

CTMC modeling for M2M/H2H coexistence in a NB-IoT Adaptive eNodeB

A. H. EL Fawal⁽¹⁾⁽²⁾, A. Mansour⁽¹⁾, M. Najem⁽³⁾, F. Le Roy⁽¹⁾, and D. Le Jeune⁽¹⁾

⁽¹⁾ Lab-STICC, UMR 6285 - CNRS, ENSTA Bretagne, 29806 Brest, France

⁽²⁾ AUL University, Beirut, Lebanon

⁽³⁾ AUCE University, Beirut, Lebanon

E-mail: ahmad.el-fawal@ensta-bretagne.org; mansour@ieee.org; mohamadnajem@auce.edu.lb; frederic.le_roy@ensta-bretagne.fr; denis.le_jeune@ensta-bretagne.fr

Abstract—The next generation of mobile systems are expected to support the new promising Machine-to-Machine (M2M) technology carried by the advance of Internet of Things (IoT) devices. In the near future, an exponential growth of the number of M2M devices is expected due to their ubiquity. In normal situations, a limited bandwidth in Narrow Band-Internet of Things (NB-IoT) technology may help in improving IoT requirements effectively. However, in emergency and disastrous moments, M2M expected storms lead inevitably to network saturations.

In this manuscript, we propose a novel Adaptive eNodeB (A-eNB), which solves the network overload problem gradually, while keeping Human-to-Human (H2H) traffic not to be affected dreadfully. The network adaptation is provided through a dynamic NB-IoT bandwidth reservation aiming to increase the number of M2M connections accessing NB-IoT network with minimal overload congestion problems.

A Continuous-Time Markov Chain (CTMC) is proposed helping the H2H/M2M coexistence to become more approachable especially during disaster scenarios.

Our results show that by leasing 18 resource blocks using an A-eNB for NB-IoT traffic can result a completion rate of 98% on M2M traffic throughout emergency scenarios.

Keywords— LTE-A, NB-IoT, M2M, H2H, CTMC, Markov chain, eNodeB.

I. INTRODUCTION

Throughout the last decade, Internet of Things (IoT) has paved the way to a prominent game changer known as Machine-to-Machine (M2M) communications. This new trend influences the human lives by delegating control in many fields to autonomous and semi-autonomous machines. Promoting dummy machines to smart adaptive cognitive-machines, allows them to monitor, track and control various parts in civil lives over distance, thus reducing the human resources, mistakes and mood-swing drawbacks in future businesses. Following this aspect, the IoT becomes a promising technology with around 50 Billion connections in 2020 which enforces an unavoidable coexistence among Human-to-Human (H2H) and M2M traffics in one Long Term Evolution-Advanced (LTE-A) network [1]; Consequently, an efficient radio access strategy becomes one of the most challenges for mobile operators, researchers and the 3rd Generation Partnership Project (3GPP) community [2]. They sound ultimately keen on conducting several studies and researches to identify the mutual impact among M2M and H2H communications.

LTE-A was coined initially to support H2H communications (e.g., web-browsing, phone calls, internet televisions, etc.). M2M devices (e.g., smart cities, mobile health, smart meters, etc.), with their sporadic small packet sizes, pour their payloads into the same LTE-A channel but with different aspects and specifications.

On one hand, H2H traffic enforces telecommunication operators to broaden the LTE-A bandwidth, thus providing higher data rates which implies a significant complexity on the future model design. On the other hand, M2M traffic requires a narrower bandwidth due to low data rates which paves the way to low complexity models. Due to these differences, it would be an imperfect match for these two traffics to share the same LTE-A bandwidth. 3GPP introduces a licensed spectrum technology so-called “Narrowband-Internet of Things” (NB-IoT). NB-IoT technology occupies 180 KHz within the LTE-A bandwidth, which paves the way to a significant reduction on model complexity consequently a minimum cost model. But with a diversity of applications a flood of devices trying to access the network and send their payloads becomes unavoidable. Inevitably, the synchronized manner of certain devices serving the same type of application will cause a performance degradation on NB-IoT services.

In this manuscript, we propose an Adaptive eNodeB for NB-IoT technology (A-eNB) while addressing all aforementioned challenges. Our proposed A-eNB solves the M2M overload congestion gradually, while maintaining the Quality of Service (QoS) of H2H traffic within the acceptable standards. The network adaptation is provided through a dynamic NB-IoT bandwidth re-allocation with the objective of maximizing the number of M2M connections while maintaining H2H traffic. The main contributions of this paper are the following:

- A new Adaptive eNodeB architecture for NB-IoT technology.
- An evaluation of the A-eNB on H2H and M2M traffics.
- A Continuous Time Markov Chain (CTMC) model is used as stochastic process tool to characterize H2H/M2M coexistence based on analytic equations.
- Many simulations are constructed to validate our model.

II. STATE OF THE ART

By 2021, Ericsson predicts that there will be around 28 billions of global connections [3]. When the number of devices peaked exponentially, a transformation effect on the bandwidth becomes a must. This effect is accumulative either as a result of connected devices load with higher bandwidths for huge time horizons, or as consequence of a burst of autonomous devices with their sporadic payloads. In [4], A proposed Small Data Transmission scheme (SDT) has been introduced which enables devices in an idle state to transmit a small data packet without following a Radio Resource Control (RRC) connection setup process. Although, an improvement on the maximum supported devices which have insufficient radio resources in NB-IoT can be achieved, but on the flip side, the larger the number of devices the greater the bandwidth needed, that wasn't discussed. In [5], a real drive test measurement was conducted with an application based on a User Equipment (UE) performing one UpLoad (UL) transmission with a payload of 128 or 256 bytes similar to the Mobile Autonomous Reporting (MAR) traffic model. The authors concluded that NB-IoT technology can provide coverage for more than 95% of the devices due to its maximum coupling loss (164 dB).

