On the Effective Capacity of MTC Networks in the Finite Blocklength Regime

Mohammad Shehab, Endrit Dosti, Hirley Alves, and Matti Latva-aho
Centre for Wireless Communications (CWC), University of Oulu, Finland
Email: firstname.lastname@oulu.fi

Abstract—This paper analyzes the effective capacity (EC) of delay constrained machine type communication (MTC) networks operating in the finite blocklength (FB) regime. First, we derive a closed-form mathematical approximation for the EC in Rayleigh block fading channels. We characterize the optimum error probability to maximize the concave EC function and study the effect of SINR variations for different delay constraints. Our analysis reveals that SINR variations have less impact on EC 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 EC 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 maximization of an objective function whose parameters are determined according to 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.

I. INTRODUCTION

Nowadays, the road to the fifth generation of mobile communication is being paved by leaps and bounds. 5G should be able to support new features such as ultra reliable transmission and massive machine-to-machine (M2M) communication [1]. MTC has gained an increasing interest in recent years [2]. MTC networks are expected to connect massive number of devices which communicate with high reliability and minimum latency to support mission critical applications, envisaged as *Ultra-Reliable Communication* (URC) [1], [3]–[5]. Furthermore, cooperative transmission in MTC is a hot topic for research so far [6].

Unlike traditional communication systems, which are based on infinite coding schemes, MTC networks are designed to communicate on short packets and delay limited basis. Such demands stimulated a revolutionary trend in information theory studying communication at finite blocklength (FB) [1], [4], [7]–[9]. In [4], a per-node throughput model was introduced for AWGN and quasi-static collision channels where interference is treated as additive Gaussian noise while considering average delay. In [9], Yang et al. characterized the attainable rate as a function of blocklength and error probability ϵ for block fading channels.

To model the delay requirements in MTC networks, we resort to the effective capacity metric which was introduced in [10] to provide an indication of the maximum possible arrival rate that can be supported by a network subject to a particular latency requirement. A statistical model for a single node effective rate in bits per channel use (bpcu) for a certain error probability with delay exponent was discussed in [8] for Rayleigh block fading channels where the channel coefficients are assumed to be fixed for a block of T_f symbols. However, a closed-form expression for the EC was not provided in these discussions. Latency-throughput tradeoff was characterized in [11] for cellular networks exploiting the EC theory. Musavian et al. analyzed the EC maximization of secondary node with some interference power constraints for primary node in a cognitive radio environment with interference constraints [12]. To the best of our knowledge, EC for FB packets transmission in multi-node MTC scenario has not been investigated. For convenience in this paper, we refer to one machine terminal as node.

In this paper, we derive a mathematical expression for EC in quasi-static Rayleigh fading for delay limited networks. This leads us to characterize the optimum error probability which maximizes the 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.

II. SYSTEM LAYOUT

A. Network model

We consider a transmission scenario in which N nodes transmit packets with equal power to a common controller through a Rayleigh block fading collision channel with blocklength T_f as shown in Fig. 1. The received vector $\mathbf{y}_n \in \mathbb{C}^n$ of node n is given by

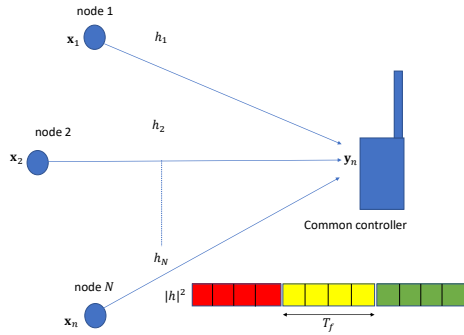$$\mathbf{y}_n = h_n \mathbf{x}_n + \sum_{s \neq n} h_s \mathbf{x}_s + \mathbf{w}, \tag{1}$$

Fig. 1. Network Layout.

where $\mathbf{x}_n \in \mathbb{C}^n$ is the transmitted packet of node n, h_n is the fading coefficient for node n which is assumed to be quasi-static with Rayleigh distribution and thus, remains constant over T_f symbols which span the whole packet duration. 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 of unit variance. Given the SNR ρ of a single node, the SINR of any node n is

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

We assume that all nodes are equidistant from the common controller and CSI is available at each node and the controller. Thus, as the number of nodes increases, the sum of Rayleigh distributed fading envelopes of $N - 1$ interfering nodes becomes $\sum_s |h_s|^2 \approx N - 1$ [13] and the interference resulting from nodes in set s can be modeled as in [4] where (2) reduces to

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

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 [4] we aim to provide a performance benchmark for such networks without interference coordination.

