An Analytical Model for Assessing the Performance of NB-IoT

Romain Barbau[‡§], Vincent Deslandes[‡], Gentian Jakllari[§], Jérôme Tronc[‡] and André-Luc Beylot[§]
[‡]Airbus Defence and Space - France
surname.name@airbus.com
[§]IRIT-INPT/ENSEEIHT - France
surname.name@toulouse-inp.fr

***Abstract*—As part of its ongoing effort to standardize 5G and become a key enabler of the Internet of Things, 3GPP introduced NB-IoT – an LPWAN access technology for massive Machine Type Communications (mMTC) – in Release 13. Rooted in LTE, NB-IoT introduces several innovation aimed at meeting the 5G IoT requirements, especially in terms of capacity. Already a commercial success with 94 deployments in over 35 countries, the performance of NB-IoT remains, however, poorly understood.**

In this paper, we introduce a theoretical model based on M/D/1-PS queues for assessing the performance of NB-IoT as function of different network settings and protocol configurations. Our model is the first to accurately capture the random access procedure for connection establishment (the *contention phase*) and the entire communicate process, with signalling, connection release and uplink, downlink messages (the *congestion phase*).

After using simulations to demonstrate its accuracy, we use our model to assess the performance of NB-IoT as function of its key parameters. Our analysis reveals a) the NPRACH/NPUSCH ratio has to be chosen carefully and take into account the traffic distribution, and b) for small packet sizes, the NB-IoT signalling becomes a significant bottleneck to the number of UEs that can be supported.

I. Introduction

Decades of advances in the semi-conductor industry, following Moore's law, have lead to miniaturized devices – looking more like things – with significant computing, sensing, and communication capabilities. Connecting all these things that can interact with the physical world into the cyberspace – the Internet of Things (IoT) – has the potential to disrupt almost all aspects of our society – from agriculture [1], [2], to healthcare [3], to manufacturing [4] and beyond.

In light of these developments, the 3rd Generation Partnership Project (3GPP) standardized Narrowband IoT (NB-IoT), a Low Power Wide Area Network (LPWAN) access technology for massive Machine Type Communications (mMTC), in Release 13 [5]. It introduces several innovations aimed at better meeting the 5G mMTC requirements in terms of coverage, energy efficiency, cost and number of devices supported.

NB-IoT is already a commercial success, with 94 deployments (2.5x as many as LTE-M) in over 35 countries [6]. Nevertheless, its performance is not fully understood, with very few analytical studies [7], [8], [9] and very scarce real-world evaluations [10]. A mathematical model is introduced in [8] to predict the maximum performance (probability of success) for particular sets of configuration parameters. However, the model assumes no retransmissions. A Semi-Markov model is introduced in [9] that takes into account the collisions in the random access channel (NPRACH), but not the failures due to congestion in the data channels. The coverage and capacity of NB-IoT in a rural environment are analyzed in [10] using the configuration and location of commercially deployed LTE sites. However, the specific deployment did not include a scheduler adapted to NB-IoT for single-tone transmissions, limiting the generality of the conclusions.

In this paper, we introduce the first theoretical model that fully captures the functionality of NB-IoT and can evaluate its performance in terms of capacity – the key metric of mMTC. The model is highly flexible, capable of taking into account 1) a *scenario*: message size, coverage condition (average code rate), traffic type, cell size, and 2) a *configuration*: the setting of uplink/downlink frames, use of user plane optimization, repetitions, subcarriers bandwidth, light signalling and the number of re-transmissions. To achieve this, we introduce a queuing model comprising two parts operating in tandem: a) the **Contention Phase** – it models an UE establishing a connection to the eNB using the random access channel – the Narrowband Physical Uplink Shared CHannel (NPRACH); b) the **Congestion Phase** – it models an UE, upon its establishing a connection to the eNB, being allocated resources on the dowlink control channel, the Narrowband Dowlink Control CHannel (NPDCCH), and transmitting to/receiving from the eNB on the Narrowband Physical Uplink Shared CHannel (NPUSCH) and the Narrowband Physical Downlink Shared CHannel (NPDSCH), respectively.

