Machine-Type Communication (MTC)

Machine-Type Communication (MTC) denotes the broad area of wireless communication with sensors, actuators, physical objects and other devices not directly operated by humans. Different types of radio access technologies are targeting MTC. For Long Term Evolution (LTE), it has emerged as an important communication mode during the recent standard evolution. The research and development efforts made to enhance LTE in a way to support MTC clearly indicate the need for the wireless system architecture to address MTC. As the role of MTC is expected to grow in the future, there is a good opportunity in the development of a 5G wireless system to address MTC from the very beginning in the system design.

Use cases and categorization of MTC

The general use case of low-rate MTC

MTC use cases exist in a wide range of areas. They are mainly related to large numbers of sensors monitoring some system state or events, potentially with some form of actuation to control an environment. One example is automation of buildings and homes, where the state e.g. of the lighting, heating, ventilation and air condition, energy consumption, are observed and/or controlled. There are also wide area use cases, such as environmental monitoring over larger areas, monitoring of some infrastructure (e.g. roads, industrial environments, ports), available parking spaces in cities, management of object fleets (e.g. rental vehicles/bicycles), asset tracking in logistics, monitoring and assistance of patients. There are use cases that comprise remote areas, such as in smart agriculture. In the context of the use cases described in Chapter 2, MTC appears as an important, if not the crucial, element in (1) autonomous vehicle control, (3) factory cell automation, (6) massive amount of geographically spread devices, (10) smart city, (12) teleprotection in smart grid network and (15) smart logistics/remote control of industry applications.
**Use case: the connected car**

The connected car has gained a lot of attention during the recent years, as it enables new services and functionalities for the automotive industry based on the use of wireless communications, and, most particularly, cellular systems. Only these systems are capable of providing the wide area coverage and performance demanded by automotive applications, including both human and machine type of communication. For Human-Type Communication (HTC), the challenge is to provide to passengers in the vehicle with comparable mobile broadband connectivity performance as can be found in stationary environments. In the automotive context, MTC refers to the exchange of information between machines that can be located in vehicles, user devices or servers, with little or no human interaction. The scope of MTC in the automotive domain encompasses a wide range of applications including road safety and traffic efficiency (e.g. highly autonomous driving), remote processing or remote diagnostics and control, among others. Some automotive applications in the area of MTC, such as road safety and traffic efficiency, require ultra-reliable connections with stringent requirements for latency and reliability, as the timely arrival of information can be critical for the safety of passengers and vulnerable road users. Furthermore, a highly reliable and widely available connectivity to the cloud can allow some functions, such as video processing, audio recognition or navigation systems to be carried out remotely by cloud servers instead of the central processing unit in the vehicle. Remote processing has not only the potential to increase the processing power beyond the vehicle capabilities but also to enable a continuous service improvement during the vehicle’s lifetime. Other applications such as remote diagnostics and control are based on the transmission of small telemetry and command messages, and therefore, do not possess stringent requirements
Use case: the smart grid

The smart grid represents an evolution of the electric power grid into an immensely complex cyber-physical system that will rely on decentralized energy production, as well as near-real-time control and coordination between the energy production and consumption. A fundamental enabler of the smart grid is the reliable, two-way wireless MTC. In the downlink, the smart grid should be able to send commands and polls. The communication design is more challenging in the uplink, as it needs to coordinate a large set of partially or fully uncoordinated transmissions and therefore research attention is more focused on the uplink. An exemplary MTC device in the smart grid is the smart electricity meter. At present, smart electricity meters are primarily used by electricity providers only for availability monitoring and billing. However, as the Distributed Energy Resources (DERs), such as wind turbines and solar panels, increase their share of energy generation, the role of the smart meter is expected to become more complex and communication-intensive [3]. Specifically, there can be increased need to grid state estimation, where the meter should frequently monitor and report the power quality parameters, such as e.g. power phasors, which enables real-time estimation and control of the grid state. Due to their sheer number within a given region, the smart meters represent a showcase of massive MTC. The Smart Grid also features instances of ultra-reliable MTC, as many of the devices should reliably and very timely report critical events in the grid, such as outage or islanding of a micro-grid.
**Fundamental techniques for MTC**

Traditionally, the evolution of cellular wireless communication systems has been centered around broadband communication and provision of increasingly high data rates. The emergence of mMTC and uMTC changes that focus, as the target scenarios do not require excessively high data rates, but rather new modes of connectivity to a massive number of simple devices and/or support extremely reliable connections. Although the performance requirements for mMTC and uMTC are vastly different, they point toward the revision of the same set of communication-theoretic mechanisms. In this section, two of those mechanisms that are promising for the radio access part are discussed: (1) creation of short packets, where the data and the associated control information are of comparable size, and (2) non-orthogonal protocols for distributed access.

