Song et al. EURASIP Journal on Wireless Communications and Networking (2016) 2016:140
DOI 10.1186/s13638-016-0636-y

EURASIP Journal on Wireless Communications and Networking

REVIEW Open Access

# Survey of radio resource management issues and proposals for energy-efficient cellular networks that will cover billions of machines

Qipeng Song[1,2*], Loutfi Nuaymi[1,2] and Xavier Lagrange[1,2]

**Abstract**

A huge increase of machines attached to wireless networks is expected in the next few years. A large part of these machines will be covered by some wireless wide area networks. The arrival of cellular M2M (machine-to-machine) communication poses new requirements due to its specific characteristics. For most of the cellular M2M applications, the essential requirement is low energy consumption level or high energy efficiency. This survey provides a global view of the network technologies previewed for cellular M2M. In this survey, we study the existing classifications of M2M applications according to different criteria in the literature. The comparison of traffic characteristics between M2M and human-to-human is also proposed. Quality of service (QoS) requirements for typical M2M applications are resumed. The advance of reference M2M network architectures proposed by the Standard Development Organization (SDO) is investigated. We identify two possible effort directions to improve the energy efficiency for cellular M2M. The first one is to evolve the current existing 3rd Generation Partnership Project (3GPP) Consortium cellular networks to effectively support MTC (Machine Type Communication). The other direction is to design M2M-dedicated networks from scratch, which are often called low-power wide-area (LPWA) networks. We review, compare and categorize the proposals related to energy issues of cellular M2M mainly over the period 2011–2015 for the first direction. We introduce the development of LPWA networks for the other research directions. We highlight that the cooperative relaying, the design of energy-efficient signaling and operation, the new radio resource allocation schemes, and the energy-efficient random access procedure are the main points of improvement. It is important to jointly use the aforementioned approaches, for example, joint design of random access control and radio resource allocation, to seek for a trade-off between energy efficiency and other system performances.

**Keywords:** 5G, M2M, Energy-efficiency, LPWAN

## 1 Introduction

Machine-to-machine communication (M2M), also known as Machine Type Communication (MTC), is an emerging technology allowing devices to mutually communicate without (or only limited) human intervention, which is expected to gain more popularity in the next decade and be an integrated part of the future wireless networks [1, 2]. As an example, Ericsson estimates that 2 out of 50 billion MTC devices in 2020 will be connected by cellular technology [3].

MTC presents lots of its own characteristics different from traditional human-to-human (H2H) or Human Type Communication (HTC): uplink-centric applications, short but more frequent transmission, large number of devices, difficulty to change battery, and so on [4]. Therefore, to well accommodate MTC traffic in the future wireless networks, two possible approaches are envisaged:

- Design from scratch of M2M-dedicated networks, i.e., the emerging Low-Power Wide-Area Network

---

*Correspondence: qipeng.song@telecom-bretagne.eu
[1] Department of Network, Security and Multimedia, 2 Rue de la Chataigneraie, 35510 Cesson Sévigné, France
[2] IRISA, 263 avenue du Général Leclerc, 35042 Rennes, France

(LPWAN). A representative example is the LoRaWAN (LoRa Wide Area Network) [5] proposed by LoRa Alliance [6]
- Evolution from existing wireless networks, which consists of adapting 3rd Generation Partnership Project (3GPP) cellular networks to support MTC traffic apart from HTC traffic, for example the Long Term Evolution (LTE)-M [7].

Cisco estimates that the LPWA and evolved 3GPP networks will have a dominant role for handling MTC traffic in the future. It is expected that 29 % of MTC devices will be served by the LPWA networks and 77 % of M2M connections will be served by the 3GPP networks (including 2G/3G/4G, shown in Fig. 1). The reason is that 3GPP cellular networks, compared with LPWA networks, have ubiquitous coverage, largely deployed infrastructure, mature user subscription/management system, and so on.

For both the aforementioned approaches, the challenges to be solved are the same: MTC subscription, network/overload control (also called massive access control), security in M2M, diverse quality of service (QoS) provisioning, energy efficiency, etc. Recently, energy efficiency-related research has attracted more and more attention, since it is deemed as a key performance indicator that determines if MTC is accepted as a promising technology [8, 9]. Note that energy efficiency actually covers the device side and network side. For cellular M2M, the network side energy efficiency [10, 11] is not a principal constraint; hence, it is not within the scope of this article. Instead, MTC devices are usually battery-operated, transmit small data, and require a long battery lifetime [12]. The device side energy efficiency is a key problem to make 3GPP cellular networks as a competitive solution for MTC. Thus, we put more focus on advance research about the cellular MTC energy-efficiency issue, especially in radio access networks.

With regard to cellular M2M-related surveys, Taleb and Kunz [13] focus on MTC devices subscription control and network congestion/overload control. Chen and Lien [14] talk about research efforts for efficient MTC and explore various M2M-related issues such as deployment, operation, security, and privacy. Andres et al. [15] make a survey of proposals improving the operation of random access channel of LTE/LTE-A and evaluate the energy consumption of LTE random-access procedure. Poncela et al. [16] identify the limitations of 4G for MTC (signaling, scheduler) and resume the improvements of LTE/LTE-A to handle M2M traffic. Several review papers [17, 18] discuss MTC in 3GPP LTE/LTE-A networks, introduce M2M use cases in detail, and identify the challenges with regard to M2M over LTE/LTE-A, e.g., random access congestion, resource allocation with QoS provisioning. Wang et al. [19] survey and discuss various remarkable techniques, in terms of all components of the mobile networks (e.g., data centers, macrocell, femtocell), towards green mobile cellular networks. Ismail et al. [20] investigate energy efficiency from the perspective of network operators and mobile users. Yang et al. [21] make a survey about software-defined wireless network (SDWN) and wireless network virtualization (WNV) for the future mobile wireless networks, which helps define the future mobile wireless network architecture to tackle with heterogeneous traffic.

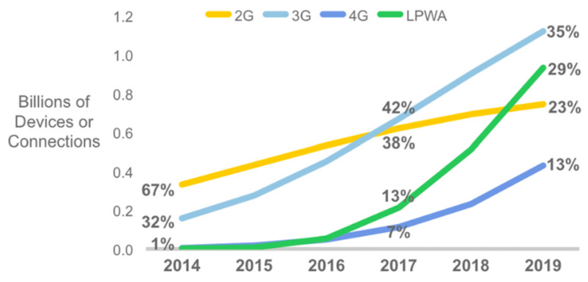

**Fig. 1** Global M2M growth and migration from 2G to 3G and 4G. Source: [84]

To our best knowledge, a comprehensive survey about device side energy-efficiency issues in radio networks used for cellular M2M service is still not available in the literature. Therefore, the goal of this article is to compare and categorize existing M2M-related energy-efficiency proposals before discussing the trends for cellular M2M research. In addition, in this article, we want to provide a short overview of cellular M2M applications, detail different types of classification of M2M services, and propose a synthesis for the QoS demands. We also review the advances of LPWAN, which are MTC-dedicated networks, which today experience a rapid development. The rest of this article is organized as follows. Section 2 presents the typical cellular M2M applications and several classifications according to different criteria and introduces a QoS requirement table for some typical cellular M2M applications. Section 3 compares the differences between H2H and M2M in terms of traffic characteristics. Section 4 first talks about conventional M2M solutions in cellular networks then presents the advance of reference M2M network architecture. Section 5 resumes the development of LPWAN. Section 6 presents, categorizes, and compares all found proposals related to energy issues for MTC in cellular networks. Section 7 gives the conclusions obtained from this survey.

## 2 M2M applications classifications

3GPP has listed a part of existing M2M applications shown in Table 1, including security, intelligent transport system, payment, health, remote maintenance/control, metering, and consumer devices. Besides, frequent M2M

**Table 1** MTC applications categorization (non-exhaustive) according to [1, 17, 18]

| Service area | MTC applications |
|---|---|
| Security | Surveillance systems |
| | Backup for landline |
| | Control of physical access |
| | Car/driver security |
| Intelligent transport system | Fleet management |
| | Order management |
| | Pay as you drive |
| | Assess tracking |
| | Navigation |
| | Traffic information |
| | Road tolling |
| Payment | Point of sales |
| | Vending machine |
| | Gaming machines |
| Health | Monitoring vital signals |
| | Supporting the aged or handicapped |
| | Web access telemedicine points |
| | Remote diagnostics |
| Remote maintenance/control | Elevator control |
| | Lighting |
| | Pumps |
| | Industrial automation |
| | Vehicle diagnostics |
| Metering | Power/gas/water |
| | Heating |
| | Grid control |
| | Industrial metering |
| Consumer devices | Digital photo frame |
| | E-book |
| Other futuristic applications | Information ambient society |
| | Robotic applications |
| | Environment monitoring |

applications are bicycle-sharing system [22], logistics application, and insurance [22]. Given that M2M applications are so various, it is impossible to develop a unique platform to support all these applications in an economical way. For example, the metering device is desired to be simple but the video surveillance device needs different and powerful codec, needs different protocols, and should have different radio capabilities. If only one solution or hardware platform is adopted for all the services, the consequence is a MTC machine with a high complexity, which is like a smart phone today [23]. Thus, it is important to classify the existing M2M applications and propose improvement according to different requirements. In addition, an appropriate M2M classification helps identify QoS features and other works. In this section, we present all classification schemes found in the literature, each of which may serve for specific research purpose.

### 2.1 Classification according to reliability and quantity of connected machines

According to reliability and quantity of connected machines, the project METIS divides M2M applications into two categories: mMTC and uMTC [24]. mMTC refers to massive MTC and provides connectivity for a large number of cost and energy-constrained devices. Sensor and actuator deployments can be in a wide area for surveillance and area covering measurements, but also co-located with human users, as in body-area networks. The main attribute of this service is the massive number of connected devices, where the required rates decrease as the number of devices grows significantly. uMTC addresses the needs for ultra-reliable, time-critical services, e.g., V2X (vehicle-to-vehicle/infrastructure) applications and industrial control applications. Both examples require reliable communication, and V2X additionally requires fast discovery and communication establishment. The main attribute is high reliability, while the number of devices and the required data rates are relatively low.

