Prioritized Random Access with Dynamic Access Barring for RAN Overload in 3GPP LTE-A Networks

Jen-Po Cheng and Chia-han Lee
Research Center for Information Technology Innovation
Academia Sinica
Taipei, Taiwan

Tzu-Ming Lin
Informations and Communications Laboratory
Industrial Technology Research Institute
HisnChu, Taiwan

***Abstract*—The rapid growth in the number of machine-type communications (MTC) devices causes the radio access network (RAN) to overload when a large number of MTC devices try to access the radio resources in a very short period. A Prioritized Random Access (PRA) scheme is proposed to efficiently solve the RAN overload problem and provide quality-of-service (QoS) for different classes of MTC devices in 3GPP LTE-A networks. This is achieved by pre-allocating random access channel (RACH) resources for different MTC classes with class-dependent backoff procedures and preventing a large number of simultaneous RACH attempts by using dynamic access barring (DAB). Simulation results show that, unlike Extended Access Barring (EAB) and the current LTE-A medium access control (MAC) scheme, the proposed PRA architecture with the DAB scheme guarantees QoS, i.e., high success rate and low access delay, even under the worst case RAN overload in LTE-A networks.**

I. Introduction

The number of devices with machine-type communications (MTC) is going to be much larger than that of human-to-human (H2H)-based communications in the near future due to the paradigm shift in modern telecommunications towards the Internet of Things [1]. MTC, also known as machine-to-machine (M2M) communications, enables many applications such as domotics, e-health, and smart grid without human interposition. However, those applications will urge for new approaches to efficiently support such huge number of devices since current cellular networks are designed for a moderate amount of H2H traffic targeting at voice transmissions.

A technical report addressed by the 3rd Generation Partnership Project (3GPP) has shown a predicted number of MTC devices such as smart meters to exceed 30000 per cell, compared to 50 per cell in the average number of active H2H devices [2]. As a result, congestions in different aspects of cellular networks, e.g., radio access network (RAN), core network (CN), and signalling network, are caused by a high number of MTC devices trying to access the radio resources simultaneously [3]. For example, some metering devices report recurring data at synchronous time intervals to support centralized entities. This concern has recently attracted serious attention in the standardization progress of 3GPP Long Term Evolution-Advanced (LTE-A). In order to enable M2M communications, some primitive issues and potential solutions are being updated in 3GPP technical reports [3], [4], but, still none of them provides efficient solutions. Also, in [5], an early effort to improve the MTC performance under LTE-A from physical layer (PHY), medium access control (MAC) layer, to CN was presented. However, this proxy-based MTC scheme is hardly compatible to the current release of LTE-A.

The RAN overload issue in 3GPP LTE-A is caused by many MTC devices trying to transmit data to a base station, called Enhanced Node B (eNB) in LTE-A, within a very short period of time. It is one of the key issues in MTC since the air interface is the first and the last miles of M2M communications. In [2], simulation results showed that the access delay using the current LTE-A medium access control scheme may be unacceptable when the number of MTC devices exceeds 30000 per cell. For disaster alarms, even 0.01 seconds could be very critical. Nevertheless, few papers have investigated the RAN overload issues of M2M communications in 3GPP LTE-A networks. In [6], a grouping-based radio resources management was proposed based on the quality-of-service (QoS) requirement and the packet arrival rate of each MTC group. However, this scheme cannot handle a surge of MTC access attempts, which is typical in RAN overload scenarios, because of failing to consider the number of MTC devices in each group in their proposed control procedure. Currently Extended Access Barring (EAB) [7] is considered as the only scheme that could control the potential surge of access attempts, and therefore it may be introduced as a baseline solution to RAN overload in the future. The basic idea of EAB is that the user equipments (UE) belonging to certain access classes indicated by the network-broadcasting information are not permitted to access the network as long as EAB is activated. However, the enabling mechanism and practical barring procedures for EAB are still in development.

