Machine to machine (M2M) communication impacts on mobile network capacity and behaviour

S. Počuča, D. Giljević
Nokia Siemens Networks, Zagreb, Croatia
sergej.pocuca@nsn.com, darko.giljevic@nsn.com

Abstract – Today there is a challenge how to handle all network aspects of M2M [1] communication. In this paper it is discussed how to smartly and effectively handle M2M communication in future networks. There are 2 main topics which should be considered. One would be network needs which should be prepared for M2M communications. The second one is that M2M devices need to become much more optimized for mobile communications and M2M applications.

I. INTRODUCTION

Smart objects will scale far beyond smart phones and smart devices [2] because over 1 Billion mobile based M2M subscribers are expected by 2015 (Figure 1).

Figure 1. Traffic forecast for GSM M2M Subscription (Source: NSN GSM traffic forecast, June 2011)

The number of transactions (short data calls) will be more than 10 folded in next few years (Figure 2) but the increase in total data traffic is expected to be marginal (Figure 3).

Figure 2. Global GSM transactions (Source: NSN GSM traffic forecast, June 2011)

Figure 3. Global GSM Data traffic by device type (Source: NSN GSM traffic forecast, June 2011)

Some applications transmit only the minimal amounts of data and only intermittently. Intermittent data updates can create a major signaling overhead on the mobile network. This has already been a problem for smart phone users. The biggest challenge is how to manage M2M network usage coming with exponential raise of objects (Figure 4).

Figure 4. Raise of “smart” devices is exponential

There are various M2M applications which will influence mobile traffic heavily [3], like:

- Smart meters (power, gas, water)
- Smart vehicles (road tolling, real time traffic info to vehicles)
- Vending machines
- Payment (unmanned gas stations, parking meters)
- Security alerting/reporting (in-car modems creating an emergency call in case of an accident)
- People tracking (kids, elderly)
These M2M applications have different requirements, e.g.

- some are stationary, some are mobile,
- some exchange small amount of data rarely, some have real-time requirements for data transfer,
- some communicate with one or more M2M servers, some with each other,
- some may have identifiers such as MSISDN, some may not.

It is important to understand the characteristics of M2M applications

- when selecting the right packet core features,
- when dimensioning the network,
- when developing protocol to communicate with end devices.

In section 2 challenges and impacts on mobile network and end users by introducing M2M object communications are described. Section 3 will give overview of traditional solutions to increase capacities of the mobile network but as well other theoretical solutions to optimize mobile networks for M2M communication in more appropriate way. In section 4, status and outlook is outlined to show what Telco vendors like Nokia Siemens Networks (NSN) have been done by developing software solution and features in radio network, what 3rd Generation Partnership Project (3GPP) has been standardized for future Packet Core networks in respect of M2M communication. Finally it is discussed how to develop and use new application protocol features in M2M application platform to communicate with end devices in more optimized way to save the network capacities when M2M traffic starts to grow heavily.

II. IMPACT OF HIGH M2M OBJECTS TRAFFIC TO A MOBILE NETWORK

A. Impact to radio network

Impacts of high M2M traffic to a radio network can be summarized as follows:

- huge increase in number of Packet Switched (PS) data calls – potential Control channel signaling channel (CCCH) congestion, and reduction of the network quality for other users sharing the same resources,
- transmission delays and lower throughput,
- M2M traffic is typically one directional, Uplink (UL) dominating, but due to acknowledgements in Downlink (DL) the transactions load will be in both directions and impacting all users,
- huge increase in SMS messages delivered in short intervals causing Stand-alone Dedicated Control Channel (SDCCH) signaling channel congestion,
- increase in paging load dominated by stationary smart M2M objects,
- high number of non-urgent M2M transactions in uplink reduce the network service quality,
- urgent M2M transactions need higher Quality of Service (QoS),
- short M2M data calls (e.g. acknowledgements) reduce the network service quality in downlink.

B. Impact to Packet Core

Impacts to a Packet Core network can be summarized as follows:

- User Equipment (UE) density can be manifold compared to Human to Human (H2H) devices,
- data transfer tends to happen upwards the network, as opposite to H2H communication,
- PS core dimensioning is impacted, depending on following factors: Number of M2M devices added in the network, M2M UE accessibility requirements from the network side and how frequently the M2M UE needs to send or receive data,
- when an M2M device needs to be reachable by the network it usually means it needs to be attached to the PS core system. For a LTE device, this is a must and for a 2G/3G device a Circuit Switching (CS) Short Message Service (SMS) can alternatively be used to trigger the device active,
- when an M2M device can autonomously decide when to establish a connection, it can do attach, send/receive data and then detach or it can stay attached all the time. M2M devices that do Attach/Detach upon need can share and thus save bearer capacity from the packet core, but they generate less signaling, particularly if they are stationary. M2M devices that do Attach/Detach upon need can share and thus save bearer capacity from the packet core, but they generate more signaling (transactions) in the packet core. Thus, optimization of the packet core resources for M2M communication means finding the right balance between bearer and transaction capacity. The balancing depends how often UE needs to exchange data with the network (Figure 5).

