M2M SCHEDULING OVER LTE

Challenges and New Perspectives

Antonis G. Gotsis, Athanasios S. Lioumpas, and Angeliki Alexiou

Machine-to-machine (M2M) communications over cellular networks pose significant challenges as a result of the large number of devices, small data transmissions, and vast applications range. Current solutions based on general packet radio service (GPRS) access proved to be inadequate for supporting the M2M ecosystem. Therefore, advanced cellular network releases, such as long-term evolution (LTE) and LTE-Advanced (LTE-A), should efficiently cater to M2M communications. However, the increase in signaling overhead and diverse quality-of-service (QoS) requirements calls for the development of novel flexible scheduling algorithms. In this article, we present the challenges in facilitating M2M scheduling over existing and future cellular infrastructures, review the related proposals, provide some initial solutions, and identify new perspectives, which pave the way for efficient and smooth migration to M2M-enabled broadband cellular systems.

Digital Object Identifier 10.1109/MVT.2012.2204544
Date of publication: 28 August 2012

M2M Markets

A steady growth is forecast worldwide for M2M cellular markets. Current and future cellular systems are expected to play a critical role in the successful deployment of M2M, since they provide essential benefits, such as an established global infrastructure and cost-efficient

 1556-6072/12/$31.00©2012IEEE

communication modules. Recent market reports predict more than 500 million embedded M2M connections by 2015 [1]. They also indicate a high number of M2M devices per cell as one of the most demanding challenges. For example, in the next few years, the number of smart-metering devices per cell in a typical urban environment is estimated to be in the order of tens of thousands [2]. The M2M applications portfolio includes a vast number of diverse smart metering, health monitoring/alerting, and intelligent transportation scenarios that must be efficiently supported [3]. To this end, third generation partnership project (3GPP) and IEEE standardization organizations have both initiated related study items and working groups for facilitating such applications [4], [5] through the evolving system releases.

Overview of Existing M2M over Cellular Solutions

Today, most M2M applications use GPRS capabilities and specific services such as the short message service since they provide a manageable, cost-efficient way for M2M deployment as long as the number of devices remains relatively small. GPRS uses a packet-radio principle for carrying end users' packet-data protocol such as Internet protocol (IP) information to/from GPRS terminals and/or external networks. GPRS is designed for bursty traffic, usually characterizing applications such as Internet browsing or e-mail reading. According to the global system for mobile communications standard [6], resources for GPRS traffic can be reserved statically or dynamically, whereas a combination of both is possible.

As mentioned earlier, GPRS was designed for bursty traffic, a kind of traffic also generated by several M2M applications. Today, almost all available M2M applications and services are based on GPRS, mainly for the following reasons, as argued by the vendors and providers:

- immediate M2M business entry
- low-cost and convenient deployment
- ubiquitous and international operability
- roaming between mobile operators
- GPRS is a proven real-world tested technology, which is open and standardized.

GPRS is a low-cost and well-established technology, but it also has several limitations, which raises serious considerations for its suitability for future M2M applications. More specifically, the capacity of a GPRS cell depends on several parameters, such as the partitioning technique utilized or the reuse pattern. However, assuming only GPRS traffic, the spectral efficiency of a GPRS cell usually does not exceed 100–150 kb/s/cell/MHz. In addition, if voice users are assumed to be active, the number of supported data users becomes limited (<30) [7]. Therefore, it is obvious that the GPRS capacity is limited for supporting the envisioned M2M applications and services with thousands of devices per cell. Another issue regarding GPRS networks is that it mandates that the connection is initiated by the device in the GPRS network [8].

