Technical Specification Group Services and System Aspects **_TSGS#10(00)0654_**
Meeting #10, Bangkok, Thailand, 11-14 December 2000

Source: **BT**
Title: **3G.IP Study on Architecture Enhancements Post R5**
Document for: **Discussion**
Agenda Item: **6.2**

3G.IP has had some study of techniques to improve the "IP friendliness" of the 3GPP core network, and has produced a comparison study which is published on the internet at http://www.3gip.org and is herewith copied to TSG-SA.

BT is not aware of a more suitable place to send this document for further discussion in 3GPP on Post-R5 issues, however BT believes that TSG-SA should try to identify such a place during TSG-SA#10.

TITLE:	**Post R5 Architecture Scenarios for the PS domain**
DATE:	**September 26, 2000**
SOURCE:	**3G.IP Post 2000 Ad Hoc Group**
ABSTRACT:	This draft document is proposed to be placed on the public pages of the 3G.IP server (according to the Munich meeting conclusions). It provides a brief description of the 3G.IP goals on post 2000 architecture and a brief description of the three long term scenarios under consideration, identifies advantages and disadvantages of each scenario individually, and provides a table comparing the scenarios according to a common set of criteria.
RECOMMENDATION:	Use this document for review on the 3G.IP exploder

Table of contents

1. Introduction

At its April meeting, 3G.IP agreed to establish an ad-hoc working group on the development of an all IP architecture for mobile networks beyond Release 2000. This includes objectives to define the high level direction and evolution of future releases of the all IP network beyond year 2000 and also to promote alignment between wireless and fixed IP technologies.

The goal of the work of this ad-hoc group is to define the target architecture for the future, without being constrained in the first instance by specific releases. In addition, the new work includes a goal to actively promote a common IP based wireless system for third generation mobile communications technology.

A major part of the work is the evaluation of various architecture proposals according to a number of specific criteria. This implies a need to review alternative architecture proposals, such as those within 3GPP2, and consider possible harmonisation options.

Among the proposals studied within this 3G.IP ad hoc group, two scenarios were considered as short term evolutions from 3GPP R2000 : SGSN Server and One Tunnel Approach. These two architectures are currently under consideration by 3GPP SA2 in a feasibility study [6]. Consequently they are not presented in this document. The following sections are focused on the long term view of UMTS core network, with three major scenarios : Mobile IP to the RNC, Edge Mobility Architecture (EMA), and a MGW/S&G-PSC architecture. It should be pointed out that no decisions have yet been taken within 3G.IP as to the preferred among these options.

In each scenario, the CSCF entity is a SIP server, as it is defined in 3GPP for multimedia call control.

Section 2 shows the basic architectural principles guiding the work within 3G.IP.

Sections 3 to 5 include the description of the long term scenarios so far identified by 3G.IP. These sections provide a brief description of the scenarios and their advantages and disadvantages.

Section 6 addresses a comparison table with a common set of criteria for all three scenarios.

Abbreviations and references are provided in sections 7 and 8.

2. 3G.IP framework for Post R2000 architectures

2.1. Scope of Work

The work on the evolution of the all-IP architecture beyond Release 00 should consider, but should not be limited to

- Evolution from the 3GPP release 00 starting point
- Alignment between IP solutions being adopted by fixed networks and mobile networks
- Harmonization with 3GPP2 IP architecture

- Further development of alternative access means

The work should also aim to maintain backwards compatibility with over the air PS domain control plane signalling so that Release 2000 terminals continue to work in the post Release 2000 networks. In addition, the following has been agreed by the management-steering group of 3G.IP:

- The 3GPP target architecture for Release 00 should be adopted as the initial baseline
- The focus of 3G.IP is the PS domain, however 3G.IP cannot ignore that there are CS domain issues.
- 3G.IP should not expend any effort on CS domain.
- 3G.IP should not let backward compatibility with CS domain terminals unduly constrain development of the PS domain.

2.2. Architecture principles

2.2.1. Evolution from the 3GPP release 00 starting point

It is important to have a smooth evolution from existing R00 terminals and infrastructure, and protect related investments taken. Major changes to the IM Subsystem have not been considered.

2.2.2. Alignment between IP solutions being adopted by fixed networks and mobile networks

This is a key area of alignment, both at network and application layer. It is important to derive common mobility requirements that could be adopted for wireless and fixed IP technologies, based on studies and review of mobility solutions on already existing mobile networks (i.e. the R00 releases) and fixed networks.

This is needed to identify properly mobility, security, multimedia call control & service control requirements before progressing the work on harmonized networks solutions for an all IP architecture post R00. It is also needed to ensure true seamless roaming and efficiency between wireless and fixed technologies.

2.2.3. Harmonization with 3GPP2 IP architecture

This is important for global harmonization of infrastructure, both for economies of scale for operators and for more benefits to subscribers.

2.2.4. Further development of alternative access means

Seek ability to add additional access networks in a way transparent to the upper layers of the architecture.

2.2.5. *Reduce cost of core network infrastructure*

This reduction is based on the use of off-the-shelf IP technology and economies of scale.

SGSNs and GGSNs currently have high intelligence levels which means that it is expensive to enhance capacity. If capacity enhancements can be made using off-the-shelf IP routers and the mobile-specific functions can be rolled out separately then we can provision our networks more cheaply

2.2.6. *Reduce cost of UTRA/EDGE access network infrastructure*

This reduction is based on the use of off-the-shelf IP technology.

2.2.7. *Scope of support for existing circuit-based services*

Services shall only be ported on to the all-IP network where there is a significant ongoing requirement to support these services into the future.

2.2.8. *Legacy network interworking*

Interworking with many different types of legacy network could become very complex. Interworking schemes should be harmonised to as great an extent as possible, and avoided completely unless there is significant market need.

