An Efficient Scheme for MTC Overload Control based on Signaling Message Compression

Tarik Taleb[1] and Adlen Ksentini[2]
[1]NEC Europe Ltd., Heidelberg, Germany
[2]IRISA/University of Rennes 1, Rennes, France
Emails: tarik.taleb@neclab.eu; aksentin@irisa.fr

Abstract – Whilst Machine Type Communication (MTC) represents an important business opportunity for mobile operators, mobile operators fear the congestion that could come with the deployment of billions/trillions of MTC devices. In this paper, we present a new mechanism that anticipates system overload due to MTC signaling messages in 3GPP networks. This mechanism proactively avoids system congestion by compacting the signaling message content for a group of MTC devices sharing redundant Information Elements (IE) by creating a profile ID for this group. Furthermore, along with this solution, we propose a dynamic grouping solution, which groups MTC devices with common subscription features in order to control the MTC signaling traffic when the network is overloaded. Simulation results show that compared to the Access Class Baring (ACB) mechanism, introduced by 3GPP, our proposed solution can avoid system overload without dropping MTC signaling messages, which is highly beneficial for MTC applications requiring service reliability.

I. Introduction

Deploying Machine Type Communication (MTC) over cellular networks offers several advantages not only for the Mobile Network Operators (MNO), but also for MTC application developers and providers. Whilst for MNOs, deploying MTC applications would generate new revenue streams, for MTC providers it gives the opportunity to target a larger population of users, including mobile users. With the intention to exploit the potential opportunities raised by a global MTC market over cellular networks, 3GPP groups are defining 3GPP network and system improvements that support MTC in the Evolved Packet System (EPS) [1]. However, the main challenge associated with the deployment of MTC over 3G/LTE is the support of high load, particularly signaling introduced by a potential number of MTC devices. System overload may occur at both the Radio Access Network (RAN) and the Evolved Packet Core (EPC), due to simultaneous signaling messages from many MTC devices. This situation may have a tremendous impact on the operations of a mobile network. Signaling congestion (overload) may happen due to a malfunction in the MTC server (e.g., MTC devices rapidly trying to reconnect to a remote server which is down) or application (e.g., synchronized recurrences of a particular procedure in the application) and/or due to massive attempts from a potential number of MTC devices to attach/connect to the network all at once [2]. Most signaling congestion avoidance and overload control mechanisms proposed in the context of MTC over cellular networks implement one of the following approaches: (i) segregate MTC traffic from the normal UE (User Equipment) traffic in order to prioritize the network access for the two UE types; i.e., this helps to anticipate the congestion which may happen due to MTC traffic; (ii) when congestion occurs, apply some back-off mechanisms rejecting MTC traffic at the RAN equipment (eNB) or at the EPS nodes (e.g., Mobility Management Entity (MME), Serving Gateway (S-GW) or even Packet Data Network Gateway (P-GW)).

Similar in spirit to the concept of ROHC (Robust Header Compression), in this paper we present a novel mechanism to decrease the size of signaling messages of a group of MTC devices sharing common IEs by replacing these latter by a profile ID. As second step of this solution, a group of MTC devices that share common subscriber features is dynamically created in order to allow the network to have better control of MTC traffic when the system is operating under specific conditions. The key features of this solution are *i*) dynamic creation of a profile ID to characterize events/scenarios whereby a mass of messages with some common IEs are exchanged between two nodes; *ii*) storage of profile ID is temporary to make efficient usage of available resources; and *iii*) actual amount of data exchanged between nodes is reduced to cope with core network overload. The same solution also enables a dynamic creation of group ID for UEs with common subscription features and/or similar behavior towards network to optimize usage of network interfaces, to reduce amount of signaling, and to reduce processing load at core network nodes.
The remainder of this paper is structured as follows. Section II highlights some related work on MTC and 3GPP system overload. Section III presents the envisioned solution. Section IV evaluates the performance of the proposed solution and discusses the obtained results. Finally, the paper concludes in Section V.

