as non-real-time, periodic, and event-driven.

\*Correspondence: [mohammad.shehab@oulu.fi](mailto:mohammad.shehab@oulu.fi)  
Centre for Wireless Communications (CWC), University of Oulu, Oulu, Finland

- MTC QoS requirement is different from HTC (i.e., different reliability and latency requirements).

In this context, MTC has gained an increasing interest in recent years via employing new multiple radio access technologies and efficient utilization of spectrum resources to improve reliability and robustness [13–15]. Another research topic that has gathered much attention is the cooperative transmission in MTC, where in [11], the authors proposed a location-based cooperative strategy to reduce the error outage probability, but without study of the latency aspects.

Traditional communication systems are based on Shannon theoretic models and utilize metrics such as channel capacity or ergodic capacity [16]. Unlike classical systems, URLLC networks are designed to communicate on short packets in order to satisfy extremely low latency in real-time applications and emerging technologies such as e-health, industrial automation, and smart grids whenever data sizes are reasonably small such as sensor readings or alarm notifications. In the finite blocklength regime, the length of metadata is of comparable size with the length of data. Such demands stimulated a revolutionary trend in information theory studying communication at finite blocklength (FB) [17–19]. In this context, conventional metrics (e.g., channel capacity or ergodic capacity) become highly suboptimal [17]. For this reason, the maximum achievable rate for quasi-static fading channels was characterized in [19] as a function of blocklength and error probability  $\epsilon$ . In [9], the authors analyzed the effect of using smaller resource blocks on error probability bounds in OFDM. The effect of relaying of blocklength-limited packets was studied and compared to direct transmission in [20, 21] where the authors concluded that relaying is more efficient than direct transmission in the FB regime specially with an average channel state information (CSI). Furthermore, the authors of [3] introduced a per-node throughput model for additive white Gaussian noise (AWGN) and quasi-static collision channels. Therein, average delay is considered, and interference is treated as AWGN.

To model the delay requirements in URLLC and MTC networks, we resort to the effective capacity (EC) metric which was introduced in [22]. It indicates the maximum possible arrival rate that can be supported by a network with a target delay constraint. In [18], the authors considered quasi-static Rayleigh fading channels and introduced a statistical model for a single node effective rate in bits per channel use (bpcu) for a certain error probability and delay exponent which reflects the latency requirement. However, throughout the paper, a closed-form expression for the EC was not provided. Exploiting the EC theory, the authors of [23] characterized the latency-throughput trade-off for cellular networks. In [24],

Musavian and Le-Ngoc analyzed the EC maximization of secondary node with some interference power constraints for primary node in a cognitive radio environment with interference constraints. Three types of constraints were imposed, namely average interference power, peak interference power, and interference power outage. The fundamental trade-off between EC and consumed power was studied in [25] where the authors suggested an algorithm to maximize the EC subject to power constraint for a single node scenario. In [26], we studied the per-node EC in MTC networks operating in quasi-static Rayleigh fading proposing three methods to alleviate interference, namely power control, graceful degradation of delay constraint, and the joint method. To the best of our knowledge, EC for FB packet transmission in multi-node MTC scenario has not been investigated until part of the work in this journal was presented in [26], which will be depicted here with extra details.

Based on its intuition, the EC theory provides a mathematical framework to study the interplay among transmit power, interference, delay, and the achievable rate for different wireless channels. In this paper, we derive a mathematical expression for EC in quasi-static Rayleigh fading for delay-constrained networks. Our results depict that a system can achieve higher reliability with a negligible sacrifice in its EC. We consider dense MTC networks and characterize the effect of interference on their EC. We propose three methods to allow a certain node maintain its EC which are (i) power control, (ii) graceful degradation of delay constraint, and (iii) joint model. Power control depends on increasing the power of a certain node to recover its EC which in turn degrades the SINR of other nodes. Our analysis proves that SINR variations have limited effect on EC in networks with stringent delay limits. Hence, the side effect of power control is worse for less stringent delay constraints and vice versa. We illustrate the trade-off between power control and graceful degradation of delay constraint. Furthermore, we introduce a joint model which combines both of them. The operational point to determine the amount of compensation performed by each of the two methods in the joint model is determined by maximization of an objective function leveraging the network performance.

The motivation beyond this paper is to provide a solid understanding of the trade-off between power, delay, and reliability in MTC networks in the finite blocklength regime. Our objective is to pave the road for utilizing short packets in 5G and machine-type networks. Extra plots that were not present in [26] are illustrated to provide the reader with full understanding of the objective function in joint compensation and the compensation process itself. Moreover, we extend the analysis in [26] by solving the optimization problem to obtain the optimum error probability which maximizes the EC in the ultra-reliable

region. We also characterize the trade-off between reliability and EC which shows that we can obtain a huge gain in reliability in return for a slight reduction in EC.

The rest of the paper is organized as follows: in Section 2, we introduce the system model and define some concepts such as communication at finite blocklength and EC. A closed form for the EC in quasi-static Rayleigh fading is derived in Section 3, where we also show the effect of interference on the per-node EC in multi-node MTC networks. Next, Section 4 depicts the optimization problem to maximize the EC in the ultra-reliable (UR) region. We present the interference alleviation methods and the trade-off between them in Section 5, while the results are depicted in Section 6. Finally, Section 7 concludes the paper. Table 1 includes the important abbreviations and symbols that will appear throughout the paper.

## 2 Preliminaries

### 2.1 Network model

We consider a transmission scenario in which  $N$  nodes transmit packets with equal power to a common controller through a quasi-static Rayleigh fading collision channel with blocklength  $T_f$  as shown in Fig. 1. For convenience in this paper, we refer to one machine terminal as node. Given that all nodes transmit at the same time slot, the controller attempts to decode the transmitted symbols arriving from all nodes. When the controller decodes one node's data, the other streams appear as interference to it. For this model, imagine that a node needs to rise its rate temporarily for a critical reason. Later on, we study the interference alleviation scenarios for one node at a certain time slot while all nodes also keep transmitting at the same time.

Based on our network model, the received vector  $\mathbf{y}_n \in \mathbb{C}^n$  of node  $n$  is given by:

$$\mathbf{y}_n = h_n \mathbf{x}_n + \sum_{s \neq n} h_s \mathbf{x}_s + \mathbf{w}, \quad (1)$$

where  $\mathbf{x}_n \in \mathbb{C}^n$  is the transmitted packet of node  $n$ , and  $h_n$  is the fading coefficient for node  $n$  which is assumed to be quasi-static with Rayleigh distribution. This implies that the fading coefficient  $h_n$  remains constant for each block of  $T_f$  channel uses which span the whole packet duration and changes independently from one block to another. The index  $s$  includes all  $N - 1$  interfering nodes which collide with node  $n$ , and  $\mathbf{w}$  is the additive complex Gaussian noise vector whose entries are defined to be circularly symmetric with unit variance. Given the signal-to-noise ratio  $\rho$  of a single node, the signal-to-interference-plus-noise ratio of any node  $n$  is

$$\rho_i = \frac{\rho}{1 + \rho \sum_s |h_s|^2}. \quad (2)$$

**Table 1** List of abbreviations and symbols

bpcu	Bits per channel use
EC	Effective capacity
Max	Maximize
PDF	Probability density function
QoS	Quality of service
SINR	Signal-to-interference-plus-noise ratio
s.t	Subject to
URC	Ultra-reliable communication
$C(\rho h ^2)$	Shannon capacity
$D_{\max}$	Maximum delay
$\mathbb{E}[\ ]$	Expectation of
EC	Effective capacity
$EC_{\max}$	Maximum effective capacity
$\mathcal{L}(\epsilon, \lambda)$	Lagrangian function
$N$	Number of nodes
$Pr()$	Probability of
$P_{\text{out\_delay}}$	Delay outage probability
$Q(x)$	Gaussian $Q$ -function
$Q^{-1}(x)$	Inverse Gaussian $Q$ -function
$T_f$	Blocklength
$V(\rho h ^2)$	Channel dispersion
$e$	Exponential Euler's number
$ h ^2$	Fading coefficient
$\ln$	Natural logarithm to the base $e$
$\log_2$	Logarithm to the base 2
$r$	Normalized achievable rate
$\mathbf{w}$	Additive white Gaussian noise vector
$\mathbf{x}_n$	Transmitted signal vector of node $n$
$\mathbf{y}_n$	Received signal vector of node $n$
$z$	Fading random variable
$\alpha$	Collision loss factor
$\alpha_c$	Compensation loss factor
$\alpha_{c_0}$	Operational point of compensation loss factor
$\alpha_t$	Total loss
$\gamma_c$	Compensation gain
$\theta$	Delay exponent
$\epsilon$	Error probability
$\epsilon_t$	Target error probability
$\epsilon^*$	Optimum error probability
$\eta_\alpha$	Compensation loss priority factor
$\eta_\theta$	Delay priority factor
$\rho$	Signal-to-noise ratio
$\rho_c$	Compensation SNR
$\rho_{c_0}$	Operational point of compensation SNR
$\rho_i$	Signal-to-interference-plus-noise ratio
$\rho_s$	SINR of other non-compensating nodes

