

# TOWARD MASSIVE MACHINE TYPE CELLULAR COMMUNICATIONS

ZAHER DAWY, WALID SAAD, ARUNABHA GHOSH,  
JEFFREY G. ANDREWS, AND ELIAS YAACOUB

## ABSTRACT

Cellular networks have been engineered and optimized to carrying ever-increasing amounts of mobile data, but over the last few years, a new class of applications based on machine-centric communications has begun to emerge. Automated devices such as sensors, tracking devices, and meters, often referred to as machine-to-machine (M2M) or machine-type communications (MTC), introduce an attractive revenue stream for mobile network operators, if a massive number of them can be efficiently supported. The novel technical challenges posed by MTC applications include increased overhead and control signaling as well as diverse application-specific constraints such as ultra-low complexity, extreme energy efficiency, critical timing, and continuous data intensive uploading. This article explains the new requirements and challenges that large-scale MTC applications introduce, and provides a survey of key techniques for overcoming them. We focus on the potential of 4.5G and 5G networks to serve both the high data rate needs of conventional human-type communication (HTC) subscribers and the forecasted billions of new MTC devices. We also opine on attractive economic models that will enable this new class of cellular subscribers to grow to its full potential.

## INTRODUCTION

The realization of smart cities in which homes, vehicles, and mundane objects are endowed with sensing and communication capabilities will accelerate toward 2020. MTC deployments and services are expected to grow exponentially and create a multi-billion dollar industry spanning a broad range of vertical sectors including transportation, utilities, health, environment, and security.

Machine-type deployments will generate many new and diverse forms of data traffic with varying requirements in terms of delay, per-link and total bit rate, reliability, energy consumption, and security/privacy. While traditional machine-type connectivity has relied on short-range wireless technologies such as Bluetooth, moving toward large-scale deployments requires broader interconnection capabilities that are best enabled by the wide-area coverage of cellular network infrastructures, especially with the current research, development, and standardization efforts toward a global 5G network. In this article, we focus on

“massive MTC,” where a very large number of devices (at least 10x greater than the current number of cellular subscribers), with varying quality-of-service (QoS) requirements must connect to the cellular network. Realizing this in practice is contingent upon transforming the design, planning, and operation of cellular networks in terms of scalability and efficiency to handle the emerging challenges of the diverse and dense machine-type deployments.

The key motivation for deploying MTC over cellular is the predicted massive number of MTC devices that will be deployed requiring significant aggregated data rates with global geographic area coverage. Cellular networks are the natural choice to handle a major part of this emerging traffic due to their already existing infrastructures, wide area coverage, and high-performance capabilities. Certainly, this will also be complemented with MTC connectivity over capillary networks and low-power wireless local area networks. Therefore, MTC over cellular is no longer a matter of choice; there are already MTC subscriptions supported by cellular operators worldwide, enhanced techniques are currently under standardization by 3GPP (e.g., see [1, 2]), industry forecasts predict steady exponential growth for MTC devices in the next five years, and MTC has been recommended as one of few central use cases for 5G development [3].

The open questions toward massive MTC deployment over cellular have to do with when it will happen, and how it should be designed and planned. To this end, there are major roadblocks that need to be addressed at both technical and economic levels. MTC services are a new paradigm for revenue generation for cellular operators, yet these services cannot be easily integrated into existing cellular subscription and activation models. One key challenge is to develop market models for understanding the interactions between mobile operators, service providers, vertical sector customers, and regulatory agencies. To this end, achieving appropriate equipment and service pricing will be key to the growth of MTC services. Also, the possibility of having mobile virtual network operators (MVNOs) that do not own the cellular spectrum, but rather lease it from spectrum owners, is another new twist to MTC over cellular.

Following are key questions that need to be addressed in order to enable large-scale machine-type device deployments over cellular networks:

Zaher Dawy is with the American University of Beirut.

Walid Saad is with Virginia Tech.

Arunabha Ghosh is with AT&T Labs.

Jeffrey Andrews is with the University of Texas at Austin.

Elias Yaacoub is with the Arab Open University.

Digital Object Identifier: 10.1109/MWC.2016.1500284WC

Requirements	HTC over cellular	MTC over cellular
Uplink	Uplink is usually more lightly loaded and power-constrained	For many MTC applications, the main bottleneck; high signaling overhead and extreme power constraints
Downlink	The main bottleneck for high data rate services, since most traffic comes from the core network	Needs to be able to deep sleep, but wake up on command for network-initiated communication
Subscriber load	Relatively few (< 100) simultaneous devices per cell	Many (>> 100) simultaneous devices per cell with traffic uploading that can be event-triggered, periodic, or continuous
Device types	Relatively homogeneous, smart phones and data consumption devices like tablets	Extremely heterogeneous device landscape that includes environmental sensors, utility meters, wearable devices, and many unforeseen applications
Delay requirements	Defined service classes by 3GPP, vary between real-time conversational and best effort data	Very diverse delay requirements, ranging from emergency/time critical to very delay tolerant applications
Energy requirements	Flexible energy requirements due to the ability to recharge daily	Many ultra-low energy applications that require extreme power consumption measures
Signaling requirements	Signaling protocol overhead is not a concern and the design provides reliable mobility and connection management mechanisms	Application-dependent signaling protocols, with extremely efficient overhead signaling and contention resolution
Architectural requirements	Well-understood hierarchical cellular architecture with standardized interfaces between access and core network elements	Wide area coverage may require integration of data aggregators with multihop relaying; relaxed requirements for handover and roaming support

**Table 1.** MTC versus HTC requirements in the context of cellular networks.

- How are MTC services different from standard HTC services? What are key use cases for MTC deployments over cellular?
- How will the operation/performance of cellular networks be impacted by the dense deployment of diverse MTC devices and services?
- Are the challenges to handle MTC over cellular real engineering challenges, or will they actually be handled quite easily by already expected 4.5G and 5G cellular evolutions?
- If not, how to enhance/evolve/customize the design, planning, and operation of next generation cellular networks to meet the requirements of MTC devices/services? What potential techniques can be leveraged to intelligently manage or offload MTC traffic?
- How should cellular operators converge on attractive economic models and pricing schemes in collaboration with the various existing business constituents?
- How to provide incentives to MTC service providers to adopt cellular networks as a competitive connectivity option in terms of cost and quality versus other options such as proprietary or WiFi-based networks?

