8 Future challenges in efficiently supporting M2M in LTE standards

Andres Laya (UPC-KTH), laya@kth.se
Kun Wang (CTTC), kun.wang@cttc.es
Luis Alonso (UPC), luisg@tsc.upc.edu
Jesus Alonso-Zarate (CTTC), jesus.alonso@cttc.es
Jan Markendahl (KTH), janmar@kth.se

Universitat Politècnica de Catalunya (UPC), Spain
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain
KTH Royal Institute of Technology (KTH), Sweden

Abstract

This chapter will describes some of the key challenges that are faced to make LTE suitable for Machine-to-machine (M2M) communications. These are related to handling large number of devices, attaining very low energy consumption, ensuring low-complexity protocols to bring down the cost of devices, and extending coverage to facilitate one-hop transmissions. Beyond being technical feasible, any network optimization must be approved by network operators; therefore, an additional techno-economic perspective will be presented in this chapter. Business related requirements and implications that emerge from the traffic mix (humans and machines) will also be covered, including scalability and deployment cost issues, as well as migration options from current solutions.

Keywords

Techno-economic challenges; device cost; business roles; technology migration
The main objective of this chapter is to illustrate the challenges related to the support of M2M communication in LTE, a radio access technology that was designed to accomplish mobile broadband demand and at first glance seems unsuitable for M2M communications, with elevated hardware price due to its complexity and far more bandwidth than the required in most M2M applications. Moreover, the coverage of 4G is not widely available yet, in contrast to 2G technologies that are considered ubiquitous and universal [1].

On the other side, the technology attributes such as higher spectral efficiency can reduce the operator’s cost to deliver service compared to technologies such as GPRS and UMTS; which can compensate for the low revenues per unit from M2M devices.

Moreover, in some industrial segments, devices are expected to have long device lifetimes, with many devices estimated to remain operational for more than ten years [2]. In such cases, LTE can be regarded as the safe choice for the future. According to the OECD [1], 2G networks are planned to be decommissioned within the next five to fifteen years and therefore, acquiring GSM modules for M2M solutions might result in a loss of connectivity services in case of spectrum refarming or if an operator decides to shut down a legacy network [3,2].

This chapter is devoted to present and describe the imminent challenges to overcome in the technical and business domain to efficiently support M2M communication in LTE network.
8.2 Main Technical Challenges and Existing Solutions

This section focuses on the technical challenges related to large number of devices, low energy consumption, low cost LTE M2M devices and enhanced coverage, which are the key challenges to be addressed for efficiently supporting M2M in the LTE system.

8.2.1 Handling a very large number of devices

A huge number of devices are predicted to connect to the communication networks in the foreseeable future, with some forecast predicting figures of 50 billion devices connected by the year 2020 [4]. This figure does not even consider the coexistence with the current (and future) human users. Such increasing M2M traffic load may result in radio network and signaling congestion, which will inevitably increase delay and packet loss, and furthermore, limit the adoption of M2M services in the market. Besides, the conventional Human-to-Human (H2H) service may be affected or even total service outage may happen. So how to support a large number of devices per cell and secure the network availability is a key challenge the LTE system needs to handle.

Access control is an effective mean to prevent the overload of networks. 3GPP release 8 specified the Access Class Barring (ACB) mechanism to perform access control [5], where all users are divided into different Access Classes (AC). In case the network suffers overload, the core network can notify the eNodeB to reject some accesses from some classes of users (e.g., those users with low priority requirements). If an UE intends to access the network, it must first check if it is barred by the network. If it is not, then is able to perform the Random Access (RA)
procedure and send service request to the eNodeB. If the eNodeB received the notification of rejecting such class UE, it will reject the service request and inform the UE to wait some time before the next access attempt. In this way, preventing traffic overload in the access channel of the radio interface.

3GPP release 11 enhanced the ACB to a mechanism referred to as Extended Access Barring (EAB) for M2M. In EAB, when the eNodeB receives the overload notification, it will broadcast the condition to the M2M devices. In such cases, M2M devices will not access the network until the eNodeB notify that the network is capable to handle M2M requests. Therefore, M2M devices with EAB will not send service requests and even random access requests to the network in contrast to the ACB mechanism. Figure 8.2.1 depicts the ACB and EAB in LTE release 8 and 11.

![Figure 8.2.1: Access Control in LTE](image-url)
In addition to the EAB, other methods have been identified to cope with the Random Access Channel (RACH) congestion [6]:

- Separate RACH resources for M2M and H2H devices when they coexist in the same cell, so the congestion incurred by M2M load increase will not affect the conventional H2H UEs accesses.
- Dynamically allocate additional RACH resources for the M2M devices in the case that the network can foresee the overload caused by M2M load increase in advance (e.g. in the time-controlled M2M applications).
- Assign M2M specific back off time to delay the RA attempts or re-attempts of M2M devices, so the congestion can be relieved.
- Allocate the M2M devices in different access time slots, so the M2M devices can only perform RA attempts or re-attempts in their own time slots, which will decrease the probability of congestion.
- Page the M2M devices by the network when the M2M server needs information from the M2M devices or can predict what time the M2M devices may send information according to the time controlled characteristic. The network thus can control the M2M devices’ accesses according to its load level. This scheme is so called pull based.

