3GPP Evolved Packet Core support for distributed mobility anchors

Control enhancements for GW relocation

Wolfgang Hahn
Corporate Research
Nokia Siemens Networks
wolfgang.hahn@nsn.com

***Abstract*—This paper discusses the 3GPP Evolved Packet Core (EPC) as a deployment of a distributed data plane architecture. Distributed GWs in connection with user mobility may result in the need to optimize data routing by GW relocation. It is proposed to include the GWs itself in the decision on GW relocation what is currently done in the control plane node (mobility management entity) only. Proposals are made to allow the relocation also for active mode devices (user activity detection), to detect the occurrence of a non optimal routing situation and to detect situations where a relocation should be suppressed to avoid particularly poor user experience. These improvements are compared with other solutions.**

Keywords - 3GPP; Gateway; Evolved Packet Core; distributed mobility anchors; architecture; LTE

I. Introduction

Operatos are faced with a huge data traffic increase in their networks. Especially for mobile networks forecasts show a high exponential grows for the next decade. This is the background to investigate new solutions that aim to manage the anticipated traffic demands in a cost efficient way.

A. Local GW support in 3GPP EPC

Motivated by the direct Internet access of Femto cells (Home eNBs) what provides an network offload both to the mobile network operator radio and core network 3GPP studied also the offload possibilities when accessing the operator (macro cell) radio network. This kind of offload provides breakout from the core network. In Rel.10 the feature Selective IP Traffic Offload (SIPTO) has been defined to allow the selection of Internet gateways (Packet Data Network GW or PGW) near to the radio access node the mobile device (UE) is attached. SIPTO is based on an enhanced GW selection that has the capability to select a GW depending on the location, see [1]. By this solution data traffic is forwarded on the shortest path out of the mobile operators NW and bypasses the operator's core and service network.

B. Distributed mobility managementdiscussions in IETF

There are ongoing attempts to provide mobility solutions with localized mobility anchors that are distributed over a NW topology in contrast to centralized anchors in a hierarchical model. Benefits of such architectures may arise from more optimal routing for local traffic and lower delay due to shorter distances to data sources like content servers. As an example dynamic mobility anchoring is presented in [2]. In this concept the UE is establishing new flows with new IP addresses that are assumed to be allocated optimally (local). Old flows are still anchored in the original anchor (acting as a home agent) using the old IP Address. This solution is compared below with the here presented one. The Mobility Extensions for IPv6 (MEXT) WG in the IETF has been chartered to work on the distributed mobility topic, see [3] and [4]. A target is to investigate the key concerns driving the need for a distributed mobility solution.

C. Tradeoff between local anchor service continuity and optimal routing/optimal offloading

The SIPTO feature in Release-10 didn't change any mobility support for data connections in general. So mobility can still be supported for "offloaded" traffic. But when the mobility anchor is located near the base station mobility of the user may lead to non optimal routing. So a main issue of localized mobility anchors is if mobility of localized traffic should be supported or not and if - how it can be avoided that non optimal routing situation will occur if the UE is moving away from the local anchor point. For SIPTO traffic 3GPP Rel. 10 introduced a solution that the MME can force a release of a (SIPTO) PDN connection indicating with a special release cause that the UE may reestablish the connection. The reestablished connection would then be anchored in a new (then again an optimal GW). So there is a tradeoff between supporting SIPTO session continuity and maintaining optimal SIPTO offload. The control has the choice to either select the PGW close to the UE to enable optimal SIPTO offload at the expense of mobility (if session continuity has to be stopped for PGW relocation) or to select P-GW deeper in the network to enable mobility at the expense of SIPTO offload efficiency.

II. Proposed enhancements for GW relocation control

In the following subsections three methods are proposed how the mobility anchor (in 3GPP networks the S/P-GW) itself can support and trigger it's relocation. It is described how this could be done for active mode devices (user activity detection), how to detect the occurrence of a non optimal routing situation and finally to detect situations where a relocation should be suppressed to avoid particularly poor user experience. These improvements are compared with other solutions.

A. GW Relocation in active mode (inactivity measurement)

The EPC SIPTO approach to force a release of a PDN connection has the disadvantage that the MME is not aware of UE activity if the UE is in active mode. The MME can only be sure not to interrupt ongoing data connections if it releases PDN connection during IDLE mode of the user devices. Hence a typical enforcement to use a new Serving GW and PDN GW (optimally combined in one node) would be implemented during a Tracking Area Update procedure (TAU) during IDLE mode of the UE when the UE has entered a new area of the radio network. In [1] this is explained “As a result of UE mobility (e.g. detected by the MME at TAU or SGSN at RAU), the MME may wish to redirect a PDN connection towards a different GW that is more appropriate for the UE's current location. When the MME decides upon the need for GW relocation, the MME deactivates the impacted PDN connections indicating "reactivation requested"…”
But especially the behavior of new smart-phones may prevent that the UE will enter the IDLE mode. Different running applications that make use of the “always on connectivity” e.g. for sending periodically reports for presence information or keep alive information keep the UE active.