**Data and control for short packets**

The success of broadband wireless communication systems is largely based on the methods for reliable transmission that follow the principles of information theory. Those principles are applicable when each transmitted packet contains a large amount of data, due to the following two features: (1) large data means that one can use methods (codes, modulation) that are applicable in an asymptotic case to guarantee reliable transmission under a constraint of the total energy used for transmission; (2) the size of control information is small compared to the size of data, as shown in Figure 4.2(a) such that, even if the control information is sent suboptimally (e.g. repetition coding), its overall effect on the system performance is negligible. These features have led to a common approach in designing broadband communication, in which data is transmitted using optimized and sophisticated methods, while the transmission of control information has been largely left to a heuristic design. This approach for creating packets for broadband communication needs to be revised when the amount of transmitted data

![Diagram](image-url)
Massive MTC

Design principles

The basic design principle of massive MTC is to exploit that mMTC services are delay-tolerant and consist of transactions with small amounts of data. These relaxed requirements can enable extensive sleep cycles for devices (to enable long battery lifetimes), define low-complexity transmission modes (to enable low device costs) and define extra-robust low-rate transmission (to enable extended transmission range). Since the total data volume of massive MTC is rather small (compared to e.g. multimedia services like video), even a very large number of devices is expected to generate (on average) manageable traffic volumes for a mobile network that is also dimensioned for mobile broadband services. However, still considerations have to be given for mMTC with high density of devices when it comes to handling the control signaling, context handling in the network, as well as overload of system resources in access peaks when large device populations try to access the network simultaneously.

Technology components

As mentioned previously, the desired features for a massive MTC system are low device complexity, long battery lifetime, and scalability and capacity. In the following the technology components that address these features are presented.
**Features for low device complexity**

The complexity of a device is related to the performance that is expected for the communication. Massive MTC services transmit typically infrequently small amounts of data and have relaxed requirements in terms of required data rate and transmission reliability. This provides opportunities to exploit the relaxed performance requirements to simplify the transmission mode and reduce device complexity. A significant evaluation on how device complexity can be simplified has been provided in [7] for LTE, and the features listed below are already addressed for LTE evolution in Releases 12–13. However, the general findings are independent from the specific radio access technology and are explained in the following.

Transmitting at wide bandwidth can provide high peak data rate at the costs of device complexity. When the transmission and reception bandwidth used by the device is bounded, the costs can be reduced compared to wide-bandwidth devices. Therefore, for mMTC devices it is desired to have a transmission mode with a limited device bandwidth. Already at bandwidths in the order of 1 MHz, as it is used in e.g. Bluetooth design, very low device complexity is achievable. It shall be noted that the total system bandwidth provided by the 5G system can be much wider, and may be used by other devices targeting e.g. high peak rates. A further cost reduction can be achieved by limiting the peak data rate in order to limit the amount of allocated buffer. The number of antennas that a device includes directly affects the device complexity, so that a low-complexity transmission mode should not depend on multiple device antennas being present. Further, a device that needs to transmit and receive simultaneously requires a duplex filter to separate the transmit signal from the receiver. If a device is alternately transmitting or receiving, like in time-division duplex or half-duplex frequency division duplex, the costs of a duplex filter can be avoided. Finally, it is desirable to limit the transmit power of a device, so that the power amplifier can be embedded onto the integrated circuit, thereby avoiding the need for a separate external power amplifier. For this purpose, for LTE Release 13 a new device power class is defined, where the device output is limited to around 20 dBm.
Features for service flexibility
MTC services comprise typically only small amounts of data that are transmitted per transaction. However, it is not easy to define the maximum transaction size for MTC services. Even more, services can be easily updated during the lifetime of a device, e.g. the monitoring of some process may be based on infrequent status reports. However, after a few years in operations, the service may be updated via over-the-air configuration or software update. As a result, the amount of data transmitted per device, the frequency of transmissions, and the priority of message may change over time. Therefore, a flexible access design is needed in order to enable flexible service provisioning. Even if some upper bound in capabilities may be set by the category of the device, some flexibility in service provisioning shall be catered for.
Features for coverage extension

