3GPP TR 23.829 V10.0.1 (2011-10)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) (Release 10)

Keywords
3GPP, IP Access, Internet, LIPA, SIPTO

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

Support of Local IP access for the Home (e)NodeB Subsystem and of Selected IP traffic offload for the Home (e)NodeB Subsystem and for the macro layer network is required in TS 22.220 [3] and TS 22.101 [2].

This Technical Report describes the following functionalities:

- Local IP access – LIPA – to residential/corporate local network for Home (e)NodeB Subsystem;
- Selected IP traffic offload – SIPTO – (e.g. Internet traffic) for Home (e)NodeB Subsystem;
- Selected IP traffic offload (e.g. Internet traffic, corporate traffic) for the macro network (3G and LTE only).

The report is intended to analyse the architectural aspects to achieve these objectives and to gather the technical content until it can be included in the relevant technical specifications.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 22.101: "Service principles".

[3] 3GPP TS 22.220: "Service requirements for Home NodeBs and Home eNodeBs".

[4] 3GPP TS 23.203: "Policy and charging control architecture".

[5] 3GPP TS 23.060: "General Packet Radio Service (GPRS); Service description; Stage 2".

[6] 3GPP TS 23.401: "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access".

[7] 3GPP TS 33.320: "3GPP Security aspect of Home NodeB and Home eNodeB".

[8] 3GPP TS 29.303: "Domain Name System Procedures; Stage 3".

[9] 3GPP TS 29.274: "Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] apply.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] apply.

4 Examples for Styles

4.1 Scenarios

4.1.1 LIPA scenarios

According to TS 22.220 [3], LIPA breakout is performed in the same residential/enterprise IP network. Figure 4.1.1.1 illustrates this breakout at a Local GW (L-GW) in the residential/enterprise IP network.

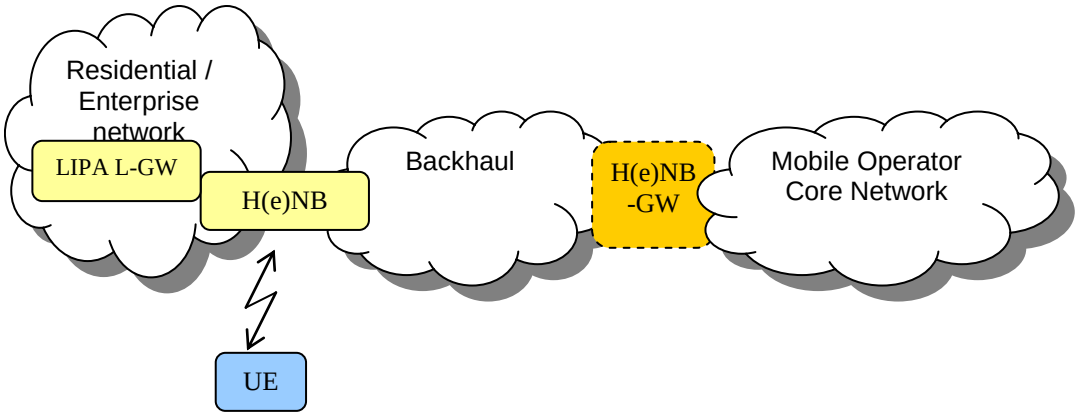

Figure 4.1.1.1: LIPA breakout in the residential/enterprise IP network

4.1.2 SIPTO scenarios

According to TS 22.101 [2], SIPTO for Macro-Cellular breakout point is close to the UE's point of attachment to the access network, and that it shall be possible to support mobility for offloaded traffic, which means that the breakout point is "at or above RAN". Moreover, SIPTO for H(e)NodeB Subsystem allows the breakout to be located either in the residential/enterprise network as LIPA, or "above" H(e)NodeB in the hierarchical view of the mobile operator network i.e. in the backhaul or at the H(e)NodeB-GW.

As a consequence, two types of breakout architectures are distinguished:

- Architectures with breakout "at or above RAN" (covering macro and some H(e)NodeB SIPTO scenarios);
- Architectures with breakout "in the residential/enterprise IP network" (covering LIPA and some H(e)NodeB SIPTO scenarios).

In addition, selected IP traffic offload for the Home (e)NodeB Subsystem may support the following three scenarios:

- Scenario 1: Home (e)NodeB Subsystem and backhaul are provided by the same operator;
- Scenario 2: Home (e)NodeB Subsystem and backhaul are provided by different operators;
- Scenario 3: Local Breakout point (L-PGW) for LIPA/SIPTO is located in a private address domain, e.g. behind a NAT gateway.