Enabling M2M on LTE

The limited capacity of the second-generation cellular systems has forced the M2M industry toward higher-capacity systems. The most promising candidate technology, 3GPP LTE offers higher capacity and more flexible radio resource management (RRM) compared with high speed packet access data technologies. However, LTE has been designed for broadband applications, while most M2M applications transmit and receive small amounts of data, leading to an unreasonable ratio between payload and required control information and nonoptimized transmission protocols. Moreover, important aspects, such as the need for low-energy devices or lower latencies, have to be considered for M2M communications. Therefore, efforts have been exerted by 3GPP under the umbrella of machine-type communications study and work items to overcome the deficiencies of LTE with respect to M2M communications [9]. Besides the necessity for supporting a large number of M2M devices, the efforts also focus on addressing the vast diversity of M2M service characteristics, need for energy efficiency, and coexistence with current communications systems.

Packet scheduling constitutes the key RRM mechanism for minimizing the overall system resources usage while guaranteeing individual QoS service requirements and, thus, has been given great attention at 3G and beyond systems. Particularly in LTE, the flexible resources organization in time–frequency resource elements or physical resource blocks (PRBs) renders efficient scheduling more decisive in system performance. To optimally allocate the PRBs to the users' equipment (UEs) and/or machine-type communication devices (MTCDs), the scheduler should exploit channel and traffic dynamics on a fast time scale, ideally per transmission time interval (TTI). Hence, associated uplink (UL) and downlink (DL) signaling channels for carrying the channel quality, traffic, and allocation information are necessary for facilitating scheduling. The unique characteristics of M2M traffic, such as the large number of devices and the bursty low-rate load nature, perplex scheduling as both complexity and signaling are heavily increased. It is worth mentioning that besides the flexible resource allocation capabilities of LTE, features such as relaying in LTE-A offer additional benefits for M2M communications, such as coverage extension for low-energy devices. This article aims at presenting the challenges of M2M scheduling over current and future 3GPP cellular networks as well as proposing new research directions on this topic.

Signaling and Scheduling Limitations for M2M over LTE

Regarding M2M application scenarios, such as smart metering, e-health, and intelligent transportation, the UL direction is expected to dominate the traffic load. Therefore, we will focus on UL scheduling.

Signaling in LTE Scheduling

In LTE, UL scheduling takes place at the base station [eNB (eNodeB)], which is the 3GPP term in a centralized manner, and the allocation decisions are communicated to the UEs through appropriate control channels. In particular, each UE sends scheduling requests to the eNB via L1/L2 control signaling, i.e., the physical UL control channel (PUCCH), requiring access to the UL shared channel.

Each UE is assigned one PUCCH, therefore, at the presence of a large number of MTCDs, as in M2M scenarios, a shortage of PUCCH resources is possible. Moreover to cater to channel-aware scheduling, the per-device channel quality information needs to be sent to the eNB; this information is also carried in a UL control channel. As the number of devices grows, so does the associated signaling load.

Based on the scheduling requests, the eNB decides the PRB-to-UEs (or MTCDs) allocation at each TTI and sends this information (in other words, the scheduling grants) to the MTCDs through the corresponding physical DL control channel (PDCCH). The PDCCH physical channel is loaded into the first one to three orthogonal frequency division multiplexing (OFDM) symbols of the DL time–frequency grid, consuming system resources. According to the 3GPP specifications, up to ten UEs (or MTCDs) may be loaded in a single subframe. Hence, PDCCH is unable to support hundreds of MTCDs demanding simultaneous access to the shared channel in future M2M scenarios.

LTE also provides a random access transport channel in the UL, mainly utilized for initial radio-link establishment, handover, and synchronization acquiring. It may also carry UL scheduling requests if no dedicated PUCCHs have been assigned. This is done through a contention-based mechanism contrary to the contention-free scheduling. Each UE competing for access may use one of the 64 (significantly lower than the expected volume of MTCDs) available orthogonal sequences per cell. If more messages should be transmitted, collisions will occur.

It is obvious from the above discussion that the signaling for supporting UL scheduling in M2M scenarios becomes prohibitive, and the current mechanisms may not effectively support MTCDs in LTE cells. Modifications on existing approaches as well as the design of new solutions for reducing the signaling overhead are important in future M2M-enabled LTE systems.