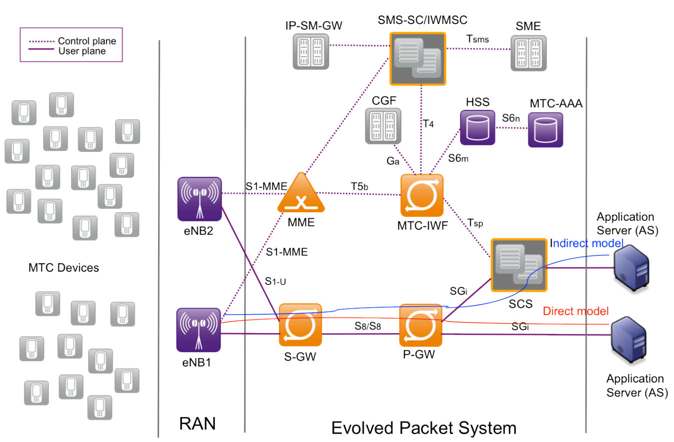

Figure 1. A typical MTC architecture as studied in 3GPP [1].

II. Related Work

Fig. 1 shows the MTC network architecture, as currently envisioned by 3GPP [1]. It consists of three main domains, namely the MTC device domain, the communication network domain, and the MTC application domain. In the network domain, most important nodes of a 3GPP EPS network are shown. Table 1 provides a brief description of the most important EPS nodes, shown in Fig. 1. The MTC application

978-1-4799-1353-4/13/$31.00 ©2013 IEEE

domain consists of MTC servers, under the control of the mobile network operator or a MTC provider.

Table 1. EPS's most important nodes.

Node	Description
eNB	Evolved Node B, the LTE's base station.
MME	Mobility Management Entity, a control plane entity for all mobility related functions, paging, authentication, bearer management in EPS.
MTC-IWF	MTC Interworking Function hides the internal Public Land Mobile Network (PLMN) topology and relays or translates signaling protocols used over Tsp to invoke specific functionality in the PLMN.
HSS	Home Subscriber Server, main database containing subscription-related information.
S-GW	Local mobility anchor for intra-3GPP handoffs.
P-GW	Packet Data Network Gateway, interfaces with the Packet Data Network (e.g., Internet).
SCS	Services Capability Server is the entity that connects MTC application domain to the network domain.

Two new entities related to MTC recently emerged in the 3GPP architecture. They are namely, MTC-IWF (Inter-Working Function) and SCS (Services Capability Server). A MTC-IWF may be a standalone entity or a functional entity of another network element. The MTC-IWF hides the internal PLMN (Public Land Mobile Network) topology and relays or translates signaling protocols used over the Tsp[1] interface to invoke specific functionality in the PLMN. SCS is an entity that connects to the 3GPP network to communicate with MTC devices and the MTC-IWF entity.

To differentiate MTC traffic from the classical traffic, most proposed solutions group the MTC devices into groups or clusters according to different metrics/features (e.g., low mobility, QoS requirement [10], belonging to macro or femtocell [9]). After grouping the MTC devices, there are two methods for prioritizing the access to the RAN, and avoiding the RACH (Radio Access CHannel) overload. The first one consists of defining "grant time periods" when MTC devices are authorized to connect to the network. The second method consists in defining specific low levels parameters for separating RACH resources for MTC and non-MTC devices. In addition to grouping MTC devices, there are also other solutions to anticipate the system overload by rejecting MTC device attach requests if there are not sufficient network resources, or by grouping the signaling messages from a group of MTC in one common bulk signaling message. Indeed, in some situations, there is need to reduce the MTC traffic by a specific amount implementing admission control at eNBs or even at MTC devices [5]. Indeed, admission control can be activated at eNBs upon receiving a congestion signal from the EPS nodes (e.g., MME, HSS, etc). Alternatively, it can be communicated to the MTC devices as in the 3GPP Access Class Baring (ACB) solution. ACB is a solution, which effectively reduces the collision probability of transmitting the bulk of preambles at the same RACH resource. Based on the parameters broadcasted by eNBs, a UE determines whether it is temporarily barred from accessing the cell. An access class barring factor or access probability (p) determines the probability that access is allowed. If a random number n generated by the UE is equal to or greater than p, then access is barred for a mean access barring time duration. In legacy ACB scheme, there are 16 access classes. AC 0-9 represents normal UEs, AC 10 represents an emergency call, and AC 11-15 represents specific high priority services, such as security services, public utilities (e.g., water/gas suppliers). A UE may be assigned one or more access classes depending on the particular cell access restriction scheme. In [12], the EAB (Extended Access Baring) is introduced for MTC, whereby a higher value of p and access class baring duration could be assigned to MTC devices in order to reduce the contention on RACH resources, since the MTC devices will be likely blocked by the small probability p. Similar in spirit to this concept, the authors proposed a congestion-aware admission control solution in [8]. The proposed solution selectively rejects signaling messages from MTC devices at RAN following a probability that is set based on Proportional Integrative Derivative (PID) controller, and is derived at a particular EPS node (e.g., MME). To further reduce signaling generated by MTC devices, the authors also proposed in [13] an optimized triggering solution for low mobility MTC devices. This solution renders the triggering operation less costly in case of triggering low mobility MTC devices, by limiting the triggering operation to a specific network area and also by reducing the number of involved network nodes, mainly avoiding MME. Other solutions aiming for coping with signaling congestion control in case of MTC are discussed in [14].

