

# Fixed Rate Statistical QoS Provisioning for Markovian Sources in Machine Type Communication

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

**Abstract**—In this paper, we study the trade-off between reliability and latency in machine type communication (MTC), which consists of single transmitter and receiver in the presence of Rayleigh fading channel. We assume that the transmitter does not know the channel conditions, therefore it would be transmitting information over a fixed rate. The fixed rate transmission is modeled as a two-state continuous-time Markov process, where the optimum transmission rate is obtained. Moreover, we conduct a performance analysis for different arrival traffic originated from MTC device via effective rate transmission. We consider that the arrival traffic is modeled as a Markovian process namely Discrete-Time Markov process, Fluid Markov process, and Markov Modulated Poisson process, under delay violation constraints. Using effective bandwidth and effective capacity theories, we evaluate the trade-off between reliability-latency and identify QoS (Quality of Service) requirement, and derive lower and upper bounds for the effective capacity subject to channel memory decay rate limits.

## I. INTRODUCTION

The next generation of communication systems will bring new business models and revolutionary changes in several industry chain processes due to ultra-reliable low latency communication (URLLC) and massive MTC (mMTC) use cases, enabling the industrial Internet of Things (IoT). As a result, there are foreseeable benefits in terms of efficiency, safety and sustainability for the industry and society. Therefore, it can be easily inferred that MTC is one of the core elements in the IoT revolution that will be required for achieving ambitious goals planned for future wireless networks [1].

Ultra-reliable low latency communication (URLLC) aims to provide communication services where stringent QoS requirements are imposed in terms of reliability and latency. This type of communication is normally used in mission critical applications such as vehicle-to-vehicle communication (V2V), remote surgery and critical links between industrial processes. On the other hand, mMTC refers to the kind of network services that can support large number of connected smart devices. It can be used to cover monitoring, automation and infrastructure of buildings, smart agriculture, logistics, tracking and fleet management. The devices are normally used to collect and forward information in both real as well as non-real time modes. These connected devices are expected to be greater than 50 billion by 2020 [2].

MTC (mMTC and URLLC) has attracted much interest in the recent years which promises a huge potential and

market growth for the technology. Furthermore, research and development suggest that cellular networks are suitable to largely satisfy the requirements of MTC in terms of global reach, QoS, scalability, security and diversity. A traffic model is a stochastic process that matches the behaviour of physical quantities of measured data traffic. Current cellular networks are based on the standard traffic model, which is designed and optimized for typical behaviour of human subscribers. Moreover, the traffic behaviour of MTC is different, since it is mostly uplink dominant; uses short as well as less number of packets; is (Non) real-time, periodic and event driven. It is usually bursty in nature (suddenly the volume of data flow increases in response to trigger of certain events) and coordinate in nature (i.e. simultaneous access attempts from many machines reacting to the same events), while HTC (Human Type Communication) is uncoordinated one. The QoS requirements of MTC are different in terms of security, reliability and latency [3].

Consequently, there is a prominent need for traffic models that can not only capture the behaviour of MTC traffic but can also provide the adequate communication services for both types of traffic with required QoS. Traffic models are mainly characterized into source and aggregated traffic models. The source traffic model captures the behaviour of individual users or sources. These models are often based on Poisson process which is an appropriate choice for capturing the properties of individual MTD (Machine Type Devices) generated traffic. On the other hand, the aggregate traffic model where a large number of MTD are assigned to one aggregator for capturing the traffic pattern of an individual user due to its homogenous and coordinated nature of the traffic [4].

The key design issue for traffic generated by MTD is how to achieve acceptable performance and efficiently use delay bound guarantees for timely transmission within the specified QoS requirements in a wireless channel. Wireless channel is highly time varying channel due to the random changes in the environment or obstacles. These changes in wireless signal usually lead to variations in the strength of the received signal, which directly affects the QoS guarantee in a MTC. Therefore, deterministic delay bound QoS constraints are usually difficult to satisfy. Alternatively, the statistical QoS provisioning is a powerful tool to characterize and implement delay bound QoS guarantees for wireless real time traffic [5].

In this context, effective capacity is a statistical QoS provisioning metric and can be defined as the maximum constant arrival rate that a given time varying service process can support while providing statistical latency guarantees [6]. It is derived from the large deviation theory and incorporates the statistical QoS constraint by capturing the decay rate of the buffer occupancy probability for the queue length. Effective capacity ensures the QoS guarantee on the constant arrival rates. Here, we are particularly interested to conduct a throughput analysis of random and bursty source traffic patterns by using Markovian source models including discrete time Markov, Markov fluid and Markov modulated Poisson sources with effective capacity.

