Congestion-Aware MTC Device Triggering

Adlen Ksentini[1], Tarik Taleb[2], Xiaohu Ge[3] and Hu Honglin[4]
[1] IRISA/INRIA/University of Rennes 1, Rennes, France
[2] NEC Europe Ltd., Heidelberg, Germany
[3]Huazhong University of Science and Technology, Wuhan, China
[4]Shanghai Research Center for Wireless Communications, Shanghai, China
Emails: aksentin@irisa.fr; talebtarik@ieee.org, xhge@mail.hust.edu.cn, honglin.hu@wico.sh

Abstract –This paper describes a device triggering optimization technique for controlling system overload when deploying massive Machine Type Communication (MTC) devices in 3GPP-based cellular networks. Triggering a large number of MTC devices can dramatically overload the underlying transport network system and incur congestion in both the Radio Access Network (RAN) and the Evolved Packet Core (EPC). The proposed solution aims at controlling the rate of device trigger requests that MTC servers can generate in order to reduce the system overload. For this purpose, we propose that the Mobility Management Entity (MME), or an alike core network node, computes the device trigger rate that alleviates congestion, and communicates this value to the MTC-Interworking Function (MTC-IWF) element that enforces MTC traffic control, via admission control or data aggregation, on the device trigger request rate received from the different MTC servers. The proposed solution is evaluated through computer simulations and encouraging results are obtained.

I. Introduction

Due to the wide range of potential applications, Machine Type Communications (MTC) or Machine-to-Machine (M2M) communications are attracting the interest of mobile network operators, equipment vendors, MTC specialist companies, and research bodies. To facilitate convergence among these different stakeholders, different standardization groups started working on MTC. Among these groups, the 3GPP System Architecture working group (SA2) has been highly active in defining and determining 3GPP network- and system-level improvements to support MTC in the Evolved Packet System (EPS) [1]. The aim of these standardization activities is to identify 3GPP network enhancements required to support a large number of MTC devices in the 3GPP network domain and to provide necessary network enablers for MTC applications.

Although MTC defines promising business opportunities for mobile operators, a massive deployment of MTC devices, without an adequate engineering of their associated traffic and signaling, could result in potential congestion of mobile networks. Congestion could appear due to simultaneous signaling messages from MTC devices. Such situation occur when: (i) there is a malfunction in a MTC server; e.g., MTC devices trying to reconnect multiple times to the server which is down; (ii) massive attempts from MTC devices to attach/connect to the network all at once (e.g., to report the detection of a particular event); (iii) massive device trigger requests from different MTC servers. Aiming at alleviating congestion at mobile core network, the Core Network Overload (CNO) Study Item was initiated in 3GPP [2], providing system requirements and introducing a few solutions. Existing solutions for MTC signaling congestion avoidance and overload control could be classified into two classes:

- Proactive solutions, categorized, in turn, into two other categories. The first category solutions separating MTC traffic from the non-MTC traffic using techniques such as the grouping and clustering of MTC devices [3][4]. Network access is then differentiated for each MTC group allocating different “grant time periods” for each group or setting up different RAN access parameters in order to prioritize the non-MTC traffic. Approaches of the second category reduce the amount of signaling traffic by aggregating the MTC requests either at RAN (i.e., Bulk signaling) [5] or by creating profile ID [6] that replaces common Information Elements (IEs) used by a group of MTC devices, which allow reducing the size of signaling messages and accordingly associated processing load.
- Reactive solutions, which react to congestion, e.g., caused by several simultaneous MTC attach/connect requests, only after its occurrence rejecting (mainly through admission control [7][8]) or delaying these requests at RAN.

