Interference Statistics Approximations for Data Rate Analysis in Uplink Massive MTC

Sergi Liesegang, Olga Muñoz, and Antonio Pascual-Iserte
Dept. Signal Theory and Communications - Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
Emails: {sergi.liesegang, olga.munoz, antonio.pascual}@upc.edu

***Abstract*—Machine-type-communications have attracted a lot of interest in the past years. They rely on interactions between devices with no human supervision. This will help to the advent of a plethora of applications such as the Internet-of-Things. Part of the research within this field deals with coordinating the access of a large number of devices to the network, the so-called *massive* machine-type-communications. In this paper, we focus on the evaluation of the data rate for that scenario, based on an approximation of the statistics of the aggregated interference that depends on the sensors activity. We will consider that the sensors can be in either active or sleep mode, modeled as a Bernoulli random variable. This results in an aggregated interference that follows a discrete distribution whose computation becomes unfeasible with the number of devices. That is why two alternatives are presented to replace the original magnitude and work with an analytic closed form expression approximating the actual statistics. Our approaches are derived using the Chernoff bound and a Gaussian approximation based on Lyapunov's central limit theorem. The average rate is found in both cases and compared with the actual values in different setups. Monte-Carlo simulations will be used for this task.**

***Index Terms*—Machine-type-communications, average data rate, Bernoulli, Chernoff bound, Lyapunov's condition**

I. Introduction

Machine-type-communications (MTC) will play an important role in future generations of mobile networks [1]–[3]. This is the reason why they have received a lot of attention during the past years. MTC consist of a set of interactions between devices with barely or even no human supervision [4]. According to [5], the number of connected terminals is expected to grow exponentially. MTC will lead the creation of lots of unprecedented applications and will open the door for new challenges and investigations. Emerging concepts like the Internet-of-Things (IoT) will be one out of the many possibilities this field will allow [6], [7].

In this paper, we will focus on a specific scenario where a large number of devices, such as sensors that gather information, try to access the network. Hence, an uplink (UL) communication is considered. This setup is usually referred to as *massive* MTC (mMTC) [8] and, like in common human-to-human communications, a study of the data rate is appropriate. In this framework, it would be very useful to obtain closed form expressions that take into account the particularities of the system, namely the sensors activity.

The work presented in this paper has been carried out within the framework of the project 5G&B RUNNER-UPC (TEC2016-77148-C2-1-R (AEI/FEDER, UE), funded by the Spanish Ministry of Science, Innovation and Universities) and is partially funded by the Catalan Government (2017 SGR 578 - AGAUR)

In real scenarios, these kind of devices are in a sleep mode for most of the time, which makes the situation of all of them transmitting simultaneously extremely rare. Hence, it is imperative to include their state in the data rate analysis. In particular, we focus on the statistics of the aggregated interference that the signal of each sensor experiences when it is received at the collector node (CN). Due to the discrete nature (given the on/off states of the sensors) of the aggregated interference, in this paper we propose two approximations of its statistics, namely one based on the Chernoff upper bound and the other on a Gaussian distribution. This way, the complexity is significantly reduced and a closed form analytic expression is provided. These expressions will be finally used for the evaluation of the UL average data rate with the help of standard numerical integration methods. For setups with fixed activity, closed-form expressions can be also obtained [9].

The remaining of this paper is structured as follows. In Section II, the system model, the average data rate and the probability distribution are described. Then, the two alternative expressions for this last magnitude are proposed in Section III. Simulations are described in Section IV and are used to compare the experimental and the approximation results. Finally, Section V is dedicated to conclusions.

II. System Model

In the following, we will consider a single-cell system with one data CN and M sensors deployed uniformly around it, i.e. inside the coverage region defined by a circle of radius R centered at the CN location. Such setup is depicted in Fig. 1 for $M = 9$. The UL received signal can be modeled as

$$y = \sum_{i=1}^{M} h_i x_i + n, \tag{1}$$

where $y \in \mathbb{C}$ is the received signal, $n \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ the corresponding Gaussian complex and circularly symmetric noise, $x_i \sim \mathcal{N}_{\mathbb{C}}(0, P_i)$ the transmit signal[1], and $h_i \in \mathbb{C}$ the individual channels with $i \in \{1, M\}$. Note that these last terms, later defined, will be different for each sensor as they depend on the distance and devices have different locations.

In this context, the date rate of the link for a certain sensor i can be calculated using Shannon's well-known formula, i.e.

$$R_i = \log_2(1 + \rho_i), \tag{2}$$

[1]Thanks to the Gaussian assumption, the data rate can be defined in terms of the SINR only and in a closed form expression, as indicated in (2).