Based on MAR traffic models, the results in [6] show that more than 72K devices can be supported by NB-IoT cells. However, the 72K result triggers a response of many questions: what if all these devices are trying to send their payloads simultaneously during a disaster?, what is the saturation point of the limited NB-IoT bandwidth? and how we can deal with the expected saturation during a disaster situation?

In [7], a Control Plan (CP) solution has been specified in NB-IoT technology, where a device in an idle state transmits a data packet through the RRC connection setup procedure to reduce the signaling overhead. But if we take into consideration that: (a) the maximum number of repetition for each RRC message is 128 repetitions during one transmission [8], (b) the signaling overhead on the up-link radio resources can be significantly increased by a large number of devices in NB-IoT networks, we conclude clearly that the radio resources to support loads of devices placed in deep coverage or cell edge can be insufficient, and requires additional research to address this problem efficiently.

In [9], the authors propose an energy-efficient resource allocation algorithm to optimize the energy efficiency while guaranteeing the QoS provisioning through joint channel selection and power allocation design. First, the available resources are scheduled by each UE in an energy-efficient way. Since UEs are only interested in improving their own performance, a game-theoretic approach is proposed with the aim to model the distributed resource allocation problem as a non-cooperative game. In [10], an adaptive dropout deep computation model with crowdsourcing is presented for big data feature learning in industrial Internet of Things. Dropout was proposed to prevent over-fitting for deep learning models. It is especially effective to learn the large-scale deep neural networks with a small number of training samples.

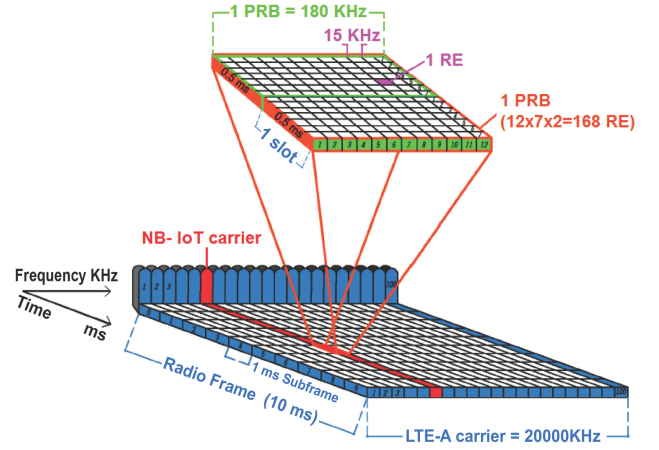


Fig. 1. Bandwidth Limitation (BL) for NB-IoT technology within LTE-A bandwidth, where PRB: Physical Resource Block, RE: Resource Element

One of the targets of 3GPP to introduce NB-IoT technology, is to serve better deep coverage IoT devices (e.g. basement). Using a repetition transmission scheme in NB-IoT technology can be considered as a serious burden to its limited bandwidth. All previous studies focus on the capacity and the way to connect a huge number of devices in a NB-IoT system effectively. This is an important step towards IoT requirements, but there is a prominent aspect that merits to be improved; It is bandwidth and its limitation. Therefore, it can be essential to study the efficient use of bandwidth in NB-IoT technology or to find an adaptive way to extend it with a minimal cost on LTE-A bandwidth. In this manuscript, we are motivated to study the limitation of a NB-IoT bandwidth while proposing an effective solution for it.

III. LTE-A AND NB-IOT DATA-RATE

In order to study LTE-A and NB-IoT bandwidths and their limitations, especially their mutual influences, we propose a presentation of the time-frequency resources and the correlation with data-rates for both H2H and M2M traffics.

In LTE-A, time-frequency resources are subdivided according to the following structure:

In time domain, the largest unit of time is the radio frame (10 ms), which is subdivided into ten subframes (1 ms). Each subframe is split into two slots (0.5 ms). Each slot comprises seven symbols (0.5/7 ms) [11], as shown in Fig. 1.

In frequency domain, resources are grouped in units, such that one unit of:

- One sub-carrier (a small channel spaced at 15 KHz with the adjacent channel) for a duration of one symbol is termed as a Resource Element (RE) with 15 KHz spacing.
- 12 sub-carriers for a duration of one slot is termed a Resource Block (RB) with 180 KHz bandwidth.
- 12 sub-carriers for a duration of one sub-frame is termed a Physical Resource Block (PRB) with 180 KHz bandwidth.

Thus a RB comprises $7 \times 12 = 84$ REs, while a PRB comprises $7 \times 12 \times 2 = 168$ REs.