This article focuses on these key questions, highlighting the main roadblocks and discussing possible paths forward. There is not a “one size fits all” solution, and thus the techniques presented complement each other to handle the diverse classes and use cases of massive machine-type deployments over cellular infrastructure.

## MTC DESIGN REQUIREMENTS AND THEIR IMPACT

The requirements to support MTC services over cellular are inherently different and more diverse than traditional HTC services. MTC services add

several challenging new constraints to the range of service requirements in cellular networks, as summarized in Table 1.

Cellular operators will face performance quality challenges if they open their networks up to MTC service providers without taking necessary measures, at both the planning and operational levels, which comes with performance and cost implications. To demonstrate the potential impact of MTC devices on current cellular networks, we present sample results from an experimental case study conducted by the authors under realistic operational assumptions.

We used drive-testing equipment to quantify experimentally the signaling overhead of two standard MTC services over the air interface of a live 3G cellular network: smart metering, which reflects static MTC traffic, and low-speed vehicular sensing, which reflects mobile MTC traffic [4]. We utilized short message service (SMS) connections for messages smaller than 248 bytes (denoted as “SMS-type”), and packet switched file uploads for larger message sizes (denoted as “email-type”). For each SMS-type and email-type connection established, we capture all the exchanged control signaling messages over the air interface in both uplink and downlink (connection management, measurement reports, security commands, mobility management, etc.), determine the size of each message, and then compute the signaling load. The signaling overhead is then calculated as the ratio of the signaling load to the sum of the signaling and traffic loads.

The measured results are summarized in Fig. 1 [4]. It is interesting to note that the signaling overhead is close to 100 percent for a range of application payload data sizes; this demonstrates the significant signaling overhead from MTC devices and motivates the need for efficient signaling reduction techniques. In general, the size of the

Cellular operators will face performance quality challenges if they open their networks up to MTC service providers without taking necessary measures, at both the planning and operational levels, which comes with performance and cost implications.

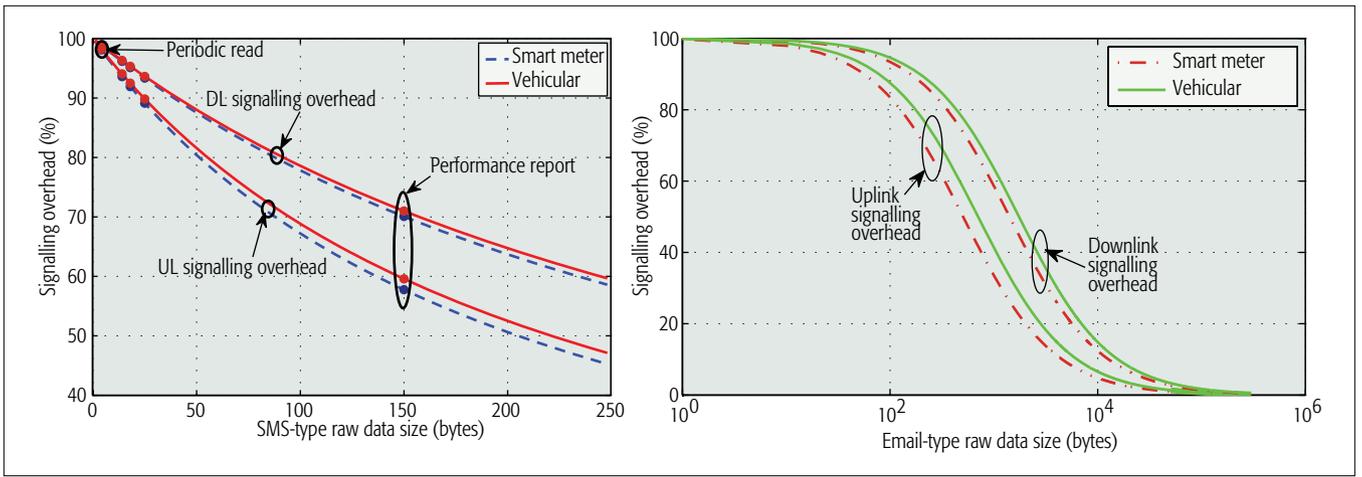


FIGURE 1. Left: signaling overhead for SMS-type MTC connections carrying short data for smart metering and vehicular sensing. Right: signaling overhead for email-type MTC connections with larger data sizes.

downlink control messages is considerably larger than the uplink, leading to higher signaling overhead. In addition, vehicular applications incur higher signaling overhead due to more frequent cell reselection and handover procedures. Finally, the signaling overhead corresponding to email-type connections notably exceeds the SMS-type, thus rendering it inefficient for small MTC messages. For example, for a 100 byte message, the uplink overhead is around 70 percent and 85 percent for SMS-type and email-type connections, respectively. Therefore, the application's payload raw data size and upload periodicity are prime factors that determine the best MTC connection type to minimize the signaling overhead.

## KEY TECHNICAL CHALLENGES AND DESIGN STRATEGIES

Due to the diversity of MTC applications, it is not feasible to devise a single technical solution for all possible applications. As noted at the outset, we focus on massive MTC deployments. Therefore, MTC applications such as vehicle-to-vehicle communications, local indoor automation, or low density fixed sensor deployments are not a main focus since they can be efficiently handled via alternative technologies such as IEEE802.11p (for V2V), IEEE802.15.4, WiFi, and/or wired broadband. We now consider some of the envisioned MTC applications.