8.2.2 Supporting low energy consumption solutions for M2M

Low energy consumption is always a desirable requirement for battery-operated devices. It is even more crucial for M2M devices when deployed in large number, in widespread areas than difficult maintenance tasks related to battery exchange or manual recharge. Therefore, high energy efficiency is another key challenge the LTE system needs to cope with.
Instead of monitoring the control channel continuously, the LTE utilizes the Discontinuous Reception (DRX) scheme to switch off the UE’s radio circuitry for a period of time, thus saving power consumption [7]. For the bursty traffic, DRX scheme inevitably increases the latency and delay, so degrades the Quality of Service (QoS). But for some time controlled or latency tolerant M2M applications, the DRX scheme is very appropriate and can significantly help save power consumption [8]. For the small data transmission of M2M applications, LTE can use Adaptive Modulation and Coding (AMC) and Uplink Power Control (UPC) mechanisms to decrease the over high data rate performance and agree with the low data rate of such applications, thus reducing energy wastage [9].

In addition, according to the M2M characteristics, signaling can be optimized in terms of energy consumption. Some M2M applications feature low mobility, where M2M devices may not change the location frequently and thus Track Area Update (TAU) cycles could be increased and thus signaling reduced. For the time controlled M2M applications, downlink data transmissions could be planned to perform after the TAU procedures [6], so that some access signaling can be saved and longer DRX sleeping period can be designed.

Besides, other communication techniques, such as Cooperative Communication (CC) and Device to Device (D2D) communication, could be used to improve the energy efficiency of M2M devices. CC can increase the data throughput and enhance the energy usage in wireless networks [10] by means of transforming a long distance transmission to the multiple consecutive short distance transmissions and multiple transceivers. In the former case, CC decreases the exponential pass
loss and saves the transmit power. In the latter case, CC utilizes multiple inputs and outputs (the coordinated multi-point (CoMP) transmission or reception in LTE) to achieve the spatial diversity or cooperative diversity, which lets the senders transmit data with low power while achieving the same required data error probability. LTE release 12 studies the D2D scheme for the proximity and public safety services [11]. D2D communications allow devices to communicate directly with each other without having to route traffic through the eNodeB, a proximity gain thus can be achieved, where the proximity of devices helps reduce the energy consumption, and achieve both high data throughput and low communication delays.

8.2.3 Supporting very low cost M2M UEs

Most operators targeting the evolution of networks to Long Term Evolution (LTE) own legacies such as General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS) networks. In order to reduce the total operating expense of the networks, they are trying to simplify multiple Radio Access Technologies (RAT) into a sole LTE network.

On the other side, the emerging M2M is envisioned to be the new revenue generation opportunity for the conventional operators, so the 3GPP has already started the study of supporting M2M in LTE system [12]. However, in comparison with the promising LTE, legacy technologies still hold advantages to support some M2M applications. This is the case for the low-end M2M, which implies that each machine only performs low data rates service and contributes low average revenue to the operators. Such low-end M2M can be served by GPRS sufficiently due to its
ubiquitous coverage and enough service capability. Furthermore, customers are likely to choose the cheap GPRS modems rather than the relatively expensive LTE modems with over high performance, which will lead to increasing deployment of M2M devices in the GPRS networks. But this is not desirable for the LTE operators. So, how to reduce the cost of LTE modem and attract low-end machines moving to the LTE networks is a challenge to be addressed [13], [14].

The item of provision of low-cost M2M devices based on LTE [14] studies the RAN specifications from Rel-10 and assesses the feasibility of defining a new type of UE which is suitable for the low-end M2M market in terms of the price and performance. To begin with, the study item lists the requirements for the solutions devised to provide low-cost M2M devices based on LTE. Such requirements can be grouped into three aspects:

- The price of low-end M2M devices based on LTE should be comparable with that of the devices based on the GPRS.
- The performance of low-end M2M devices based on LTE should be higher than that of the devices based on the GPRS in terms of data rate, spectrum efficiency and power consumption.
- The low-end LTE devices should hold backward compatibility and have no impacts on other conventional devices and the existing network architecture.

The intuitive solution to decrease the device cost is to reduce the UE complexity and simplify its implementation with controlled performance degradation. So the study then analyzes the current cost structure of a reference LTE UE [14] and
identifies the complexity simplifying methods which may bring significant UE cost savings. Such methods include:

- **Reduction of maximum bandwidth**: A normal UE can support a maximum bandwidth of 20 MHz and reducing the maximum bandwidth can significantly save the UE cost according to the data collected from different companies [14]. Reduction of maximum bandwidth can simplify the baseband processing, which may utilize the lower complexity of Fast Fourier Transform (FFT)/ Inverse Fast Fourier Transform (IFFT) and receiver. But reducing the maximum bandwidth can only save small cost on the Up Link (UL), because the cost of UL processing block takes a small portion of the total baseband cost, and the Radio Frequency (RF) component cost is affected slightly by the bandwidth.