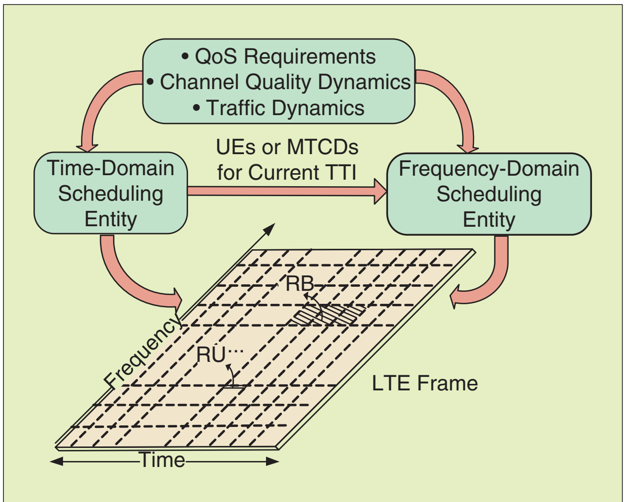

FIGURE 1 The LTE packet scheduling framework.

The LTE Scheduling Framework

In general, scheduling (as an RRM procedure) is not part of the standardization work, rather it is an implementation-specific issue. However, signaling is standardized, thus any scheduling proposal should be in line with the set of control requirements. Toward this purpose, a generic packet-scheduling framework has been proposed in 3GPP [10], according to which 1) a set of UEs is picked by a time-domain scheduling entity for packet transmission in the current frame and 2) a frequency-domain scheduling entity allocates the PRBs to the selected UEs (Figure 1). Both entities make the allocation decisions based on metrics expressing the channel quality, experienced packet delays, and QoS tolerances. Leaning on this flexible framework, several schemes have been devised for dynamically allocating resources to UEs with heterogeneous QoS requirements, such as non-real-time and real-time services (file transfer protocol downloading/uploading, Web surfing, and video streaming) [11]. A set of nine QoS classes has been prescribed in 3GPP specifications [QoS class identifiers (QCIs)], classifying services (or radio bearers) based on the resource type [guaranteed bit rate (GBR)/non-GBR], priority order, packet delay budget, and packet loss rate characteristics.

With M2M scenarios coming into play, however, the scheduling entities have to deal with extremely diverse QoS criteria. For example, delay tolerance may span from tens of milliseconds (vehicles collision) to several minutes (environmental monitoring). In addition, the error rate tolerances may scale accordingly. Thus, forming specific QoS classes/clusters is not an easy task. As scheduling exploits channel and traffic dynamics, the presence of a huge number of MTCDs may induce further processing delays and impose storage constraints into the base station. Developing practical scheduling schemes that support a large number of MTCDs without deteriorating the performance of standard LTE services is therefore a challenging issue.

Existing Proposals for M2M Scheduling over LTE

Since M2M-enabled LTE is a recent research area, the bibliography on M2M scheduling is still scarce.

Group-Based Scheduling

In [12], a reduced complexity approach for managing radio resources and scheduling is proposed. It is based on the formation of MTCD groups or clusters, where each cluster is associated with a prescribed QoS profile. Then MTCDs are transparently connected to and managed by LTE, since eNBs control the cluster entities. Scheduling prioritization is imposed on a cluster basis, a policy that significantly reduces complexity and overhead. Cluster formation is dictated by the packet arrival rate and maximum tolerable jitter.