Most above-cited research work addresses “northbound” system overload that may arise from MTC devices towards mobile network. However, “southbound” system overload that may arise from MTC device triggers is not sufficiently addressed. Indeed, MTC-IWF (Inter-Working Function) is responsible for forwarding trigger requests from MTC servers to the core network, so that the core network pages the relevant MTC devices. Each device trigger issued by a MTC server is followed by a number of signaling messages exchanged among different EPC nodes (e.g., Short Message Service- Service Center (SMS-SC), Home Subscriber Server /Home Location Register (HSS/HLR), and MME). Thus, for a potentially large number of MTC devices, without an adequate engineering of the MTC device trigger, the operator network, particularly MME, may get overloaded. There is therefore need for appropriate mechanisms to mitigate system overload that may be due to device triggers.

In this paper, we address this issue by proposing a new solution to prevent network congestion that may occur when MTC devices respond to a trigger generated by the Services Capability Server (SCS), through MTC-IWF, which is connected to the MTC servers. The proposed solution consists of implementing a congestion prevention mechanism at MME, which notifies and instructs MTC-IWF to reduce the device trigger rate whenever core network is deemed to be congested by MTC device triggers. In response, MTC-IWF reduces the rate of the MTC device

978-1-4799-2003-7/14/$31.00 ©2014 IEEE

trigger by either rejecting SCS requests or aggregating these requests in bulk.

The remainder of this paper is structured as follows. Section II gives an overview of the MTC architecture and the device trigger procedure as currently specified by 3GPP SA2 group. 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.

II. Related Work

A. MTC architecture

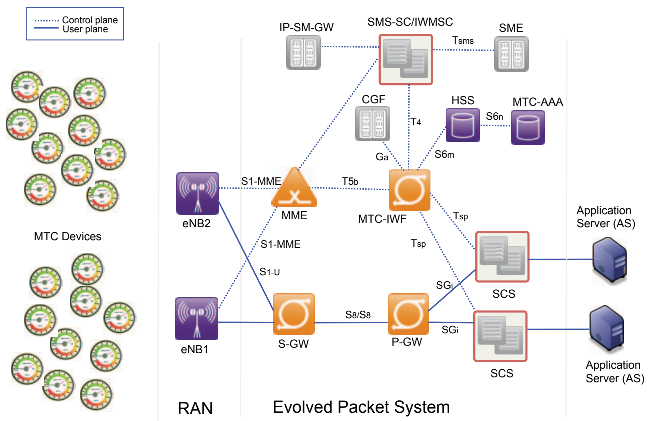

Figure 1. Envisioned MTC architecture as studied in 3GPP.

A simple illustration of the current 3GPP MTC network architecture [1] is shown in Fig. 1. Three distinct domains, namely the MTC device domain, the communication network domain, and the MTC application domain, constitute this architecture. Table 1 describes most important nodes that compose the 3GPP EPS (Evolved Packet System) network. The MTC application domain consists of MTC servers, under the control of the mobile network operator or an independent 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 MTC-relevant entities have been added to the 3GPP architecture: MTC-IWF and SCS. A MTC-IWF may be a standalone entity or a functional entity of another network element. MTC-IWF hides the internal PLMN (Public Land Mobile Network) topology and relays or translates signaling protocols used over the Tsp interface to invoke specific functionality in PLMN. SCS is an entity that connects to the 3GPP network to communicate with MTC devices and the MTC-IWF entity. Depending on the number of MTC servers and their locations, one MTC-IWF entity can be connected to more than one SCSs.

B. Device Triggering procedure

Based on the MTC architecture described in [1] and showed in Fig. 1, there are different trigger variants: (i) device triggering procedure over Tsp, which support trigger request delivery using T5 or T4 interfaces; (ii) device triggering using direct model over user plane; (iii) device triggering with OMA push. Note that the current version of the 3GPP specification does not detail the T5 triggering procedure. Hence, we will focus in this paper on the procedure using T4 trigger delivery and based on the Tsp interface. Fig. 2 shows the details of the Tsp-based MTC trigger procedure. For simplicity and without loosing generality, we omitted presenting few details. For more details, interested reader can refer to [1]. Security issues [9][10] are also not discussed in this paper.