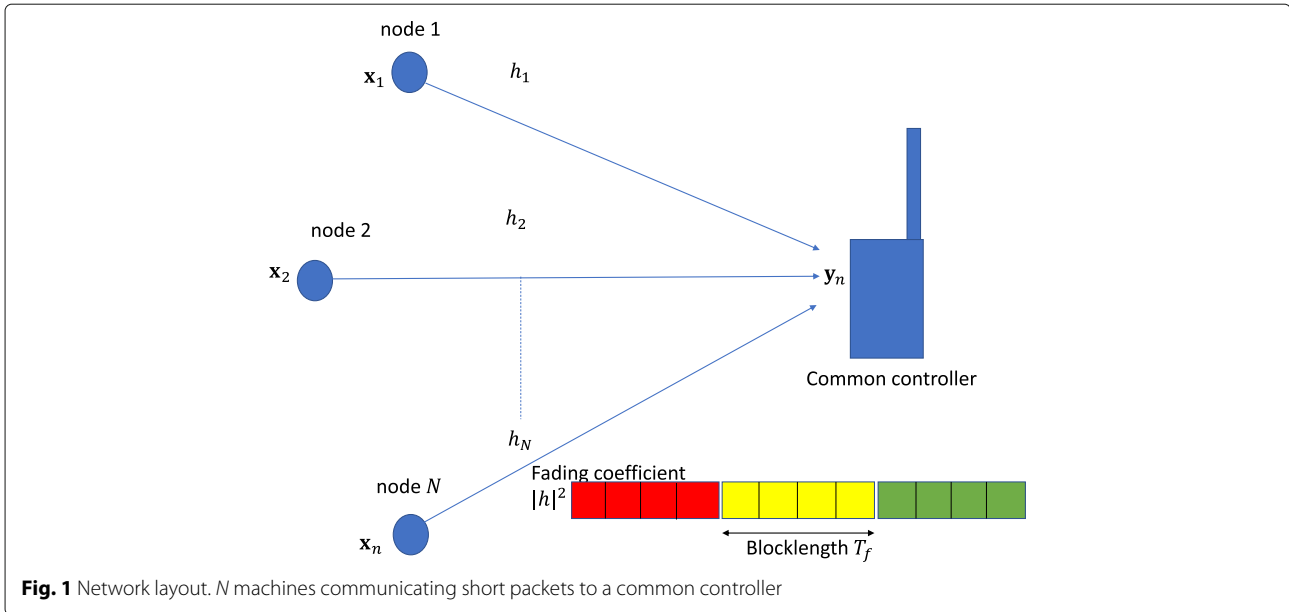


Fig. 1 Network layout.  $N$  machines communicating short packets to a common controller

To simplify the analysis, we assume that (i) each node always has a packet to transmit (buffer is always non-empty), (ii) all nodes are equidistant from the common controller (i.e., same path loss), and (iii) the fading coefficients  $h_s$  are independent and identically distributed and perfectly known to the receiver. Thus, as the number of nodes increases, the sum of Rayleigh-distributed fading envelopes of  $N - 1$  interfering nodes becomes  $\sum_s |h_s|^2 \approx N - 1$  [27], and the interference resulting from nodes in set  $s$  can be modeled as in [3] where (2) reduces to:

$$\rho_i = \frac{\rho}{1 + \rho(N - 1)}. \quad (3)$$

Note that CSI acquisition in this setup is not trivial, and its cost is negligible whenever the channel remains constant over multiple symbols. Additionally, as in [3], we aim to provide a performance benchmark for such networks without interference coordination.

### 2.2 Communication at finite blocklength

In this section, we present the notion of FB transmission, in which short packets are conveyed at rate that depends not only on the SNR, but also on the blocklength and the probability of error  $\epsilon$  [17]. In this case,  $\epsilon$  has a small value but not vanishing. For error probability  $\epsilon \in [0, 1]$ , the normalized achievable rate in bits per channel use is given by:

$$r \approx C(\rho_i |h|^2) - \sqrt{\frac{V(\rho_i |h|^2)}{T_f}} Q^{-1}(\epsilon), \quad (4)$$

where

$$C(t) = \log_2(1 + t) \quad (5)$$

is Shannon's channel capacity for sufficiently long packets, while

$$V(t) = (1 - (1 + t)^{-2}) (\log_2 e)^2 \quad (6)$$

denotes the channel dispersion which appears for relatively short packets ( $T_f < 2000$  channel uses) [18],  $Q(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$  is the Gaussian Q-function, and  $Q^{-1}(t)$  represents its inverse,  $\rho_i$  is the SINR, and  $|h|^2$  is the fading envelope.

The channel is assumed to be Rayleigh quasi-static fading where the fading coefficients remain constant over  $T_f$  symbols which spans the whole packet duration. For Rayleigh channels [28], the fading coefficients  $Z = |h|^2$  have the following probability density function distribution:

$$f_Z(z) = e^{-z}. \quad (7)$$

### 2.3 Effective capacity

The concept of EC indicates the capability of communication nodes to exchange data with maximum rate and certain latency constraint and thus guarantees QoS by capturing the physical and link layer aspects. A statistical delay violation model implies that an outage occurs when a packet delay exceeds a maximum delay bound  $D_{\max}$ , and its probability is defined as [22]:

$$P_{\text{out\_delay}} = \Pr(\text{delay} \geq D_{\max}) \approx e^{-\theta \cdot EC \cdot D_{\max}}, \quad (8)$$

where  $\Pr(\cdot)$  denotes the probability of a certain event. Conventionally, a network's tolerance to long delay is measured by the delay exponent  $\theta$ . The network has more tolerance to large delays for small values of  $\theta$  (i.e.,  $\theta \rightarrow 0$ ), while for large values of  $\theta$ , it becomes more delay strict.

For the infinite blocklength model, the EC capacity is defined as:

$$EC(\rho_i, \theta) = -\frac{\ln \mathbb{E}_Z \left( e^{-\theta T_f C(\rho_i | h^2)} \right)}{T_f \theta}, \quad (9)$$

In quasi-static fading, the channel remains constant within each transmission period  $T_f$  [21], and the EC is subject to the finite blocklength error bounds and thus according to [18] can be written as:

$$EC(\rho_i, \theta, \epsilon) = -\frac{\ln \psi(\rho_i, \theta, \epsilon)}{T_f \theta}, \quad (10)$$

where

$$\psi(\rho_i, \theta, \epsilon) = \mathbb{E}_Z \left[ \epsilon + (1 - \epsilon) e^{-T_f \theta r} \right]. \quad (11)$$

In [18, 29], the effective capacity is statistically studied for single node scenario in block fading but never to a closed form expression. It has been proven that the EC is concave in  $\epsilon$  and, hence, has a unique maximizer. In what follows, we shall represent the EC expression for quasi-static Rayleigh fading.

### 3 Effective capacity analysis under finite blocklength

**Lemma 1** *The effective capacity of a certain node communicating in a quasi-static Rayleigh fading channel is approximated by:*

$$EC(\rho_i, \theta, \epsilon) \approx -\frac{1}{T_f \theta} \ln [\epsilon + (1 - \epsilon) \mathcal{J}], \quad (12)$$

with

$$\mathcal{J} = e^{\frac{1}{\rho}} \rho^d \left[ \left( \frac{c^2}{2} + c + 1 \right) \Gamma \left( d + 1, \frac{1}{\rho} \right) - \frac{c}{2} (c + 1) \rho^{-2} \Gamma \left( d - 1, \frac{1}{\rho} \right) \right], \quad (13)$$

where  $d = \frac{-\theta T_f}{\ln(2)}$ . Also, let  $c = \theta \sqrt{T_f} Q^{-1}(\epsilon) \log_2 e$  and  $x = \sqrt{1 - \frac{1}{(1 + \rho_i z)^2}}$ , and  $\Gamma(\cdot, \cdot)$  is the upper incomplete gamma function ([30], § 8.350-2).