3. Mobile IP to the RNC

In this scenario, mobile IP is used as a mobility management mechanism down to the level of the RNC, thus as a macro-mobility scheme with a relatively fine resolution. This is unlike offering mobile IP as a service, which is already possible with UMTS release 99.

The base mobile IP (MIP) protocol according to RFC 2002 [1] is assumed to be well known and is not described in detail in the following. This protocol provides mobility support for IPv4. Routing is typically triangular, mobile host terminated traffic is intercepted by the home agent (HA) and tunnelled to the care of address (COA), while mobile originated traffic is sent without encapsulation (using the mobile’s home IP address as source address) directly to the destination. If the COA has changed following movement, the old tunnel needs to be torn down and a new tunnel needs to be set up from the HA to the new COA.

There are several add-ons to the base protocol. These add-ons can be applied to avoid triangular routing and to improve handover performance [2]. To avoid having to send registrations back to the HA with every change of COA, the tunnel can be divided into 2 (or more) parts, a more static tunnel from the HA to a newly introduced gateway foreign agent (GFA), and a tunnel between GFA and COA [3]. As long as movement is contained to a region covered by the GFA, only registrations to the GFA are required upon change of COA (this is why this approach is referred to as regional registration in [3]). Where routers are configured to accept only packets with topologically correct source addresses (ingress filtering), reverse tunnelling from the mobile host to the HA must be applied [4].

For a number of reasons, mobility support in IPv6 looks quite different from that in IPv4, which will further complicate an already tricky migration from IPv4 to v6, unless MIPv6 is used as a starting point. Given the differences, the two cases are discussed separately in the following. Apart from the individual advantages and disadvantages listed below, an eventual choice between MIPv4 and v6 would also depend on "external" factors. For instance, a potential harmonisation between 3GPP and 3GPP2 would point to MIPv4, since this is the starting point in 3GPP2. On the other hand, the 3GPP IM subsystem is using IPv6, and it appears not to be advisable to carry IPv6 over MIPv4. Of course, 3GPP2 will need to solve the IPv6 problem too, and it could be that their solution is also appropriate for 3GPP.

3.1. MIPv4 as per RFC 2002 and Add-ons

Due to address space constraints, the care of address (COA) is typically a foreign agent (FA) COA rather than a co-located COA. Use of an FA COA has the additional advantage of avoiding tunnelling overhead over the air interface. The scenario we are looking at is therefore a home agent (HA) somewhere in the home domain (it is shown here in an ISP external to the PLMN, but it could be within the PLMN), FAs at the RNCs, and MIPv4 compliant mobile hosts (MH, also mobile equipment, ME). The following picture shows this scenario, where it is assumed that an SGSN server is required for control traffic. For MIP compliant mobile nodes, the GGSN is essentially a border router.

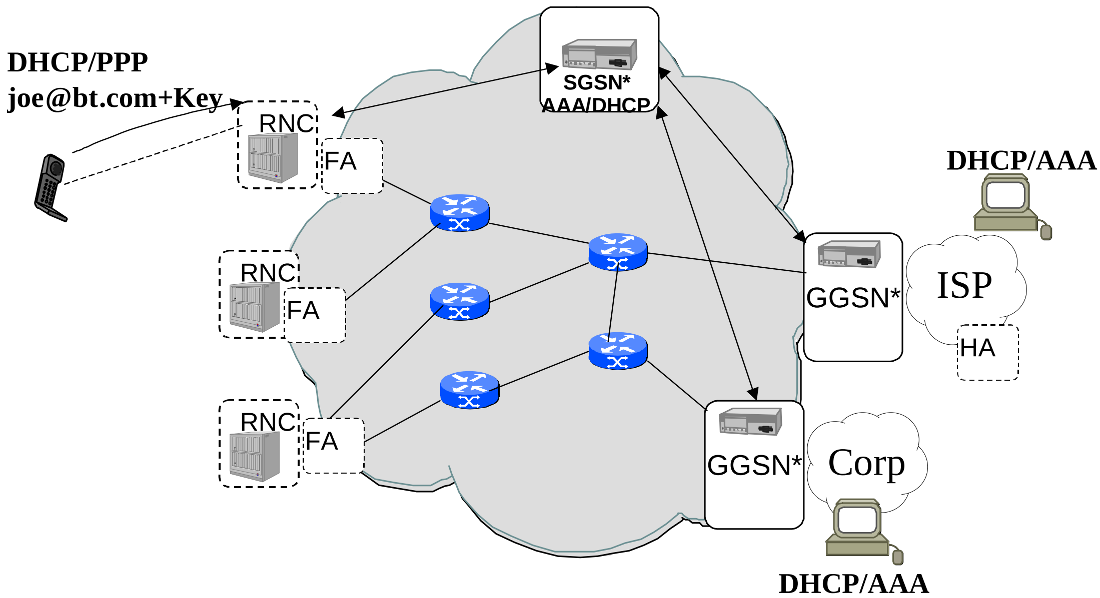

Fig. 1. Mobility Support with Mobile IPv4.

3.1.1.1. Advantages

1. Base protocol already around in IETF for a few years, increased momentum now partly due to adoption of mobile IP by cdma2000 proponents.
2. In case of FA COA, no encapsulation overhead over the air interface (unless reverse tunnelling is applied).