B. Communication at Finite Blocklength

We start by presenting 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 ϵ [7]. In this case, ϵ has a small value but not vanishing. For error probability $\epsilon \in [0, 1]$, the normalized achievable rate in bpcu is given by

$$\begin{aligned} r \approx &\log_2(1 + \rho_i |h|^2) \\ &- \sqrt{\frac{1}{T_f}\left(1 - \frac{1}{(1 + \rho_i |h|^2)^2}\right)}Q^{-1}(\epsilon)\log_2(e), \end{aligned} \tag{4}$$

where $Q(\cdot) = \int_\cdot^\infty \frac{1}{\sqrt{2\pi}}e^{\frac{-t^2}{2}}\,dt$ is the Gaussian Q-function, and $Q^{-1}(\cdot)$ represents its inverse, ρ_i is the SINR and $|h|^2$ is the fading envelope.

III. EFFECTIVE CAPACITY ANALYSIS UNDER FINITE BLOCKLENGTH

The concept of EC indicates the capability of communication nodes to exchange data with maximum rate and certain latency constraint. An outage occurs when a packet delay exceeds a maximum delay bound D_{max} and its probability is defined as [10]

$$P_{out\_delay} = Pr(delay \geqslant D_{max}) \approx e^{-\theta \cdot EC \cdot D_{max}}. \tag{5}$$

Conventionally, a network's tolerance to long delay is measured by the delay exponent θ. The network has more tolerance to large delays for small values of θ (i.e., $\theta \to 0$), while for large values of θ, it becomes more delay strict. For example, a network with unity EC and an outage probability of 10^{-3} can tolerate a maximum delay of 691 symbol periods for $\theta = 0.01$ and 23 symbol periods when $\theta = 0.3$. In quasi-static fading, the channel remains constant within each transmission period T_f [14], and the EC is [8]

$$EC(\rho_i, \theta, \epsilon) = -\frac{1}{T_f\theta}\ln\left(E_{z=|h|^2}\left[\epsilon + (1 - \epsilon)e^{-T_f\theta r}\right]\right). \tag{6}$$

Lemma 1. *The effective capacity of a certain node communicating in a Rayleigh block fading channel is given by*

$$EC(\rho_i, \theta, \epsilon) \approx -\frac{1}{T_f\theta}\ln\left[\epsilon + (1 - \epsilon)\,\mathcal{J}\right], \tag{7}$$

with

$$\mathcal{J} = \sum_{m=0}^{M} c^m \int_0^\infty (1 + \rho_i z)^d \frac{x^m}{m!}e^{-z}\,dz, \tag{8}$$

where $d = \frac{-\theta T_f}{\ln(2)}$. Also let $c = \theta\sqrt{T_f}Q^{-1}(\epsilon)\log_2 e$ and $x = \sqrt{\left(1 - \frac{1}{(1+\rho_i z)^2}\right)}$.