Effective capacity [7] has extensively been used over the past few years to evaluate the trade-off among the reliability, latency, security and energy efficiency. For example, the authors [8] evaluate the trade-off between reliability and energy efficiency under QoS constraints. In [9], an optimal power allocation scheme is proposed to maximize the energy efficiency under given QoS constraint. In [10], the authors assess the performance of security in MTC networks using secure statistical QoS provisioning. In [11], the authors have considered fixed-rate transmission modelled as a two-state (ON/OFF) continuous-time Markov chain and utilize effective capacity to analyse energy efficiency.

In this paper, we conduct performance analysis of optimum fixed rate model. By using effective capacity theory, we aim at meeting the adequate reliability and latency requirement in point-to-point MTC network. We follow the recent contribution in [12] to incorporate Markovian arrival traffic so that its impact on the optimum transmission rate performance of the network can be evaluated. We derive upper and lower bounds of the effective capacity for high and low channel memory values. Our work is different from [12] as we assume a fixed rate transmission over Rayleigh fading channel which is modelled as continuous time Markov chain. This model allows us to identify the level of reliability and latency that each transmission possesses and therefore design the appropriate optimum transmission rate by formulating an optimization problem, that not only maximizes the effective capacity but also increases the allowed maximum average arrival rate at sources. Hence the satisfactory reliability and latency requirement are achieved under statistical QoS constraints which are satisfied by the system.

## II. PRELIMINARIES

### A. System Model

In this paper, we consider single transmitter and receiver in the presence of Rayleigh fading. The input output relation of the channel model can be expressed as

$$y(t) = h(t)x(t) + n(t), \quad (1)$$

where  $x(t)$  and  $y(t)$  are the complex valued input and output signals, respectively, and  $n(t)$  is zero mean, circularly-symmetric, complex Gaussian noise. Finally,  $h(t)$  is the Rayleigh fading coefficient which denotes the multiplicative

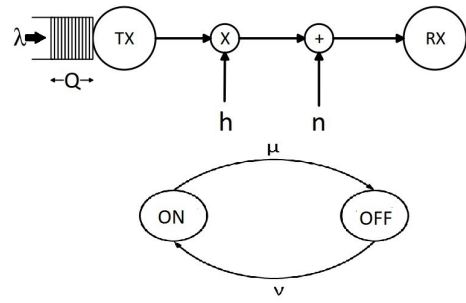


Fig. 1. System Model, where  $\lambda$  is the arrival rate at the source, and  $Q$  is the length of the queue at the transmitter. The channel coefficient is denoted as  $h$  and AWGN noise as  $n$ . Transmission in wireless link is also modeled as a continuous-time Markov chain with ON and OFF states, whose transition rate of fading states are measured by  $\nu$  and  $\mu$ .

fading component expressing the attenuation and phase shift experienced in the channel, and it is assumed to be a zero-mean complex Gaussian process. Therefore,  $z(t) = |h(t)|^2$  has an exponential distribution and assume unit variance.

In this paper, we assume that the receiver is able to estimate the channel coefficient  $h(t)$ , whereas the transmitter does not know this information. Therefore the transmitter would be transmitting information over a fixed rate  $\mathcal{R}$  bits/s. The wireless channel changes slowly and hence,  $h(t)$  stays constant over each coding block, the instantaneous channel capacity of Gaussian channel model is given by

$$C(t) = \log_2(1 + \gamma z(t)) \text{ bits/s}, \quad (2)$$

where  $\gamma$  is the average transmitted signal-to-noise ratio. When  $\mathcal{R} < C(t)$ , then the channel is considered to be in ON state. Hence, the transmitted message is decoded correctly and reliable communication is accomplished. While  $\mathcal{R} \geq C(t)$ , then the channel is considered to be in OFF state; outage arises and retransmission is needed.

The transition rate from ON to OFF state is denoted by  $\nu$ , and OFF to ON state is  $\mu$  as illustrated in Fig. 1. Then, we can write that the channel ON state probability is equivalent to  $Pr\{z(t) > \Psi\} = \int_{\Psi}^{\infty} e^{-z} dz = e^{-\Psi} = \frac{\nu}{\nu + \mu}$ , where  $\Psi = \frac{2^{\mathcal{R}} - 1}{\gamma}$ , and the OFF state probability is as  $Pr\{z(t) \leq \Psi\} = \int_0^{\Psi} e^{-z} dz = 1 - e^{-\Psi} = \frac{\mu}{\nu + \mu}$ . Hence, we have  $\nu = \kappa e^{-\Psi}$  and  $\mu = \kappa(1 - e^{-\Psi})$ . These equations are necessary to determine  $\nu$  and  $\mu$ . It is noted that the channel memory of these two-state Markov process exponentially decay at rate  $\nu + \mu = \kappa$  [11].

