Evolution of the Resource Reservation Mechanisms for Machine Type Communication Over Mobile Broadband Evolved Packet Core Architecture

Marius Corici, Jens Fiedler, Thomas Magedanz, Dragos Vingarzan
Fraunhofer FOKUS Institute
Kaiserin Augusta Allee 31,
10589 Berlin, Germany
{marius-iulian.corici, jens.fiedler, thomas.magedanz, dragos.vingarzan}@fokus.fraunhofer.de

***Abstract*—With the deployment of novel wireless broadband technologies such as LTE and LTE-Advanced, new types of devices will be able to communicate remotely such as sensors and actuators, enabling the deployment of new services and applications generically named Machine-to-Machine (M2M) communication. However, the current core network architectures are designed only for Human-to-Human communication, lacking the fundamental support for the machine type communication patterns for functions like mobility and resource management, charging control etc. This article introduces a novel solution for mitigating with the basic support for resource management for M2M through wireless device based customization and grouping and introduction of cached based resource reservation and event notification mechanisms. Furthermore, the concept is exemplified on the 3GPP Evolved Packet Core architecture and evaluated as a testbed realization based on the Fraunhofer OpenEPC testbed.**

Resource Reservation, Machine Type Communication, M2M, 3GPP Evolved Packet Core, Bearer Model

I. Introduction

With the gradual deployment of new radio technologies such as LTE and LTE-Advanced, enabling increased capacity in specific areas of the network, the broadband communication becomes available for a massive number of devices.

Because of this increase in throughput and the delay reduction over the wireless environment, along with the new devices available such as smartphones and tablets, the users are gradually adopting the packet based broadband services expecting an exponential increase in their data traffic.

On the other side, due to the development of wireless devices able to interact with the physical environment such as sensors and actuators, new Machine-to-Machine services are foreseen enabling remote communication for other industries such as energy, automotive, transportation and logistics as well as for security, health and education. These new services are expected to increase the number of connected devices with at least two orders of magnitude in the next years [1].

The rapid deployment of the machine devices augmenting the acceptance of data services by mobile users will require a new level of scalability to be achieved in the subscription based handling including mobility and resource management.

The current network architecture is designed with the human communication in mind. Each device has a unique and independent subscription for which a specific subscription profile is maintained in the network. The system is prepared any time to reserve resources for the mobile device according to the information received from the services and applications and the information in the subscription profile. The actual reservation and release of the resources is synchronous with the session establishment and termination, as prior to that moment the network is not aware that the subscriber will initiate a service and on the level of resources required by that service.

With the evolution towards machine type communication, the new devices will multiply the same functionality i.e. they will communicate in the same way with the network entities. These multiple communications will be part of the same service e.g. fire alarm sensors in a building requiring that they will be handled in the same manner by the network while current solutions handle them independently.

Also machine communication has predictable resources which are required from the network at one moment in time. This is due to the limited number of services which are deployed in these devices e.g. a street video camera does not require time variable bandwidth for social networking services. Furthermore, in some scenarios, machine communication has predictable time intervals for data exchange e.g. initiation time, interval to next data exchange and duration of data exchange which cannot be assumed for any human communication.

In order to reduce the redundant network procedures related to the resource reservations for the machine type devices described, this paper proposes a new resource reservation concept based on functional grouping of devices and the predictability of the resources required in the mobile core networks bringing a new level of scalability.

The information structures and communication procedures related to the new concept are then exemplified using the 3GPP Evolved Packet Core (EPC) architecture, reference architecture for the future mobile communications. The concept here presented integrates in the EPC as an extension of the policy based control of the resource reservations addressing specific wireless devices.

The remainder of this paper is organized as follows: Section II provides the background of the proposed method while Section III describes the concept and an example evaluation of the solution. Section IV describes the exemplification of the concept on the 3GPP architecture followed by the description of the Fraunhofer OpenEPC testbed and in Section V conclusions are provided.