III. MTC Overload Control based on Signaling Message Compression

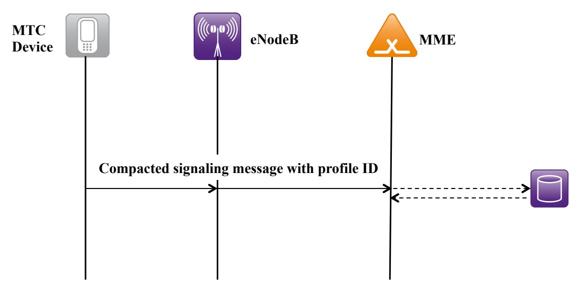

Figure 2. Basic idea behind the envisioned solution.

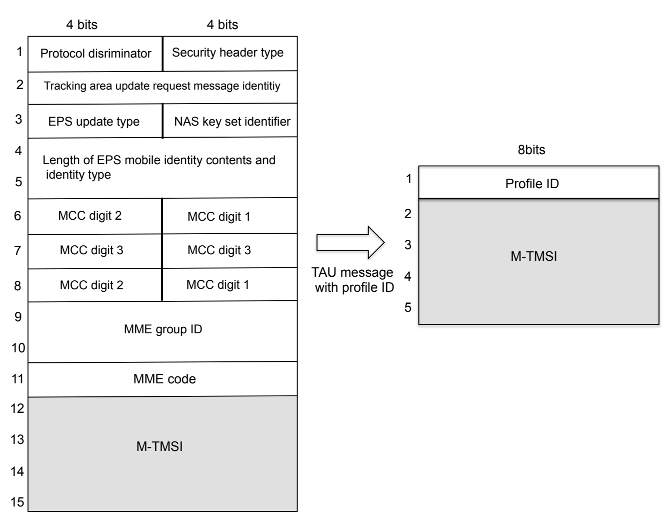

Figure 3. Example of profile ID creation: TAU message.

Stemming from the observation that there could be many scenarios whereby a mobile network node (e.g., eNB, MME, P-

[1] Names of interfaces and nodes may change as MTC relevant standards are evolving.

GW, MTC-IWF, etc) needs to deal with sending a mass of messages (i.e., at both user plane and control plane) from a group of UEs (e.g., MTC devices) with some IEs in common to the same target node at nearly the same time, in [5], the authors proposed holding such messages at a node for a specific time or till a number of messages is received, aggregating their content while avoiding duplicate IEs, and handling them in bulk to the receiving node. In this way, the message content can be considerably compacted. Moreover, the effort of parsing the parameters of many messages is also reduced to a minimum, which shall reduce by a large factor the time spent for processing the messages at the receiving node. With this regard, it shall be noted that IETF has initiated some activities on handling diameters messages in bulk [6] and they have a work charter dedicated for diameter maintenance and extensions [7].

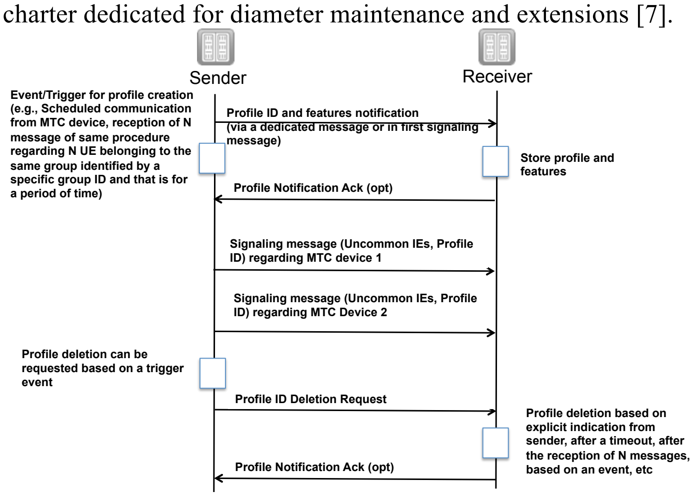

Figure 4. Compressing messages with common IEs using a unique profile ID created and managed following a specific logic using the group ID of relevant UEs.