Coverage is normally defined as the maximum range or path loss at which a certain throughput limit can still be upheld. With the latency tolerance of mMTC, this must not any longer be the case but a degradation of the throughput is acceptable. In fact, LTE Release 13 is specifying 15 dB coverage gain by means of time repetition. The resulting low bit rate is not a problem in itself, however, since the device energy consumption is very dependent on the time during which the device cannot stay in sleep mode, the longer transmission times are affecting the battery life. Therefore, the coverage extension and battery life requirements are somewhat contradictory and it is difficult to meet them at the same time. One way of doing this is to make use of the massive number of devices and allow for some of them to function as simple relays and greatly improve the link budget for devices in challenging coverage. Clearly, the relays dissipate more power, but improve the overall connectivity. This has been evaluated in the Madrid propagation map developed in METIS [19] at a 2 GHz carrier frequency and a bandwidth of 1 MHz, where the devices acting as relays are operating in the same frequency band and with the same device output power of 23 dBm, and still send their own traffic (for more details see [20]). Further, the simulations are static, limiting to two-hop relaying and the uplink and downlink considered separately. In Figure 4.4 it is seen that the drop rate is greatly reduced when MTC devices are enabled to act as relays to other MTC devices. The drop rate refers to those devices that are dropped by the network if a certain throughput cannot be upheld. Therefore, this relates to the coverage since devices are dropped if located outside a certain cell radius. Although this evaluation does not give the coverage enhancement in number of dB, MTC device relaying is obviously very promising for improving the coverage. More importantly, it does so without extending the transmission times, which is beneficial for the device battery lifetime, as discussed below.
Features for long battery lifetime

In broad terms, the device energy consumption is proportional to the transmission and reception time of the device during which it cannot power down to a conservative sleep state [21]. For very long battery lifetimes, also the self-discharge rate of the battery plays a role. The discontinuous reception (DRX) technique was developed to reduce the reception time of the device. In LTE Release 10 DRX cycle lengths are configurable up to a maximum value of 2.56 s, which means that the device is only required to listen for paging e.g. once per such interval and not continuously. To reduce the transmission time, many proposals for 5G rely on the working assumption that contention-based transmission of data is beneficial [22]. That is, omitting the RRC Connection Setup procedure and transmitting the payload (and associated control overhead) at once. In the most favorable case, the device would not even have to obtain uplink synchronization before transmission. This is applicable to waveforms that require no or relaxed synchronization such as FBMC or UFMC (see Chapter 7), as an unsynchronized uplink will cause no or little inter-subcarrier interference, or to cases where the timing advance could be estimated, e.g. reusing a previous value for stationary devices. Figure 4.5 depicts in different curves different reporting periodicities for the uplink payload of 125 Bytes. From Figure 4.5 it can be seen that the largest gain is achieved by extending DRX cycles beyond 2.56 s, while contention
based transmission (labeled “no UL sync”) provides an additional, but significantly smaller, extra gain. The gains of extended DRX cycles are especially for longer uplink reporting periodicities (for more details see [22]). The longer the reporting periodicity, the smaller the gain from contention-based transmission and the larger the gain from longer DRX cycles since the paging monitoring is the dominant part of the device energy consumption. For too frequent reporting, every minute in this case, the 10 years battery lifetime requirement is simply not fundamentally possible. The largest gains of contention-based transmissions are therefore found for the case of 5-minute reporting. In this case, the extension of the DRX cycle from 2.56 s to 300 s gives an improvement of 20 times longer battery life (see [21] for a calculation of the battery lifetime), and the contention-based transmission can further increase this up to a factor of 25 times. Note, however, that in practice these results represent an upper limit, as one also needs to account for the overhead in the RRC Connection Setup (addresses, security, etc.), which also consumes transmission resources [23]. To further separate out the gains of omitting the RRC Connection Setup signaling from those of not having to obtain uplink synchronization, a third alternative is included in which Random Access is used only to obtaining the Timing Advance in Random Access Response, which is sent with a fixed timing (this alternative is labeled “improved LTE/OFDM”). It is seen that the majority of the additional gain (i.e. increase from 20 to 25 times) comes from omitting the initial RRC signaling, since in this case the gain is 24 times of that of the reference case. Still, the largest gain (20 times) overall comes from monitoring paging that is performed less often.