Proof. For Rayleigh envelope of pdf $f_{|h|^2}(z) = e^{-z}$, the EC expression in (6) can be written as

$$EC(\rho_i, \theta, \epsilon) = -\frac{1}{T_f\theta}\ln\left(\int_0^\infty \left(\epsilon + (1 - \epsilon)e^{-\theta T_f r}\right)e^{-z}\,dz\right). \tag{9}$$

From (4), we have

$$e^{-\theta T_f r} = e^{-\theta T_f \log_2(1+\rho_i z)}e^{\theta\sqrt{T_f(1-\frac{1}{(1+\rho_i z)^2})}Q^{-1}(\epsilon)\log_2 e}. \tag{10}$$

Elaborating, we attain

$$e^{-\theta T_f \log_2(1+\rho_i z)} = (1 + \rho_i z)^d, \tag{11}$$

$$e^{\theta\sqrt{T_f(1-\frac{1}{(1+\rho_i z)^2})}Q^{-1}(\epsilon)\log_2 e} = e^{cx}, \tag{12}$$

where $e^{cx} = \sum_{m=0}^\infty \frac{(cx)^m}{m!}$. It follows from (10), (11) and (12) that the expression in (9) can be written as a truncated sum of m terms leading to (7) and the approximation becomes equality as $M \to \infty$. $\qquad\square$

The infinite series in (8) can be truncated to a finite sum of terms and we evaluate the accuracy of the expression

noting that the accuracy increases with the number of terms. However, it is noticed that when examining for different network parameters (N, ρ, θ, T_f), the accuracy for expanding 1 term is 92.7%, 2 terms is 99% and 99.9% for 3 terms only. Henceforth, in our analysis, 3 terms will be enough. Thus (7) provides a closed-form approximation for the EC in Rayleigh block fading when $M = 2$. Moreover, the expectation in (6) is proved to be convex in ϵ [8]. Thus, it has a unique minimizer ϵ^*, which is consequently the EC maximizer. We define the optimum value of error probability ϵ^* which maximizes the EC in Rayleigh fading channels as

$$\epsilon^*(\rho_i, c, d) = \arg \min_{0 \leqslant \epsilon \leqslant 1} \ \epsilon + (1 - \epsilon) \, \mathcal{J}. \qquad (13)$$

To obtain the maximum effective capacity EC_{max}, we simply insert ϵ^* into (7).

IV. PERFORMANCE ANALYSIS

We elaborate the effect of interference by plotting the per-node EC obtained from Lemma 1 for 1, 5 and 10 nodes in Fig. 2. The network parameters are set as $T_f = 1000$, $\rho = 2$, and $\theta = 0.01$ and the channel is assumed to be Rayleigh. It is obvious that the per-node EC decreases when increasing N as more interference is added. Notice that the EC curves are concave in ϵ and hence, have a unique maximizer which is obtained from (13) and depicted in the figure. Another observation worth mentioning is that the optimum probability of error ϵ^* which maximizes the EC becomes higher when increasing the number of nodes. Notice that in Fig. 2, we assume $M = 2$ in (8) which renders an accurate approximation to (6).

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 [4]. For this model, imagine that a node needs to raise its EC in order to meet its QoS constraint. We study the interference alleviation scenarios for one node at a certain time slot, while other nodes also keep transmitting at the same time.

A. Power control

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

$$\rho_c = \rho \, (1 + \rho(N - 1)). \qquad (14)$$

When a certain node transmits with SNR of ρ_c, its EC is the same as in the case when transmitting with SNR equals to ρ while other nodes are silent. The method of power control is simple; however, it causes extra interference into other nodes due to the power increase of the recovering node. From (14), we define the SINR of other nodes colliding in the same network (nodes in set s) after the compensation of one node as

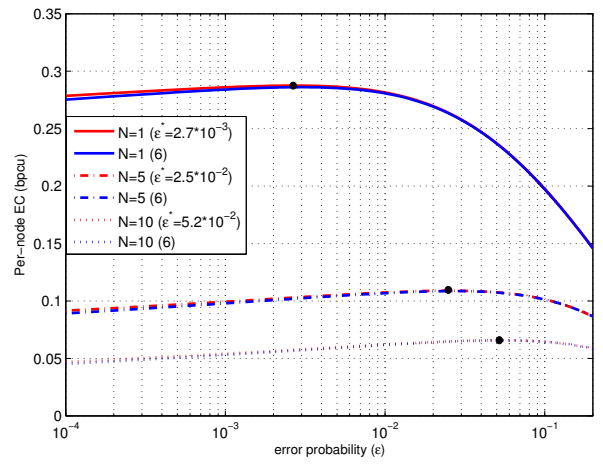$$\rho_s = \frac{\rho}{1 + \rho_c + \rho(N - 2)} = \frac{\rho}{1 + \rho \, (\rho + 1)(N - 1)} \qquad (15)$$

Fig. 2. EC as a function of error outage probability ϵ for different number of nodes, with $T_f = 1000, \rho = 2$, and $\theta = 0.01$.