*Proof* Please refer to Appendix 1.  $\square$

**Lemma 2** *There is a unique local and global maximizer in  $\epsilon$  for the per-node EC in quasi-static Rayleigh fading channels which is given by:*

$$\epsilon^*(\rho_i, c, d) = \arg \min_{0 \leq \epsilon \leq 1} \psi(\rho_i, c, d) \approx \epsilon + (1 - \epsilon) \mathcal{J} \quad (14)$$

*Proof* The expectation in (10) is shown to be convex in  $\epsilon$  in [18] independent of the distribution of channel coefficients  $z = |h|^2$ . Thus, it has a unique minimizer  $\epsilon^*$  which is consequently the EC maximizer given by (14).  $\square$

Note that  $c = \theta \sqrt{T_f} Q^{-1}(\epsilon) \log_2 e$  is not a function of  $z$ , so it can be taken out of the integration which simplifies the optimization problem. To obtain the maximum per-node effective capacity  $EC_{\max}$ , we simply insert  $\epsilon^*$  into (12).

Having obtained the closed-form solution for EC, we proceed with studying the effect of multi-node interference on the per-node EC. We elaborate the effect of interference in quasi-static Rayleigh fading by plotting the per-node EC obtained from Lemma 1 and (3) for 1, 5, and 10 machines in Fig. 2. The network parameters are set as  $T_f = 1000$ ,  $\rho = 2$ , and  $\theta = 0.01$ . It is obvious that the per-node EC decreases when increasing the number of machine  $N$  as more interference is added. Notice that the EC curves are concave in  $\epsilon$  as envisaged by [18] and hence have a unique maximizer which is obtained from Lemma 2 and depicted in the figure. Another observation worth mentioning is that the optimum probability of error  $\epsilon^*$  which maximizes the EC becomes higher when increasing the number of machines. Notice that Fig. 2 confirms that Lemma 2 renders an accurate approximation to (10).

### 4 Maximization of effective capacity in the ultra-reliable region

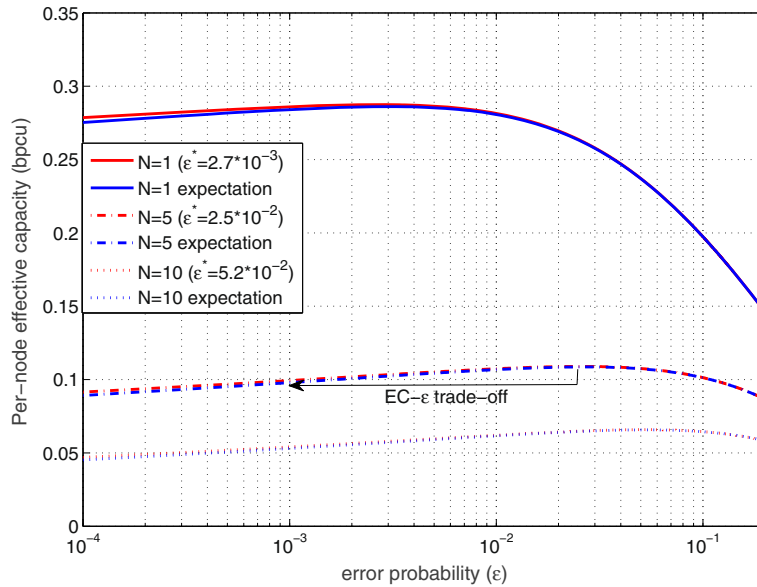
Taking a closer look at Fig. 2, we observe the trade-off between the per-node EC and error probability  $\epsilon$ . It is apparent that we can earn a lower error probability  $\epsilon$  by sacrificing only a small amount of EC. For example, we observe that for the five-node network operating in quasi-static Rayleigh fading channel, if we tolerate a decrease in the EC from 0.11 to 0.1 bpcu, the error probability  $\epsilon$  improves to  $10^{-3}$  instead of  $2.5 \times 10^{-2}$ . Thus, sacrificing only 9% of the EC maximum value boosts the error probability by nearly 1250% and hence leads to a dramatic enhancement of reliability. This observation is an open topic for analysis of the EC- $\epsilon$  trade-off with a target of maximizing the EC with some error constraint reflecting the reliability guarantees.

In this section, we discuss the EC maximization in ultra-reliable (UR) region (i.e., when the probability of error is extremely low). We maximize the EC so that the error outage probability stays below a very small target value  $\epsilon_t$ . We define the optimization problem as:

$$\begin{aligned} \max \quad & EC(\rho_i, \theta, \epsilon) \\ \text{s.t.} \quad & \epsilon \leq \epsilon_t, \end{aligned} \quad (15)$$

which according to (40) in Appendix 1 can be interpreted to

$$\begin{aligned} \min \quad & \psi(\epsilon) \approx \epsilon + (1 - \epsilon) \sum_{n=0}^2 \int_0^{\infty} (1 + \rho_i z)^d \frac{(cx)^n}{n!} e^{-z} dz, \\ \text{s.t.} \quad & \epsilon \leq \epsilon_t. \end{aligned} \quad (16)$$



**Fig. 2** Interference effect. Per-node effective capacity in bpcu as a function of error outage probability  $\epsilon$  for different number of nodes, with  $T_f = 1000, \rho = 2,$  and  $\theta = 0.01$

The solution of (16) renders the operational EC which guarantees ultra-reliability according to the error constraint. Operational EC in this case is smaller than or equal to the maximum effective capacity  $EC_{\max}$  obtained from Lemma 2. Due to the convexity of  $\psi(\epsilon)$  [18], the first derivative of  $\psi(\epsilon)$  is positive if  $\epsilon$  is greater than the global maximum and vice versa. Thus, we can check if the optimum solution is given by  $\epsilon_t$  or not through the first derivative of  $\psi(\epsilon)$ . To elucidate, we write down the Lagrangian of (16) as:

$$\mathcal{L}(\epsilon, \lambda) = \psi(\epsilon) + \lambda(\epsilon - \epsilon_t), \tag{17}$$

where  $\lambda$  is the Lagrangian multiplier. This leads to the following Karush-Kuhn-Tucker (KKT) conditions [31]

$$\frac{\partial \mathcal{L}}{\partial \epsilon} = \frac{\partial \psi(\epsilon)}{\partial \epsilon} + \lambda = 0 \tag{18a}$$

$$\lambda(\epsilon - \epsilon_t) = 0. \tag{18b}$$

From the second condition, if  $\lambda$  is greater than zero, this means that the constraint is active, where  $\epsilon^* = \epsilon_t$  and  $\frac{\partial \psi(\epsilon)}{\partial \epsilon} |_{\epsilon=\epsilon_t}$  is indeed negative. Reversing this conclusion, we can infer whether the constraint is active or not from the sign of  $\frac{\partial \psi(\epsilon)}{\partial \epsilon} |_{\epsilon=\epsilon_t}$  so that

$$\epsilon^* = \begin{cases} \epsilon_t & \frac{\partial \psi(\epsilon)}{\partial \epsilon} |_{\epsilon=\epsilon_t} < 0 \\ \arg \min_{\epsilon \geq 0} \psi(\epsilon), & \frac{\partial \psi(\epsilon)}{\partial \epsilon} |_{\epsilon=\epsilon_t} > 0 \end{cases}. \tag{19}$$

The first derivative of  $\psi(\epsilon)$  with respect to  $\epsilon$  is derived as follows. From (40), we have:

$$\begin{aligned} \mathcal{J} &= \sum_{n=0}^2 \int_0^\infty (1 + \rho iz)^d \frac{(cx)^n}{n!} e^{-z} dz \\ &= \mathcal{J}_1 + c\mathcal{J}_2 + \frac{c^2}{2}\mathcal{J}_3, \end{aligned} \tag{20}$$

where

$$\begin{aligned} \mathcal{J}_1 &= e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) \\ \mathcal{J}_2 &= e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) - \frac{1}{2} e^{\frac{1}{\rho}} \rho^{d-2} \Gamma\left(d - 2, \frac{1}{\rho}\right), \\ \mathcal{J}_3 &= e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) - e^{\frac{1}{\rho}} \rho^{d-2} \Gamma\left(d - 2, \frac{1}{\rho}\right), \end{aligned} \tag{21}$$

Let  $\delta = \theta \sqrt{T_f} \log_2 e$ , then

$$\begin{aligned} \frac{\partial \psi(\epsilon)}{\partial \epsilon} &= 1 + (1 - \epsilon) \frac{\partial \mathcal{J}}{\partial \epsilon} - \mathcal{J} \\ &= 1 + (1 - \epsilon) (\delta q(\epsilon) \mathcal{J}_2 + \delta^2 Q^{-1}(\epsilon) q(\epsilon) \mathcal{J}_3) - \mathcal{J}, \end{aligned} \tag{22}$$

and  $q(\epsilon) = \frac{\partial Q^{-1}(\epsilon)}{\partial \epsilon} = -\sqrt{2\pi} e^{\frac{Q^{-1}(\epsilon)}{2}}$  is the first derivative of  $Q^{-1}(\epsilon)$  w.r.t  $\epsilon$ . Substituting, we get

$$\frac{\partial \psi(\epsilon)}{\partial \epsilon} = 1 - (1 - \epsilon) \delta \sqrt{2\pi} e^{\frac{(Q^{-1}(\epsilon))^2}{2}} (\mathcal{J}_2 + \delta \mathcal{J}_3) - \mathcal{J}. \tag{23}$$

## 5 Methods

Given that all nodes transmit at the same time slot, the controller attempts decoding the transmitted symbols arriving from all of them. When the controller decodes one node's data, the other streams appear as interference to it [3]. For this model, imagine that a node needs to raise its EC in order to meet its QoS constraint. At first glance, applying successive interference cancellation at the base station would seem to be an attractive solution. However, this will result in extra delay for lower priority nodes where the decoder must wait for the higher priority packets to be decoded first to perform interference cancellation which dictates parallel decoding [32]. We study the interference alleviation scenarios for one node at a certain time slot, while other nodes' packets also are transmitted and decoded at the same time as a lower bound worse case.