- **Single receive RF chain**: A normal UE has two antennas and two receive RF chains to realize the receive diversity, so we may reduce the UE cost by using only one receive RF chain, which can reduce the receive filtering cost and the baseband processing functional blocks in terms of FFT, the channel estimator and data buffer memory. On the other hand, using one receive RF chain reduces the receive sensitivity of the machine, which leads to the shrinkage of the Down Link (DL) coverage and the loss on spectral efficiency.

- **Reduction of peak rate**: To support a higher peak data rate, the UE needs to support a larger maximum Transport Block Sizes (TBS) for DL and UL, a larger maximum number of simultaneously assigned Physical Resource Blocks (PRB) or higher modulation orders. So if we reduce the peak data rate, the UE cost can be reduced by means of simplifying the UL processing, turbo coding and Hybrid Automatic Repeat Request (HARQ) buffering.
• **Reduction of transmit power:** A normal UE needs to realize a maximum transmit power of 23 dBm, which is usually achieved through an added power amplifier stage. So reducing the output power or even removing the power amplifier stage completely is a method to reduce the UE cost. But on the other side, a UE with reduced transmit power will decrease its UL coverage. And to maintain the UL coverage, the UE may need to utilize the low level modulation coding and low speed data rates to perform transmissions, which then result in the low spectral efficiency.

• **Half duplex operation:** In the Frequency-division duplexing (FDD) LTE system, a normal UE can transmit and receive data simultaneously, which is realized by a duplexer. When the UE only needs to operate in a half-duplex mode, such duplexer can be replaced by a switch, which can reduce the UE cost. And as in half duplex mode the UE does not need to provide the processing power and memory for both downlink and uplink operations, the complexity of and memory of the baseband could be reduced too. But it should be remarked that the eNodeB still work in a full duplex mode, and the scheduler should not schedule the half-duplex UE to transmit and receive in the same time. So such half duplex method improves the complexity of the scheduler in the eNodeB.

• **Reduction of supported downlink transmission modes:** For the LTE DL, a Rel-10 UE supports one layer of spatial multiplexing and up to 9 kinds of transmission modes. If we reduce the supported transmission modes to the basic two kinds, we then may remove the Demodulation Reference Signal (DMRS) based channel estimation, the Pre-coding Matrix Indicator (PMI) computation and simplify the Multiple Input and Multiple output (MIMO)
detection/equalization algorithm. But as we remove the pre-coding scheme, so the DL performance may be affected due to lack of pre-coding gain.

8.2.4 Providing enhanced coverage for M2M devices

The complexity simplifying methods can help to reduce the cost of M2M devices based on LTE, but some of them, such as Single receive RF chain and Reduction of transmit power, will inevitably decrease the coverage of DLs and ULs. Furthermore, some machines are deployed in the extreme coverage circumstances, such as the basements of buildings, where the signals will suffer extremely higher attenuation than the normal deployments through the wireless channels. So, how to improve the coverage in such deployments is another new challenge for the LTE system to efficiently support the M2M. The study item in LTE Rel-12 [14] also investigates the possible methods to conquer this challenge.

The UEs in such deployments usually have fixed locations, where they will have no time period to move to the good coverage regions. It is also assumed that such UEs have low data rates and can tolerate a large delay. According to such characteristics, some coverage improvement methods are proposed as follows:

- **HARQ retransmission:** when the receiver cannot decode the received message correctly, it will not discard the message but ask the sender to retransmit the message again through the HARQ mechanism. Each retransmission carries the same data information but maybe different redundancy bits (different redundancy versions), so the receiver can combine the different version messages from different transmissions to perform decoding incrementally, which will improve the probability of
correctly decoding messages and resist the high attenuation of wireless channels. HARQ thus can be used to improve the coverage.

- **TTI bundling**: Transmission Time Interval (TTI) is the time unit for the eNodeB to scheduling the UL and DL data transmissions. TTI bundling functionality allows the sender to transmit the same data through the same radio resources on the consecutive TTIs, each TTI for one data transmission with different redundancy versions. So the receiver can improve the probability of decoding messages correctly and the coverage. In comparison with the HARQ retransmission, TTI bundling can avoid the control channel overhead used for acknowledgements.

- **Repetition**: HARQ retransmission and TTI bundling can be generalized to the repetition, where data transmissions can be repeated on different TTI or different frequency bands. Similarly, the control information can also use repetition to improve the coverage, such as the functionality of HARQ -ACK repetition.

- **Code spreading**: through the code spreading, the original code with little bits can be spread to a larger code with more bits, which is then transmitted through more TTIs. Such method can increase the transmission resistance to the channel attenuation and improve the coverage.

