Asymptotic Throughput Analysis of Massive M2M Access

Andrea Zanella, Andrea Biral, Michele Zorzi
Department of Information Engineering, University of Padova, via Gradenigo 6/B, Padova, 35131, ITALY
e-mail: {zanella, biraland, zorzi}@dei.unipd.it.

Abstract—5G systems are expected to be able to handle channel access from a massive number of low cost machine-type devices (MTDs), requiring intermittent connectivity to a network. Ideally, these devices should be able to transmit their short message either immediately without prior connection establishment (random access) or with a lightweight connection establishment (access reservation). In this paper, we analyze the capacity of a system where a large number of devices transmit simultaneously to a single receiver, capable of performing multiple packet reception (MPR) by means of advanced decoding techniques, such as successive interference cancellation (SIC). We derive a simplified mathematical model that allows us to determine the average number of signals that can be successfully decoded by the receiver as a function of the number of overlapping transmissions and the MPR capabilities of the system. We observe that, according to our analysis, a receiver capable of performing perfect SIC and MPR can theoretically decode an arbitrarily large number of simultaneous transmissions by proportionally reducing the per-user data rate, in such a way that the aggregate system capacity remains almost constant.

I. INTRODUCTION

Machine-to-Machine (M2M) communications are expected to play a major role in the transition from the current cellular standard to next generation networks. Indeed, while the ad hoc reengineering of existing systems (GSM, UMTS, LTE) recently carried forward by 3GPP and some research groups [1]–[3] may initially satisfy the rising demand of wireless access from Machine-Type Devices (MTDs), the Internet of Things revolution will likely require new solutions designed from scratch, in order to become a reality. Today’s radio systems, in fact, have been originally designed to support relatively few connections with high data rates. Thus, they are intrinsically unsuitable for M2M communications, which are conversely characterized by sporadic transmissions of small data packets from possibly a very large number of devices, with heterogeneous QoS specifications.

This vision motivates the study of new cellular protocols and network architectures natively designed to support both M2M applications and conventional Human-to-Human services. However, while the fundamental limits for broadband systems are generally well understood, a similar level of understanding for M2M systems is still lacking. Specifically, the different nature of M2M traffic with respect to conventional UE transmissions calls for novel solutions, whose design, in turn, requires a deep and solid comprehension of the fundamental properties and characteristics of the M2M communication paradigm. In this respect, studies that adopt a clean-slate and standard-agnostic approach can provide insights on the basic aspects of M2M communication, thus shedding light on the fundamental performance limits of these types of systems.

As a matter of fact, the centrally coordinated access mechanism, adopted by most of today’s wireless cellular systems, presents severe limitations in the M2M context. Indeed, with machine-type applications typically generating payloads of less than 1 kB, the signaling commonly used to set up and maintain mobile broadband data connections may become the bottleneck of the system, substantially limiting the performance of both M2M and conventional services. On the other hand, uncoordinated schemes have been proved to be able to asymptotically reach the efficiency of coordinated access protocols, without the burden of the signaling overhead associated to the latter. Some such evidence is given in [4], where the performance of coordinated and uncoordinated transmission strategies for multiple access is analyzed. That work shows that, for the typical length of M2M data payloads, uncoordinated access schemes support more devices than coordinated access mechanisms, because of the lower signaling overhead.

In the context of random access schemes, a promising solution to increase the capacity of wireless networks based on uncoordinated access schemes consists in providing Multi-Packet Reception (MPR) capabilities to the receiver, which will hence be able to decode multiple overlapping packets by using different techniques [5], [6]. A possible method to enable MPR without excessively increasing the complexity of the end devices consists in using pseudo-orthogonal spreading codes to decrease the power spectral density of the transmission and, hence, limit the interference at the receiver. This is particularly effective if combined with Successive Interference Cancellation (SIC) at the receiver, which allows the receiver to (partially) remove the contribution of all decoded signals from the compound received signal, thus possibly enabling the successive decoding of other overlapping transmissions. The combination of MPR and SIC can hence enable the correct decoding of a large number of transmissions that overlap in time and frequency, thus dramatically increasing the probability of successful channel access in random access wireless systems [7]. The downside is that the spreading techniques enlarge the bandwidth required to transmit the signal which, for a fixed system bandwidth, decreases the effective bit rate that can
be offered to the higher layers, while on the other hand SIC requires rather sophisticated receivers.

