

# QoS Provisioning in Large Wireless Networks

Marios Kountouris\*, Nikolaos Pappas†, and Apostolos Avranas\*

\*Mathematical and Algorithmic Sciences Lab, Paris Research Center, Huawei France

†Department of Science and Technology, Linköping University, Sweden

Email: {marios.kountouris,apostolos.avranas}@huawei.com, nikolaos.pappas@liu.se

**Abstract**—Quality of service (QoS) provisioning in next-generation mobile communications systems entails a deep understanding of the delay performance. The delay in wireless networks is strongly affected by the traffic arrival process and the service process, which in turn depends on the medium access protocol and the signal-to-interference-plus-noise ratio (SINR) distribution. In this work, we characterize the conditional distribution of the service process given the point process in Poisson bipolar networks. We then provide an upper bound on the delay violation probability combining tools from stochastic network calculus and stochastic geometry. Furthermore, we analyze the delay performance under statistical queueing constraints using the effective capacity formulation. The impact of QoS requirements, network geometry and link distance on the delay performance is identified. Our results provide useful insights for guaranteeing stringent delay requirements in large wireless networks.

**Index Terms**—Poisson bipolar networks, stochastic geometry, stochastic network calculus, effective capacity, delay, QoS.

## I. INTRODUCTION

Data traffic has been growing tremendously over the last decade, fueled by the ubiquity of smart devices and bandwidth-demanding applications. Current wireless networks are confronted with an avalanche of heterogeneous traffic with diverse requirements in terms of rate, reliability and latency. Emerging mobile communication systems will not only be designed to provide enhanced spectral efficiency and coverage, but they should also meet the delay requirements of new delay-sensitive applications, such as industrial control, automated driving and healthcare. Different applications are expected to have very diverse QoS requirements in terms of throughput and delay. QoS provisioning is instrumental for next-generation low latency networks; yet it is particularly challenging mainly due to time-varying wireless channels and spatio-temporal randomness in traffic arrivals and interferers locations. Ensuring deterministic (hard) QoS guarantees would most likely result in extremely conservative performance. As a result, providing statistical (soft) QoS guarantees, in terms of effective bandwidth/capacity and bounds in queue length and delay violation probability, stands as a powerful approach to characterize delay QoS provisioning in wireless networks [1]–[4].

There have been several attempts at quantifying delay in wireless networks, and queueing theory has been instrumental in providing exact backlog and delay characterization. However, queueing analysis is largely restricted to networks with a single or few interacting queues with random arrivals and typically provides the average delay rather than the worst-case delay, which is of cardinal importance in mission-critical

applications. The delay in wireless networks is strongly affected by the queueing process and the service process of the packets. The latter is mainly governed by the access protocol and the link quality, i.e., the received SINR. The SINR in turn critically depends on the link distance and the network geometry on which the path loss and the fading characteristics are dependent upon. Although the locations of nodes in wireless networks are traditionally modeled by regular grids, the spatial node distribution in emerging networks (e.g. heterogeneous cellular networks) is irregular. Stochastic geometry and point process theory have recently proved to be powerful mathematical tools for analyzing and designing large wireless networks with spatial randomness. These approaches have focused on metrics such as coverage probability and spatial average rate, and calculate spatial averages by considering a snapshot of the network. Analyzing the delay in large spatial networks is very challenging due to the correlation of interference [5]–[7] and the large number of interacting queues. The local delay, i.e., the random number of transmission attempts until a packet is successfully transmitted to its target receiver, is proposed in [8] and extended in [9]. Local delay assumes fully backlogged nodes, thus it captures the transmission delay, but not the queueing delay. Throughput maximization subject to delay constraints in two-tier spatial networks is studied in [10].

In this paper, we investigate the delay performance of large wireless networks in the presence of statistical QoS constraints, which are imposed as limits on the delay violation and buffer overflow probabilities. We start by characterizing the distribution of the service process in Poisson bipolar networks. The results are exploited in order to derive a bound on the delay violation probability using tools from stochastic network calculus [2], [11], [12] and the effective capacity. For relevant results, we perform space-time scale separation and consider static or low-mobility spatial networks. This results in temporal interference correlation and the conditional rate given the point process will vary from node to node. Thus, both delay violation probability and effective capacity are random variables whose statistics ought to be found. Our analytical results show the effect of inter-node distances, spatial randomness, and QoS constraints on the delay violation and effective capacity performance of large spatial networks.

## II. SYSTEM MODEL

### A. Network Model

We consider a communication network in which the locations of the (potential) transmitters are modeled as a homoge-

neous Poisson point process (PPP)  $\Phi = \{x_i\} \subset \mathbb{R}^2$  of intensity  $\lambda$ . Each transmitter has an associated receiver at fixed distance  $r$  in a random direction (denoted by  $R_x$  for a transmitter  $x$ ). This model is commonly referred to as Poisson bipolar network.