3.1.1.2. Disadvantages

1. Triangular routing (this is not only inefficient, it may also seriously affect voice quality due to unequal delays on the two links, which may cause echo cancellers not to work properly).
2. Reverse tunnelling over HA, if the IP source address needs to be part of the RNC address range.
3. Mobility results in tearing down and setting up of tunnels originating at the HA, which is potentially far away, thus adds handover latency due to signalling delays. This can be mitigated through introduction of GFAs, as outlined above. Other reasons that have an impact on handover performance include tradeoffs to be made between the frequency of agent advertisements (currently limited to 1/s) and resulting overhead, particularly if these need to be sent on expensive broadcast or pilot channels. Furthermore, in the case of co-located COA, allocation of new COA through DHCP is slow.
4. Route optimisation is an add-on to the mobile IP protocol and can only be used if correspondent hosts (CH) have this add-on implemented.
5. Additional overhead due to encapsulation in the fixed network and also over the air interface, if reverse tunnelling with encapsulating delivery style from MH to FA is applied (this delivery style must be selected for broadcast and multicast packets [4]).
6. QoS:
    - If DiffServ is used, HA needs to copy TOS/traffic class field of incoming packets into outer header;
    - If IntServ is used, we have to solve the problem arising from the fact that the tunnel between the HA and COA hides the RSVP control messages inside the tunnel, such that intermediate routers can not perform QoS related processing. Furthermore, change of COA requires RSVP session re-establishment. ([9], section 11.8.1.3, deals with this issue in more detail.)
7. Security issues (e.g. problem of key distribution when wanting to apply route optimisation, as security associations are required between the HA and correspondent hosts).
8. Location confidentiality: Even if no route optimisation is applied, the correspondent host might be able to detect the (logical) location of the FA (see [9], section 8.6.2.4).

3.2. Mobile IPv6 specific issues

Mobility is considered to be an inherent feature of IPv6, and so is route optimisation. A detailed description of mobility support in IPv6 can be found in [5]. Due to the abundance of IPv6 addresses and autoconfiguration features, it was not deemed to be necessary to introduce foreign agents, and therefore, the scenario looked at is as shown in Fig. 1 without FAs. Unfortunately, this results in additional messages and header overhead to be carried over the air interface. At least the header overhead could in theory be compressed away (fully, if routing headers are used, probably only partially, if IP in IP encapsulation is used, need to wait for the output of the IETF ROHC WG for details) [1].

[1] This is provided that no IP-layer encryption is applied end-to-end.

Triangular routing is assumed to be less a problem than in IPv4, since all IPv6 clients should be able to perform route optimisation. Therefore disadvantage number 1 of MIPv4 is not listed here again and disadvantage 3 applies only partially, thanks to improved handover performance. Should there be circumstances where security constraints may hinder route optimisation (which is FFS), these disadvantages would also apply to MIPv6. Since mobile hosts can insert the COA instead of the home IP address as source address in the respective IP header field (the home address would then be put into the home address destination option), problems with ingress filtering can be avoided. However, if location confidentiality is required and thus the COA cannot be used, reverse tunnelling has to be performed if ingress filtering is applied.

3.2.1.1. Advantages

1. Mobility is an inherent feature in IPv6.
2. Route optimisation to avoid triangular routing is an inherent feature in IPv6, and works faster than in IPv4, as binding updates are generated by the mobile node, not by the home agent.
3. MIPv6 includes features to minimise loss of in-flight packets during handover (similar features are included in MIPv4 route optimisation).
4. No foreign agents need to be deployed.

3.2.1.2. Disadvantages

1. Reverse tunnelling over HA, if ingress filtering is performed and location confidentiality is required (as far as the latter is concerned, disadvantage 8 of MIPv4 is also an issue here).
2. The handover delay performance may be insufficient.
3. Encapsulation overhead (or routing headers, if route optimisation applied) carried over the air interface, needs to be compressed away.
4. Unlike MIPv4, binding updates for route optimisation are carried over the air interface (on the plus side, they can be piggybacked onto traffic packets, so better from a fixed network perspective than MIPv4).
5. QoS:
   - If DiffServ is used, HA needs to copy TOS/traffic class field of incoming packets into outer header. This problem seems to disappear in case of route optimisation, since routing headers are used instead of encapsulation.
   - If IntServ is used, and routing is over the HA, the problem of hidden RSVP control messages arises as in MIPv4. This problem seems to disappear in case of route optimisation, although, according to [9], “there is a mismatch in the addressing information in the RSVP control messages and in the IP header which causes routing problems. This can be resolved as long as the RSVP layer at both the CH and ME are aware of the MEs COA”. Again, change of COA requires RSVP session re-establishment. ([9], section 11.8.1.3, deals with this issue in more detail.)

3.3. Summarising Comments

At first glance, introducing Mobile IP as a macro-mobility protocol for UMTS looks like replacing one tunnelling protocol with another for the sake of an IETF badge. Depending on the applicability of route optimisation, however, MIPv6 may get away with some of the routing inefficiencies of GTP. Running MIP to the RNC would also take the SGSN and the GGSN out of the data-path (unless we want to use the GGSN, which we need to keep for backward compatibility anyway, in the data-path as some kind of border router). However, the HA represents a new bottleneck/single point of failure dealing with both control traffic and user data (albeit the latter only temporarily, when route optimisation is applied). What SGSN and GGSN control functionality would have to remain and where it would have to be situated is yet to be established (there could be an SGSN server, for instance), and migration to such a solution is certainly not straightforward.

It may also make sense to move to MIP for the sake of convergence (i.e. use of MIP as a common macro-mobility management scheme for different access technologies), and harmonisation with the 3GPP2 architecture. As far as the latter is concerned, note that MIP terminates currently at the PDSN in 3GPP2, and not at the equivalent of the RNC, as proposed here. Furthermore, 3GPP2 use MIPv4, whereas MIPv6 may be the more obvious solution for 3GPP, given that the IM subsystem is based on IPv6.

Issues of particular concern are QoS, overhead over the air interface, and migration from GTP to MIP. There may be security issues as well.