### B. Throughput of delay constrained networks

It is assumed that before transmission, data generated by random sources is stored in First in First out (FIFO) buffer as illustrated in Fig. 1. Thus, delay may occur in transmitting the data because of the long waiting time of data in the buffer. Moreover, the delay overflow probability is given by [6]

$$\Pr\{D \geq d\} \approx \zeta e^{-\theta a(\theta)d}, \quad (3)$$

where  $D$  is the queueing delay,  $d$  is the delay threshold,  $\theta$  is the delay QoS constraint,  $a(\theta)$  is the effective bandwidth and  $\zeta$  is the probability of non-empty buffer. Larger value of

$\theta \rightarrow \infty$  implies that stringent QoS constraint is imposed, while for loose QoS constraint the value  $\theta \rightarrow 0$  is small [13].

In this paper, we consider three types of Markov arrival sources namely Discrete-Time Markov source (DTMS), Fluid Markov source (FMS) and Markov Modulated Poisson source (MMPS). These Markov sources are using two-state ON and OFF model. In ON state, data arrives with rate  $\lambda$  bits/block and the OFF state refers to no arrival of data as illustrated in Fig. 1. For these sources, effective bandwidth provides a mean to characterize the minimum constant service rates required to support the random arrival of data into the buffer constrained to some statistical QoS requirements, namely buffer violation probability in (3). Let the time accumulated arrival process at instant  $t$  be  $A(t) = \sum_{k=1}^t a(k)$ . Then the effective bandwidth is defined as [6]

$$a(\theta) = \lim_{t \rightarrow \infty} \frac{1}{\theta t} \log \mathbb{E}\{e^{\theta A(t)}\}, \quad (4)$$

Effective capacity ( $C_E$ ) is the dual concept of effective bandwidth, where it defines the maximum constant arrival rate that a given time-varying service process can support in order to guarantee a statistical QoS requirement specified by the QoS exponent  $\theta$ . The effective capacity for a given QoS exponent is obtained from [14]

$$C_E(\gamma, \theta) = - \lim_{t \rightarrow \infty} \frac{1}{\theta t} \log \mathbb{E}\{e^{-\theta S[t]}\}, \quad (5)$$

where  $S[t] \triangleq \sum_{k=1}^t P[k]$  is the time accumulated service process and  $\{P[k], k = 1, 2, \dots\}$  shows the discrete time stationary and ergodic stochastic service process. So  $P[k] = \mathcal{R}$ , when it is ON state and otherwise 0 in OFF state.

We formulate a two-state continuous-time Markov chain fixed rate effective capacity model for a given statistical constraint, where  $\theta$  as

$$C_E(\gamma, \theta, \mathcal{R}) = \frac{1}{2\theta} [\theta \mathcal{R} + (\nu + \mu) - \xi], \quad (6)$$

where  $\xi = \sqrt{(\theta \mathcal{R} + (\nu + \mu))^2 - 4\nu\theta \mathcal{R}}$ , and (6) represents a simplified version of the effective capacity.

**Remark 1.** *The upper bound of effective capacity is expressed as below*

$$C_E = \lim_{\kappa \rightarrow \infty} \frac{1}{2\theta} [\theta \mathcal{R} + (\nu + \mu) - \xi] = \mathcal{R}e^{-\Psi}, \quad (7)$$

describes that  $C_E$  is asymptotic at the high channel memory decay rate towards  $\mathcal{R}e^{-\Psi}$ , which is free from QoS constraints  $\theta$ . Meanwhile, there is a lower bound for  $C_E$ , which approaches zero for very low channel memory decay rate as follows

$$C_E = \lim_{\kappa \rightarrow 0} \frac{1}{2\theta} [\theta \mathcal{R} + (\nu + \mu) - \xi] = 0. \quad (8)$$

It is fairly hard to investigate the performance of the communication system when the arrival of data and channel characteristics are random in nature with certain target guarantees. We assume the maximum average arrival rate is the throughput metric of Markovian arrival sources that can

support wireless fading channel and satisfy QoS guarantees illustrated in (3). Thus, QoS constraints are fulfilled when the effective bandwidth of the arrival process is equal to the effective capacity of service [6], therefore

$$a(\theta) = C_E(\gamma, \theta, \mathcal{R}), \quad (9)$$

Then we can find maximum arrival rate that can support fixed rate transmissions at given  $\gamma$  and  $\theta$ .