Time Granularity of Scheduling

In LTE, full dynamic channel- and QoS-aware scheduling per TTI is optimal in terms of system performance. For M2M-enabled LTE, this would be the case as well, although a large number of MTCDs imposes serious constraints to the applicability of such highly dynamic approaches. A similar problem has been identified in the LTE standard as well, regarding the provision of voice over the IP (VoIP) services. LTE should support hundreds of VoIP UEs that generate small amounts of periodic data traffic. In the literature, semipersistent scheduling schemes have been proposed for effectively dealing with special traffic characteristics [13]. Semipersistent scheduling makes an allocation decision for a longer time period; thus, it is unnecessary to inform the UEs on a TTI basis. Due to the similarity of the traffic nature of VoIP with M2M traffic, semipersistent scheduling is a possible candidate for M2M scenarios.

New Results and Long-Term Perspectives

QoS Clustering and LTE Scheduling

Although the group-based proposal lowers the signaling burden, the issue of QoS classes definition for MTCDs or, in other words, the introduction of new MTCD-QCIs on top of LTE ones, raises several issues, with the most important one stemming from the fact that the number of different classes/clusters must be infinite. This means that both the number and range of each class must be carefully chosen to successfully capture the diverse traffic characteristics of M2M applications. Moreover, in contrast to current policies, a dynamic formation of QoS classes, according to a particular application scenario, may be more appropriate for M2M communications given that the MTCDs topology and individual characteristics are not a priori known.

To develop a novel analytical (or combined analytical/empirical) model for MTCD-QCIs formation, we first need to understand the effect of clustering granularity on system performance. In a recent work [14], two simple scheduling algorithms were proposed that do not use the clustering idea but assign resources to MTCDs based on each individual delay budget. This work illustrated the importance of the number of different M2M classes for the performance of the scheduler. Utilizing classes that span the whole range of M2M QoS requirements is essential. Figure 2 depicts the performance of a priority-based scheduler (MTC devices with lower delay tolerance are served first), when different number of classes are used. It can be seen that, using the standard QCI classes (as specified for the existing LTE services), the percentage of MTC devices violating their constraints are significantly higher than the case where different classes are used. More specifically, using classes that arise from quantizing the whole range of delay tolerances (e.g., uniform quantization between the minimum and maximum required delay), the performance of the scheduler significantly increases. Standardizing the number of required QCIs for M2M communications is considered as an essential task.

Low-Complexity Scheduling and M2M Bandwidth Estimation

We next put an emphasis on a low-complexity solution for scheduling MTCDs (single and grouped) by extending [12]. In particular, we adopt the time-controlled M2M feature, which dictates that M2M device groups are scheduled at specific periodic intervals, similar to the semipersistent VoIP paradigm, and assume Poisson-like generated traffic. We model the particular scheduling as an M/D/1 queuing problem and following the ideas of such a queuing framework, we develop an analytical model that relates the scheduling period, average offered traffic load, and QoS requirements, in terms of packet delay budget and packet-dropped rate as in LTE. The proposed model may be utilized for 1) tuning the scheduling decisions, i.e., decide which is the minimum scheduling period for an

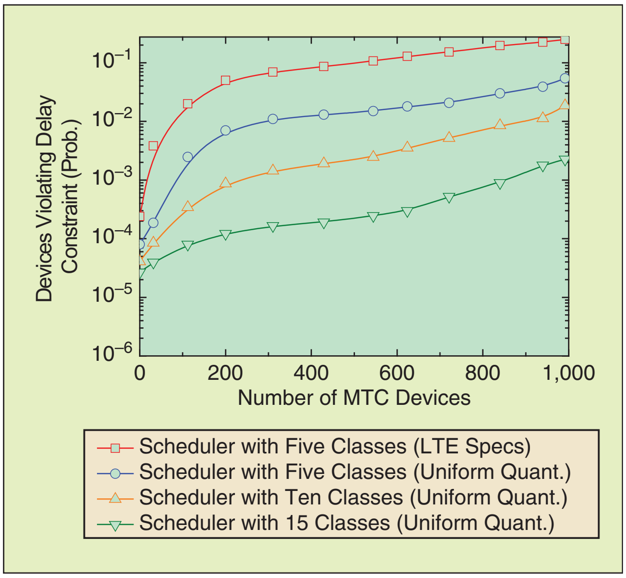

FIGURE 2 The effect of the number of M2M classes on the scheduler's performance.