The mentioned figures are just rough estimates because other factors like mobility, latency and blocking tolerance requirements would also have impacts. It actually depends on the application requirements how the devices should communicate with the network. This means that the devices need to be much more flexible to use the network resources in a way tailored better for the application load. More accurate dimensioning is achieved.
by a creating traffic mix for each M2M application what could be reused as a basis for dimensioning.

C. Impact to end user experience

Impacts to an end user experience can be summarized as follows:

- Network could be busy which can follow to “missed call alerts”
- Lower data throughput
- SMS message not sent

III. SOLUTIONS

Traditional solutions would be to invest into end to end capacity to:

- add signaling channels,
- add traffic channels,
- expand GPRS territory,
- split the cells,
- split the Base transceiver station (BTS) sites,
- add transport capacity,
- add PCU capacity,
- add Gb interface capacity.

Such capacity extensions are cost consuming and increase Total Cost of Ownership (TCO), and more optimal solutions can be proposed.

In order to save money to operators and to prepare the network for M2M communication, it’s possible to implement changes in Radio network to optimize and reduce the load in the network.

In addition to preparation of network needs, the M2M devices need to become much more optimized for mobile communications and M2M applications.

IV. STATUS AND OUTLOOK

In this section it is described what NSN has implemented in the radio network and in Cumulocity M2M application platform[4][5][6], and what is to come including Packet core standard features proposed by 3GPP.

A. Radio network based solution

NSN proposal is based on introduction of following software solutions shown in Figure 6:

Precise Paging – It reduces paging load by sending paging first to the smart object’s last known BTS site. It can reduce M2M signaling load by up to 70%.

Priority Class based QoS - QoS profiles for smart objects according to the need minimize the impact to other users and maximize the speed for urgent transactions. M2M objects with low QoS class can increase data throughput for other users by 50%. High QoS class can be used for urgent health and safety transactions.

Smart Resource Adaptation - Short data calls need only one traffic channel (radio timeslot). It can bring up to 5 times more M2M transaction in downlink with existing network.

The described features can bring to significant reduction in TCO for operators.

In 3GPP Release 9 (R9) and earlier [7][8], for machine type communication nothing is defined and an M2M UE is just like any device. In 3GPP R10 [9] [10] an M2M UE can indicate low priority towards RAN/CN. In case of an overload, Mobility Management Entity (MME) and/or Serving Gateway (SGW) / Packet Gateway (PGW) do specific actions for low priority UEs, e.g.

- rejection of Mobility Management / Session Management requests,
- allocation of a long periodic tracking area update (TAU) timer in MME,
- sending a back-off timer from MME to UE which prevents signaling by UE until the back-off timer expires,
- adding low priority indication for charging in SGW/PGW.

The overload control (shown in Figure 7.) can be applied on network element level, or be specific for certain Access Point Names (APN) only.

B. 3GPP standard features for M2M devices

In 3GPP R11, several new items are being discussed currently:

- Architectural changes: An Interworking Function (IWF) is added between Machine Type
Communication (MTC) server and 3GPP system. MTC is 3GPP name for M2M.

- Device triggering: Non Access Stratum (NAS) signaling or Mobile Terminating - Short Message System (MT-SMS) could be selected. Offline device triggering (non-Attached devices) is postponed to Release 12.
- Identifications: System shall support both Mobile Subscriber ISDN Number (MSISDN) and MSISDN-less UEs.
- PS-Only: System should be able to work even without Mobile Switching Center(MSC)/Visitor Location Register (VLR) functions.

The R11 MTC standardization topics (Figure 8.) are expected to finalize in 2012. NSN is actively participating to the standardization work.

![Diagram of M2M Architectural Improvements in 3GPP R11](image)

**Figure 8.** M2M architectural improvements in 3GPP R11

### C. Solution in application platform layer for communication with M2M objects

In order to optimize the traffic in mobile network, the protocol between devices and application needs to be much more fault tolerant (if it runs on low priority, it has to robustly handle errors) thus NSN has designed the protocol for communication between devices and NSN M2M application platform - Cumulocity (Figure 9.) based on Hyper Text Transfer protocol (HTTP)/Representational state transfer (REST) paradigm and agent implementation. The platform manages M2M objects in a way that a device is just an HTTP client towards Cumulocity. Today many devices don't have it implemented but NSN sees many benefits of using HTTP and its extensions for this communication, and the fact that European Telecommunications Standards Institute (ETSI) is going to standardize it for M2M, is a good direction[11] [12].