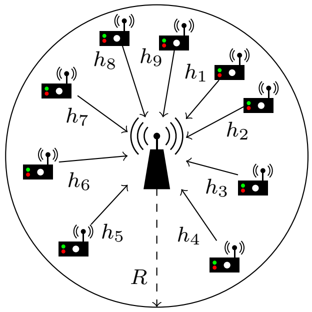

Fig. 1. System setup for $M = 9$

where ρ_i is the associated SINR that can be expressed as

$$\rho_i = \frac{P_i|h_i|^2}{1 + \sum_{j \neq i} \Gamma_j}. \tag{3}$$

Note that throughout this work we will assume channels h_i to be fixed and known by the receiver[1]. Thereby, all randomness will be given by the sensors activity. Hence, to model such behavior, we introduce the random variables (RV) Γ_j in the previous expression. They are defined as follows

$$\Gamma_j = \beta_j P_j |h_j|^2, \quad j = 1, \ldots, M \tag{4}$$

where $\beta_j \sim$ *Ber*(p_j) and p_j is the transmission probability of the j-th device, not to be confused with its transmit power P_j. Thus, p_j corresponds to the percentage of time that the j-th sensor is transmitting. Nevertheless, and for the sake of clarity in the presentation, we will consider all sensors to have the same activity, i.e. $p_j = p$. Thanks to the introduction of this model, the state of the sensor can be included in the analysis. Note that since we focus on actual data transmission, in (3) we have assumed that the reference sensor for which we want to evaluate the data rate, name sensor i, is active, i.e. $\beta_i = 1$.

Moreover, when denoting $\tilde{\Gamma}_i = \sum_{j \neq i} \Gamma_j$ as the aggregated interference for sensor i, the data rate from (2) reads as

$$R_i(\tilde{\Gamma}_i) = \log_2 \left(1 + \frac{P_i|h_i|^2}{1 + \tilde{\Gamma}_i} \right), \tag{5}$$

which is also a RV and whose statistics can be computed by the probability mass function (pmf) of the aggregated interference, i.e. $p_{\tilde{\Gamma}_i}(\tilde{\gamma}_i)$. Note that since $\tilde{\Gamma}_i$ is random, also the instantaneous rate in (5) will be random. Thereby, in this work we will focus on the evaluation of the average rate, i.e. $\mathrm{E}[R_i(\tilde{\Gamma}_i)]$. To this end, we will first derive an analytic approximated expression of the statistical distribution of $\tilde{\Gamma}_i$.

The distribution of $\tilde{\Gamma}_i$ is the result of a series of discrete convolutions (since $\tilde{\Gamma}_i$ is the sum of a set of independent RVs, where each one only takes two possible values) and can be very difficult to calculate. In fact, the number of operations needed grows exponentially with the number of interfering devices (here $M - 1$) and, therefore, it becomes quickly unfeasible even for a small M. That is why in the next section we propose two different alternatives to express the previous statistics through continuous approximations.

[1]Given the static positions of the sensors, it is feasible to estimate the path loss component as well as the fading coefficients at the receiver side.

III. Probability Distribution Alternatives

The main purpose of this work is to find an alternative for the pmf of the aggregated interference $\tilde{\Gamma}_i$ corresponding to sensor i. We will address this issue in two different ways. First, by means of the Chernoff upper bound and, later, with the help of a Gaussian distribution. Both approximations will lead to a continuous expression of the pmf, i.e. the probability distribution function (pdf) $f_{\tilde{\Gamma}_i}(\tilde{\gamma}_i)$. Accordingly, the average data rate will then read as

$$\mathrm{E}[R_i(\tilde{\Gamma}_i)] = \sum R_i(\tilde{\Gamma}_i) p_{\tilde{\Gamma}_i}(\tilde{\gamma}_i) \approx \int R_i(\tilde{\Gamma}_i) f_{\tilde{\Gamma}_i}(\tilde{\gamma}_i) d\tilde{\gamma}_i. \tag{6}$$

Note that the integral in (6) will be computed numerically as an analytic expression is difficult to be obtained.