- **Low rate coding**: the low data rate coding with more redundancy information increases the probability of correctly decoding in the receiver side and can be used to improve the coverage.

- **Low modulation order**: the coding with low modulation order holds relatively lower Signal to Interference and Noise Ratio (SINR) threshold,
which can be used to conquer the higher propagation loss and improve the coverage.

- **RLC segmentation**: the Radio Link Control (RLC) can segment the large data packets into smaller packets, and each small packet can be coded with low rate or low modulation order. Such low data rate transmission can be used to improve the resistance to the channel attenuation and the coverage.

- New decoding techniques, such as correlation or reduced search space decoding, can be used to improve the probability of correctly decoding received messages during a relatively worse wireless circumstances and extend the coverage.

- **Power boosting / PSD boosting**: power-boosting technique can be used by the eNodeB to improve the power on the DL transmissions to the M2M devices. Furthermore, for the eNodeB or UEs with limited power level, the Power Spectral Density (PSD) boosting technique can be used to put all the power together on some decreased bandwidth, which effectively increases the transmit power density on such bandwidth. The increased transmit power can improve the coverage.

- **Relaxed requirement**: some control channel performance requirements for the normal UEs are sometimes too high for the UEs in extreme deployments. For example, after sending a Physical Random Access Channel (PRACH) preamble, the UE need to receive the Random Access Response (RAR) in a time window. Maybe the UE needs more time to accumulate the multiple copies of RAR to decode it correctly, so the window size should be relaxed, otherwise the UE may not decode the RAR and fail the random access.
• **Small cells**: implementing more cells is another method to extend the coverage. Macro cells are essentially deployed to improve the coverage or increase the network capacity. Additional, low power nodes can be deployed to establish some small cells in hotspots or coverage holes. Such heterogeneous deployment can provide very high data traffic for the hotspots or extended coverage for faraway placed M2M devices.

8.3 Integrating M2M traffic into a human-centric system: A Techno-Economic Perspective

M2M traffic in fundamentally heterogeneous and challenging to classify in a rough level; for this reason, there will never be a one-solution-fits-all in M2M and different strategies should be considered depending on the service requirements and scenarios. Considering the ICT historical evolution from the operators’ perspectives and their revenues model, there was an early pricing strategy for voice communications based on charging per usage. With the introduction of mobile data services, operators restructured their billing strategy for charging flat rates for data and later on they adopted data blocks models. M2M might be the next inflexion point forcing a new pricing structure; especially if we consider that most M2M device will only transmit very low amount of data but applications can generate higher profits.

8.3.1 The impact of a larger number of devices

M2M communications are mostly associated to small and sporadic data transmission and predictions agree that it will represent a density of devices orders of magnitude higher than H2H communications [4] [15], i.e., the latest forecasts published by Cisco Systems indicate that approximately 5.1% of the global mobile
traffic will be M2M-related by 2017; on the same report, it is estimated that M2M devices will represent 17% of the global number of mobile connections [15].

Besides the air interface optimizations described previously in section 8.2.1, scalable connectivity and hierarchical network architectures are needed in order to support the M2M traffic in the network [16]. Devices can be directly connected by means of cellular links, but in terms of scalability, the implementation of aggregation points plays an important role to allow the connection of larger number of devices. Such aggregation points can be also referred to as M2M gateways and are capable to collect and process data from M2M devices [16].

As explained in [16], the increase in number of devices will force a more spread adoption of hierarchical solutions in cellular networks. Such architectures take place when large cells are deployed to provide coverage and allow device mobility and additional small cells are also deployed in order to enhance the reliability and increase the capacity in area with higher demand. Additionally, it is mentioned the benefits of integrated multiple access radio technologies, including the underlying advantage of using additional spectrum (licensed or unlicensed) in order to cope with the capacity demand imposed by the increasing number of devices.

8.3.2 The integration of LTE and capillary networks as a scalable solution

Capillary solutions for M2M refer to networks deployed using short-range wireless communication technologies, e.g., ZigBee, WiFi, and Bluetooth. These technologies are use in M2M to connect devices in specific deployments. Moreover, Wi-Fi has
become a widely adopted technology and the emerging low power solutions can make it suitable for M2M communications.

Although they offer the advantages of low-cost and low-power end devices, the limited range and mobility make them unfeasible for some M2M applications. Therefore, capillary solutions are typically integrated with WAN networks, e.g., cellular networks, through M2M gateways, creating hierarchical architectures.

