View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Aaltodoc Publication Archive

Aalto University

Aaltodoc

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Author(s): Shariatmadari, Hamidreza & Iraji, Sassan & Jäntti, Riku

Title: Delay Analysis of Network Architectures for Machine-to-Machine Communications in LTE System

Year: 2014

Version: Post print

Please cite the original version:

Shariatmadari, Hamidreza & Iraji, Sassan & Jäntti, Riku. 2014. Delay Analysis of Network Architectures for Machine-to-Machine Communications in LTE System. 21st International Conference on Telecommunications (ICT) 2014, Workshop on M2M Solutions and Services. 492-496.

Rights: © 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other work.

All material supplied via Aaltodoc is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Delay Analysis of Network Architectures for Machine-to-Machine Communications in LTE System

Hamidreza Shariatmadari[1], Sassan Iraji[1], Andrés Laya[2], Omer Anjum[3], Riku Jäntti[1], Zexian Li[3], Carl Wijting[3]
[1]{firstname.lastname@aalto.fi}, Aalto University, Finland
[2]{lastname@kth.se}, KTH Royal Institute of Technology, Sweden
[3]{firstname.lastname@nokia.com}, Nokia Research Center, Finland

***Abstract*—Machine-to-machine communications has emerged to provide autonomic communications for a wide variety of intelligent services and applications. Among different communication technologies available for connecting machines, cellular-based systems have gained more attention as backhaul networks due to ubiquitous coverage and mobility support. The diverse ranges of service requirements as well as machine constraints require adopting different network architectures. This paper reviews three M2M network architectures to integrate machines into the LTE system and analyzes their associated communication delays. It also presents how the appropriate networks can be selected for some machine-to-machine applications, fulfilling their latency constraints.**

***Index Terms*—Machine-to-Machine Communications, LTE, Network Architecture, Capillary Network, Latency**

I. Introduction

As the number of mobile subscribers is becoming saturated, particularly in developed countries, the next revenue generator in mobile industry is expected to be from machine-to-machine (M2M) services. According to Cisco Internet Business Solutions Group, there are already more autonomous devices connected to the Internet than people, approximately 12.5 billion devices in 2010 alone. Their predictions account for 25 billion connected devices by 2015, and 50 billion by 2020 [1]. Special units for M2M business and technology have been dedicated by different stakeholders from the multinational mobile operators, service providers, and mobile vendors.

M2M communications or machine-type communications (MTC) is defined as the necessary infrastructures for data communications among machines without the need for direct human intervention. Commercial applications for M2M services include: security, tracking, health monitoring, car-to-car communications, payment, smart grid, automatic energy metering, and remote maintenance/monitoring [2]. M2M applications are diverse with features, such as: low cost, supporting a large number of devices, small and infrequent data transmissions, high reliability, and real-time operation.

The traffic characteristics and requirements for M2M applications are generally different from regular human-to-human (H2H) wireless communications. In many M2M applications, small and infrequent data is generated from a mass number of devices which imposes a higher traffic volume on the uplink. The latency constraints in these applications may be more important than supporting high data rate transmissions, which requires efficient data collection and distribution. The distinct features in M2M applications have raised new technical challenges, such as M2M network architecture, security, energy efficiency, resource allocations, delay restriction, and scalability [3]. The communication networks, in particular wireless cellular systems, play an important role in supporting M2M services and widely deploying them. With the widespread introduction of the LTE and decommissioning of legacy systems, migration of M2M devices to the LTE system is under investigation by many cellular operators and standardization organizations.

In order to meet different M2M service requirements and machine constraints in the LTE system, it is necessary to employ advanced network architectures [4]. Several research efforts have began investigating the possible network improvements to accommodate M2M devices in cellular systems. As an example, the LOLA project focused on solutions to achieve low-latency in wireless communications [5]. The network access delays for real-time machines in the LTE system are assessed in [6]. However, in some applications like automation systems, constrains rely more on link delays between machines than network access delay. This paper analyzes the link establishment and data transmission delays between machines in three proposed M2M network architectures and analyzes their associated delays. These architectures consist of data transmissions through the cellular system, communications with support of direct link between machines, and integration of capillary networks. Furthermore, some typical M2M applications are mentioned as examples to check the feasibility of mentioned networks to support their latency goals. The paper also addresses the existing challenges in M2M communications for their large scale deployment.