A. Approximation Based on the Chernoff Upper Bound

The first proposal relies on the Chernoff bound and follows a similar derivation to that in [10]. In particular, we will first find the cumulative distribution function (cdf) with this approach for later obtaining the pdf from its derivative. We will use the notation $F(\tilde{\gamma}_i)$ for the cdf of $\tilde{\Gamma}_i$, i.e. $F(\tilde{\gamma}_i) = \Pr\{\tilde{\Gamma}_i \leq \tilde{\gamma}_i\}$. Nevertheless, we will use the Chernoff bound on the complementary cdf (ccdf) with notation $\bar{F}(\tilde{\gamma}_i) = 1 - F(\tilde{\gamma}_i)$.

Let us start then by applying Markov's inequality [11] to the moment generating function e^t for all $t > 0$:

$$\bar{F}(\tilde{\gamma}_i) = \Pr\{\tilde{\Gamma}_i > \tilde{\gamma}_i\} = \Pr\{e^{t\tilde{\Gamma}_i} > e^{t\tilde{\gamma}_i}\} < e^{-t\tilde{\gamma}_i}\mathrm{E}[e^{t\tilde{\Gamma}_i}]. \tag{7}$$

Since all β_j are independent, by defining $a_j = P_j|h_j|^2$ we will have that

$$\begin{aligned} e^{-t\tilde{\gamma}_i}\mathrm{E}[e^{t\tilde{\Gamma}_i}] &= e^{-t\tilde{\gamma}_i} \prod_{j \neq i} \mathrm{E}[e^{t\beta_j a_j}] \\ &= e^{-t\tilde{\gamma}_i} \prod_{j \neq i} [p_j e^{t a_j} + (1 - p_j)]. \end{aligned} \tag{8}$$

For a general $\delta > 0$, later defined, by choosing $t = \ln(1+\delta)$, which ensures t will be positive, the previous expression yields

$$\begin{aligned} &(1+\delta)^{-\tilde{\gamma}_i} \prod_{j \neq i} [p_j (1+\delta)^{a_j} + (1 - p_j)] \\ &\qquad \leq (1+\delta)^{-\tilde{\gamma}_i} \prod_{j \neq i} e^{p_j[(1+\delta)^{a_j} - 1]}, \end{aligned} \tag{9}$$

which follows from $x + 1 \leq e^x$. Finally, for $a_j \in (0, 1]$ we have that $(1+\delta)^{a_j} - 1 \leq \delta a_j$ and, thus

$$\bar{F}(\tilde{\gamma}_i) \leq (1+\delta)^{-\tilde{\gamma}_i} e^{\sum_{j \neq i} \delta a_j p_j} = (1+\delta)^{-\tilde{\gamma}_i} e^{\delta \tilde{\mu}_i}, \tag{10}$$

where $\tilde{\mu}_i$ is the statistical mean of $\tilde{\Gamma}_i$ and reads as

$$\tilde{\mu}_i = \sum_{j \neq i} p_j a_j = p \sum_{j \neq i} a_j. \tag{11}$$

Regarding δ, we opted for choosing it equal to the first moment $\tilde{\mu}_i$ as it is strictly positive in this scenario ($a_j > 0$) and captures the basic statistics of $\tilde{\Gamma}_i$. Note also that the previous holds only when normalizing all a_j so that the maximum value is not higher than 1. Finally, the ccdf yields

$$\bar{F}(\tilde{\gamma}_i) \leq e^{\tilde{\mu}_i^2} (1 + \tilde{\mu}_i)^{-\tilde{\gamma}_i} = \bar{F}_C(\tilde{\gamma}_i) = 1 - F_C(\tilde{\gamma}_i), \tag{12}$$

where C stands for Chernoff. The approximation that we propose is to take as the pdf the derivative of the previous bound of the cdf, i.e. $F_C(\tilde{\gamma}_i)$, where the derivative is scaled conveniently with a factor K such that the area of that pdf is equal to 1. Hence, the pdf can be expressed as

$$f_{C,\tilde{\Gamma}_i}(\tilde{\gamma}_i) = K\frac{dF_C(\tilde{\gamma}_i)}{d\tilde{\gamma}_i} = Ke^{\tilde{\mu}_i^2}(1+\tilde{\mu}_i)^{-\tilde{\gamma}_i}\ln(1+\tilde{\mu}_i). \quad (13)$$

Consequently, to obtain the average data rate we will replace the pdf in (6) and evaluate the integral numerically. This will be done in Section IV, where simulations are shown.

