Towards fundamental limits of bursty multi-user communications in wireless networks

Jean-Marie Gorce, Philippe Mary, Jean-Marc Kélif

TOWARDS FUNDAMENTAL LIMITS OF BURSTY MULTI-USER COMMUNICATIONS IN WIRELESS NETWORK

Jean-Marie Gorce[1] and Philippe Mary[2] and Jean-Marc Kélif[3]

[1] Univ Lyon, INSA Lyon, Inria, CITI, F-69621 Villeurbanne, France
[2] INSA de Rennes, IETR UMR CNRS 6164, Rennes, France
[3] Orange Labs, France
jean-marie.gorce@insa-lyon.fr

ABSTRACT

Considering an isolated wireless cell containing a high density of nodes, the fundamental limit can be defined as the maximal number of nodes the associate base station can serve under some system level constraints including maximal rate, reliability, latency and transmission power. This limit can be investigated in the downlink, modeled as a spatial continuum broadast channel (SCBC) as well as in the uplink modeled as a spatial continuum multiple access channel (SCMAC).

In this short paper, we summarize the different steps towards the characterization of this fundamental limit, considering four figures of merit: energy efficiency, spectral efficiency, latency, reliabilty.

1. INTRODUCTION

The recent evolution of the cellular market towards internet of things (IoT) connecting objects instead of humans induces significant changes in the mathematical modeling of wireless networks. Indead, the IoT paradigm relies on *bursty but massive* distributed transmissions, where billions of communicating objects will be spread over large radio cells, transmitting only few packets per day, per month or even per year. In such a scenario, the classical fundamental limit of communication systems derived since the seminal work of Claude E. Shannon [1] needs to be revised. As we will see, the capacity, or the capacity region in case of multi-user communications, becomes less important in regard to other metrics.

It is worth that Shannon's second theorem on capacity can be expressed as a fundamental tradeoff between energy efficiency (EE) and spectral efficiency (SE). For a real Gaussian channel the received power needs to verify $P \geq \left(2^{\eta_S/2}-1\right) N_0$, where η_s, the spectral efficiency, is the number of bits-per-channel-use and N_0 the receiver noise. The normalized energy efficiency is given by $\eta_E = \eta_S/(P/N_0)$. But this fundamental limit is achievable in the asymptotic regime, i.e. when the encoding time spread over an infinite number of channel uses. This is nothing but a steady-state analysis and relies on a set of assumptions, not further valid in the context of this study:

1. A unique source and a unique destination are considered.
2. The traffic is characterized by a fixed and continuous data rate.
3. Synchronization and transmitter identification costs are dropped off since the communication is assumed to be established.
4. The encoding time is infinite and so is the latency.

Information theory has been extended since its infancy to multi-user scenarios including the multiple access channel (MAC) or the broadcast channel (BC) well appropriate to study the uplink and the downlink, respectively. Most noticeable results with these schemes have been provided in the asymptotic regime, providing the expression of achievable rates region [2] characterizing achievable joint rates. For Gaussian channels, these capacity regions are known even for a large number of nodes and strategies based on superposition coding and successive decoding are capacity achieving [2].

But in the context of IoT cells with bursty traffic, these joint rates are not representative of the real scenarios that have to be supported. It comes out to be necessary to study these fundamental limits in the finite block length (FBL) regime where the information transmitted in the network is divided in finite size quantities that have to be transmitted under finite time transmission constraints. Yuri Polyanskiy et al [3] recently studied the fundamental limits of the point to point Gaussian channel in the FBL regime, paving the way to the study of latency and reliability constraints from a fundamental point of view. At the best of our knowledge the unique extension of this work for Gaussian channels has been done by Molavian-Jazi and Laneman [4] for the Gaussian MAC. We analysed the Gaussian BC in [5].

In addition to the FBL regime it is also necessary to study the behavior of BC/MAC models when the nodes are modeled through a spatial distribution as we proposed in [6, 7] to establish the fundamental EE-SE limit under equal rates condition in the asymptotic regime. We called these models SCMAC or SCBC, where SC stands for *spatial continuum*. This fundamental limit can be interpreted as the equivalent of the asymptotic Shannon capacity for a wireless cell.

If these results provided interesting insights, latency and reliability where kept off the study. Indeed, the asymptotic regime relies on error free but infinite latency transmission. So, the next question relates on the price of introducing a latency constraint. Such a constraint comes essentially with a reliability penalty, but in addition, the capacity region also shrinks.