The IEEE and The IETF are two of the main Standards Developing Organizations (SDOs) that are working in this area, with IEEE focusing on the physical and medium access control layer and IETF on the upper layers, but excluding the application layer. The IEEE standards that are considered for M2M communications are IEEE 802.15.4 (which is used by ZigBee), IEEE 802.15.11 (which is used by WiFi) and IEEE 802.15.1 (which is used by Bluetooth). The main enhancement considered in 802.15.11 and 802.15.1 is addressing power consumption and resulted in Bluetooth Low Energy, and IEEE 802.11 Low Power. On the other hand, IEEE 802.15.4 is split in several versions [17]:

- **IEEE 802.15.4e.** Extension to the original IEEE 802.15.4 MAC layer, to support industrial applications, e.g. factory automation, smart buildings;
- **IEEE 802.15.4f.** Active RFID systems for bi-directional communications;
- **IEEE 802.15.4g.** Support for Smart Utility Networks.
- **IEEE 802.15.4k.** Critical infrastructure monitoring and ultra-low power operation.

IETF’s efforts in the area of capillary M2M are centered on IETF 6LoWPAN (IPv6 over Low Power Wireless Personal Area Networks), IETF ROLL (Routing Over Low power
Lossy Networks) and IETF CoRE (Constrained RESTful Environments) [17]. 6LoWPAN enables the use of IPv6 over constrained networks such as 802.15.4, by means of an adaption layer that performs, among others, header compression and packet fragmentation. IETF ROLL addresses routing in low power wireless networks, while IETF CoRE aims at extending web services to devices in constrained environments, which cannot use a full IP protocol stack, by defining an application transfer protocol (CoAP) for M2M applications.

At the moment, the way forward from an industry perspective is combining some of the above-mentioned standards and protocols into a standardized protocol stack that will meet the strict requirements of M2M applications. In particular, the envisioned protocol stack is comprised of the IEEE 802.11.4-2006 PHY layer, the IEEE 802.11.4e MAC layer, with the IETF protocols, 6LoWPAN, ROLL and CoAP integrated on top [18]. However, for applications that require higher data rates or seamless integration with existing products, WiFi is a viable candidate, considering both the efforts being done in reducing energy consumption and also the high adoption rate of this technology. Nevertheless, there are some technical challenges that still need to be addressed, for instance:

- **Interference management.** The densification of capillary network will lead to a negative impact on the performance due to the lack of effective interference coordination mechanisms in the overcrowded license-free 2,4 Ghz band. The design of efficient interference management techniques is still an active area for further research.
Handling large number of devices in the capillary domain. The IEEE 802.11 and 802.15.4 standards suffer from congestion when the number of devices is very high. This is due to the use of Carrier Sensing Multiple Access (CSMA) at the Medium Access Layer (MAC), which leads to very bad performance under heavy traffic loads. Therefore, the design of new ways of handling a great number of simultaneous connected devices becomes necessary.

Within the EXALTED project [19], an interesting solution has been proposed, a M2M Gateway as part of the LTE network architecture, which is capable of connecting the capillary networks with the core network. One of the clear benefits of these M2M Gateways is to increase the coverage of the wireless network. An M2M Gateway has the important role of enabling the interconnection of network operating under different radio technologies.

Other functionalities that an M2M Gateway can provide are protocol translations (between short-range technologies and LTE), resource management, device management, preliminary data processing and storage, data compression and aggregation, etc.

For Mobile Network Operators, integrating a M2M Gateway in the network architecture enable a scalability solution in order to handle large number of devices in dense areas or extend the coverage in remote location. Moreover, it could allow M2M Device management and monitoring through the M2M Gateway, enhancing the range of service that can be offered to M2M users (as further described in Section 8.4.3).
8.3.3 Technology migration and deployment strategies

Currently, there are two debated strategies regarding cellular networks for M2M communications. The first strategy corresponds to the establishment of dedicated legacy (i.e., GSM) networks for M2M communications. The positive side is that it could allow separating M2M and H2H traffic and managing each of them independently. Nonetheless, the current amount of traffic generated by M2M devices might not provide enough revenues in order to cope with the recurrent operation and maintenance costs and, moreover, support additional license of the spectrum for this exclusive purpose [2]. The second strategy correspond to an utter migration from legacy networks in order to allow spectrum refarming and deployment of networks capable to provide the require capacity for M2M and H2H communications seamlessly.

The traditional strategy for operators is to deploy outdoor macro and small cells, which are intended to cope with the capacity requirements but in many cases, fall short to provide the capacity a quality of service for indoor-generated traffic.

In countries with high mobile broadband demand, mobile network operators deploy denser heterogeneous 3G and 4G networks. Future deployment trends are having a strong focus on the feasibility of efficient network planning through denser network deployment in high demand areas in order to cope with the increasing capacity requirements. On [20] it is clearly stated that the main enabler of data increase in the last decades is directly related to smaller cell deployments.

In order to reduce the cost-per-bit and allow a more effective reuse of bandwidth, smaller cell deployments have been studied and results presented in [21] and [22].
conclude that small cells and LTE Femtocells bring benefits mainly on the reductions of deployment costs, since up to 74% on network-related costs have been reached with the available bandwidth is small (i.e., 5MHz). The benefits also depend on the preexisting infrastructure of the operator, since higher saving are feasible in cases where when Femtocells are not considered as part of the operator network, therefore, it does not constitute an additional cost for the operator.