### OVERVIEW OF USE CASES AND REQUIREMENTS

Table 2 captures the main properties and requirements for various key use cases for M2M over cellular, while also highlighting effective strategies for optimized operation.

Navigating through the various use cases and their properties and requirements, we can extract the following set of key challenges and opportunities:

- Deployment-related challenges include massive density, arbitrary locations, shadowed indoor, and high mobility; opportunities include use cases with static fixed locations.
- Traffic-related challenges include continuous uploading, critical real-time uplink and/or downlink connections, and very large message size; opportunities include use cases with (very) delay tolerant traffic.

- Energy-related challenges include ultra-low energy consumption and high energy efficiency; opportunities include possibly automated battery charging.
- Protocol-related challenges include low overhead and scalable signaling with deep sleep modes, yet timely paging or network-initiated device triggering for time-critical downlink connections.
- Complexity-related challenges include transmitter/receiver design requirements, computational/storage capabilities, and device longevity through backward compatibility, to maintain functionality over a long period of time even as cellular technologies evolve.

In the remainder of this section, we discuss effective strategies that can shape the design of next-generation cellular technologies and standards to support large-scale M2M deployments. These solutions complement ongoing activities in 3GPP standardization to develop new air interfaces to better support MTC devices, in addition to ongoing activities by the IEEE 802.11ah Task Group to enhance wireless local area networks to support dense machine-type deployments [5].

## EFFICIENT OVERHEAD SIGNALING PROTOCOLS AND PROCEDURES

As now established, a main technical challenge is the amount of overhead signaling generated by dense MTC devices. 3GPP has identified the causes for uplink signaling congestion due to a high number of devices trying almost simultaneously to attach to the network, to activate, modify, or deactivate a connection, or to perform cell reselection and handover procedures. There are several 3GPP proposals that present interesting ideas to reduce the signaling congestion problem [6], as well as basic research such as [7]. These include dedicated random access channel (RACH) formation, time-controlled policies with slotted access, prioritized connection rejection schemes, and network triggered polling. For example, in a smart metering MTC scenario, although the device is in a fixed location, it will report back measurement control messages to the network. In this case, it is possible to use spare bits in the connection request message to signal to the network that the

Use cases	Deployment properties	Main requirements	Solution strategies
Smart utility metering	Fixed and known locations; medium to high density	Low rate; periodic traffic; small message sizes; uplink data	<ul style="list-style-type: none"> <li>- Data aggregation</li> <li>- Long sleep cycles</li> <li>- Large link budget to cover deep indoor</li> </ul>
Industrial automation	Fixed and known locations; low to medium density; hot spots	Low rate; periodic or continuous traffic; small message sizes; uplink data and downlink control with critical timing	<ul style="list-style-type: none"> <li>- Data aggregation</li> <li>- Multi-homed connections for reliability</li> <li>- High priority bearer service for reliability</li> </ul>
Video surveillance	Fixed and known locations; low to medium density	Medium to very high rate; continuous traffic; large message sizes; uplink data	<ul style="list-style-type: none"> <li>- Data aggregation</li> <li>- Event-based uploading</li> <li>- Reservation based scheduling</li> </ul>
Environmental sensing and actuation (forests, oceans, transportation, agriculture)	Fixed and arbitrary locations; high to massive density	Low rate; periodic or event triggered traffic; small message sizes; uplink data; downlink control; ultra-low energy consumption; low complexity	<ul style="list-style-type: none"> <li>- Energy harvesting</li> <li>- Deep sleep modes</li> <li>- Low overhead uplink signaling protocols</li> <li>- Timely paging for machine terminated traffic</li> </ul>
Wearable sensing (health, activity, emotion, location)	Mobile and arbitrary locations; medium density	Low rate; periodic or event triggered traffic; small to large message sizes; delay tolerant to time critical uplink data; none to time critical downlink control; low energy consumption; low complexity	<ul style="list-style-type: none"> <li>- Event-based uploading</li> <li>- In-device caching with delay tolerant uploading</li> <li>- WiFi offloading in indoor scenario</li> <li>- Multi-mode operation to support emergency situations</li> </ul>
Vehicular sensing (vehicle indicators, safety, driving behavior, congestion levels, alarms, surrounding)	Mobile and arbitrary locations; high density	Low to medium rate; periodic or event triggered traffic; small to very large message sizes; uplink data; downlink control; high mobility	<ul style="list-style-type: none"> <li>- Event-based uploading</li> <li>- In-vehicle caching with opportunistic uploading and collection point offloading</li> <li>- Delay-based service priority for scheduling</li> <li>- Multi-mode handover and cell reselection protocols</li> </ul>

**Table 2 .** A summary of key use cases for MTC over cellular.

device is stationary, and thus disable all mobility-related signaling.

Another, and closely related, requirement for many MTC devices is extreme energy conservation. This includes many types of meters, sensors, and trackers, where recharging batteries may be difficult or impossible. This in turn requires most of the time to be spent in deep sleep modes, with either periodic or network triggered wake up. The advantage of periodic wake up is that it is simple; however, the period needs to be increased notably. For example, deep sleep times can now reach up to 10 min. in 3GPP Rel. 13, but this is still a few orders of magnitude too often for many MTC applications. Such infrequent wake ups conserve energy, but prohibit urgent or unpredictable control messages from the network. Also, such messages are essential for certain industrial automation, environmental monitoring, or mobile health MTC application scenarios. For example, if a gas leak is detected, all sensors need to be polled to determine its location and possibly take immediate action. Thus, a more attractive option is to devise network-triggered mechanisms that can wake up sleeping devices quickly when needed, for example through passive reception of such triggers (similar to RFID) which is closely related to the emerging area of simultaneous wireless power and information transfer. Such wireless power transfers are implausible for high rate communication, but could be used to provide MTC devices with energy to receive simple and infrequent wake-up signals while in deep sleep.