Nonetheless, the combined use of MPR and SIC appears particularly appealing in the context of M2M, where the aim is to guarantee high channel access probability to many nodes that have low transmit rates [8]. In this paper we investigate the limiting performance of such a scenario when the population of transmitters grows, while the per-user bitrate decreases. By building upon the model developed in [7], we find an approximate model to estimate the maximum throughput of a wireless cellular system, where the receiver is capable of performing MPR and perfect SIC. The throughput here is defined as the average number of transmissions that can be successfully decoded per packet duration. The theoretical results provided by the model are validated against simulation results, which show that the model tends to slightly overestimate the actual system throughput, but is able to capture the general trend when changing the scenario, i.e., the statistical distribution of the received signal powers and the MPR capabilities of the receiver. Furthermore, the model shows that the throughput tends to grow linearly with the MPR capabilities of the system that, in turn, are inversely proportional to the per-user bitrate. This tradeoff implies that the product of throughput and bitrate asymptotically tends to a constant which depends on the statistical distribution of the signal powers at the receiver.

The remainder of this paper is organized as follows. Sec. II overviews the prior work related to this subject. In Sec. III the asymptotic analysis of an MPR system with SIC capabilities is presented. Then, theoretical and simulation results for such a system are illustrated in Sec. IV. Finally, Sec. V draws our conclusions.

II. RELATED WORK

As mentioned, the introduction of uncoordinated access schemes with MPR capabilities at the receiver may significantly impact the performance in M2M scenarios since it allows the detection of multiple overlapping signals in the event of a collision, thus making it possible to increase the system throughput even when transmitters are not centrally coordinated. The relevance of such a signal-capture phenomenon in mobile radio systems has been recognized since long, producing a quite rich literature. A first attempt at the analysis of MPR-capable wireless systems was made in [9], where the stability of the ALOHA system is studied under infinite-user single-buffer assumptions. The paper, however, models the number of transmissions successfully decoded in a slot as a random variable that only depends on the number of overlapping signals, without considering the signal power distribution. Another simplistic capture model is considered in [10], where the authors assume that all signals are decoded if the sum rate of all the transmitting users does not exceed the capacity of the channel, whereas in [11] a signal is captured whenever the strongest interferer is sufficiently far apart from the target receiver, according to a statistical geometry approach.

Successively, the analysis of the capture effect was extended and made more realistic by including in the capture model basic physical propagation phenomena and the cumulative character of interference, considering the random distribution of the signal powers at the receiver, and introducing a capture criterion based on the Signal-to-Interference-plus-Noise-Ratio (SINR) of the signals [7], [12], [13].

On the other hand, the use of interference cancellation to improve the channel access probability has been investigated in [14], where the authors present an enhanced version of Slotted ALOHA, called Frameless ALOHA, that exploits SIC to increase the number of packets that can be successfully transmitted over a single slot. This protocol takes inspiration from the rate-less code paradigm: users transmit replicas of their packets in the slots of a frame, whose length is changed at each contention round according to the outcomes of the previous round. The transmission slots are picked at random, according to a Bernoulli process of given probability. Every packet includes pointers to all its replicas so that, upon decoding one packet, the receiver can cancel the corresponding signal from the compound signals received in all the slots where a replica of the packet was sent. This can lead to the resolution of new packets, thus increasing the system throughput. The effectiveness of this approach is confirmed by the results provided in [14], where the protocol is shown to achieve an asymptotic efficiency around 87%, much larger than the classical Slotted ALOHA performance.