The small scale fading between two nodes is independent and identically distributed (i.i.d.) across time and space (unless otherwise stated) and is exponentially distributed (Rayleigh fading). The transmit power at all nodes is fixed to 1. The large-scale path loss function is denoted by  $\ell(x) : \mathbb{R}^2 \rightarrow [0, \infty]$  and is assumed to be a non-increasing function of  $\|x\|$  and  $\int_{B(o,d)} \ell(x) dx < \infty, \forall d$ , where  $B(o,d)$  is the ball of radius  $d$  centered around the origin  $o = (0, 0)$ . In this paper, we focus on a non-bounded model  $\ell(x) = \|x\|^{-\alpha}$ ,  $\alpha > 2$ . Time is divided into discrete slots of equal duration and transmission attempts are synchronized. We focus on the interference-limited case, but our analysis can be easily extended including background noise.

The received signal-to-interference ratio (SIR) in time slot  $t$  is given by

$$\text{SIR}_{R_x,t} = \frac{h_{xR_x}^t \ell(r)}{\sum_{y \in \Phi_t \setminus \{x\}} h_{yR_x}^t \ell(y - R_x)} \quad (1)$$

where  $h_{xy}^t$  is the small-scale fading coefficient between nodes  $x$  and  $y$  in time slot  $t$ , and  $\Phi_t \subset \Phi$  is the set of active interferers in time slot  $t$ . The interference at time slot  $t$  can be alternatively written as

$$I_{R_x,t} = \sum_{y \in \Phi_t \setminus \{x\}} h_{yR_x}^t \ell(y - R_x) \mathbb{1}(y \in \Phi_t) \quad (2)$$

where  $\mathbb{1}(\cdot)$  is the indicator function.

The success probability of a typical link is given by

$$\mathbb{P}^{1o}(\text{SIR}_{R_o,t} > \xi) = \lim_{\delta \rightarrow \infty} \frac{\sum_{x \in \Phi \cap B(o,\delta)} \mathbb{P}(\text{SIR}_{R_x,t} > \xi \mid \Phi)}{\lambda \pi \delta^2}. \quad (3)$$

We focus on the typical link with a transmitter located at the origin and we drop the time and node subindexes whenever evident.

We denote  $P_s(\xi) \triangleq \mathbb{P}(\text{SIR} > \xi \mid \Phi)$  the success probability given the point process (i.e., conditioned on the location of interferers) and that the transmitter of interest is active, which is taken over the fading and the random access scheme. The conditional probability  $P_s(\xi)$  can be interpreted as the mark of a virtual typical link placed at the origin, whose distribution is given by  $\mathbb{P}^{1o}(P_s(\xi) > x)$ ,  $x \in [0, 1]$ , where  $\mathbb{P}^{1o}$  is the reduced Palm probability (for a PPP we have  $\mathbb{P}^{1o} = \mathbb{P}$ ) [13].

## B. Traffic Model

We consider a system-theoretic stochastic model as in [14], which involves a queueing system with stochastic arrival and departure processes described by bivariate stochastic processes  $A(\tau, t)$  and  $D(\tau, t)$ , respectively. A fluid-flow traffic model is adopted and the system starts with empty queues at  $t = 0$ . The number of bits arriving at the queue at a discrete time

instant  $i$  is modeled by the arrival process  $a_i$ . For successful transmissions, the service process is equal to the instantaneous capacity. In case of transmission errors, the service is considered to be zero as no data is removed from the queue. The departure process  $d_i$  describes the number of bits that arrive successfully at the destination and depends on both the service process and the number of bits waiting in the queue. Acknowledgments and feedback messages are assumed to be instantaneous and error-free. To avoid data loss, data is stored in a buffer or queue, in which it will remain for a random time until the receiver indicates that data was successfully decoded. At each time slot, node  $j \in \Phi$  independently transmits with probability  $p_j$ . The steady-state probability  $p_j$ ,  $\forall j \in \Phi$  depends on the arrival process of  $j$ -th node, the transmit probability of the other nodes  $p_i$ ,  $\forall i \in \Phi, i \neq j$ , and the channel of all nodes. A node remains idle when there is no traffic arrival and the queue is empty due to the early arrival and late departure assumption. The aforementioned two events can be assumed independent under non-saturated or light traffic conditions. Unless otherwise stated, we set  $p_i = p$ ,  $\forall i \in \Phi \setminus \{o\}$ .