The notion of sleep mode should also be expanded to differentiate between the radio communications sleep cycle and the sensing/computing sleep cycle. That is, an MTC device can switch its communications interface off but continue sensing, or an MTC device can switch both sensing and communications interfaces off depending on the use case requirements. Revisiting the

gas leak example, the monitoring MTC devices can sense frequently with very low energy consumption and wake up only when a measurement exceeds a predefined threshold.

In addition to signaling overhead reduction, there are opportunities to exploit multiple radio access technologies (RAT) with technology-specific customizations for better support and improved performance efficiency. Ideas in this direction include recent 3GPP activities on a new cellular ultra-low complexity and throughput air interface [2], low cost LTE UE (user equipment) with reduced receive bandwidth to 1.4 MHz (Cat-0 UE category) [1], control and data plane separation over multiple cells that belong to the same or different RATs, and MTC device allocation to different RATs depending on service requirements and RAT capabilities.

### DATA AGGREGATION AND MULTIHOP CONNECTIVITY FOR MTC DEVICES

Hierarchical data aggregation is a key solution strategy to collect, process, and communicate data in MTC use cases with static devices, especially if the locations of the devices are known, such as smart utility meters or video surveillance cameras. This requires the planning and deployment of aggregators that act as masters to collect information from pre-defined clusters of MTC devices over capillary short-range. The communication from the MTC devices to the aggregators can take place over multihop device-to-device (D2D) connections. D2D is expected to be an integral part of 4.5G and 5G technologies, with direct benefits to MTC connectivity and use cases, especially to overcome link budget problems for ultra-low power devices (e.g., see [3]). Relevant research challenges include MTC device clustering, transmission mode selection, in addition to interference and power management. Therefore,

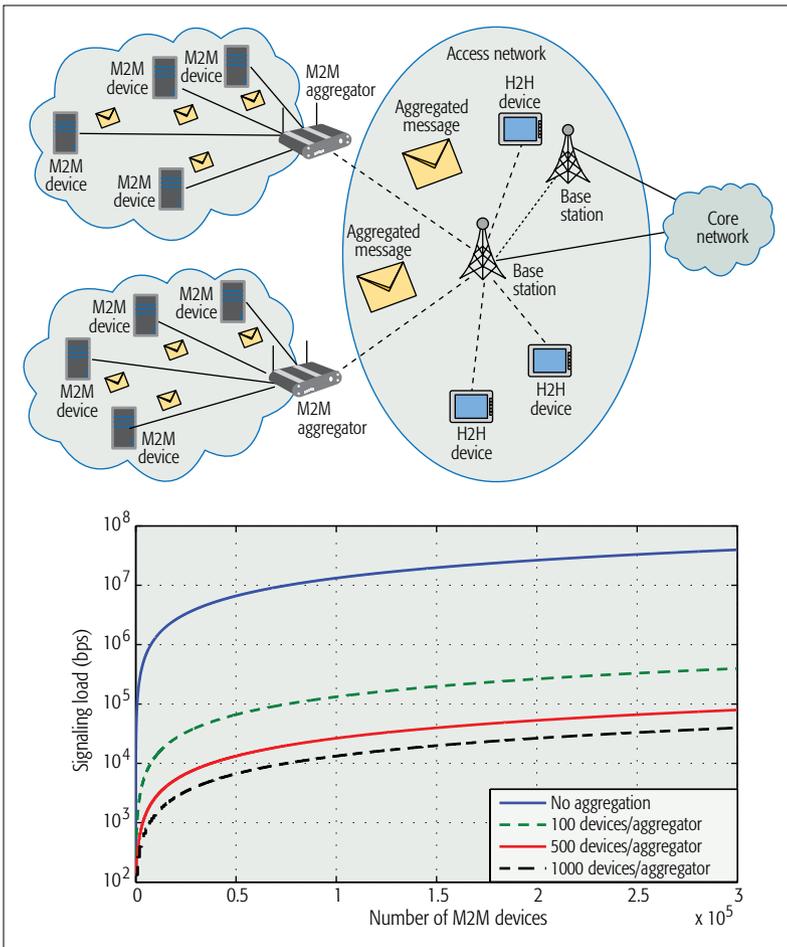


FIGURE 2. Top: M2M aggregation scenario. Bottom: uplink signaling load as a function of aggregation level and number of M2M devices for a given traffic load.

aggregators must be equipped with short range wireless interfaces to connect to MTC devices (such as LTE-Direct, WiFi-Direct, Bluetooth, or other proprietary solutions) in addition to cellular interfaces to relay traffic from MTC devices to the network. There is also potential for utilizing mmWave connectivity between MTC devices and aggregators for specific applications that require ultrahigh bit rates over very short distances. However, this requires addressing major research challenges related to the design of circuit components and antennas that meet the energy and complexity requirements of MTC devices.

For MTC use cases with static devices but arbitrary locations (e.g., deployment of sensors in a field), there is a need to devise self-organizing clustering mechanisms with either structured or arbitrary distribution of aggregator nodes. This is a complex problem, especially with the support of multihop connectivity among the devices and the aggregators. An attractive solution approach is machine learning-based spectral clustering where similarity matrices are determined based on MTC device information and utilized to perform clustering. Another more dynamic approach is coalitional game theory that allows not only clustering based on similarity, but also implementation of cooperative communication techniques that exploit the existing similarities.

Data aggregation has many potential benefits for many MTC over cellular use cases. It reduces

uplink signaling overhead by reducing the number of radio connections into any given receiver, reduces transmit power by reducing the link distance, and extends coverage range.