In this paper, a Prioritized Random Access (PRA) scheme combined with a novel dynamic access barring (DAB) is proposed to efficiently tackle the overload problems for MTC operating under 3GPP LTE-A networks. The rest of this paper is organized as follows. In Section II, the LTE-A system and the classification of MTC traffic are introduced. The detailed

978-1-4673-0040-7/11/$26.00 ©2011 IEEE

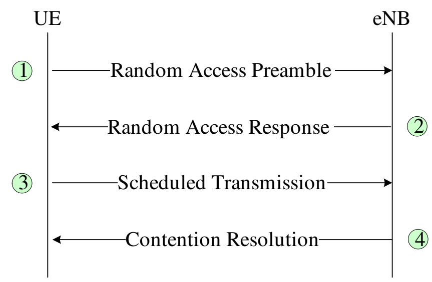

Fig. 1: Random access procedure in LTE-A [8].

design of the proposed PRA architecture is presented in Section III. The performance of the proposed PRA architecture is evaluated using computer simulation, and the results are presented and discussed in Section IV. Finally, Section V concludes this paper with future work.

II. System Setup

In this section, we briefly review the random access (RA) procedure in LTE-A and describe the RAN overload issue. An MTC traffic classification is proposed in order to help cope with this problem.

A. LTE-A Environments

Prior to sending data, MTC devices get attention of an eNB through the random access channel (RACH) by using the random access procedure defined in 3GPP LTE-A [8]. eNB constantly releases some RACH opportunities according to the amount of currently available radio resources. An MTC device then uses those RACH opportunities to ask eNB for radio resources by a four-step procedure as shown in Fig. 1.

The first step is for a UE to transmit a message called Msg1, which contains one of the 64 random access preambles generated by a Zadoff-Chu sequence for allowing those simultaneously transmitted Msg1's to be decoded by eNB [9]. Then, the UE waits to receive the random access response (RAR) within the RA response window. After successful decoding of Msg1's, eNB replies an RAR containing an RA-preamble identifier or a Backoff Indicator subheader called Msg2 at the second step. If the RAR contains an RA-preamble that matches the one it transmitted in Msg1, the UE then conveys Msg3 with a UE identifier to eNB. After receiving Msg3, eNB replies Msg4 to confirm that the connection is successfully established and ends the RA procedure. For further details on the random access procedure in LTE-A, see [8].

A collision is detected by MTC devices if RAR is not received within the RA response window. This happens if two or more MTC devices select the same preamble such that eNB is unable to decode any of the preambles and will not send RAR [2]. If RAR is not received, UE shall randomly backoff for a period of time according to a value drawn uniformly between 0 and the Backoff Parameter Value. After that, the UE retransmits Msg1 until the maximum number of retransmissions is reached. Msg3 and Msg4 may not be received due to the imperfection of PHY, so they may be retransmitted for a given maximum number of hybrid automatic repeat request (HARQ) transmissions.

TABLE I: Classification of MTC traffic.

Class name	Application exemplar	QoS requirement
H2H	Voice call	Hardly effected by MTC
Low priority	Consumer electronics / fleet management	Strict delay
High priority	E-care	Strict delay
Scheduled	Smart meters	Delay tolerant
Emergency	Seismic alarms	Extremely short delay

Due to the limited number of preambles, a large amount of MTC devices would easily select the same preambles, causing RAN overload. Our proposed PRA scheme solves such server congestion successfully.

B. Classification of MTC Traffic

According to the RA procedure mentioned above, eNB does not know UE identifiers until Msg3's are successfully received. This makes eNB unable to give access priorities to UE's in order to guarantee QoS. To cope with this, we consider that MTC devices are classified into five categories: H2H, low priority, high priority, scheduled, and emergency as shown in Table I. Human-to-human communications are expected to be unaffected by MTC traffic. The category of emergency includes the traffic having the features of very-low frequency of occurrence, extremely short delay constraints, and high channel access success rate requirements. MTC devices in the third category have periodic, scheduled traffic with delay-tolerance such as in smart grid-related applications. A large number of these MTC devices reports data periodically to eNB during a short period of time, and this burst of MTC traffic is the main factor causing RAN overload. For applications with QoS constraints, usually tasks are classified into high or low priorities, which classes are also included here. Note that the scheme proposed in this paper can be easily extended to more priority classes.

III. The PRA Architecture

The PRA scheme solves the RAN overload problem and provides QoS for different classes of MTC devices. This is achieved by pre-allocating RACH resources for different MTC classes while preventing a large number of simultaneous RACH attempts. The proposed PRA architecture is composed of two main components: virtual resource allocation with class-dependent backoff procedures and dynamic access barring.