While modeling the random access channel is a well-studied problem, the congestion phase raises several challenges, some unique to NB-IoT. First, the number of UE-to-eNB connections that can be established during NPRACH is not necessarily constrained by the available capacity in the NPDCCH / NPUSCH / NPDSCH queues. As a result, congestion can occur, leading to failures even if NPDCCH / NPUSCH / NPDSCH are contention-free. A similar situation exists in LTE, modeled using queues with impatient customers but assuming exponential service times [7]. However, the service time in NB-IoT is constant, depending only on the payload size and the data rate. Second, unlike LTE, in NB-IoT the eNB shares the Resource Block (180 kHz bandwidth) among the UEs – thus, all communications are served in parallel. As a result, the well-established FIFO policy, used to model LTE [7], does not apply here. To better capture the functionality of NB-IoT, we model the congestion phase using an M/D/1-PS queue. It uses constant service times, relaxing the exponential distribution assumption, and process sharing (PS) for the

NPDCCH / NPUSCH / NPDSCH queues. Unfortunately, no theoretical result exists regarding the impatience probability of an M/D/1-PS queue. In short, we address this challenge by approximating the impatience probability by the probability of a communication being served after the impatience time. We show that an approximate solution of a more accurate model, M/D/1-PS, better captures the functionality of NB-IoT than the exact solution of an M/M/1 model.

Our main contributions may be summarized as follows:

- We introduce the first theoretical model that comprises two parts operating in tandem the contention and congestion phases of NB-IoT (§ III). The model is flexible and can evaluate the performance in terms of capacity as a function of the network settings and protocol configuration.
- We develop a simulator that implements the details of NB-IoT as defined in Release 15, available to the community as open-source [11], and use it to demonstrate the accuracy of our analytical model (§ IV-A). The data validates our choice of modeling NB-IoT exactly, even if the mathematical solution is approximate, rather than using an approximate model with an exact mathematical solution.
- We use our model to assess the impact of the key parameters of NB-IoT on its performance in terms of capacity (§ IV-B), as defined in the 5G IoT requirements [12]. Our analysis reveals that a) the ratio between the contention and the communication phases needs to be adapted to the traffic characteristics to avoid a significant drop in performance, and b) for small packet sizes, the kind generated by many IoT applications [13], the NB-IoT signalling becomes a significant bottleneck to the number of UEs that can be supported.

II. NB-IoT Communication

NB-IoT is organized in carriers of 180 kHz bandwidth, called Physical Resource Blocks (PRB), which can be stacked for scaling purposes. In this section, we detail the functionality of one PRB and describe the different phases that make up an NB-IoT communication.

A. Awakening Phase

The User Equipment (UE) synchronizes in time and frequency with eNodeB (eNB) thanks to the Narrowband Primary and Secondary Synchronisation Signals (NPSS and NSSS). It uses the information broadcasted in the Narrowband Physical Broadcast CHannel (NPBCH) frames by eNB to acquire information necessary to the Access Phase (below). The synchronization and broadcast signals constitute 25% of the downlink PRB.

B. Access Phase - Contention

To notify the eNB that it wishes to transmit, the UE must send a preamble in the uplink PRB. To this end, part of the PRB resources is dedicated to preambles, referred to as the **Narrowband Physical Random Access CHannel (NPRACH)**. Periodically, there is a random access opportunity (RAO) containing between 12 and 48 connection possibilities. Among these possibilities, the UE chooses an index on which to transmit its preamble at random.

If two or more UE's choose the same preamble index, there will be a collision. The UE's involved in a collision will retransmit after a randomly selected period of time. If the limit of the retransmissions is reached without a success, the UE abandons the effort. The backoff parameters – backoff window value and retransmission limit – are provided continuously by eNB through the broadcast channel, NPBCH. If the preamble is transmitted with success, the UE moves on to the data phase.

C. Data Phase - Congestion

In the data phase, messages are exchanged on allocated resources – there is no contention. Nevertheless, failures can still occur due to congestion: an eNB unable to allocate resources for a communication within a given amount time. The data phase comprises three channels:

(1) Narrowband Physical Uplink Shared CHannel

NPUSCH comprises all the space not occupied by the NPRACH in the uplink PRB. It is organized in Resource Units (RU). An eNB allocates RU's of different time-frequency shapes through UL Grant messages [14]. The different types of RU make it possible to adjust either the flow rate or the received SNR. A UL Grant specifies a number of useful bits to transmit, N_{bits}, between 16 and 2536 (Rel-15), and a number of RU's, between 1 to 10, depending on the coverage conditions of the UE.