To demonstrate the effectiveness of aggregation in reducing uplink signaling load, we present sample experimental results that quantify the performance tradeoffs of aggregation over a live 3G cellular network using a real-time drive testing setup. Data aggregation is modeled by the transmission of a large combined email-type message instead of multiple small messages from individual MTC devices. Figure 2 quantifies the signaling load as a function of the number of MTC devices connected to the aggregator [8], which we can see decreases as the aggregators become more numerous. Recent theoretical results derive bounds on the optimal number of hops, that is, the number of levels of aggregators, finding that it is linearly proportional to the path loss exponent and outage target, and inversely proportional to the required SINR [9].

The proliferation of MTC aggregators in high density geographic areas will lead to ultra-dense small cell-like deployments with capacity, coverage, backhaul, and interference management requirements. In such scenarios, multi-RAT solutions with self-organizing features are expected to be highly effective, for example, see [10] for a novel 3GPP-compliant architecture to reduce radio and core network congestion via home femtocells.

Finally, the role of an aggregator need not be limited to communications relaying but should be extended to support radio network intelligence (such as resource allocation, interference management, spectrum sharing, etc.) and local computing for higher efficiency. For example, aggregators can apply advanced data reduction techniques to compress the combined data before uploading, or apply feature extraction and event detection techniques to decide whether to upload the data or simply drop. This cross-layer fusion between communications and computing intelligence is pivotal for achieving an advanced degree of efficiency, robustness, and scalability.

## IN-DEVICE PROCESSING WITH INTELLIGENT RESOURCE MANAGEMENT

Distinguishing characteristics of some key MTC use cases are traffic predictability, event-based monitoring, and high delay tolerance. These facilitate the implementation of various dynamic in-device processing techniques in order to enhance the overall communications efficiency and reduce the load on the cellular network infrastructure.

**For Smart Utility Meters or Aggregators:** The traffic is normally highly periodic over time and nearly deterministic in terms of data size. For such traffic, reservation based scheduling with lower signaling overhead (e.g., similar to semi-persistent scheduling for voice over LTE [11]) is attractive as it allows the cellular operator to dimension which radio resources are used and when from a large number of MTC devices per cell. For example, the MTC devices can be allocated the needed resources to upload their traffic in off-peak hours in coordination with the MTC application service provider.

**For MTC Use Cases Such as Vehicular Sensing:** Sensors can collect data from the vehicle and its sur-

roundings on a continuous basis with in-device caching and *long-term opportunistic scheduling* [12]. Depending on the delay tolerance of the collected data, the transmitting machine-type device can be allocated cellular resources depending on the current network load in the various cells that the vehicle is passing through. To control that the traffic caching delay does not exceed the service quality bound, the connection's service priority can be dynamically increased over time based on an adaptive algorithm. If the service delay bound is relatively flexible, then the cached data over a certain time window can even be offloaded over pre-deployed *data collection points* (e.g., WiFi access points) installed in specific locations that are either public (e.g., street crossings) or private (e.g., parking lots).

**For Use Cases Such as Environmental Sensing via Static MTC Devices or Wearable Sensing via Mobile MTC Devices:** A main application is monitoring with *event-triggered communications* (e.g., a health emergency or alarming temperature level). In this case, the device will need to mainly upload data, which could be highly urgent, as soon as an event is detected. Thus, the cellular operator needs to provide a new high priority uplink service for critical MTC uploading; this can directly build on the existing 3GPP emergency call procedures. The need for critical timing in MTC communications extends to downlink cellular, for example, for sensor control in emergency situations. This can be achieved via network-triggered device wake-up and control mechanisms, as discussed previously.

**In Scenarios with Limited Coordination Between the Cellular Operators, MTC Service Providers, or MTC Customers:** There is a need to develop *adaptive machine learning* algorithms that can accurately predict the nature of the traffic and transmission patterns. This information will then be used to develop *self-organizing resource management mechanisms* that can exploit the learned data to optimize performance without disrupting HTC connection quality. A promising approach in this direction is to exploit low-complexity and distributed mathematical frameworks, such as matching theory [13], which enable optimized allocation of resources to devices, while accounting for locally learned data.

## MEETING ENERGY AND LONGEVITY REQUIREMENTS

In this section, we focus on two of the most demanding operational constraints for MTC that cannot be adequately addressed via incremental enhancements: ultra-low energy consumption and device longevity. These challenges are naturally coupled with complexity constraints at the device level in terms of communications, computing, and storage capabilities.

Due to battery capacity limitations, *energy harvesting* techniques can provide MTC devices with the needed energy to maintain connectivity and functionality over long periods of time without the need for recharging. The choice of energy harvesting technique (e.g., solar, vibration, heat) will depend on the MTC device's hardware design, function, operational, and deployment conditions. For example, health-related wearables could harness the body's heat or vibrations, whereas an outdoor sensor could use a very small solar panel. Although energy harvesting is currently a hot topic for wireless applications in general, it seems

that massive MTC is a particularly attractive application of energy harvesting since the amount of total required energy is fairly small, and the need to avoid recharging is so great.

*Longevity* is an important and usually overlooked operational challenge that is quite specific to MTC over cellular. Unlike smartphones, which are replaced every two to three years on average and are rarely still in use after five years, utility meters and other MTC sensors routinely require a guaranteed lifetime of 15 years or more before replacement. The business model becomes unattractive if an MTC service provider has to replace devices as new cellular technologies are standardized for the faster churning commercial HTC applications. We suggest two main approaches to guarantee longevity. The first is to have cellular operators provide guarantees to maintain backward compatibility for a given period of time, for example, 20 years. Although this solution is transparent to the MTC devices, it greatly restrains future innovation and flexibility, and could have damaging long-term effects. However, these effects could possibly be reduced by introducing new MTC technologies on different bands, while maintaining backward compatibility on legacy bands.