This problem can be solved if the decision to relocate the PGW is shifted from the MME to the S/PGW or if the GW sends additional relevant information about the UE activity status of the SIPTO connections to the MME. In the latter case the MME has a final decision role or can override a PDN GW decision.
To decide for a release and reconnection procedure the PGW determines an optimal point in time by monitoring user inactivity phases: For PDN connection selected as not optimal routed (see next section) the PGW checks the UE activity (e.g. by setting an inactivity timer). If no traffic is carried in the connection (during the timer interval) the PGW starts the releases procedure for the connection. For this it supports a new release cause towards the MME that would trigger the MME to send the “PDN deactivation with reactivation requested” indication to the UE. (The MME has to perform a session management release cause mapping between GTP messages and NAS messages.) The MME may take into account own criteria to decide, if the optimization procedure should be carried out and in negative case reject the bearer deactivation procedure (with a new cause value) towards the PGW.

B. Relocation decision: Check for Non Optimal Routing

It is not stated in 3GPP standards how the MME shall detect if a non optimal traffic routing situation has established. This function can be taken over by the PGW (and optimally combined with the functions described in A and C).

Different options exist to detect (and release) non-optimal routed connections. As long as the a combined (local/distributed) S/PGW can serve the base stations the UE is connected to the situations can be regarded as optimal. If the UE is further roaming a new SGW may be selected and the GW splits into distinct SGW and PGW (connected via IP tunnel interface, S5). This is a trigger to check for a PGW relocation.

A solution is that the PGW is configured with SGW addresses for which routing is assumed to be optimal. In the case of a SGW relocation procedure (during a TAU or HO procedure) the PGW checks, if the new SGW belongs to the list of optimal SGWs. If not the PGW marks the PDN connection to be released.

A network example is shown in Figure 1. Here it is assumed that during UE mobility events always a local SGW is used by a SGW relocation procedure. This allows to select also a local PGW with optimal routing (probably collocated with the SGW) in case the UE will establish new PDN connections. In the following elaborations only the old - only one - PDN connection is considered. For GW selection issues with multiple PDN connections see [5].

During the movement of the UE several HOs occur that lead to SGW changes. A typical case can be mobile terminals used in a train. The UE is first attached at GW1 that provides PGW and SGW functionality for the UE. After the first SGW relocation GW1 provides the PGW only and GW2 the SGW. After each SGW relocation the PGW of GW1 performs the checks for the routing situation as described above. The PGW implements this procedure after receiving a “Modify Bearer Request” procedure from a new SGW (see SGW relocation in [1], section 5.5.1.1.3). If this check is done for GW4 (SGW4) the GW1 recognizes that the UEs PDN connection(s) are non-optimal. Due to this it checks for a time period of UE inactivity and starts the releases of the connection towards the MME. The PGW provides proper information (like a release cause) that again triggers the MME to release the PDN connection of the UE indicating (with a release cause in a NAS messages) that the UE shall re-establish the connection.

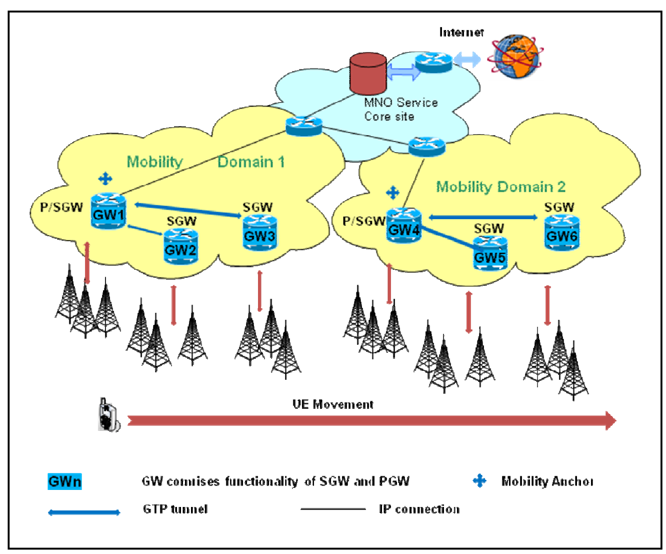

Figure 1. GW relocation between mobility domains

If the UE reestablishes the PDN connection the GW4 becomes both the PGW and SGW. The PGW4 is now the new mobility anchor that serves the UE more optimally in the new location/network domain.

Instead of managing “optimal SGW” lists per PGW alternative solutions would be to couple the “optimal mobility domain” decision with the IP routing topology of the NW. This would avoid managing the SGW lists per PGW for which mobility is supported and simplify the management of the feature in a “self-configuration” way. See the following examples:

- The local PGW can assume all SGWs in the same IP sub-network as optimal GWs with that mobility can be supported. For other SGWs routing is assumed to be non-optimal and a new PGW allocation should be enforced. This would then also take care of the additional delay which an additional router-hop would introduce.
- This schema can be also enhanced to “mobility optimal sub-networks”, e.g. that sub-NW addresses are configured in the GW for which mobility can be supported.
- In many cases mobile network GWs are containing IP router functionality. This can be used to determine how easily the SGW can be reached (the PGW may uses the number of routing hops to the SGW or the routing cost to the destination, information that can be delivered by the routing protocols).