The cumulative arrival and departure processes for any  $0 \leq \tau \leq t$ , measured in bits of the flow during time interval  $[\tau, t)$ , are defined as

$$A(\tau, t) = \sum_{i=\tau}^{t-1} a_i, \quad \text{and} \quad D(\tau, t) = \sum_{i=\tau}^{t-1} d_i. \quad (4)$$

For lossless first-in first-out (FIFO) queueing systems, the delay  $W(t)$  at time  $t$ , i.e., the number of slots it takes for an information bit arriving at time  $t$  to be received at the destination, is defined as

$$W(t) = \inf\{u > 0 : A(0, t)/D(0, t+u) \leq 1\}. \quad (5)$$

and the delay violation probability is given by

$$\Lambda(w, t) = \sup_{t \geq 0} \mathbb{P}[W(t) > w].$$

## III. SERVICE PROCESS CHARACTERIZATION

The instantaneous rate or capacity  $C_t$  of the channel at time  $t$  can be expressed as a function of the instantaneous SNR or SIR at this time. Assuming flat-fading, Gaussian codebooks and ideal link adaptation, the instantaneous rate can be expressed as

$$C_t = N \log(1 + \text{SIR}_t) \quad (\text{nats/s}) \quad (6)$$

where  $N$  is the number of transmitted symbols per time slot. The symbol rate is usually related to the bandwidth  $W$  as  $N = 2W$  (Shannon-Hartley theorem). In the remainder, we assume  $N = 1$  to simplify notation and we reincorporate this parameter into the equations in the numerical results.

The service process (or cumulative capacity) through period  $(\tau, t]$  is defined as

$$S(\tau, t) \triangleq \sum_{i=\tau}^{t-1} C_i, \quad (7)$$

and is a random variable with cumulative distribution function (cdf)  $F_{S(\tau,t)}(x) = \mathbb{P}(S(\tau,t) \leq x)$ ,  $x > 0$ .

If  $C_i$  and  $C_j$ ,  $i \neq j$ , are independent, then  $f_{S(\tau,t)} = f_{C_{\tau+1}} * \dots * f_{C_t}$ , where  $*$  denotes the convolution operation, i.e.,  $(f * g)(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)dy$ . Hence,

$$F_{S(\tau,t)}(x) = \int_{-\infty}^x f_{S(\tau,t)}(y)dy. \quad (8)$$

An upper bound can be derived using Young's inequality, which states  $\|f * g\|_r \leq \|f\|_p \|g\|_q$  for  $1/p + 1/q = 1/r + 1$ ,  $f \in L^p(\mathbb{R}^d)$  and  $g \in L^q(\mathbb{R}^d)$ . When all marginal distributions are identical ( $F_{C_i} \sim F_C, \forall i$ ), the probability density function (pdf) of the service process is given by the  $n$ -fold convolution with  $n = t - \tau$ , i.e.,  $f_{S(\tau,t)} = f_C^{*n}$ .

When the length of the period  $t - \tau$  is large,  $F_{S(\tau,t)}(x)$  converges to a normal distribution (Central Limit Theorem)

$$F_{S(\tau,t)}(x) \approx Q\left(\frac{x - \mathbb{E}[S(\tau,t)]}{\sigma^2[S(\tau,t)]}\right), \quad (9)$$

with mean  $\mathbb{E}[S(\tau,t)] = \sum_{i=\tau+1}^t \mathbb{E}[C_i]$  and variance  $\sigma^2[S(\tau,t)] = \sum_{i=\tau+1}^t \sigma^2[C_i]$ , where  $Q(x) \triangleq \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ .

If  $C_i$  and  $C_j$ ,  $i \neq j$ , are not independent, the exact calculation of the service rate distribution seems to be highly involved. For that, we resort to the Fréchet-Hoeffding bounds on copulas [15], [16], which give that cdf of the cumulative capacity satisfies

$$F_{S(\tau,t)}^l(z) \leq F_{S(\tau,t)}(z) \leq F_{S(\tau,t)}^u(z) \quad (10)$$

where

$$F_{S(\tau,t)}^u(z) \triangleq \inf_{\sum_{i=\tau+1}^t z_i = z} \left[ \sum_{i=\tau+1}^t F_{C_i}(z_i) \right]_1,$$

$$F_{S(\tau,t)}^l(z) \triangleq \sup_{\sum_{i=\tau+1}^t z_i = z} \left[ \sum_{i=\tau+1}^t F_{C_i}(z_i) - (t - \tau - 1) \right]^+.$$

where  $[f]^+ = \max(f, 0)$ ,  $[f]_1 = \min(f, 1)$ , and  $F_{C_i}(z) = \mathbb{P}(\text{SIR}_i \leq e^z - 1) = F_{\text{SIR}_i}(e^z - 1) = 1 - P_s(e^z - 1)$ ,  $z > 0$ .

### A. Moment Generating Function

The service rate distribution can also be characterized via the moment generating function (MGF)  $M_{S(\tau,t)}(\theta)$ ,  $\theta \in \mathbb{R}$ , which is given by

$$M_{S(\tau,t)}(\theta) \triangleq \mathbb{E}\left[e^{\theta S(\tau,t)}\right] = \int_{-\infty}^{\infty} e^{\theta z} dF_{S(\tau,t)}(z).$$

For the independent case, we have  $\overline{M}_{S(\tau,t)}^{\text{ind}}(\theta) = \prod_{i=\tau}^{t-1} \overline{M}_{C_i}(\theta)$  and for the i.i.d. case, we have  $\overline{M}_{S(\tau,t)}^{\text{iid}}(\theta) = (\overline{M}_{C_i}(\theta))^{t-\tau}$ , where  $\overline{M}_{S(\tau,t)}(\theta) = M_{S(\tau,t)}(-\theta)$ .