### 2.2 Classification according to the level of mobility and dispersion

According to the level of mobility and dispersion, M2M applications also could be categorized into four categories: dispersed and fixed application, dispersed and mobile application, concentrated and fixed application, and concentrated and mobile application [22]. The dispersion refers to the area that the M2M devices are spread out over. The mobility measures whether the device is stationary or whether it needs to move around. The typical example for dispersed and fixed application is a smart grid where sensors are deployed in a large scale at a fixed location. Logistic applications are representative for dispersed and mobile application. For example, the sensors to track the cargo in the container may spread and move anywhere. Figure 2 shows the result of classification for some typical M2M applications according to this criterion.

### 2.3 Classification according to delay tolerance level

According to the delay tolerance level, M2M applications are divided into four classes: class 1 (elastic applications), class 2 (hard real-time applications), class 3 (delay-adaptive applications), and class 4 (rate-adaptive

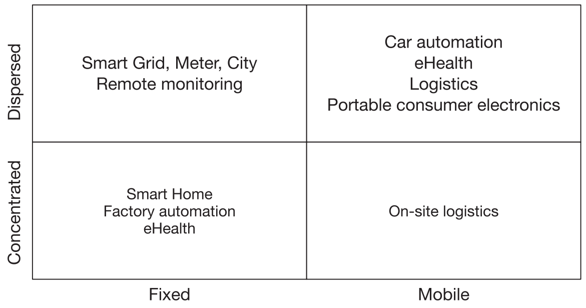

**Fig. 2** M2M application taxonomy by mobility and dispersion. Source: [22]. This classification is not always true. Some applications may be in different cases in some situations

applications) [25]. The class 1 applications are generally rather tolerant of delays, for example, file downloading of remote MTC devices from MTC servers. The class 2 applications need their data to be served within a given delay constraint. The typical example of class 2 application is vehicle and asset tracking. Similar to class 2, the class 3 applications are usually delay sensitive, but most applications of class 3 can be made rather tolerant of occasional delay-bound violation and dropped packets. The class 4 applications adjust their transmission rates according to available radio resources while maintaining moderate delays.

### 2.4 Classification according to data reporting mode

According to data reporting mode, M2M applications are classified into five categories [9, 26, 27]: time-driven, query-driven, event-driven, continuous-based, and hybrid-driven. Time-driven M2M applications refer to those applications where machines periodically turn on their sensors and transmitters to transmit the collected data. Query-driven applications reply to certain instructions from MTC application servers by transmitting data. This type of applications allows packet omissions, as adjacent data reports usually contain redundant information. Event-driven applications react to certain critical query or event. Normally, applications fall into this category when they use priority alarm messages (PAM). Continuous-based M2M applications make the devices send their data continuously to the remote server at a pre-specified rate. Hybrid-driven is a combination of the aforementioned three types.

### 2.5 QoS feature for typical M2M applications

In Table 2, a QoS requirement table in terms of data rate, latency, and message priority is given for some representative cellular M2M applications, based on different references found in the literature.

## 3 M2M traffic characterization

### 3.1 Comparison of M2M and H2H traffic

A comprehensive comparison between M2M and H2H is resumed in Table 3. The illustration of this difference helps to rethink the design of some principles and the optimization guideline. Here are some explanations about Table 3. First, the representative device in H2H communication is a smartphone, which is equipped with more and more computational capacity. The complexity of M2M devices is various: in application such as remote monitoring, it could be in format of sensor with a transceiver and a simplified processor. In intelligent transport system (ITS), it could be regarded as a smartphone without a screen. Second, the experiment results in [4] reveal that compared to H2H communication, cellular M2M traffic suffers from a higher packet loss ratio, and the reason may be the adverse deployment location and the lack of UI (e.g., screen) to show the signal strength in its place. Third, since at present, a majority cellular M2M applications are based on GSM/Universal Mobile Telecommunications System (UMTS) technologies, MTC mainly supports short message service (SMS) or data reporting service. We could imagine more innovation M2M services when 4G network is largely rolled out.

**Table 2** QoS feature and cellular M2M service. The values we propose are based on different references. We give only indicative values

| M2M service | Data rate | Latency | Priority |
|---|---|---|---|
| Surveillance system | 64.000 b/s [86] | Small | Medium |
| Urgent notification | Small | Less than 1 s | High |
| Fleet management | Less than 500 B | Very small | High |
| Pay as you pay | Small | Very small | High |
| Smart metering | 500–1000 B per message | 15 s–15 min [87] | Low |
| Grid automation | 10–100 kps [16] | 0.1–2 s [16] | High |
| Monitoring vital signals | Less than 200 B per message [88] | Small | High |
| Monitoring in emergency | Less than 200 B per message [88] | Small | High |
| Industrial automation | Small | Less than 5 ms [89] | High |
| Vending machine control | Small | Small | Medium |

**Table 3** Difference between M2M and H2H

| Item | M2M | H2H |
|---|---|---|
| Delay range | 10 ms∼several minutes [50] | 250 ms (voice) to few seconds (email for example) |
| Device composition | GSM/UMTS/LTE module,<br>extension slots, USB, memory, CPU, etc. | GSM/UMTS/LTE/Wifi module<br>GPS, Bluetooth, USB, memory, CPU, flash storage, etc. |
| Packet loss radio | Relatively high [4] | Low |
| Mobility | Most of the M2M devices (90 % according to [90]) are stationary. | Humans are very rarely considered fixed in practical mobile networks |
| Support service | Mainly SMS or data reporting | SMS/voice/Web/multimedia, etc. |
| Session duration /frequency | Short but more or less frequent [91], depending on the applications: monitoring, transport or others | Long but less frequent |
| Uplink | MTC traffic is mainly generated in uplink | Traditionally less traffic in uplink but increase rapidly with the flourishing of interactive applications such as social network |
| Downlink | Less traffic except for some application requiring interaction between sensors and MTC servers, for example consumer electronics use case | Currently most traffic, for instance, Web browsing and multimedia |
| Message size | Generally very short.<br>In some cases could increase, for example, if video sequences are uploaded | Typically big, especially for multimedia and real-time transmission |
| Number of devices | Hundreds or thousands of devices per base station | At most hundreds of UE, typically tens of UEs per base station [15] |
| Battery life requirement | Up to a few years, especially for deployment locations with difficult access | Order of days or weeks,<br>Human could easily recharge their device |
| Key metrics for user experience | Energy efficiency, latency | Delay, throughput, packet loss |

Not all MTC applications have the same characteristics and not every optimization is suitable to all applications; therefore, features are defined to provide some structure to the customer and the network is then tuned accordingly to needs.

In many applications, saving energy for machines is more important than increasing the throughput because machines usually transmit small data but have limited electric energy [12]. The energy efficiency is deemed as a key performance indicator that determines if M2M communication is accepted as a promising communication technology [8]. One of the important requirements in cellular M2M system is extremely low power consumption [28]. The hard QoS guarantee is deemed as one of the most important requirements since disasters occur if timing constraints are violated for some MTC applications [29]. Energy efficiency is the key in M2M communications, since machine devices are generally powered by batteries [9]. A critical issue in M2M communications is energy efficiency as typically the machine devices are powered by batteries of low capacity and thus it is the key to optimize their consumption [9].

### 3.2 M2M traffic-related research efforts

The first large-scale measurement about M2M traffic over an actual cellular network is in [4]. The possible impacts of MTC traffic on H2H traffic over cellular networks are evaluated in [30]. Traditionally, traffic models are classified as source traffic model and aggregated traffic models [31]. Source traffic model is precise but not scalable with number of MTC devices, and aggregated traffic model is less complex but not precise. A couple Markov-modulated Poisson process (CMMPP) model is proposed by combing the respective advantages of source traffic and aggregated traffic models.

## 4 Conventional M2M in 3GPP networks and standardization of M2M architecture

### 4.1 Conventional M2M solution in 3GPP networks

The 3GPP GSM cellular networks have been regarded as an ideal carrier for M2M, for the small data transmission, low data rate and energy efficiency, and low-cost hardware for MTC devices. Thus, lots of cellular-based commercial solutions [32] have been proposed via GSM using SMS or GPRS before 2010. However, for a long-term view, GSM is not the best choice for MTC. There are many reasons for that. First, the operators have the plan of *spectrum refarming*, that is, the spectrum resource will be allocated to the future generation of a cellular network with a higher spectrum efficiency. For example, AT&T recently announced the closure of its GSM networks in 2017 [33]. Second, GSM is not able to handle the future massive number of MTC devices and cannot guarantee the QoS requirements for some M2M applications. Third, GSM can not satisfy the increasing demand for high data rate in M2M. Another limitation is that GSM requires MTC devices initiating connections [34], which can not satisfy the device trigger requirement [35]. Thus, a shift from 2G to 3G/4G or more advanced standards can be expected in the next decade for the already deployed commercial M2M solution.

The 3G family, UMTS and HSPA, is not a suitable technology for MTC because of the power efficiency and cost of the modem. Overall, it is an overkill technology in terms of design since it provides much more than needed [36]. As the roll out of 4G (mainly LTE and LTE-A) networks, the 4G technology is progressively attractive for MTC, among other reasons, the Orthogonal Frequency-Division Multiple Access (OFDMA) air interface allows the scaling of bandwidth according to needs. However, the modem cost and global coverage are still issues to be solved.

As discussed in a previous section, M2M applications can be classified into four categories: time-driven, query-driven, event-driven, and hybrid-driven. In both time-driven and event-driven types, the M2M device initiates the communication and uploads the gathered data in the form of either SMS or packet data. When M2M devices are self-triggered by an expected event, they first send uplink preambles to establish Radio Resource Control (RRC) connection. With the establishment of RRC connection, M2M devices connect to the core network (CN). Then, M2M devices establish connection with M2M server in TCP/application layer, which involve many transmission overhead.

In query-driven, the MTC device responses to the query from the MTC application server. For this kind of trigger mechanism, 3GPP identifies three possible models: direct model, indirect model, and hybrid model [37].