### III. MAXIMUM ARRIVAL RATE OF MARKOVIAN SOURCES

#### A. Discrete - Time Markov Sources

In the DTMS model, the arrival of data in the buffer is discrete in time. We assume a simple two-state ON/OFF model. In ON state  $\lambda$  bits arrive in the buffer, whereas no arrival of data in OFF state. Effective bandwidth in this case is defined by [12]

$$a(\theta) = \frac{1}{\theta} \log_e \left( \frac{1}{2} (p_{11} + p_{22} e^{\theta \lambda} + \sqrt{(p_{11} + p_{22} e^{\theta \lambda})^2 - 4(p_{11} + p_{22} - 1)e^{\theta \lambda}}) \right), \quad (10)$$

where  $p_{11}$  denotes the probability of staying in OFF state, while  $p_{22}$  determines the probability of ON state. The transition probabilities from one state to another are denoted by  $p_{21} = 1 - p_{22}$  and  $p_{12} = 1 - p_{11}$ .  $P_{\text{ON}}$  is the probability of ON state in the steady state regime, which is used to the calculate average arrival rate as

$$\lambda_{\text{avg}} = \lambda \cdot P_{\text{ON}} = \lambda \frac{1 - p_{11}}{2 - p_{11} - p_{22}}, \quad (11)$$

when the queue is in steady state then arrival rate is equal to the departure rate [12].

We substitute the effective bandwidth expression of discrete time Markov source (10) in (9) and solve as follows

$$(p_{11} + p_{22} e^{\lambda \theta} - 2e^{\theta C_E(\gamma, \theta)})^2 = (p_{11} + p_{22} e^{\lambda \theta})^2 - 4(p_{11} + p_{22} - 1)e^{\lambda \theta}. \quad (12)$$

After solving (12) for  $\lambda$ , we obtain maximum ON state arrival rate as

$$\lambda^*(\theta) = \frac{1}{\theta} \log_e \left( \frac{e^{2\theta C_E(\gamma, \theta)} - p_{11} e^{\theta C_E(\gamma, \theta)}}{(1 - p_{11} - p_{22}) + p_{22} e^{\theta C_E(\gamma, \theta)}} \right). \quad (13)$$

By incorporating (11) we can express the maximum average arrival rate as a function of QoS exponent, effective capacity fading channel and state transition probabilities as [12]

$$\lambda_{\text{avg}}^*(\gamma, \theta) = \frac{P_{\text{ON}}}{\theta} \log_e \left( \frac{e^{2\theta C_E(\gamma, \theta)} - p_{11} e^{\theta C_E(\gamma, \theta)}}{1 - p_{11} - p_{22} + p_{22} e^{\theta C_E(\gamma, \theta)}} \right). \quad (14)$$

Let us further simplify our analysis by assuming  $p_{11} = 1 - s$  and  $p_{22} = s$ , we obtained a simplified version of the source model in which single parameter  $s$  measure the burstiness of the source [6].

$$\lambda_{\text{avg}}^*(\gamma, \theta) = \frac{s}{\theta} \log \left( \frac{e^{\theta C_E(\theta)} - (1 - s)}{s} \right). \quad (15)$$

## B. Markov Fluid Sources

In this model, the data continuously arrive in the buffer and the effective bandwidth is expressed by [6]

$$a(\theta) = \frac{1}{2\theta} \left[ \theta\lambda - (\alpha + \beta) + \sqrt{(\theta\lambda - (\alpha + \beta))^2 + 4\alpha\theta\lambda} \right], \quad (16)$$

where  $\alpha$  shows the transition rate from OFF state to ON state and  $\beta$  is the transition rate from ON state to OFF state. Then, we attain the steady state probability of being ON as

$$P_{\text{ON}} = \frac{\alpha}{\alpha + \beta}. \quad (17)$$

We follow a similar procedure as for the DTMS to determine the maximum average arrival rate of two-state ON/OFF model for the MFS as [12]

$$\lambda_{\text{avg}}^*(\gamma, \theta) = P_{\text{ON}} \frac{\theta C_E(\gamma, \theta) + \alpha + \beta}{\theta C_E(\gamma, \theta) + \alpha} C_E(\gamma, \theta). \quad (18)$$

## C. Markov Modulated Poisson Sources

In Markov modulated Poisson process, the arrival of data in the buffer is a Poisson process, whose intensity is controlled by a continuous time Markov chain. In OFF state, there is no arrival of data which means arrival intensity is zero. On the other hand,  $\lambda$  is the arrival intensity in ON state. In this case the effective bandwidth is defined as [6]

$$a(\theta) = \frac{1}{2\theta} \left[ (e^\theta - 1)\lambda - (\alpha + \beta) \right] + \frac{1}{2\theta} \sqrt{((e^\theta - 1)\lambda - (\alpha + \beta))^2 + 4\alpha(e^\theta - 1)\lambda}. \quad (19)$$