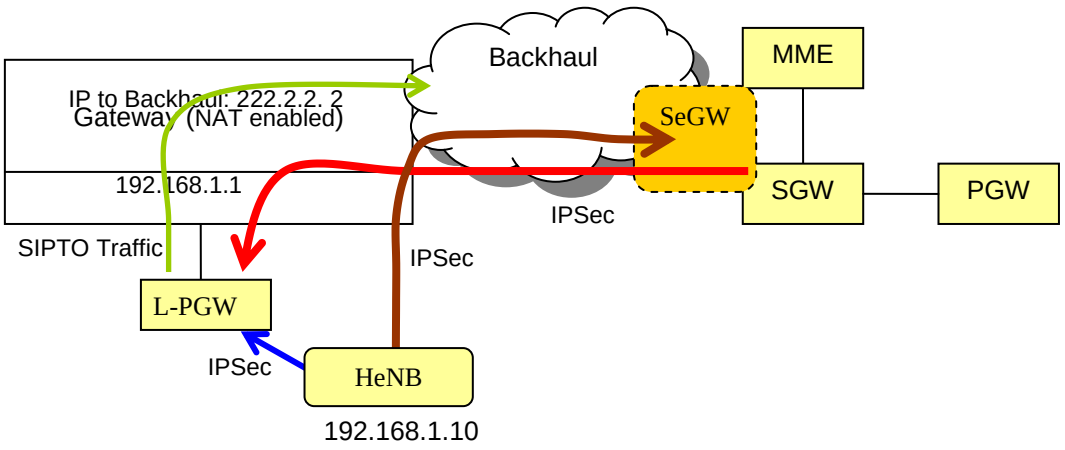

Figure 4.1.2.1: A potential deployment case for scenario 3 where the local breakout point is behind a NAT gateway

4.2 Key issues

4.2.1 Legal interception

Editor's note: This needs to be checked with SA WG3.

- Whether the Home (e)NodeB Subsystem provides Legal Intercept (LI) functionality for Local IP Access to the Home;
- Location of Legal Intercept (LI) functionality for Selected IP traffic offload for the Home (e)NodeB Subsystem;
- The legal interception requirements for LIPA, for SIPTO from H(e)NodeB Subsystem, and SIPTO from macro network can be different;
- Whether the Mobile Operator is in charge of legal interception or whether and how to assist the Backhaul Operator to perform legal intercept (e.g., by making the Mobile Operator's Core Network aware of the IP address assigned to the LIPA or SIPTO PDN connection).

4.2.2 QoS

- Whether QoS for LIPA and SIPTO traffic is based on static policies (no Gx to Home (e)NodeB).

4.2.3 Single/multiple PDN support

Multiple PDN support is not available in all UEs. The solutions have to consider the following cases:

- Single PDN support: Only one PDN connection is used;
- Multiple PDN support: Multiple PDN connections are used simultaneously.

4.2.4 Deployed behind NAT

The solutions for LIPA/SIPTO should consider the deployment scenario where the local breakout point (L-PGW) for LIPA/SIPTO is behind a NAT gateway.

4.2.5 Operator control for SIPTO

- Can SIPTO occur dynamically? or only statically?
- How do operators select the traffic to offload?

4.3 Architectural requirements

The solutions for local IP access and selected IP traffic offload for Home (e)NodeB Subsystem shall fulfil the service requirements described in TS 22.220 [3] in addition to the following requirements:

a) ability for the UE to request a LIPA PDN using:

- a well-defined APN; or
- a specific indication independent of the APN.

The solutions for selected IP traffic offload for the macro network shall fulfil the service requirements described in TS 22.101 [2].

The solutions for Selected IP Traffic Offload for macro (3G and LTE) shall fulfil the following architectural requirements:

- It shall be possible to perform traffic offload without user interaction.
- For UTRAN, the traffic offload shall be performed on or above the RNC node.
- The impact on the existing network entities and procedures by introducing traffic offload shall be minimized.

The H(e)NodeBs supporting LIPA shall be able to provide Intranet type access to the home based network.

NOTE: If the home based network provides a route to other private networks or to the public internet, then these networks may be accessible via LIPA.