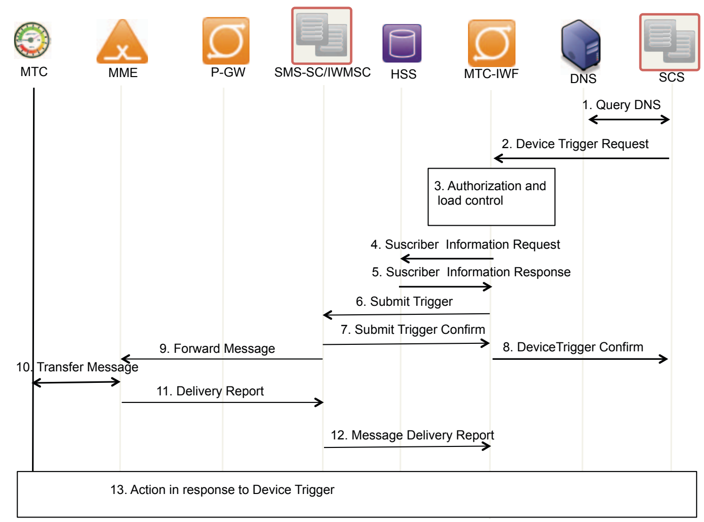

Figure 2. T4-based device trigger procedure.

Through SCS, the MTC server initiates a device trigger procedure for a given MTC device. If SCS does not know which MTC-IWF to contact, it performs a Domain Name Server (DNS) query using the External Identifier or using a locally configured MTC-IWF identifier to obtain the relevant IP address. SCS then sends the Device Trigger Request message to MTC-IWF including the following information: an external identifier or Mobile Station Integrated Service Digital (MSISDN), SCS identifier, trigger reference number, validity period, priority, Application Port ID and trigger payload. In the trigger payload, SCS includes information destined for the MTC application, and information to route the request message to the MTC application. The Application Port ID is set to address a triggering function within the MTC device. When receiving the request message, MTC-IWF checks whether SCS is authorized to send trigger requests. If this check fails, MTC-IWF sends back a Device Trigger Confirm message including the reason for the failure condition and the procedure stops at this step. Otherwise, MTC-IWF sends a Subscriber Information Request message to the HSS/HLR to (i) know if SCS is authorized to trigger the MTC device; (ii) resolve the External Identifier or MSISDN to International Mobile Subscriber Identity (IMSI); (iii) find the "routing information" stored at HSS, which include the identities of the MTC device's serving core nodes. After that, HSS/HLR sends back to MTC-IWF the Subscriber Information Response (i.e., IMSI and/or MSISDN and related "Routing information" including the serving node(s) identities, cause) message. If the cause value indicates that SCS is not authorized to send a trigger message to this MTC device (or group of MTC devices), or there is no valid subscription information, MTC-IWF sends a Device Trigger Confirm message with a cause value indicating the reason for failure and the procedure stops at this step. Otherwise, MTC-IWF initiates the T4 trigger delivery procedure, involving SMS-SC, P-GW and particularly MME in order to localize the MTC device. Finally, MTC-IWF sends the Device Trigger Report (External Identifier or MSISDN and trigger reference number) message to SCS with a cause value indicating the trigger delivery outcome (e.g. succeeded, unknown or failed and the reason for the failure). Upon receiving the trigger, the MTC device may initiate a communication with SCS or directly with the MTC server. In [6], the authors proposed an optimization of this procedure avoiding the involvement of MME for MTC devices with low mobility features, assuming the ability of MTC-IWF to page MTC devices. It is important to recall that each device trigger issued by a MTC server is followed by a number of signaling messages exchanged among different EPC nodes (e.g., SMS-SC, HSS/HLR, and MME), MME being the most impacted node. It is therefore the objective of this paper to mitigate congestion and system overload at the core network in the event of multiple MTC device triggers being issued by different MTC servers.