The paper is organized as follows. In Section II, three possible network architectures for M2M communications are presented and the communication delays are assessed. In Section III, the delay constrains for some M2M application are covered and they are mapped to the feasible network architectures. Finally, conclusions and future lines of research are drawn in Section IV.

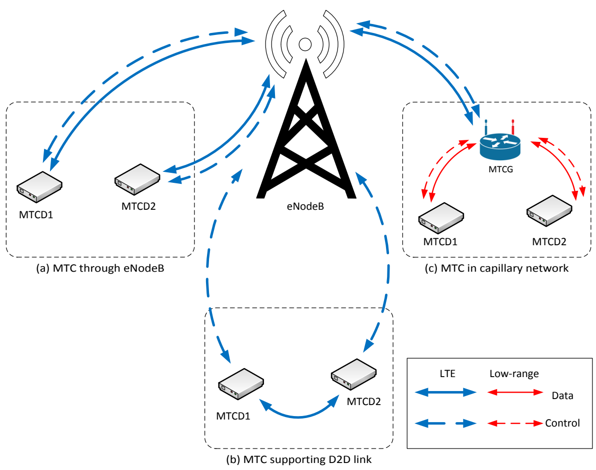

Fig. 1: The LTE-based network architectures for M2M communications

II. Delay Analysis in the LTE-based M2M Network Architectures

Various network architectures have been proposed to integrate machines into the LTE system and fulfill their requirements [4][7]. The proper architecture should be chosen for each application according to service requirements and machine constraints. In this Section, the delay performances of three different network architectures depicted in Fig. 1 are analyzed. In each of these systems, it is assumed that two MTC devices (MTCDs) exist; MTCD1 intends to establish a link with MTCD2 and start communicating. It is also assumed that both machines are served by a single eNodeB or a MTC gateway (MTCG). This is a common scenario for many M2M systems whose devices are located in a small area. The delays for link establishment and data transmission are our focus in this context.

A. *Data transmission through eNodeB*

M2M connectivity can be established simply by utilizing the existing LTE network architecture designed for H2H communications. This requires that machines are equipped with LTE-based radio transceivers. This design, as shown in Fig. 1(a), entails that all control signals (depicted with dash lines in the Fig.) and data transmissions (drawn with solid lines in the Fig.) from devices are passed through the radio access network. In this architecture, devices should request a link establishment with the eNodeB to obtain radio resources for transmitting or receiving data. Devices initiate link establishment by preforming random access procedure. The eNodeB does paging and triggering a device to perform link establishment procedure, in case it needs to transmit data to the device. The network provides dedicated radio resources and informs the device to utilize them for its communication.

In the above scenario, it is assumed that both machines are in LTE IDLE state, and MTCD1 intends to establish a connection with MTCD2 through the serving eNodeB. Hence,

TABLE I: Link establishment and data transmission latencies through eNodeB.

	Procedure	Latency Estimates
Link Establishment	Link establishment for MTCD1:	C-plane establishment delay (≈ 50 ms+2Ts1c*) U-plane establishment delay (≈ 15 ms+2Ts1u**)
	Data transmission for requesting a link connection with the destination machine:	Message transmission + processing time (1+4 ms)
	Paging the MTCD2:	Sending paging message on PBCH (≈ 40 ms)
	Link establishment for MTCD2:	C-plane establishment delay (≈ 50 ms+2Ts1c) U-plane establishment delay (≈ 15 ms+2Ts1u)
	Total delay:	187-265 ms
Data Transmission	Data transmission from MTCD1 to the eNodeB:	(1 ms)
	Data processing at eNodeB:	(4 ms)
	Delay for the next available downlink subframe:	(0-4 ms)
	Data transmission from eNodeB to MTCD2:	(1 ms)
	Data processing at MTCD2:	(4 ms)
	Total delay:	9-13 ms