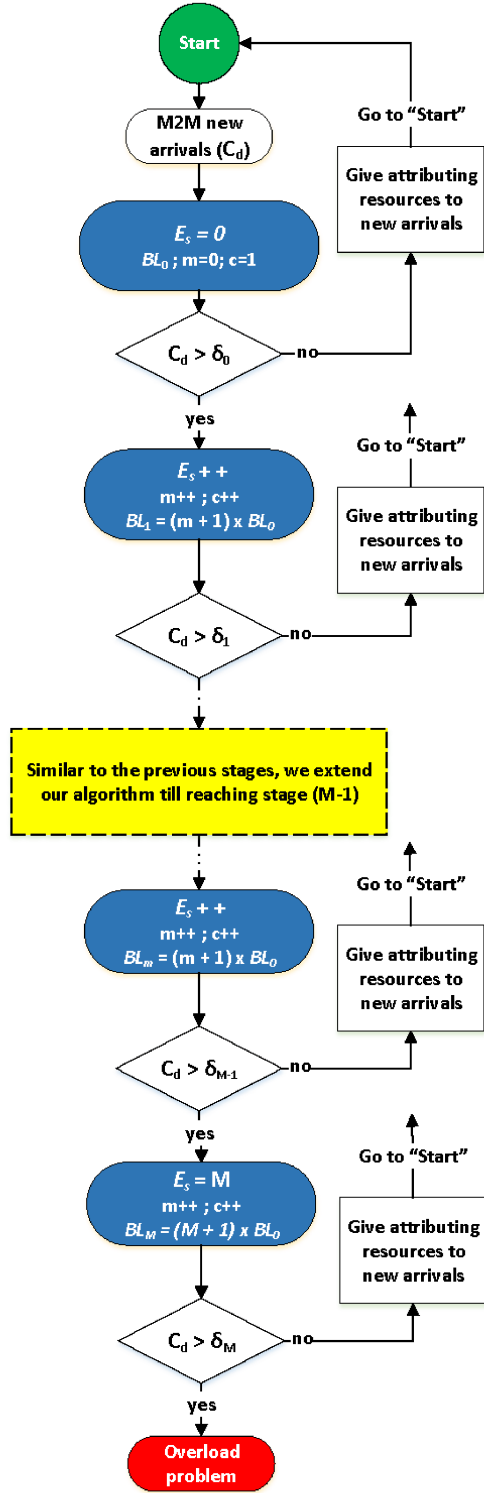


Fig. 2. Flow chart for NB-IoT Adaptive eNode-B over normal cycle: from initial stage till overload problem, where E_s : Emergency stage number, BL_m : Bandwidth Limited at the m^{th} stage, $BL_0 = 200KHz$: Initial Bandwidth Limited, m : current stage number, M : Maximum stage number, C_d : number of M2M connected devices, δ_m : threshold at the m^{th} stage, c : number of system resources.

Moreover, a PRB represents the minimal unit that can be scheduled for a UE to send or receive data; Now, If we recall that it is expected to have more than 52K devices trying to send their payloads simultaneously during a disaster event, that explains the necessity of studying the maximum data rate in classical LTE-A and NB-IoT technologies:

- (a) In LTE-A, each RE can carry 6 bits in the best modulation scheme (64QAM). Consequently, the maximum data-rate can be estimated as follows:

$$R_{max} = n \times o \times l \times p \times e \quad (1)$$

Where R_{max} : is the maximum data rate, n : number of sub-carriers, o : number of OFDMA symbols, l : number of slots, p : number of PRBs, e : number of carried bits per RE.

Let's assume that we have a 20 MHz total bandwidth (with a single antenna). As one PRB is equivalent to 180 KHz, then there are 100 PRBs available with 2 MHz guard-band in 20 MHz total bandwidth. So, the maximum data-rate for the 100 PRBs is about 100 kb/ms (100 Mbps approx.).

- (b) The previously computed maximum data rate is usually dedicated to H2H users. However, NB-IoT technology dedicates in 3GPP Rel-13, a 180 KHz of the total bandwidth for M2M communications. Following the same previous calculation but with 1 PRB, QPSK modulation and half-duplex mode; the maximum data-rate in NB-IoT technology is reduced to 150 Kbps for the M2M traffic.

If we compare the storm rates in the case-study presented in section VI-A and summarized in Table I to the maximum data-rate in NB-IoT technology (150 Kbps), we conclude that the available bandwidth will suffer from a huge degradation which requires an efficient solution as discussed in the next section.

IV. NB-IoT ADAPTIVE eNodeB

All the above calculations shed the light on the importance of an intelligent solution which operates effectively to absorb any expected signaling storm. To this end, we propose a promising approach, which extends the classical functionality of eNodeB and includes an adaptive control of the bandwidth based on the M2M load status throughout the network. The proposed A-eNB manages automatically both H2H and M2M traffics. Based on the instantaneous requests of M2M, a part of the total bandwidth will be dedicated to fulfill these sudden needs.

The Adaptive eNodeB for NB-IoT networks depicted in Fig. 2, is designed to operate in different conditions, depending on the number of M2M connected devices C_d throughout the network:

A) Initial stage

In normal situations, an A-eNB works similar to any legacy eNodeB in LTE-A/NB-IoT networks. A part of the total bandwidth is dedicated to NB-IoT traffic. For instance, a total of 20000 KHz (100 PRBs) will be divided into a basic Bandwidth-Limited ($BL_0 = 200KHz$) reserved for M2M devices, and the remaining bandwidth (99 PRBs) are dedicated for H2H devices.