(2) Narrowband Dowlink Control CHannel

NPDCCH is organized in narrowband control channel elements (NCCE). Either one or two NCCEs are mapped to a downlink subframe. Each NCCE carries essentially three kinds of control messages, called DCI (Downlink Control Information):

- Uplink (UL) Grant information (Format N0)
- Downlink (DL) scheduling information (Format N1)
- Indicator of paging or SI update (Format N2)

UL Grant is used to allocate resources for uplink communication. DL scheduling informs the UE it has data to receive in the NPDSCH. And Format N2 indicates either paging or the broadcast news has changed.

(3) Narrowband Physical Downlink Shared CHannel

NPDSCH is used to carry data traffic to UEs. One DL scheduling uses 1 to 10 subframes (without repetitions), it is UE specific and contains a full MAC-PDU. The amount of information bits, the lenght and the repetitions of the scheduling depends on the coverage conditions as for RUs in UL.

III. Analytical Model

The objective of this work is to evaluate the capacity of NB-IoT in terms of arrival rate, necessary for the calculation of the connection density as defined by the ITU [12]. In this section, we propose an analytical model that considers both the connection (random access) and the communication part (contention free but prone to congestion) of the protocol.

A. Model overview

To evaluate the performance of NB-IoT, we introduce the queuing model depicted in Fig. 1. Recall that communication in NB-IoT involves a two-phase process (§ II), each prone to

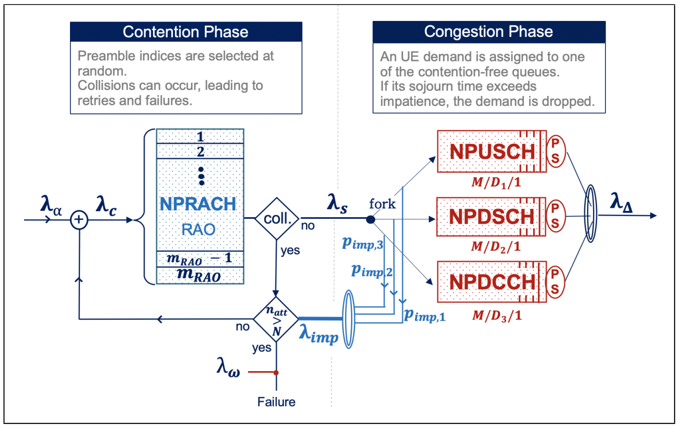

Fig. 1: Overview of the analytical model.

failures. The maximum capacity of the cell is reached for an overall failure probability of 99% [12].

Computing the probability of success, $p_{success}$ – the ratio between the successful flow rate, λ_Δ, and the incoming flow rate, λ_α – requires modelling both sources of failure. In the following two subsections, we model the two independent failure sources - collisions in NPRACH, and congestion in NPDCCH / NPDSCH / NPUSCH.

B. Preamble Contention Phase

In this section, we focus on the first part (NPRACH) of our queuing model (Fig. 1). We assume that new UE demands for connection initialization follow a Poisson process with mean rate λ_α. Failures due to collisions in NPRACH or congestion in NPDCCH / NPDSCH / NPUSCH are added to the flow of arrivals in NPRACH, leading to a total rate of λ_c

UE's wait for a Random Access Opportunity, **RAO**, every NPRACH period, δ_{RAO}. In every RAO there is a fixed number of subcarriers available, m_{RAO}[1]. Each UE chooses at random a subcarrier to transmit the preamble. Thus the average number of attempts per RAO is, $N_c = \lambda_c \cdot \delta_{RAO}$.