C. Relocation decision: application awareness

When checking the PDN connections that are candidates to be released the PGW may include additional information in the decision about the content of the connection. Such information may be provided by traffic inspections functions (TDF/DPI) . The PGW may then take into account special operator policies how to distinguish different SIPTO connections, e.g.

- Exclude PDN connections with special content from being mobility/routing optimized (like secured VPN connections). For a small number of traffic non optimal routing may be acceptable for an operator compared with the poor user experience when loosing IP connectivity.
- Avoid interruption of ongoing sessions with dedicated QoS and policy control (even if no activity is recognised in these sessions). The gateway may check if the PDN connection contains dedicated bearers and/or sessions with a policy control and charging function (PCRF). This may be the case for special applications like voice over LTE. In this case the PDN GW may be not changed to avoid the need for a relocation or reestablishment of the context in the GW and PCRF. Also the user device would have to reregister its new IP address after a PGW relocation with the application function.
- Set special (longer) inactivity timers depending on the applications running in the PDN connection. The timers are used to check the activity level of the user before releasing (the check described in subsection A). This way an acceptable user experience can be achieved.

Only a PGW based implementation can take into account the used applications in the PDN connections. This could solve difficulties that can result from changing the IP addresses for certain applications. For those cases routing optimization may be suppressed.

D. Advantages of the proposed improvements

A PGW based GW relocation control provides the following advantages:

- Also for always active UE it is possible to force them to use optimal GWs (the MME would do this preferably in IDLE mode only).
- The UE gets an explicit trigger when to request a new PDN connection/IP address.
- The NW is in full control over the usage of local PDN connection and doesn’t depend on UE behavior or UE policies when to use a new/local IP address (which is the case in the dynamic mobility anchoring concept [2]).
- It is possible to force all traffic through one S/P GW (if the APN allows) regardless if the UE moves over a long time or not. This is an advantage over the dynamic mobility anchoring concept [2] what may generate NW overhead due to the need to maintain many parallel tunnel connections to many mobility anchors.
- Compared to the existing MME based solution there is no need have a core NW topology information in the MME to check for SGW relocations whether the new SGW is still in an acceptable routing domain for distributed PGW.
- Furthermore a PGW based implementation can take into account the user content in the PDN connections and cope better with issues that can result from changing the IP addresses for some applications.

III. CONCLUSION

Advantageous solutions are discussed how the Evolved Packet Core GWs can decide on a GW relocation or support the MME in such a decision. Fundamental information like user activity or the used services are only available in the data path and hence can be provided by the GW only. This allows mitigation of poor user experience that may result from traffic offload features when data path updates are made by GW relocation. In addition the GW can learn the network topology in a self configuration manner by using routing information and avoid additional management effort of providing topology information to the MME. The impact of this on standardization is low: Basically only new cause codes for PDN connection release messages have to be defined for the PGW/SGW to MME interface.

Nokia Siemens Network related research has received funding from European CELTIC MEVICO project.

ABBREVIATIONS

APN	Access Point Name
BS	Base Station
DPI	Deep Packet Inspection
EPC	Evolved Packet Core (in the EPS)
EPS	Evolved Packet system, LTE RAN and EPC
GW	Gateway, S/P-GW
SGW	Serving GW
PGW	PDN GW
LTE	Long Term Evolution
MME	Mobility Management Entity
PDN	Packet Data Network
SIPTO	Selective IP Traffic Offload
TAU	Tracking Area Update
TDF	Traffic Detection function
UE	User Equipment, mobile device

REFERENCES

[1] 3GPP TS 23.401: "GPRS Enhancements for E-UTRAN Access"

[2] Seite, P., Ed. and P. Bertin, "Dynamic Mobility Anchoring", draft-seite-netext-dma-00.txt , March 2010

[3] B. Patil, C. Williams, J. Korhonen “draft-patil-mext-dmm-approaches-00 - Approaches to Distributed mobility management using Mobile IPv6 and its extensions”, March 2011

[4] Dapeng Liu, Hidetoshi Yokota, Charles E. Perkins, Melia Telemaco, Pierrick Seite, H. Anthony Chan, Wassim Haddad, Hui Deng, Elena Demaria, Carlos Bernardos, Jun Song, “Distributed/Dynamic Mobility Management(DMM) - Problem Statement” IETF 79 Proceedings, Beijing, China November 07-12, 2010

[5] Wolfgang Hahn “Flat 3GPP Evolved Packet Core - Improvement for multiple network connections” submitted to WPMC 2011 - International Workshop on Mobility Management for Flat Networks, Brest, France, October 03 – 06, 2011