The H(e)NodeBs supporting LIPA shall be able to provide access to the multicast groups that are active on the home based network:

- A H(e)NodeB supporting LIPA shall allow UEs to join multicast groups active on the home based network.
- It shall be possible for a H(e)NodeB supporting LIPA to forward multicast traffic from the home based network to the UE and from the UE to the home based network.

5 Architecture solutions

5.1 Architectural principles

The following architectural principles apply to all Local IP access and Selected IP traffic offload solutions:

- For traffic going through the mobile operator's Core Network, the SGW/SGSN User Plane functions are located within the Mobile Operator's Core Network;
- Mobility management signalling between the UE and the network is handled in the Mobile Operator's Core Network;
- Session management signalling (bearer setup, etc.) for LIPA, SIPTO traffic and traffic going through the mobile operator's Core Network terminates in the Mobile Operator's Core Network;
- Reselection of a UE's offload point for SIPTO traffic that is geographically/topologically close to the user shall be possible during idle mode mobility procedures.

Editor's Note: Mobility to non-3GPP accesses should be considered.

5.2 Solution 1 – Local IP Access and Selected IP Traffic Offload solution based on traffic breakout performed within H(e)NodeB using a local PDN connection

5.2.1 Applicability

This solution supports the following scenarios:

- Local IP access for HNB and HeNodeB Subsystem
- Selected IP traffic offload for HNB and HeNodeB Subsystem

This solution is applicable for breakout "in the residential/enterprise IP network".

5.2.2 Architectural principles

5.2.2.1 General principles

Common principles applying to both UMTS and EPS:

- Separate PDN connection(s) is assumed for traffic going through the mobile operator's Core Network;
- Pre-Rel-9 UEs that support Multiple PDN connections can simultaneously access LIPA, SIPTO and mobile operator's Core Network PDN connections;
- For LIPA traffic a Local P-GW function or Local GGSN function for EPS and UMTS, respectively is located within the residential/enterprise network;
- For traffic going through the mobile operator's Core Network, the P-GW/GGSN is located within the core network;
- LIPA PDN can be identified by a well-defined APN;
- Mobility management signalling between UE and network is handled in the core network;
- Session management signalling (Bearer setup, etc.) terminates in the core network;
- Before LIPA or SIPTO PDN connection is established, the UE is authenticated, authorized and registered by the core network;
- The paging function for LIPA/SIPTO traffic is located in the Core SGSN/MME;
- For active UE's, mechanisms to optimize the routing of the EPS/UMTS bearers used for LIPA traffic is to be studied, allowing the user plane to bypass the Core SGW and SGSN;
- The existing procedures for PDP context/PDN connectivity activation are used to establish LIPA with minor changes to determine if LIPA is enabled/disabled for the UE, to perform L-GW selection at the SGSN/MME, and to provide correlation information to enable the direct path between H(e)NodeB and L-GW;
- The existing procedures for PDP context/PDN connectivity deactivation can also be used to deactivate LIPA PDP context/PDN connectivity.

Additional principles applying to UMTS only:

- (none)

Additional principles applying to EPS only:

- (none)

5.2.2.2 Architectural functions

NOTE: Although this clause is EPC-oriented, the architectural functions respectively handled by P-GW and S-GW can be extended respectively to GGSN and SGSN in the case of GPRS core.

5.2.2.2.1 LIPA

P-GW functions for the support of LIPA services

They are a subset of the functions of the EPC PGW:

- per UE policy based packet filtering and rate policing/shaping;
- UE IP Address assignment;
- Direct Tunnelling between L-GW and RAN in connected mode.

These functions are included in a Local GW (L-GW) that is logically part of the Access Network (E-UTRAN or UTRAN). The L-GW for LIPA shall be located in the H(e)NodeB subsystem.

SGW functions for the support of LIPA services

It is FFS whether IDLE mode downlink packet buffering and initiation of network triggered service request procedure should be local to the H(e)NodeB, leading to two S-GWs per UE (one in Core Network and one in H(e)NodeB subsystem or transport backhaul network), which is not in line with current TS 23.401 [6] architecture principles, or whether this function should be in the Core Network.

MME impacts for the support of LIPA services:

The SGSN/MME supports the following ESM functions for LIPA:

- The H(e)NodeB provides the IP address of the L-GW to the SGSN/MME in UE-associated signalling. The SGSN/MME uses the information from the H(e)NodeB to potentially override the normal L-GW selection algorithm.