To compute the collision rate, we consider each subcarrier. Let X_j be the random variable representing the number of preambles sent on the j^{th} subcarrier per RAO. The probability, $\mathcal{P}_k$, that k UE's select the same preamble index, j, follows a binomial law independent of j:

$$\mathcal{P}_k = P(X_j = k) = \binom{N_c}{k} \cdot \left(\frac{m_{RAO}-1}{m_{RAO}}\right)^{N_c-k} \cdot \left(\frac{1}{m_{RAO}}\right)^k \quad (1)$$

Thus, the mean number of UE communication demands per unit of time admitted through the contention phase, λ_s, can be computed as follows:

$$\lambda_s = \frac{m_{RAO} \cdot \mathcal{P}_1}{\delta_{RAO}} \quad (2)$$

C. Congestion phase

In this section, we describe how we model the congestion phase and how to compute the congestion probability. Note that, every UE communication demand succeeding through the contention phase induces a workload on the contention-free channels. The eNB schedules this workload depending on the available resources in the UL and DL frames. The service time, the time it take of the respective UE to transmit its packets, depends on the workload and the available resources. Should the service time increase beyond a certain level, the UE abandons and declares failure.

1) **Model:***:* We model the three contention-free channels – NPDCCH, NPDSCH and NPUSCH – using queues with impatient customers. Each queue has an arrival rate, a service rate and a type of service, and a customer represents an UE communication. When a customer is not served before the impatience time, it abandons the queue. Therefore, *the impatience probability is the congestion probability*.

Queues with impatient customers are widely used in literature, including for modeling LTE [7]. Compared to [7], we make two contributions:

First, we assume a constant service rate, rather than an exponential rate. While this assumption makes the model more complex mathematically, it is more realistic since the queue service rate is deterministic, dependent on payload size and frame configuration.

Second, we model the NPDCCH, NPDSCH and NPUSCH queues using Processor Sharing (PS), as opposed to FIFO [7]. This choice is justified by the fact that, in NB-IoT, the eNB does not wait the end of a communication to begin to allocate resources for another one – all communications are served at the same time. One consequence of this service model is that the more customers are there in the queue, the lower is its service rate.

Formally, we model the NPDCCH / NPDSCH / NPUSCH of the congestion phase as an M/D/1-PS processor sharing queue. Unfortunately, no theoretical result exists regarding the impatience probability of as an M/D/1-PS. Therefore, our approach is to approximate it by the probability of a communication to be served after the impatience time. To this end, we compute the following equation on the sojourn time cumulative distribution, which is obtained from the probability density function [15] divided by s (integration):

$$\mathcal{P}(\lambda,\mu,T)^{M/D/1-PS} = \mathcal{L}^{-1}\left\{\frac{(\lambda+s)^2\cdot(1-\rho)\cdot e^{-\frac{\lambda+s}{\mu}}}{s\left(s^2+\lambda\cdot(\lambda+2s-\rho\cdot(\lambda+s))\cdot e^{-\frac{\lambda+s}{\mu}}\right)}\right\}(T) \quad (3)$$

with $\mathcal{L}^{-1}$, the Inverse Laplace Transform, λ the arrival rate, μ the service rate and ρ their ratio, at the impatience points introduced by the timeout timers of the NB-IoT standard.

In the following, we describe how to obtain λ, μ and T and compute the impatience probabilities.

2) **Arrivals - $\boldsymbol{\lambda}$***:* We model the queues with exponential arrivals. It is the same law we used to represent communication arrival in the cell λ_α. The parameter that characterizes this exponential distribution is λ_s (Eq.2).

3) **Services rate - $\boldsymbol{\mu}$***:* Considering the same amount of payload bytes to send, $\boldsymbol{B}$, service time varies only in function of the Code Rate (CR) used depending on coverage condition.

[1]We consider a single coverage enhancement (CE) zone.

Recognizing that, from a coverage distribution, we can compute an average CR and deduce an average communication workload per queue W_{NPUSCH}, W_{NPDSCH} and W_{NPDCCH}.

Finally, we obtain the service rates – D_1, D_2, D_3, – when introducing the amount of resources available in each channel:

$$\begin{cases} D_1 = \frac{Nb_{RU} \cdot \left(1 - \frac{\lceil R_{ra} \cdot d_p \rceil \cdot m_{RAO}}{48 \cdot \delta_{RAO}}\right)}{W_{NPUSCH}} \\ D_2 = \frac{750 \cdot \frac{G-1}{G}}{W_{NPDSCH}} \\ D_3 = \frac{750 \cdot N_{ncce} \cdot \frac{1}{G}}{W_{NPDCCH}} \end{cases} \tag{4}$$

with R_{ra}, d_p, N_{ncce} and G, respectively, the number of repetition per preamble, the duration of the preamble, the number of NCCE per grants and the ratio between NPDSCH and NPDCCH.