### 5.1 Power control

The method of power control is based on increasing the SNR of node  $n$  to allow it recover from the interference effect. Let  $\rho_c$  be the new SNR of node  $n$ , while the other nodes still transmit with SNR equal to  $\rho$ . Then, we equate the SINR equation in (3) to the case where no collision occurs ( $N = 1$ ) to obtain:

$$\rho_c = \rho (1 + \rho(N - 1)). \quad (24)$$

When a certain node transmits with SNR of  $\rho_c$ , its EC is the same as in the case when transmitting with SNR equals to  $\rho$  while other nodes are silent. The method of power control is simple; however, it causes extra interference into the other nodes due to the power increase of the recovering node.

From (24), we define the SINR of other nodes colliding in the same network (nodes in set  $s$ ) after the compensation of one node as:

$$\rho_s = \frac{\rho}{1 + \rho_c + \rho(N - 2)} = \frac{\rho}{1 + \rho(\rho + 1)(N - 1)} \quad (25)$$

Now, we are interested in comparing the per-node EC in three cases: (i) no collision, (ii) collision without compensation, and (iii) collision with compensation of one node. Here, we look for the maximum achievable per-node EC in each case. Define the collision loss factor  $\alpha$  as the ratio between the maximum effective capacity  $EC_{\max}$  in case of collision and in case of no collision as:

$$\alpha = \frac{EC(\rho_i, \theta, \epsilon_i^*)}{EC(\rho, \theta, \epsilon^*)} \quad (26)$$

where  $\epsilon^*$  and  $\epsilon_i^*$  are the optimal error probabilities for the cases of no collision ( $N = 1$ ) and collision without compensation, respectively, and both are obtained from (19).

To determine the effect of compensation of one node on the other nodes, we define the compensation loss factor  $\alpha_c$  as the ratio between maximum EC of other nodes (set  $s$ ) in case of one node compensation and in case of no compensation, that is:

$$\alpha_c = \frac{EC(\rho_s, \theta, \epsilon_s^*)}{EC(\rho_i, \theta, \epsilon_i^*)} \quad (27)$$

where  $\epsilon_s^*$  is the optimum error probability obtained from (14) when the SINR is set to  $\rho_s$ . To understand the effect of increased interference on the network performance, we study the effect of SINR variations on EC for different delay constraints.

**Proposition 1** *SINR variations have comparably limited effect on EC when the delay constraint becomes more strict and vice versa.*

*Proof* Please refer to Appendix 2.  $\square$

Furthermore, we include the compensation factor  $\gamma_c$  as the ratio of the maximum achievable EC of the compensated node after and before compensation which is expressed as:

$$\gamma_c = \frac{EC(\rho, \theta, \epsilon^*)}{EC(\rho_i, \theta, \epsilon_i^*)} = \frac{1}{\alpha} \quad (28)$$

where  $\gamma_c$  is a gain factor (i.e.,  $\gamma_c \geq 1$ ). Finally, we define the total loss factor  $\alpha_t$  as the ratio between the maximum attainable effective capacity of colliding nodes in the system (set  $s$ ) when a node compensates to the maximum attainable EC of these nodes if they were not colliding at all, that is:

$$\alpha_t = \frac{\alpha_c}{\gamma_c} = \alpha \cdot \alpha_c \quad (29)$$

### 5.2 Graceful degradation of the delay constraint

Here, we discuss how to compensate for the decrease in the per-user EC for the multiuser interference scenario by changing the value of delay constraint  $\theta$ . More specifically, we determine how the delay exponent  $\theta$  should be gracefully degraded to obtain the same  $EC_{\max}$  as if the target node was transmitting without collision. This represents the cases where a node has flexible QoS constraint delay wise, so that the EC could be attained given a slight variation on the overall delay as envisioned in [1]. Let  $\theta$  be the original delay exponent and  $\theta_i$  represent the new gracefully degraded one;  $\theta_i$  is obtained by solving:

$$EC(\rho, \theta, \epsilon^*) = EC(\rho_i, \theta_i, \epsilon_i^*) \quad (30)$$

where  $\epsilon_i^*$  is the maximizer of EC for the parameters  $\rho_i$  and  $\theta_i$ , and  $\epsilon_i^*$  is the optimum error probability for  $\rho_i$  and  $\theta_i$ . The solution of (30) renders the necessary value of  $\theta_i$  to compensate for the EC decrease due to collision in

this case. Notice that (30) can be solved numerically to obtain the necessary value of  $\theta_2$  to compensate for the rate decrease due to collision in this case.

### 5.3 Joint compensation model

To mitigate the side effects of power control and graceful delay constraint degradation, we apply a joint model in which both methods are partially employed. Define the operational SINR in power controlled compensation for nodes in set  $s$  as  $\rho_{s_o}$ , where  $\rho_{s_o}$  lies on the interval  $[\rho_s, \rho_i]$ . Using (25), the operational SNR for the recovering node can be written as:

$$\rho_{c_o} = \frac{\rho}{\rho_{s_o}} - 1 - \rho(N - 2), \tag{31}$$

and the operational point of the compensation loss factor  $\alpha_{c_o}$  is

$$\alpha_{c_o} = \frac{EC(\rho_{s_o}, \theta, \epsilon_{s_o}^*)}{EC(\rho_i, \theta, \epsilon_i^*)} \tag{32}$$

where  $\epsilon_{s_o}^*$  is the optimum error probability obtained from (14) for the parameters  $\rho_{s_o}$  and  $\theta_1$ .  $\alpha_{c_o}$  is considered to be the loss factor caused by the part of compensation performed via power control.

Next, we perform the rest of compensation via graceful degradation of  $\theta$  as in Section 5.2. To obtain  $\theta_2$ , we solve:

$$EC(\rho, \theta, \epsilon^*) = EC\left(\frac{\rho_{c_o}}{1 + \rho(N - 1)}, \theta_2, \epsilon_2^*\right) \tag{33}$$

From (33), we compute the necessary value of  $\theta_2$  to continue the compensation process via graceful degradation of the delay constraint.

Now, we propose an objective function leveraging the network performance for the joint model. First, we define the priority factor  $\eta_\alpha$  as a measure of the risk of decrease in EC of nodes in set  $s$  when the compensating node boosts its transmission power. In other words, the higher the value of  $\eta_\alpha$ , the more important it is not to allow much degradation of EC of nodes in set  $s$ , and hence, we try not to compensate via power control and shift compensation towards  $\theta$  graceful degradation. On the other hand, we define the priority factor  $\eta_\theta$  as a measure of strictness of the delay constraint (i.e., the higher the value of  $\eta_\theta$ , the more strict it is not to degrade delay constraint and hence the less we are allowed to relax  $\theta$  to get higher EC for the compensating user). Thus, we can formalize our objective function as the summation:

$$\eta = \eta_\alpha \alpha_{c_o} + \eta_\theta \theta_2 \tag{34}$$

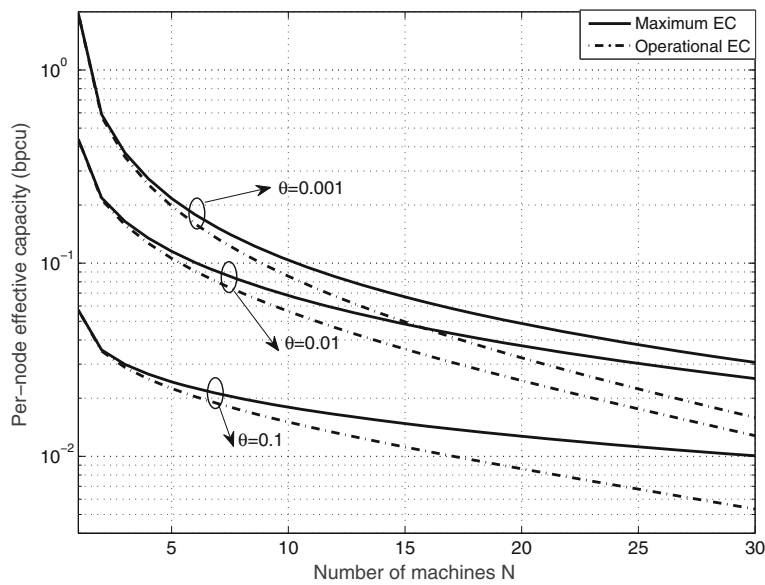
where  $\alpha_{c_o}$  and  $\theta_2$  are the operational point. Now, we choose this operational point to satisfy:

$$\begin{aligned} \eta_{\max} &= \max_{\theta_2 \geq 0} \eta_\alpha \alpha_{c_o} + \eta_\theta \theta_2 \\ \text{s.t. } &\rho_s \leq \rho_{s_o} \leq \rho_i \end{aligned} \tag{35}$$

where the solution to this problem gives the optimum operational point which can be found from (31), (32), and (33).