A. Virtual Resource Allocation

According to the MTC traffic classification mentioned in Sec. II-B, we pre-assign different amount of virtual resources for different MTC classes as shown in Fig. 2. Every slot in the diagram is a virtual resource mapped to a RACH opportunity, which may be released from different LTE-A subframes. For example, in Fig. 2 the scheduled class can access one RACH

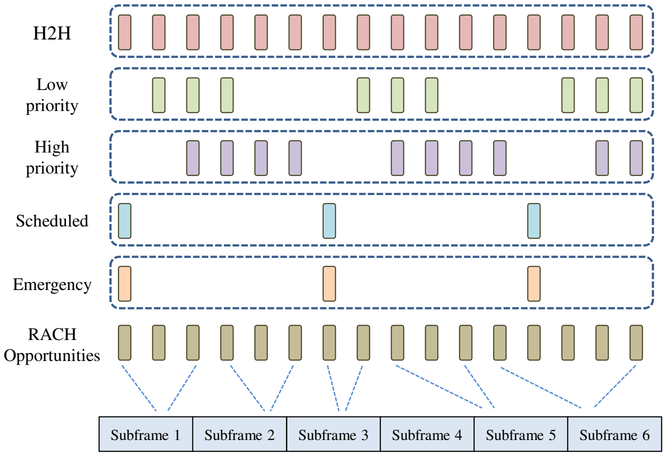

Fig. 2: Virtual resource allocations for five classes of MTC devices.

resource every six RACH opportunities. Virtual resources are assigned to MTC devices according to the following designs:

1) H2H can use all the available RACH opportunities.
2) The emergency and the scheduled traffics share the same virtual resource allocation.
3) Low/high priority traffic and emergency/scheduled traffics are assigned different virtual slots.

A dedicated resource allocation to emergency will result in extremely low utilization since emergency tasks are very rare. Therefore, we let emergency and scheduled traffics share the same virtual slots and then use back-off schemes to resolve collisions. In contrast, we separate resources of scheduled traffics and low/high priority traffics for the benefit of providing different techniques to delay-tolerant and delay-constrained MTC devices.

When failing to receive Msg2 containing its RA-preamble within the RA response window, a MTC device will detect a collision and back off for a random number N_B of virtual resources before retransmitting Msg1. N_B is designed as

$$N_B = \begin{cases} \mathrm{Uni}[0, 2^{\lambda} - 3] + 2, & \text{H2H, Scheduled,} \\ \mathrm{Uni}[0, 2^{\lambda} - 1], & \text{Emergency, L/H priority,} \end{cases} \tag{1}$$

where λ is the backoff exponent (note that each class has its own backoff exponent) and Uni$[a, b]$ means uniform distribution with support $[a, b]$. The backoff exponents are designed such that the emergency class has the highest priority among all MTC classes followed by the H2H class due to their short delay requirements. On the other hand, the classes of high priority, low priority, and scheduled employ exponential backoff.

In the practical deployment, eNB decides the virtual resource allocation according to the knowledge of the statistics of the MTC devices. This information may be stored in certain system information blocks (SIB) of the eNB, and a MTC device reads the SIB to acquire the updated realization of the virtual resource allocation. The SIB also allows emergency

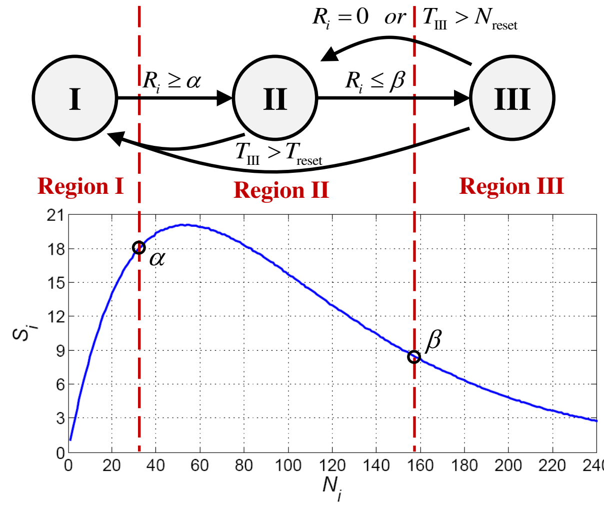

Fig. 3: State transition diagram with the enabling/disabling mechanism in DAB. T_{III} denotes the time spent in state II/III, $T_{\mathtt{reset}}$ the time to reset DAB, and $R_i = \frac{1}{2}(S_i + S_{i-1})$.