Whilst handling messages with common IEs in bulk have advantages, its main drawback is the delay it adds in processing the messages. The purpose of the solution proposed herein is to achieve the goal of the “bulk message handling solution” in reducing the amount of traffic sent on EPS interfaces, but without compromising the delay in handling messages. The proposed solution defines methods for creating and managing, in a dynamic way, profile identifiers referring to a set of IEs that are common in messages relevant to a group of UEs or MTC devices, identifiable by a unique group ID, and replacing the common IEs with the created profile ID (Fig. 2). Before describing the core idea of the proposed solution, it shall be noted that there are many messages that can be subject to the idea of the solution. Considering messages on interfaces using Diameter and just to name a few, we can consider Update Location Request and Cancel Location Request as in [3], and CC-Request (Credit Control- Request – CCR) Command as in [4]. From the format of these messages, it can be said that for UEs belonging to the same group, it is likely to have many common IEs regarding these UEs and/or their bearers. Creating a profile ID to refer to these common IEs and sending it instead of all common IEs would definitely reduce the amount of traffic exchanged on the respective interfaces. The importance of replacing common IEs in messages by a profile ID, similar in spirit to ROCH, becomes more significant knowing that the size of messages is increasing with every release of the specifications. Fig. 3 depicts the case of creating profile ID for a Tracking Area Update (TAU) message. In the current 3GPP specification, a TAU message consists of mandatory fields, worth 15 bytes, and a set of optimal fields. If we focus only on MTC devices that are associated with the same MME, the only parameter that is device specific is the M-TMSI (MME Temporary Mobile Subscriber Identity). Therefore the other IEs are common and can be grouped and replaced by a profile ID. The idea of signaling message compression can be applied between any two entities that exchange messages between them over a particular interface and as part of a particular procedure, as shown in Fig. 4.

In Fig. 4, based on a particular trigger or event, such as 1) scheduled communication from MTC devices (or certain applications of smart phones known by the network a priori) or 2) reception of a number of messages of the same procedure with common IEs and regarding a number of UEs belonging to the same group identified by a specific group ID and that is within a period of time; the sender creates a profile identified by a unique ID referring to a set of attributes (e.g., common IEs). The profile ID can be a random value or a function of the group ID of relevant UEs and other metrics. The group ID can be explicitly indicated in the messages, inferred from the identifiers (and/or other information elements) of relevant UEs, inferred from subscription data of UEs downloaded on demand or a priori from HSS or another relevant node, or inferred from a mapping between the relevant procedure and the locations (e.g., cells, tracking areas, service areas, etc.) of the relevant UEs.

As second step, the sender communicates the profile ID and its features to the receiver, optionally along with instructions on when to delete the profile at the receiver, event type, etc. This notification can be either in the form of a dedicated signaling message or it can be inserted in the first relevant message sent after the profile creation. The profile notification message can be optionally acknowledged by the receiver. In response, the receiver stores the profile ID and its attributes. For the subsequent messages relevant to the profile, the sender does not insert common IEs; instead it inserts only the profile ID. In this way, the amount of communication on the interface between the two entities can be reduced. Storage of the profile and relevant information at the receiver and/or sender can be deleted either via a dedicated signaling message or a trigger can be sent in the last message relevant to the profile sent from the sender to the receiver. Alternatively, the receiver/sender can delete information on the profile after a timeout during which no relevant message is received, after a specific timeout initially indicated by the sender or a third party, or after receiving a total number of messages, or based on an event detected by the receiver/sender.