II. BACKGROUND

Current all-IP based architectures have as goal to offer a uniform and independent resource management for all the wireless connected devices in which the resources are reserved synchronously with the services establishment and at the level negotiated between the specific wireless device and the specific application as depicted in Figure 1.

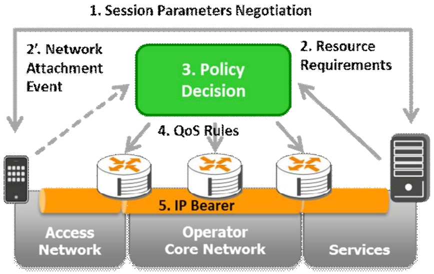

Figure 1 – Current Resource Reservation Concept

For each data session, the mobile device negotiates first with the service provider the session parameters (1). Then the resource requirements are transferred to a policy decision entity (2)[1][4][5]. For network attachment, the event is transmitted to the policy decision entity (2'). By using information from the subscription profile and the requirements from the services, a policy decision is made on resources that have to be reserved on the data path (3), decision which is send as QoS rules to the data path entities (4). The data path entities reserve the resources on the data path in the form of an IP bearer (5) spanning from the mobile device to the services [3], [8].

For machine type communication [9], the described mechanism is inadequate and causes several problems, especially in terms of flexibility and efficiency.

1. Each bearer is associated with a session, for which a service node needs to keep track of. When a wireless node is connected to the network, a full session establishment, including bearer reservation needs to be performed before the wireless device can exchange data with a service (Steps 1 to 5 from Figure 1).
2. The high number of connected devices causes an extreme overhead in bearer administration and in the core network. Each device has at least its own default bearer established when it is attached to the network.
3. A bearer has to be established and terminated due to an event. This event has to be propagated through the full signaling path. However, in case of machine communication, the communication of some of these events is redundant as the wireless device communicates during predetermined time intervals.

As example network architecture, the 3GPP Evolved Packet Core (EPC) was chosen, due to its capability of providing all-IP connectivity over LTE and other heterogeneous access networks. EPC includes the Policy and Charging Control (PCC) Architecture [7], responsible for performing resource reservation control i.e. gating and resource reservations. The part of the PCC which interacts with the core network elements and also with the services – generically named Application Function (AF) – is the Policy and Charging Rule Function (PCRF). The PCRF makes policy decisions based on the requirements from the applications or attachment events from the mobile device named User Equipment (UE).

The decision is transmitted as Policy and Charging Control (PCC) or QoS Rules to the data path entities which than install these rules in order to establish the communication *"bearer"*. For example, in case of LTE access, depicted in Figure 2, the entities include the UE itself, the LTE base station (eNodeB), the Serving Gateway (S-GW) and the Packet Data Network Gateway (PDN GW). These entities have to ensure that the data flows included in the specific bearer are receiving the required resources while forwarding the data packets. For communicating with the eNodeB on the control path a Mobility Management Entity (MME) is used.

The QoS rules are the main control information exchanged between the data path entities and the policy decision node. The PCC rules contain additionally charging information. They consist of one or more traffic flow templates (TFT) including the identification of the data flows by source and destination IP addresses and ports and transport protocol (e.g. TCP). Additionally they contain information on how the data flow has to be handled i.e. the QoS class classification, the aggregated bit rate and the guaranteed bit rate along with preemption priority, handling priority etc.

Each of the data path entities has to keep track of all the bearers assigned to each UE. When data packets are received, they are matched against the TFTs in order to determine how the data packet has to be further processed.

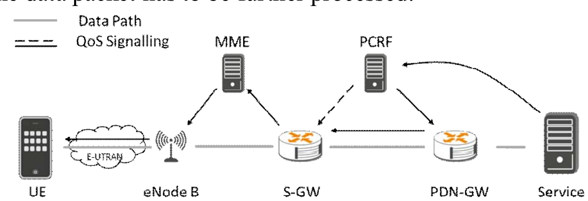

Figure 2 – Resource Reservation Signalling in EPC

When the UE attaches to the network, a default bearer is established allowing only a limited communication of the device. When the PCRF receives the attachment event, it makes a policy decision and installs the specified rules on the data path entities.