Recently, interference cancellation was also applied to cognitive radio networking. In [15], the authors investigate a scenario where primary (licensed) users and secondary (unlicensed) users share the same spectrum in order to communicate with their respective receivers. Primary users are assumed to implement a retransmission-based error control technique, which implies the retransmission of a copy of the failed packets over subsequent time slots. This behavior gives rise to redundancy in the interference generated by the primary user to the secondary user, which can be exploited to design secondary transmission policies. The basic idea is that, if a secondary receiver detects the primary user message in a given initial transmission, then it can use this knowledge to cancel the primary interference in the subsequent slots in case of primary retransmissions, thus achieving a larger secondary throughput. Indeed, this result is confirmed by the comparison between the performance of the above-mentioned technique and those of a couple of suboptimal policies that do not exploit interference cancellation.

In this paper, we focus on the maximum number of transmissions that can be successfully decoded in a single slot, without considering the possibility of exploiting retransmissions in successive slots to further increase the performance. Instead, the analysis takes into account the MPR capabilities of the receiver, and the effect of signal power distribution.

III. ASYMPTOTIC ANALYSIS OF MASSIVE ACCESS CAPACITY

As mentioned, the model presented in this paper builds upon the mathematical framework developed in [7], which provides the expression of the probability mass distribution of the number \( r \) of signals that can be successfully decoded
when \( n \) transmissions overlap in time and frequency at the receiver. The reference model is reported in the following, for the reader’s convenience.

Denoting with \( P_i \) the power of the \( i \)th signal at the receiver, the SINR is given by

\[
\gamma_i = \frac{P_i}{\sum_{j \neq i} P_j + N_0}
\]

(1)

where \( N_0 \) is the background noise power, which will be henceforth neglected for simplicity. Signal \( i \) is assumed to be successfully decoded if

\[
\gamma_i > b,
\]

(2)

where \( b > 0 \) is the so-called capture threshold of the system. Note that, without SIC, the maximum number of signals that can be successfully decoded is limited to \( \lceil \frac{1}{b} \rceil \). Therefore, MPR capabilities require \( b < 1 \) or SIC, or both.

We consider a scenario where \( n \) terminals (machine type devices) simultaneously transmit packets of equal size to a common receiver, the Base Station (BS), which is capable of MPR and perfect SIC. More specifically, we assume that signal decoding at the BS is an iterative process. At each iteration, all signals that satisfy the capture condition (2) will be successfully decoded. Hence, the corresponding contributions can be cancelled out of the compound received signal. The entire decoding procedure is then repeated on the remaining signal, until all overlapping transmissions are successfully detected, or no signal satisfies (2).

The received signal powers at the BS are modelled as independent and identically distributed (iid) random variables \( \{ P_j, j = 1, \ldots, n \} \), with common Cumulative Distribution Function (CDF) \( F(x) \), \( x \geq 0 \). The number of correctly decoded signals at the first iteration can be expressed as [7]:

\[
\rho = \sum_{j=1}^{n} \chi \{ \gamma_j > b \}
\]

(3)

where \( \chi \{ A \} = 1 \) if condition \( A \) holds true and 0 otherwise. Taking the expectations of both sides, we hence get

\[
r = n \Pr [ P_j > I_0 ]
\]

(4)

where \( I_0 = \sum_{i \neq j} P_i b \). Eq (4) is the average number of signals that are captured by the BS before performing SIC. Now, in [7], a simple recursive method is proposed to estimate the mean number of signals that can be decoded at each successive SIC iteration. Being the basis of the analysis developed in this paper, we here recollect the recursive method, considering only the case with zero-residual interference after cancellation (i.e., assuming \( z = 0 \) according to the notation used in [7]).

A. Recursive throughput expression with SIC

First of all, (4) is simplified by replacing \( I_0 \) with its first order approximation \( I_0 = (n - 1)b E \{ P_j \} \), which gives:

\[
\tilde{r} = n \Pr [ P_j > (n - 1)b E \{ P_j \}] .
\]

(5)

We then denote as \( \tilde{r}_h \) the estimation of the mean number of signals that are captured at the \( h \)th iteration, where \( h = 0, 1, \ldots \). In particular, from (5), we get:

\[
\tilde{r}_0 = n \Pr [ P_j > I_0 ] .
\]

(6)

These signals are then removed from the set of \( n \) overlapping signals, so that after the first SIC operation, the remaining signals are (approximately):

\[
n_1 = n - \tilde{r}_0 = n \Pr [ P_j \leq I_0 ] = n F(I_0) .
\]

(7)

Note that all these signals have power less than or equal to \( I_0 \), otherwise they would have been decoded at the previous iteration.