The above results are for a typical cell-edge throughput of 23 kbps. The cell edge users have the lowest battery lifetime and one should seek to fulfill the 10-year battery lifetime for those devices. Devices in the cell center or other favorable locations benefit from short transmission times and can obtain significantly longer battery lifetimes then devices at the cell edge. If extended coverage modes (e.g. by means of time repetition) is considered, with a data rate of around 1 kbps, the gain of contention-based transmission is always below 1% and therefore insignificant.
Ultra-reliable low-latency MTC

Design principles

Reliable low-latency design is often needed in a control-related communication context. This can be remote control of machinery (e.g. tele-surgery or operations in hazardous environments) or factory cell automation. In these use cases, data messages are typically short control messages, e.g. 100–1000 bits that need to be transmitted within very strict delay bounds. For example, for industrial automation requirements can be as strict as requiring guaranteed end-to-end packet transmissions within 1 ms. Required reliability levels can be in the 99.999th percentile (i.e. \((1 - 10^{-5})\) but may need to even reach levels up to 99.9999999th percentile (i.e. \((1 - 10^{-9})\)) in extreme cases. The following Figure 4.10 shows the design objective of ultra-reliable low-latency communication in contrast to typical mobile broadband objectives. The mobile broadband systems commonly focus on metrics related to median and peak performance, as well as certain modest percentile (e.g. the 95th or 99th percentile) in the performance distribution. For uMTC the focus is rather on a very high percentile, given by the reliability requirement, which ensures that at this level the required delay can be met. Improving the transmissions that are already within the delay bound, which is equivalent to a higher data rate, is not a goal per se.

When ultra-high reliability needs to be attained at a system level, then one way to proceed is to derive reliability requirements for each of the modules that constitute the system. For a data packet, such requirements can be sublimed as, e.g. “transfer of data packets that have at most B bytes with a delay D less than L seconds in 99.99% of the
attempts”. This creates a rather simple criterion to see whether the system meets the requirement or not. However, the problem with this criterion is that the data transmission model needs to report a failure whenever this simple and rigid criterion is not met. The data transmission module may still be able to send something and the overall service at the system level does not need to fail. In order to achieve this type of operation, one needs to reconsider the way in which a certain communication service is composed. The concept of Reliable Service Composition [13], discussed in Section 4.1.2.2, enables different versions of a service according to the reliability at which the connectivity can be provided.

In order to illustrate the idea, let us consider RSC in the case of V2V communication. It is noted that the percentages used in the example are provisional, only for illustration. The basic version of the service is available 99.999% of the time. In the V2V setting, the basic version could involve transmission of a small set of warning/safety messages without certification. The fact that the set of messages transferable in the basic mode is limited can be used to design efficient transmission mechanisms that use low rate. An enhanced version of the service is available 99.9% of the time, includes limited certification and guarantees the transfer of a payload of size D1 within time T1 with probability 99.9%. The full version is available 97% of the time, includes full certification and guarantees for transfer of payload of size D2 > D1 within time T2 < T1 with probability 97%. The key issue in making RSC operational is to have reliable criteria to detect which version the system should apply at a given time, i.e. have suitable indicators of availability and reliability.
Technology components

The desired features for an uMTC system are reliable low latency, and availability indication. In the following the technology components, among them D2D communications, that address these features are presented.

Features for reliable low latency

In wireless transmission, Rayleigh fading adds significant signal fluctuations, which increases the risk of temporary outage and packet losses. This can be compensated with a fading margin added to the average SNR. In order to achieve high levels of reliability, a significant margin needs to be added, e.g. 50 dB–90 dB for reliability levels of $1 \times 10^{-5}$ to $1 \times 10^{-9}$ (see Figure 4.11). Adding diversity to the transmission with independently faded signal components provides robustness against fading losses. With higher level of diversity, the fading margin can be significantly reduced. With a diversity order of 8 or 16, the fading margin for a reliability of $1 \times 10^{-9}$ can be reduced from 90 dB to 18 dB or 9 dB respectively. Diversity can be achieved in dimensions space, frequency and time. Due to the low-latency requirement, the time dimension cannot be exploited. Overall, diversity is one of the key technology components to enable reliability on short time scales in wireless communication system with fluctuating channel properties [28][29]. Other sources of uncertainty that challenge reliability is interference caused by other transmissions. Orthogonal
Figure 4.11  Required fading margins for ultra-reliable low-latency transmission.