- Direct model. The first and the most straightforward deployment paradigm is the direct model, where the application server (AS) connects directly to an operator network in order to communicate with the M2M devices without using the services of any external service capability server (SCS)
- Indirect model. The second deployment paradigm is the indirect model, in which the AS connects indirectly to an operator network through the services of an SCS in order to utilize additional value-added services for M2M (e.g., control plane device triggering)
- The third deployment paradigm is the hybrid model, where the AS uses the direct model and indirect model simultaneously in order to directly connect to an operator network to perform direct user plane communications with the M2M devices while also using an SCS.

### 4.2 Standardization of reference M2M Architecture

Given that there is not a consensus about MTC reference architecture, the Standards Developing Organizations (SDO) and research community have proposed a few proposals. The European Telecommunications Standards Institute (ETSI) provides a general M2M reference architecture with the purpose of designing an access and transmission technology independent service middle layer [38]. Nowadays, the architecture-related works are transferred to oneM2M. The project oneM2M published their reference architecture at the beginning of 2015, which is similar but different than that of ETSI M2M. 3GPP proposes a MTC reference architecture with focus on improvement of the core network [35]. IEEE 802.16p gives an overall architecture for M2M [39]. The authors of [40] review and compare the aforementioned architectures then propose a hybrid reference model. Since the reference architectures of 3GPP and IEEE 802.16p are functionally equivalent, mainly the efforts of ETSI, oneM2M, and 3GPP are presented.

### 4.3 ETSI M2M reference architecture

The ETSI proposes a high-level reference architecture for M2M communication, which is illustrated in Fig. 3 and composed by a device-and-gateway domain and a network domain [38]. This architecture consists of the following components: (1) M2M-D, a device running M2M applications usually embedded in a smart device and replies to requests or sends data; (2) M2M area network, a capillary network (e.g., small-scale home environment) composed by individual M2M-D leveraging short-range communication technologies (e.g. IEEE 802.15.1, Zigbee, Bluetooth, etc.); (3) M2M-G, a proxy responsible for interworking for M2M area network and network domain; (4) access network, a network that provides access to the core network for the devices (i.e. M2M-D and M2M-G) in the device-and-gateway domain, it can be, among others, in the form of the access network of 3GPP, xDSL, satellite, and WiMAX; (5) core network, a network that provides various services such as IP connectivity, network control, interworking, and roaming between M2M-A and M2M-D. It includes, but is not limited to, 3GPP CN, ETSI TISPAN CN, and 3GPP2 CN; (6) M2M-A, M2M applications, application services that run the service logic and use M2M-SC via application programming interfaces; and (7) M2M-SC, a network node providing M2M functions to M2M-A and hiding network specificities for M2M application development.

### 4.4 oneM2M reference architecture

Apart from the efforts of ETSI M2M, other regional SDOs conduct standardization activities. To avoid the risks of divergence, the oneM2M partnership project was established in 2012 by leading regional SDOs such as ETSI (Europe), TIA and ATIS (North America), CCSA (China), TTA (Korea), and ARIB and TTC (Japan). The objective of oneM2M is to prepare, approve, and maintain globally applicable, access-independent technical specifications and reports related to M2M solutions, with initial focus on the service layer. The functional architecture of oneM2M is shown in Fig. 4. There exist mainly three functions: Application Entity (AE), Common Services Entity (CSE), and Network Services Entity (NSE). Application Entity is an entity in the application layer that implements an M2M application service logic and equivalent to the M2M-A in ETSI reference architecture (shown in Fig. 3). The Common Services Entity represents an instantiation of a set of *common service function* (i.e., M2M service subscription management, device management). A Network Services Entity hides the implementation of underlying communication networks and provides services to the CSEs. The oneM2M reference architecture reuses principles and solutions from ETSI M2M: the entity AE is like the M2M-A in Fig. 3. The CSE is equivalent to the M2M-SC in Fig. 3. The infrastructure is similar to the ETSI network domain and the field domain is like the device and gateway domain of ETSI.

### 4.5 3GPP reference MTC architecture

The main contribution of ETSI M2M overall architecture is to standardize the resource structure representing the information contained in M2M-SC, but ETSI has not specified the standardization for M2M area network, access network, and core network.

The efforts of 3GPP about MTC architecture are summarized in [35, 37]. Different from that of ETSI, the focuses of 3GPP are mainly cellular wireless network, especially the access and core network. The proposed reference architecture is shown in Fig. 5.

The enhancement made by 3GPP supports the device trigger function. To this end, two new network nodes

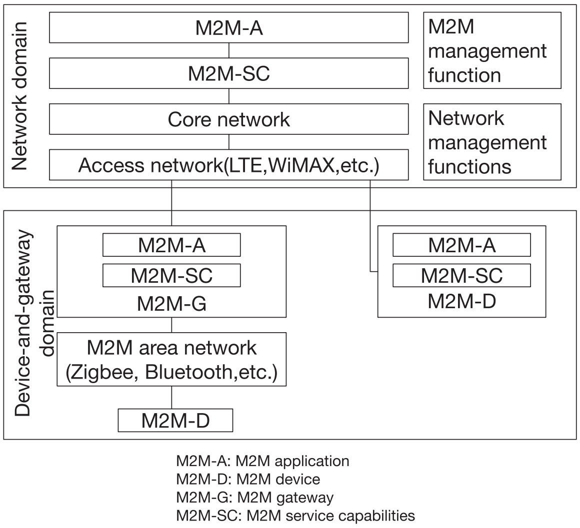

**Fig. 3** ETSI reference M2M communication architecture. (Modified based on [38])

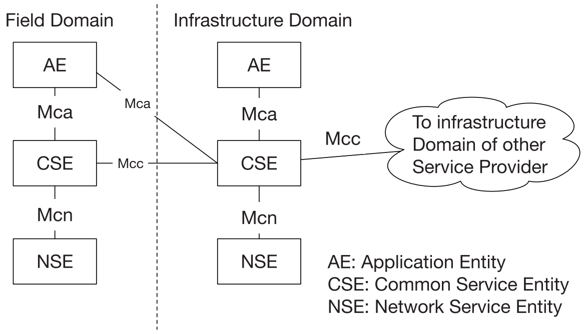

**Fig. 4** oneM2M reference functional M2M architecture. Source: [85]

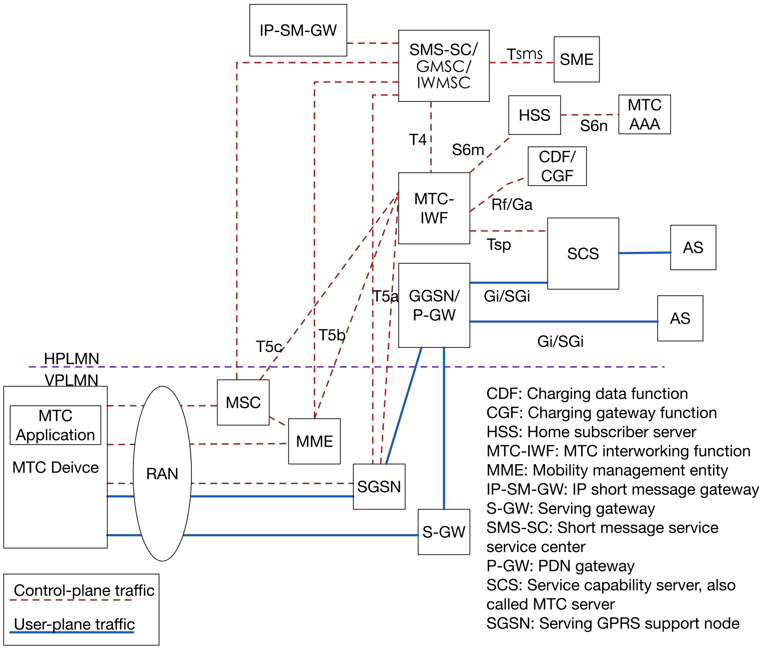

**Fig. 5** 3GPP reference MTC architecture. (based on [35])

(MTC-IWF and SCS) and a series of reference points related to these two nodes are introduced. The first node, MTC Interworking Function (MTC-IWF), hides the internal PLMN (Public Land Mobile Network) topology and relays or translates signaling protocols used over Tsp (shown in Fig. 5) to invoke specific functionality in the PLMN. The main functions of MTC-IWF are to authorize the SCS before communication establishment with the 3GPP network, receive a device trigger request from SCS, select the most efficient and effective device trigger delivery mechanism, etc. The SCS is an entity that connects to the 3GPP network to communicate with MTC devices and the MTC-IWF in the HPLMN. This entity offers capabilities to be used by one or multiple MTC Applications, and is controlled either by the mobile operator or a MTC service provider.

### 4.6 M2M architecture in the literature

Combining the respective advantages of proposals of 3GPP and ETSI, Lo et al. propose a cellular-centric M2M service architecture based on LTE-A specification [40], which is shown in Fig. 6. Their proposal is a four-tier architecture:

- Tier 1: the tier 1 consists of M2M applications and servers (namely M2M-A and M2M-S).
- Tier 2: the most distinguished change is at tier 2, in which a new functional entity M2M-R (M2M relay function) is introduced. M2M-R is an extension of the conventional LTE-A relay functionality. This

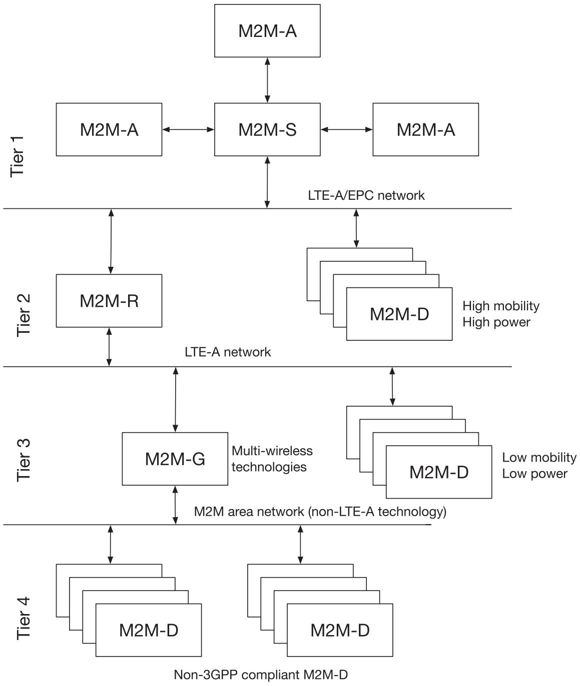

**Fig. 6** Hierarchical LTE-A cellular-centric M2M service architecture. Source: [40]

extension enables LTE-A relay node to act as a M2M data concentrator.
- Tier 3: the most important functional entity at tier 3 is M2M-G, which is actually a gateway to serve M2M devices non-3GPP compliant. The full-fledged 3GPP MTC-G implementation has not been standardized ant thus is still an open research issue.
- Tier 4: the tier 4 consists of those non-3GPP compliant M2M devices (e.g., devices using ZigBee)

The proposal in [40] actually leverages M2M gateway (M2M-R) to support non-3GPP compliant devices. The introduction of the M2M-R aggregates the packets from a large number of M2M devices into a single large packet, adds system capacity, and then reduces transmission.