Repeating this reasoning, the average number of signals decoded at this new iteration can be approximated as:

\[
\tilde{r}_1 = n_1 \Pr [ P_j > I_1 | P_j \leq I_0 ] ,
\]

(8)

where \( I_1 = b(n_1 - 1)E \{ P_j | P_j \leq I_0 \} \) is the approximate expression of the aggregate interference power of \( n_1 - 1 \) signals, scaled by the capture threshold \( b \), and \( E \{ x \mid y \} \) denotes the conditional expectation of \( x \) given \( y \).

After the generic \( h \)th SIC iteration, we then have

\[
n_h = n - \sum_{j=0}^{h-1} \tilde{r}_j ; \quad I_h = b(n_h - 1)E \{ P \mid P \leq I_{h-1} \} ;
\]

(9)

Finally, the approximate normalized throughput after \( K \) SIC iterations is given by

\[
\tilde{S}_n^{(c)}(K) = \sum_{j=0}^{K} \tilde{r}_j .
\]

(10)

B. Fixed point throughput expression with unlimited SIC

Starting from the recursive expression in (9), we here derive an approximate expression for the system throughput when the number of allowed SIC iterations is unlimited, i.e., \( K = \infty \). In other words, we suppose that the decoding process stops only when no further signal satisfies the capture condition.

We rewrite \( n_2 \) as follows

\[
n_2 = n_1 - \tilde{r}_1
\]

\[
= n_1 \Pr \{ P_j \leq I_0 | P_j \leq I_0 \}
\]

\[
= n \Pr \{ P_j \leq I_0 \} \Pr \{ P_j \leq I_1 | P_j \leq I_0 \}
\]

\[
= n \Pr \{ P_j \leq \min \{ I_0, I_1 \} \}
\]

\[
= n \Pr \{ P_j \leq I_1 \}
\]

\[
= n F(I_1) .
\]

(11)

where the second and third rows are obtained using (8) and (7), respectively, while the second-to-last row follows from the fact that \( I_k \) is always non-increasing in \( k \). Repeating iteratively, we get that the average number of signals that remain to

\[\footnote{A preliminary formulation of this expression was presented in [8].}\]
be decoded after $K$ SIC cycles can be estimated using the following recursive expressions

$$n_K = n F(I_{K-1}), \quad I_K = (n_K - 1) b E[P|P \leq I_{K-1}], \quad (12)$$

for $K = 1, 2, \ldots$, where

$$E[P|P \leq I_{K-1}] = \frac{\int_{0}^{I_{K-1}} x f(x) dx}{F(I_{K-1})}$$

is the average power of any still undecoded signal, being $f(x)$ the (unconditional) probability density function of the signal received powers.

The system throughput (10) after $K$ SIC iterations, hence, can be expressed as the difference between the number $n$ of overlapping transmissions and the number of residual signals after $K$ SIC iterations, i.e.,

$$S_n(K) = n - n_K. \quad (13)$$

Letting $K$ grow to infinity, the maximum achievable throughput can thus be approximated as

$$S(n) = n - \lim_{K \to \infty} n_K = n(1 - F(I_\infty(n))) \quad (14)$$

where $I_\infty(n)$ is equal to the fixed-point of (12), i.e.,

$$I_\infty(n) = (n F(I_\infty(n)) - 1) b E[P|P \leq I_\infty(n)], \quad (15)$$

provided it exists, and $I_\infty(n) = 0$ otherwise.