TABLE I
CASE-STUDY M2M STORMS

Group #	M2M device Type	Message size (Bytes)	Rate (msg/day)	Number of devices (K)	Storm rate (Kbps)	Number of storms (Storm/day)
1	Asset tracking	50	100	20	1600	500
2	Assisted medical	100	8	20	3200	40
3	Environmental monitoring	200	24	20	6400	120

During this stage, M2M devices are allowed to access BL_0 reserved for NB-IoT to send their payloads via different IoT applications, while H2H devices use the remaining bandwidth to exchange their information.

B) Emergency stages

In the case of disaster, a huge number of M2M devices saturate the initial NB-IoT bandwidth (BL_0) by their signaling storm briefly. When C_d cross the initial saturation threshold δ_0 (configurable), the bandwidth dedicated to M2M starts to increase from $BL_0 = 200KHz$ to $BL_1 = 400KHz$, to allow more M2M devices to access the network.

If C_d cross a next saturation threshold δ_m , caused by an additional M2M signaling storm, the A-eNB adapts gradually its bandwidth from BL_1 till it reaches:

$BL_{(M)} = [(M + 1) \times BL_0]$, as shown in Fig. 3.

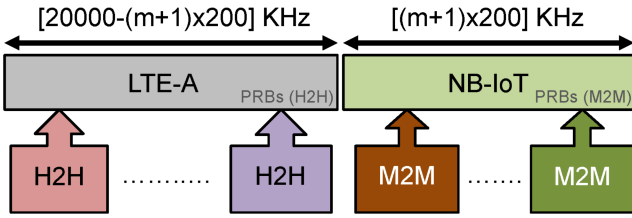


Fig. 3. Adaptive bandwidth during an emergency scenario, where m : current stage number (0 to M), PRB: Physical Resource Block.

For a given stage m , we can calculate:

- Total number of PRBs reserved for M2M devices:
 $p_{m2m} = m + 1$.
- Maximum M2M data-rate: following equation (1) the maximum M2M data-rate could be estimated as follows:

$$R_{m2m} = n \times o \times l \times p_{m2m} \times e \quad (2)$$

- Total number of PRBs reserved for H2H devices:
 $p_{h2h} = 100 - p_{m2m}$
- Maximum H2H data-rate: similar to equation (2):

$$R_{h2h} = n \times o \times l \times p_{h2h} \times e \quad (3)$$

C) Reset stage

Once the number of M2M connections starts decreasing, A-eNB reduces iteratively the NB-IoT bandwidth BL_m until achieving the initial state with BL_0 at the end of the disastrous events, and consequently starts operating similar to any traditional e-NodeB.

Following this methodology, the A-eNB has the adaptability and robustness to lend progressively a temporary bandwidth: $(M + 1)$ folds of 200 KHz to M2M devices tentative use during disaster scenario. Inevitably, this solution can soak up most of M2M storms with a minimal impact on H2H devices.

V. CTMC ANALYTICAL MODEL

A Markov chain is a stochastic model describing a sequence of possible events for a system that could be represented as a set of states. Each state has a probability, called steady-state probability $\pi_{(i,j,k,etc.)}$ which represents the probability to be in a given state $s(i,j,k,etc.)$.

In our CTMC model, any state $s(i,j,k,etc.)$ represents the system state after receiving an ongoing request or a termination of service for an application.

Our methodology to build this model consists of four steps:

- First, we use CTMC as a stochastic method to describe the sequence of possible events built on different states and probabilities for various applications (e.g., H2H and M2M requests).
- Second, we generate the equilibrium equations which rule the probability of transitioning from one state to another.
- Third, by applying these probabilities and states in a linear system and solving it, we calculate all steady-state probabilities.
- Fourth, using the aforementioned probabilities, we propose some metrics to characterize the performance of the network, with the aim to evaluate M2M and H2H traffics.

A. CTMC model for one traffic

As a preliminary step, we use a CTMC model for a single traffic which mimics a simulation model proposed in [12]. Then, in VI-B, an analysis for both the analytical and simulation results are conducted in order to validate the accuracy of the simulation model that will be used in our scenarios.

The CTMC model has the following assumptions:

- Arrivals suppose to follow a Poisson's distribution with an average arrival rate (λ) and move the process from state (i) to ($i + 1$).
- Service times assumed to follow an exponential distribution with rate parameter (μ).
- A single server ($c = 1$) serves one event only in the smallest time interval.
- When the service is completed, the served request leaves the system and the number of ongoing services in the system reduces by one (i) to ($i - 1$).

1) *Representing the model as a set of states:* The system moves from one state to another if an event occurs (increasing/decreasing of i), as shown in Fig. 4.

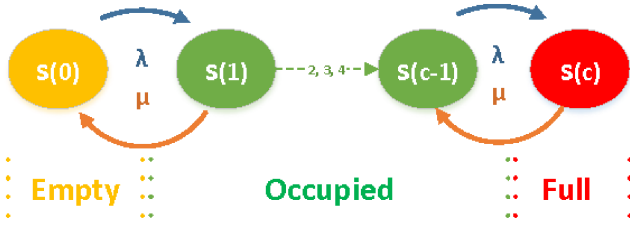


Fig. 4. Simple CTMC states, where s : state, c : system resources, λ : average arrival rate, μ : service rate .