Similar to the previous sources models, we determine the maximum average arrival rate for the MMPS source model as [12]

$$\lambda_{\text{avg}}^*(\gamma, \theta) = P_{\text{ON}} \frac{\theta[C_E(\gamma, \theta) + \alpha + \beta]}{(e^\theta - 1)\theta C_E(\gamma, \theta) + \alpha} C_E(\gamma, \theta). \quad (20)$$

## IV. OPTIMAL FIXED RATE TRANSMISSION

Fig. 2 shows the maximum average arrival rate of Markovian sources as a function of transmission rate  $\mathcal{R}$  for fixed values of QoS constraint  $\theta$ ,  $\kappa$ ,  $\gamma$  and  $P_{\text{ON}}$ . It is clearly visible that the maximum average arrival rate is maximized as transmission rate increases, but after certain limit the effective capacity gradually starts to decrease due to fixed value of  $\gamma$ . If one transmits with higher rate, it will degrade the overall throughput. Therefore, efficient use of transmission rate boosts the performance of communication system. Furthermore, it is noticed that optimum transmission rate which maximizes the  $\lambda_{\text{avg}}$  has similar value for all markovian models. We conclude that optimizing the effective capacity with respect to the transmission rate, allows for high link throughput while allowing larger arrival rates as we shall see next.

The optimization problem to maximize  $C_E$  subject to a non-negative transmission rate

$$C_E^*(\gamma, \theta, \mathcal{R}) = \max_{\mathcal{R} \geq 0} \left\{ \frac{1}{2\theta} [\theta\mathcal{R} + (\nu + \mu) - \xi] \right\} \quad (21)$$

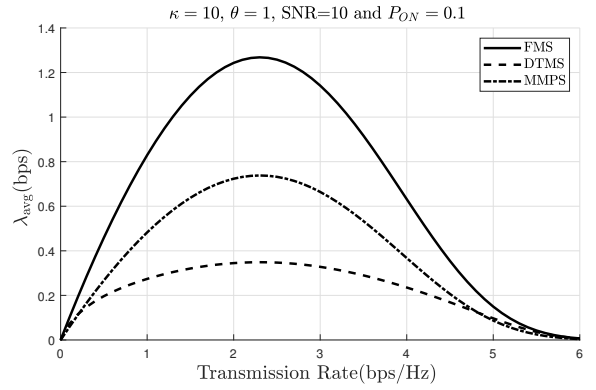


Fig. 2. Maximum average arrival rate of Markovian sources as a function of transmission rate  $\mathcal{R}$  for fixed values of QoS constraint  $\theta$ ,  $\kappa$ ,  $\gamma$  and  $P_{\text{ON}}$ .

Now, taking first derivative of (21) and equating it to zero yields the optimum value of  $\mathcal{R}$ .

$$\frac{\partial C_E^*(\gamma, \theta, \mathcal{R})}{\partial \mathcal{R}} = \frac{\theta - \frac{2\theta(\mathcal{R}\theta + \kappa) - 4\theta\nu + \frac{2^{2+\mathcal{R}}\mathcal{R}\log(2)\nu}{\gamma}}{\xi}}{2\theta} = 0. \quad (22)$$

After some algebraic manipulation, we have

$$\gamma \xi - \gamma(\mathcal{R}\theta + \kappa) + 2\gamma\nu - 2\mathcal{R}\nu \log(2) 2^{\mathcal{R}} = 0 \quad (23)$$

Further, we simplify (23) to reach

$$\mathcal{R}^* = \frac{\gamma(\xi - \kappa + 2\nu)}{\gamma\theta + 2\nu \log(2) 2^{\mathcal{R}^*}}. \quad (24)$$

However, closed form solution for (24) does not exist. In order to obtain optimum transmission rate  $\mathcal{R}^*$ , we use fzero Matlab root function.

## V. RESULTS AND DISCUSSION

Herein, Fig. 3 shows the effective capacity as a function of channel memory  $\kappa$  for different value of  $\gamma$  and  $\theta$ , when  $\mathcal{R} = 3$  bps. High values of  $\kappa$  measure the channel memory decays fast which corresponds to good channel state and bears higher transmission rate but has less probability of this event occurred. We observe that the effective capacity is maximized at high level of channel memory rate for low  $\gamma$  curve as compared to high  $\gamma$  when stringent QoS constraint is imposed and then it saturates. It is shown in Remark 1 that channels with high channel memory  $\kappa$  produce an asymptotic boost in effective capacity up towards an upper bound, which is physical phenomena that can measure with  $\kappa$  but do not control. We conclude that, larger channel memory  $\kappa$  or  $\gamma$  above a certain level will not increase effective capacity.