Similarly, when the UE needs to communicate with a specific service a default bearer is established. For this, the UE negotiates with the application the resources required. Then this information is transmitted to the PCRF which makes a policy based decision and installs the rules on the data path.

For the installation of the rules, the PCRF communicates directly with the PDN-GW and directly or indirectly with the S-GW, depending whether the data path control protocol between S-GW and PDN-GW. The S-GW pushes the rule down the data path via the MME to the eNodeB, which may also forward it to the UE. This process is depicted in Figure 2.

At the end of this procedure, the required resources are reserved on the complete data path for either the default or the dedicated bearers.

The concept here maps directly to the resource reservation communication mechanism and addresses the issue of extended stored information and redundant signaling for machine type communication.

III. CONCEPT

This article proposes a novel resource reservation concept in which the specific information stored in the entities and the communication between them is simplified and adapted to the machine type communication. Concept is depicted in Figure 3.

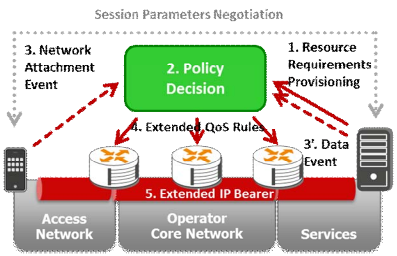

Figure 3 – Novel Resource Reservation Concept

A wireless connected machine device such as a sensor or an actuator has only a limited number of levels of communication. Compared to the human controlled devices, the machine device communicates the same type of information with the same required level of resources as its relationship with the physical environment does not change i.e. it will sense or modify the same physical features. Also, it usually communicates only with a single service platform acting as an application proxy.

From this perspective, the data flows have the same characteristic whenever the wireless device communicates e.g. same ports and same destination IP. For this reasons, there is no need for a negotiation of session parameters prior to the initial data exchange, reducing the signaling over the wireless core and the afferent required resources.

Instead the resource requirements are send by the service platform to the policy control entity at service provisioning (1). When multiple wireless devices are introduced in the network, independent of any session, the service platform transmits general rules on how resources have to be reserved for the specified devices including the information required for matching the different flows.

As this type of services are addressing only a specific set of devices, the resource requirements can be transmitted in the form of governance policies to the core network.

The information is processed by the policy decision entity (2) which makes specific internal policies based on the identity of the wireless connected devices. Because the policies are based on the device identity (and thus are not general working policies in the policy decision entity), they are affecting only specific devices, thus enabling classification and grouping allowing the parallel support for the current mechanisms.

Through this pre-provisioning of the resources required, the information and the policy decision is brought to the core network prior to any service establishment, thus reducing the duration of actual service establishment with the time interval for signaling between device and service and the one for signaling between the service and policy control entity.

This interval can be reduced if the information is transmitted also to the data path entities which establish the required data bearers. However, this implies pre-reservation of resources on the data path which increases the problem of scalability.

As the multiple devices of the same service share the same connectivity characteristics – e.g. maximum allowed bit rate etc. – they can share the same subscription profile. The policy decision made during resource requirements provisioning can be stored together with the subscription profile, thus not requiring novel data storage.

During the attachment procedure to the network of any device including the machine devices, the policy decision entity is notified (3). It makes a policy decision based on the information stored at the resource requirements provisioning and transmits the specific QoS rules to the data path entities (4) enabling the establishment of the IP bearer (5).

Also the device may require to remain connected in the network, with no communication bearer associated, waiting for a communication trigger from the physical environment or from the services. In this case, the specific bearers are installed for the communication based on a data event (3).