Other research studies focus on the benefits to offload data traffic from the mobile network by means of WiFi or Femtocells [23]. Moreover, the widespread adoption of Wi-Fi technology had incentive the integration of Wi-Fi with small cells. This is usually referred to as Integrated Femto-WiFi solution (IFW). The Femtocell radio interface can be used for delay-constrained services, while throughput-demanding services can be served by the Wi-Fi radio interface.

Furthermore, M2M is all about enabling services and for that reason it is extremely important to understand the relationship between important, key players that will interact in the M2M ecosystem beside mobile network operators. Some of the most relevant players are presented next [21]:

- **Mobile Network Operators.** Mobile Network Operators (MNO) corresponds to the providers of wireless services that are in charge of the control and operation of mobile networks as well as the management of customer relations, where customers refer to end users with User Equipment. Conventionally, MNOs are used implement complete vertical solutions and handling all different parts of the network, a change from historical value.
chains is happening in the telecom ecosystem towards new business networks. This shift has made MNOs reconsider their position and roles in the M2M communications and connectivity, as will be further discussed in Section 8.4.1.

- **Network Vendors.** Network Vendors (NV) correspond to the manufacturers of the telecommunication equipment that delivery the technical assets for the MNOs. In recent years, NVs have expanded their traditional roles in the sector and it is nowadays common to find NVs in charge of the operation and maintenance of the networks on behalf of MNOs. In this sense, MNOs outsource the operation and maintenance of the network in order to focus on the customer relationship business. Related to the M2M ecosystem, important NVs in the market have developed vertical M2M solutions to handle device connectivity; their strong value relies on the fact the NVs have global and well-established relations with MNOs that give them a natural and beneficial position to promote the adoption of their solutions.

- **Managed Service Partner.** The term Managed Services refers to end-to-end solutions that lessen MNOs tasks and allows them to focus on their customer relationships. Managed Service Partners (MSP) usually are devoted to tasks related to operation and maintenance of the network. In general, MSPs are third party actors that are capable to reduce the operation cost from MNOs. The most representative case of MSP correspond to the NVs that are in charge of the operation of the network, MNOs trust this task to then mainly because is carried by the same producer of the equipment.

- **Data Aggregators.** Along this chapter, it has been discussed the challenge of connecting large number of devices, but the challenge does not ends there;
there are interesting ongoing efforts studying effective method to collect, store, analyze and process the data generated by all those connected devices. Based on the sensitivity of the M2M data, MNOs or third party actors are capable of handling this role but focusing on more detailed forms of analytical data processing needed. The challenges related to data aggregation in M2M correspond to privacy and data ownership, authenticity and security and scarcity of data processing skills [24].

- **M2M Service Providers.** The M2M service providers are in fact the actors driving the M2M ecosystem and correspond to those companies that firms offering over-the-top solutions based on M2M communications. These SPs normally prefer to adapt a service enablement kit in order to be able to offer their services over existing platforms unless the solution is either too complicated or implemented in a vertical mode that needs to be totally isolated regardless of the service enablement kits.

Still is not clear how these players will interact in the future, even when several solutions are already available in the market, there is no clear dominance of one solution and the position and relationship among them varies depending on the application and the scenario in which they are implemented.

In the next section, the business implications of M2M in LTE will be presented, highlighting the future challenges to overcome in order to enable the massive adoption of such solutions.

### 8.4 Business implications for M2M in LTE

The low traffic expected to be generated by most M2M devices will inevitably result in extremely low average revenues per device and the main revenues are
expected to come from the applications build on top of the transmitted data [25]. The major benefit of connected devices is directly related to seamless data availability. Moreover, in M2M applications, consumers are part of the aftermarket and deliver usage data and feedback to product manufacturers and M2M service providers, allowing co-creation of values [26], [27]. Therefore, when the data represents valuable information and it is processed appropriately, it can be exploited as a product.

It should be clear at this stage that there will not be one solution that fits all the technology requirements imposed by such a heterogeneous set of devices [1]. Following the same reasoning, both the technology and economic development of M2M communications will require a complex set of faculties that are unlikely to be covered by one single player in the industry. As discussed in section 8.3.3, different players are pushing their view and solutions in order to position themselves in the market [28]. More importantly, new players are also entering the market and positioning themselves in the M2M ecosystem.

The shift in the relationship between players can be compared with the mobile broadband case, where the access to affordable and high performance technology opened the market for Over-the-Top (OTT) services and applications, resulting in a new set of actors that dominate the ICT market, shirking the traditional structure and control of MNOs [28].

M2M communications are expected to be the next shifting point in the telecommunications sector, impacting on an extensive range of industries and
services. The open debate is still on who will provide these services and what would be the relationship between the different actors [21].