4) **Impatience times - $\boldsymbol{T}$***:* Impatience represents the failure due to starvation of resources in the NPUSCH, NPDSCH or NPDCCH. For NPDCCH, we use the duration of the timing window for grant messages, defined at the UE side. For NPUSCH and NPDSCH we use the maximum delay configurable in grants (respectively $128ms$ in N0 and $64ms$ in N1). We multiply by the number of messages per channel to get the impatience per communication, T_{NPUSCH}, T_{NPDSCH} and T_{NPDCCH}.

5) **Impatience probability***:* Now, we can leverage Eq. (3) to compute the impatience probabilities:

$$\begin{cases} p_{impa,1} = 1 - \mathcal{P}(\lambda_s, D_1, T_{NPUSCH})^{M/D/1-PS} \\ p_{impa,2} = 1 - \mathcal{P}(\lambda_s, D_2, T_{NPDSCH})^{M/D/1-PS} \\ p_{impa,3} = 1 - \mathcal{P}(\lambda_s, D_3, T_{NPDCCH})^{M/D/1-PS} \end{cases} \tag{5}$$

To compute the average number of communications in congestion, $\boldsymbol{\lambda_{imp}}$, we recognize that, once a congestion happens in one queue, its workload is removed from the other queues. Thus, the queue with the highest congestion probability drives the global impatience probability:

$$\lambda_{imp} = \lambda_s \cdot \left(\max_{k \in \{1,2,3\}} p_{impa,k} \right) \tag{6}$$

D. Overall success probability

To determine the capacity of an NB-IoT cell, it is necessary to know the probability of success of a communication given an incoming flow. The success probability, $\boldsymbol{p_{success}}$, is linked to the probability that the communication fails, $\boldsymbol{\mathcal{P}_{outage}}$, as follows:

$$p_{success} = 1 - \mathcal{P}_{outage} \tag{7}$$

Recall that, if a communication demand fails, either due to collision or congestion, the UE will retry after a randomly selected interval. The UE abandons if it exceeds the maximum number of retries, $\boldsymbol{N}$.

The failure probability for a single attempt, $\boldsymbol{\mathcal{P}_f}$, is the inverse of the joint probability of the communication succeeding through the contention phase (§ III-B) and the congestion phase (§ III-C):

$$\mathcal{P}_f = 1 - \left(1 - \frac{\lambda_c}{\lambda_s}\right) \cdot \left(1 - \frac{\lambda_{imp}}{\lambda_c}\right) \tag{8}$$

TABLE I: Parameter Settings

Parameter	Setting	Parameter	Setting
G	$[1.5, 2, 4, 8, ..., 64]$	d_p	6.4 ms
δ_{RAO}	$[20, 40, ..., 2560]$ ms	N_{NCCE}	2
m_{RAO}	$[12, 24, 36, 48]$	B	Variable

The outage probability is defined as the probability that the communication fails $\boldsymbol{N}$ times in a row:

$$\mathcal{P}_{outage} = \mathcal{P}_f^N \tag{9}$$

A key aspect of the model is that, due to the retries, the load, λ_c, is higher than the input charge, λ_α. To calculate λ_c, it is necessary to determine the average number of attempts per communication. The number of required attempts is computed as follows:

$$N_{Tx}(\lambda_c) = (1 - \mathcal{P}_f) \cdot \sum_{k=0}^{N} (k+1) \cdot \mathcal{P}_f^k + N \cdot \mathcal{P}_f^N \tag{10}$$

Finally, we can now compute the communication flow arriving into NPRACH by solving the following iterative equation:

$$\lambda_c = \lambda_\alpha \cdot N_{Tx}(\lambda_c) \tag{11}$$

At every iteration λ_c is updated and so is $p_{success}$. Eventually, the model converges to a stable state, outputting the success probability of a communication depending on the incoming flow rate and the cell configuration (Table. I).

IV. Performance Evaluation and Analysis

The performance evaluation is organized in two parts. In the first part, we evaluate the accuracy of our analytical model using simulations. In the second part, we evaluate analytically the performance of NB-IoT as function of different configuration parameters using our model.