In a wireless communication system, it is generally presumed that by increasing signal power, the performance of the system improves. Fig. 4 illustrates the maximum average arrival rate  $\lambda_{\text{avg}}^*$  of Markovian models as a function of  $\gamma$  for different values of  $P_{\text{ON}}$  when we set QoS exponent  $\theta = 1$  and  $\kappa = 50$ . It is clear that as the  $\gamma$  increase throughput also improves. The increase in burstiness of sources (i.e., when  $P_{\text{ON}}$  decreases) implies that the data arrives less frequently which

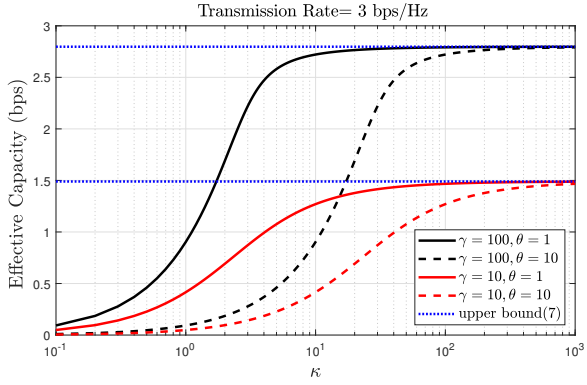


Fig. 3. Effective capacity as a function of  $\kappa$  for different values of  $\gamma$  (SNR) and  $\theta$  when  $\mathcal{R}=3$ .

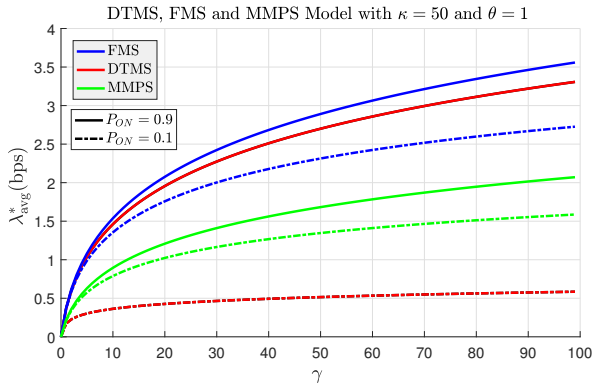


Fig. 4. Maximum average arrival rate  $\lambda_{\text{avg}}^*$  as a function of  $\gamma$  (SNR) for distinct Markov source models with different value of  $P_{\text{ON}}$ .

means that the source is bursty. Moreover, the arrival rate in ON state becomes higher.

The low value of  $\theta$  is examined in Fig. 5, we plot the maximum average arrival rate  $\lambda_{\text{avg}}^*$  as function of QoS constraint  $\theta$  for distinct Markov source models with different value of source characteristic  $P_{\text{ON}}$  when we set  $\gamma = 10$  and  $\kappa = 50$ . It is noted that higher value of  $\lambda_{\text{avg}}$  can be sustained for the loose QoS constraints. Hence, throughput does not depend on channel memory and source characteristics. When the QoS constraints becomes too stringent, the throughput tends to zero, since assigned arrival rate must satisfy short delay requirement. This reduction in  $\lambda_{\text{avg}}^*$  is more severe in bursty sources (i.e., when  $P_{\text{ON}} = 0.1$ ). Moreover, maximum average arrival curve of MMPS sources is generally smaller and declines fast, when stringent QoS constraint are imposed. This can be the natural attributed to the much more randomness/burstiness tolerated at high  $\gamma$  regime for the DTMS and FMS. Furthermore, MMPS throughput is less affected by burstiness because MMPS sources exhibit a higher level of variation in this sense and can be regarded as a more bursty source.

Fig. 6 includes subplots for the outage that occurs due to delay violation probability as a function of  $\gamma$ , QoS exponent  $\theta$  and source characteristic  $P_{\text{ON}}$ . It is obvious that the reliability and latency tradeoff is described in terms of delay violation

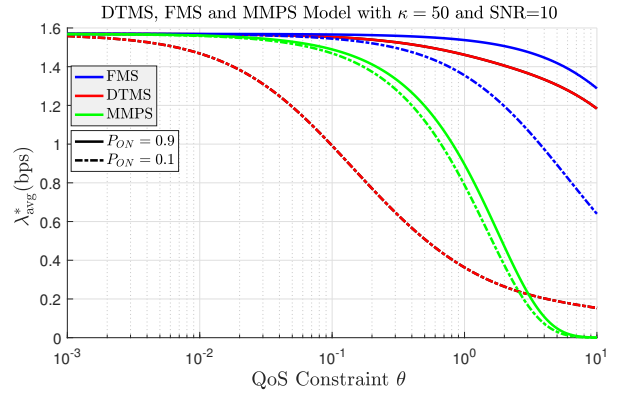


Fig. 5. Maximum average arrival rate  $\lambda_{\text{avg}}^*$  as a function of  $\theta$  for distinct Markov source models with different value of  $P_{\text{ON}}$ .