**Benefits are:**
- Device doesn't need to keep connection/packet data protocol (PDP) context open to receive commands. It’s possible to send them a flash SMS for wake up and they'll retrieve operations that they need to carry out,
- no Virtual Private Network (VPN) infrastructure is needed in order to be able to connect to the device,
- dynamic IP addressing is fully supported,
- it gives good security, where no hardening of device is needed,
- it’s possible also to do auto-provisioning. (E.g., text credentials to the device and then it will upload its topology to the system,
- link between SIM card and tenant could come from connectivity platform of the operator (if present),
- Web based data compression between devices and Cumulocity can be used, e.g. Internet Engineering Task Force (IETF) Constrained Application Protocol (CoAP) [13] as protocol.

To interface with M2M devices and other M2M-related data sources from M2M Cumulocity platform, a driver software called agent is required. M2M devices come with a wide variety of protocols, parameters and network connectivity options. Protocols to devices range from low-level serial links to full-blown IT protocols such as web services. Today's M2M standards [14] rarely define exactly how to access particular readings of particular sensors or manipulate particular controls.

Devices may be connected through mobile networks and gateways. To shield machine-to-machine applications from this diversity, Cumulocity uses so-called agents (Figure 10). An agent is a function that fulfils different responsibilities for a given vendor and type of devices among which there is a function that enables secure remote communication in various network architectures.

![Diagram of Device Agent Responsibilities](image)

**Figure 10.** Device agent responsibilities

Managed agents are based on the Cumulocity run-time environment and are deployed and started by Cumulocity in the cloud. Such agents connect from the cloud to the devices in the sensor network using the device-specific protocol. This variant can be chosen whenever there is a secure and internet-enabled communication path to the devices – either directly supported by the device protocols or through a VPN infrastructure (Figure 11).
Non-managed agents run on hosts in the local sensor network. Such hosts could be mobile phones, gateways or other devices where the software is pre-installed and pre-configured. The agent software is based on whatever runtime environment the hosts support, hence it cannot be managed from the Cumulocity core (but it might be managed through other technologies like over-the-air programming).

![Diagram of supported agent architectures](image)

**Figure 11.** Supported agent architectures

Cumulocity uses a simple and secure reference protocol based on REST (i.e. HTTP) and JSON [15], which can be used from a wide variety of programming environments down to small embedded systems. To support near-real-time scenarios, the protocol is designed around a "push" model, i.e., data is sent as soon as it is available.

To provide a reliable M2M service, it’s necessary to ensure that the M2M devices are available to the M2M applications. This responsibility is split up in the following way:

- Cumulocity keeps track of agent availability.
- The agents keep track of the devices that they manage.

For tracking agent availability, Cumulocity records the time stamp of the last communication with the agent. If the last communication occurred more than a configurable interval ago, the agent is considered unavailable. If there is no actual need for a request between agent and platform, an agent may send regular alive messages to fulfill this interval. This mechanism supports also cases where bandwidth is critical and connectivity is unreliable. E.g., in smart metering, it is sometimes enough to get connectivity only once in a day. Avoiding that the connection is always open and under traffic, the cost of connectivity would be decreased.

V. CONCLUSION

Smart objects will scale far beyond smart phones and smart devices and operators need to prepare their network for upcoming traffic which is more than 10 times higher by number of transactions and in parallel to attract the application developers for the M2M ecosystem with M2M application platforms while it is in an early stage.

To save on network capacities and reduce the risks of coming M2M communication, operators can introduce new functionalities in radio and packet core networks to support optimal behavior of telecom network for M2M communication.

Benefits of introducing the described functionalities are improved customer satisfaction (e.g. data throughput and network accessibility), improved network quality (e.g. higher paging success rate and lower multi timeslot blocking) and more efficient use of existing network hardware resources which all leads to reduced Total Cost of Ownership for operators. In the end the network prepared for M2M communications is more attractive for application developers, which can benefit also with existence of NSN M2M Cumulocity platform which takes care and handle the interaction and communication with objects/machines.

ACKNOWLEDGMENT

The work presented in this paper was supported by Nokia Siemens Networks and NSN M2M team. Especially we would like to thank a colleague, Andre Eickler for his constructive hints and professional review of this paper.

REFERENCES

[7] 3GPP, M2M Study Report- 3GPP TR 22.868 V8.0.0, Rel. 8, March 2007
[8] 3GPP, M2M Security Aspects for Remote Provisioning and Subscription Change- 3GPP TR 33.812 V9.0.0, Rel. 9, December 2009
[9] 3GPP, MTC Service Requirements- 3GPP TS 22.368 V1.1.1 Rel. 10, November 2009
[10] 3GPP, System Improvements for Machine-Type Communications Release 10, 3GPP TR 25.888 V1.0.0 July 2010
[14] ETSI TS 102 689, Machine-to-Machine communications (M2M); M2M service requirements, Technical Specification, v1.1.1, August 2010