2) *Generating the equilibrium equations:* By considering new arrival events with an average rate λ and a service rate μ , we can create the equilibrium equations based on [13]. The system that could be in one of the following three cases:

- Empty state: $i = 0$, includes one state only and its equilibrium equation is given:

$$\lambda\pi_{(i)} = \mu\pi_{(i+1)} \quad (4)$$

$\pi_{(i)}$: is the steady-state probability in a given i state.

- Occupied state: $0 < i < c$, when the equilibrium equation depends on two states (entering the state and leaving the state):

$$(\lambda + \mu)\pi_{(i)} = \lambda\pi_{(i-1)} + \mu\pi_{(i+1)} \quad (5)$$

- Full state: $i = c$, includes one state and its equilibrium equation is given as follows:

$$\mu\pi_{(i)} = \lambda\pi_{(i-1)} \quad (6)$$

The three previous equations can be represented in one general equation as follows:

$$(\alpha\lambda + \beta\mu)\pi_{(i)} = \beta\lambda\pi_{(i-1)} + \alpha\mu\pi_{(i+1)} \quad (7)$$

Where $\alpha = 0$ indicates the Full state (otherwise $\alpha = 1$) and $\beta = 0$ represents the Empty state (otherwise $\beta = 1$).

3) *Linear system solution:* The previous general equation (7) can be written in the following form:

$$\mathbf{A}\Pi = 0 \quad (8)$$

Where $\Pi = [\pi_{(0)}, \pi_{(1)}, \dots, \pi_{(c)}]^T$: represents the steady-state probability vector, and $\mathbf{A}$ represents a square matrix with $(c + 1) \times (c + 1)$ dimensions, which is considered as coefficient in the linear system.

The linear system in (8) should be solved while respecting the following parameter equations:

$$\sum_{i=0}^c \pi_{(i)} = 1 \quad (9)$$

$$0 \leq \pi_{(i)} \leq 1 \quad (10)$$

According to the nature of equilibrium equation system, shown in (4), (5), (6) and (7), we can conclude that: $\text{Rank}(\mathbf{A}) = c$.

By replacing the first row of $\mathbf{A}$ by the coefficient of equation (9), we obtained the following modified system:

$$\mathbf{B}\Pi = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (11)$$

In this way, $\mathbf{B}$ becomes a full rank $(c + 1) \times (c + 1)$ matrix.

Finally, following huge number of simulations constructed with $c = 3$ and $c = 6$ while changing $1 \leq \lambda_1, \lambda_2 \leq 10$ and $1 \leq \mu_1, \mu_2 \leq 10$ in each simulation, we show that all times $\det(\mathbf{B}) \neq 0$ and the unique solution obtained from (11) was satisfying the constraint (10).

B. CTMC model for M2M/H2H traffics

Now we extend our model to deal with two variables (i, j) which denote the number of ongoing services for two traffics H2H and M2M respectively.

1) *Representing the system as a set of states:* The system moves from one state to another if an event occurs (increase or decrease for i or j). The two traffics assumed to follow Poisson distribution which has two average arrival rates (λ_1, λ_2) while the two service rates (μ_1, μ_2) assumed to follow an exponential distribution. The general CTMC model is shown in Fig. 5.

2) *Generating the generic equilibrium equation:* Similar to section V-A2 but while considering two ongoing services instead of one, we end up with the following general equation:

$$\begin{aligned} &(\alpha\lambda_1 + \alpha\lambda_2 + i\beta\mu_1 + j\beta\mu_2)\pi_{(i,j)} \\ &= \beta\lambda_1\pi_{(i-1,j)} + \beta\lambda_2\pi_{(i,j-1)} \\ &+ (i+1)\alpha\mu_1\pi_{(i+1,j)} + (j+1)\alpha\mu_2\pi_{(i,j+1)} \end{aligned} \quad (12)$$

Where $\alpha = 0$ indicates the Full state (otherwise $\alpha = 1$) and $\beta = 0$ represents the Empty state (otherwise $\beta = 1$).

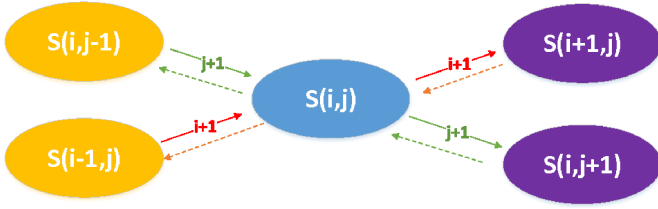


Fig. 5. Generic CTMC states, where s : state, i : number of ongoing service for H2H requests, j : number of ongoing service for M2M requests.

3) *Linear system solution*: The previous general equation (12) can be solved similar to section V-A3 while considering s_t as dimension for the steady-state probability vector Π and $(s_t) \times (s_t)$ as dimensions for the square matrix $\mathbf{B}$. s_t could be calculated as follows: $s_t = \frac{(c+1) \times (c+2)}{2}$

For example, for $c = 3$, the dimension for Π is 10 and for $\mathbf{B}$ is $(10) \times (10)$.

Because $0 \leq i + j \leq c$, we can consider the upper triangle part of the probability matrix $i \times j$ only:

$$P = \begin{bmatrix} 00 & 01 & 02 & 03 \\ 10 & 11 & 12 & 13 \\ 20 & 21 & 22 & 23 \\ 30 & 31 & 32 & 33 \end{bmatrix}$$

We build the steady-state probability vector Π ($s_t = 10$) as its dimension with the following order:

$$\Pi = [00 \ 01 \ 02 \ 03 \ 10 \ 11 \ 12 \ 20 \ 21 \ 30]$$

Then, we build the square $(10) \times (10)$ matrix $\mathbf{B}$ following the same order. Finally, using equation (11), we can solve the full rank matrix $\mathbf{B}$.