*Ts1c (2-15 ms), **Ts1u (1-15 ms)

the MTCD1 should first establish a link with the eNodeB and request that the network send paging message to the MTCD2 for establishing a link. The eNodeB sends a paging message for the MTCD2 over the Physical Broadcast Channel (PBCH) to perform random access procedure. When both devices successfully establish links with the eNodeB and transit to LTE ACTIVE state, they can start communicating with each other through the network. The required steps for link establishment and data transmissions between machines, with estimated delays associated to them are described in TABLE I. The delay for link establishment in the LTE system between a machine and eNodeB consists of Control plane (C-plane) and User plane (U-plane) delays. The C-plane delay depends on the state of the machine: LTE IDLE, LTE ACTIVE, Radio Resource Control (RRC) IDLE, RRC CONNECTED, network configuration, and employing either frequency division duplexing (FDD) or time division duplexing (TDD) frame structure [6]. Supporting discontinuous reception mode adds an additional delay when a machine intends to transmit or receive data [8]. The U-plane delay depends on the scheduling policy, buffering, transmission time interval , frame alignment, and number of data retransmissions. The estimated delays are calculated from the first time instance when the initiating MTCD has channel access. It is assumed that the TDD frame structure is employed in the network and the reception of all data are successful without the need for data retransmissions [9]. The delay associated with data transmission between two machines compromises receiving the data from the source machine and delivering it to the destination machine. To exclude additional delay from the network traffic, it is assumed that the traffic load in the network is low, thus adequate radio resources can be allocated for machines without any constrain.

B. Data transmission through network-assisted device-to-device link

The data transmission delay between machines in the previous architecture can be significantly reduced by supporting direct communication between machines, referred to as device-to-device (D2D) communication. The cellular system is responsible for the link establishment between machines and allocating the required radio resources for them. Compared to other low-range communication solutions operating over unlicensed radio bands, the advantage of this peer-to-peer connection over cellular system is that the network can control the interference condition in the whole system. In addition, the coverage of cellular systems is higher which eliminates the cost of further investment for deploying new access points. The cellular system can also establish secure connections between machines without the need for manual device pairing [6]. The network architecture supporting D2D connection is depicted in Fig.1(b). The support of D2D in the LTE system has not yet been standardized and is still under development, hence the exact delay for link establishment is not yet clear. However, a simple proposed scheme is that machines first become connected to the eNodeB, perform some link measurements and report the results to the eNodeB. If the D2D link is feasible, the eNodeB provides the required resources and informs the machines to utilize the direct link connection. The setup delay for this scheme is higher compared to the previous mentioned network architecture, however, the data transmission delay is lower since data is not passed through the network. The complexity of this scheme is also higher as the eNodeB should constantly monitor the link quality and force machines to switch back to data transmission through the network, if the quality of direct link is not satisfactory. The link establishment and data transmission procedures, and estimated delays for this scheme are provided in TABLE II.

TABLE II: Link establishment and data transmission latencies with D2D support.

	Procedure	**Latency Estimates**
Link Establishment	Link establishment for MTCD1:	C-plane establishment delay (≈ 50 ms+2Ts1c) U-plane establishment delay (≈15 ms+2Ts1u)
	Data transmission for requesting a link connection with the destination machine:	Message transmission+ processing time (1+4 ms)
	Paging the MTCD2:	Sending paging message on PBCH (≈ 40 ms)
	Link establishment for MTCD2:	C-plane establishment delay (≈50 ms+2Ts1c) U-plane establishment delay (≈15 ms+2Ts1u)
	Link quality measurements and reporting the information:	(1+4+1+4 ms)
	Total delay:	200-278 ms
Data Transmission	Data transmission from MTCD1 to MTCD2:	(1 ms)
	Data processing at MTCD2:	(4 ms)
	Total delay:	5 ms