The distribution of the service process can be calculated

using Gil-Pelaez theorem [17]

$$F_{S(\tau,t)}(z) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \Im[e^{-jtx} M_{S(\tau,t)}(jt)] \frac{dt}{t}$$

$$= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{\log M_{S(\tau,t)}(t) - tx} \frac{dt}{t}$$

with  $c \in \mathbb{R}_{>0}$  in the convergence strip of the cumulant generating function  $\log M_{S(\tau,t)}(t)$ . The distribution can be conveniently evaluated numerically using Lévy's inversion theorem [18].

### B. Mellin transform

For calculating the delay violation probability using stochastic network calculus, the service process is characterized in terms of its Mellin transform (MT).

The MT of a nonnegative random variable  $X$  is defined as  $\mathcal{M}_X(s) \triangleq \mathbb{E}[X^{s-1}] = \overline{M}_{\log X}(s-1)$  for any  $s \in \mathbb{C}$  for which the expectation exists. For the service process, we have

$$\mathcal{M}_{S(\tau,t)}(s) \triangleq \mathbb{E}[(S(\tau,t))^{s-1}] = \int_{-\infty}^{\infty} z^{s-1} dF_{S(\tau,t)}(z).$$

According to (9) and (10), the MT of the service process for the independent and dependent case is given by, respectively

$$\mathcal{M}_{S(\tau,t)}^{\text{ind}}(s) \approx \int_{-\infty}^{\infty} z^{s-1} dQ\left(\frac{z - \mathbb{E}[S(\tau,t)]}{\sigma^2[S(\tau,t)]}\right),$$

$$\mathcal{M}_{S(\tau,t)}^{\text{dep},l}(s) = \int_{-\infty}^{\infty} z^{s-1} dF_{S(\tau,t)}^l(z) \leq \mathcal{M}_{S(\tau,t)}^{\text{dep}}(s)$$

$$\leq \int_{-\infty}^{\infty} z^{s-1} dF_{S(\tau,t)}^u(z) = \mathcal{M}_{S(\tau,t)}^{\text{dep},u}(s)$$

where the upper and lower bounds hold for  $s < 1$ .

## IV. DELAY PERFORMANCE IN STATIC NETWORKS

In this section, we analyze the QoS performance in terms of delay violation probability and effective capacity. We consider a static network, where the random locations of nodes do not vary with time, and perform the analysis given the locations of the nodes, i.e., conditioned on  $\Phi$ . The conditional success probability  $P_s(\xi)$  is a random variable that depends on the spatial distribution, which implies that the conditional service rate varies from node to node. This means that some nodes have an arbitrarily small service rate and consequently an arbitrarily large delay. Instead of deriving spatial averages for the delay metrics, we aim at obtaining the spatial distribution of the delay violation probability and of the effective capacity. For that, we derive the conditional service rate and calculate the delay metrics for each draw of points in the space. As a result, the delay violation probability and the effective capacity are random variables and we are interested in deriving their distribution.

### A. Delay Violation Probability

In this section, we obtain an upper bound on the delay violation probability using a statistical characterization of the arrival and service processes in the exponential (or SIR) domain [14]. First, we convert the cumulative processes in

the bit domain to the SIR domain (denoted by calligraphic letters) through the exponential function, i.e.,

$$\mathcal{A}(\tau, t) = e^{A(\tau, t)}, \quad \mathcal{D}(\tau, t) = e^{D(\tau, t)}, \quad \mathcal{S}(\tau, t) = e^{S(\tau, t)}.$$

An upper bound on the delay violation probability can be computed by means of the Mellin transforms of  $\mathcal{A}(\tau, t)$  and  $\mathcal{S}(\tau, t)$  [14]:

$$p_v(w) = \inf_{s>0} \{K(s, -w)\} \geq \Lambda(w) \quad (11)$$

where  $K(s, -w)$  is the steady-state kernel, defined as

$$\mathcal{K}(s, -w) = \lim_{t \rightarrow \infty} \sum_{u=0}^t \mathcal{M}_{\mathcal{A}}(1+s, u, t) \mathcal{M}_{\mathcal{S}}(1-s, u, t+w). \quad (12)$$