4. Edge Mobility Architecture – Mobility Enhanced Routing to the RNC

The Edge Mobility Architecture is presented in [10] [11]and [13]. The scenario we are looking at is depicted in the following figure:

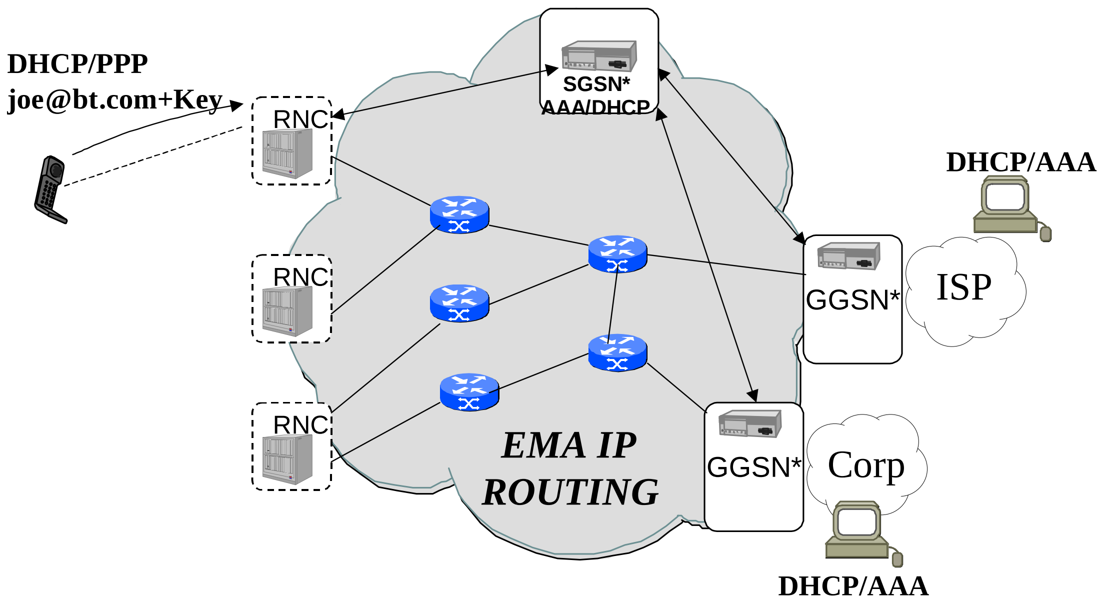

Fig. 2. Mobility Support through Edge Mobility Architecture.

The SGSN is a server dealing with signalling only and for EMA compliant mobile nodes, the GGSN is essentially simply a border router. The routing within the EMA domain is mobility enhanced. No specific assumptions on the exact routing protocol are made at this point other than that it is able to deal with mobility by

- allocating topologically correct addresses at session set-up; and
- injecting host routes into a *limited* set of "affected" routers when mobile terminals move away from the address-allocating RNC[2] (how limited depends on the mobility behaviour of hosts and the topology of the network).

Since EMA deals with optimising the "horizontal plane", mostly between access routers (here assumed to be RNCs), but does not affect the "vertical plane" (between mobile and correspondent host) beyond the RNC, it is possible to introduce EMA without affecting the terminals, that is, in a first stage, the control plane on the air interface could look exactly the same as in R2000. On the fixed network side, the PDP context could then be viewed as collapsed within the RNC/SGSN server. In addition, if the address is allocated from the operator domain space associated with the RNC then it is not necessary to create tunnels of any kind from the RNC to the SGSN and from there to the GGSN. Direct IP linkage is possible from the RNC to the internet or to other nodes such as voice gateways, etc. However, where address allocation is still necessary for third parties such as ISP's or corporate intranets, tunnelled access is required rather than the "direct access" enabled by EMA. In this case, it is possible to use the existing GPRS tunnelling protocol from the RNC to the GGSN, with EMA then managing the tunnel endpoint mobility at the RNC's, if this is found to be beneficial. From an evolution perspective, this approach is clearly advantageous, as it will for instance be possible to support standard roaming terminals in the visited network without any requirements being placed on the home network, irrespective of whether EMA "direct access" is provided in the visited network or the PDP address points to either home or visited GGSN.

A migration towards IP-based AAA mechanisms would provide further benefits. Finally, one could reuse MIP signalling messages and "horizontal MIP mechanisms" for EMA. In [13] and [12] it is not only shown how this could work, but also how MIP and EMA can coexist and provide mutual benefits, with MIP providing for inter-domain terminal mobility. Obviously, in this case, there would be an impact on the terminals and in roaming to other networks.

4.1.1.1. Advantages

1. Standard IP routers (running mobility enhanced routing) can be used for the user plane instead of MGWs (coexistence of this new protocol with existing intra-domain routing protocols is FFS).
2. Integration of back-end DHCP and AAA systems along with better integration with NAI based IP roaming using proxy AAA.
3. Services such as Multicast or Web-hosting can be run off the PLMN core.
4. Tunnelling not required, except for specific resilience situations during handover and for inter-domain mobility. Where tunnels are used, they can either be GTP or MIP tunnels.
5. Mobility transparent to the correspondent host (unlike MIP with route optimisation).
6. No encapsulation overhead.

---

[2] Strictly speaking, we are currently assuming that the SGSN allocates the address, but the prefix would point to a particular RNC.

7. Near-optimum routing[3].
8. DiffServ inherently supported. (If IntServ were to be used to provide QoS in the core, which we do not expect, then it would have to be enhanced to cater for route changes as a result of mobility).

4.1.1.2. Disadvantages

1. Topology constraints to enable scalability (a degree of hierarchy is required[4], there will only be a limited number of fully meshed core routers).
2. Mobility events will result in host route entries in certain routing tables and additional signalling traffic between routers.
3. Limited location confidentiality (dynamically assigned address reflects position at session set up).
4. Security concerns (due to opening up the core network), although these are no different from fixed ISP issues.