## 5 Low-power wide-area network

The term LPWA network refers to a network relying on low-power and wide-area connectivity technology that simultaneously supports low batter energy consumption and wide coverage area. The LPWA network is dedicated to serve battery-powered applications characterized by low throughput, delay tolerance, and being event-driven such as water-meter monitoring. Unlike the other technologies that are adapted for Internet of Things (IoT), LPWA networks are purposely designed from scratch to meet wide-area IoT application. LPWA technologies are typically narrow-band (with some exceptions) and operate in the ISM license-exempt spectrum bands. Faced with a potential huge market, lots of players propose their solutions. Some typical and already deployed proprietary technologies of LPWA network are LoRaWAN, SIGFOX, Weightless, OnRamp, etc. A comprehensive comparison about these existing LPWA solutions is shown in Table 4.

LoRa alliance has issued their first vision of LoRaWAN specification [5] in January 2015, which is regarded as a major step towards international standardization in the field of LPWA networks. Thus, LoRaWAN technology is taken as a concrete example to give a general view about LPWA networks. The network architecture is illustrated in Fig. 7, which is a star-of-stars topology. A LoRaWAN network consists of the following components [5]:

- End-device: the end-device is the element in a LoRaWAN network which is responsible for collecting and uploading information to remote network server. LoRa supported functionalities can be classified to three classes: class A (bi-directional end-devices), class B (bi-directional end-devices with scheduled receive slots), and class C (bi-directional end-devices with maximal receive slots). All LoRaWAN end-devices at least support class A. According to applications, end-devices can optionally support class B and class C.
- LoRa air interface: The LoRa air interface provides the connectivity between LoRa end-devices and gateway. It is on ISM (Industrial Scientific Medical) band and based on LoRa modulation, which is a proprietary modulation scheme. The LoRa data rate ranges from 0.3 kbps to 50 kbps. The selection of data rate is a trade-off between communication range and message duration, and communications with different data rates do not interfere with each other.
- LoRa gateway: the LoRa gateway receives the communications from the LoRa end-devices and then transfers them to a network server via the backhaul system. Note that LoRa gateways may be co-located with a cellular base station. In this way, they are able to use spare capacity on the backhaul network.
- Network server: the LoRa network server manages the network. The network server acts to eliminate duplicate packets, schedules acknowledgment, and adapts data rates (adaptive data rate scheme). The communication between the LoRa gateway and the network server is IP-based, and the underlying carrier networks can be wired or wireless, Ethernet or 3GPP cellular, public or private networks.

In order to answer the huge expected demand of cellular M2M coverage, the standardization organizations embarked on a process of standardizing narrow-band technology for use in mobile spectrum. Two possible tracks are addressed by the 3GPP. The first track is the evolution of LTE 3GPP cellular system with the objective of reducing the occupied bandwidth but still reusing the basic LTE principles. The second track is to propose a clean slate solution, which features narrow-band (NB) technologies and leverage the existing cellular infrastructure. One major difference between these two tracks relies in that whether it should redesign the radio interface and multiple access control mechanism for cellular M2M networks. As an effort in the first track, the 3GPP developed LTE-M specification in Rel-12 [41] with introduction of a new low complexity device category (Cat-0). The device complexity of Cat-0 is 50 % of the previously defined Cat-1, which is the basic LTE terminal defined in the first LTE Release (Rel-8). Nowadays, 3GPP is considering to further optimize LTE-M in Rel-13: (1) bandwidth of 1.4 MHz and less complexity [41] and (2) a narrow-band evolution of LTE-M with bandwidth 200 kHz [7]. For the clean slate solutions, the main idea is to sacrifice the data rate in order to gain energy efficiency and coverage extension. They are supposed to satisfy the following requirements: deployment in a small bandwidth (e.g. 200 kHZ), ultra low-cost terminal (less than 5 dollars), ultra-long battery life, and coverage extension of 20 dB with existing cellular technologies. The typical solutions include Narrow Band M2M (NB M2M), Narrow

**Table 4** The comparison among LPWAN solutions (All solutions are on ISM band, extracted from [92])

| | LoRaWAN | NWave | OnRamp | SIGFOX | Telensa | Weightless-N | Weightless-P | Amber Wireless |
|---|---|---|---|---|---|---|---|---|
| Range (km) (Caveat) | 15–45 flat;<br>15–22 suburban;<br>3–8 urban | 10 | 4<br>(but claims 25× competition) | 50 rural;<br>10 urban | Up to 8 | 5+ | 2+ urban | Up to 20 |
| Band (MHz) | Spread;<br>varies by region | sub-GHz | 2.4 GHz | 868;902 | 868/915<br>470 (China) | Sub-GHz | Sub-GHz | 434, 868, 2.4GHz |
| Symmetric up/down | Depends on mode. Can be | No | No(4:1) | No | Yes | Uplink only | Not yet determined | |
| Data rate | 0.3–50 kbps (adaptive) | 100 bps | 8 bps–8 kpbs | 100 bps | low | 30–100 kbps | up to 100 kpbs | up to 500 kbps |
| Max nodes (Caveat) | Depends;<br>millions/hub | Million/base | "10s of 1000s" | Millions/hub | 150,000/server (moving to 500,000) | | 32767 NWs,<br>65535 hubs each,<br>16 M edge device per NW | 255 network of 255 nodes |
| Operational model | Public or private (expect 80 % public) | Public or private | Public or private | Public | Public | Public or private | Public or private | |
| Standard status (if any) | LoRa: proprietary<br>LoRaWAN: yes | Weightless-N | No | No | No<br>(perhaps in future) | Yes | In process | |

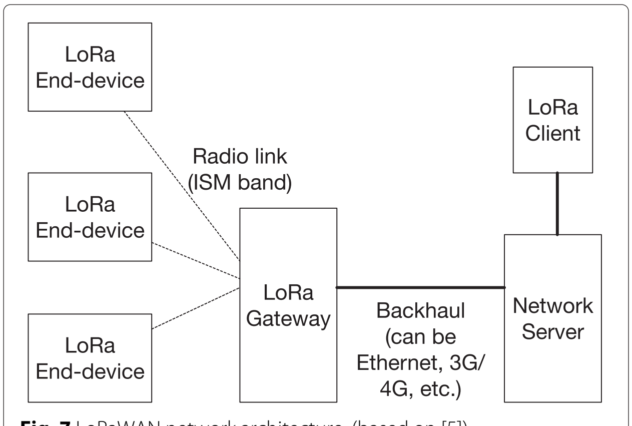

**Fig. 7** LoRaWAN network architecture. (based on [5])

Band OFDMA, and Cooperative Ultra Narrow Band (C-UNB) [42]. The deployment options include re-farming GSM spectrum, LTE band guard, and leftover fragments of spectrum during re-farming of 2G/3G to 4G.

When these standards will be available, the cellular M2M connectivity solutions may be more competitive, since they not only fulfill the requirements of extended coverage and long battery life, but also have the advantage of being able to operate in currently existing cellular network, thus requiring no additional deployment of antennas, radio, or other hardware. On the contrary, the proprietary (at least for the time-being) technologies such as Sigfox, On-Ramp, and Semtech require a dedicated network and maintenance team to deploy and maintain their services, which increases operational complexity for the operator. However, their M2M solution is currently available for the operators and starts to occupy some share of the market. In addition, some of the proprietary technologies such as LoRa have the plan to adapt their technology running on licensed spectrum and were submitted to GERAN [43] to keep their competitiveness.

## 6 Review about research proposals for energy-related cellular M2M

To achieve energy efficiency at the device side, the research community has done lots of efforts. In this section, we present, categorize, and compare all found proposals related to energy issues for MTC in cellular networks. The energy issues may refer to energy saving, energy efficiency, or power efficiency/saving. Note that we just concentrate on the research efforts about PHY/MAC layer. Obviously, the new routing algorithms [44, 45] help to achieve energy efficiency, especially for mobile ad hoc networks, but this is beyond the scope of this survey. In addition, there also exist mathematical works that are not categorized into the following section, but they provide useful design guidelines to help improve device side energy efficiency. For example, in [46], the transmission energy is modeled as a function of transmission power, packet size and link capacity. A cumulative distribution function (CDF) of energy consumption for large-scale MTC is derived by using stochastic geometry and Poisson point process. In [47], the comparison in terms of energy and power efficiency between uncoordinated and coordinated multiple access strategies are conducted. In [48], the work [47] is extended by considering various packet size and imperfect power control. The result of classification and comparison among the proposals presented in this survey is resumed in Table 5.

### 6.1 Cooperative relaying

Cooperative relaying, also called cooperative design [49], can be interpreted as the process of devices helping each other to jointly achieve a goal more efficiently than each device could do on its own. The first possible form is group-based and relay (in fact, at most two hops) mechanism, which is very useful for energy-saving, massive access control. A general and common description about group-based and relay transmission is illustrated in Fig. 8: all MTC devices are classified into several groups (some references call group as cluster). A certain device in each group is selected as the coordinator (also can be called cluster header, group head) according to some criteria (e.g., QoS requirements [50], link quality, or location). The MTC devices other than the coordinators transmit packets to their allocated coordinators, which relay the received packets to the BS (multiple-hop communication). In this model, there exist two links: MTCD-to-coordinator (actually MTCD-to-MTCD link, since the coordinator is by nature still a MTC device) and coordinator-to-BS. A basic issue is the interference between the two aforementioned links. This problem can be addressed by interference-based topology control algorithm [51], specific scheduling, the use of different frequencies between the two communications (but this is pricey in frequencies), or even the use of two different protocols as proposed in [12]. In addition, how to enable and make the MTCD-to-MTCD link efficient is addressed by device-to-device (D2D) communication [52, 53] and multiple-hop communication [54].