B. Approximation Based on the Gaussian Distribution

The other alternative for the approximation of the pmf of the aggregated interference is based on a Gaussian distribution, denoted by the kernel $\phi(\tilde{\gamma}_i)$ with mean $\tilde{\mu}_i$ and variance $\tilde{\sigma}_i$. The former moment is already defined in (11) and the latter can be obtained similarly as follows

$$\tilde{\sigma}_i^2 = \sum_{j\neq i} p_j(1-p_j)a_j^2 = p(1-p)\sum_{j\neq i} a_j^2. \quad (14)$$

The validity of this approximation relies on a special version of the central limit theorem (CLT), which states that $\tilde{\Gamma}_i$ converges to a Gaussian distribution when the number of addends $M-1$ is very high (true for mMTC [8]), and these addends are independent but not necessarily equally distributed as long as Lyapunov's condition is fulfilled [12]. Note that in our scenario, the RVs that are added have different statistics since they correspond to sensors located at different distances from the CN, which implies that they are not equally distributed. Moreover, to emphasize such claim, the accuracy of this approximation will be later evaluated in Section IV.

Note that the previous approach assumes $\tilde{\Gamma}_i$ to have an infinite support, although it is actually lower and upper bounded by 0 and $I_i = \sum_{j\neq i} a_j$, respectively. These are the extreme cases for the aggregated interference. Consequently, we must use a truncated kernel with unit area. The same applies in the case of the Chernoff bound, where we also need to normalize the expression by its area within the limits 0 and I_i.

Therefore, in order to get the average data rate we need to substitute $f_{\tilde{\Gamma}_i}(\tilde{\gamma}_i)$ by $\phi(\tilde{\gamma}_i)$ in (6) and solve the integral. However, as already mentioned, a closed form expression for $\mathrm{E}[R_i(\tilde{\Gamma}_i)]$ would be difficult to find. That is why its computation is reserved to numerical methods.

Proof of Lyapunov's condition: Lyapunov's CLT is a variant of the classical CLT where the sum of a sequence of independent RVs with different statistics converges in distribution to a standard Gaussian RV under a certain condition. That condition is the following: for some $\epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{\tilde{\sigma}_i^{2+\epsilon}} \sum_{j\neq i} \mathrm{E}\Big[\big|a_j\beta_j - a_jp_j\big|^{2+\epsilon}\Big] = 0, \quad (15)$$

where $n = M-1$ is the number of addends. It can be shown that the individual terms can be upper bounded by

$$\mathrm{E}\Big[\big|a_j\beta_j - a_jp_j\big|^{2+\epsilon}\Big] \leq a_j^{2+\epsilon}p_j(1-p_j). \quad (16)$$

Thereby, the condition in (15) will be satisfied as long as the following ratio tends to zero for $n\to\infty$

$$\frac{1}{\tilde{\sigma}_i^{2+\epsilon}} \sum_{j\neq i} a_j^{2+\epsilon}p_j(1-p_j) = \frac{1}{\big(p(1-p)\big)^{\epsilon/2}}\left(\frac{\|\tilde{\boldsymbol{a}}_i\|_{2+\epsilon}}{\|\tilde{\boldsymbol{a}}_i\|_2}\right)^{2+\epsilon}, \quad (17)$$

where the notation $\tilde{\boldsymbol{a}}_i$ refers to the vector containing the interfering values. From now on, assuming the term $p(1-p)$ is not close to 0, we will prove the remaining ratio tends to zero with the help of the following relations between norms:

$$\|\boldsymbol{x}\|_r \leq \|\boldsymbol{x}\|_q \leq n^{\left(\frac{1}{q}-\frac{1}{r}\right)}\|\boldsymbol{x}\|_r, \quad (18)$$

for any vector $\boldsymbol{x}$ of length n and $0 < q < r$. This, translated to our scenario, leads to the following inequalities

$$1 \geq \left(\frac{\|\tilde{\boldsymbol{a}}_i\|_{2+\epsilon}}{\|\tilde{\boldsymbol{a}}_i\|_2}\right)^{2+\epsilon} \geq n^{1-(2+\epsilon)/2}, \quad (19)$$

where the right hand side tends to zero with $n\to\infty$ and any positive value of ϵ. Besides, it is straightforward to see that this bound is tight for $\epsilon\to\infty$ as $\|\tilde{\boldsymbol{a}}_i\|_\infty = \max_j \mathrm{a}_j < \infty$, and the ratio $\|\tilde{\boldsymbol{a}}_i\|_\infty/\|\tilde{\boldsymbol{a}}_i\|_2$ will be always smaller than 1 for any $n > 1$. In fact, for $n\to\infty$ the speed of convergence is much faster as $\|\tilde{\boldsymbol{a}}_i\|_2\to\infty$. This concludes the proof.