The other approach is to develop MTC devices with software defined radio (SDR) capabilities to facilitate a certain degree of re-configurability via software updates. However, this is still complex since the devices will be designed without knowing how technology standard releases will evolve 10 to 20 years later; in addition, the devices might have complexity constraints to handle re-configurable hardware. Therefore, the objective should be to converge at some middle ground that allows for a certain level of reconfiguration combined with a certain level of backward compatibility from the operators. For example, physical layer functions can be fixed, whereas MAC and network layer protocols and procedures can be software configurable.

## ECONOMICS OF MTC DEPLOYMENTS

The deployment of MTC services over cellular networks introduces new economic challenges that can significantly differ from existing models. Indeed, the diversity of the MTC service offerings will require rethinking how existing cellular network economic models operate, as discussed in this section.

### OVERVIEW OF

### ECONOMIC CONSIDERATIONS AND CHALLENGES

A summary of the key economic challenges for MTC deployment is provided in Table 3. Hereinafter, an MTC customer is defined as a user who subscribes to MTC services provided directly by an operator, while an MTC provider is a dedicated third-party that offers MTC services that are eventually delivered over the operator's network.

Cellular operators and mobile virtual network operators (MVNOs) will be required to deliver MTC services to their own customers as well as customers of third-party MTC providers. MVNOs are virtual operators that do not own spectrum or infrastructure, but rather lease it from the owner. Operators and MVNOs must consider how to

The deployment of MTC services over cellular networks introduces new economic challenges that can significantly differ from existing models. Indeed, the diversity of the MTC service offerings will require rethinking how existing cellular network economic models operate.

Economic aspects	HTC over cellular	MTC over cellular
Market players	Operators; MVNOs; regulatory bodies	Operators; MTC providers/vertical sector customers; MVNOs; regulatory bodies
Market model	Hierarchical model with pre-defined leader-follower structure	Hierarchical model with varying leader-follower structure
Role of operators	Deliver connectivity and wireless services	Deliver own and third-party MTC services on top of connectivity and existing services
Pricing plans	Well-defined pricing plans	MTC service dependent; smart data pricing; net neutrality questions
Data caps	"Standardized" data caps per individual subscriber	Caps not viable for all MTC services; need for aggregate data caps per service
Profit or cost sharing	Not common	Cooperation between operators and MTC providers

**Table 3.** MTC versus HTC economic aspects in the context of cellular networks.

build economically profitable partnerships with MTC providers, such as a transportation service provider or a smart grid utility, while building their own market base. The relationship between operators and MVNOs must also be revisited in light of MTC deployments. For example, operators might also consider MVNOs as a cost-effective way to offload MTC traffic, depending on how machine-type services will grow and how such growth will impact the technical operation of the network. A key economic question here would be how to manage the lease of infrastructure and spectrum to MVNOs for MTC purposes.

In addition, the role of regulatory bodies in ensuring a gradual and effective deployment of MTC services must be analyzed. Examples of such involvement can include questions on whether such bodies might want to provide certain incentives, monetary or in terms of spectrum, for operators to expedite the deployment of MTC services. How such incentives must be designed will constitute an important challenge for MTC deployment.

Last but not least, there will be a need for cooperation and possibly revenue sharing mechanisms between operators and third-party MTC providers. Cooperation between operators themselves may also be needed to regulate and better manage the expected massive deployments. Such cooperation and profit sharing mechanisms are minimal and even non-existent in classical cellular markets. The role of MTC customers in driving such profit sharing mechanisms can also be critical and must be analyzed.

### BASIC STUDY OF AN MTC MARKET STRUCTURE

The MTC market structure will closely follow the hierarchical model of existing HTC markets. In such markets, a leader-follower structure exists in which the economic decisions made at an upper hierarchy (e.g., operators) can propagate to the lower hierarchies (e.g., customers) and vice versa.

In MTC, in contrast to the static HTC market structure, the exact leader/follower role can be dynamic, depending on the scenarios. For example, when leasing infrastructure, operators are naturally leading the decisions of MVNOs, MTC

providers, and MTC customers. However, when new services are launched by an MTC provider, such decisions will lead and impose changes to the operators' own network. This dynamic hierarchical model motivates a game-theoretic approach to analyze the MTC market. A natural tool is the framework of a Stackelberg game, which must handle two unique challenges:

- A multi-level, multi-leader, multi-follower structure as opposed to the classical two-level, single-leader structure.
- Dependence of the objectives of the players on externalities such as the technical decisions made.

Indeed, the interplay between the network economics and the technical challenges must be properly factored in for effective MTC deployment.

To provide a basic insight on such a framework, we investigate, based on our work in [14], a multi-leader, multi-follower structure for studying the economics of deploying new services in a system having a large population of customers with heterogeneous requirements. Here, competing service providers (SPs) aim to deploy similar MTC services. The goal is to analyze optimal prices and customer selections. Using this framework, as shown in Fig. 3, we can analyze a duopoly market in which SP 1 offers two services, service 1 and service 2, and another service provider, SP 2, offers only service 1. SP 1 is evaluating how much to economically invest in service 2. Each service is associated with a unique spectrum, hence the objective functions depend exclusively on each service's own load. We use  $G_0$  to represent the maximum QoS offered by SP 1's service 2;  $G_0$  depends on the economical effort needed to deploy service 2 by SP 1 (to deploy a higher QoS service, one needs a higher  $G_0$ ). Then, we let  $G_0$  vary from 0 (i.e., no new service) to 5 (two services by SP) thus evaluating the SPs' revenues. From Fig. 3, we can see that the income of SP 1 is boosted by the new service only when  $G_0 > 3$ . When  $G_0 < 3$ , SP 2 can approximately limit the loss of users by decreasing the price asked for the service. In this case, the effect of the implementation of the new service by SP 1 is to slightly decrease the profits of SP 2 (therefore improving the utilities of the users), instead of increasing the revenues of SP 1. Clearly, within the MTC market structure, pricing will play an important role, as discussed next.