C. Data transmission through capillary network

Some of M2M applications have been already deployed utilizing low-range wireless communication technologies, such as IEEE 802.15.4, Bluetooth, and low-power Wi-Fi, benefiting from cheap radio transceivers. In contrast, the current LTE-based radios are more expensive, which hinder to utilize these radios in cheap Machines. An easy solution to incorporate the existing machines in the LTE system is the introduction of MTCGs. A gateway can form a capillary network and provide the connectivity for MTCDs. In this architecture which is shown in Fig.1(c), the MTCG exchange data locally between machines, while it provides the connectivity to the Internet by passing data through the LTE system. Most of the short-range technologies operate over unlicensed bands, which prevent wasting cellular radio resources for local data exchanges. In addition, the MTCG can aggregate data from multiple machines and deliver the aggregate data to the eNodeB which is an efficient solution for applications involved with a large number of machines generating small amount of data. The communication delay in capillary networks depends on the underlying short-range technology and employed communication protocol. The IEEE 802.15.4 standard is a promising technology for the low-cost wireless industrial automation and control systems. Among various communication protocols designed for this technology, WirelessHART provides secure, reliable and low-latency communications [10]. It benefits from the time division multiple access and channel hopping to prevent undesirable collisions and mitigate the interference. It also supports multi-hop communications, while data transmission over each link is occurred during 10 ms time slot.

In the WirelessHART, when a device joins the network, it can transmit data to other available devices without specific link establishment. However, the latency for this type of data transmission can be high and vary during the time, which is not desirable in most of the control systems. In order to have a low-delay communication between two devices, the originating device should send a request to the gateway indicating the destination device and the data traffic characteristics. The gateway allocates required resources for both devices and informs them [11]. When the communication session is established, the devices can start communicating with each other through the gateway. TABLE III shows the procedures of local link establishment and data transmissions, and the estimated delays for each step. Delays in this network depends on the network configuration, number of active devices, and the allocated radio resources for each device in the system. The delays are estimated with the assumptions that: radio resources are optimally provided for both machines, i.e., transmissions

TABLE III: Link establishment and data transmission latencies within capillary network.

	Procedure	Latency Estimates
Link Establishment	Transmission of link request connection by MTCD1:	(10+10 ms)
	Delay to access the next downlink timeslot:	Variable
	Sending a command to MTCD2:	(10+10 ms)
	Transmission of confirmation from MTCG to MTCD1:	(10+10 ms)
	Total delay:	minimum 60 ms
Data Transmission	Data transmission from MTCD1 to MTCG:	(10 ms)
	Data transmission from MTCG to MTCD2:	(10 ms)
	Total delay:	20 ms

occur in consecutive time slots, and machines can process the received data in the same time slot that the data is received.

III. Mapping different applications into network architectures

As stated in Section I, M2M communications enable wide variety of applications, supporting machine to server, and machine to machine communications [2]. This section describes three M2M applications (monitoring and automation, intelligent transportation and smart grids), requiring low latency communications between machine entities. The latency requirements are assessed and feasible network architectures for them are proposed.

A. Monitoring and Automation applications

M2M communications has a great potential to enable monitoring and automation both in home and industrial environments. In monitoring scenarios, distributed sensors monitor the environment and report abnormal conditions to prevent excessive damages. Automation systems are mostly involved with data exchanging between sensors and actuators. Benefiting from wireless technologies in these systems can significantly reduce the wiring costs and enhance flexibility. Sensor devices in automation systems are generally required to report with small amount of data periodically, and also react on special events and send alarm messages. The latency requirements vary according to applications, however, the data type can be categorized into two main groups: normal and time-critical data; while the latency requirements for the latter are usually tighter. The requirements for these two data types are as follows [12]:

- End-to-end delay in general monitoring and automation systems, like process automation, power plants must be below 200 ms;
- End-to-end delay in time-critical monitoring automation systems, like synchronized drives must be below 20 ms.

According to these requirements, it is evident that latency goals for general monitoring systems can be met by all the mentioned network architectures in the previous Section. However, the tight latency requirements for the time-critical applications are restricted to utilize only LTE-links, either through the network or direct links. The support of D2D in these applications is highly recommended since devices are usually in fixed positions and the qualities of direct links do not change frequently.

B. Intelligent Transportation systems

The safety in transportation systems can be improved by vehicle collision detection and avoidance systems. Vehicles equipped with radio transceivers can receive and transmit messages or warnings among other vehicles or, receive similar messages from roadside infrastructure. Vehicles involved in a collision can also report the accident and their locations to emergency departments for necessary actions. The latency requirements depend on the service types such as [13]:

- Road safety services need low end-to-end latency ranging from 20 to 100 ms;
- Efficient mobility services may tolerate latencies from 500 to 1000 ms.