In order to be able to determine data events, the data path entities have to be provisioned by the policy decision entity with the information on which data traffic has to be monitored during the attachment of the wireless device to the network.

Also, the communication of some machine devices has a specific time characteristic i.e. is established at a specific time of day or after a specific time interval from the previous communication and has a given duration.

For these devices, QoS rules can be transmitted prior to any data path event (for pre-determined session initiation) or can contain a recurrence timer (for recurrent establishments). For the prior QoS policies, there is no need of any signaling during the communication related to resource reservation. For recurrent policies, data path entities may maintain information for multiple sessions and only notify the policy entity in case of exceptions e.g. devices did not connect as expected. In this case, the resource reservation procedures are executed at the first communication without any subsequent signaling.

For the devices for which the communication duration can be pre-determined, there is no need for communication termination signaling as the installed bearers may have a given validity duration. For all the other devices, the data path entities have to signal to the policy decision entity that the communication has ended and the policy decision entity has to remove the QoS rules. Compared to the current signaling, no signaling between wireless device and service and between service and core network is required for session termination.

IV. 3GPP EXEMPLIFICATION AND OPENEPC TESTBED

A. 3GPP EPC Exemplification

For the 3GPP Evolved Packet Core the concept is implementable directly as an extension of the Policy and Charging Control (PCC) architecture [7]. The concept is separated into two distinct and independent procedures: one for the provisioning of the service and one for the establishment and termination of the communication with the wireless device.

The service provisioning procedure, depicted in Figure 4, is novel to the EPC as currently the communication between core network and service platforms is supported only in relationship with a data session. However, it follows the same logical structure, in order to reduce the necessary modifications for the real deployment of the concept.

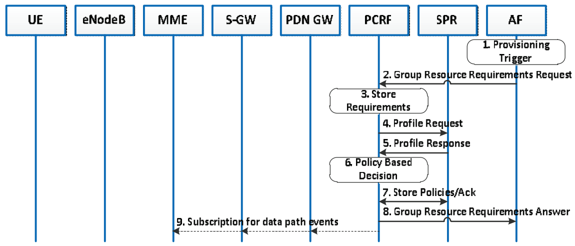

Figure 4 – Resource Reservation Provisioning

When an Application Function has new or updated resource requirements for some wireless devices (1), it sends to the PCRF a "Group Resource Requirements Request" which contains the service identification and the description of the resources which will be required by the UEs for data exchange with the specific application (2). This information is stored in the PCRF (3). The subscription profile for the group of devices is fetched from the Subscription Profile Repository (SPR) (4 and 5). The PCRF makes a policy decision on the resource levels that should be allocated to the wireless devices when they attach to the network and want to communicate with the given service (6). The policy decision may have multiple levels based on the network location and on the communication time of the wireless device. These policies are stored in the user profile in the SPR (7) to be used when data path events occur. The resource requirements processing is acknowledged by the PCRF to the AF (8).

Optionally, the PCRF may subscribe to data path events related to the devices for which provisioning was executed (9). This may be executed as part of the provisioning with a single signaling session for the complete group of wireless devices. However, for any new attachment the PCRF receives a notification, thus this step can be executed when devices attach to the network.

At the end of this procedure, the PCRF has all the information related to the resource requirements of the specific application for the specific group of devices. Because of this information, there is no need for further communication when the actual service is established.

For the same device, multiple service platforms may send requirements. Also session establishment requirements can be also received superseding the policies stored in the SPR. Because of these multiple sources of resource requirements which augment each other in terms of required resources and may be conflicting in some situations, the solution with a policy decision was chosen as the simple updating of the subscription profile may create multiple conflicts. Also the SPR is maintained specific to the operator domain, thus not opening an interface to the outside environment.