Editor's note: Alternatively, the L-GW selection is performed with enhancements to the DNS mechanism. It is FFS how this can be achieved and then, which of the two alternatives for L-GW selection should be preferred. This editor's note also applies to other text occurrences referring to L-GW selection.

- The granularity of LIPA control is per APN and per CSG. A new LIPA_enabled flag is defined for per APN and per CSG in the user's subscription, where the flag indicates whether LIPA is enabled/disabled for the user in that CSG.

It is FFS whether the MME may need adaptations to the EMM procedures regarding inter-HeNodeB mobility.

Indications to UE

If indications are required to the UE on whether the PDN connection for LIPA traffic can be initiated and/or on the APN to request, multiple solutions exist. The choice of the solution can be left as FFS.

- A list of CSG IDs or cell IDs statically configured in the UE/USIM, e.g. based on provisioning. This method best suits the case for residential LIPA access but may not be well suited for a corporate network for example if some of the H(e)NodeBs support LIPA and others do not. It also does not work for the hybrid cells where the UE is not a member and open cells.
- Informing the UE via NAS, i.e., the MME includes an indication in the NAS message towards the UE when it establishes a connection whether LIPA is supported at this cell. The UE can decide whether to initiate a LIPA PDN connection based on this indication. For example, the MME may be informed of the cell's capability to support LIPA when the UE establishes a connection in the initial UE message or the relocation request acknowledgement.
- Including the LIPA capability in RAN layer signalling. An indication of the LIPA support can be included in the SIB broadcasted by the RAN nodes or included in RRC signalling.

5.2.2.2.2 SIPTO for H(e)NodeB

P-GW functions for the support of H(e)NodeB SIPTO services

They are the same as LIPA case, plus:

- FFS: per user charging and inter-operator accounting.

These functions are included in a Local GW (L-GW). It is FFS whether the L-GW for H(e)NodeB SIPTO may be located in the H(e)NodeB or whether, mainly due to LI, charging and/or security reasons, it shall be located above H(e)NodeB.

Mobility-related functions are FFS.

SGW functions for the support of H(e)NodeB SIPTO services

Same as LIPA case.

MME impacts for the support of H(e)NodeB SIPTO services

Same as LIPA case.

HeNodeB to Core Network control interface for H(e)NodeB SIPTO services

Same as LIPA case.

5.2.2.3 Activation/Deactivation mechanism for LIPA and SIPTO

5.2.2.3.1 LIPA

The following is the summary on how to activate LIPA:

- The MME can select the L-GW close to or co-located with HeNodeB in initial attach or UE requested PDN connectivity establishment based on UE subscription, as described in clause 4.3 bullet a.

The following is the summary on how to deactivate LIPA:

- the UE can initiate PDN disconnection;
- the MME/L-GW can trigger deactivation of the PDN connection for the UE reusing the current mechanism in the EPS.

NOTE: It is FFS whether further mechanisms are needed to trigger the procedures to deactivate the LIPA connectivity after the UE has moved out of the HeNodeB.

The same principles as above apply for UMTS.

5.2.2.3.2 SIPTO

The following is the summary on how to activate SIPTO:

- The MME/SGSN can select the PGW/GGSN close to the RAN in initial attach or UE requested PDN connectivity establishment based on UE subscription information and the requested APN (any additional information used by the MME/SGSN to select the PGW/GGSN is FFS).