Using the same logic used in creating profile IDs, a node such as MME, SGSN, MSC/VLR can create unique group IDs, in a dynamic way, for UEs based on the frequency at which their relevant messages of a particular procedure (e.g., mobility management procedure such as attach request, RAU (Routing Area Update), LAU (Location Area Update), TAU requests; session management procedure such as PDN (Packet Data Network) connectivity request, activate PDP context request, etc) are transmitted. For instance, when a MME receives a number N of RAU request messages from N different UEs with a certain priority level during a specific period of time, the MME may create a group ID that will refer to this set of UEs and the ones that have similar subscription features and will be sending RAU request messages over another specific period of time. As a summary, in this proposed solution, based on subscriber profile received from HSS, MME or relevant core network node creates in a dynamic way a group ID to refer to UEs with particular subscription features and particular behavior towards network. MME or relevant core network node shares the group ID with other network nodes such as S-GW, P-GW, PCRF (Policy Charging and Rules Function), eNB, etc. These nodes make a local binding between the group ID and one or multiple UE's IDs. The group ID is used by eNBs for enforcing (extended) access class barring e.g., to deal with mobility management messages, by S-GW, P-GW, PCRF for e.g., handling session management messages and/or policy and charging control messages in bulk to optimize the usage of network interfaces.

IV. Performance evaluation

A. Simulation scenario

Table 2. Simulation parameters.

Parameter	Value
Number of macrocell	10
Number of MTC devices per group	300
Signaling message size (Byte)	350
Signaling message size after creation Profile ID	70
Buffer size at the MME	100

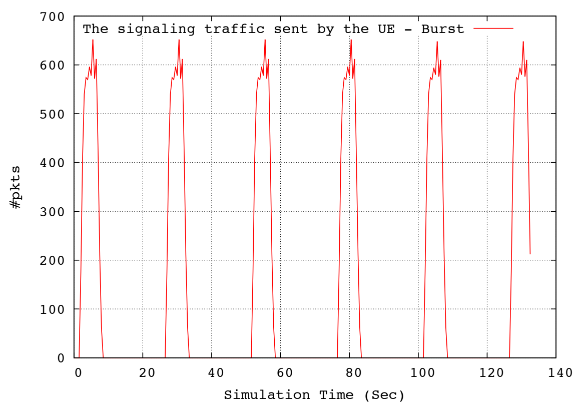

Figure 5. Traffic sent by one MTC device (burst traffic model).

We implemented the proposed solution using the NS3 simulator. Without any specific purpose in mind, we simulated a system of 1 MME, 1 S-GW and 10 eNBs. The MTC traffic is modeled as a bursty traffic (Fig. 4) and random traffic (Fig. 5) which represents the connection requests originating from the MTC devices. Burst traffic represents the case of MTC devices that periodically connect to report events to the remote server, while random traffic illustrates the case of deploying different independent MTC applications. Each cell contains the same number of MTC devices belonging to one group, which implies that each eNB will generate the same amount of traffic to the MME. Other simulation parameters are shown in Table 2.

We compare the performance of the proposed mechanism against the ACB mechanism proposed by 3GPP and the baseline solution whereby neither profile ID creation nor an admission control is enforced. It shall be recalled that ACB is using admission control at the RAN level, where MTC traffic is bared from accessing the channel with a probability p. For the simulation, two p values were considered, namely 0.1 and 0.3. The second value is more aggressive in rejecting MTC signaling messages at the eNB. In the proposed solution, we assume that the group is identified and the ID is attributed to this group.

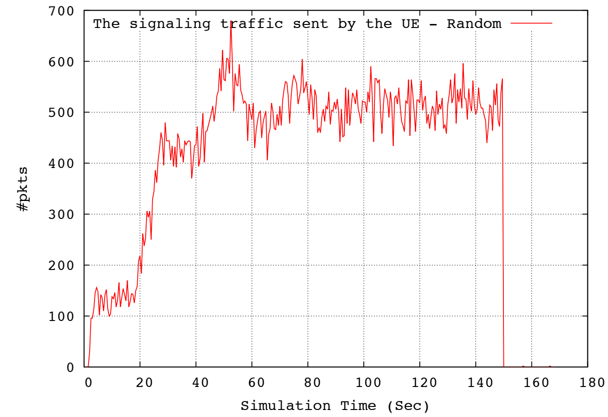

Figure 6. Traffic sent by one MTC device (random traffic model).