4.2. Summarising Comments

In terms of the fixed network infrastructure of the mobile network, this solution may be more demanding in terms of evolution than the other scenarios (although MIP to the RNC may be similarly complex). On the other hand, by tying EMA into the existing mechanisms in R2000, in particular using the same control plane over the air interface, an impact on terminals can be avoided. Further phases could then include a movement towards MIP-based mechanisms, to provide heterogeneous inter-domain mobility (e.g. between a home-network based on short-range wireless access and a wide-area cellular network).

EMA gets rid of mandatory tunnels, provides near-optimum routing (thus improves delay behaviour), and eases QoS management (at least as far as DiffServ is concerned). Cost reduction is possible thanks to reliance on “off-the-shelf” routers and reuse of skills and processes due to integration of back-end systems. Furthermore, the solution is completely transparent to the corresponding hosts (unlike MIP with route optimisation). Finally, the solution lends itself very well to provide fixed mobile convergence, within an EMA mobile-enhanced routing domain if applied on its own, inter-domain if applied together with MIP.

5. MG/S&G-PSC Architecture

5.1. Introduction

Section 5 presents a third scenario on a post-R00 architecture for the PS domain re-using the concepts developed in the IETF MEGACO working group to offer data and multimedia services. The main idea here has been to consider the work done on PSTN-IP interworking in the MEGACO WG and to adapt it for the Core Network to interconnect UMTS Terrestrial Radio Access Network (UTRAN) and externals networks to an IP CN.

---

[3] It is not optimum because of the hierarchy required in the topology.

[4] Note though that a certain degree of hierarchy exists normally in Internet designs anyway.

This solution relies on Media Gateways in the transfer plane (MG) and Packet Switched Controllers (PSC) in the control plane. The Serving PSC (S-PSC) is an evolution of multiple control functions of SGSN with H.248 interfaces. The Gateway PSC (G-PSC) can be viewed as an evolution of the control functions of GGSN, integrating the MGC for interconnection and H.248 interfaces. The functional content of these entities may be extended to support new interfaces/ functions (e.g. S-PSC interfaced to SIP proxy).

Signalling gateways (SG) will also be needed, where relevant, for conversion of transport protocols for signalling exchanges.

A more thorough study on mobility aspect (handover, roaming, etc.) is needed to identify the impacts on MG functionality and on the H.248 protocol used between the control plane (including PSC in this section) and the transfer plane.

5.2. Discussion

5.2.1. Major characteristics

The proposed post-R00 architecture has the following distinctive characteristics:

- Separation for transport resources, control (mobility management, call control) and service and application layers (e.g. Camel, APIs, etc.).
- S-PSC and G-PSC: among their functionalities, they handle macro-mobility management, Media Gateways control and session control.
- Signalling gateways (SG) to transfer the signalling from/to UTRAN to/from PSC and for the interconnection towards external networks.
- Media gateways (MG), handled by PSCs, to convert media formats from the radio part to the IP network and vice-versa. In a similar manner, MGs are needed between the CN and external networks.
- CSCF is in charge of multimedia call control (SIP server).
- Mobility management:

- Is split between the access network (micro-mobility[5] handling, i.e. soft handover) and the core network (macro-mobility[6] management, i.e. hard handover, SNRS Relocation)
- Is supported by the enhanced HSS functions in the core network.
- Is based on user identity, using IMSI (complementary identities may be optionally considered - e.g. NAI), in the signalling exchanges within the core network.
- Takes into account UMTS terminal states (connected, idle or detached) and using of the paging mechanism.

[5] This means local mobility towards the air interface management without any control done by the core network.

[6] Macro mobility triggers procedures in the core network to govern moves of UMTS terminals in the UTRAN.

- Provides seamless accesses to subscribers databases using protocol such as LDAP from the various entities in the core network (CSCF, PSCs, etc.).

Concerning transfer plane (between MGs), the proposed architecture is based on :

- IP protocol
- Common IP transport network for wireless and wireline subscribers
- Use of standard IP solutions (encapsulation, security, support of multicast, etc.)
- Some MGs with specific functions for the access networks guarantee the independence of access networks (possibility to connect various accesses) for the Core network.
- Optimum routing in the PS domain (e.g. local traffic) made possible – this is a necessary evolution from the R00.

The proposed architecture provides answers to 3G.IP requirements selected for the post-R00 study. Especially the following points are supported :

- MGs shall be common to different call servers (PSCs, MSC servers, etc.), providing a true common IP core network.
- The IP Core network is independent from the access networks. Multiple accesses should be considered such that UTRAN, GERAN, Hiperlan 2, etc. This can include wireline accesses in addition to the existing radio accesses.
- Route optimization is supported for all kind of traffics (local or not, intranets or Internet, etc.). Optimum routing requires evolution from the current distribution of functions in GPRS R'00. It can be achieved by the separation of the control and the transfer plane, which is enables a better distribution of functions in the control pane independently from the transfer plane.
- The operation and equipment cost of the network is reduced because of the use of a common backbone for all services. Moreover the splitting between transport /control /services allows a flexible network deployment, with greater independence between layers.
- The Core Network provides a global interconnection towards any external network. In using MGs to transcode flows, the Core Network is perfectly independent from the technologies used in external networks. MGs just need to be adapted to support interconnections with other networks.
- Guarantee of an optimal use of the UMTS radio interface for signalling and transport. This can be done in promoting "compact procedures".
- To maintain the core network control over the access resources when needed (i.e. registration procedures, hard handovers, …). Also simple and fast procedures (especially for the optimal use of radio interface) should be defined. A detailed analysis should be performed to list which functionalities are needed and where they should be located in order to improve the radio link efficiency (to minimize signalling between the terminal and the 3G core network).
- To insure backward compatibility with R00 packet domain terminals.

5.2.2. Overview

Next figure gives an overview of France Telecom proposed post-R00 architecture for PS domain.