5.2.3 Architecture variants

5.2.3.1 Architecture variant 1 for LIPA

5.2.3.1.1 General

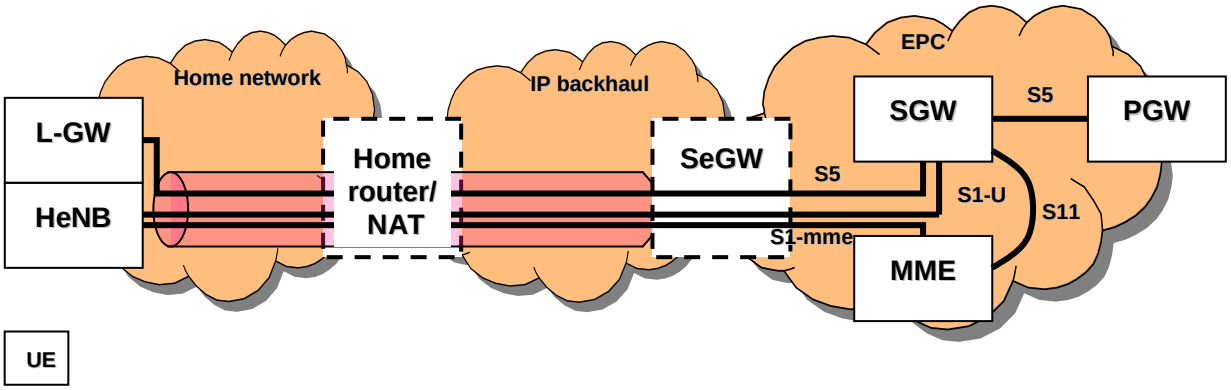

Figure 5.2.3.1.1.1: LIPA solution for HeNodeB using local PDN connection

The salient features of the architecture in Figure 5.2.3.1.1.1 are the following:

- a Local PDN Gateway (L-GW) function is collocated with the HeNodeB;
- the MME and SGW are located in the EPC;
- a Security Gateway (SeGW) node is located at the edge of the operator's core network; its role (according to TS 33.320 [7]) is to maintain a secure association with the HeNodeB across the IP backhaul network that is considered insecure;
- a Home router/NAT device is located at the boundary of the home-based IP network and the IP backhaul network, as typically found in DSL or cable access deployments today;
- for completeness also depicted is an external PDN Gateway (PGW) located in the operator's core network. It is used for access to the operator services;
- Paging of Idle mode UEs is triggered by sending the first downlink user packet or a "dummy" packet on S5. All other downlink user packets (or all user packets in case paging is performed with a "dummy" packet) are buffered in the L-GW. The Paging procedure is the same as in TS 23.401 [6]; when UE enters Connected mode, the packet buffered in SGW is forwarded on S1-U, whereas the packets buffered in the L-GW are forwarded to the HeNodeB on the direct path. In case the UE is paged by a "dummy" packet, the packet is discarded at the HeNodeB/L-GW;

Editor's note: It is FFS which of the two paging alternatives should be preferred.

- For mapping of the E-RAB IDs in the HeNodeB with the EPS Bearer IDs in the L-GW, the S5 PGW TEID (user plane) parameter is used as correlation information i.e. it is signalled across S1-MME to the HeNodeB. Candidate messages include INITIAL CONTEXT SETUP REQUEST or E-RAB SETUP REQUEST, etc. (refer to the call flows for illustration of messages carrying this parameter);
- with S5-PMIP the S5 PGW GRE parameter is used as correlation information;

Editor's note: support for S5-PMIP is FFS.

- S5 may be tunnelled in the same IPSec tunnel as S1-MME and S1-U or in a separate IPSec tunnel (depicted in Figure 5.2.3.1.1.1 is the case with one common IPSec tunnel);
- IKEv2 mechanisms are used to request one IP address each for the HeNodeB and the L-GW function. The assigned L-GW address is signalled to the MME via S1-MME in UE-associated signalling messages. The MME uses the information from the H(e)NodeB to override the normal L-GW selection algorithm, etc.

Editor's note: alternatively, the L-GW selection is performed with enhancements to the DNS mechanism. It is FFS how this can be achieved and then, which of the two alternatives for L-GW selection should be preferred. This editor's note also applies to other text occurrences referring to L-GW selection.

Depicted in Figure 5.2.3.1.1.2 is the equivalent LIPA architecture for HNB femto cells with S4-SGSN.

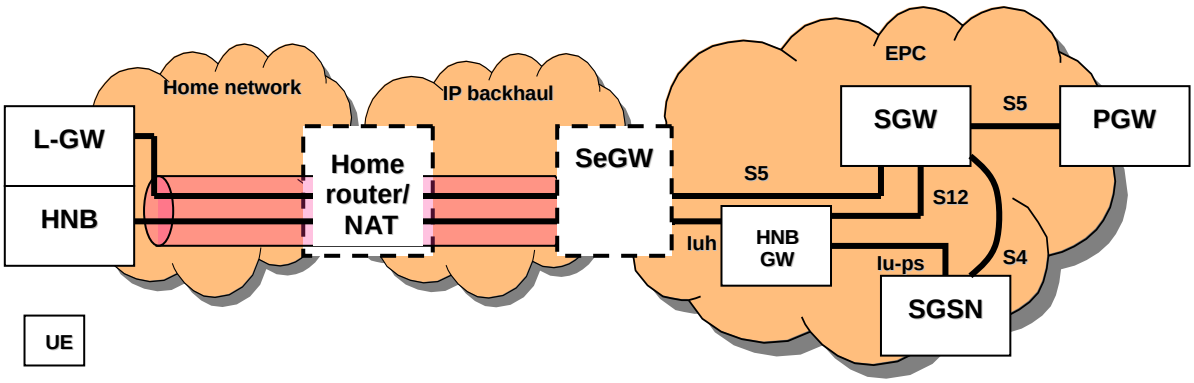

Figure 5.2.3.1.1.2: The equivalent LIPA solution for HNB using local PDN connection