B. Results

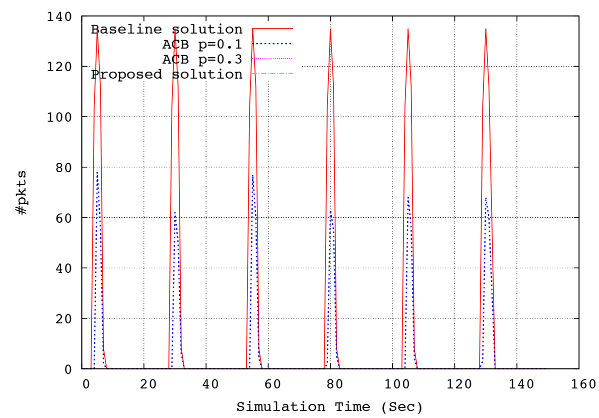

Figure 7. Packets dropped at the MME (burst traffic case).

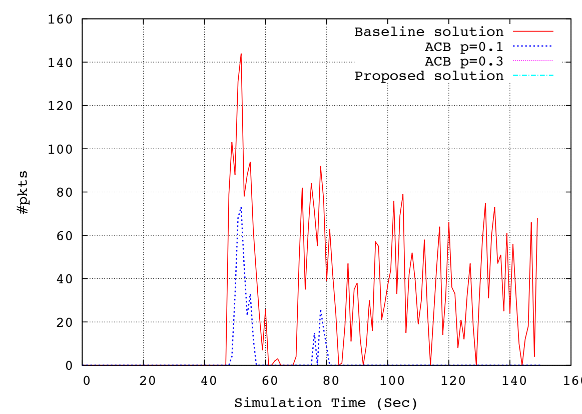

Figure 8. Packets dropped at the MME (random traffic case).

Figs. 7 and 8 show, for both traffic models, the number of dropped packets at the MME. For both models, we clearly see that most of the packets are dropped in the case of using the classical approach and ACB (p=0.1). This is attributable to the fact that both mechanisms fail in reducing the system load caused by the MTC signaling traffic. In ACB, setting p=0.1 is not enough to avoid dropped packets at the MME. In the contrary, both proposed solution and the ACB (p=0.3) mechanism avoid overloading the MME. The proposed solution succeeds in ensuring no system overload by reducing the size of the signaling messages to be carried by the core network nodes, while ACB (p=0.3) rejects the traffic at the *eNB* in order to reduce the signaling traffic in the core network.

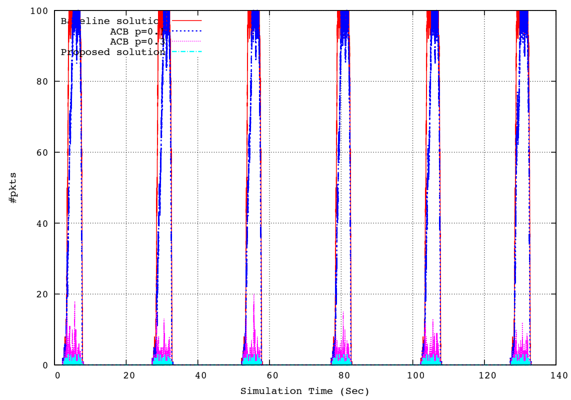

Figure 9. MME queue length (burst traffic case).

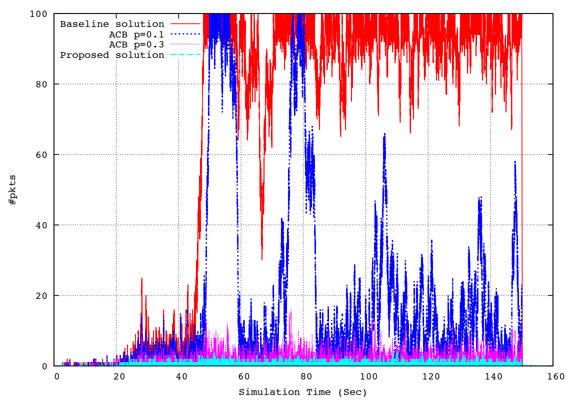

Figure 10. MME queue length (random traffic case).

Figs. 9 and 10 illustrate the evolution of the MME queue length throughout the simulation for both traffic models. As expected from the results of Figs. 7 and 8, both the baseline approach and ACB (p=0.1) maintain high queue length. In case of burst traffic, the queue size grows rapidly, causing frequent overflows and hence packets are dropped. Sometimes, the number of dropped packets exceeds 100. In case of random traffic, the queue size is high and frequently reaches the maximum in case of the baseline approach. In contrast, the proposed solution maintains low queue size, thanks to the creation of profile ID which reduces the size of signaling packets and hence enhances the MME service time.