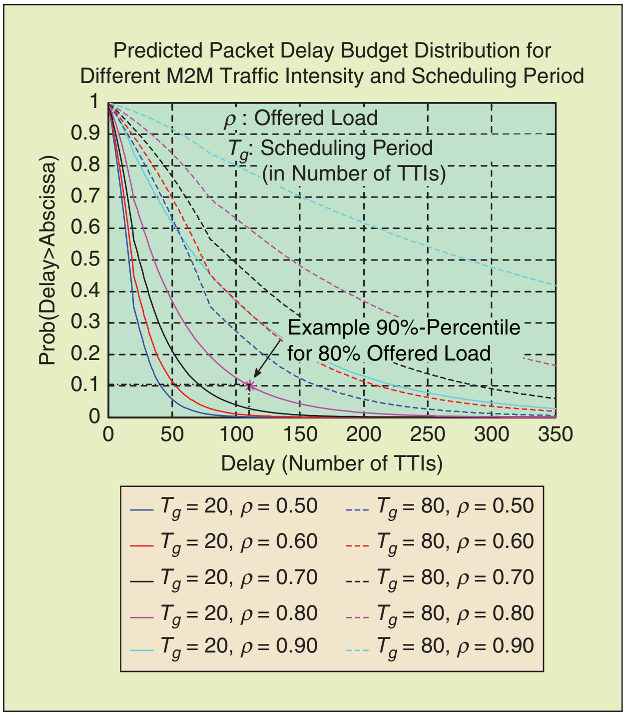

FIGURE 3 Low-complexity M2M scheduling tuning and prediction example results.

M2M cluster to meet the probabilistic QoS targets and 2) estimating the minimum LTE bandwidth reserved for M2M to support different M2M loads with prescribed QoS requirements. In Figure 3, indicative results for the packet delay distribution corresponding to variable M2M traffic loads and scheduling periods are illustrated. One may easily observe how for a specific scheduling/load scenario, e.g., the 90% percentile of the packet delay budget may be predicted. Besides the analytical tractability of this approach, such a periodic-like scheduling scheme is efficient in terms of signaling because each MTCD needs to be informed about the period when it will be active.

This is the first step in understanding the viability of such scheduling schemes on M2M-enabled LTE systems. On the other hand, hybrid full dynamic/semipersistent schemes for traffic mixes, including legacy LTE UEs and MTCDs, could be more appropriate. A thorough study of these issues is necessary for identifying the inherent performance-complexity tradeoffs of various scheduling schemes. Moreover, studies targeting different mix loads are necessary for understanding the real performance bounds of LTE or other wide cellular networks that facilitate M2M services.

Longer-Term Perspectives

The issues we have presented so far mainly call for modifications on the existing protocols. However, to efficiently support M2M over LTE, more radical changes may be needed in the future, regarding the radio access network architecture and procedures. An important question that the future system designer needs to answer is whether an evolutionary or a revolutionary approach would be more appropriate in the advancement process of cellular systems like LTE for them to efficiently provide M2M communications.

As a first example, the initial access procedure of LTE (Figure 4) involves many steps that increase overhead and latency and may be avoided for M2M data transmission. Simpler access procedures (evolutionary approach) or completely separated standard LTE- and M2M-oriented control and data channels (revolutionary approach) may be needed. Second, so far, we have assumed that M2M data are loaded on the current LTE frame. However, the LTE frame is designed for data-hungry applications; thus, the minimum resource element is 180 kHz, significantly larger than the needs of typical M2M applications. Hence, alternative physical-layer oriented solutions will be needed, such as aggregating data from multiple devices into a single resource block or employing advanced hierarchical modulation techniques. Finally, to minimize layer-3 signaling and latency, a hybrid contention- and schedule-based scheme may be applied, incorporating the merits of low-complexity random access and high-performance centralized access at higher loads.