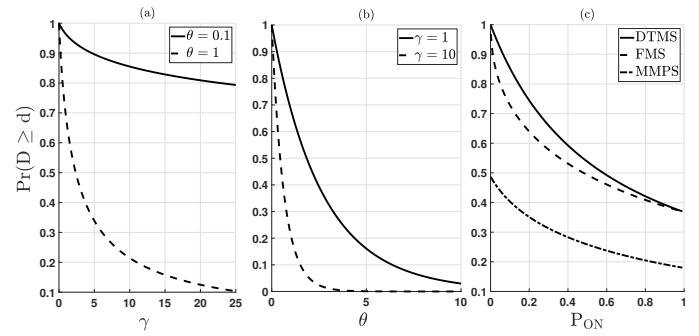


Fig. 6. Delay violation probability as function of (a)  $\gamma$  assuming fixed  $\theta$ , (b)  $\theta$  assuming fixed  $\gamma$ (SNR) and (c)  $P_{\text{ON}}$ .

probability. Enhancing reliability requires large values of  $\gamma$ , tight QoS requirements and sources should behave less bursty. The relation between reducing latency and enhancing reliability is a trade-off subject to the design requirements. In Fig. 6(a), as  $\gamma$  increases, good channel quality is available for optimal transmitting rate, which tends to boost reliability. Fig. 6(b) shows that as QoS requirement increases, the waiting time of data in buffer shorten which uplifts reliability level. In Fig. 6(c), delay violation probability is plotted as a function of  $P_{\text{ON}}$  for fixed arrival rate of 1 bps and fixed QoS requirements  $\theta = 0.01$ . Burstiness is measured from the average arrival of data in ON state. It is observed that more bursty sources degrade the maximum average arrival rate, which directly rises delay violation probability. Furthermore, the MMPS sources tolerate lower delay violation probability. Hence, the reliability of MMPS is greater as compared to DTMS and FMS in the presence of random/bursty sources.

In Fig. 7, we investigate the impact of optimum effective capacity  $C_E^*$  on the random/bursty sources under QoS Constraints. It is clearly observed that bursty sources (i.e., less  $P_{\text{ON}}$ ) have higher arrival rate in ON state because of maintaining average arrival rate nondecreasing function as defined (11). However, with the same departure rate, it is challenging to keep throughput non-decreasing when QoS constraints are imposed. Therefore, the higher arrival rate is required to achieve  $C_E^*$  when the source becomes bursty

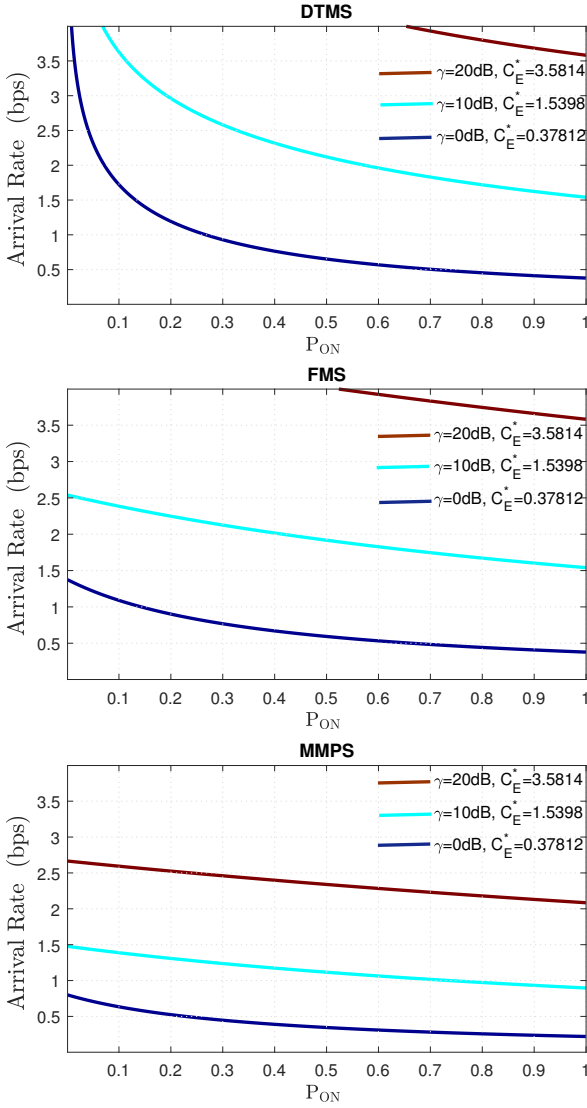


Fig. 7. Arrival rate vs  $P_{ON}$  for different  $C_E^*$  of Markovian sources.