A. Model Validation

1) **Methodology:** We developed a discrete event-driven simulator in Python that implements NB-IoT communications at a Resource Block level of abstraction, as described in Fig. 1. To evaluate the accuracy of our analytical model, we use two performance metrics:

A **The probability of success**, $p_{success}$, as defined in Eq. 7

B **Capacity**, λ_{max}, defined as the maximum load for which the probability of success is 99%, as per the 5G quality of service requirements for the IoT segment [12].

Finally, we compare the results of the simulations with those of our model as well as the model proposed in [7].

2) **Probability of success***:* Figure 2 shows the probability of success computed via simulations, our analytical model and the model in [7]. The data shows the waterfall behavior of the success probability, characteristic of communication systems with re-transmissions. Our model predicts the collapse point of the success probability with high accuracy, while [7] is overly optimistic. This can be explained by the fact that our model (§ III-C) uses constant service times for the NPDCCH / NPDSCH / NPUSCH queues as opposed to approximating it using a Poisson process, as in [7]. A slight difference is observed between the simulation and our model. It results from the fact that the model assumes the arrivals at the NPDCCH /

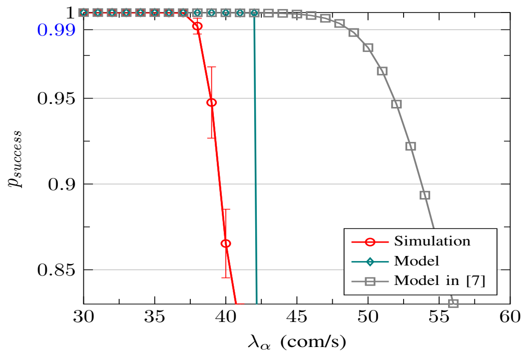

Fig. 2: Comparison Models for $p_{success}$ (250 Bytes) with 95% confidence interval.

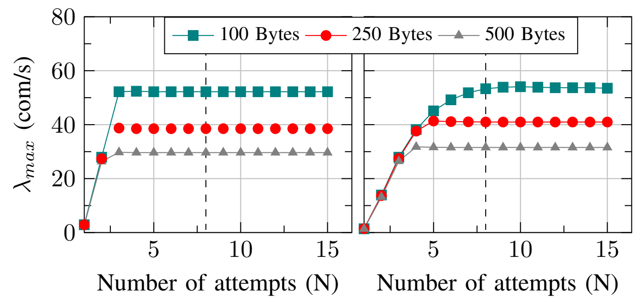

Fig. 4: Left: $\delta_{RAO} = 80ms$. Right $\delta_{RAO} = 160ms$.

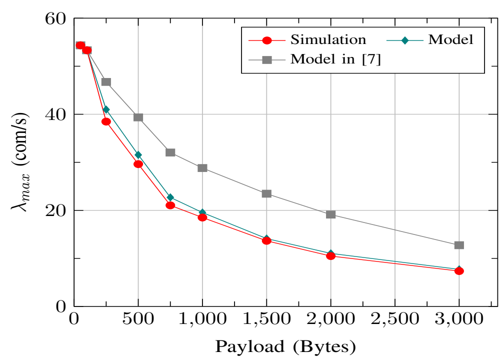

Fig. 3: Models and Simulation Comparison for $\delta_{RAO} = 160ms$ and $m_{RAO} = 24$ subcarriers.

NPDSCH / NPUSCH queues follow a Poisson process. In the simulations, as in reality, the arrivals can be bursty – several connection requests can successfully complete the NPRACH phase simultaneously.

3) **Capacity analysis***:* Figure 3 shows the NB-IoT capacity in terms of communications per second as function of the payload and the NPRACH configuration. The data shows that in all configurations our analytical model closely matches the simulations. [7], on the other hand, always overestimates the network capacity.

4) **Key takeaway***:* The simulation study shows that our model is capable of accurately capturing the behaviour of NB-IoT. Fundamentally, it means that it is better to model a system exactly – our choice of using fixed service times for the NPDCCH / NPDSCH / NPUSCH queues – even if the mathematical solution is approximate, than using an approximate model – assuming exponential service time [7] – even if it can be solved exactly.

B. NB-IoT performance assessment

In the following, we use our analytical model to assess the performance of NB-IoT. We focus on two parameters which have a central role in the protocol:

1) The number of attempts, representing the trade-off between reliability and latency.
2) The ratio of NPRACH and NPUSCH, representing the trade-off between preamble contention and data congestion.