8.4.1 Is there a need for a change in operators’ mindset?

Previous studies (presented in [29] and [30]) examined different cases related to M2M-related services in the market. Findings show how it is simpler to analyze the values and benefits of isolated solutions but more importantly, it is highlighted how the implementation of successful services always incur in radical changes from traditional business thinking. Especially in term of connectivity services, since telecommunication providers could not find a feasible structure to generate value from the M2M service. Therefore, alternative relation emerged in the market in order to successfully provide the services. In most cases, M2M solutions based on mobile communication are hampered due to the lack of a proper M2M business model, since the traditional vision for MNOs is based on a provider-consumer perspective like the one shown at the left on Figure 8.4.2 (a).

As it has been explained in Section 8.3.3, M2M services are provided in a complex constellation of players and the end consumer is rarely involved in a direct interaction with MNOs. The M2M ecosystem tends to be more alike to the one shown on Figure 8.4.2 (b).
Figure 8.4.1: Change in the position of players and the direct relation with end-consumers.

8.4.2 The relationship between business challenges and engineers

Business studies on M2M have a strong focus on the analysis of the roles for incumbent and new actors needed in order to set-up and manage M2M services and, more importantly, which are the key values that generate profits for the different stakeholders involved in each application. There are key challenges that still need to be addressed for M2M in the business domain, as pointed out by the OECD report [1]. For instance, system and network architectures proposed for M2M rarely consider the role and market power of existent player in the M2M ecosystem that could definitively affect the decision making and future trends.

Also, current M2M services in the market remain as dedicated solutions developed and devoted to a single application, triggered by the high level of integration required with the targeted industry, i.e., solutions for energy efficiency are developed with direct focus on the energy sector but connected vehicle services are developed with the car manufacture’s perspective. Even within the same industrial sector, solutions are generally tailored for a specific company, leading to an increasing market fragmentation. Unavoidably, the consequence of these bespoke solutions is a higher development cost for M2M applications [31].

Another relevant aspect is the fact that M2M services are often provided in a complex and dynamic value chain, where the traditional provider-customer model adopted by MNOs does not apply [32], as presented in the previous section. Such complex interaction of players pose a challenge to the analysis of the real values and benefits for all the players involved. The real economic benefits are yet
unclear for many applications. This has directly affected the adoption of large-scale solutions in M2M like smart cities in Europe; where there is no enough value for one player in order to take the lead and drive the ecosystem of players [30]. Specific M2M applications can be successfully deployed when their value is clear and the business thinking is adapted to the new market perspective. But the integration of solutions and large-scale application remain as an open research challenges.

Two key challenges that directly limit any technical solution for M2M can be highlighted at this stage:

- It is difficult to change the position of traditional players that have significant market power.
- Some industrial sectors are dominated by players that are use to have the control position.

The challenge related to the change of position, there are lessons to be learned by engineering companies that manage to take new positions in the market. The case refers to Ericsson, originally a network vendors, that is now succeeding in their strategic move toward support of services based connected devices. An excellent example is the “Connected Vehicle” case, presented in [28], it describes the partnership between the car manufacturer Volvo and Ericsson and it shows the new added value that involved different business thinking. The solution has the following value proposition:

- The car manufacturer will be able to offer value added services either themselves or through third-party actors. Furthermore they can gather
valuable data on the operation of the car, which will improve maintenance and spare parts management.

- The drivers gain access to new range of applications and services.
- Third-party actors have a single channel to reach the end-users.

Ericsson, who originally based their engineering business on selling network equipment to MNOs, has started to expand the set of activities and offers managed solutions to industrial actors interested in solutions based on M2M; providing an M2M service platform, on which the car manufacturer builds its services. In figure 8.4.2, it is shown the change of position achieved by Ericsson, which was traditionally position only as a network vendors, following the description depicted in Figure 8.4.1 (a) and (b).

![Figure 8.4.2: Redefined position for M2M service offering.](image)

The example above shows the new type of business thinking applied in the M2M world, which is based on partnerships and re-thinking of roles.

The challenge related to the control position of industrial player is evidenced in the study carried out in [30], the M2M projects related to smart cities (Stockholm
Royal Seaport) and smart homes (Swedish e-home project in Halmstad) encountered similar barriers when dealing with different industrial sectors. Both projects pose the feasibility to have a common, and shared, network infrastructure in order to reduce the deployment and maintenance costs, lowering the entry barriers for new services based on M2M communications.

However, the interesting finding is that in both cases, dominant industry players like energy or media content companies showed unwillingness to share any type of infrastructure and are prepared to have their own system and in some cases, they even consider owning dedicated LTE networks for such purposes. There are several reasons adding to this fact, one is related to each company internal structure; they are use to operate in a certain manner and sharing assets will required deep restructuring, and since they already have a profitable business, there is no motivation to do so. Other reason is related to the direct customer and billing relationship, claiming that if the infrastructure is shared, it will be hard to differentiate the services and it will be more difficult to maintain their customer base. In any case, the barrier is more related to indisposition to change their position, since the analysis presented in [29] conclude that in cases where players are able to maintain the customer and billing control, and even more, they are proposed to share only with companies form different industrial sectors that do not represent any competition, the opinion remain invariable.