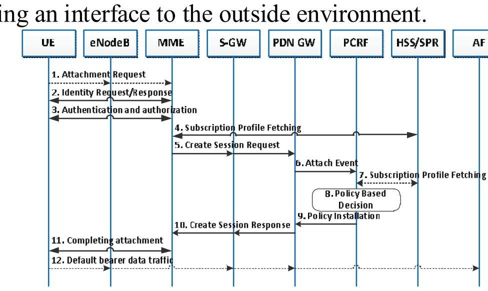

Figure 5 – Attachment Procedure

When the device attaches to the network it receives a default bearer considering the group subscription profile installed at provisioning phase and other dedicated bearers which it is known that the device will use for communication immediately after the attachment, as depicted in Figure 5.

When the device initiates the attachment (1) it is identified, authenticated and authorized to communicate over the wireless environment (2 and 3) and then the MME fetches the subscription profile (4) if not already available from other wireless device from the same group. Then a request for a session is send by the MME to the S-GW and PDN GW containing the required resources from the subscription profile (5) which is then forwarded as an attach event to the PCRF (6). The PCRF fetches the subscription profile (7) if not already available and makes a policy decision (8) which is then installed on the data path (9 and 10) before the attachment is completed (11). This allows the wireless device to have established all the bearers that it requires for communication (12) including a default bearer.

Using this procedure, a default bearer has to be established only for the first device from the group in the specific location. Other devices will use the same bearer for which the resources are augmented to fit the new user in without requiring any policy decision. This implies that the procedure for subsequent devices can skip the steps between 5 and 10 and by this reducing the attachment time. It is supposed that the devices in the same group will not behave in a competition manner for the resources.

The bearer establishment for a specific application is highly reduced compared with the state of the art by relying on the information already available in the system. In Figure 6, the dedicated bearer establishment upon uplink data traffic is depicted. For default bearer and downlink data traffic similar procedures are considered.

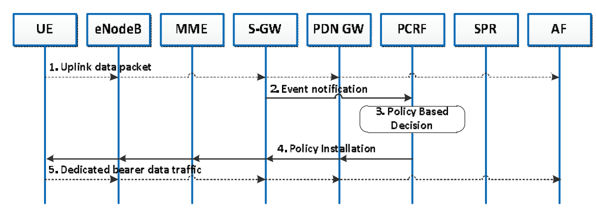

Figure 6 – Dedicated Bearer Resource Reservation Procedure

In this procedure it is assumed the UE will send one or more uplink data packets (1) to an application function using the default bearer as the only one allocated. An event notification is send by the S-GW or the PDN GW to the PCRF (2) which based on the group subscription profile makes a policy decision (3) which is then installed on the data path (4) using standard EPC mechanisms allowing the communication of the device over a dedicated bearer (5) without requiring any establishment signaling from the mobile device or from the service platform.

The termination of a dedicated bearer follows the same procedure based on event notification from the data path in case the termination time cannot be pre-determined or is implicit in all the entities when the duration of the communication was pre-determined.

The dedicated bearer may be allocated for all the devices in a specific area in case the group subscription profile includes the information that when one of the wireless devices needs to communicate, the rest will follow. In this way, a single set of rules is exchanged from the PCRF with the data path entities, thus reducing the necessary signaling through the core network.

B. OpenEPC Testbed

For testing and evolving the network architecture in a real testbed implementation, including the evolution of the bearer model concept presented in this paper Fraunhofer FOKUS developed the OpenEPC toolkit depicted in Figure 7 [10]. OpenEPC provides the complete resource reservation mechanism from the 3GPP EPC and PCC standards along with mobility management, security and subscriber based handling in both IPv4 and IPv6 environments.

A UE is able to connect to the OpenEPC platform using an LTE type of access in which the radio link is simulated using WiFi technology and the core network includes the GTP forwarding between an eNodeB stub and a S-GW and a control path through an MME. It may also connect to a public 3G network for which the data traffic is tunneled to a UMTS S-GW or a direct WiFi connection to an evolved Packed Data Gateway (ePDG) [6]. These ePDG and S-GWs are using GTP or PMIP protocol for the mobility management procedures with the PDN GW.