## 6 Results and discussion

Figure 3 depicts the operational and maximum EC as a function of the number of nodes  $N$  for  $T_f = 1000$ ,  $\rho = 10$  dB,  $\epsilon_t$  different values of  $\theta$ . The figure reveals that operating in the UR region necessitates a considerable



**Fig. 3** Effective capacity in the ultra-reliable region. Operational and maximum effective capacity as a function of the number of machines  $N$  for  $T_f = 1000$  and  $\rho = 10$



sacrifice in the EC. Specifically, as the number of nodes increases and the SINR decreases, the amount of EC sacrifice becomes more significant as the gap between operational and optimum EC becomes higher. For example, at  $N = 30$ , the nodes lose up to 50% of its EC to maintain reliability.

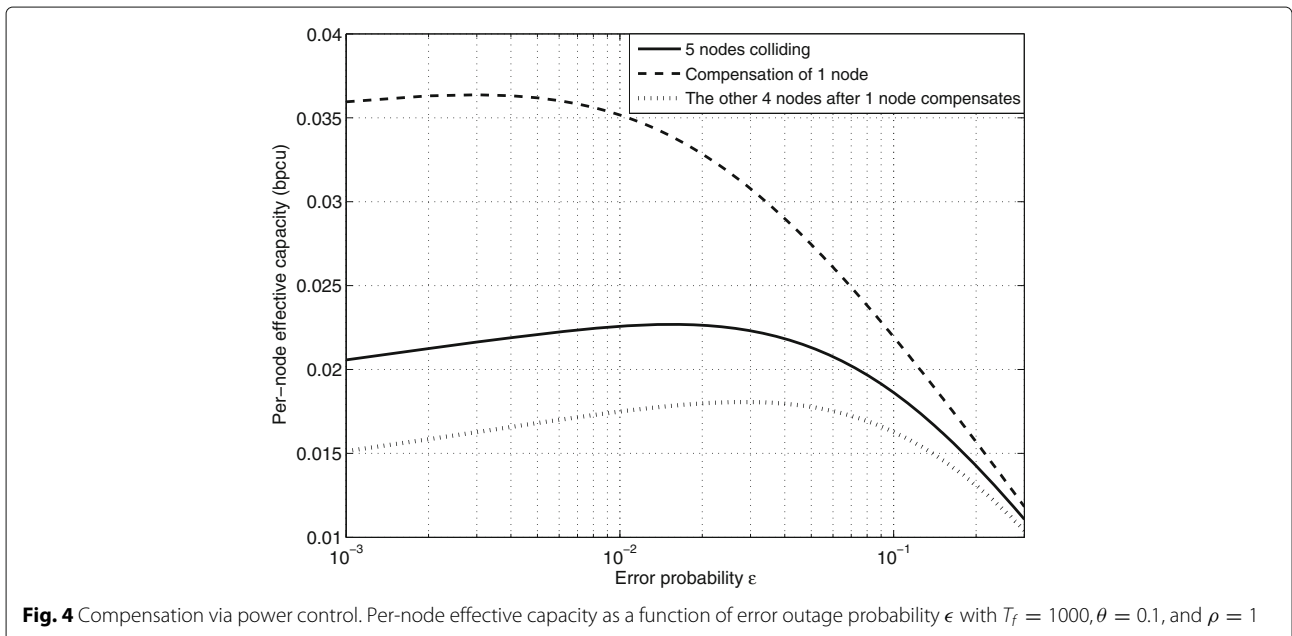
As an example, consider three colliding nodes where  $T_f = 500, \theta = 0.01$ , and  $\rho = 0.5$ . Applying (24), we get  $\rho_c = 1$ . Hence, the interference effect is canceled for a certain node by boosting its SNR from 0.5 to 1. To elucidate more, Fig. 4 shows the effect of the collision of five nodes with and without compensation for  $T_f = 1000, \theta = 0.1$ , and  $\rho = 1$ . We plot the per-node EC before the compensation of one node. Then, we compare it to the effective capacities of the four remaining nodes after one node compensates using (24). The figure also shows the EC of the compensating node.

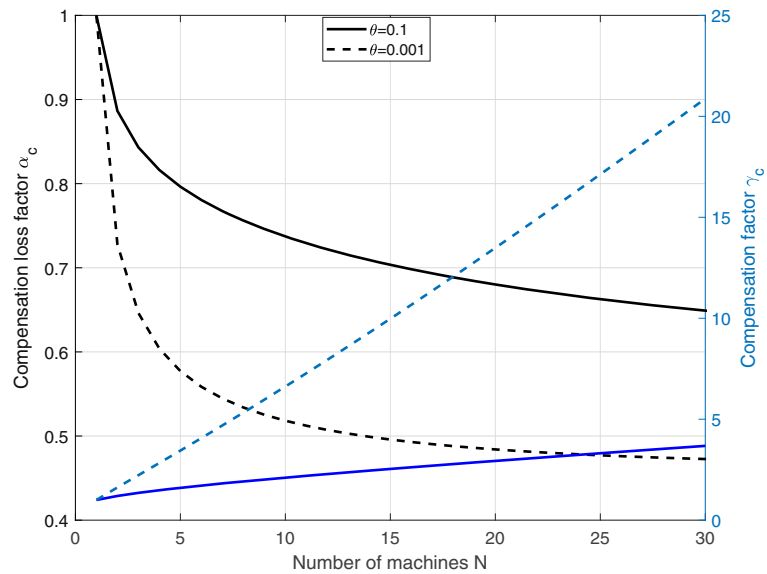
Consider  $T_f = 1000$  and  $\rho = 1$ , then the left axis of Fig. 5 depicts the compensation loss factor  $\alpha_c$  for different number of nodes  $N$  with  $\theta = 0.1$  and  $0.001$ . The figure shows that  $\alpha_c$  is lower for smaller values of  $\theta$ . Hence, the effect of compensation appears to be more severe for less stringent delay constraints. This follows from Proposition 1, which states that SINR variations have less of an effect on delay strict networks and vice versa. Finally, we notice that the compensation loss factor decreases rapidly for a less dense network. The right axis of Fig. 5 shows the compensation factor  $\gamma_c$  vs the number of nodes in the system  $N$ .  $\gamma_c$  appears to have a linear behavior as a function of  $N$ . That is, the effect of compensation for the compensated node increases linearly with  $N$ . The rate by which  $\gamma_c$  increases is faster for smaller  $\theta$ . The compensation factor

$\gamma_c$  (compensation gain) is higher for less stringent delay constraint (less  $\theta$ ). It appears that  $\alpha_c$  and  $\gamma_c$  are inversely correlated to each other as envisioned by (28).

The collision loss factor  $\alpha$  and the total loss factor  $\alpha_t$  are also depicted in Fig. 6. As observed from the figure, the total loss is nearly the same as the collision loss for small  $N$ . The gap starts to appear and becomes almost constant for high number of nodes. This gap is tighter in the case of small  $\theta$  (i.e., the collision loss  $\alpha$  is more dominant). Also, there is a higher loss (both collision and total) in case of less stringent delay constraint  $\theta$ . Thus, the effect of collision and compensation is more annoying in case of less stringent delay constraint (smaller  $\theta$ ). Furthermore,  $\alpha$  and  $\alpha_t$  have the same behavior as  $\alpha_c$ . They decrease rapidly for small number of nodes and tend to be constant for high  $N$ .