C. Optimal number of simultaneous transmissions

Clearly, (14) linearly increases with $n$ as long as (15) does not admit any other solution than $I_\infty(n) = 0$. In this respect, a simple functional analysis reveals that (15) always admits positive solutions when $n$ exceeds a certain threshold. For values of $n$ larger than such a threshold, the throughput given by (14) starts decreasing. As a proof of concept, we plot in solid lines in Fig. 1 the throughput estimate given by (13) for different values of $K$, as reported in the legend, and the asymptotic value given by (14) for $K = \infty$, when varying $n$. The marks in the figure, instead, correspond to the simulation results in the same conditions. These results have been obtained by considering the Rayleigh Fading (RF) scenario defined in [7], where the received signal powers are iid exponential random variables with unit mean. Furthermore, the capture threshold has been set to $b = 0.02$.

From the figure, we observe that the throughput approximation is excellent for relatively small values of $K$, while it tends to overestimate the actual performance when $K \to \infty$. More importantly, we note that the throughput equals the number of simultaneous transmissions $n$ up to a critical value $n^*$, beyond which the performance decreases sharply. However, according to (14), $S_n(K) < n$ only if $F(I_\infty(n)) > 0$ and, hence, $I_\infty(n) > 0$. Therefore, the critical value $n^*$ corresponds to the maximum $n$ for which the fixed-point equation (12) does not admit any positive solution. To find such a value, we can express $n$ as a function of the unknown $x = I_\infty(n)$, so that the fixed-point equation (15) yields

$$n(x) = \left(\frac{x}{b E[P|P \leq x]} + 1\right) \frac{1}{F(x)}. \quad (16)$$

Now, it is easy to realize that the right-hand side of (16) is continuous and positive, grows to infinity for $x \to 0$ and for $x \to \infty$, and therefore has a minimum for some positive $x^*$. As an example, in Fig. 2 we report the graph of $n(x)$ when varying $x$ for the RF scenario, for which we get

$$n(x) = \left(\frac{x(1 - e^{-x})}{b(1 - e^{-x}(x + 1))} + 1\right) \frac{1}{1 - e^{-x}}. \quad (17)$$

To find the global minimum, we set to zero the derivative in $x$ of (16). After some algebra, we get

$$F(x) E[P|P \leq x] = f(x) \left(x^2 + b E[P|P \leq x]^2\right) \quad (18)$$

whose solution(s) can be found using numerical methods.2

Replacing $x^*$ into (16), and rounding down the result, we finally get the critical value $n^*$ of $n$ after which the approximate throughput starts decreasing. As a matter of fact, it is easy to prove that $S(n^*+1) < S(n^*)$. In fact, being $n^*$ the maximum

2In general, (18) may admit multiple solutions. In this case, $x^*$ is the solution that minimizes $n(x^*)$ as given by (17)
value of \( n \) for which (15) does not admit positive solutions, we have that the approximate throughput will equal \( n \) for any \( n \leq n^* \), in particular \( S(n^*) = n^* \). Therefore, we need to prove that

\[
S(n^* + 1) < n^*.
\]

Using (14) in (19), we get

\[
(n^* + 1)(1 - F(x(n^* + 1))) < n^* \rightarrow (n^* + 1)F(x^*) > 1,
\]

where in the rightmost inequality we approximate \( x(n^* + 1) \) with \( x^* \). Now, using (15), we get

\[
((n^* + 1)F(x^*) - 1)b \mathbb{E}[P|P \leq x^*] = x^* > 0,
\]

from which we get \( (n^* + 1)F(x^*) > 1 \), as required by (20).

Summing up, the approximate throughput (14) is maximized when the number of simultaneous transmissions equals the critical threshold \( n^* \), which depends on the capture threshold \( b \) of the receiver, and on the CDF \( F(x) \) of the received signal powers. Hence, given the statistical distribution of the signals, the maximum achievable throughput can be approximated as

\[
S^*(b) = \max_n S(n^*(b)) \simeq n^*(b),
\]

where \( n^*(b) \) is the critical value of \( n \) for a given \( b \).

### IV. Results

In this section we exploit the results derived in the previous section to gain some insight on the optimal performance of a random access system with a receiver capable of performing MPR of all signals with SNR above the threshold \( b \), and SIC, in an iterative manner. More specifically, we are interested in investigating the asymptotic performance of the throughput as the number of SIC cycles \( K \) increases.