The following is the summary of differences compared to the architecture for HeNodeB femto cells described in Figure 5.2.3.1.1.1:

- HeNodeB and MME replaced by HNB and SGSN, respectively;
- Presence of HNB GW; it is connected to the HNB, SGSN and SGW via Iuh, Iu-ps and S12, respectively;
- S11 replaced by S4.

The candidate protocol messages for this architecture are the following:

- The "Optimal Routing" information (S5 PGW TEID or S5 PGW GRE) may be carried in the RAB ASSIGNMENT REQUEST message (defined in RANAP);
- On Iu, the L-GW address and the S5 Protocol Type parameters may be carried in the INITIAL UE MESSAGE message (defined in RANAP);
- On Iuh, the L-GW address and the S5 Protocol Type parameters may be carried in the HNB REGISTER REQUEST message (defined in HNBAP) or in the UE REGISTER REQUEST message (defined in HNBAP).

5.2.3.1.2 LIPA PDN connection establishment

The following procedure illustrates the setup of LIPA PDN connection via the UE requested PDN connectivity request procedure. Similar changes would also apply to setup of LIPA PDN connection in the attach procedure.

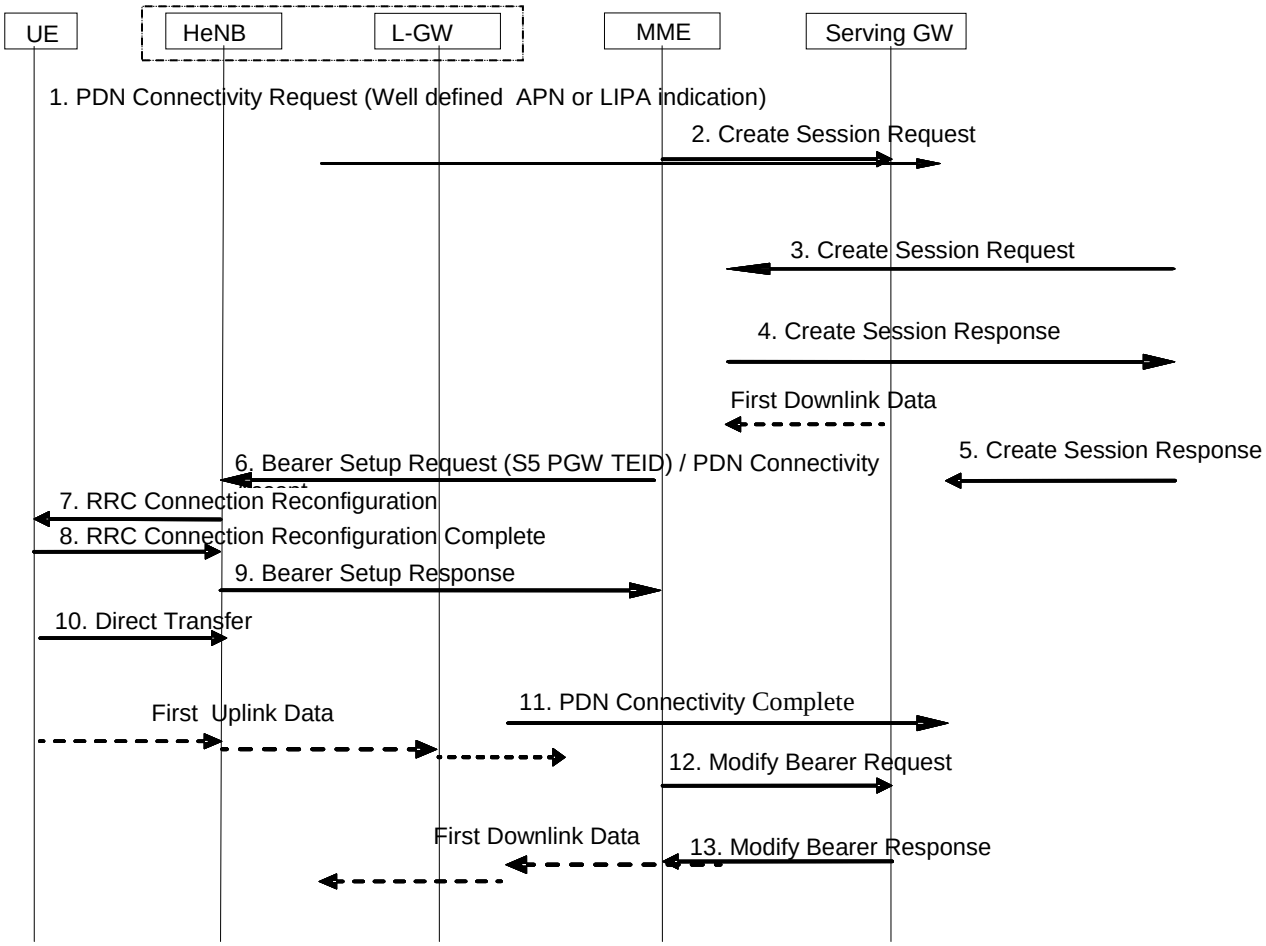

Figure 5.2.3.1.2-1: UE requested PDN connectivity to LIPA

In comparison with the existing call flow for UE requested PDN connection, the following steps are worth additional explanation:

1. UE initiates PDN connectivity request to establish PDN connection. A well defined APN or a special LIPA indication is included to indicate the desire of LIPA.

NOTE 1: It is FFS whether a well defined APN is enough.