III. Proposed solution

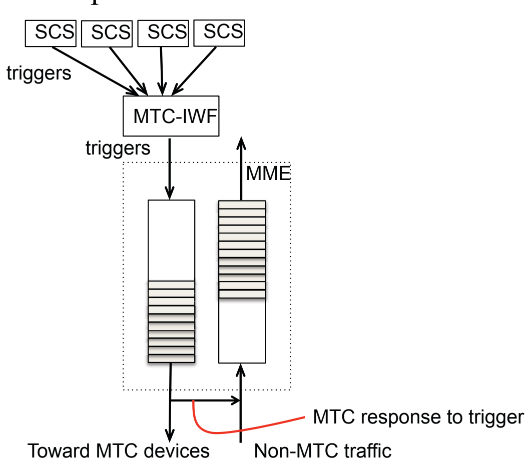

Figure 3. The queuing dynamics of the system composed by the MME and the MTC-IWF.

The key idea behind the proposed solution consists in implementing a congestion prevention mechanism at MME, which requests MTC-IWF for a reduction in the rate of MTC device triggers when it detects that congestion is about to occur. In response, MTC-IWF reacts by either rejecting SCS requests or aggregating these requests.

Fig. 3 represents the queuing dynamics of the system, particularly focusing on MME and MTC-IWF. We assume that MME maintains two separate queues, one for signaling traffic generated by UEs (including MTC devices) and another for requests generated from outside PLMN. Furthermore, we consider that after each successful MTC device trigger request, generated by MTC-IWF, a MTC attach/connect request is received by MME during the next period. Therefore, when $\lambda(t)$ MTC device triggers are successfully issued by MTC-IWF during a period t, MME will receive $\lambda(t)$ MTC attach requests from relevant MTC devices during the upcoming period $t+1$. This assumption helps in modeling and controlling the MTC uplink signaling traffic at MME, by controlling the device trigger rate at MTC-IWF.

In addition to the normal traffic (i.e., the non-MTC traffic, noted by ω), the total traffic received and processed by MME during the period t is $\theta(t)=\omega(t)+\lambda(t)$. Our aim is to control λ as we cannot control the regular traffic. According to Fig. 3, we estimate the number of packets in the queue at the beginning of the next period (t+1) as follows:

$$q(t+1)=\theta(t)+q(t)-\mu(t)$$

where $\mu(t)$ denotes the service rate, which is considered constant over time.

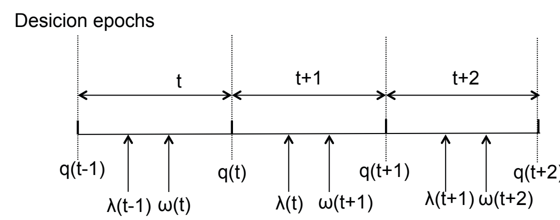

Figure 4. Detailed steps of the proposed solution.

It is worth noting that no information on the distribution of λ, ω and μ is required. In contrast, the only required information is the number of arrived packets during a period t. Denoting by q_0 the reference queue size, beyond which MME is deemed overloaded, the objective of the proposed mechanism is to control the rate of device triggers (λ) during the time interval t, so that the queue size $q(t+1)$ in the upcoming time interval is around the reference q_0. For this objective, we denote by $e(t+1)$ the error that needs to be minimized during the period $t+1$. It is computed as:

$$e(t+1)=q(t+1)-q_0$$

The objective of the proposed control mechanism is to keep the values of $e(t+1)$ in the vicinity of zero. Accordingly, we set $e(t+1)$ to zero and replace $q(t+1)$ by its formula, we obtain:

$$\theta(t)+q(t)-\mu(t)-q_0=0$$

As our aim is to control $\lambda(t)$, we need to formulate $\theta(t)$ in a way that ensures $e(t+1)=0$. Hence, we obtain:

$$\theta(t)=\mu(t)+q_0-q(t)$$

Given the lack of prior knowledge of the values of ω, we use a prediction mechanism based on the Exponential Weighted Moving Average (EWMA), which requires the two previous values of ω. The predicted value ϖ is obtained as follows:

$$\varpi(t)=\alpha\cdot\omega(t-1)+(1-\alpha)\cdot\omega$$

where α is a predetermined factor. Therefore, the values of $\lambda(t)$ that keep $e(t+1)$ in the vicinity of zero can be obtained as follows:

$$\lambda(t) = \begin{cases} \theta(t) - \varpi(t), & \varpi(t) < \theta(t) \\ 0, & Otherwise \end{cases}$$

Fig. 4 shows how the proposed solution functions. At the end of each period, MME computes $\lambda(t)$ that needs to be used by MTC-IWF during the next period. For instance, at instant (t) indicating the end of the first period, MME derives the value of $\lambda(t)$ which ensures that $e(t+1)$ is minimized at time instant $(t+1)$. This value is communicated to MTC-IWF through the T5a/T5b interface. According to the notified value, MTC-IWF rejects triggers if the total number of device triggers received from SCS exceeds $\lambda(t)$. Another way to reduce the number of triggers is by aggregating some of them and processing them in bulk.

IV. Performance evaluation

Table 2. Simulation parameters.

Parameter	Value
Number of macro cells	10
Number of SCS (attached to one MTC server)	10
Number of MTC devices per cell	1500
Period t	1 sec
λ_1 - both scenarios	6
λ_2 - first scenario/second scenario	60/30
μ	80
α	0.5
q_0	80
MME buffer size	100

In order to evaluate the performance of the proposed solution, we used a discrete event simulator based on C programming language. We simulated a system of one MME, one MTC-IWF, n eNodeBs, and n SCS. We assume that each SCS generates m device triggers following a Poisson distribution with intensity λ_1. The non-MTC traffic, coming from eNodeBs, is also assumed to follow a Poisson distribution with intensity λ_2. We envision two scenarios: (i) the first one considers that the ratio (denoted as ρ) between the generated SCS trigger requests and the non MTC signaling traffic is equal to one, i.e., the amount of MTC traffic is equal to that of non-MTC traffic; (ii) the second one considers that ρ=0.5, i.e., the amount of non-MTC traffic is twice that of MTC traffic. The simulation duration is set to 150 seconds; a duration long enough to ensure stable network performance. The duration of time intervals (t) is set to 1s; a period long enough to assess the rate of trigger requests that shall reduce imminent congestion at MME. Table 2 summarizes the simulation parameters.

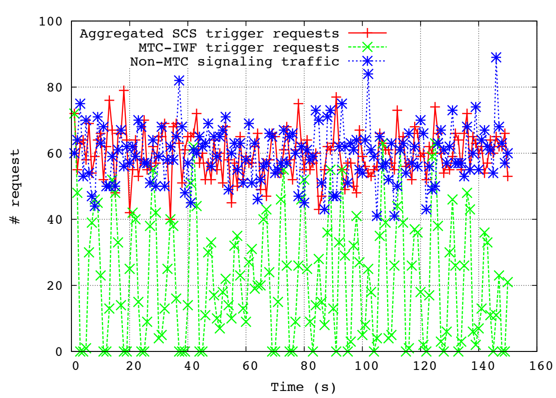

Figure 5. Trigger request and non-MTC signaling traffic – ρ=1.

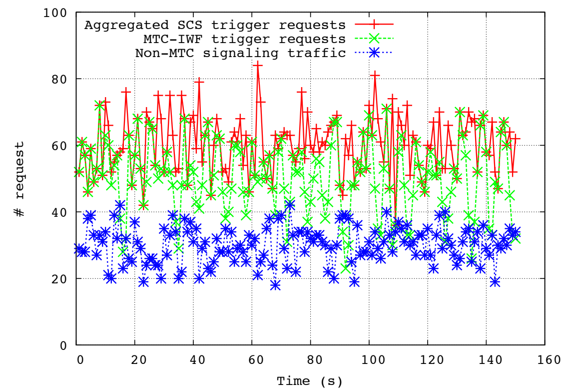

Figure 6. Trigger request and non MTC signaling traffic – ρ=0.5.