The S-GW, the ePDG and the PDN GW contain installation points for the PCC rules. Additionally the testbed includes a PCRF which is able to communicate with different application functions and is able to retrive and store the information on a Subscription Profile Repository (SPR) integrated with a Home Subscriber Server (HSS). The PCRF is able to receive requirements of the service platforms and events from the data path and to make policy decisions related to the resources that have to be reserve on the data path and to transmit these decisions to the data path gateways.

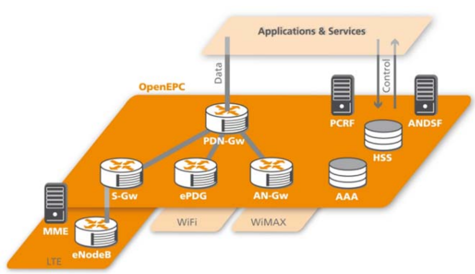

Figure 7 – Dedicated Bearer Resource Reservation Procedure

In the testbed, the PCRF functionality was extended with group functionality and pre-provisioning functionality as exposed in the previous concept section. The group based functionality was obtained using ranges of IP addresses e.g. all the devices on a group will receive an IP address from a specific prefix for which a single PCC and QoS rule is allocated on the data paths of the various devices.

However an in-depth testing of the scalability of the solution has to be further developed through a stress use case, as well as the implementation of the major exception cases such as UE using more resources than required.

V. Conclusions and Further Work

In this article we have presented resource management issues of machine type communication scenarios in wireless core networks exemplified on the 3GPP Evolved Packet Core (EPC). The enormous forecasted number of devices will challenge the limits of the existing mobile networks.

The new bearer model proposed in this article will address these issues by extending the bearer concept according to the expected communication pattern of connected machine type end devices, through the following mechanisms:

1. It is possible to pre-allocate bearer information from application side of the network, reducing the signaling each time the end device wants to communicate.
2. Bearers are valid not only for a single devices, but for a group of devices of same type and service. This reduces the number of bearers and enables new methods of resource optimization, as QoS traffic aggregation is now possible inside a group of end devices.
3. Bearers now are extended by a time function dimension. This means it is only valid at certain points in time. This leads to better resource exploitation inside the core network without requiring extensive signaling.

As an impact of the new bearer model, also the charging model needs to go under revision for the reason of flexibility and efficiency for M2M scenarios, considering both online and offline charging extensions.

The bearer model also has an impact on the session management outside the EPC, i.e. in the application function, which may be represented by a cloud or M2M service platform. For this a novel level of harmonization of presenting the information from sensors and broadcasting commands to the actuator devices has to be considered.

Further work in the area of resource reservation in the future mobile networks will aim at these different features and will provide adaptations to the EPC for M2M also in the mentioned areas.

References

[1] Cisco® Visual Networking Index (VNI), : Forecast and Methodology, 2010-2015, www.cisco.com;

[2] 3GPP TS 29.212, “Policy and charging control over Gx/Sd reference point”, www.3gpp.org;

[3] 3GPP TS 29.213, “Policy and Charging Control signalling flows and Quality of Service parameter mapping”, www.3gpp.org;

[4] 3GPP TS 29.214, “Policy and charging control over Rx reference point”, www.3gpp.org;

[5] 3GPP TS 29.215, “Policy and Charging Control (PCC) over S9 reference point; stage 3”, www.3gpp.org;

[6] 3GPP TS 23.402, “Architecture enhancements for non-3GPP accesses”, www.3gpp.org;

[7] 3GPP TS 23.203, “Policy and Charging Control Architecture”, version www.3gpp.org;

[8] 3GPP TS 23.207, “End-to-end Quality of Service (QoS) concept and architecture”, www.3gpp.org

[9] 3GPP TR 23.888, “System improvements for Machine Type Communications (MTC)” www.3gpp.org;

[10] Fraunhofer FOKUS OpenEPC toolkit, http://www.openepc.net/