The idea of group-based communication is presented in [55] but without clarifying grouping/clustering algorithms. In fact, the algorithms of device grouping and the coordinator are a key factor influencing energy efficiency. A series of K-means (K-means clustering aims to partition $n$ observations into $k$ clusters in which each observation belongs to the cluster with the nearest mean) which derived grouping and coordinator selection algorithms are proposed in [56]. Ho and Huang [12] propose a two-stage mechanism to minimize the energy consumption of all MTC devices. The first stage consists of MTC device grouping and coordinator selection. The criteria

**Table 5** Categorization and comparison of energy/power saving-related proposals

| Category | Subcategory | Reference | Principle for energy saving | Drawback | Notes |
|---|---|---|---|---|---|
| Cooperative design | Clustering and relay | [55]<br>[56]<br>[12]<br>[57] | Group devices into clusters; Cluster head relays the messages for other cluster member in the same group | High energy consumption for cluster head; Scheduling and resource issues in order to manage the interference; Delay increase cause of relay | It is possible to combine cooperative relaying with other emerging technologies for the device-to-device link |
| | Cooperation between MTC server and MTC devices | [9] | Adjust MTC device setting according to context | High complexity for MTC devices | |
| | M2M gateway | [23]<br>[58]<br>[60] | Similar to clustering and relay, except that M2M gateway is a special network node instead of a MTC device | Installation and deployment cost of M2M gateway for operators | Reduce implementation complexity for MTC device.<br>Possible to use LoRa gateway as M2M gateway. |
| Design of energy efficient signaling and operation | Modified DRX and Idle state | [62]<br>[63] | Make MTC devices stay in low power mode as long as possible | High delay | Simple method to achieve energy saving |
| | Extending paging cycle | [1]<br>[2]<br>[64] | Make MTC devices stay in low power mode as long as possible | High delay | |
| | Reduction of RRC Inactivity timer | [65] | Make MTC devices stay in low power mode as long as possible | Impact on H2H service | |
| | Group-based and M2M- dedicated paging mechanism | [66] | Group paging for MTC devices | May reduce the paging capacity of H2H; Scalability and bacward-compatibility issues. | |
| | Removal of unnecessary activities | [66] | Remove activities related to mobility management (MM) for MTC Device | Applicable uniquely for M2M application with no or low mobility | Reduce the cost of MTC device and energy consumption |
| | Disconnect MTC-device from network when inactive | [67] | Instead of staying in low power mode, turn devices radio off | High delay | |
| Radio resource allocation and packet | Formulation of optimization problem | [71]<br>[72] | Convert radio resource allocation into an optimization problem with constraints | High complexity; Scalability issues; Possible impact for human | |

**Table 5** Categorization and comparison of energy/power saving-related proposals *(Continued)*

| Category | Subcategory | Reference | Principle for energy saving | Drawback | Notes |
|---|---|---|---|---|---|
| scheduling | | | | users. | |
| | Optimized with periodicity | [73]<br>[74] | Leverage the periodicity of MTC | Applicable uniquely to periodic M2M applications;<br>Not easy applied in the presence of different period values. | |
| | Packet scheduling | [76]<br>[29] | Propose packet scheduling adapted for cellular MTC context | Compatibility with H2H;<br>May need modifications of the specification | Ensure QoS requirements |
| Energy efficient random access and MAC | New random access protocol | [15] | ALOHA-based random access protocol is not the best option for MTC | It may be difficult to find a specified random access protocol suitable for all cellular users | |
| | Fixed time alignment | [78] | Leverage the low mobility of MTC | Only applicable for M2M with no mobility | Reduce the access collision |
| | Transmit message in MAC PDU or preamble | [23] | Transmit message directly in access reservation stage | Scalability issues | Reduce the signaling overhead |

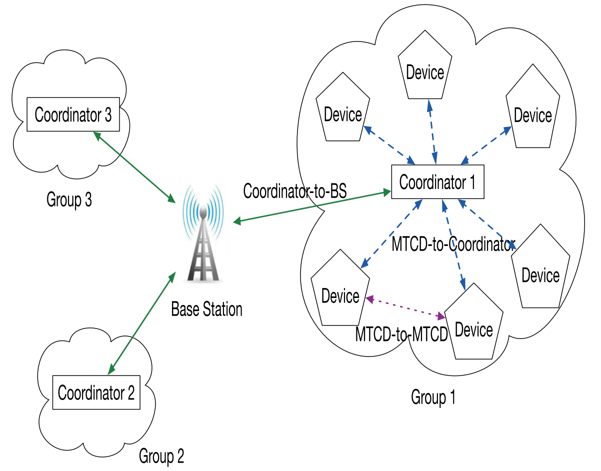

**Fig. 8** Illustration of M2M devices clustering

for grouping and coordinator selection in each group are the minimization of energy consumption of this group. The second stage is that BS performs power allocation for each coordinator to further reduce energy consumption. However, it is difficult to obtain the closed-form solution for the formulated problem, the proposal of [12] could achieve suboptimal result. An implementation of clustering and relay is presented in [57]: intra-cluster communication uses CSMA/CA protocol with multiple-phases while resource reservation-based protocol is used for communication between the cluster head and BS. The drawback of the group-based and relay design relies in that (i) although it is globally energy efficient for all the MTC devices, the cooperative relaying causes the cluster head to consume more energy than others; (ii) each MTC device should be equipped with multiple transceiver (e.g., OFDMA transceiver for coordinator-to-BS link, TDMA transceiver for MTCD-to-coordinator link), since every device is possible to be selected as the coordinator.

The second form of cooperation is to introduce M2M gateway (may be called proxy [23]). The M2M gateway serves as an intermediary node to collect and process data for neighbor MTC devices. Thus, topologically, M2M gateway is very similar to the aforementioned cluster header, except that M2M gateway does not have its own data to transmit and may have a permanent energy source. It is preferred to use half-duplex M2M gateway to avoid self-interference and reduce implementation complexity [25]. The use of M2M gateway helps reduce the number of accessing devices, signaling overhead and contention, and thus helps improve energy efficiency. Chen and Wang [23] give a simple work flow for MTC with M2M gateway. Pereira and Aguiar [58] consider the use of smartphones as M2M gateway between the BS and MTC devices. The existing scheduled airliners used as relays between ground devices and satellites are presented in [59]. It is also possible to add a gateway-like element node called M2M facilitator [60] between 3GPP RAN and core network, which is shown in Fig. 9. M2M devices nominally communicate with the M2M server. However, the M2M devices only communicate with the base station and enter in sleep mode after the session with the base station. The base station then transfers received data to the M2M facilitator. Finally, the M2M facilitator is in charge of data transmission, retransmission, and session termination with the M2M server. Since the devices communicate only with the M2M facilitator and the latter has no energy constraint, the protocol stack at the device side can be significantly simplified to save energy consumption. The cost is that a fraction of protocol stack complexity is transferred to the M2M facilitator. The inconvenient points of M2M gateway-related proposals are: (i) the operator should deploy M2M gateways to serve MTC devices, which may be of high cost and (ii) not flexible and dynamic compared with a clustering-based group.

Compared with the group-based and relay mechanism, the use of M2M gateway may be more energy efficient for devices. This is actually due to that a part of energy consumption of MTC devices is transferred to the M2M gateway, which may have a permanent energy source.

The third possible form of cooperation mechanism is context-based communication between the MTC servers and MTC devices. The philosophy of this axis is intelligent algorithms can be deployed at application level to help M2M devices and overall network adjust their settings to improve energy efficiency. According to this philosophy, the authors propose a context-aware framework based on the context concept in [9]. M2M devices send context-related information to the MTC server, and the MTC server judges the data reporting mode (time-driven, query-driven, event-driven, or hybrid) and QoS feature (real-time, priority, and accuracy) then returns a set of output useful for the setting device's behavior. This output involves inter-arrival time, average packet size, transmitting power, packet omission, etc. that could extend the operative lifetime of M2M devices.

### 6.2 Design of energy-efficient signaling and operation

An intuitive solution to achieve energy-efficiency for MTC is to design energy-efficient signaling and simplify

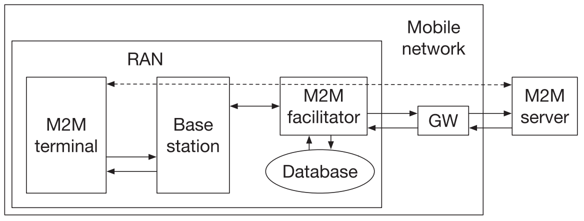

**Fig. 9** Un-peer-to-peer cellular MTC architecture with the M2M facilitator. Source: [60]

the operations leading to energy wasting. The key philosophy is to maximize the duration in a low-power state and reduce or even delete all operations unnecessary for MTC. The improvement possibilities are summarized as follows:

(1) Adaption of Discontinuous Reception (DRX) and idle state: 3GPP has incorporated some energy/power-saving mechanisms into their specification for H2H devices, for instance, DRX [61]. Due to long inter-arrival time between traffic sessions, MTC devices stay in LTE idle states most time as it is designed as a low-power state. In the idle state, MTC devices go to sleep to save battery and wake up periodically to inquire any system information (SI) update or downlink packet arrival via paging mechanism. However, the aforementioned mechanisms are still insufficient for MTC device, and DRX and idle state are adapted to improve device power saving in [62]. The potential of trading high delay for reducing MTC devices battery consumption is studied in [63].
(2) Extending paging cycle: 3GPP considers to extend paging cycle for M2M devices [1, 2, 64], but simulation results show that extending paging cycle beyond 2.56 s reduces the power consumption significantly for M2M traffic with small or medium values of inter-arrival time, but has no effect when the paging cycle is more than a limit [65]. In addition, extension of the paging cycle always increases the packet buffering delay.
(3) Reduction of RRC inactivity timer: the network keeps MTC device in connected mode even after the last packet delivery due to the RRC inactivity timer in which MTC device still keep a high power, thus the impact of variation of RRC inactivity timer is explored and proved to be better than extending paging cycle [65].
(4) Group-based and M2M-dedicated paging mechanism: to increase paging capacity, multiple terminals are allocated with the same paging occasion (PO). H2H and M2M terminals may occupy the same certain paging occasion. If both are paged together, a large number of H2H terminals and MTC devices need to wake up at the same PO, which results in terminal power wasting. The solution to avoid this kind of power wasting is paging targets with group ID or device ID at dedicated paging occasion allocated uniquely for M2M devices [66];
(5) Removal of unnecessary operations: even in the idle state, the MTC devices are not actually inactive: they are supposed to run activities related to mobility management (e.g., TAU procedure for LTE). Since MTC presents low mobility feature, it is possible to remove these unnecessary operations (periodic AS measurement and NAS LAU/RAU/TAU procedure) for power saving [66].
(6) Disconnect MTC device from network when inactive: since the device power consumption in idle state is not negligible, it is possible to turn the LTE radio off (i.e., disconnect the device from the network) to obtain further power saving [67] and this is proved to be outperforming than the approach of reducing RRC inactivity timer. However, the gain of power saving is with cost of higher delay related to downlink packet. In addition, the network may also need to store the downlink data, and the MTC device identities if the MTC device is off.