Figs. 5 and 6 plot the aggregated SCS trigger requests, the MTC-IWF trigger requests and the non-MTC signaling traffic for both envisioned cases, ρ=1 and ρ=0.5, respectively. Note that the MTC-IWF trigger requests are controlled by the MME reduction message which contains information on the value of λ(t) to use. In Fig. 5, we remark that the MTC-IWF request rate is lower than the incoming SCS requests, meaning that MTC-IWF has applied rate regulation by rejecting parts of SCS requests. This is due to the fact that, along with the non-MTC traffic, the SCS trigger requests increase the MME queue size, as in Figs. 7 and 8. This triggers MME to request MTC-IWF to reduce traffic. However, when the non-MTC traffic is half the SCS trigger request (Fig. 6), the MTC-IWF requests rate is merely equal to the SCS trigger request rate. This is intuitive as MME is lightly congested in this scenario compared to the precedent one. Accordingly, the number of rejected SCS requests is lower in this scenario.

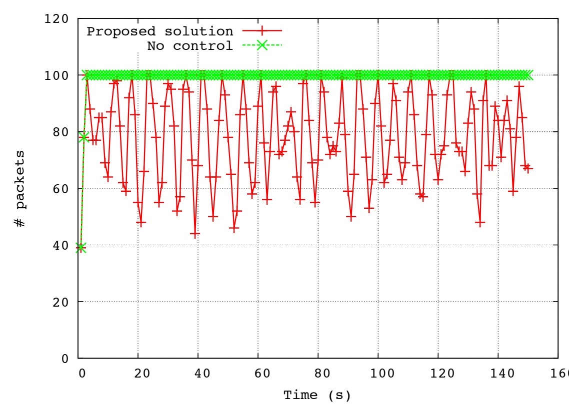

Figure 7. MME queue size – ρ=1.

Figs. 7 and 8 show the variation of the MME queue size when the proposed solution is used and no control mechanism is applied for both scenarios ρ=1 and ρ=0.5, respectively. We observe that using the proposed solution helps in maintaining the queue size around the reference value (80) across the simulation time. In contrast, when there is no control mechanism, MME experiences frequent queue overflows, particularly in the first scenario. Consequently, traffic packets, including non-MTC traffic, get discarded at MME, which degrades the Quality of Service (QoS) of non-MTC traffic.

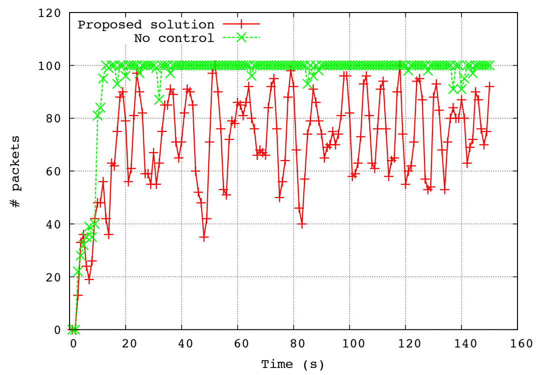

Figure 8. MME queue size – ρ=0.5.

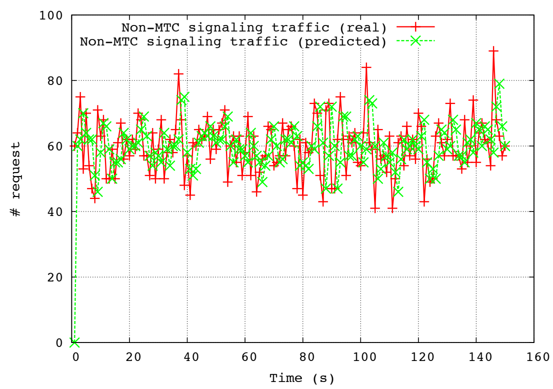

Figure 9. Non-MTC signaling traffic: predicted versus real – ρ=1.