4) *Performance metrics*: We use two performance metrics in order to identify how far the degradation on the traffic could occur during emergency scenarios and the impact on the network. Using our proposed solution we show how congestion problems could be alleviated .

- *Service Completion Rate (S_c)*: It gives the number of completed requests per time interval and it is based on the service rate μ and the number of ongoing requests for a certain application (i):

$$S_c = \sum_{ij} i \mu \pi_{(i,j)} \quad (13)$$

- *Resource Utilization (R_u)*: This metric gives the probability of the system to be busy serving the arrivals in terms of the number of utilized PRBs in each state:

$$R_u = \sum_{ij} \frac{i+j}{c} \pi_{(i,j)} \quad (14)$$

In the next section, we translate the previous methodology into a case-study, then we simulate different scenarios with the aim to evaluate the A-eNB and its response throughout critical events.

VI. SIMULATIONS AND RESULTS

A. Case-study:

In real life, the emergency events such as terror attacks, natural disasters and multiple accidents are not predictable at all. With vague scenarios and lack of statistics and researches about the behavior of M2M devices throughout emergency scenarios, we built a case-study in our previous work [14] based on some use-cases such the one found in [5] and 3GPP technical reports [15] as summarized in Table I.

If we compare again the different previous storms with the maximum up-link data-rate in NB-IoT (150 Kbps), we conclude that the available bandwidth will suffer from a huge degradation which requires an effective solution for this problem.

B. Model accuracy

We construct many simulations using the model proposed in [12] which can generate similar traffic under the same conditions and parameters with some modifications to make it suitable to our scenarios. Then, an analysis for both analytical and simulation results are conducted in order to validate the accuracy of the model by calculating the Mean Absolute Error (MAE): $MAE = \frac{1}{n} \sum_{t=1}^n |x_{c(t)} - x_{s(t)}|$ and the Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_{c(t)} - x_{s(t)})^2}$$

The comparison reveals that the proposed model results fall into the acceptable error limits comparing to the analytical results with a $MAE_{max} = 0.0592$ when $c = 6$ and a $MAE_{max} = 0.0177$ when $c = 25$.

C. Simulation parameters:

The purpose of this section is to evaluate the A-eNB by measuring the S_c for different arrival rates in different cases. To this end, we use the simulation model presented in [12] to create different scenarios that validate our work.

Based on our case-study in section VI-A, in which we consider three different groups selected according to the parameters set to different models, we turn the group storms into different average arrival rates $\lambda_1 = 1.6$ for Group1 storm, $\lambda_2 = 3.2$ for Group2 storm, $\lambda_3 = 6.4$ for Group3 storm and $\lambda_w = 11.2$ for worst-case storm (the three storms simultaneously).

We assume that all storms have the same service rate: $\mu_1 = \mu_2 = \mu_3 = \mu_w = 1$ in order to focus on the other parameters. Additionally, we assume that the observation time intervals are small enough to have one arrival or departure at one time interval with a total simulation time 1000 time intervals.

D. Comparison and result discussions:

1) *Using an eNode-B*: To mimic a disaster scenario similar our use-case, we consider only a regular eNodeB with a fixed number of resources ($c = 1$) reserved for M2M traffic, while considering different storms with $\lambda_1 = 1.6$, $\lambda_2 = 3.2$, $\lambda_3 = 6.4$ and $\lambda_w = 11.2$. In the results shown in Fig. 6, we realize a huge degradation in the service completion

rate when moving from Group1 storm ($\lambda_1 = 1.6$) with a 56% completion rate till reaching the worst-case scenario ($\lambda_w = 11.2$) with only 9% completion rate.

2) *Using an A-eNB*: To test the adaptability of our A-eNB during disaster scenarios, we consider the same different storms with $\lambda_1 = 1.6$, $\lambda_2 = 3.2$ and $\lambda_3 = 6.4$ while enabling the flexibility of the bandwidth till it absorbs all previous storms. The results shown in Fig. 7 reveal that by adapting the NB-IoT bandwidth from ($c = 1$) used usually in a regular eNodeB, to ($c = 10$) promoted in an A-eNB, we can eliminate all M2M storms with a completion rate of 98%.

E. Worst-case scenario:

Due to the synchronized behavior of M2M devices, we expect to have worst-case scenarios in which the three M2M storms influence the network simultaneously with a $\lambda_w = 11.2$. By simulating this case, the results spot that the service completion rate reach its lowest value with 9%. Meanwhile, it peaks to 98% when implementing the A-eNB while ($c = 18$) only.

F. M2M and H2H coexistence scenarios:

After the validation of the proposed model for a single traffic, we extend our exploration to include M2M and H2H traffics, with the aim to study the mutual impact of both traffics in normal, dense area and emergency scenarios.

The proposed simulation is based on the simulation model proposed in [12], but with the following modifications:

- We assume to have one LTE-A network which serves both H2H and M2M traffics. Each traffic has an average arrival rate (λ_1, λ_2) with a service rate (μ_1, μ_2) for each traffic respectively.
- Both traffics have the same priority.
- A bandwidth of 5 MHz ($c = 25$) is used in order to stress the traffics to the maximum.