Consider (30) with  $N = 5, \theta_1 = 0.05, \rho = 1$ , and  $T_f = 1000$ , we get  $\theta_i = 0.023$ . Thus, by gracefully degrading the delay constraint from 0.05 to 0.023, we attain the same value for the maximum effective capacity  $EC_{max} = 0.066$  as was depicted in Fig. 4 in [26]. Here, in Fig. 7, we illustrate the graceful degradation of the delay constraint by plotting the delay outage probability  $P_{out\_delay}$  as a function of the maximum delay bound  $D_{max}$  before and after compensation. The figure shows that for a delay outage probability of  $10^{-3}$ , the compensation process is performed by extending the allowable delay  $D_{max}$  from 3600 to 4600 channel uses. We perform a limited delay extension ( $\approx 25\%$ ) because the rise in EC partially compensates the graceful degradation of  $\theta$  in (8). Note that the optimum error probabilities have different values in each case due to the change in SINR in (13) [26].



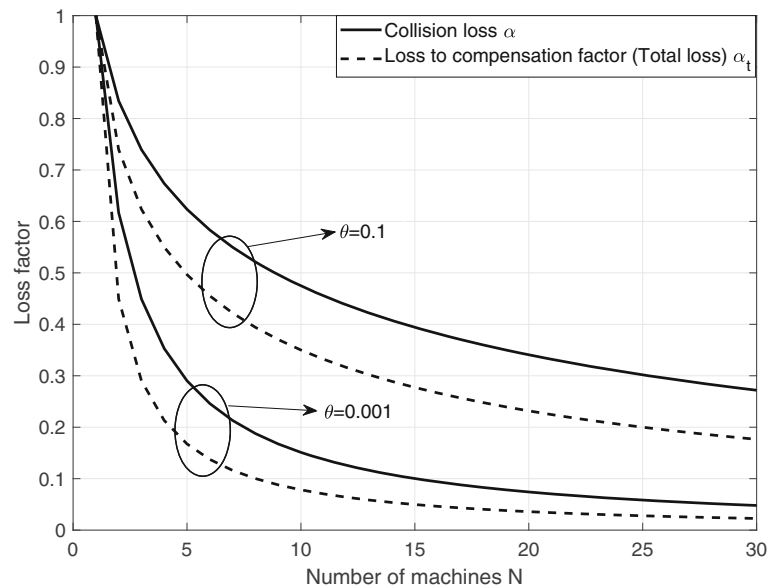


**Fig. 5** Power-controlled compensation loss and compensation factors. Compensation loss factor  $\alpha_c$  and compensation factor  $\gamma_c$  as a function of the number of nodes  $N$ , for distinct QoS exponents

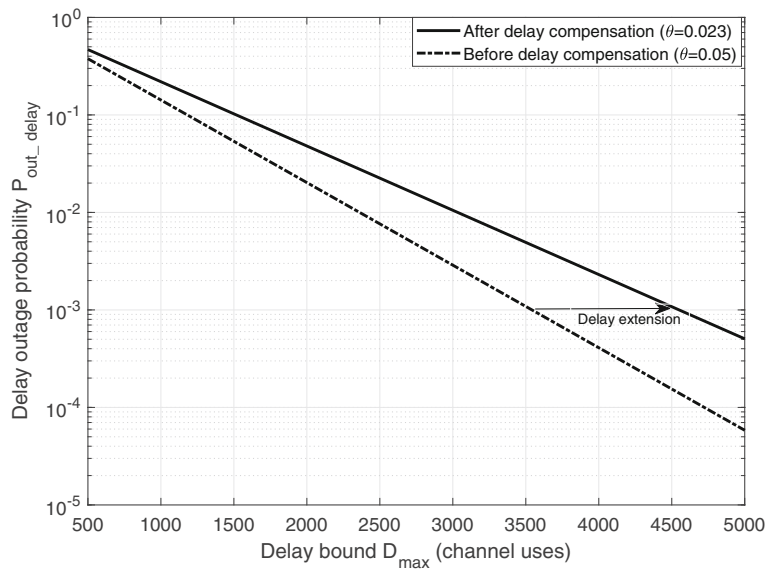
Figure 8 illustrates different operational points for the joint model for different blocklength  $T_f$  where  $N = 5$ ,  $\rho = 1$ ,  $P_{out\_delay} = 10^{-3}$ , and  $\theta_1 = 0.1$ . For example, when  $T_f = 700$ , we select the operational point  $\alpha_{c_o} = 0.9$  and  $\theta_2 = 0.075$ . This implies that a part of compensation will be performed via power control, which leads to a 10% loss in EC of other nodes (set  $s$ ). Then, the rest of the compensation will be performed by gracefully degrading its  $\theta$  from 0.1 to 0.075. The maximum delay of the recovering

node remains 2500 channel uses before and after recovery as restoring the EC compensates for the decrease in  $\theta$  in (8). The figure also shows that for smaller packet sizes, the amount of losses due to compensation is higher.

According to the system parameters, certain values of the priority factors  $\eta_\alpha$  and  $\eta_\theta$  may produce a concave maximization problem for the objective function  $\eta$ . For an MTC network with 15 devices where  $T_f = 1000$ ,  $\rho = 2$ ,  $\theta_1 = 0.1$ ,  $\eta_\alpha = 1$ , and  $\eta_\theta = 4$ , the optimum value of



**Fig. 6** Power-controlled loss factors. Loss factors  $\alpha$  and  $\alpha_t$  vs number of nodes  $N$

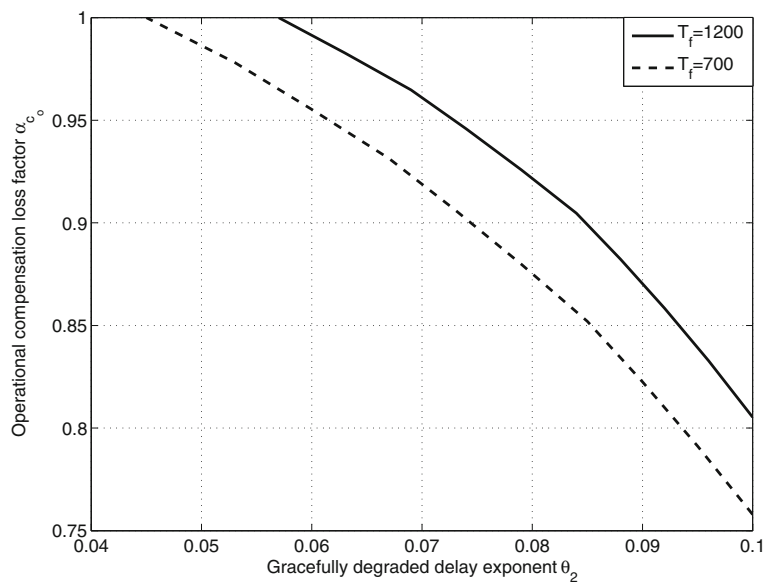


**Fig. 7** Compensation via graceful degradation of delay constraint. Graceful degradation of delay constraint  $\theta$  in case of five nodes colliding, where  $T_f = 1000$  and  $\rho = 1$

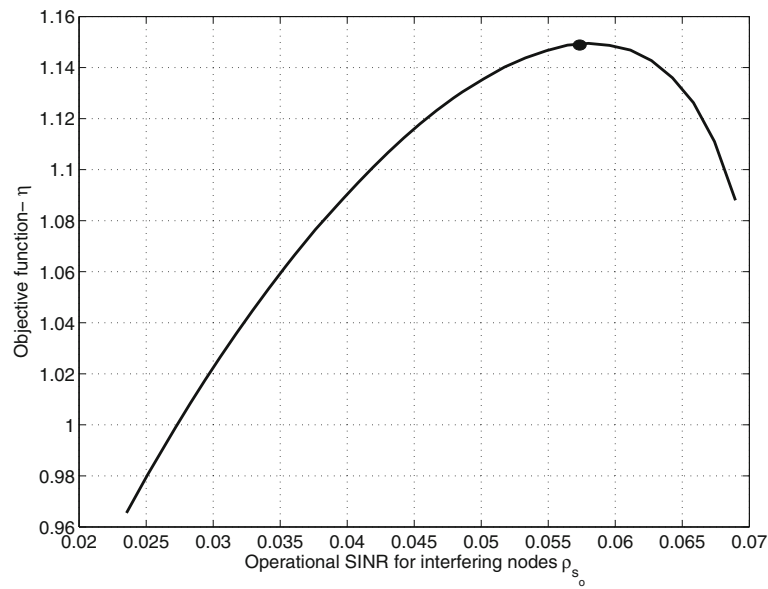
$\rho_{s_0}$  which maximizes the objective function  $\eta$  will be 0.057 according to Fig. 9. This value corresponds to the operational point  $\alpha_{c_0} = 0.9397$  and  $\theta_2 = 0.053$ . The SNR of the recovering node becomes  $\rho_{c_0} = 8.08$ . In other words, in order to maximize the network throughput according to the given priority factors, the compensating node boosts its SNR from 2 to 8.08 and gracefully degrades its delay exponent from 0.1 to 0.053. This results in only 6% loss

in EC of other nodes as depicted in Fig. 10. Priority factors are left for the designer’s preferences depending on the reliability or latency requirements.