To determine the effect of compensation of one node on the other nodes, we define the compensation loss factor α_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

$$\alpha_c = \frac{EC(\rho_s, \theta, \epsilon_s^*)}{EC(\rho_i, \theta, \epsilon_i^*)} \qquad (16)$$

where ϵ_s^* is the optimum error probability obtained from (13) when the SINR is set to ρ_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. Differentiating (6) with respect to ρ_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 ρ_i. Differentiating once more with respect to θ

$$\frac{\partial}{\partial \theta} \left(\frac{\partial EC}{\partial \rho_i} \right) = -\frac{\mathcal{K} T_f r e^{-T_f \theta r}}{(\epsilon + (1 - \epsilon)e^{-T_f \theta r})^2}, \qquad (17)$$

which is strictly negative and thus, validating our proposition. $\square$

Consider $T_f = 1000$ and $\rho = 1$, then Fig. 3 depicts the compensation loss factor α_c for different number of nodes N with $\theta = 0.1$ and 0.001. The figure shows that α_c is lower for smaller values of θ. Hence, the effect of compensation appears to be more severe for less stringent delay constraints. This follows from Proposition 1 where 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.
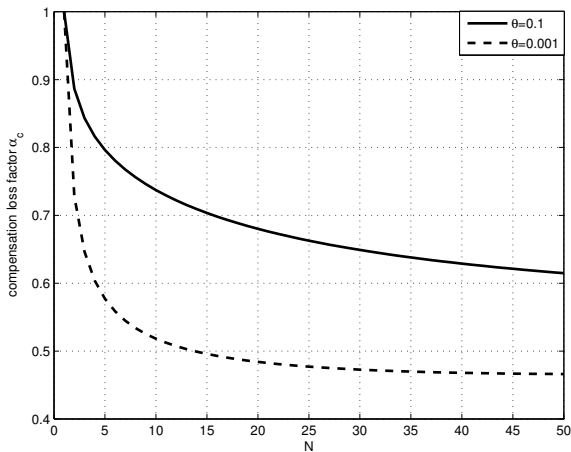
Fig. 3. Compensation loss factor α_c vs number of nodes N

B. Graceful degradation of the delay constraint

Here we determine how the delay exponent θ 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 θ be the original delay exponent and θ_i represent the new gracefully degraded one; θ_i is obtained by solving

$$EC(\rho, \theta, \epsilon^*) = EC(\rho_i, \theta_i, \epsilon_i^*) \qquad (18)$$

where ϵ_i^* is the maximizer of EC for the parameters (ρ_i, θ_i) and ϵ_i^* is the optimum error probability for (ρ_i, θ_i). The solution of (18) renders the necessary value of θ_i to compensate for the EC decrease due to collision in this case. In Fig. 4, we illustrate the graceful degradation of the delay constraint as a function of the target error outage probability. Consider (18) 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 $EC_{max} = 0.066$. For a delay outage probability of 10^{-3}, this corresponds to extending the allowable delay D_{max} from 3600 to 4600 symbol periods. We perform a limited delay extension ($\approx 25\%$) because the rise in EC partially compensates the graceful degradation of θ in (5). Note that the optimum error probabilities have different values in each case due to the change in SINR in (8).