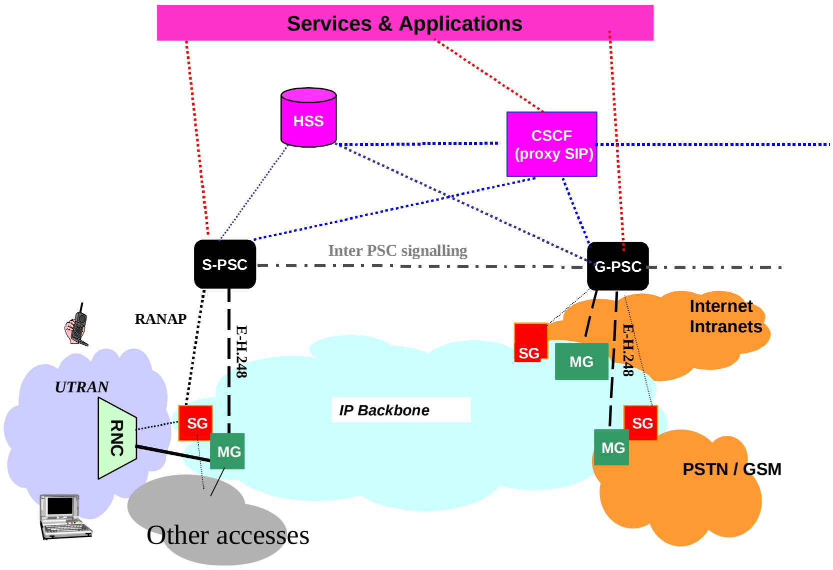

Fig. 4 : Post-R00 MG/S&G-PSC architecture for PS domain

Protocol selection is briefly presented in section 5.4.

5.2.3. Description of major equipment

The proposed architecture is mainly based on two families of equipment, the first one concerns servers in the control plane and second one is focused on MGs. This part details major functionalities for those equipment (S-PSC, G-PSC, MGs, SG).

S-PSC objectives is to administrate the transport layer (toward the control of MGs) and to realize the management of those transport resources in relation to services selected by terminals. S-PSC cover a lot of functions from the admission control of the network to the selection of an relevant QoS. They can be viewed (in relation to G-PSC) as a kind of « mapping interfaces » between services layer requirements and the optimal uses of transport resources.

Serving Packet Switched Controller

S-PSC must support:

- Admission control (UTRAN access, network attach, control of services triggering, etc.)
- Session management (all process to support services activation by terminal in the core network, QoS negotiation, etc.),
- Mobility management (inter-MGs, etc.),
- Charging and Billing (in relation to MGs which are controlled by this S-PSC),
- Lawful interception (LI),

- Control of MGs (resources management, to improve routings, etc.),
- Control of UTRAN (radio bearer, air interface encryption, etc.),
- Choice of G-PSC (based on APN required by terminals) for the support of terminals,
- Connection with services layers towards interfaces (i.e. API, Camel, …),
- Control of MGs (one or several) to support calls and to manage ressources on MGs (to order new ressources allocation for a call, to release ressources, …)
- Allocation of address (towards DHCP, Radius or …).

Gateway Packet Switched Controller

G-PSC work in strong relation to S-PSC. G-PSC are servers for control, mainly dedicated to the interconnection to other networks. The global control of the transport resources is achieved by the combination of both S-PSC and G-PSC.

Another main objective of G-PSC is to act as the controller of anchor points (physically represented in this scenario by MGs which are associated to a G-PSC) of communications. This is important for calls in case of mobility.

They play the same role as S-PSC concerning « interfaces » between services part and transport administration.

G-PSC provide following functions:

- Control of one or several MGs,
- Charging and Billing,
- routing management in the transport layer (MGs) in relation with S-PSC,
- Control of the interconnection to other networks (i.e. relation with Radius servers, etc.),
- Security control,
- Allocation of address (in relation with S-PSC),
- MGC functions (e.g. SIP to ISUP, etc.) to interconnect circuit switched networks
- Session management (shared processes with S-PSC). Both G-PSC and S-PSC must act together to support a call. For example this is important to provide optimal routing within the core network.

Media Gateways

First of all, one must note that functions and interfaces of MGs associated with G-PSC may be different from functions and interfaces of MGs with S-PSC. However even if they are not the same, MGs share some common characteristics. They always interconnect two transport medias which can be identical in some cases (i.e. 3G.IP core network and Internet) or different (i.e. 3G.IP core network and UTRAN based on ATM). MGs should be controlled by S-PSC and G-PSC at the same time in using some signallings between PSCs.

Several types of MGs have been identified :

- Between UTRAN and the 3G.IP core network.

- Between the 3G.IP core network and external networks (2G PLMN, PSTN, intranets, Internet, …).

Main functions of those MGs are :

- Routing in and outside the core network,
- Support of QoS mechanisms,
- Collection of Charging data (volume of data, loss of packets, …),
- Reservation and release of resources,
- Maintenance of resources,
- Ability to provide state of available resources,
- Mapping and transcoding different media for dissimilar networks (i.e. from IP to ATM and vice versa),

However some functions are specific to the type of MGs. For example :

- LI procedures for MGs which interconnect the UTRAN,
- Security process for those which realize interconnection with external networks.

There is a need to study those specific aspects. This is FFS.

Signalling Gateway

The Signalling Gateway (SG) receives/sends native signalling at the edge of the IP CN. In this architecture, the SG relays the UTRAN / external networks signalling to the PSC (interconnection to PSTN, public networks, Internet, intranets). It may be co-located with MGs. This SG should be transparent for the signalling protocol at the application level : it is involved in the conversion of the signalling transport protocols. The SG between the UTRAN and the S-PSC provides inter-working for the protocol stack over the Iu interface. It is used only for the transport of signalling and no inter-working is needed for the application protocols (e.g. RANAP).

Note : it was mentioned that an alternative option was that the SG could perform native protocol conversion. The comparison needs further study.