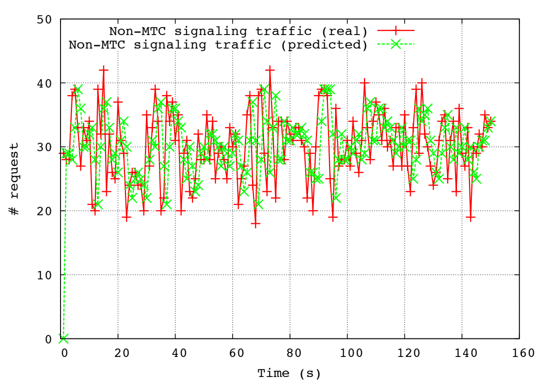

Figure 10. Non-MTC signaling traffic: predicted versus real – ρ=0.5.

Figs. 9 and 10 show the evolution of the non-MTC traffic, and compare the predicted value by our proposed solution and the real traffic, for both scenario ρ=1 and ρ=0.5, respectively. We clearly remark that the predicted values have the same behavior as the real traffic, which indicates that the adopted EWMA technique accurately predicted the incoming non-MTC traffic and validate the underlying assumptions.

V. Conclusion

In this paper, we introduced the problem of "southbound" system overload that may arise from the underlying MTC device trigger procedure when massive MTC devices are triggered at nearly the same time by different MTC servers. We then presented a solution that allows MME control and regulates the MTC device trigger requests rate and have it enforced at MTC-IWF. Indeed, MME implements a control mechanism which detects imminent congestion and mitigates it by reducing incoming MTC device trigger requests from the MTC-IWF entity. In response, this latter uses admission control to decide on immediately accommodating, postponing, or completely rejecting MTC device trigger requests from MTC servers. The obtained simulation results have demonstrated that the proposed solution serves its design goals.

VI. Acknowledgment

Prof. Xiahu Ge and Prof. Hu Honglin were supported by the the National Natural Science Foundation of China (NSFC) under the grants 61271224, NFSC Major International Joint Research Project under the grant 61210002, and the Ministry of Science and Technology (MOST) of China under the grants 0903, the Hubei Provincial Science and Technology Department under the grant 2011BFA004, the Fundamental Research Funds for the Central Universities under the grant 2013QN136.

References

[1] 3GPP TS 23.682 "Architecture enhancements to facilitate communications with packet data networks and applications" Version 11.4.0, 2012-06..

[2] Study on Core Network Overload Solutions; TR 23.843 V0.4.0 (2011-04).

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

[4] S-Y. Lien and K-C. Chen, "Massive Access Management for QoS Guarantees in 3GPP Machine-to-Machine Communications", IEEE Commun. Letter, Vol. 15, No. 3, Mar. 2011.

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

[6] T. Taleb and A. Ksentini, "On Alleviating MTC overload in EPS", in Elsevier J. on Ad Hoc Networks, available online 25 March 2013. (to appear).

[7] A. Ahmed, A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Congestion Control for Machine Type Communications," in Proc. IEEE ICC 2012, Ottawa, Canada, Jun. 2012.

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

[9] O. Mansouri, L. Mokdad, J. Ben Othmane, M. Loualem, "Detecting DOS attacks in WSN based on clustering technique", IEEE WCNC 2013, Shangai, China, 2013.

[10] L. Mokdad, J. Ben Othmane, B. Yahia, S. Niagne, "Performance Evaluation tools for QoS MAC protocol for wireless sensor networks," in Elsevier. J. On Ad Hoc Networks, 86-99 (2014)