We therefore anticipate three issues to be addressed in the framework of dense IoT networks. Firstly, the cost of the transient regime necessary to synchronize and detect which nodes are transmitting cannot be neglected. This corresponds to an additional information to be transmitted that should be considered in large scale low rate networks. Secondly, transmitting small packets over finite time slots leads necessarily to transmission errors. The minimal error can be modeled in information theory by working in the finite block length (FBL) regime, introducing an additional latency-reliability tradeoff.

2. SYSTEM MODEL

Let consider a unique BS, serving a cell area denoted by $\Omega \subset \mathbb{R}^2$. We denote by $(\Omega, \mathcal{A}, m)$ the corresponding measurable space with $\mathcal{A}$ the Lebesgue $\sigma-$ algebra and m the Lebesgue measure. Let x be a point in Ω.

Without lack of genericity, the BS is assumed to be located at point $(0, 0)$. Nodes appear randomly in time and space on Ω. As such, they are not described by a discrete set but through a spatial density $u(x)$ (the density of users per m^2 per time unit). Denoting by $\tilde{u}(x, t)$ a realization of the ergodic random process with probability density function (pdf) $u(x)$, then, for any subset $B \in \mathcal{A}$, the average number of users per time unit is given by

$$U(B) = \lim_{T \to \infty} \frac{1}{T} \int_{t \in T} \int_{x \in B} \tilde{u}(x, t) \cdot dx \cdot dt. \quad (1)$$

The global average number of nodes associated with the whole cell service area Ω is denoted by U_T.

Definition 2.1 (Requested rate density) *The requested rate density $\rho(x) : \Omega \to \mathbb{R}$ is a measurable function that represents the information rate spatial density requested at point x.*

- Note 1: $\rho(x)$ is normalized[1] by the system bandwidth and is expressed in bits per channel use (bpcu) per m^2.
- Note 2: The requested rates are either in downlink (SCBC) or uplink (SCMAC) modes.

The cell sum-rate per channel use is denoted the spectral efficiency (SE) of the cell:

$$\eta_s = \int_{\Omega} \rho(x) \cdot m(dx). \quad (2)$$

[1] $\rho(x)$ and related quantities $(R(.)...)$ are given in bit-per-channel-use (bpcu) throughout this paper. For practical applications, the results can be converted in bits or in bps, with appropriate system parameters.

The uniform rate condition corresponds to the case where each node requests the same quantity of information denoted by I_0, with

$$\rho(x) = \frac{I_0 \cdot u(x)}{N_{cu}}, \quad (3)$$

where N_{cu} is the number of channel uses per time unit, allocated to the system. In this special case, one have $\eta_s = \frac{I_0 \cdot U_T}{N_{cu}}$.

3. UNIFORM CAPACITY IN THE ASYMPTOTIC REGIME

The access capacity region is defined in [6] for the uplink and the downlink, as the set of rate spatial distributions for which a joint encoding-decoding scheme exists such that the transmission error tends to 0 when the encoding time tends to infinity (characterized by the number of channel uses n). Compared to the classical Shannon's approach, the asymptotic regime herein integrates another dimension: when $n \to \infty$, the number of nodes tends to infinity but individual rates tend to 0 since each packet is spread over an infinite time. Note that the cell's sum-rate tends to its average spectral efficiency. In [6], based on an iterative splitting process, the sum-rate that the cell can achieve when a continuum of users is considered has been obtained. The corresponding fundamental limit is expressed through the total transmisison power required to serve a given distribution and was found identical for the GSCBC and the GSCMAC with transferable powers [7]:

Theorem 3.1 (GSCBC and GSCMAC fundamental limit) *The achievable EE-SE tradeoff for a given rate spatial density $\rho(x)$ is given by*

$$\eta_e \leq \left[a \int_{\nu_m}^{\nu_M} t \cdot f_\nu(t) \cdot e^{a \cdot \eta_s \cdot G_\nu(t)} \cdot dt \right]^{-1}, \quad (4)$$

where

- $a = 2\log(2)$.
- *$f_\nu(t)$ and $G_\nu(t)$ are respectively the probability density function (pdf) and the complementary cumulated density function (ccdf) of the rate as a function of the equivalent noise (an exemple is represented in Fig.1).*

4. EXTENSIONS IN THE NON ASYMPTOTIC REGIME

The former result was derived in the asymptotic regime under no latency constraint. Modeling IoT networks requires additional constraints that will reduce the fundamentl limit.