Finally in Fig. 11, we compare the EC vs the delay constraint  $\theta$  for fixed and variable rate transmissions. The SNR and the blocklength are set as  $\rho = 1$  and  $T_f = 1000$ , respectively. It can be observed that for high values of  $\theta$ , fixed-rate transmission performs strictly better. The figure



**Fig. 8** Joint compensation trade-off. Trade-off between compensation loss factor via power control  $\alpha_{c_0}$  and graceful degradation of delay constraint  $\theta_2$  for different operational points



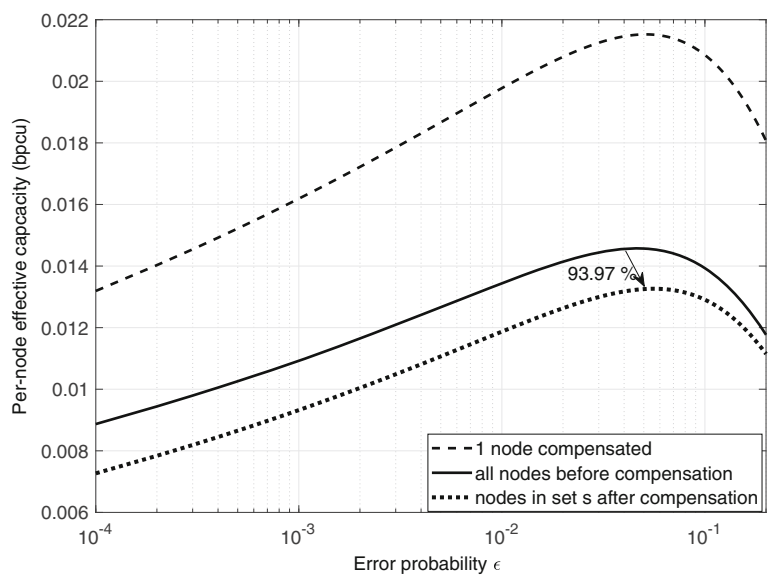
**Fig. 9** Throughput of joint compensation. Throughput  $\eta$  for  $T_f = 1000, \theta = 0.1, \rho = 2$ , and  $N = 15$

also confirms the fact that EC degrades with the increase of delay constraint  $\theta$ .

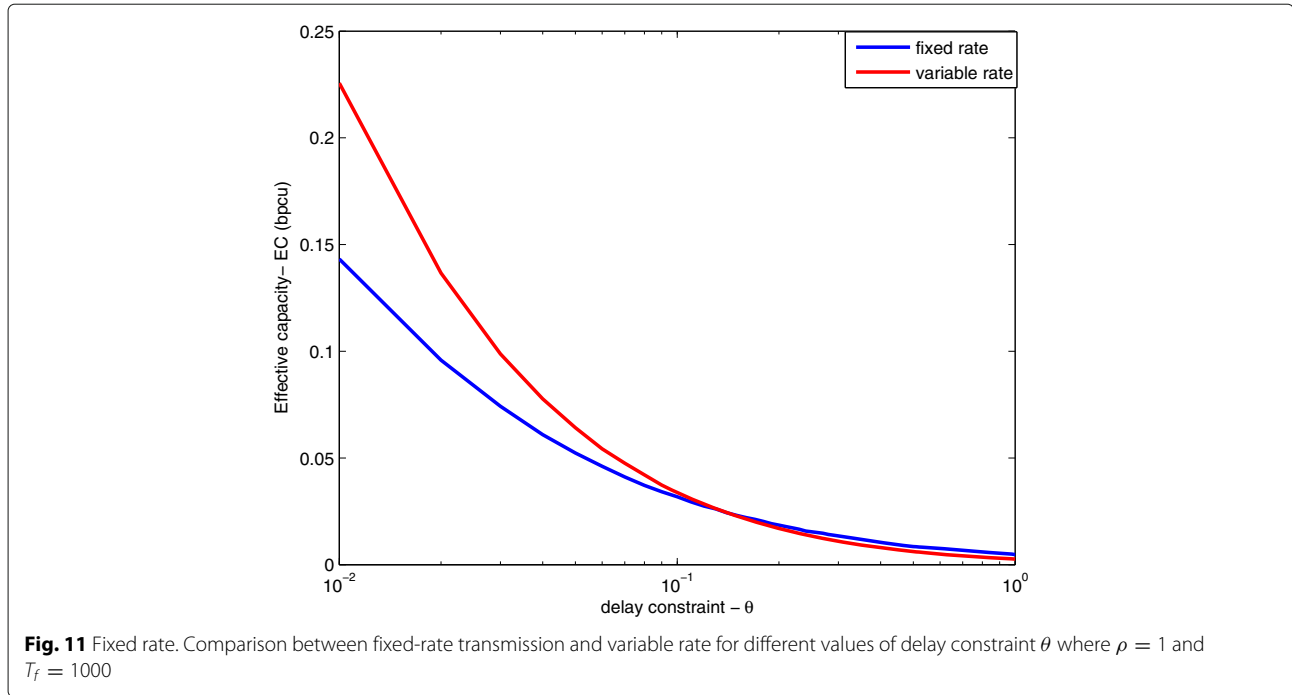
### 7 Conclusions

In this work, we presented a detailed analysis of the EC for delay-constrained MTC networks in the finite blocklength regime. For quasi-static Rayleigh fading channels, we proposed an approximation for the EC and defined the optimum error probability. We characterized the optimization

problem to maximize EC with error constraint which showed that there is a relatively small sacrifice in EC for high SINR. Our analysis indicated that SINR variations have minimum effect on EC under strict delay constraints. In a dense MTC network scenario, we illustrated the effect of interference on EC. We proposed power control as an adequate method to restore the EC in networks with less stringent delay constraints. Another method is graceful degradation of delay constraint, where we showed that a



**Fig. 10** Joint compensation. Per-node effective capacity as a function of error probability  $\epsilon$  before and after joint compensation for  $T_f = 1000, \theta_1 = 0.1, \rho = 2$ , and  $N = 15$



very limited extension in the delay bound could successfully recover the EC. Joint compensation emerges as a combination between these two methods, where an operational point is selected to maximize an objective function according to the networks design aspects. Finally, we concluded that for high values of  $\theta$ , fixed-rate transmission performs strictly better. As future work, we aim to analyze the impact of imperfect CSI on the EC and coordination algorithms that maximize EC with fairness constraints.

**Appendix 1: Proof of Lemma 1**

For Rayleigh envelope of PDF given in (7), the EC expression in (10) can be written as:

$$\psi(\rho_i, \theta, \epsilon) = \int_0^\infty \left( \epsilon + (1 - \epsilon)e^{-\theta T_f r} \right) e^{-z} dz. \tag{36}$$

From (4), we have

$$\begin{aligned} \exp(-\theta T_f r) &= \exp(-\theta T_f \log_2(1 + \rho_i z)) \times \\ &\exp\left(\theta \sqrt{T_f \left(1 - \frac{1}{(1 + \rho_i z)^2}\right)} Q^{-1}(\epsilon) \log_2 e\right). \end{aligned} \tag{37}$$

Elaborating, we attain

$$\exp(-\theta T_f \log_2(1 + \rho_i z)) = (1 + \rho_i z)^d, \tag{38}$$

$$\exp\left(\theta \sqrt{T_f \left(1 - \frac{1}{(1 + \rho_i z)^2}\right)} Q^{-1}(\epsilon) \log_2 e\right) = \exp(cx). \tag{39}$$

We resort to the second order Taylor expansion to obtain  $e^{cx} = 1 + cx + \frac{(cx)^2}{2}$  and place it into (37), then (36) becomes:

$$\begin{aligned} \psi(\rho, \theta, \epsilon) &= \epsilon \int_0^\infty e^{-z} dz + (1 - \epsilon) \left( \int_0^\infty (1 + \rho z)^d e^{-z} dz \right. \\ &\quad \left. + c \int_0^\infty (1 + \rho z)^d x e^{-z} dz + \frac{c^2}{2} \int_0^\infty (1 + \rho z)^d x^2 e^{-z} dz \right) \end{aligned} \tag{40}$$

The first integral reduces to unity. Applying Laurent's expansion for  $x$  [33], we attain:

$$x \approx 1 - \frac{1}{2(1 + \rho z)^2}. \tag{41}$$

Replacing (41) into (40) yields

$$\begin{aligned} \psi(\rho, \theta, \epsilon) &= \epsilon + (1 - \epsilon) \left[ e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) \right. \\ &\quad \left. + c \left( e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) - \frac{1}{2} e^{\frac{1}{\rho}} \rho^{d-2} \Gamma\left(d - 2, \frac{1}{\rho}\right) \right) \right. \\ &\quad \left. + \frac{c^2}{2} \left( e^{\frac{1}{\rho}} \rho^d \Gamma\left(d + 1, \frac{1}{\rho}\right) - e^{\frac{1}{\rho}} \rho^{d-2} \Gamma\left(d - 2, \frac{1}{\rho}\right) \right) \right]. \end{aligned} \tag{42}$$

After manipulating (42) and inserting it into (10), we obtain (12).