1) **Impact of the number of attempts:** Figure 4 shows the capacity of the network as function of the number of attempts for different values of payload and the NPRACH period. The data points to two interesting conclusions. First, with very few attempts, the network suffers from losses during the random access phase (NPRACH), preventing many UE's from establishing a connection to the eNB. As a result, the network capacity drops. Second, regardless of the configuration, there exists a number of attempts beyond which the network capacity does not improve any further.

Therefore, while according to the standard, the eNB can assign a value ranging between one to 100, our analysis shows that the number of attempts need to be limited to a far smaller range, between 5 and 8. For the rest of our study, we set this parameter to 8.

2) **Impact of the NPRACH, NPUSCH ratio:** Figure 5 shows the capacity of an NB-IoT network as function of $r_{UL} = \frac{NPRACH}{NPUSCH}$. In this experiment, we measure the capacity in bytes/s so as to get a representative idea of the load in useful bytes successfully transmitted over the entire physical resource block (PRB). The data shows that for the same traffic load, bigger, thus less numerous, messages are better served when r_{UL} is small. This can be explained by the fact that less messages means less contention during NPRACH. The opposite is true for small messages.

The impact of r_{UL} can be dramatic, e.g., choosing the optimal value for 500-byte traffic ($\approx 8\%$) reduces the cell capacity by a factor of 2 in case of a 100-byte traffic.

The behaviour with 100-byte messages is particularly interesting and counter-intuitive at first. The data shows a flat zone between $r_{UL} = 0.15$ and 0.5, meaning that neither contention nor congestion are defining the cell capacity. A careful analysis reveals that impatience is triggered on the downlink channel even though we are considering essentially uplink traffic. Recall that a certain number of control messages have to be sent regardless of the payload size (connection procedure). As a result, there is an upper bound (around 65) on the number of communications that can be supported at any point in time.

This result suggests that for small-packet traffic, NB-IoT needs a connection procedure lighter than the current User Plane optimization. Our finding is all the more relevant to the development of NB-IoT when considering that many IoT applications are expected to generate small packets.

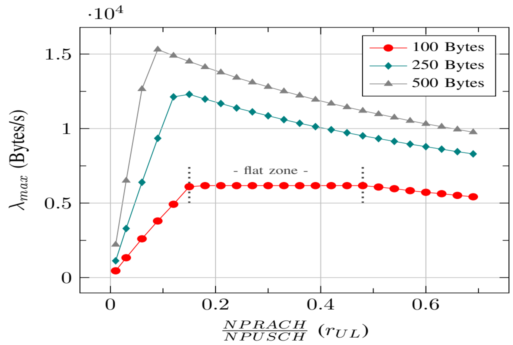

Fig. 5: Models and Simulation Comparison

V. Related Work

Madueño et al. [7] were the first to introduce an analytical model for assesing the performance of LTE, from which NB-IoT was inspired. However, the model simplifies the congestion phase of the protocol: it assumes exponential times for the service time and a First In First Out (FIFO) type of service. Our model uses deterministic service rates and Processor Shared (PS) type of services, more accurately capturing the functioning of the NB-IoT congestion phase. Feltrin et al. [8] introduce a theoretical model assuming no retransmissions, an important part of NB-IoT. Furthermore, they did not consider the resource starvation on the control channel. A model to evaluate NB-IoT performance against SNR by varying only the number of repetitions and the bandwidth used for transmissions is introduced in [16]. Our study complements this work by adding the variation in the sizes of payloads, the number of attempts and the NPRACH/NPUSCH ratio.

The work in [17] introduced an approach for estimating the required SNR using Shannon's information theory. Following up, [18] linked the required SNR to BLER and coding schemes through MATLAB simulations and a curve fitting model. The impact of repetitions on the performance NB-IoT has been studied in [19]. We relied on these works to design our cell configuration.

Finally, coverage and capacity of NB-IoT and LTE-M in a rural environment are analyzed in [10] using the configuration and location of commercially deployed LTE sites. The work shares one of our findings that the downlink transmissions strongly degrades the NB-IoT performance. However, their capacity estimates are very pessimistic because the platform on which the analysis was carried out did not include a scheduler adapted to NB-IoT for single-tone transmissions. A theoretical model has the advantage of abstracting out implementation shortcomings which can be specific to an operator.