These challenges are open issues still to be address for services based on M2M communication, considering the position of important players, the increasing market power of new emerging companies and differentiated maker structure that varies on each country.
On the next section, business models options for M2M are discussed, focusing on the difficulties perceived by MNOs in order to be able to monetize from devices that do not necessarily generate the minimum required revenues.

8.4.3 Business models options for M2M

It is agreed that M2M will expand the business possibilities in several industries, allowing new business model opportunities to take place; as it is stated on the OECD report [33], some examples of this new business models could be:

- **Pay as you drive insurance.** Allow charging drivers based on distances, location and behavior. This could significantly reduce the cost associated to the insurance fee.

- **Products as services.** As explained in the report, there are already companies providing light as a service (charging per lumen) or energy-saving, which charging according to the saving they generate. It is expected that M2M solutions use this type of business model to a large extend.

Nevertheless, defining effective business models for M2M in different industry sectors is a difficult problem that has not been solved in the market yet [33]. There are key challenges that to be addressed. For example, many of the current solutions in the market have been developed as a complete vertical system, from connectivity to service provisioning [31]. Additionally, M2M services are frequently part of a dynamic value constellation where a provider-customer model does not apply [32].
All the discussion of new business models is centered on the M2M solutions and little in mentioned about alternative business models for MNOs, especially in LTE networks, in relation in M2M communications. The reason for it is that when the traditional value network for the telecommunication sector is considered, the roles of MNOs is usually constrained to provisioning of technology to allow data transfer; shrinking the relevance and value of their service. The real limitation for MNOs is that revenues on M2M communications are simply too low to turn a profit just from the connectivity, even with the most efficient operation.

MNOs face an important challenge, since their traditional revenue stream is founded on charging based on high amount of traffic per users. But M2M devices usually generate very low traffic and MNOs cannot base the charging on the traffic volumes, hence alternative revenue models needs to be considered [25].

In traditional, human-centric, communication users can be charged per devices or small group of devices (SIM cards) such as smartphones and tablets, and according to a subscription. The interaction between end users and MNOs is only related to billing and customer services in case of connectivity issues. In this sense, consumers have to adapt to the service offers given by the MNO or change to another operators.

The M2M case is entirely different, since the agreements are reached between MNOs and M2M service providers that have to manage large groups of devices with limited resources. International organizations like the 3GPP, ETSI and OECD have
identified several of the particular requirements from M2M users in order to effectively handle their devices [34] [35]:

- Access to network status in order to differentiate network failures from device malfunctions.
- Ability to change the communication provider without replacing the SIM card inside each device, and ability to have national roaming in case of network failure.
- Devices must be ready to operate out-of-the-box, but should be activated remotely once is ready to be used.
- Roaming capabilities should be seamless in terms of cost.
- Guarantee network availability for long periods. In case of LTE, guarantee increasing coverage areas.

What this rationale tries to explain is the fact that MNOs should not only rely on their connectivity provision to monetize on M2M; on the contrary, solution-based business models, as recommended by Analysys Mason, could be more effective [36]. A solution-based business model implies that pricing on M2M should never be focused on bytes per devices; instead, MNOs should focus on providing the necessary support they can provide in the operation of the services.

An example in market of this approach is Telenor Connexion\textsuperscript{1} [36], which uses the solution-based business model, basing the price of the product according to the requirements of the M2M service provider in terms of the level of service, type of information transferred over the network, managed service for devices, consulting and support in the integration process and security requirements. The main

\textsuperscript{1} More details available at Telenor Connexion’s website. \url{http://www.telenorconnexion.com/}
objective should be to focus on service enablement, i.e., provide advanced M2M device management like remote software management, simplify communication between applications and devices, and simplify data collection [37].

8.5 Conclusions

Through this chapter, technical and business aspect related to the support of M2M communications over LTE networks have been described, highlighting the current works and describing the different alternatives that are currently under consideration both in the academia and in the industry sector. Some of these alternatives are summarized below:

- Small cells could be the most cost-effective solution in ultra-dense scenarios. But there is an general unwillingness from many MNOs to deploy small cell or indoor solutions
- The role of mobile operators is not always fixed and many times replaceable; the emergence of “independent” indoor network operators. Moreover, NVs are a strong force in the market, able to manage the connectivity of devices.
- Intermediary actors have a strong position, they control service platform and have relations with end-users and service providers.

M2M is about services based on communication, not about communication services. The heterogeneity of the solutions results in a complex ecosystem that involves many additional players that dynamically interact with MNO. Moreover, the current understanding in the market is that MNOs will not drive the M2M ecosystem but rather support the solution development.
References


[11] 3GPP TR 22.803, "Feasibility study for Proximity Services (ProSe), Release 12, ".


[34] OECD (2012), "Machine-to-Machine Communications: Connecting Billions of Devices,".

[35] 3GPP TS 22.368, "Service requirements for Machine-Type Communications (MTC) Stage 1, Release 12."