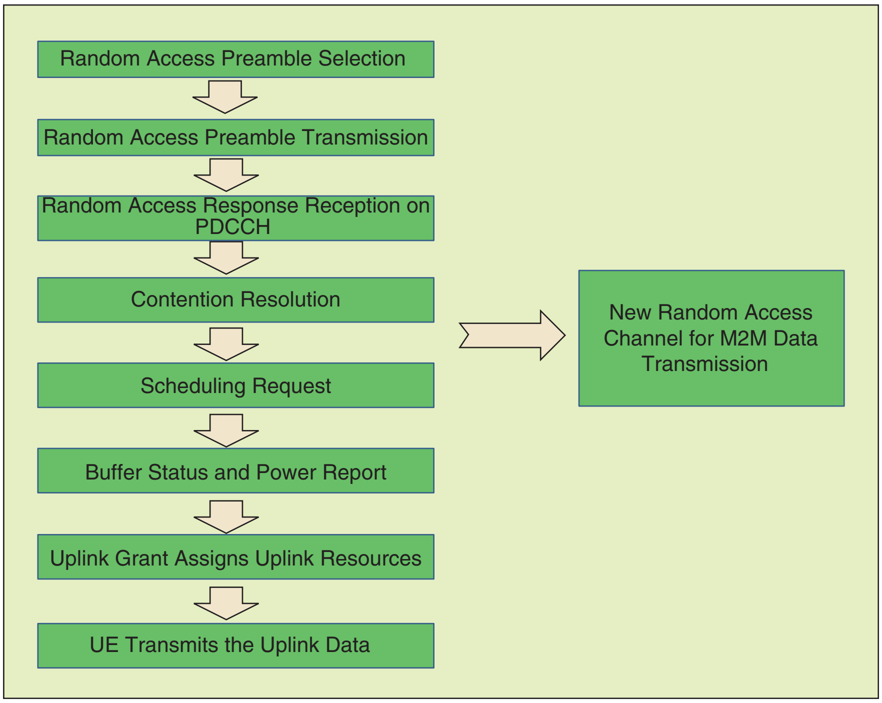

FIGURE 4 New random access channel for M2M data transmission.

Conclusions

Supporting M2M communications over advanced next-generation cellular infrastructures, such as LTE and its descendant, LTE-A, opens up a wide range of applications. An important enabler for M2M is the design of efficient and dynamic packet scheduling schemes. In this article, we first explained why second-generation systems are not expected to carry the increasing M2M load and then recognized the limitations of the existing scheduling approaches with respect to the distinctive M2M traffic characteristics. Then we reviewed the proposals appearing in the literature for supporting this type of traffic in an advanced cellular network, such as device clustering and semipersistent/fixed grant scheduling. We presented new results on possible scheduling solutions and explained why there are still significant challenges to deal with. We finally argued that, contrary to the presented evolutionary changes, revolutionary paths may also be needed for fully enabling M2M capabilities over LTE and LTE-A.

Acknowledgments

This work has been performed in the framework of the ICT project, ICT-5-258512 EXALTED, which is partly funded by the European Union. The authors acknowledge the contributions of their colleagues, although the views expressed are those of the authors and do not necessarily represent the project. This research has also been cofinanced by the European Union [European Social Fund (ESF)] and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALIS.

Author Information

Antonis G. Gotsis (agotsis@unipi.gr) received his diploma in electrical and computer engineering and his Ph.D. degree in wireless communications from the ECE School of the National Technical University of Athens (NTUA), Greece, in 2002 and 2010, respectively. His thesis regarded optimization techniques for RRM over advanced broadband wireless access networks. From September 2002 to December 2010, he was with the Mobile Radio-Communications Laboratory of NTUA working as a project engineer and a research assistant. In January 2012, he joined the Department of Digital Systems, University of Piraeus, working as a postdoc researcher. His research interests include M2M communications over LTE networks and interference management techniques beyond LTE and LTE-A.