VI. Conclusion

The new mechanics introduced in NB-IoT provide extensive modularity in order to meet the 5G IoT requirements in terms of coverage, latency, power consumption, density and complexity. In this paper, we have introduced a highly flexible theoretical model for assessing the performance of NB-IoT as function of network settings and protocol configuration. After demonstrating the accuracy of the model using a simulator we have made available as open-source, we evaluated the performance of NB-IoT as function of its key parameters. The data shows that failing to find the right balance between the random access and the contention-free channels can lead to a halving of the cell capacity. We also revealed that for small payload, the capacity bottleneck is not the number of connection opportunities but the lack of resources in the control channel (NPDCCH). Our results show the enabling value of our model in configuring an NB-IoT cell.

References

[1] K. A. Patil and N. R. Kale, "A model for smart agriculture using iot," in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 2016, pp. 543–545.

[2] Z. Yan and H. Gang, "Design of intelligent water metering system for agricultural water based on nb-iot," in *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2019, pp. 1665–1669.

[3] W. Manatarinat, S. Poomrittigul, and P. Tantatsanawong, "Narrowband-internet of things (nb-iot) system for elderly healthcare services," in *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, 2019, pp. 1–4.

[4] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, "Smart factory of industry 4.0: Key technologies, application case, and challenges," *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[5] . GPP. Release 13 overview. [Online]. Available: https://www.3gpp.org/release-13

[6] "Mobile IoT commercial launches." [Online]. Available: https://www.gsma.com/iot/mobile-iot-commercial-launches/

[7] G. C. Madueño, J. J. Nielsen, D. M. Kim, N. K. Pratas, Stefanović, and P. Popovski, "Assessment of LTE Wireless Access for Monitoring of Energy Distribution in the Smart Grid," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 675–688, March 2016.

[8] L. Feltrin, M. Condoluci, T. Mahmoodi, M. Dohler, and R. Verdone, "Nb-iot: Performance estimation and optimal configuration," in *European Wireless 2018; 24th European Wireless Conference*, May 2018, pp. 1–6.

[9] H. Bello, X. Jian, Y. Wei, and M. Chen, "Energy-delay evaluation and optimization for nb-iot psm with periodic uplink reporting," *IEEE Access*, vol. 7, pp. 3074–3081, 2019.

[10] M. Lauridsen, I. Z. Kovacs, P. Mogensen, M. Sorensen, and S. Holst, "Coverage and capacity analysis of lte-m and nb-iot in a rural area," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[11] R. Barbau. (2019) NB-IoT simulator with M/D/1-PS queues in python. [Online]. Available: https://github.com/Romainbarbau/NB-IoT_Simulator_with_M-D-1-PS_queues

[12] ITU, "Guidelines for evaluation of radio interface technologies for imt-2020," vol. Report M.2410, October 2017.

[13] A. Sivanathan, D. Sherratt, H. H. Gharakheili, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Characterizing and classifying iot traffic in smart cities and campuses," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2017, pp. 559–564.

[14] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (e-utra); physical channels and modulation (release 15)," vol. 3GPP TS 36.211, no. V15.6.0, June 2019.

[15] W. G. J.L. van den Berg, O.J. Boxma, "Sojourn times in the m/g/1 queue with deterministic feedbackk," vol. COMMUN. STATIST.-STOCHASTIC MODEL, no. 5, 1989.

[16] R. Cruz, A. Coelho, R. Campos, and M. Ricardo, "A theoretical model for planning nb-iot networks," in *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2019, pp. 1–4.

[17] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. Navarro-Ortiz, and J. M. Lopez-Soler, "An analytical performance evaluation framework for nb-iot," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7232–7240, Aug 2019.

[18] S. Ravi, P. Zand, M. El Soussi, and M. Nabi, "Evaluation, modeling and optimization of coverage enhancement methods of nb-iot," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2019, pp. 1–7.

[19] L. Feltrin, M. Condoluci, T. Mahmoodi, M. Dohler, and R. Verdone, "Nb-iot: Performance estimation and optimal configuration," in *European Wireless 2018; 24th European Wireless Conference*, May 2018, pp. 1–6.