Since the coverage area for this scenario could be large, relying on capillary networks would require high expenditures related to network rollout. Supporting D2D communications also seems unfeasible since the number of links can be high and require excessive efforts to establish, unless there are ways to support local broadcast, or the D2D link can be established in advance. The effective network solution seems to be transmitting all data through the LTE network.

C. Smart Grids

The current electric power systems still rely on technologies implemented many decades ago. The next-generation of such systems are usually referred to as smart grids which would integrate data communication networks in order to handle the systems efficiently. Such networks must be high-speed and reliable in order to transmit any alarm or critical event to control centers. Different networks, technologies and their requirements have been studied, as presented in [14], which include wide area networks, field area networks and home area networks. It is important to highlight that a smart grid system is very complex and requires the interconnection of different networks, connecting different units, such as electricity meters, electrical substations, remote terminals, etc. Nevertheless, an overall delay constrains for three major traffic categories (in terms of delay) can be identified as [14]:

- End-to-end delay for protection information must be less than 8 ms;
- End-to-end delay for monitor and control information must be less than 1 s;
- End-to-end delay for operation and maintenance information must be less than 10 s.

Interestingly, the end-to-end delay for protection information is too restrictive for all the network architectures presented in Section II, unless dedicated and permanent resources are reserved for this use with support of D2D links, in which the link establishment delays are reduced. For the other two categories, any of the presented architectures can meet the delay restrictions and therefore the selection of the most suitable architecture will depend on each specific deployment scenario. But based on the coverage extension, solutions based on cellular networks represent a better choice in terms of deployment costs.

IV. Conclusion

In order to support a variety of M2M applications with different system requirements, it is necessary to adopt novel mechanisms aiming to provide efficient data transmissions and meet the system requirements. This paper describes the latency performance of M2M link in three different network architectures. The proper network architecture can be selected in accordance to the latency constraints for the underlying application. In this paper, the latency analysis is obtained under the assumption of low traffic load from limited number of machines in the system. Further studies taking into account impacts of high traffic and the large number of machines are required. Also the study can be further enhanced by considering other parameters, such as power consumption, resource utilization, operational expense, and capital expenditure.

Acknowledgment

This research work has been partially supported by EIT ICT Labs under M2MRISE project.

References

[1] http://share.cisco.com/internet-of-things.html, accessed: 15/04/2014.
[2] 3GPP TS 22.368 V10.00, "Service requirements for machine-type-communications (MTC)," March 2010.
[3] K. Chen and S. Lien, "Machine-to-machine communications: technologies and challenges," *Ad Hoc Networks*, 2013.
[4] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, 2012.
[5] LOLA project, http://www.ict-lola.eu/, accessed: 15/04/2014.
[6] N. Nikaein and S. Krea, "Latency for real-time machine-to-machine communication in LTE-based system architecture," in *11th European Wireless Conference on Sustainable Wireless Technologies (European Wireless)*, pp. 1–6, 2011.
[7] S. Singh and K. Huang, "A robust M2M gateway for effective integration of capillary and 3GPP networks," in *IEEE 5th International Conference on Advanced Networks and Telecommunication Systems (ANTS)*, pp. 1–3, 2011.
[8] C. Bontu and E. Illidge, "DRX mechanism for power saving in LTE," *IEEE Communications Magazine*, pp. 48–55, 2009.
[9] 3GPP TR 25.912 V9.0.0, "Feasibility study for evolved universal terrestrial radio access (UTRA) and universal terrestrial radio access network (UTRAN)," March 2010.
[10] S. Han, X. Zhu, A. Mok, D. Chen, and M. Nixon, "Reliable and real-time communication in industrial wireless mesh networks," in *17th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, pp. 3–12, 2011.
[11] HART Communication foundation, "Network management specification HCFSPEC-085 revision 1.2," Released March 2009.
[12] IEC 62439, "High availability automation networks," Feb 2008.
[13] National Highway Traffic Safety Administration , CAMP Vehicle safety Communications, Vehicle Safety Communications Project, Task 3 Final Report, "Identify Intelligent Vehicle Safety Application Enabled by DSRC, DOT HS 809 859," DOT HS 809 859, Washington DC, March 2005.
[14] M. K. Wenye Wang, Yi Xu, "A survey on the communication architectures in smart grid," *Computer Networks 55, Elsevier*, pp. 3604–3629, 2011.