4.1. Finite Time Transmission

The asymptotic regime used in Th. 3.1 means that all packets are transmitted simultaneously in the network to or from all nodes and spread over an infinite time. Let now be imposed a finite time constraint formalized as follows:

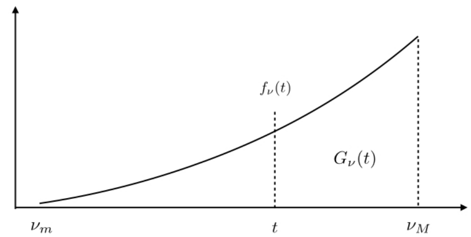

Figure 1. The rate request is distributed according to the density $f_\nu(.)$, where $\nu(x)$ for a node located in x is the equivalent noise power given by the ratio between the receiver noise and the channel gain $\nu(x) = N_0/g(x)$. Then, $G_\nu(t)$ is the fraction of the cell with an equivalent noise power greater or equal to t. This distribution for a regular circular cell is represented where the x axis is the equivalent noise and the y axis represents the rate density.

Property 4.1 *A multi-user network with packets of constant information quantities I_0, is said FTT constrained if each transmission is constrained to be done in at most n^* channel uses.*

The average spectral efficiency (i.e. the sum-rate) of the cell is noted $\bar{\eta}_s$ and is nothing but:

$$\bar{\eta}_s = \frac{I_0 \cdot U(\Omega)}{N_{cu}}. \tag{5}$$

In the downlink, when BS transmits packets of I_0 bits in n^* channel uses, individual rates are given by $\eta_u = \frac{I_0}{n^*}$. The BS can achieve the target spectral efficiency by using superpostion coding (SC) to simultaneously transmit several packets. Therefore, optimizing the transmission strategy relies on optimizing the scheduling of the different packets in order to minimize the total power.

Let us define a frame of length N_{cu}, made of L slots, each slot $s_l; \forall l \in \{1, 2, \ldots, L\}$ being made of n^* channel uses. For consistency, all time units are counted in channel uses. In the downlink, the BS has in its queue a random number of packets to be transmitted to a set of nodes $\mathcal{N}_{U,L}$, with $\lim_{L\to\infty} \frac{|\mathcal{N}_{U,L}|}{L \cdot n^*} = U(\Omega)$.

In each slot s_l, a subset of nodes $\mathcal{N}_u(l) \subset \mathcal{N}_{U,L}$ is selected and the corresponding packets are simultaneously transmitted. Let $N_{u,l} = |\mathcal{N}_{u,l}|$ be the number of nodes allocated to the slot s_l, the spectral efficiency is :

$$\eta_s(l) = N_{u,l} \cdot \eta_u. \tag{6}$$

To achieve the desired rate $\eta_u = I_0/n^*$, the SINR at each decoding step has to verify $\gamma^* \geq e^{a\eta_u} - 1$ according to the Shannon channel capacity. To be optimal, the power allocation is computed such that each node gets a SINR exactly equal to γ^*. Using a superposition coding strategy, each node decodes its signal after having decoded all greater signals.

Under these conditions of equal rate, the minimal power associated to each slot is obtained as:

Theorem 4.1 (Sum-power in the downlink) *For a given set of users indexed by $\{1, \ldots, N_{u,l}\}$, the minimal sum-power necessary to transmit reliably to each node independent informations of size I_0 in n^* channel uses, is given by:*

$$P_m(l) = \gamma^* \sum_{\bar{k}=1}^{N_{u,l}} \nu_{\bar{k},l} \cdot (1+\gamma^*)^{\bar{k}-1} \tag{7}$$

with $\gamma^* = e^{aI_0/n^*} - 1$, and $\bar{k} = N_{u,l} - k + 1$.

Therefore, comparing numerically Th. 4.1 to Th. 3.1 allows to evaluate the performance loss due to the FTT constraint.

4.2. Finite Block Length Coding

In the former section, we assumed n^* to be large enough to have error-free transmission. This is not the case in a band limited system.

Strassen [8] studied the finite blocklength regime and has shown for a Discrete Memoryless Channel (DMC) that:

$$\log M^*(n, \epsilon) = nC - Q^{-1}(\epsilon)\sqrt{nV} + O(\log n) \tag{8}$$

where n, ϵ and M are the blocklength, the error probability and the codebook size, respectively. The second term in the right hand side of (8) is the rate penalty, namely the back-off from the channel capacity. This expression gives the fundamental trade-off between latency and reliability for short packet transmissions. Unfortunately, Strassen's result was not compliant to be generalized, particularly to the AWGN channel, generally to any type of channel with input constraints. Almost half a century later, Polyanskiy *et al.* [3] has generalized Strassen's approach, i.e. treating the mutual information density as a random variable and using second order statistics to approximate the channel coding rates in the FBL regime, for various channel models such as the AWGN channel, for which the following holds:

Theorem 4.2 *(Polyanskiy et al., 2010) For the AWGN channel with signal-to-noise ratio (SNR) γ, for $\epsilon \in (0, 1)$ and for equal, maximal and average-power constraints, the following holds,*

$$\log M^*(n, \epsilon, \gamma) = nC(\gamma) - Q^{-1}(\epsilon)\sqrt{nV(\gamma)} + O(\log n) \tag{9}$$

where $C(\gamma)$ is the Shannon capacity defined as

$$C(\gamma) = \frac{1}{2}\log(1+\gamma) \tag{10}$$

and

$$V(\gamma) = \frac{\gamma}{2}\frac{\gamma+2}{(\gamma+1)^2}\log^2 \mathrm{e}. \tag{11}$$

However, the extension of this result to the case of multi-user channels is not straightforward. The exact capacity regions of Gaussian MAC or Gaussian BC are not known but achievable regions for the 2-user BC or 2-user MAC have been characterized recently in [4, 5]. As a first approximation, the use of a superposition coding strategy [5] allows to extrapolate the FBL expression in Th. 4.2 to compute a modified SNR threshold γ^* as a function of some latency and reliability constraints. This γ^* can be used in Th. 4.1 to modify the fundamental limit.

4.3. Additional real time constraints

In the former paragraph, we detailed how to combine the dense nodes distribution assumption and a finite time transmission constraint. To model real-time transmissions, where the nodes have to transmit or receive instantaneously, additional issues have to be considered. In the downlink, the main challenge is to indicate to the nodes if they have something to decode and how it is encoded. Clearly, signaling cannot be ignored and is even a part of the information to be transmitted. In the uplink, the randomness of the active nodes distribution and their identification is necessary. Stochastic geometry coupled with FBL information theory can provide insights about the impact of randomness [9]. From a protocol point of view, integrating data transmission and signaling as a whole, approaches based on compressive sensing [10, 11] are appealing although their optimality is not granted.

5. CONCLUSION

In this paper, we paved the way to derive fundamental limits of IoT cells where dense distributions of users are spread over large cells and where a sporadic traffic has to be supported. Starting from the EE-SE fundamental limit in the asymptotic regime, we defined the different steps towards the derivation of the EE-SE loss in the FBL regime and also how to balance it with respect to the reliability-latency trade-off.

6. ACKNOWLEDGMENTS

This work has been partly supported by Orange Labs under CRE $nF05151$, and by the French National Agency for Research (ANR) under project Arburst.

7. REFERENCES

[1] CE Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 1948.

[2] Thomas M Cover and Joy A Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[3] Y. Polyanskiy, H.V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, pp. 2307–2359, December 2010.

[4] E. MolavianJazi and J.N. Laneman, "A Second-Order Achievable Rate Region for Gaussian Multi-Access Channels via a Central Limit Theorem for Functions," *IEEE Transactions on Information Theory*, vol. 61, pp. 6719–6733, December 2015.

[5] Ayse Unsal and Jean-Marie Gorce, "The Dispersion of Superposition Coding for Gaussian Broadcast Channels," in *IEEE Information Theory Workshop 2017*, Kaohsiung, Taiwan, Nov. 2017.

[6] Jean-Marie Gorce, H Vincent Poor, and Jean-Marc Kelif, "Spatial continuum model: Toward the fundamental limits of dense wireless networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2016.

[7] Jean-Marie Gorce, Yasser Fadlallah, Jean-Marc Kelif, H Vincent Poor, and Azeddine Gati, "Fundamental limits of a dense iot cell in the uplink," in *15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2017.

[8] V. Strassen, "Asymptotische Abschaetzungen in Shannon's Informationstheorie," in *Transactions of the Third Prague Conference on Information Theory*, 1962, pp. 689–723.

[9] Harpreet S Dhillon, Howard Huang, Harish Viswanathan, and Reinaldo A Valenzuela, "Fundamentals of throughput maximization with random arrivals for m2m communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4094–4109, 2014.

[10] Alyson K Fletcher, Sundeep Rangan, and Vivek K Goyal, "Necessary and Sufficient Conditions for Sparsity Pattern Recovery," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5758–5772, dec 2009.

[11] Ronggui Xie, Huarui Yin, Xiaohui Chen, and Zhengdao Wang, "Many Access for Small Packets Based on Precoding and Sparsity-Aware Recovery," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4680–4694, nov 2016.