IV. Numerical Results

In this section, several simulations will be presented to compare the actual value and the two approximations of the pmf, as well as the resulting data rate. The former will be computed using $N = 10000$ Monte-Carlo simulations in which the same sensor is sending packets and the rest are transmitting randomly. Such interfering devices will behave according to a Bernoulli RV with transmission probability p. Once the distributions are approximated analytically, the average data rate in (6) will be obtained numerically.

This simulation model connects directly with the assumption behind the analytic approximating expressions and it is related to the widely used traffic model where the number of packets at a given time instant follows a Poisson distribution. Note that the sum of a large number M of Bernoulli RVs, with parameter p tending to zero, approaches the Poisson distribution with $\lambda = Mp$, which is true for real scenarios as lots of sensors are deployed [8]. The same applies for $M-1$, the actual number of interfering devices. Also, it is worth mentioning all transmission slots have to be aligned in time (synchronized, like in slotted ALOHA [13]) and no overlapping among packet transmissions from the same sensor is permitted for our simulation results to be consistent.

Regarding the simulations, we will represent the actual and approximated distributions of the interference of a certain device by means of the cdf. This is illustrated in Fig. 2. In addition, the resulting average data rate will be shown with respect to the instantaneous value in Fig. 3. Note that a standard IoT bandwidth of 180 kHz is considered here [7]. Then, to compare both strategies, we will focus on the relative error between the average data rates, measured in percentage. More specifically, we will compute the difference between the original and the approximated (Chernoff based or Gaussian based) average data rates and we will normalize it with respect to the original magnitude. In that sense, we will present different results where we sweep the values of M and p for a fixed $p = 0.3$ and $M = 1000$, respectively. Such results are plotted in Fig. 4 and Fig. 5. Note that we assume the power of the interfering devices to be fixed, i.e. $P_j = P = 10$ W, including that of sensor i, i.e. $P_i = P$.

Furthermore, we will consider the individuals channels of the sensors to be defined as $h_i = d_i^{-\alpha/2} g_i$, where d_i is the distance between sensor i and the CN, $\alpha = 2$ is the path loss exponent, and $g_i \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ is the Rayleigh distributed fast fading coefficient that changes in each time realization. In addition, a cell radius of $R = 100$ m is assumed.

As we can see in Fig. 2, the Chernoff based approach provides a poor accuracy in the approximation of the aggregated interference distribution. On the contrary, the Gaussian approximation adapts suitably to the actual statistics and reveals a promising performance. This can also be seen in Fig. 3, where the average value differs considerably in the case of the Chernoff bound. It does not represent accurately the average of the instantaneous data rate, which in turn varies considerably with respect to the actual mean. Differently, the Gaussian approach reveals a good accuracy as it provides an almost exact average value. Besides, the same behavior can be observed when looking at the relative error (%) in Fig. 4 and Fig. 5. The Gaussian approximation shows a robust performance when varying M, as illustrated in Fig. 4. Also, it does not vary significantly when changing p in the Gaussian case as both magnitudes, the actual and the approximated ones, behave in a similar way. In these sweeps, maximum relative error values of 2.19% and 4.88% are attained when computing the average data rate, which proves that the Gaussian approximation can provide accurate results with far less operations and a sufficient number of interfering sensors. On the other hand, the Chernoff approach achieves maximum relative errors of 153.38% and 122.25% respectively, which shows that it is not adequate for replacing the actual pmf. In fact, it worsens when increasing the values of p and M as only a single statistical parameter is used for modeling the distribution and the upper bound used for the ccdf results too loose.

V. Conclusions

In this paper we have addressed the issue of how to represent the statistics of the interference seen by the sensor information collector in the UL, taking into account the activity of sensors in a wireless network. In particular, the discrete statistics of the aggregated interference have been presented and modeled through Chernoff based and Gaussian based approximations. Such approaches provide closed form analytic approximating expressions to work with and, specifically, the latter has shown to be adequate for replacing the original magnitude. Based on that, the average data rate can be computed numerically with a significant lower complexity and, in the Gaussian approach, with a promising accuracy. In fact, only the two first order moments are required for such computation.