enhancement. After successful receiving an emergency alarm, eNB marks an emergency flag to 1 (the default is 0), which can easily be implemented by inserting one bit into the SIB, and resets the flag when sufficient information has been acquired. If the flag has value 1, an MTC device belonging to the scheduled class trying to transmit Msg1 for the first time should delay its access attempt for time T_1. Such mechanism prevents emergency traffic from being submerged by other MTC attempts of different classes and provides emergency traffic higher reliability and shorter delay.

B. Dynamic Access Barring for Collisions Avoidance

Although we have allocated different resources for different classes to reduce the chance of collision, the number of MTC devices can be too large for eNB to grant enough number of channel accesses in a short time. Thus dynamic access barring is proposed for collision avoidance.

To help describe the details of DAB, let $N_{i,k}$ be the number of RA attempts that have transmitted Msg1 k times until the i-th RACH opportunity, and L be the given maximum number of Msg1 retransmissions. The total amount of RA attempts during the i-th RACH opportunity can be expressed by $N_i = \sum_{k=0}^{L+1} N_{i,k}$. Then S_i, the expected number of successfully decoded Msg1 during the i-th RACH opportunity, is a function of N_i (as shown in Fig. 3) and used as the loading indicator.

The proposed DAB operates as follows. eNB continuously monitors S_i to determine the state of the current loading by using the state transition diagram shown in Fig. 3. In our design, there are three loading states, each of which indicates a level of loading from low (I), medium (II), to high (III). eNB marks a high-loading flag in the SIB to 1 (the default is 0) if and only if the current loading state is state III. This forces MTC devices transmitting for the first time to delay their access attempts for time T_1, thus reducing N_i.

TABLE II: Simulation parameters for LTE-A RACH [2].

Parameter	Setting
Cell bandwidth	5 MHz
Cell size (radius)	2 km
PRACH Configuration Index	6
Total number of preambles	54
Maximum number of Msg1 transmissions ($L+1$)	10
Number of UL grants per RAR	3
ra-ResponseWindowSize	5 ms
Backoff Parameter Value	20 ms
HARQ retransmission probability for Msg3 and Msg4 (non-adaptive HARQ)	10%
Maximum number of HARQ TX for Msg3 and Msg4 (non-adaptive HARQ)	5

The enabling mechanism of DAB works as follows. Let $R_i = \frac{1}{2}(S_i + S_{i-1})$. Starting from state I, the loading state transits to state II when eNB observes $R_i \geq \alpha$, and further transits to state III if after some time $R_i \leq \beta$, where α and β are some thresholds. DAB is disabled (the flag is reset to 0) when one of the three conditions is satisfied: after staying in state III for $N_{\texttt{reset}}$ slots, $R_i = 0$ is observed, or the total time staying in state II and III is more than $T_{\texttt{reset}}$. To further handle the worst case RAN overload, if having not transmitted Msg1 for more than $T_{\texttt{extra}}$, an MTC device transmitting for the first time delays its access attempt for time T_2, where $T_2 \gg T_1$. With DAB, eNB is able to control the amount of RA attempts in each RACH opportunity.

IV. Performance Evaluations

In this section, we present the simulation setup and the simulation results which show the superior performance of our proposed PRA scheme for solving the RAN overload problem.

A. Simulation Setup

Table II summarizes the basic LTE-A simulation parameters as defined by [2]. For the processing latency of the RA procedure, refer to Table B.1.1.1-1 in [10]. Under this setting, eNB releases one RACH opportunity every five subframes.

In order to evaluate our proposed PRA scheme, we introduce one MTC application for each MTC class: voice call, fleet management [2], hospital e-care, smart meters [2], and seismic alarms [2] as shown in Table III. The synchronization range is defined as the range between the first and the last MTC attempts. The shorter the range, the higher in intensity the generated RA attempts is. We refer the user generation rate of voice over IP (VoIP) in [11] as the intensity of RA attempts in the voice call applications. The urban London scenario is considered and all smart meters report within a 5 minute period. This scenario is chosen because it results in the worst case RAN overload. In every traffic cycle of smart meters, we let the seismic traffic randomly appears once within the synchronization range of smart meters, which is also the worst case setup. The node number of hospital e-care is the number of special hospice beds from one of the largest hospitals in Taiwan.