Athanasios S. Lioumpas (lioumpas@unipi.gr) received his diploma in 2005 and his Ph.D. degree in 2009, both in electrical engineering, from the Aristotle University of Thessaloniki, Greece. In 2010, he joined the University of Piraeus. He has published more than 30 scientific articles in international journals and conferences, while he has been granted five patents. He is corecipient of the Best Paper Award of the Wireless Communications Symposium (WCS) in IEEE, International Conference on Communications (ICC) 2007. His research interests include low-complexity and energy-efficient mobile communications, M2M communications over LTE, and resource allocation in OFDM-based systems.

Angeliki Alexiou (alexiou@unipi.gr) received her diploma in electrical and computer engineering from the National Technical University of Athens in 1994 and her Ph.D. degree in electrical engineering from Imperial College of Science, Technology and Medicine, University of London in 2000. She was with Bell Laboratories, Wireless Research, Lucent Technologies, now Alcatel-Lucent, in Swindon, United Kingdom, first as a member of technical staff (1999–2006) and later as a technical manager (2006–2009). Since May 2009, she has been an assistant professor at the Department of Digital Systems, University of Piraeus, Greece, where she conducted research and teaches courses in the area of broadband communications and short range wireless systems. She is the chair of the working group on new air interfaces, relay-based systems, and smart antennas of the Wireless World Research Forum. Her research interests include multiple antenna systems, multihop communications, advanced signal processing, and efficient RRM for future generation wireless systems.

References

[1] "Machine-to-machine (M2M), The rise of the machines," Juniper Networks White Paper, 2011.

[2] Vodafone, "RACH intensity of time controlled devices," 3GPP, Tech. Rep., R2-102296, Apr. 2010.

[3] EXALTED Deliverable 2-1.(2011). Description of baseline reference systems, scenarios, technical requirements and evaluation methodology [Online]. Available: http://www.ict-exalted.eu

[4] 3GPP, "Service requirements for machine-type communications (MTC); Stage 1(Release 11), v.11.3.0," TS 22.368, 2011-09.

[5] *DRAFT Amendment to IEEE Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems: Enhancements to Support Machine-to-Machine Applications*, IEEE Standard P802.16p-11/0033, Oct. 2011.

[6] *Digital Cellular Telecommunications System (Phase2+)*, ETSI GSM Specification 02.60, 1999.

[7] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE Performance: Evolution Towards 3G/UMTS*, Hoboken, NJ: Wiley, 2002.

[8] M. Martsola, T. Kiravuo, and J. Lindqvist, "Machine to machine communication in cellular networks," in *Proc. 2nd Int. Conf. Mobile Technology, Applications and Systems*, 2005, pp. 6.

[9] 3GPP, "System improvements for machine-type communications; (Release 11), v.1.6.0," TR 23.888, 2011-11.

[10] 3GPP R1-072277. (2007, May). Downlink interference coordination, technical document, Nokia and Nokia Siemens Networks, Kobe-Japan [Online]. Available: http://www.3gpp.org.

[11] G. Monghal, K. I. Pedersen, I. Z. Kovacs, and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proc. IEEE Vehicular Technology Conf.* (VTC), May 11-14, 2008, pp. 2532–2536.

[12] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[13] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proc. Int. Conf. Wireless Communications, Networking and Mobile Computing (WiCom)*, Sept. 21-25, 2007, pp. 2861–2864.

[14] A. S. Lioumpas and A. Alexiou, "Uplink scheduling for machine-to-machine communications in LTE-based cellular systems," in *Proc. IEEE Globecom 2011 Conf.*, Houston, Texas, pp. 353–357.

VT