Figure 4.12  Slot structure and access delay.

multiple access (like e.g. OFDM) with coordinated channel access enables robustness against interference.
**Feature for reliability: availability indication**

The application of wireless communication systems for ultra-reliable communication use cases depends on the capability to provide reliable connectivity. The failure to comply with the reliability requirements can render the service useless as it can lead to costly damage (e.g. as in the case of failure of a smart grid or an industrial process) or cause even harm (e.g. to vehicle passengers and other traffic participants in case of road safety). At the same time, wireless communication systems are inherently subject to uncertainties of the radio environment and are today typically designed to provide only modest reliability levels. A design to provide ultra-high reliability levels at all times and in every reception scenario may lead to an overdesign of a system with little commercial viability. In order to cope with this problem and enable the provision of uMTC applications with strict reliability requirements, future 5G systems should be able to warn the application about the presence or absence of reliability according to its requirements. In this manner, the applications would be capable of using the wireless communication system only in those instances in which the reliability can be guaranteed. For example, highly autonomous driving systems could reduce the velocity in case of insufficient reliability on the wireless interface, or even prompt the driver to take control of the vehicle.

In order to enable the use of wireless communications for the provision of critical applications, such as those in the area of road safety and traffic efficiency, it is of paramount importance to ensure high reliability. Within this context, two different aspects must be taken into account:

- The probability of false alarm, that is, the probability that the link is indicated as reliable when it is not, must be kept below a maximum as specified by the application. This is determined by the precision of the channel estimation and prediction method used for the computation of the availability. More accurate channel estimation and prediction methods result in a better knowledge of the propagation channel, and therefore, improve the accuracy of the availability estimation. The results in [30] illustrate how it is possible to ensure availability estimation at user velocities with Doppler shift up to 200 Hz as long as the propagation channel can be predicted perfectly.
- The availability of the wireless communication link, that is, the probability that the link is indicated as reliable, should be maximized under the false alarm. In general, the provision of link reliability depends on the robustness of the physical layer (i.e. channel coding, diversity schemes, etc.), the amount of available radio resources, etc. In this sense, the utilization of high orders of diversity is fundamental to improve the availability of the wireless communication link.
Features enabled by D2D communications

D2D communication is an important enabler for uMTC applications from the automotive and industrial domain that have very stringent latency requirements (in the order of

---

Exemplary implementation supporting the availability indication for reliability.
milliseconds). Exploiting the D2D capabilities, two physically close-located communicating peers can take advantage of their proximity to exchange information over a direct link so that messages do not need to be relayed by a central BS. Especially for short-range communication requirements, direct communication between peers by means of D2D reduces the end-to-end latency. It is spectrally more efficient than communication through the cellular infrastructure, as this avoids the redundancy in the use of resources due to the uplink and downlink and it enables a spatial reuse of resources by reducing the interference in the service area. Furthermore, the combination of D2D with cellular infrastructure-based communication can lead to increased reliability by means of multi-path diversity.

As an example, V2X communication – that is, the direct exchange of messages between vehicles (V2V), between a vehicle and a device carried by an individual (V2D) and between a vehicle and the infrastructure (V2I) – has the potential to significantly improve the provision of road safety and traffic efficiency services, including highly autonomous driving. While cloud connectivity is necessary in order to allow vehicles to download high definition digital maps as well as real-time traffic information, direct V2X communication based on D2D improves the awareness of the vehicle beyond the capabilities of sensor technology. Modern vehicles are equipped with a variety of sensors including cameras and radars that allow them to recognize objects in their environment. Nevertheless, the range of such sensors is quite limited and insufficient for the recognition of most hazards on the road. Combining the information gathered by multiple vehicles and fixed infrastructure (e.g. traffic surveillance cameras at intersections), leads to an extended perception horizon reaching far beyond the limited field of view of a single vehicle or its driver [31]. This would enable drivers and systems for autonomous driving to recognize hazards in advance and take preventive actions much earlier, but requires cooperation (i.e. information exchange) between the traffic participants. Note that the V2X information is only relevant for other traffic participants in the proximity, but not to those located beyond a certain distance. Therefore, D2D is an important key technology for automotive uMTC services.