### 6.3 Radio resource allocation and packet scheduling strategies

Radio resource allocation and packet scheduling strategies play a key role in the overall performance of OFDMA-based wireless networks [68]. Most of the research efforts in this field are about the downlink, which can for example improve network throughput and mitigate inter-cell interference [69, 70]. However, they have limited effectiveness about energy efficiency on the MTC device side. In addition, MTC applications are usually uplink-centric; thus, it is import to design new radio resource allocation and packet scheduling schemes to achieve energy-efficiency.

Zheng et al. [25] study the radio resource allocation scheme for the five possible links in an M2M and H2H co-existence scenario: the eNB-to-UE link, the eNB-to-MTCD link, the eNB-to-MTCG link, the MTCG-to-MTCD link, and the MTCD-to-MTCD link. They first design a new radio resource partition scheme and then propose efficient radio resource allocation algorithms in each partition to mitigate co-channel interference and enhance network efficiency, which is useful for energy-efficiency, but they just consider the data rate when allocating resource for both UE and MTC devices.

Given that MTC usually features low data rate and emphasizes the delay requirement, Aijaz and Aghvami [71] extend the work [25] by formulating a bits-per-joule capacity maximization optimization problem. The resource constraint for UE is the minimum data rate and maximum tolerable packet delay for MTC devices. They propose two heuristic algorithms based on the steepest descent approach to solve the optimization problem. Aijaz et al. [72] further extend [71] by introducing the notion of statistical QoS (i.e., probability of exceeding a specific delay threshold) and solve the optimization problem with canonical duality theory (CDT).

Radio resource allocation scheme can also leverage the periodicity of MTC, since a considerable M2M users repetitively access the networks to transmit collected data. Zhang [73] proposes to use persistent resource allocation for periodic M2M applications and indicate the condition for multiplexing multiple MTC devices with different reporting periods. Madueno et al. [74] propose a periodically occurring pool of resources that are reserved for M2M communications and shared for uplink transmission by all MTC devices. Song et al. [75] propose a multiple-period polling service in LTE transport network to avoid random access procedure by leveraging the periodicity feature of M2M.

Group-based feature can be leveraged when designing M2M-compatible radio resource allocation strategies. Since MTC devices in the same cluster are assumed to have exactly the same QoS requirements, a grouping-based radio resource allocation algorithm [29, 50] is proposed for LTE-A base stations according to packet arrival rate and maximum jitter. The access grant time interval (AGTI) is periodically allocated for each MTC device cluster according to cluster priority. All the MTC devices of a same cluster occupy an equal number of resource block (RB) in allocated AGTI. The shortcomings of this proposal are as follows: (i) the number of served MTC devices is limited due to the inefficient utilization of resource; (ii) the base station only supports a unique packet size for all MTC devices; (iii) the proposal is not scalable since BS has to know in advance how many clusters are there in its coverage; and (iv) the supported QoS classes are limited.

Since the packet delay employed in [29] is a deterministic bound, Gotsis et al. [76] extend the work of [29] by using a statistical QoS (also used in [71]), which refers to the probability of exceeding a specific delay threshold. They propose an analytical model to study the performance of period scheduling algorithm in terms of statistical QoS metric with modeling the arrival traffic as Poisson process. They also enhance the periodic scheduling in [29] with queue-awareness in which devices with larger queues than the others are first granted access to the scheduled AGTI with an extra cost in complexity and signaling.

To overcome the issue of QoS classes, two uplink packet scheduling strategies are proposed in [77], which take into account both the channel conditions and the maximum allowed delay of each device; however, it suffers from increased signaling requirements and is just able to serve limited number of MTC devices. In [12] a resource allocation scheme (i.e., optimal transmit power allocation over RBs) for M2M traffic over OFDMA frames is proposed, assuming a two-hop access to the LTE network though a coordinator. This work achieves the reduction of energy consumption but does not take into account QoS issue of MTC such as the delay requirements.

### 6.4 Energy-efficient random access procedure and MAC

The currently standardized random access procedure, for example, in LTE networks, is designed and optimized for large amount of data transmission and limited UEs; thus, it suffers from a random access overload issue which leads to high collision probability and waste of energy. The improvement works about random access procedure have been attracting the attention of the research community. Current random access optimization research efforts can be resumed into two categories: (i) improvement for currently employed ALOHA random access procedure and (ii) designs of a non-ALOHA procedure.

For the first category, the focus is to reduce either signaling overhead for small data transmission or contention probability, since both reduce the transmitted bits for MTC devices and thus help energy saving. For MTC with lots of fixed-location machines, Ko et al. [78] propose a novel random access scheme based on fixed TA (timing alignment) for a OFDMA-based cellular system. The proposal is based on the assumption that the TA value between each fixed-location machine device and eNB is fixed and unchanged, and the MTC devices store the TA value acquired from the initial RA access and compare it with the TA values obtained from the subsequent random access procedure. In case of mismatch of TA values, MTC devices directly start the retransmission procedure to avoid possible collision after waiting a randomly selected backoff time. Otherwise, the devices continue the conventional procedure. However, this proposal is only applicable to M2M with stationary devices and has a limited effect on energy efficiency.

For M2M applications with small data transmission, establishing RRC connection and network connection to transmit several bits is deemed as wasteful. Thus, Chen and Wang [23] suggest either (i) use the MAC PDU that should carry the RRC signaling to carry the data and (ii) define a special preamble to transmit coded data. The drawbacks of [23] are as follows: (i) the solution based on the preambles is not very scalable due to the limited amount of available preambles and (ii) for a long-term view, the transmission of data in the control plane violate the principle of separation between the control plane and the data plane. Wiriaatmadja and Choi [79] propose to simplify the data communication procedure by allowing MTC devices to send data right after the preamble transmission without explicitly establishing a connection.

For the congestion in random access, Physical Downlink Shared Channel (PDSCH) resources of LTE are deemed sufficient in most communication scenarios. To ease the congestion on the air interface, those downlink assignments and uplink grants for MTC devices, which cannot be served by Physical Downlink Control Channel (PDCCH), can be aggregated into a transport block on PDSCH identified by a special radio network terminal identifier (RNTI) called MTC-RNTI [80]. MTC devices monitor PDCCH channel with their own cell RNTI and MTC-RNTI simultaneously. Game theory has been used for the context of cellular M2M to optimize preamble allocation [81]. In addition, a detailed random access related proposals are summarized in [15], which can be a complement of our categorization.

Random access protocols can be categorized into two families: ALOHA family and tree family [82]. Andres et al. [15] claim that ALOHA-based RACH procedure is not suitable for MTC. Instead, they mention that RACH procedure based on distributed queuing (DQ) is more promising. The concept of DQ [82] was proposed 20 years ago and then demonstrated in other literatures in terms of stability and near-optimum behavior. DQ is based on the combination of a m-array tree splitting algorithm with a smart set of simple rules that allow organizing every device in one out of two virtual queues. Due to the rules of DQ, it behaves as a random access method for low traffic loads, and it switches smoothly and seamlessly to a reservation access method as the traffic load increases. The authors conduct some ongoing research efforts applying DQ ideas within LTE/LTE-A systems. Dhillon et al. [47] suggest to implement a load-dependent access scheme wherein uncoordinated strategy is for light load and coordinated strategy for heavy load. Bontu et al. [83] propose a new uplink (UL) physical, transport, and logical channel: common traffic channel (CTCH), UL simultaneous-access shared channel (UL-SSCH), and physical uplink simultaneous access shared channel (PUSSCH). The aforementioned channels enable M2M devices to simultaneously transmit data packet in the same radio resource. They also propose to transmit control signaling through in-band transmission in the user plane control.

In fact, the cellular network evolution trend is always to seek for a trade-off between diverse performance metrics such as energy efficiency, packet delay, and user data rate. Hence, the gain of energy efficiency is inevitably with cost of a certain degradation of other performance metrics. For example, the cooperative relaying achieves energy efficiency with more packet delay due to multiple-hop transmission. The design of energy-efficient signaling and operation, such as disconnection of the MTC devices from a network when inactive, surely saves energy consumption but introduce higher delay for downlink packet, and the effort in this direction is not systematic. The energy-efficient uplink radio resource allocation and packet scheduling schemes provide a systematic manner to gain energy efficiency for MTC, but they may lead to either serving less number of MTC devices and supporting limited QoS classes or bringing more signaling messages in radio access networks. More importantly, it is difficult to design schemes simultaneously satisfying the QoS provisioning for both human and MTC users. The random access can be designed to reduce the retransmission probability for MTC users, but human users may suffer from degradation of service, due to the limited radio access resources. Therefore, it is very important to jointly apply the aforementioned approaches to gain the device side energy efficiency and seek for a trade-off with other system performance. For example, the random access procedure can be optimized by applying cooperative relaying to reduce the direct links towards the base station. The energy-efficient uplink radio resource allocation algorithms can be jointly designed with a random access procedure: allocate more resources for PRACH in case of overload or limit the number of access devices when no radio resource for data transmission, etc.