First of all, we report in the top part of Fig. 3 the optimal throughput when varying the capture threshold \( b \) in the RF scenario. The approximation given by (21) is shown in blue solid line with markers, while the red line represents simulation results. We can note that the optimal throughput is approximately equal to \( S^*(b) = \alpha/b \), where the coefficient \( \alpha \) depends on the statistical distribution of the received signal powers. Indeed, although not shown in the figure to reduce clutter, the results obtained for the other two scenarios defined in [7], i.e., pure path loss and shadowing channels, are essentially the same, except for a small difference in the coefficient \( \alpha \). Furthermore, we observe that the approximation is slightly optimistic, in particular for larger values of \( b \). This mismatch is likely due to the progressive worsening of the approximation \( I_h \) of the residual interference as the number of SIC cycles increases. Indeed, it is easy to realize that, given \( k \) residual signals sorted in increasing order of received power, and if \( a \) is the power of the strongest such signal, i.e., \( P_k = a \), the iterative decoding process stops if and only if

\[
\Pr \left[ a \leq b \sum_{j \neq k} P_j | P_k \leq a, \forall j \right] = \Pr \left[ \sum_{j \neq k} P_j \geq \frac{a}{b(k-1)} | P_k \leq a, \forall j \right] \approx Q \left( \frac{a}{b\sqrt{k-1}} - \frac{1}{\sqrt{k-1}} \mathbb{E}[P|P \leq a] \right) \tag{22}
\]

where in the last step we resorted to the central limit theorem to approximate the mean of \( k-1 \) iid random variables as a normal random variable with mean \( \mathbb{E}[P|P \leq a] \) and standard deviation \( \text{STD}(P|P \leq a) / \sqrt{k-1} \). Now, it is apparent that, for large values of \( k \), the probability (22) grows to 1 with a step that depends on the conditional standard deviation \( \text{STD}(P|P \leq a) \): the smaller the range over which the signal powers are distributed, the higher the probability that the decoding process stops with \( k \) residual signals. Now, after each SIC cycle, the powers of the remaining signals are compacted in a smaller interval, so that \( \text{STD}(P|P \leq a) \) progressively decreases. Furthermore, the larger \( b \) the sooner (22) starts growing with \( k \). This aspect is not captured by the recursive model (9) that, hence, yields an upper bound to the number of signals captured. Actually, the approximation can be improved by taking into account the model (22), though this approach is not further investigated in this paper.

Although not very accurate when the number of SIC cycles is
Therefore, the coefficient $\alpha$, interpreted as a scaling factor of the bit rate of each terminal.

However, achieving the optimal performance requires a quite exploited even using a slotted random access mechanism. Indeed, we showed that, with ideal SIC, the MPR capabilities of the receiver can be asymptotically approaches the value returned by simulations. Furthermore, we note that the aggregate system capacity remains almost constant when varying $b$, which means that the random access scheme with MPR and SIC can accommodate a variable number of terminals, provided that the transmit rates are proportionally scaled.

V. Conclusions

In this paper, we proposed an approximate analysis of the asymptotic throughput of a wireless channel in the presence of massive access from low-rate terminals, with disparate signal powers. The mathematical model, though overoptimistic in its estimate of the system performance, is capable of capturing some fundamental properties of the system. Indeed, we showed that, with ideal SIC, the MPR capabilities of the receiver can be fully exploited even using a slotted random access mechanisms. However, achieving the optimal performance requires a quite accurate control of the total number of transmitters, since the throughput rapidly decreases if the critical threshold $n^*(b)$ is exceeded.

As future work, we plan to improve the accuracy of the mathematical model and to relax some simplifying assumptions, such as ideal SIC. Furthermore, we plan to include protocol aspects into the model such as, for instance, mechanisms to limit the number of simultaneous transmitters, and techniques to exploit transmissions in multiple slots. Finally, energy aspects, which are very sensitive in M2M scenarios, will also be considered.

References