C. 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 ρ_{s_o}, where ρ_{s_o} lies on the interval $[\rho_s \ \rho_i]$. Using (15), the operational SNR for the recovering node can be written as

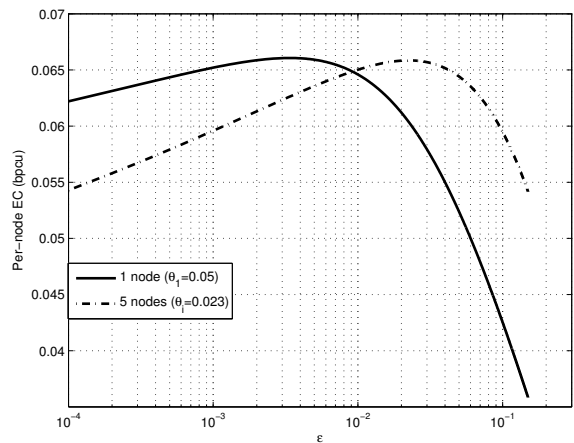$$\rho_{c_o} = \frac{\rho}{\rho_{s_o}} - 1 - \rho(N - 2), \qquad (19)$$

Fig. 4. Graceful degradation delay constraint θ in case of 5 nodes colliding where $T_f = 1000$ and $\rho = 1$.
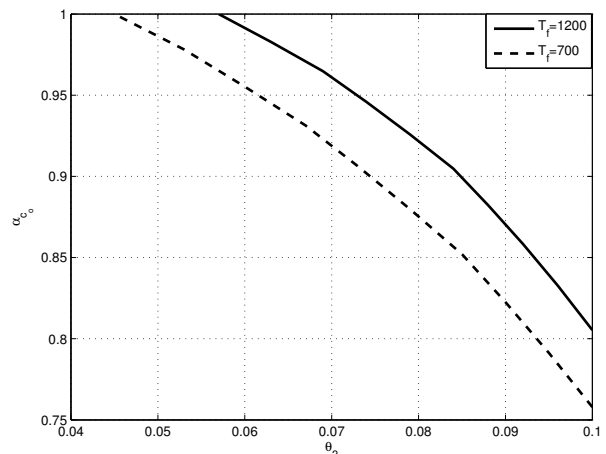

Fig. 5. Trade off between compensation loss factor via power control α_{c_o} and graceful degradation of delay constraint θ_2 for different operational points

and the operational point of the compensation loss factor α_{c_o} is

$$\alpha_{c_o} = \frac{EC(\rho_{s_o}, \theta, \epsilon_{s_o}^*)}{EC(\rho_i, \theta, \epsilon_i^*)} \qquad (20)$$

where $\epsilon_{s_o}^*$ is the optimum error probability obtained from (13) for the parameters (ρ_{s_o}, θ_1). α_{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 θ as in Section IV-B. To obtain θ_2, we solve

$$EC(\rho, \theta, \epsilon^*) = EC(\frac{\rho_{c_o}}{1 + \rho(N - 1)}, \theta_2, \epsilon_2^*) \qquad (21)$$

From (21), we compute the necessary value of θ_2 to continue the compensation process via graceful degradation of the delay constraint.

Fig. 5 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 $\alpha_{c_o} = 0.9, \theta_2 = 0.075$. This implies that a part of compensation will be performed via power control, which leads to 10% loss in EC of other nodes (set s). Then, the rest of compensation will be performed by
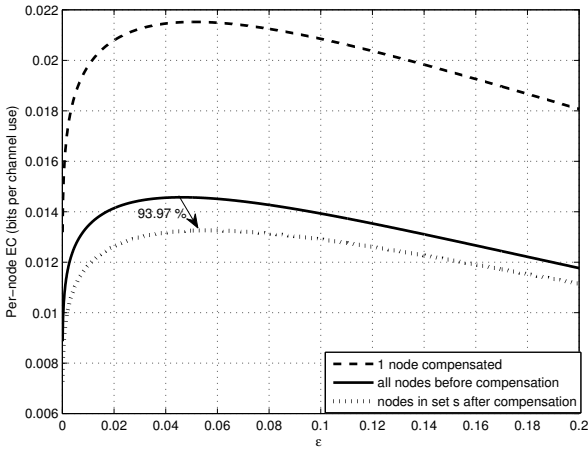
Fig. 6. Per-node EC as a function of error probability ϵ before and after joint compensation for $T_f = 1000, \theta_1 = 0.1, \rho = 2,$ and $N = 15$

gracefully degrading its θ from 0.1 to 0.075. The maximum delay of the recovering node remains 2500 symbol periods before and after recovery as restoring the EC compensates for the decrease in θ in (5). The figure also shows that for smaller packet sizes, the amount of losses due to compensation are higher.