The S1-AP message that carries the PDN connectivity request includes the following additional parameters:

- L-GW IP address assigned during establishment of the IPSec tunnel(s);
- H(e)NodeB capability to support LIPA.

MME performs LIPA authorisation of the UE to decide whether the UE is allowed to use LIPA function or not according to the UE subscription data and the LIPA capability of the HeNodeB. The LIPA subscription data may be per APN, per CSG or both. The MME rejects the PDN connectivity request if the LIPA authorisation fails.

After successful LIPA authorisation, the MME uses the L-GW address provided in S1-AP signalling to select the L-GW collocated with HeNodeB.

2. If there is a requirement to avoid IMSI storage in the L-GW (FFS), the MME omits the IMSI from the Create Session Request. The current condition in TS 29.274 [9] for not sending the IMSI ("If the UE is emergency attached and the UE is UICC-less") may need to be extended to cover LIPA.

6. The S1-AP message includes the S5 PGW TEID parameter assigned by the L-GW in step 5 for each E-RAB in the E-RAB to be Setup List.

5.2.3.1.3 Inter-HeNodeB mobility

Left intentionally empty.

Editor's note: Due to the presence of the non-optimised path (L-GW – SGW – HeNodeB) in parallel to the direct path, the architecture in Figure 5.2.3.1.1.1 can support mobility to another H(e)NodeB from the same home/enterprise network if, subsequent to the handover, it is acceptable to trombone the LIPA traffic via the operator's core network. It is FFS whether this is acceptable.