### PRICING PLANS AND DATA CAPS FOR MTC SERVICES

In a classical cellular network, well-defined pricing plans and bundles already exist. However, incremental changes to such plans will not be sufficient to handle the MTC market. For example, if an operator delivers services for both third-party MTC providers and its own MTC customers, it must consider how to price these two different offerings so as to maximize revenues, ensure service quality, and avoid competing with itself by inadvertently pushing customers to prefer the third-party providers. This also brings in interesting questions of net neutrality since operators might not be able to provide preferential services for their own customers at the behest of customers of the third-party providers.

The pricing plan will also depend on the type of the MTC service itself. For instance, critical MTC services, which require higher QoS and maintenance, may need to pay a premium as opposed to massive, non-critical services (such as smart utility meters). In addition, owing to the heterogeneity and diversity of machine-type devices and services, the usage of smart data pricing will become more prominent [15]. Such smart data pricing policies can include time-dependent and location-dependent pricing that can be tailored to specific business scenarios. For example, non-critical MTC services can benefit from time-dependent pricing plans that incentivize them to transmit their data during non-peak hours, thus also contributing to reducing the congestion of the system, as discussed earlier.

One additional challenge is the notion of data caps and volume bundles. For regular subscribers, per-user data caps are common and widely used. In contrast, for MTC services, defining per-device data caps may not be viable, particularly for critical services or for services that rely on low-data rate, small devices, such as sensors. In such cases, one may consider introducing aggregate data cap models. Such models can be set on a per MTC provider case, thus providing a fixed price based on the total amount of data that a certain provider can use within a given time period or even per location. Determining these caps will also be driven by the introduction of accurate models for MTC traffic.

Beyond pricing services, other economic challenges include investigating how to price and offer various machine-type devices, how to provision and allocate blocks of SIM cards to MTC customers with flexible activation/deactivation options, and how to perform automated on-demand device configuration.

Clearly, the massive deployment of MTC over cellular networks poses important and non-trivial economic considerations that must be jointly considered along with the technical challenges to pave the way toward the successful realization of a key requirement as part of the 5G vision.

## CONCLUSION

Cellular technologies have inherent advantages to play a central role in serving the exponentially growing number of machine-type devices, due to their wide area coverage and high capacity characteristics. Yet, this is contingent on addressing major challenges at both the technical and economical levels. In this article, we present a set of effective design strategies tailored to various MTC use cases that can enhance the design of next generation cellular technologies. In addition, we cover key economic considerations for facilitating the deployment of MTC over cellular from a business perspective, and discuss a basic study of an MTC market structure.

## ACKNOWLEDGMENTS

This work was made possible by NPRP grant 4-353-2-130 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors. Z. Dawy would like to thank Mona Jaber and Nour Kouzayha for their help in generating Fig. 1 and Fig. 2.

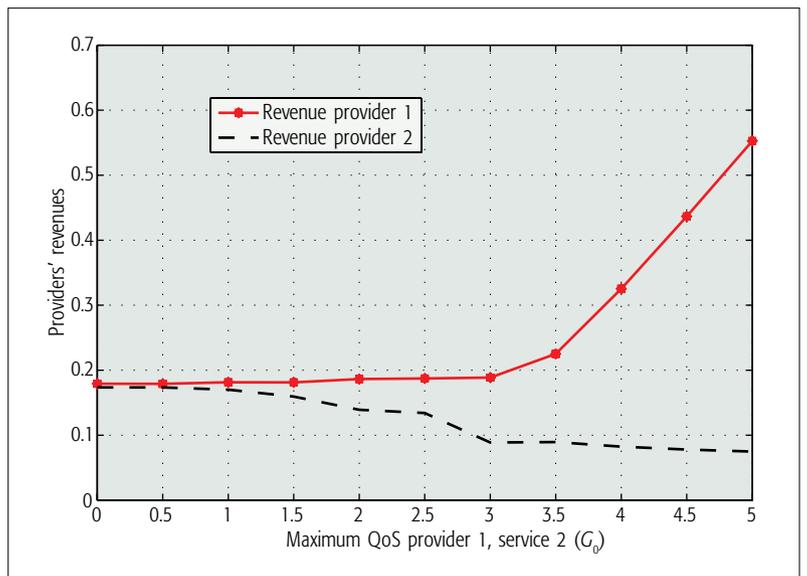


FIGURE 3. On M2M over cellular economics: Sample results for a case study with two service providers.