For the arrival process, assuming that  $\mathcal{A}(\tau, t)$  has stationary and independent increments, the MT becomes independent of the time instance, i.e.,

$$\begin{aligned} \mathcal{M}_{\mathcal{A}}(s, \tau, t) &= \mathbb{E} \left[ \left( \prod_{i=\tau+1}^t e^{a_i} \right)^{s-1} \right] \\ &= \mathbb{E} \left[ e^{\alpha(s-1)} \right]^{t-\tau} = \mathcal{M}_{\alpha}(s)^{t-\tau} \end{aligned}$$

where  $\alpha = e^a$  denotes the non-cumulative arrival process in the SIR domain. Using Chang's traffic characterization [2], we consider the traffic class of  $(\sigma(s), \rho(s))$ -bounded arrivals, whose MGF in the bit domain is bounded by

$$\frac{1}{s} \log \mathbb{E}[e^{sA(\tau, t)}] \leq \rho(s) \cdot (t - \tau) + \sigma(s) \quad (13)$$

for some  $s > 0$ . Restricting ourselves to the case where  $\rho$  is independent of  $s$  and  $\sigma(s) = 0$ , we have

$$\mathcal{M}_{\alpha}(s) = e^{\rho(s-1)}. \quad (14)$$

For the service process, as said before, we consider a static network and condition on  $\Phi$  (random but static over time). Therefore, the SIRs are conditionally independent and the random variations come from independent block fading for all active links. The MT of the (conditional) service process is given by

$$\begin{aligned} \mathcal{M}_{\mathcal{S}}(s, \tau, t) &= \mathbb{E} \left[ \left( \prod_{i=\tau}^{t-1} (1 + \text{SIR}_i) \right)^{s-1} \mid \Phi \right] \\ &= \mathbb{E} \left[ (1 + \text{SIR})^{s-1} \mid \Phi \right]^{t-\tau} = (\mathcal{M}_{\gamma}(s))^{t-\tau}. \end{aligned} \quad (15)$$

Plugging (14) and (15) into (12) and following [14], the steady-state kernel can be finally rewritten as

$$\mathcal{K}(s, -w) = \frac{(\mathcal{M}_{\gamma}(1-s))^w}{1 - \mathcal{M}_{\alpha}(1+s)\mathcal{M}_{\gamma}(1-s)}, \quad (16)$$

for any  $s > 0$  under the stability condition  $\mathcal{M}_{\alpha}(1+s)\mathcal{M}_{\gamma}(1-s) < 1$ . The upper bound on the conditional delay violation probability (11) is thus reduced to

$$p_v(w) = \inf_{s>0} \left\{ \frac{(\mathcal{M}_{\gamma}(1-s))^w}{1 - \mathcal{M}_{\alpha}(1+s)\mathcal{M}_{\gamma}(1-s)} \right\}. \quad (17)$$

The above delay bound can be calculated using the following result.

**Proposition 1.** *The MT of the service process in static networks is given by  $\mathcal{M}_{\mathcal{S}(\tau, t)}(s) = (\mathcal{M}_{\gamma}(s))^{t-\tau}$  where*

$$\mathcal{M}_{\gamma}(s) = 1 + (s-1) \int_0^{\infty} P_s(y)(1+y)^{s-2} dy, \text{ for } s < 1$$

$$\text{where } P_s(\xi) = \prod_{x \in \Phi \setminus \{o\}} \left( \frac{p}{1 + \xi r^{\alpha} \|x - R_o\|^{-\alpha}} + 1 - p \right).$$

*Proof:* See Appendix I. ■

The above semi-closed form expression requires numerical integration. For easier numerical evaluation and in order to gain insights, we provide the following upper bound on  $\mathcal{M}_{\gamma}(s)$ , which in turn provides an upper bound on the delay violation probability. Taking into account only the interference from the nearest interfering transmitter  $x_{\min} = \arg \min_{x \in \Phi \setminus \{o\}} \|x - R_o\|$ , we have

$$\begin{aligned} \mathcal{M}_{\gamma}^{u1}(s) &\leq 1 + (s-1) \int_0^{\infty} \frac{(1+y)^{s-2}}{1 + yr^{\alpha} \|x_{\min} - R_o\|^{-\alpha}} dy \\ &= 1 + \frac{(s-1) {}_2F_1(2-s, 2-s; 1-s; 1-Z)}{Z^{s-1}(2-s)} \\ &\stackrel{(a)}{\leq} 1 + (s-1)[(3-2s)Z]^{-1/2} = \mathcal{M}_{\gamma}^{u2}(s) \end{aligned}$$

where  ${}_2F_1(a, b; c; x)$  is the Gauss hypergeometric function,  $Z = r^{\alpha} \|x_{\min} - R_o\|^{-\alpha}$  and (a) follows applying Cauchy-Schwarz inequality.

Calculating the the distribution of the (conditional) delay violation probability  $\mathbb{P}^{!o}(p_v(w) > x)$ ,  $x \in [0, 1]$  is complex. The distribution of the upper bound based on nearest neighbor can be calculated as follows. Since the delay violation probability  $p_v(w)$  is a decreasing function (say  $g$ ) of the random variable  $\|x_{\min}\|$ , we have

$$\begin{aligned} \mathbb{P}^{!o}(p_v(w) > x) &\leq \mathbb{P}^{!o}(p_v^u(w) > x) = \mathbb{P}^{!o}(g(\|x_{\min}\|) > x) \\ &= \mathbb{P}^{!o}(\|x_{\min}\| > g^{-1}(x)) = e^{-\lambda \pi (g^{-1}(x))^2}. \end{aligned}$$

## B. Effective Capacity

The effective capacity is defined as the maximum constant arrival rate at a buffer that can be supported by the service process while satisfying statistical QoS requirements specified by the QoS exponent  $\theta$  [3]. For time-varying arrival rates, effective capacity specifies the effective bandwidth of the arrival process that can be supported by the channel.