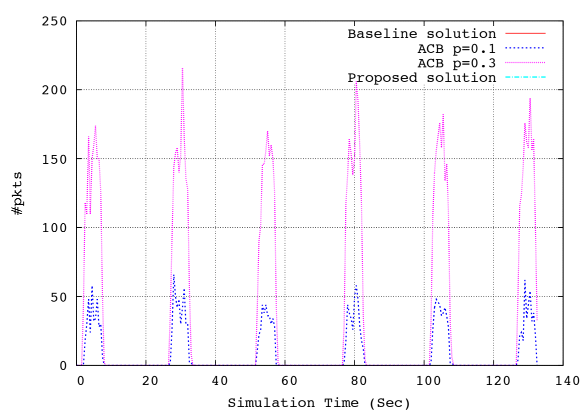

Figure 11. Dropped packets at the eNBs (burst case).

Fig. 11 shows the dropped packets at the eNBs. We observe that most of the dropped packets are experienced in the ACB mechanism, since ACB enforces admission control at eNBs. Accordingly, although the ACB (p=0.3) mechanism can avoid MME overload, it causes packet drops at the RAN level. In contrast, by reducing the message size with the creation of profile IDs, the proposed solution avoids MME overload without dropping MTC signaling packets, which is, in some cases, highly beneficial, e.g., for critical MTC applications whereby reliability is of utmost importance.

V. Conclusion

MTC defines a promising business for mobile operators. However, MTC deployment over 3GPP networks comes with important challenges. Of particular interest, mass access to the network from a potential number of MTC devices may overload EPS and degrades QoS for other ongoing services. In this paper, we introduced a solution to mitigate the issue of network overload due to MTC signaling. The proposed solution compacts the size of signaling messages sharing common information elements in order to reduce the communication and processing loads at the system interfaces and nodes, respectively. Simulation results showed that unlike ACB, the proposed solution can efficiently cope with the MTC overload without dropping MTC signaling messages, which is highly important for reliability-critical MTC applications.

References

[1] 3GPP TS 23.682 "Architecture enhancements to facilitate communications with packet data networks and applications" version 11.0.0, 2012-03-12.

[2] 3GPP TS 23.368 "Service requirements for Machine-Type Communications (MTC)"

[3] 3GPP TS 29.272 "MME and SGSN related interfaces based on Diameter protocol" 11.5.0, 2012-12-21

[4] 3GPP TS 29.212 "Policy and Charging Control (PCC) reference points" 11.7.0, 2012-12-20.

[5] T. Taleb and K. Samdanis, " Ensuring Service Resilience in the EPS: MME Failure Restoration Case," in Proc. IEEE Globecom 2011, Houston, Texas, USA, Nov. 2011.

[6] Mark Jones, "Diameter Group Signaling," IETF draft-jones-diameter-group-signaling-00, Nov. 2011.

[7] "Diameter Maintenance and Extensions (DIME) WG, http://datatracker.ietf.org/wg/dime/charter/.

[8] A. Ksentini, Y. Hadjadj-Aoul, T. Taleb, "Cellular-Based Machine-to-Machine: Overload Control," IEEE Network Magazine, Vol. 26, No. 6, Nov/Dec 2012 pp. 54-60.

[9] A-H. Tsai, L_C. Wang, J-H. Huang and T-M. Lin, "Overload Control for Machine Type Communications with Femtocells", in Proc. Of IEEE VTC Fall 2012, Quebec, Canada 2012.

[10] 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, April 2011.

[11] CMCC TSG R2-113197, "Performance comparison of access class barring and MTC specific backoff schemes for MTC", 3GPP meeting 23-27 August 2010.

[12] KPN TSG SA, "Problem of overload control with Machine Type Communications", 3GPP meeting 10-14 May 2010, Kyoto, Japan.

[13] T. Taleb and A. Ksentini, "On Alleviating MTC overload in EPS", in Elsevier J. on Ad Hoc Networks (to appear).

[14] T. Taleb and A. Kunz, "Machine Type Communications in 3GPP Networks: Potential, Challenges, and Solutions," in IEEE Communications Magazine, Vol. 50, No. 3, Mar. 2012.