1) *Normal Scenario*: In rural and suburb areas, we assume to have low requests for both M2M and H2H devices by considering the following system parameters: $0 < \lambda_1 \leq 2$, $\lambda_2 = 1$, $\mu_1 = 0.5$ and $\mu_2 = 1$.

From the results shown in Fig. 8-(a), we conclude that having 25 RBs per time interval is more than enough for such

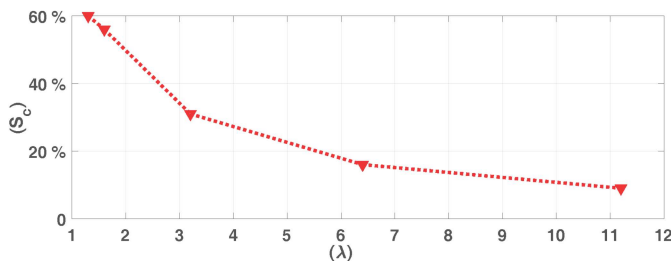


Fig. 6. Overload problem in a regular eNodeB during emergency scenario while ($c = 1$), c : represents the number of system resource, S_c : Service completion rate, λ : average arrival rate.

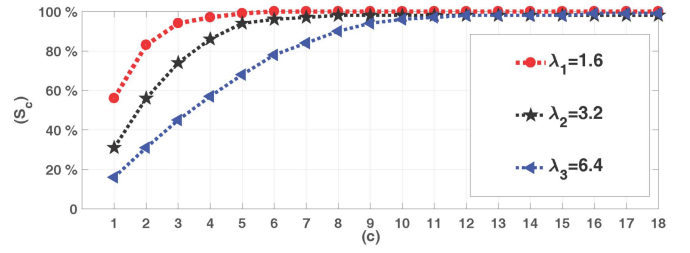


Fig. 7. Resolving M2M storms by an A-eNB during an emergency scenario with an adaptive number of system resources (c), where: A-eNB: Adaptive eNodeB, λ_1 , λ_2 and λ_3 represent the arrival rates for Group1, Group2 and Group3 storms.

arrival rates. Moreover, the resource utilization reaches 20% only, which shows the ability of the system to receive much more requests with no noticeable shortage.

2) *Dense Area Scenario*: In urban and dense areas, a huge number of M2M and H2H devices are exchanging their payloads. Consequently, to study the impact of M2M traffic on H2H traffic we consider a fixed arrival rate of H2H requests $\lambda_1 = 5$ and incremental arrival rate of M2M requests $0 < \lambda_2 \leq 20$, while $\mu_1 = 0.5$ and $\mu_2 = 1$. From the results shown in Fig. 8-(b), we can conclude the following:

- *M2M Service Completion Rate ($M2M-S_c$)*: The system is not able to serve all of M2M requests when $\lambda_2 \geq 15$. Meanwhile, at the peak, only 80% from the M2M requests are served.
- *H2H Service Completion Rate ($H2H-S_c$)*: The H2H do not suffer from any degradation till $\lambda_2 \geq 15$. At the peak, 20% from the H2H requests aren't served.
- A high utilization of the system ($R_u = 90\%$) which requires a queuing strategy to fulfill the excessive requests and to minimize the number of forced terminated services.

If we want to study the impact of H2H traffic on M2M traffic by considering the same parameters, it is obvious to have the same results because H2H and M2M devices have the same priority.

3) *Emergency Scenario*: In emergency cases, a M2M signaling storm is definitely expected as a result of connected devices. This congestion caused by the simultaneous synchronization affects both H2H and M2M devices. In this scenario, we consider the following system parameters:

$$\lambda_1 = 50, 50 \leq \lambda_2 \leq 200, \mu_1 = 5 \text{ and } \mu_2 = 1.$$

From the results shown in Fig. 8-(c), we conclude the following:

- *M2M- S_c or H2H- S_c* : The results shows that at the beginning only 22% of the requests are served. Meanwhile, at the peak only 12% from the requests are served.
- The system utilization peaks to its cut-off point with a $R_u = 100\%$.