Let  $Q$  be the the stationary queue length, then  $\theta$  is the decay rate of the tail distribution of the queue length  $Q$

$$\lim_{q \rightarrow \infty} \frac{\log \mathbb{P}(Q \geq q)}{q} = -\theta \quad (18)$$

and from Gärtner-Ellis Theorem the buffer violation probability for large  $q_{\max}$  is approximated as  $\mathbb{P}(Q \geq q_{\max}) \approx e^{-\theta q_{\max}}$ . Therefore, larger  $\theta$  corresponds to more strict QoS constraints, while smaller  $\theta$  implies looser constraints.

In block fading channels with coherence time  $T$ , the effec-

tive capacity simplifies to

$$\begin{aligned}\mathcal{R}(\theta T) &\triangleq -\frac{1}{\theta T} \log \mathbb{E} [e^{-\theta T C_t}] = -\frac{1}{\theta T} \log \mathcal{M}_\gamma(1 - \theta T) \\ &= -\frac{1}{\theta T} \log \mathbb{E} [(1 + \text{SIR}_t)^{-\theta T}]\end{aligned}$$

Using that for a positive random variable  $X$ ,  $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq t) dt$ , we can have the following alternative expression for  $\Psi(\theta T) = \mathbb{E} [(1 + \text{SIR}_t)^{-\theta T}]$

$$\Psi(\theta T) = 1 - \int_0^1 P_s(t^{-\frac{1}{\theta T}} - 1) dt.$$

The distribution of the effective capacity can be calculated using Gil-Pelaez theorem, which involves numerical integration and does not provide much insight on the behavior of the effective capacity. For that, we establish below bounds on the distribution using classical concentration inequalities. The simplest upper bound on the complementary cdf (ccdf) follows from Markov's inequality:

$$\mathbb{P}^{l_0}(\mathcal{R}(\theta T) > x) \leq \frac{\mathbb{E}[\mathcal{R}(\theta T)]}{x} \quad (19)$$

where the first moment can be upper bounded as follows:

$$\begin{aligned}\mathbb{E}[\mathcal{R}(\theta T)] &= -\frac{1}{\theta T} \mathbb{E}[\log \Psi(\theta T)] \stackrel{(a)}{\leq} -\frac{1}{\theta} \log \mathbb{E}[\Psi(\theta T)] \\ &\stackrel{(b)}{=} -\frac{1}{\theta T} \log \left( 1 - \int_0^1 \mathbb{E}^{l_0}[P_s(t^{-\frac{1}{\theta T}} - 1)] dt \right) \\ &\stackrel{(c)}{=} -\frac{1}{\theta T} \log \left( 1 - \int_0^1 e^{-\lambda \mathcal{I}} dt \right) \\ &= -\frac{1}{\theta T} \log \left( 1 - \int_0^1 e^{-\lambda \pi p r^2 \mathcal{C}(\alpha) (t^{-\frac{1}{\theta T}} - 1)^{2/\alpha}} dt \right) \\ &= -\frac{1}{\theta T} \log \left( 1 - \theta T \int_0^\infty \frac{e^{-\lambda \pi p r^2 \mathcal{C}(\alpha) y^{2/\alpha}}}{(1+y)^{1+\theta T}} dy \right)\end{aligned}$$

where  $\mathcal{I} = \int_{\mathbb{R}^2} \left( 1 - \frac{1}{1 + (t^{-\frac{1}{\theta T}} - 1) r^\alpha \|x - R_o\|^{-\alpha}} \right) dx$ , and  $\mathcal{C}(\alpha) = \frac{2}{\alpha} B(1 - 2/\alpha, 2/\alpha)$  with  $B(a, b)$  denoting the beta function. Step (a) follows using Jensen's inequality, (b) by exchanging expectation and integration order (Fubini-Tonelli's Theorem), (c) from the probability generating functional of the PPP. The first moment can be further bounded applying Cauchy-Schwarz inequality in the integral inside the logarithm.