**Appendix 2: Proof of Proposition 1**

Differentiating (10) with respect to  $\rho_i$

$$\frac{\partial EC}{\partial \rho_i} = \frac{\partial EC}{\partial r} \frac{\partial r}{\partial \rho_i} = \frac{e^{-T_f \theta r}}{\epsilon + (1 - \epsilon)e^{-T_f \theta r}} \mathcal{K},$$

where  $\mathcal{K} = \frac{\partial r}{\partial \rho_i} (1 - \epsilon)$  is strictly positive since the achievable rate  $r$  is an increasing function of the SINR  $\rho_i$ . Differentiating once more with respect to  $\theta$ :

$$\frac{\partial}{\partial \theta} \left( \frac{\partial \text{EC}}{\partial \rho_i} \right) = - \frac{\mathcal{K} T_f r e^{-T_f \theta r}}{(\epsilon + (1 - \epsilon) e^{-T_f \theta r})^2}, \quad (43)$$

which is strictly negative and thus validating our proposition.

#### Funding

This work is partially supported by Aka Project SAFE (Grant no. 303532) and by Finnish Funding Agency for Technology and Innovation (Tekes), Bittium Wireless, Keysight Technologies Finland, Kyynel, MediaTek Wireless, and Nokia Solutions and Networks.

#### Availability of data and materials

The paper is self-contained, since we provide a mathematical framework which can be reproduced with the details provided in Sections 2 to 5, and in Section 5, numerical results and parameter settings are described in details.

#### Authors' contributions

MS derived the equations and performed the system simulations. ED revised the equations and contributed to the writing of the introduction, system model, and conclusion. HA supervised and reviewed the paper, while ML directed and supervised the research. All authors participated in this work, and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

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

Received: 18 October 2017 Accepted: 30 July 2018

Published online: 10 August 2018

#### References

1. P Popovski, in *1st International Conference on 5G for Ubiquitous Connectivity*. Ultra-reliable communication in 5G wireless systems, (Akasloppolo, 2014), pp. 146–151
2. E Dosti, UL Wijewardhana, H Alves, M Latva-aho, in *IEEE ICC 2017*. Ultra reliable communication via optimum power allocation for type-I ARQ in finite block-length (IEEE, Paris, 2017), pp. 5168–5173
3. R Devassy, G Durisi, P Popovski, EG Strom, Finite-blocklength analysis of the ARQ-protocol throughput over the Gaussian collision channel. 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), Athens, 173–177 (2014)
4. Nokia, 5G for Mission Critical Communication: achieve ultra-reliability and virtual zero latency. Nokia White Paper (2016)
5. H Shariatmadari, S Iraj, R Jantti, Analysis of transmission methods for ultra-reliable communications. IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC. **2015-December**, 2303–2308 (2015)
6. E Dosti, M Shehab, H Alves, M Latva-aho, in *2017 European Conference on Networks and Communications (EuCNC)*. Ultra reliable communication via CC-HARQ in finite block-length, (2017), pp. 1–5. <https://doi.org/10.1109/EuCNC.2017.7980708>
7. Ericsson, Cellular networks for massive IoT. Ericsson White Paper (2016)
8. P Schulz, et al., Latency critical IoT applications in 5G: perspective on the design of radio interface and network architecture. IEEE Commun. Mag. **55**(2), 70–78 (2017). <https://doi.org/10.1109/MCOM.2017.1600435CM>
9. J Oestman, G Durisi, EG Stroem, J Li, H Sahlh, G Liva, in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*. Low-latency ultra-reliable 5G communications: finite-blocklength bounds and coding schemes, (Hamburg, 2017), pp. 1–6
10. A Ali, W Hamouda, M Uysal, Next generation M2M cellular networks: challenges and practical considerations. IEEE Commun. Mag. **53**(9), 18–24 (2015). <https://doi.org/10.1109/MCOM.2015.7263368>
11. H Chen, et al., in *2016 IEEE GLOBECOM*. Cooperative retransmission for massive MTC under spatiotemporally correlated interference, (2016), pp. 1–6. <https://doi.org/10.1109/GLOCOM.2016.7841698>
12. E Soltanmohammadi, K Ghavami, M Naraghi-Pour, A survey of traffic issues in machine-to-machine communications over LTE. IEEE Internet Things J. **3**(6), 865–884 (2016). <https://doi.org/10.1109/JIOT.2016.2533541>
13. A Orsino, et al., Effects of heterogeneous mobility on D2D- and drone-assisted mission-critical MTC in 5G. IEEE Commun. Mag. **55**(2), 79–87 (2017)
14. JF Monserrat, G Mange, V Braun, H Tullberg, G Zimmermann, Ö Bulakci, Metis research advances towards the 5G mobile and wireless system definition. EURASIP J. Wirel. Commun. Netw. **2015**(1), 53 (2015). <https://doi.org/10.1186/s13638-015-0302-9>
15. Y Mehmood, C Görg, M Muehleisen, A Timm-Giel, Mobile M2M communication architectures, upcoming challenges, applications, and future directions. EURASIP J. Wirel. Commun. Netw. **2015**(1), 250 (2015). <https://doi.org/10.1186/s13638-015-0479-y>
16. CE Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**(July 1928), 379–423 (1948)
17. G Durisi, T Koch, P Popovski, Toward massive, ultrareliable, and low-latency wireless communication with short packets. Proc. IEEE. **104**(9), 1711–1726 (2016)
18. M Gursoy, in *EURASIP Journal on Wireless Communications and Networking 2013*. Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime, (2013)
19. W Yang, G Durisi, T Koch, Y Polyanskiy, Block-fading channels at finite blocklength. ISWCS 2013. **9**, 410–413 (2013)
20. Y Hu, J Gross, A Schmeink, On the capacity of relaying with finite blocklength. IEEE Trans. Veh. Technol. **65**(3), 1790–1794 (2016)
21. Y Hu, A Schmeink, J Gross, Blocklength-limited performance of relaying under quasi-static Rayleigh channels. IEEE Trans. Wirel. Commun. **2**(4), 630–643 (2016)
22. D Wu, R Negi, Effective capacity: a wireless link model for support of quality of service. IEEE Trans. Wirel. Commun. **2**(4), 630–643 (2003)
23. B Soret, et al., in *2014 IEEE Globecom Workshops (GC Wkshps)*. Fundamental tradeoffs among reliability, latency and throughput in cellular networks, (2014), pp. 1391–1396. <https://doi.org/10.1109/GLOCOMW.2014.7063628>
24. L Musavian, T Le-Ngoc, in *Trans. Emerging Tel. Tech*. QoS-based power allocation for cognitive radios with AMC and ARQ in Nakagami-m fading channels, vol. 27 (Wiley, New York, 2014), pp. 266–277
25. L Musavian, Q Ni, Effective capacity maximization with statistical delay and effective energy efficiency requirements. IEEE Trans. Wirel. Commun. **14**(7), 3824–3835 (2015). <https://doi.org/10.1109/TWC.2015.2412542>
26. M Shehab, E Dosti, H Alves, M Latva-aho, in *EuCNC 2017*. On the effective capacity of MTC networks in the finite blocklength regime (IEEE, Oulu, 2017)
27. D Tse, P Viswanath, *Fundamentals of wireless communication* (2005)
28. MK Simon, A M-S, *Digital communication over fading channels: a unified approach to performance analysis*. (Wiley, New York, 2005)
29. MC Gursoy, Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime. IEEE Int. Conf. Commun. Kyoto, 1–5 (2011)
30. IS Gradshteyn, IM Ryzhik, *Table of integrals, series, and products, 7th edn*. (Academic Press, London, 2007)
31. S Boyd, L Vandenberghe, *Convex optimization*. (Cambridge University Press, New York, 2004)
32. M Shirvanimoghaddam, M Condoluci, M Dohler, SJ Johnson, On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT. IEEE J. Sel. Areas Commun. **35**(10), 2238–2252 (2017). <https://doi.org/10.1109/JSAC.2017.2724442>
33. R Rodriguez, K Irwin, J Gilman, *Complex analysis: in the spirit of Lipman Bers*. (Springer-Verlag, New York, 2012)