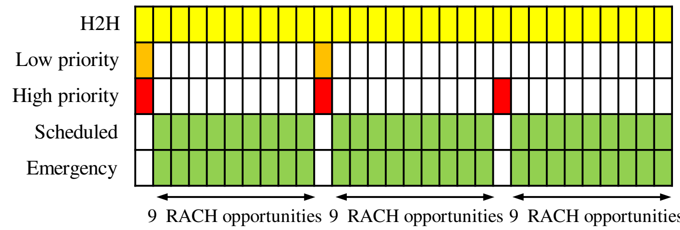

Fig. 4: Virtual resource allocation for simulation.

The PRA parameters are determined as the following. α and β are designed such that, when PRA is turned on, the number of new coming RA requests are no more than the number of channel accesses the system can grant. Then by computer simulations we can determine the value of $N_{\texttt{reset}}$. T_1 is then calculated as $N_{\texttt{reset}}\times$ra-ResponseWindowSize. $T_{\texttt{reset}}$, $T_{\texttt{extra}}$, and T_2 are parameters determined through observing simulation results.

Here we define an outage condition for MTC classes with QoS requirement of strict delay. For fleet management and hospital e-care, an access delay over 20 seconds is considered an outage since practically the data becomes useless after that.

B. Simulation Results

To evaluate the performance of the proposed PRA scheme, we consider the following performance metrics [2].

1) *Access success probability*, defined as the probability of successful completion of the RA procedure within the maximum number of Msg1 transmissions.
2) *Access delay*, defined as the delay for each RA procedure between the traffic arrival time and the completion of the RA procedure for the successfully accessed MTC devices.

The simulation results are shown in Table IV and Table V, respectively. As mentioned earlier, EAB is the best currently available solution to the RAN overload problem [7], so we compare the performance of PRA with EAB. Since EAB is still in development, we apply the settings which have been used during the standardization process of LTE-A [12]: EAB$(0.9, 4\mathrm{s})$, EAB$(0.7, 8\mathrm{s})$, and EAB$(0.5, 16\mathrm{s})$, where EAB(a, b) is defined as follows. An MTC device in the EAB barring class randomly selects a value between 0 and 1 every time before beginning the RA procedure. If the number is bigger than a, it shall back off for a period of time b and restart the EAB process until a value smaller than a is selected. We also use LTE-A as the baseline method.

By applying the virtual resource allocation in Fig. 4 with the proposed DAB, the PRA scheme achieves very high success probabilities in all MTC classes in Table IV, compared to the unacceptable low success probabilities of RA requests from both smart meters and seismic alarms using the LTE-A scheme. Since the LTE-A scheme performs well in other applications, we conclude that the RAN overload problem only happens during the synchronization range of the smart meters. Although EAB performs slightly better in smart meters and

TABLE III: Simulation setup.

MTC Class	H2H	Low priority	High priority	Scheduled	Emergency
Applications	Voice call	Fleet manag. [2]	Hospital e-eare	Smart meters [2]	Seismic alarms [2]
Node number	-	800	800	35670	126
Synchronization range	1 sec.	10 sec.	5 sec.	10 sec.	1 sec.
RA attempts per second	9	80	160	3567	126
Attempts distribution	Uniform	Uniform	Uniform	Beta(3,4)	Uniform
Traffic cycle	1 sec.	10 sec.	5 sec.	5 min.	Once
Maximum delay tolerant	-	20 sec.	20 sec.	-	-
EAB barring class	No	Yes	Yes	Yes	No
PRA Backoff indicator (λ)	3	1-4	1-4	3-6	1
PRA Exponential backoff	No	Yes	Yes	Yes	No
PRA($\alpha, \beta, N_{\mathtt{reset}}, T_{\mathtt{reset}}, T_1, T_2, T_{\mathtt{extra}}$)	(18, 8, 32, 8s, 160ms, 4s+Uni[0,1s], 5s)				

TABLE IV: Access success probability.