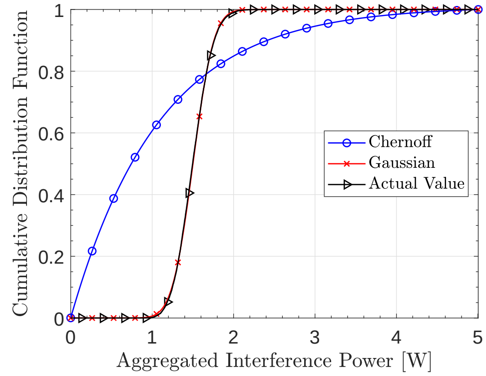

Fig. 2. Actual and approximated cdfs for $M = 1000$ and $p = 0.3$

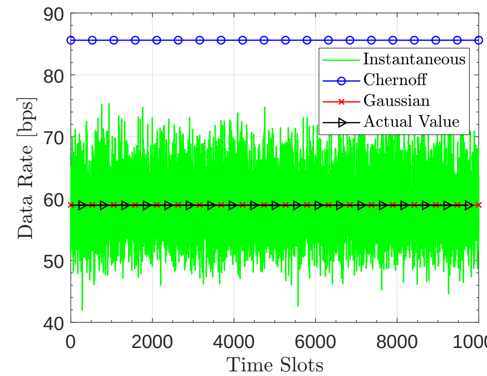

Fig. 3. Instantaneous and average data rate for $M = 1000$ and $p = 0.3$

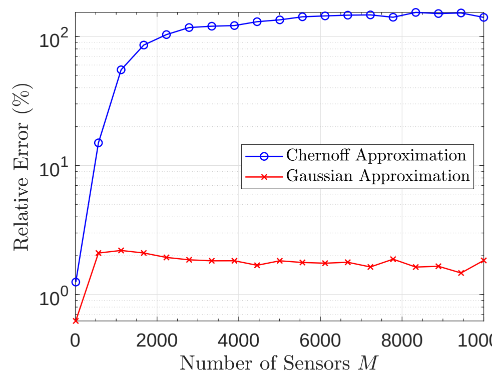

Fig. 4. Relative error in the average data rate calculation versus number of sensors M with $p = 0.3$

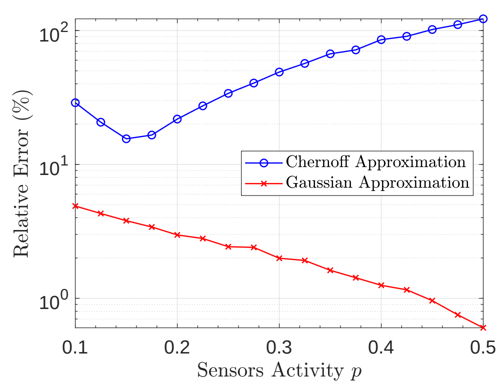

Fig. 5. Relative error in the average data rate calculation versus sensors activity p with $M = 1000$

References

[1] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu. Device-to-device communication in 5g cellular networks: challenges, solutions, and future directions. *IEEE Communications Magazine*, 52(5):86–92, May 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. What will 5g be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, June 2014.

[3] X. Xiong, K. Zheng, R. Xu, W. Xiang, and P. Chatzimisios. Low power wide area machine-to-machine networks: key techniques and prototype. *IEEE Communications Magazine*, 53(9):64–71, September 2015.

[4] H. Wang and A. O. Fapojuwo. A survey of enabling technologies of low power and long range machine-to-machine communications. *IEEE Communications Surveys Tutorials*, 19(4):2621–2639, Fourthquarter 2017.

[5] P. Kela, H. Lundqvist, M. Costa, K. Leppnen, and R. Jntti. Connectionless access for massive machine type communications in ultra-dense networks. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2017.

[6] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jntti, and A. Ghosh. Machine-type communications: current status and future perspectives toward 5g systems. *IEEE Communications Magazine*, 53(9):10–17, September 2015.

[7] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen. Narrowband internet of things: Evolutions, technologies, and open issues. *IEEE Internet of Things Journal*, 5(3):1449–1462, June 2018.

[8] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy. Massive machine-type communications in 5g: physical and mac-layer solutions. *IEEE Communications Magazine*, 54(9):59–65, September 2016.

[9] V. Raghavan, S. V. Hanly, and V. V. Veeravalli. Statistical beamforming on the grassmann manifold for the two-user broadcast channel. *IEEE Transactions on Information Theory*, 59(10):6464–6489, Oct 2013.

[10] P. Raghavan. Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *Journal of Computer and System Sciences*, 37(2):130 – 143, 1988.

[11] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[12] K. Knight. *Mathematical Statistics*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 1999.

[13] C. Namislo. Analysis of mobile radio slotted aloha networks. *IEEE Journal on Selected Areas in Communications*, 2(4):583–588, July 1984.