## 7 Conclusions

The arrival of billions of connected machines in the short and mid terms is a huge challenge for cellular networks, although not all these machines will necessarily be connected to the latter. A part of the machines will only be connected to the ad hoc networks between each other, while some other will rely on the dedicated networks of LPWAN style. Yet, a large part of the machines will be better served within 3GPP cellular networks, especially the future 5G networks. Nowadays, 2G-style GSM or GPRS cellular is often used for cellular M2M service. However, GPRS is not suitable for M2M and often not competitive compared with LPWA solution, since it can not support a large number of devices, among other reasons. LPWA technologies and cellular 3GPP solutions will be the main support used for cellular M2M. In this paper, we describe the present state of these technologies and the evolutions as expected today.

We propose a synthesis for the QoS demands and the difference of characteristics between H2H and M2M. We then review the proposals for radio coverage and service of these machines. We identify the advantage of cellular networks for this expected service. The 3GPP cellular networks have a mature infrastructure to provide a wide-coverage, high-availability service and user subscription/management system, but the shortages of 3GPP networks are relatively high energy consumption level and cost of hardware with regard to LPWA networks. These challenges are addressed by some research proposals that we summarize in this article. In terms of LPWA network such as LoRa, their significant advantage is their low energy consumption design and low-cost hardware. However, their disadvantage is that the operators should deploy dedicated infrastructure for providing LPWA-related service.

According to our survey, to improve the device side energy efficiency in the future cellular networks, we get some design guidelines as follows:

- The possible approaches to improve device side energy efficiency for cellular MTC include the following: cooperative relaying, design of energy-efficient signaling and operation, radio resource allocation and packet scheduling strategies, and energy-efficient random access procedure and MAC.

- It is a better solution to employ cooperative relaying, since it can be combined with other emerging technologies such as D2D communication, ad hoc networks research results, and LoRa technology.
- The radio resource allocation and packet scheduling schemes allows to get energy efficiency while keeping a certain level of QoS; however, it is difficult to design this kind of schemes simultaneously satisfying the QoS provisioning for both human and MTC users.
- No matter by which approach, to gain energy efficiency is always with sacrifice of other system performances such as packet delay. Thus, it is important to jointly use the aforementioned approaches, for example, joint design of random access control and radio resource allocation, to seek for a trade-off between energy efficiency and other system performances.

**Acknowledgments**
The authors would like to thank Professor Michela Meo, from Politecnico di Torino, for the fruitful discussions about this work.

**Competing interests**
The authors declare that they have no competing interests.

Received: 20 February 2016 Accepted: 11 May 2016

**References**

1. 3GPP, Service requirements for machine-type communications, TS 22.368 V11.0.0 (2010). http://www.3gpp.org/DynaReport/22368.htm
2. 3GPP, Study on RAN improvements for machine-type communication. TS 37.868 V11.0.0 (2010). http://www.3gpp.org/DynaReport/37868.htm
3. Ericsson, More than 50 billion connected devices. White paper (2011). http://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf
4. MZ Shafiq, L Ji, AX Liu, et al., A first look at cellular machine-to-machine traffic: large scale measurement and characterization. SIGMETRICS Perform. Eval. Rev. **40**(1), 65–76 (2012)
5. N Sornin, M Luis, T Eirich, T Kramp, Lorawan specification. Technical Report V.1.0, LoRa Alliance (2015). https://www.lora-alliance.org/portals/0/specs/LoRaWANSpecification1R0.pdf
6. LoRa Alliance. https://www.lora-alliance.org. Accessed 14 Oct 2015
7. R Ratasuk, N Mangalvedhe, A Ghosh, B Vejlgaard, in *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*. Narrowband lte-m system for m2m communication (IEEE, 2014), pp. 1–5
8. R Lu, X Li, X Liang, X Shen, X Lin, GRS: The green, reliability, and security of emerging machine to machine communications. IEEE Commun. Mag. **49**(4), 28–35 (2011)
9. JM Costa, G Miao, in *INFOCOM Workshops*. Context-aware machine-to-machine communications (IEEE, 2014), pp. 730–735
10. L Suarez, L Nuaymi, J-M Bonnin, An overview and classification of research approaches in green wireless networks. Eurasip J. Wirel. Commun. Netw. **2012**, 1–18 (2012)
11. C-Y Wang, C-H Ko, H-Y Wei, AV Vasilakos, A voting-based femtocell downlink cell-breathing control mechanism. IEEE/ACM Transactions on Networking. **24**(1), 85–98 (2014)
12. CY Ho, C Huang, Energy-saving massive access control and resource allocation schemes for M2M communications in ofdma cellular networks. IEEE Wirel. Commun. Lett. **1**(3), 209–212 (2012)
13. T Taleb, A Kunz, Machine type communications in 3GPP networks: potential, challenges, and solutions. IEEE Commun. Mag. **50**, 178–184 (2012)
14. K-C Chen, S-Y Lien, Machine-to-machine communications: technologies and challenges. Ad Hoc Netw. **18**, 3–23 (2014)
15. L Andres, A Luis, AZ Jesus, Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives. IEEE Commun. Surv. Tutor., 4–16 (2014)
16. J Poncela, JM Moreno-Roldan, M Aamir, B Alvi, M2M challenges and opportunities in 4G. Wirel. Pers. Commun., 1–14 (2015)
17. F Ghavimi, H-H Chen, M2M communications in 3GPP LTE/LTE-A networks: architectures, service requirements, challenges, and applications. IEEE Commun. Surv. Tutor. **17**(2), 525–549 (2015)
18. Y Mehmood, C Görg, M Muehleisen, A Timm-Giel, Mobile M2M communication architectures, upcoming challenges, applications, and future directions. EURASIP EURASIP J. Wirel. Commun. Netw. **2015**, 1–37 (2015)
19. X Wang, AV Vasilakos, M Chen, Y Liu, TT Kwon, A survey of green mobile networks: opportunities and challenges. Mob. Netw. Appl. **17**(1), 4–20 (2012)
20. M Ismail, W Zhuang, E Serpedin, K Qaraqe, A survey on green mobile networking: from the perspectives of network operators and mobile users. IEEE Commun. Surv. Tutor. **17**(3), 1535–1556 (2015)
21. M Yang, Y Li, D Jin, L Zeng, X Wu, AV Vasilakos, Software-defined and virtualized future mobile and wireless networks: a survey. Mob. Netw. Appl. **20**(1), 4–18 (2015)
22. OECD, Machine-to-machine communications: connecting billions of devices. OECD Digital Economy Papers, No. 192 (2012). doi:10.1787/5k9gsh2gp043-en
23. Y Chen, W Wang, in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*. Machine-to-machine communication in LTE-A (IEEE, 2010), pp. 1–4
24. H Tullberg, P Popovski, et al., in *IEEE Communications Magazine*. METIS system concept: The shape of 5G to come, (2015). https://www.metis2020.com/wp-content/uploads/publications/IEEE_CommMag_2015_Tullberg_etal_METIS-System-Concept.pdf
25. K Zheng, F Hu, W Wang, et al., Radio resource allocation in LTE-advanced cellular networks with M2M communications. IEEE Commun. Mag. **50**(7), 184–192 (2012)
26. JN Al-Karaki, AE Kamal, Routing techniques in wireless sensor networks: a survey. IEEE Wirel. Commun. **11**(6), 6–28 (2004)
27. LM Borges, FJ Velez, AS Lebres, Survey on the characterization and classification of wireless sensor network applications. IEEE Commun. Surv. Tutor. **16**(4), 1860–1890 (2014)
28. IEEE802.16, Machine to machine M2M communication study report. Ts. (2010). http://www.ieee802.org/16/ppc/docs/80216ppc-10_0002r7.doc
29. S Lien, K Chen, Massive access management for QoS guarantees in 3GPP machine-to-machine communications. Commun. Lett., IEEE. **15**(3), 311–313 (2011)
30. K Zheng, S Ou, J Alonso-Zarate, M Dohler, F Liu, H Zhu, Challenges of massive access in highly dense lte-advanced networks with machine-to-machine communications. IEEE Wirel. Commun. **21**(3), 12–18 (2014)
31. M Laner, P Svoboda, N Nikaein, M Rupp, in *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium On*. Traffic models for machine type communications (VDE, 2013), pp. 1–5
32. I Curran, S Pluta, in *Water Event, 2008 6th Institution of Engineering and Technology*. Overview of machine to machine and telematics, (2008), pp. 1-33
33. AT&T, Frequently Asked Questions Regarding 2G Sunset (2014). http://www.business.att.com/content/other/2G_Sunset_FAQs_2014_A1.pdf. Accessed 27 Oct 2015
34. M Martsola, T Kiravuo, J Lindqvist, in *Mobile Technology, Applications and Systems, 2005 2nd International Conference On*. Machine to machine communication in cellular networks (IEEE, 2005), p. 6
35. 3GPP, System improvements for machine-type communications. TS 23.888 V11.0.0. (2012)
36. CA-H Dohler, (ed.), in *Machine-to-machine (M2M) Communications*. Chapter 3 - overview of 3GPP machine-type communication standardization (Woodhead Publishing, Oxford, 47)