## REFERENCES

- [1] 3rd Generation Partnership Project (3GPP), Technical Specification Group Radio Access Network, Study on Provision of Low-Cost Machine-Type Communications (MTC) User Equipments (UEs) Based on LTE, Release 12, June 2013.
- [2] 3rd Generation Partnership Project (3GPP), Technical Specification Group GSM/EDGE Radio Access Network, Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things, Release 13, June 2015.
- [3] Mobile and Wireless Communications Enablers for the Twenty-Two Information Society (METIS), "Final Project Report," document number ICT-317669-METIS/D8.4, April 2015.
- [4] M. Jaber *et al.*, "On Cellular Network Planning and Operation with M2M Signalling and Security Considerations," *Proc. IEEE ICC 2014 Workshops*, June 2014.
- [5] T. Adame *et al.*, "IEEE 802.11ah: The WiFi Approach for M2M Communications," *IEEE Wireless Commun.*, vol. 21, no. 6, Dec. 2014, pp. 144–52.
- [6] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives," *IEEE Commun. Surveys Tutorials*, vol. 16, no. 1, Jan. 2014, pp. 4–16.
- [7] H. S. Dhillon *et al.*, "Fundamentals of Throughput Maximization with Random Arrivals for M2M Communications," *IEEE Trans. Commun.*, vol. 62, no. 11, Nov. 2014, pp. 4094–4109.
- [8] N. Kouzayha, M. Jaber, and Z. Dawy, "M2M Data Aggregation over Cellular Networks: Signaling-Delay Tradeoffs," *Proc. IEEE Globecom 2014 Workshops*, Dec. 2014.
- [9] D. Malak, H. S. Dhillon, and J. G. Andrews, "Optimizing Data Aggregation for Uplink Machine-to-Machine Communication Networks," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1274–90, March 2016.
- [10] M. Condoluci *et al.*, "Toward 5G DenseNets: Architectural Advances for Effective Machine-Type Communications over Femtocells," *IEEE Commun. Mag.*, vol. 53, no. 1, Jan. 2015, pp. 134–41.
- [11] J. B. Seo and V. C. M. Leung, "Performance Modeling and Stability of Semi-Persistent Scheduling with Initial Random Access in LTE," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, Dec. 2012, pp. 4446–56.
- [12] F. Mehmeti and T. Spyropoulos, "Is it Worth to be Patient? Analysis and Optimization of Delayed Mobile Data Offloading," *Proc. IEEE Infocom 2014*, April 2014.
- [13] Y. Gu *et al.*, "Matching Theory for Future Wireless Networks: Fundamentals and Applications," *IEEE Commun. Mag.*, vol. 53, no. 5, May 2015, pp. 52–59.
- [14] L. Rose *et al.*, "Pricing in Wireless Networks with Heterogeneous Services and Customers: Hierarchical Games and Dynamics," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, May 2014, pp. 4985–5001.
- [15] S. Sen *et al.*, "Incentivizing Time-Shifting of Data: A Survey of Time-Dependent Pricing for Internet Access," *IEEE Commun. Mag.*, vol. 50, no. 11, Nov. 2012, pp. 91–99.

---

## BIOGRAPHIES

ZAHER DAWY [M'97, SM'09] received the B.E. degree in computer and communications engineering from the American University of Beirut (AUB), Beirut, Lebanon, in 1998, and the M.S. and Dr.-Ing. degrees in communications engineering from Munich University of Technology (TUM), Munich, Germany, in 2000 and 2004, respectively. Since 2004, he has been with the Department of Electrical and Computer Engineering, AUB, where he is currently a professor. His research and teaching interests include wireless communications, cellular technologies, context-aware mobile sensing, computational genomics, and biomedical engineering. He received the IEEE Communications Society 2011 Outstanding Young Researcher Award in the Europe, Middle East, and Africa Region.

WALID SAAD [S'07, M'10, SM'15] received his Ph.D. degree from the University of Oslo in 2010. Currently, he is an assistant professor and the Steven O. Lane Junior Faculty Fellow at the Department of Electrical and Computer Engineering at Virginia Tech. His research interests include wireless networks, game theory, cybersecurity, and cyber-physical systems. He received the NSF CAREER award in 2013 and the Young Investigator Award from the Office of Naval Research (ONR) in 2015 for his work on M2M (ONR-N00014-15-1-2709). He has received five conference best paper awards, and he received the 2015 IEEE Fred W. Ellersick Prize.

ARUNABHA GHOSH received his B.S. from the Indian Institute of Technology Kanpur in 1992, and his Ph.D. from the University of Illinois at Urbana-Champaign in 1998. He is currently the director of the Advanced Wireless Technology Group at AT&T Labs, where he leads the development and implementation of advanced wireless technology for 4G and 5G radio access networks. Currently he is involved in various 5G related R&D activities, including mmWave channel modeling and system design, design of massive MIMO systems and network architecture for 5G network, spectrum sharing, software defined radio access networks, etc. He is also responsible for various

5G trial and testbed activities carried out in the 5G technology lab in Austin. His main area of research relates to signal processing and communication theory, but he has also worked extensively on other topics such as multi-user MIMO, massive MIMO, cloud RAN, advanced system architecture, etc. He has published extensively in these areas, and is a key contributor to AT&T's patent portfolio on this subject. Dr. Ghosh also participates in and represents AT&T in 3GPP. Dr. Ghosh has also authored several best selling books on LTE and WiMAX. He received his Ph.D. from the University of Illinois at Urbana-Champaign and his masters and bachelors from the Indian Institute of Technology.

JEFFREY G. ANDREWS [S'98, M'02, SM'06, F'13] received the B.S. in engineering from Harvey Mudd College in 1995, and the M.S. and Ph.D. in electrical engineering from Stanford University. He is the Cullen Trust Endowed Professor (#1) of ECE at the University of Texas at Austin, and Editor-in-Chief of the *IEEE Transactions on Wireless Communications*. He developed code division multiple access systems at Qualcomm from 1995-97, and he has consulted for entities including Apple, Samsung, Verizon, AT&T, the WiMAX Forum, Intel, Microsoft, Clearwire, Sprint, and NASA. He is an elected member of the Board of Governors of the IEEE Information Theory Society, and he is a co-author of 14 best paper award recipients, including the 2014 IEEE Stephen O. Rice Prize and the 2014 IEEE Leonard G. Abraham Award. He recently received the 2015 ASEE/Hewlett Packard Terman Award.

ELIAS YAACOUB [S'07, M'10, SM'14] received the B.E. degree in electrical engineering from the Lebanese University in 2002, and the master and Ph.D. degrees from the American University of Beirut in 2005 and 2010, respectively. From 2010 to 2014 he worked as a research scientist at the Qatar Mobility Innovations Center. Afterward, he worked as a consultant at Strategic Decisions Group before joining the Arab Open University as an associate professor. His research interests include wireless communications, QoS/QoE, sensor networks, smart grids, and antenna theory.