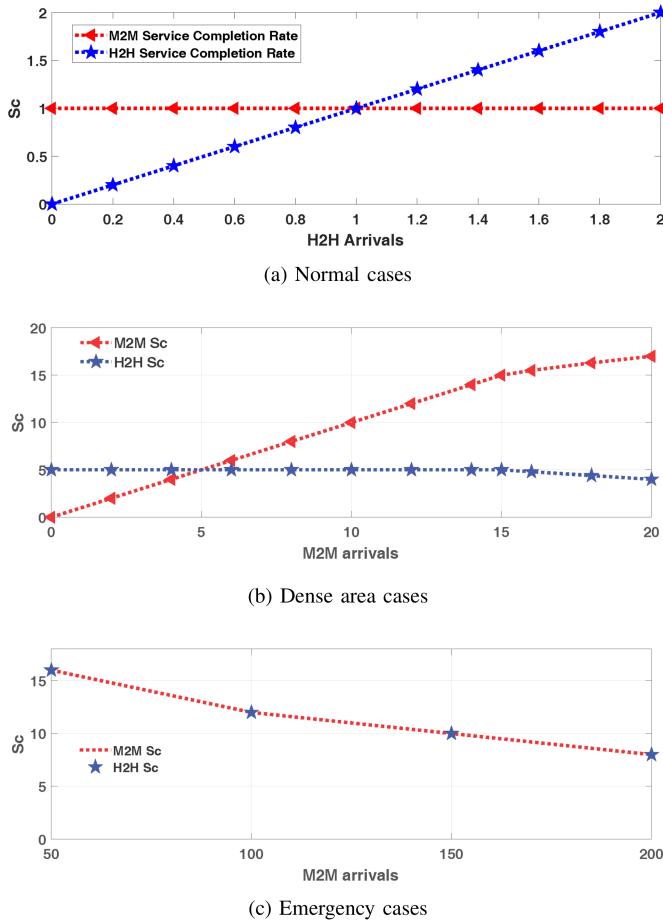


Fig. 8. M2M Service Completion Rate ($M2M-S_c$) vs H2H Service Completion Rate ($H2H-S_c$) in different scenarios.

VII. CONCLUSION

H2H and M2M traffic coexistence sharing the same LTE-A networks becomes a hot topic due to their different specificity and functionality. Consequently, mutual impacts on both M2M and H2H traffics become unavoidable. 3GPP addresses this problem with its new NB-IoT technology dedicating a limited bandwidth for M2M traffic. Unlike H2H communications, M2M devices transmit small-sized packages in different time intervals in form of synchronized storms. Therefore, one of the prominent challenges facing mobile operators is the fast saturation of the limited bandwidth with the rising of M2M devices especially in emergency cases where all these devices are attempting to send their payloads simultaneously and causing inevitably overload congestion problems.

In this manuscript, we have proposed a new concept called A-eNB as an extension to the classical eNodeB in NB-IoT networks. Our A-eNB keeps broadening the NB-IoT bandwidth gradually while shrinking LTE-A bandwidth till soaking-up all M2M storms. Moreover, this solution provides an essential resolution of presumable M2M storms especially during disaster and emergency events.

In order to study M2M and H2H traffic coexistence, a Continuous-Time Markov Chain (CTMC) is proposed to model the system behavior and study the impact on the traffics and the network.

Our simulation results show that the network will be facing a huge degradation in the service completion rate when using the classical eNodeB. This degradation reaches 9% in the worst-case scenario ($\lambda_w = 11 \cdot 2$). But using our A-eNB solution and by leasing a maximum 18 PRBs for M2M traffic of the total bandwidth in LTE-A/NB-IoT networks, can result a completion rate of 98% on all simulated M2M storms throughout emergency scenarios.

In our future work, we tend to add different priorities for both M2M and H2H traffics such as video-streaming, voice over IP, file transfer, etc. By measuring the mutual impacts, we are planning to add queuing strategy to maintain the QoS of M2M and H2H traffics.

REFERENCES

- [1] A. Mansour, R. Mesleh, and M. Abaza, "New challenges in wireless and free space optical communications," *Optics and Lasers in Engineering*, vol. 89, pp. 95–108, 2017.
- [2] I. Gudkova, K. Samouylov, I. Buturlin, V. Borodakiy, M. Gerasimenko, O. Galinina, and S. Andreev, "Analyzing impacts of coexistence between M2M and H2H communication on 3GPP LTE system," in *International Conference on Wired/Wireless Internet Communications*, pp. 162–174, Paris, France, May, 2014.
- [3] ERICSSON. (2015) Ericsson mobility report, on the pulse of the networked society. Accessed: 20-May-2017. [Online]. Available: <http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf>
- [4] S. M. Oh and J. Shin, "An Efficient Small Data Transmission Scheme in the 3GPP NB-IoT System," *IEEE Communications Letters*, vol. 21, no. 3, pp. 660–663, March 2017.
- [5] 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)*, Montreal, Canada, pp. 1–5, September, 2016.
- [6] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT system for M2M communication," in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–5, Doha, Qatar, April, 2016.
- [7] Evolved Universal Terrestrial Radio Access and Evolved Universal Terrestrial Radio Access Network, Technical specification, Release 13, 3GPP TS 36.300, Stage 2, version 13.4.0, August, 2016.
- [8] Radio Resource Control (RRC), Protocol specification, Release 13, 3GPP TS 36.331, v13.2.0, June 2016.
- [9] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-Efficient Resource Allocation for D2D Communications Underlying Cloud-RAN-Based LTE-A Networks," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 428–438, June 2016.
- [10] Q. Zhang, L. T. Yang, Z. Chen, P. Li, and F. Bu, "An Adaptive Dropout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2018.
- [11] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS Long Term Evolution*. Wiley Online Library, 2015.
- [12] M/M/1 Queuing System. (Accessed: 19-Dec-2017). [Online]. Available: <https://www.mathworks.com/help/simevents/examples/m-m-1-queuing-system.html>
- [13] G. Grimmett, D. Stirzaker, *Probability and random processes*. Oxford university press, Oxford, U.S.A., 2001.
- [14] A. H. El Fawal, A. Mansour, M. Najem, F. Le Roy and D. Le Jeune, "Adaptive LTE-M eNodeB for Emergency Scenarios," *The 8th International Conference on Information and Communication Technologies Convergence ICTC*, Jeju Island, Korea, October 2017.
- [15] "Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things, Technical Report, 3GPP TR 45.820, August, 2016." Tech. Rep.