The protocols implementation of a SG is based on the networks it has to interconnect. For example SGs between UTRAN and Core network are not the same as the ones between the Core Network and a PSTN network.

It complies to the definition of SG in the Megaco architecture and it has been specifically introduced here for the UTRAN- S-PSC interface.

The evolution of the UTRAN and of the Iu interface has not been studied in details yet. We assumed that it was unchanged for the time being. If no inter-working is eventually required at the transport level over the Iu, it will be removed.

5.3. Open issues

Relevant technical aspects should be investigated such as :

- Improvement of QoS through IP backbone
- Growing of IP security on the 3G.IP transport backbone (to avoid spoofing, sniffing, flooding, …)
- Scalability of potential protocols
- Interconnection to IPv4 networks : which migration techniques should be implemented ?

5.4. Protocols

This section provides a list of the various interfaces and protocols for this architecture. The choice of protocols needs a deeper analysis of the requirements over the various interfaces and further discussion. This section intends to highlight the hotter topics :

- MG – MG (transfer plane): GTP-U, GRE, other IP in IP encapsulations, Mobile IP v6
- MG – PSC: H.248
- Between S-PSC and G-PSC : GTP-C, SIP/SDP, subset of BICC
- PSC – CSCF : SIP
- CSCF – HSS : LDAP or other protocol
- PSC – HSS : protocol(s) to be selected among existing protocols (MAP, LDAP, etc.) or new solutions (SIP mobility, etc.)
- PSC – Application & Service Layer : CAP
- CSCF – Application & Service layer : CAP, APIs (OSA), etc.

6. Comparison Table

Note: Concerning features 1 and 13, the scenarios are assessed against a few sub-headings to these features, which provide more details.

Feature	Mobile IP to RNC	Edge Mobility	MG / S&G-PSC
1. Evolution from the 3GPP R2000 starting point a) Complexity of changes from R2000	Considerable.	Considerable.	Medium.
b) Complexity of handling R2000 and post R2000 solutions in one network	FFS	FFS	FFS
c) Support of roaming	Between (post) R2000 networks: Both the roamer has to detect whether MIP is supported or not and the network has to be able to detect whether the terminal understands MIP or not. Backward compatibility required to support GGSN in home and SGSN in visited NW scenario. Provides also for roaming to non-UMTS/EGPRS networks using MIP (providing appropriate security/AAA mechanisms in place).	Without MIP, from a terminal perspective as per R2000. It also provides additional inter-access mobility within an EMA domain. Roamed-to EMA network must provide SGSN-equivalent to enable remote access over home GGSN in plain R2000 network. If used in conjunction with MIP, considerations provided for MIP apply. In particular, it does not matter whether a roamed-to domain provides only MIP or also mobile enhanced routing as per EMA.	As per R2000. *[Impact of GGSN split FFS.]*
d) Impact on terminals	TE must implement MIP stack. Implies macro- and intermediate MM now handled by TE rather than MT.	Potentially none other than the capability to recognise whether to use direct access or tunnelled access, in a first phase, as mobile enhanced routing transparent to terminal. Same as MIP if	FFS

		used in conjunction with MIP. Use of EMA direct access capability with legacy terminals is FFS, might require conventions on access point names.	
e) Industry acceptance and maturity of proposed protocols	Acceptance in parts of the cellular industry, but not yet mature.	Proposals have been discussed in a number of fora including IETF, although as yet there is limited industry acceptance and maturity.	Good - similar option used for CS domain R00 - vendors implementations in progress. Maturity issues relating to the protocol between server and MG.
2. Alignment with fixed IP solutions	Yes.	Yes, proposed architecture is not mobile cellular specific.	Yes, same concept of separation of transport and control planes.
3. Harmonisation with 3GPP2 architecture	Potentially, but there may be version conflicts due to adoption of MIPv4 in 3GPP2. Note also that MIP goes currently only to the PDSN and not to the RN in 3GPP2.	EMA may reuse elements of the MIP architecture, limiting differences to mobility management/ routing protocol.	No
4. Further development of alternative access means	Yes, thanks to alignment with fixed IP solutions.	Yes, thanks to alignment with fixed IP solutions.	Yes, due to core network independent from access technologies as much as possible.
5. Potential for cost reduction of core network infrastructure (potential for cost increase in other areas, such as lawful interception, FFS)	Yes, thanks to use of reasonably standard IP routers for the U-plane (assuming HA = standard router + software and, where required, FA = standard router + software).	Yes, thanks to use of reasonably standard IP routers for the U-plane (they need to run new routing protocols and cope with increased number of host routes).	Yes, using generic transfer plane equipment.
6. Cost reduction of UTRA/EDGE access network infrastructure	Not addressed.	Not addressed.	Not addressed.
7. [Not relevant in this context]			
8. Simplifying and harmonising interworking	No benefits w.r.t. R2000.	Interworking not required for certain fixed IP services (e.g. multicast, see 15)	Yes. Interworking limited to some entities in the control plane and

interworking		fixed IP services (e.g. multicast, see 15)	transport plane.
9. Separation of user data and control data streams	Only while route optimisation is applied, otherwise HA both in the control and user path.	Yes.	Yes at SGSN and GGSN through servers and MGs.
10. Handling of SGSN functionality	There may be an SGSN server similar to that in the EMA case. SGSN disappears from the user plane.	SGSN is now a server responsible for IP address allocation, radio authentication, IP authentication. SGSN disappears from the user plane (user plane runs on mobility-enhanced core routers).	Yes through PSCs. SGSN and GGSN are split into servers for the control plane and MGs for the user plane. The S-PSC contains current SGSN control functionality and runs the protocol to control the MG. The G-PSC contains current GGSN control functions and runs the protocol to manage MGs.
11. IP address allocation (only primary IP address, not COA)	Typically external to the GGSN, i.e. by the network owning the HA (address then points to this network).	SGSN (address out of the address space of the serving RNC at the time of session set up). IP address in external domain can also be allocated (GTP, MIP or policy routing required for this).	S&G-PSC.
12. Use of GTP-U	No.	Possible.	Possible, FFS.
13. Tunnelling, encapsulation, and route optimisation a) Avoidance of tunnelling	Partially through use of route optimisation, where possible.	In general, yes. Tunnels may be used temporarily between RNC’s (to forward packets in flight during handover) and for remote access. Where tunnels are used, these may be MIP or GTP tunnels.	. No.
b) Encapsulation	Required (IP in IP, optionally GRE and minimal encaps.), except for MIPv6 route optimisation (uses	Not required as long as no tunnels are used.	FFS.