Now, we propose an objective function leveraging the network performance for the joint model. First, we define the priority factor η_α 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 η_α, 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 θ graceful degradation. On the other hand, we define the priority factor η_θ as a measure of strictness of the delay constraint (i.e., the higher the value of η_θ, the more strict it is not to degrade delay constraint). Thus, we can formalize our objective function as the summation

$$\eta = \eta_\alpha \alpha_{c_o} + \eta_\theta \theta_2 \qquad (22)$$

where (α_{c_o}, θ_2) is the operational point. Now, we choose this operational point to satisfy

$$\eta_{max} = \max_{\theta_2 \geqslant 0} \ \eta_\alpha \alpha_{c_o} + \eta_\theta \theta_2 \\ s.t \ \ \rho_s \leqslant \rho_{s_o} \leqslant \rho_i \qquad (23)$$

where the solution to this problem gives the optimum operational point which can be found from (19), (20) and (21).

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 ρ_{s_o} will be 0.057. This value corresponds to the operational point $\alpha_{c_o} = 0.9397$ and $\theta_2 = 0.053$. The SNR of the recovering node becomes $\rho_{c_o} = 8.08$. Thus, 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. 6.

V. CONCLUSION

In this work, we presented a detailed analysis of the EC for delay constrained MTC networks in the finite blocklength regime. For Rayleigh block fading channels, we proposed an approximation for the EC and characterized the optimum error probability. 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 very limited extension in delay limit 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. As future work, we aim to analyze the impact of imperfect CSI on the EC and coordination algorithms that maximize EC with fairness constraints.

REFERENCES

[1] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 146–151.
[2] A. Orsino et al., "Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical MTC in 5G," *IEEE Communications Magazine*, February 2017.
[3] E. Dosti, U. L. Wijewardhana, H. Alves, and M. Latva-aho, "Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length," in *IEEE ICC 2017*, Paris, France, May 2017.
[4] R. Devassy, G. Durisi, P. Popovski, and E. G. Strom, "Finite-blocklength analysis of the ARQ-protocol throughput over the Gaussian collision channel," *ISCCSP 2014*, pp. 173–177.
[5] Nokia, "5G for Mission Critical Communication: Achieve ultra-reliability and virtual zero latency," *Nokia White Paper*, 2016.
[6] H. Chen et al., "Cooperative Retransmission for Massive MTC under Spatiotemporally Correlated Interference," in *2016 IEEE GLOBECOM*, Dec 2016, pp. 1–6.
[7] G. Durisi, T. Koch, and P. Popovski, "Toward Massive, Ultrareliable, and Low-Latency Wireless Communication with Short Packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.
[8] M. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," in *EURASIP Journal on Wireless Communucations and Networking 2013*, 2013.
[9] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-Fading Channels at Finite Blocklength," *ISWCS 2013*, vol. 9, pp. 410–413, 2013.
[10] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun*, vol. 2, no. 4, pp. 630–643, 2003.
[11] B. S. et al., "Fundamental tradeoffs among reliability, latency and throughput in cellular networks," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 1391–1396.
[12] L. Musavian and T. Le-Ngoc, "QoS-based power allocation for cognitive radios with AMC and ARQ in Nakagami-m fading Channels," in *Trans. Emerging Tel. Tech.*, vol. 27, 2014, pp. 266–277.
[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.
[14] Y. Hu, A. Schmeink, and J. Gross, "Blocklength-Limited Performance of Relaying under Quasi-Static Rayleigh Channels," *IEEE Transactions on Wireless Communications*, vol. 15, 2016.