(i.e.  $P_{ON} < 1$ ). It is also observed that due to more bursty sources MMPS has a small rise in arrival rate required to support given  $C_E^*$  as compared to DTMS and FMS. Thus MMPS is more stable and convenient for modelling bursty or random source traffic.

## VI. CONCLUSIONS

In this work, we conducted the detailed performance analysis of effective transmission rate model to achieve adequate reliability and latency trade-off in single point-to-point machine type devices. Effective transmission rate model is based on effective capacity for continuous time markov chain model. We characterized the upper and lower bounds of effective capacity as a function of channel memory decay rate and defined the optimum fixed transmission rate which maximizes the throughput for Markovian source models. The results showed that source, buffer, channel memory and transition rates have major impact on the system performance when certain reliability and latency constraints are imposed. Furthermore,

increased source burstiness and stringent QoS requirement all need an increase in SNR and decay rate of channel memory to fulfill the requirement of reliable communication. Moreover, it is observed that MMPS model is more stable for more bursty sources when compared to FMS and DTMS in the presence of QoS constraints.

## ACKNOWLEDGMENTS

This work is partially supported by Academy of Finland 6Genesis Flagship (Grant no. 318927), Aka Project EE-IoT (Grant no. 319008), and by Finnish Funding Agency for Technology and Innovation (Tekes), Bittium Wireless, Keysight Technologies Finland, Kyynele, MediaTek Wireless, and Nokia Solutions and Networks.

## REFERENCES

- [1] P. Popovski, "Ultra-reliable Communication in 5G Wireless Systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, 2014.
- [2] Y. Morioka, "LTE for Mobile Consumer Devices," *ETSI Workshops on Machine to Machine standardization*, 2011.
- [3] E. Soltanmohammadi et al., "A Survey of Traffic Issues in Machine-to-Machine Communications Over LTE," *IEEE Internet of Things Journal*, Dec 2016.
- [4] E. Grigoreva et al., "Coupled Markovian Arrival Process for Automotive Machine Type Communication Traffic Modeling," in *2017 IEEE (ICC)*, May 2017.
- [5] W. Cheng, X. Zhang, and H. Zhang, "Heterogeneous statistical QoS provisioning for downlink transmissions over mobile wireless cellular networks," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 4622–4628.
- [6] M. Ozmen and M. C. Gursoy, "Wireless Throughput and Energy Efficiency with Random Arrivals and Statistical Queuing Constraints," *IEEE Transactions on Information Theory*, March 2016.
- [7] M. Shehab, E. Dosti, H. Alves, and M. Latva-aho, "Statistical QoS provisioning for MTC networks under finite blocklength," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 194, Aug 2018.
- [8] L. Musavian and Q. Ni, "Effective capacity maximization with statistical delay and effective energy requirements," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3824–3835, July 2015.
- [9] J. Tang et al., "Quality-of-Service Driven Power and Rate Adaptation over Wireless Links," *IEEE Transactions on Wireless Communications*, August 2007.
- [10] H. Alves, P. H. J. Nardelli, and C. H. M. de Lima, "Secure Statistical QoS Provisioning for Machine-Type Wireless Communication Networks," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018, pp. 1–5.
- [11] M. Ozmen and M. C. Gursoy, "Energy Efficiency of Fixed-Rate Transmissions with Markov Arrivals under Queuing Constraints," *IEEE Communications Letters*, April 2014.
- [12] F. Qasmi, M. Shehab, H. Alves, and M. Latva-Aho, "Optimum Transmission Rate in Fading Channels with Markovian Sources and QoS Constraints," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2018, pp. 1–5.
- [13] M. Shehab, E. Dosti, H. Alves, and M. Latva-aho, "On the effective energy efficiency of ultra-reliable networks in the finite blocklength regime," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2017, pp. 275–280.
- [14] Dapeng Wu and R. Negi, "Effective Capacity: a Wireless Link Model for Support of Quality of Service," *IEEE Transactions on Wireless Communications*, July 2003.