	routing headers instead).		
c) Additional *payload-related* or U-plane overhead over the air interface (i.e. ignoring signalling messages for HO, etc.)	Due to encapsulation (e.g. if co-located COA) or routing headers (MIPv6 with route optimisation). Under certain conditions, (some) overhead can be compressed away.	None.	None.
d) User plane route optimisation	Through use of MIP route optimisation. This may not always be possible, particularly in MIPv4, where CH may not have implemented it.	Yes, thanks to native IP routing.	Yes - full route optimization is the target, mechanisms yet to be established
14. What happens on change of RNC?	HA or GFA (and CH in case of route optimisation) has to be updated with IP address of new RNC.	A few routers at the edge have routing tables updated. If GTP tunnel to GGSN is used with tunnel endpoint managed by EMA, GGSN not impacted by these mobility events.	Change of MG managed by S-PSC.
15. Transparency to IP multicast protocols	No (copies of multicast packets have to be sent down each tunnel individually).	Yes.	FFS
16. Lawful interception	Needs to be sorted out.	Need to enhance fixed ISP solutions to cater for mobility.	Done by MG + PSC.
17. Location confidentiality/ privacy	May require reverse tunnelling. Route optimisation not possible if confidentiality required.	IP address points to location at session set-up. Work-around desirable to provide location confidentiality without having to resort to tunnelling.	As per R2000.
18. Security	Issues such as key distribution for route optimisation.	Issues comparable to fixed ISP case.	Use of IP security mechanisms.
19. QoS	Some improvements (DiffServ and general delay performance) in case of route optimisation. IntServ tbd.	Delay minimisation thanks to near-opt. routing. DiffServ now straightforward, IntServ in core tbd, but unlikely application of IntServ.	Use of IP QoS mechanisms and improvements for routings.

20. Collection of statistics for charging	Statistics likely to be collected by RNC, GGSN/Border router and HA.	Statistics likely to be collected by RNC and GGSN/Border router.	MG and PSC and CSCF collect statistics.

7. Abbreviations

3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project No. 2
AAA	Authentication, Authorisation and Accounting
API	Application Programming Interface
COA	Care-of Address
CS	Circuit Switched
CSCF	Call State Control Function
EMA	Edge Mobility Architecture
FA	Foreign Agent
GFA	Gateway Foreign Agent
GGSN	Gateway GPRS Support Node
G-PSC	Gateway Packet Switched Controller
GTP	GPRS Tunnelling Protocol
HA	Home Agent
LI	Lawful Interception
IETF	Internet Engineering Task Force
IM	IP Multimedia
IP	Internet Protocol
ISP	Internet Service Provider
MG	Media Gateway
MGW	Media Gateway
MIP	Mobile IP
PS	Packet Switched
PSTN	Public Switched Telephony Network QoS Quality of Service
R00	(3GPP) Release 2000
RNC	Radio Network Controller
SGSN	Serving GPRS Support Node
S-PSC	Serving Packet Switched Controller
UTRA	UMTS Terrestrial Radio Access

8. References

[1] RFC 2002 (Standards Track), *IP Mobility Support*, October 1996.

[2] draft-ietf-mobileip-optim-09.txt, *Route Optimization in Mobile IP*, February 15, 2000.

[3] draft-ietf-mobileip-reg-tunnel-02.txt, *Mobile IP Regional Registration*, March 6, 2000.

[4] RFC 2344 (Standards Track), *Reverse Tunnelling for Mobile IP*, May 1998.

[5] draft-ietf-mobileip-ipv6-12.txt, *Mobility Support in IPv6*, April 27, 2000.

[6] 3GPP TR 23.873, ver. 0.1.0, *Feasibility Study for Transport and Control Separation in the PS CN Domain*, August 2000.

[7] Tdoc S2-000888, Ericsson, *SGSN decomposition*, 3GPP TSG SA2 Meeting #13, Berlin, May 22 – 26, 2000.

[8] Tdoc S2-000635, Ericsson, *Split of the SGSN into a server and a MGW*, 3GPP S2 Drafting Meeting on R00 Architecture, Helsinki, Finland, April 10, 2000.

[9] 3GPP TR 23.923, ver. 1.2.0, *Combined GSM and MobileIP Mobility Handling in UMTS IP CN*, November 1999.

[10] A. O'Neill, G. Tsirtsis, S. Corson, *Edge Mobility Architecture*, IPCN2000 conference, Paris, May 24 – 26, 2000, provided for information only as 3G.IP 00648, 3G.IP Ystad, June 13 – 15, 2000.

[11] M. S. Corson, A. O'Neill, *An Approach to Fixed/Mobile Converged Routing*, University of Maryland, Institute for Systems Research Technical Report, TR-2000-5, March 2000, http://www.isr.umd.edu/TechReports/ISR/2000/TR_2000-5/TR_2000-5.phtml.

[12] draft-oneill-ema-mip-00.txt, *EMA enhanced Mobile IPv6/IPv4*, July 2000.

[13] draft-oneill-ema-02.txt, *Edge mobility architecture*, July 2000.