	Voice call	Fleet mana.	Hospital e-care	Smart meters	Seismic alarms
LTE-A	0.984	0.985	0.985	0.234	0.504
EAB(0.9,4s)	0.986	0.984	0.977	0.301	0.667
EAB(0.7,8s)	0.990	0.915	0.787	0.586	0.763
EAB(0.5,16s)	1.000	0.730	0.733	0.750	1.000
PRA	0.998	1.000	1.000	0.939	1.000

TABLE V: Average delay of *successful* transmissions (ms).

	Voice call	Fleet mana.	Hospital e-care	Smart meters	Seismic alarms
LTE-A	18	19	19	46	20
EAB(0.9,4s)	18	462	396	1286	21
EAB(0.7,8s)	18	1855	401	4671	21
EAB(0.5,16s)	18	1393	466	5414	23
PRA	21	127	77	2937	17

seismic alarms, it gradually trades the success probabilities of fleet management and hospital e-care for that of smart meters from $a = 0.9$ to 0.5.

Meanwhile, Table V shows that our PRA has much shorter access delays in all MTC classes compared to EAB methods. It is because the proposed DAB scheme spreads the access attempt traffics more efficiently compared to the blind spreading of EAB. The EAB $(0.9, 4\mathrm{s})$ has shorter access delay than PRA in smart meters, but it suffers from poor success probabilities in both smart meters and seismic alarms.

In summary, unlike the LTE-A scheme resulting in unnecessary low delays in applications of fleet management, hospital e-care, and smart meters, our PRA scheme makes a common good performance in all MTC applications while satisfying their own QoS requirements. While EAB partially solves the smart meters problem in LTE-A by introducing longer delay, our proposed PRA scheme beats EAB in both success probability and latency.

V. Conclusions

In this paper, the RAN overload issue in LTE-A has been addressed. The Prioritized Random Access architecture for solving the RAN overload problem in 3GPP LTE-A networks has been presented. This scheme utilizes the different traffic characteristics of MTC services to allocate RACH resources with class-dependent backoff procedures in order to satisfy different QoS requirements. To overcome the worst case RAN overload, the dynamic access barring scheme has been proposed. It has been shown by computer simulations that PRA provides higher reliability and shorter delay for emergency and smart meter services compared to EAB and LTE-A. PRA also differentiates high and low priority traffics and maintains quality of H2H communications. In summary, the proposed PRA scheme is able to achieve high access success probability in all five different classes of services while maintaining reasonably low access delay. Furthermore, the proposed enabling/disabling mechanisms in dynamic access barring show a full potential to be applied to EAB.

Acknowledgements

This work was supported by Industrial Technology Research Institute of Taiwan under Grant J520001001.

References

[1] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[2] 3GPP TR 37.868 V0.7.0, “Study on RAN improvements for machine-type communications,” Oct. 2010.

[3] 3GPP TS 22.368 V11.1.0, “Service requirements for machine-type communications,” Apr. 2011.

[4] 3GPP TR 23.888 V1.3.0, “System improvements for machine-type communications,” Jun. 2011.

[5] Y. Chen and W. Wang, “Machine-to-machine communication in LTE-A,” in *IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, Sept. 2010, pp. 1–4.

[6] S.-Y. Lien, K.-C. Chen, and Y. Lin, “Toward ubiquitous massive accesses in 3GPP machine-to-machine communications,” *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[7] 3GPP TSG WG2, “R2-113650: Report of 3GPP TSG WG2 meeting no.73bis,” Apr. 2011.

[8] 3GPP TS 36.321 V10.0.0, “Evolved universal terrestrial radio access (E-UTRA) medium access control (MAC) protocol specification,” Dec. 2010.

[9] 3GPP TS 36.211 V10.1.0, “Evolved universal terrestrial radio access (E-UTRA) physical channels and modulation,” Mar. 2011.

[10] 3GPP TR 36.912 V10.0.0, “Feasibility study for further advancements for E-UTRA (LTE-Advanced),” Mar. 2011.

[11] M. Wemersson, S. Wanstedt, and P. Synnergren, “Effects of QoS scheduling strategies on performance of mixed services over LTE,” in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007)*, Sept. 2007, pp. 1–5.

[12] Alcatel-Lucent Shanghai Bell, “R2-105623: Comparison on RAN loading control schemes for MTC,” Oct. 2010.