A lower bound on the ccdf of the effective capacity can be found using the Paley-Zygmund or the reverse Markov inequality, which however involves the calculation of the mean and the variance of  $\mathcal{R}(\theta T)$ . Using the upper bound on the Mellin transform based on the nearest interferer, we can establish the following lower bound on the ccdf:

$$\begin{aligned}\mathbb{P}^{l_0}(\mathcal{R}(\theta T) > x) &= \mathbb{P}^{l_0}(\mathcal{M}_\gamma(1 - \theta T) < e^{-x\theta T}) \\ &\geq \mathbb{P}^{l_0}(\mathcal{M}_\gamma^u(1 - \theta T) < e^{-x\theta T}) \\ &= \mathbb{P}^{l_0} \left( Z^{-1/2} < \zeta \right) = 1 - e^{-\lambda \pi r^2 \zeta^{\frac{4}{\alpha}}}\end{aligned}$$

where  $\zeta = \frac{e^{-x\theta T} - 1}{\theta T(1+2\theta T)^{-1/2}}$ .

## V. NUMERICAL RESULTS

In this section, we validate the above analysis and provide numerical evaluation of the delay performance. The duration of a slot is set to  $T = 1$  ms and the blocklength is  $N = 100$ . For the Poisson bipolar network we have a density of  $\lambda = 1$  node/km<sup>2</sup> and pathloss exponent  $\alpha = 3.5$ .

We start by validating the upper bound on the delay violation probability with Monte Carlo simulations. In Figure 1, we compare the delay violation probability and its bound for link distance  $r = 0.3$  km. We corroborate that the analytical bound follows the trend of the simulated curve, having a difference of about 1 – 2 ms (equivalent to one to two slots).

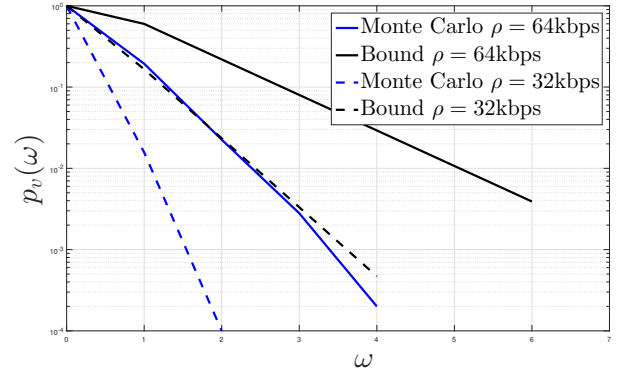


Figure 1. Delay violation probability and associated bound as a function of the target delay for  $\rho = 32$  kbps and  $\rho = 64$  kbps.

In Figure 2, we compare the distribution of the delay violation probability and that of the analytical bound for  $\rho = 64$  kbps  $r = 0.2$  km. We observe that the analytical bound becomes tighter for  $\omega$  increasing.

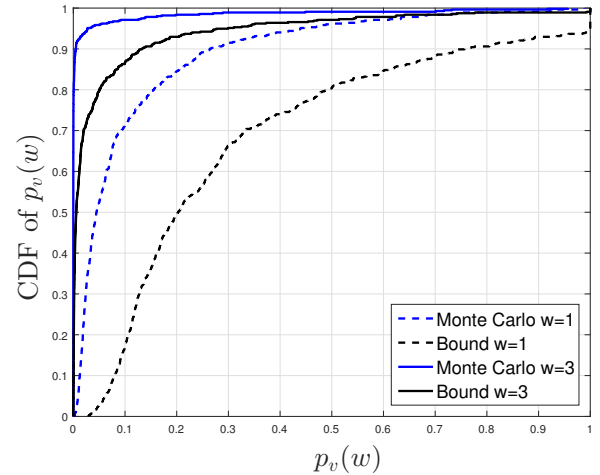


Figure 2. Distribution of delay violation probability for  $\omega = 1$  ms and  $\omega = 3$  ms.

Finally, in Figure 3, we plot the delay violation probability and its analytical bound as a function of the inter-node distance  $r$  for two different values of  $\omega$ . As expected, the more

stringent the delay constraint is, the closer the transmitter and its intended receiver should be. Alternatively, for fixed link distance, tighter delay constraints can be guaranteed for lower density of interferers  $\lambda$ .

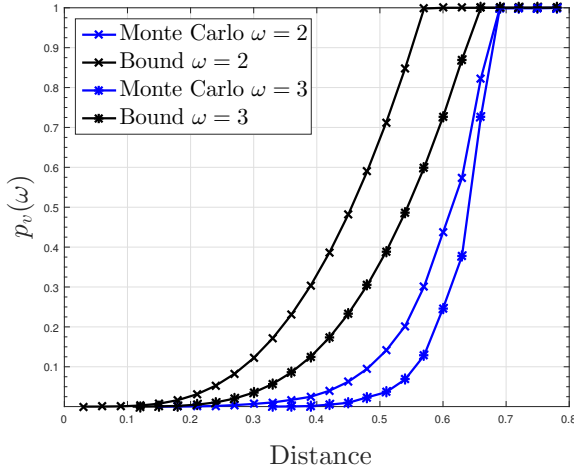


Figure 3. Delay violation probability and associated bound as a function of the link distance  $r$  for  $\rho = 64$  kbps.