37. 3GPP, Architecture enhancements to facilitate communications with packet data networks and applications. Technical Report TS 23.682 V 12.4.0. (2015). http://www.3gpp.org/DynaReport/23682.htm
38. ETSI, Machine-to-machine communications (M2M); functional architecture. TS 102 690, ETSI. (2011)
39. IEEE802.16p, Machine-to-machine (M2M) system requirements document. Technical Report. IEEE 802.16 Broadband Wireless Access Working Group (2011). http://ieee802.org/16/m2m/#10_0004
40. A Lo, Y Law, M Jacobsson, A cellular-centric service architecture for machine-to-machine (M2M) communications. Wirel. Commun., IEEE. **20**(5), 143–151 (2013)
41. Nokia, LTE-M—optimizing LTE for the Internet of Things. White paper (2015). http://networks.nokia.com/sites/default/files/document/nokia_lte-m_-_optimizing_lte_for_the_internet_of_things_white_paper.pdf
42. 3GPP, Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT). TR 45.820. 3rd Generation Partnership Project (3GPP) (2015). http://www.3gpp.org/DynaReport/45820.htm
43. F Rayal, Shaping cellular IoT connectivity: emerging technologies in wide-area Connectivity (2015). http://www.xonapartners.com/wp-content/uploads/2015/07/Shaping-Cellular-IoT-Connectivity.pdf. Accessed 1 Dec 2015
44. A Dvir, AV Vasilakos, Backpressure-based routing protocol for dtns. ACM SIGCOMM Comput. Commun. Rev. **41**(4), 405–406 (2011)
45. P Li, S Guo, S Yu, AV Vasilakos, Reliable multicast with pipelined network coding using opportunistic feeding and routing. IEEE Trans. Parallel Distrib. Syst. **25**(12), 3264–3273 (2014)
46. M Khoshkholgh, Y Zhang, K Shin, V Leung, S Gjessing, in *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*. Modeling and characterization of transmission energy consumption in machine-to-machine networks (IEEE, 2015), pp. 2073–2078
47. HS Dhillon, HC Huang, et al., Power-efficient system design for cellular-based machine-to-machine communications. Wirel. Commun. IEEE Trans. **12**(11), 5740–5753 (2013)
48. Q Song, L Nuaymi, X Lagrange, in *Wireless Communications and Networking Conference Workshops (WCNCW), 2016 IEEE*. Evaluation of multiple access strategies with power control error and variable packet length in m2m (IEEE, 2016)
49. A Bartoli, M Dohler, J Hernández-Serrano, A Kountouris, D Barthel, in *NETWORKING 2011 Workshops*. Low-power low-rate goes long-range: the case for secure and cooperative machine-to-machine communications (Springer, 2011), pp. 219–230
50. S Lien, K Chen, Y Lin, Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. IEEE Commun. Mag. (2011)
51. XM Zhang, Y Zhang, F Yan, AV Vasilakos, Interference-based topology control algorithm for delay-constrained mobile ad hoc networks. IEEE Trans. Mob. Comput. **14**(4), 742–754 (2015)
52. Y Niu, Y Li, D Jin, L Su, AV Vasilakos, A survey of millimeter wave communications (mmWave) for 5g: opportunities and challenges. Wirel. Netw. **21**(8), 2657–2676 (2015)
53. Y Niu, C Gao, Y Li, L Su, D Jin, AV Vasilakos, Exploiting device-to-device communications in joint scheduling of access and backhaul for mmWave small cells. IEEE J. Sel. Areas Commun. **33**(10), 2052–2069 (2015)
54. T Meng, F Wu, Z Yang, G Chen, AV Vasilakos, Spatial reusability-aware routing in multi-hop wireless networks. IEEE Trans. Comput. **65**(1), 244–255 (2016)
55. RY Kim, in *Information and Communication Technology Convergence (ICTC), 2010 International Conference On*. Snoop based group communication scheme in cellular machine-to-machine communications (IEEE, 2010), pp. 380–381
56. C Tu, C Ho, C Huang, in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*. Energy-efficient algorithms and evaluations for massive access management in cellular based machine to machine communications (IEEE, 2011), pp. 1–5
57. A Azari, G Miao, in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference On*. Energy efficient MAC for cellular-based M2M communications (IEEE, 2014), pp. 128–132
58. C Pereira, A Aguiar, Towards efficient mobile M2M communications: survey and open challenges. Sensors. **14**(10), 19582–19608 (2014)
59. S Plass, M Berioli, R Hermenier, in *Future Network & Mobile Summit (FutureNetw), 2012*. Concept for an M2M communications infrastructure via airliners (IEEE, 2012), pp. 1–8
60. Y Chen, Y Yang, in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*. Cellular based machine to machine communication with un-peer2peer protocol stack (IEEE, 2009), pp. 1–5
61. 3GPP, Evolved Universal Terrestrial Radio Access (EUTRA); User Equipment (UE) procedures in idle mode. TS 36.304 V11.3.0, 3GPP. (2013). http://www.3gpp.org/ftp/Specs/html-info/36304.htm
62. M Gupta, SC Jha, AT Koc, R Vannithamby, Energy impact of emerging mobile internet applications on lte networks: issues and solutions. IEEE Commun. Mag. **51**(2), 90–97 (2013)
63. T Tirronen, A Larmo, J Sachs, B Lindoff, N Wiberg, in *Globecom Workshops (GC Wkshps), 2012 IEEE*. Reducing energy consumption of lte devices for machine-to-machine communication (IEEE, 2012), pp. 1650–1656
64. 3GPP, Study on machine-type communications MTC and other mobile data applications communications enhancements. TS 23.887 V12.03, 3GPP (2013). http://www.3gpp.org/DynaReport/23887.htm
65. SC Jha, AT Koc, et al., in *Communications and Networking (BlackSeaCom), 2013 First International Black Sea Conference On*. Power saving mechanisms for M2M communication over LTE networks (IEEE, 2013), pp. 102–106
66. H Chao, Y Chen, J Wu, in *GLOBECOM Workshops*. Power saving for machine to machine communications in cellular networks, (2011), pp. 389–393
67. SC Jha, AT Koc, R Vannithamby, in *Communications Workshops (ICC), 2014 IEEE International Conference On*. Device power saving mechanisms for low cost MTC over LTE networks (IEEE, 2014), pp. 412–417
68. AG Gotsis, NT Koutsokeras, P Constantinou, in *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*. Radio resource allocation and packet scheduling strategies for single-cell ofdma packet networks (IEEE, 2007), pp. 1847–1851
69. D López-Pérez, X Chu, AV Vasilakos, H Claussen, On distributed and coordinated resource allocation for interference mitigation in self-organizing lte networks. IEEE/ACM Trans. Networking (TON). **21**(4), 1145–1158 (2013)
70. M Ding, D López-Pérez, R Xue, AV Vasilakos, W Chen, in *Communications (ICC), 2014 IEEE International Conference On*. Small cell dynamic TDD transmissions in heterogeneous networks (IEEE, 2014), pp. 4881–4887
71. A Aijaz, AH Aghvami, in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*. On radio resource allocation in LTE networks with machine-to-machine communications (IEEE, 2013), pp. 1–5
72. A Aijaz, M Tshangini, MR Nakhai, et al., Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees. Commun., IEEE Trans. **62**(7), 2353–2365 (2014)
73. Y Zhang, Tree-based resource allocation for periodic cellular M2M communications. Wirel. Commun. Lett., IEEE. **3**(6), 621–624 (2014)
74. GC Madueno, C Stefanovic, P Popovski, Reliable reporting for massive M2M communications with periodic resource pooling. Wirel. Commun. Lett., IEEE. **3**(4), 429–432 (2014)
75. Q Song, X Lagrange, L Nuaymi, in *Vehicular Technology Conference (VTC Fall)*. An efficient M2M-oriented network-integrated multiple-period polling service in LTE network (IEEE, 2015), pp. 1–6
76. AG Gotsis, AS Lioumpas, A Alexiou, in *Globecom Workshops (GC Wkshps), 2012 IEEE*. Evolution of packet scheduling for machine-type communications over LTE: algorithmic design and performance analysis (IEEE, 2012), pp. 1620–1625
77. AS Lioumpas, A Alexiou, in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*. Uplink scheduling for machine-to-machine communications in LTE-based cellular systems (IEEE, 2011), pp. 353–357
78. K Ko, M Kim, et al., A novel random access for fixed-location machine-to-machine communications in OFDMA based systems. IEEE Commun. Lett. **9**, 1428–1431 (2012)
79. DT Wiriaatmadja, KW Choi, Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks. IEEE Trans. Wirel. Commun. **14**(1), 33–46 (2015)
80. B Yang, G Zhu, W Wu, Y Gao, M2M access performance in LTE-A system. Trans. Emerg. Telecommun. Technol. **25**(1), 3–10 (2014)
81. M Hasan, E Hossain, D Niyato, Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches. IEEE Commun. Mag. **51**(6), 86–93 (2013)
82. W Xu, G Campbell, in *Communications, 1992. ICC'92, Conference record, SUPERCOMM/ICC'92, Discovering a New World of Communications., IEEE*

*International Conference on*. A near perfect stable random access protocol for a broadcast channel (IEEE, 1992), pp. 370–374
83. CS Bontu, S Periyalwar, M Pecen, Wireless wide-area networks for Internet of Things: an air interface protocol for IOT and a simultaneous access channel for uplink IOT communication. IEEE Veh. Technol. Mag. **9**(1), 54–63 (2014)
84. Cisco, Global Mobile Data Traffic Forecast Update, 2014-2019. White paper. Cisco Visual Networking Index (2015)
85. oneM2M, Functional Architecture. Technical Report TS 0001, V.1.6.1. (2015). http://www.onem2m.org/images/files/deliverables/TS-0001-Functional_Architecture-V1_6_1.pdf
86. R Ratasuk, J Tan, A Ghosh, in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*. Coverage and capacity analysis for machine type communications in LTE (IEEE, 2012), pp. 1–5
87. 3GPP, Smart grid traffic behaviour discussion. TSG-RAN R2-102340 (2012)
88. Health Informatics - PoC Medical Device Communication - part 00101: Guide—guidelines for the use of RF wireless technology (2008)
89. R Ratasuk, A Prasad, Z Li, A Ghosh, M Uusitalo, in *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference On*. Recent advancements in M2M communications in 4G networks and evolution towards 5G (IEEE, 2015), pp. 52–57
90. D Boswarthick, O Hersent, O Elloumi, *M2M Communications : a Systems Approach*. (Wiley-Blackwell, Oxford, 2012)
91. J Chou, Machine-to-machine (M2M) communications using short message services (SMS). Google Patents. EP Patent App. EP20,110,869,333 (2014). http://www.google.com/patents/EP2732565A1?cl=en
92. B Moyer, Low Power, Wide Area: A Survey of Longer-Range IoT Wireless Protocols (2015). http://www.eejournal.com/archives/articles/20150907-lpwa. Accessed 13 Oct 2015