## VI. CONCLUSIONS

We have investigated the delay performance of large wireless networks in the presence of statistical QoS constraints. We have characterized the distribution of the conditional delay violation probability and effective capacity in Poisson bipolar networks. Our results provide useful insight into providing delay guarantees in random spatial networks. From a broader perspective, this paper is a first attempt to combine stochastic network calculus with stochastic geometry as a means to quantify the delay in wireless networks with spatial randomness.

## ACKNOWLEDGMENT

The work of N. Pappas was supported in part by the Center for Industrial Information Technology (CENIIT).

## APPENDIX I PROOF OF PROPOSITION 1

We start by deriving the conditional success probability  $P_s(\xi)$ , i.e., the probability that a transmission will be successful by exceeding  $\xi$  conditioned upon  $\Phi$  [19].

$$\begin{aligned}
 P_s(\xi) &= \mathbb{P}(\text{SIR}_t > \xi \mid \Phi) \\
 &= \mathbb{P}(h_{oR_o}^t r^{-\alpha} > \xi I_{R_o,t} \mid \Phi) \\
 &\stackrel{(a)}{=} \mathbb{E}(\exp(-\xi r^\alpha I_{R_o,t}) \mid \Phi) \\
 &= \mathbb{E}\left(\exp\left(-\sum_{x \in \Phi \setminus \{o\}} \xi r^\alpha h_{xR_o}^t \|x - R_o\|^{-\alpha} \mathbb{1}(x \in \Phi_t)\right) \mid \Phi\right) \\
 &= \prod_{x \in \Phi \setminus \{o\}} \mathbb{E}\left(\exp(-\xi r^\alpha h_{xR_o}^t \|x - R_o\|^{-\alpha} \mathbb{1}(x \in \Phi_t)) \mid \Phi\right) \\
 &\stackrel{(b)}{=} \prod_{x \in \Phi \setminus \{o\}} \left(\frac{p}{1 + \xi r^\alpha \|x - R_o\|^{-\alpha}} + 1 - p\right).
 \end{aligned}$$

where (a) and (b) follows because the fading coefficients are i.i.d. random variables with exponential distribution of unit mean.

Therefore, using integration by parts, we have

$$\begin{aligned}
 \mathcal{M}_\gamma(s) &= \int_0^\infty (1+y)^{s-1} d\mathbb{P}(\text{SIR} < y \mid \Phi) \\
 &= -(s-1) \int_0^\infty (1+y)^{s-2} \mathbb{P}(\text{SIR} < y \mid \Phi) dy \\
 &= 1 + (s-1) \int_0^\infty (1+y)^{s-2} P_s(y) dy.
 \end{aligned}$$

## REFERENCES

- [1] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.
- [2] —, *Performance Guarantees in Communication Networks*. London, UK: Springer-Verlag, 2000.
- [3] 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, July 2003.
- [4] J. Tang and X. Zhang, "Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2318–2328, June 2008.
- [5] R. K. Ganti and M. Haenggi, "Spatial and temporal correlation of the interference in aho ad hoc networks?" *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 631–633, Sept. 2009.
- [6] R. K. Ganti and J. G. Andrews, "Correlation of link outages in low-mobility spatial wireless networks," in *Proc. 44th Asilomar Conference on Signals, Systems and Computers*, Nov. 2010, pp. 312–316.
- [7] Y. Zhong, W. Zhang, and M. Haenggi, "Managing interference correlation through random medium access," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 928–941, Feb. 2014.
- [8] F. Baccelli and B. Blaszczyszyn, "A new phase transitions for local delays in MANETs," in *Proc. IEEE INFOCOM*, Mar. 2010.
- [9] M. Haenggi, "The local delay in Poisson networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1788–1802, Mar. 2013.
- [10] Z. Chen, N. Pappas, M. Kountouris, and V. Angelakis, "Throughput analysis of smart objects with delay constraints," in *Proc. IEEE 17th Intern. Symp. WoWMoM*, June 2016.
- [11] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. London, UK: Springer-Verlag, 2008.
- [12] M. Fidler and A. Rizk, "A guide to the stochastic network calculus," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 92–105, First quarter 2015.
- [13] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. I*, 2nd ed. New York, NY, USA: Springer-Verlag, 2003.
- [14] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "A (min, x) network calculus for multi-hop fading channels," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 1833–1841.
- [15] R. Nelsen, *An Introduction to Copulas*. New York, USA: Springer-Verlag, 2006.
- [16] F. Sun and Y. Jiang, "Further properties of wireless channel capacity," 2016. [Online]. Available: <https://arxiv.org/abs/1502.00979v2>
- [17] J. Gil-Pelaez, "Note on the inversion theorem," *Biometrika*, vol. 38, pp. 481–482, Dec. 1951.
- [18] W. Feller, *An Introduction to Probability Theory and Its Applications, vol. II*, 2nd ed. NY, USA: Wiley & Sons, Inc., Jan. 1991.
- [19] M. Haenggi, "The meta distribution of the SIR in Poisson bipolar and cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2577–2589, Apr. 2016.