5.2.3.1.4 S1 Release procedure

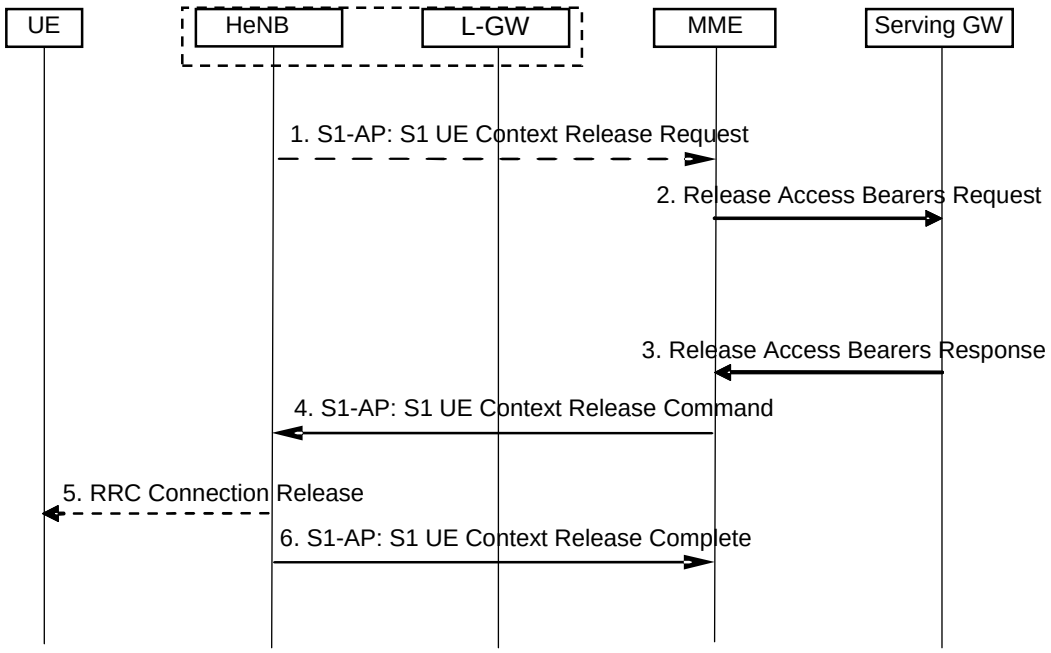

Figure 5.2.3.1.4.1: S1 Release procedure when LIPA PDN connection for UE exists

There is no impact on the S1 release procedure, except that in step 4, the HeNodeB informs the L-GW that UE is in Idle mode, so that L-GW enables the S5 path for paging.

5.2.3.1.5 UE Triggered Service Request procedure

Figure 5.2.3.1.5.1 shows the service request procedure when a LIPA PDN connection has been set up for the UE.

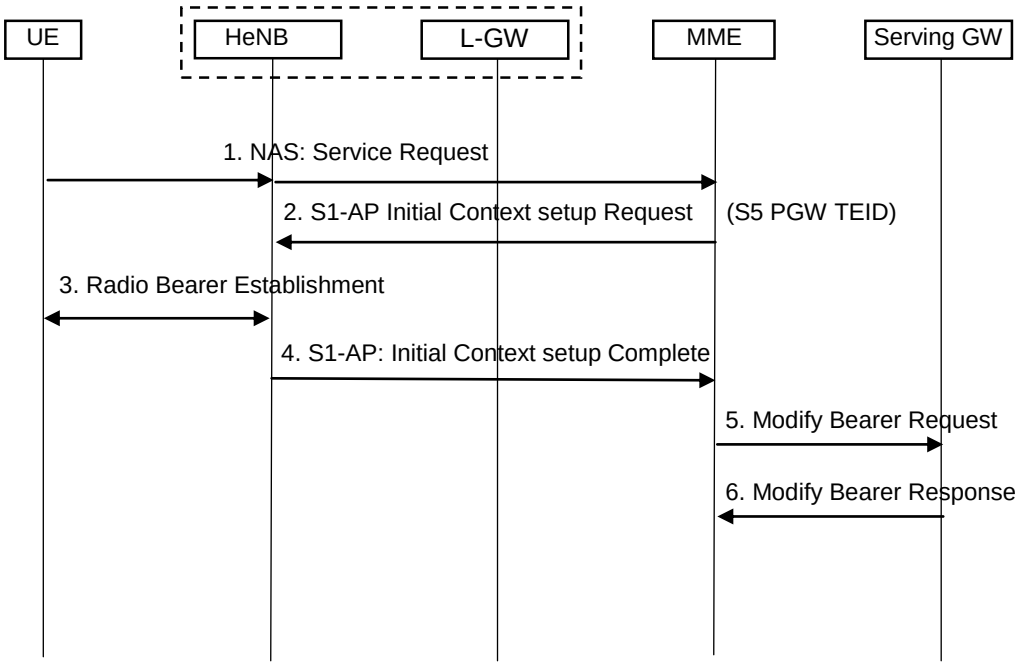

Figure 5.2.3.1.5.1: UE Triggered Service request procedure if LIPA PDN connection existed

When the UE initiates Service Request procedure to enter connected mode and there is an activated LIPA PDN connection, the MME includes the S5 PGW TEID for each E-RAB in the E-RABs to be Setup List in the S1-AP message.

5.2.3.1.6 L-GW Triggered Service Request procedure

Figure 5.2.3.1.6.1 shows the L-GW triggered service request procedure when a LIPA PDN connection has been set up for the UE.

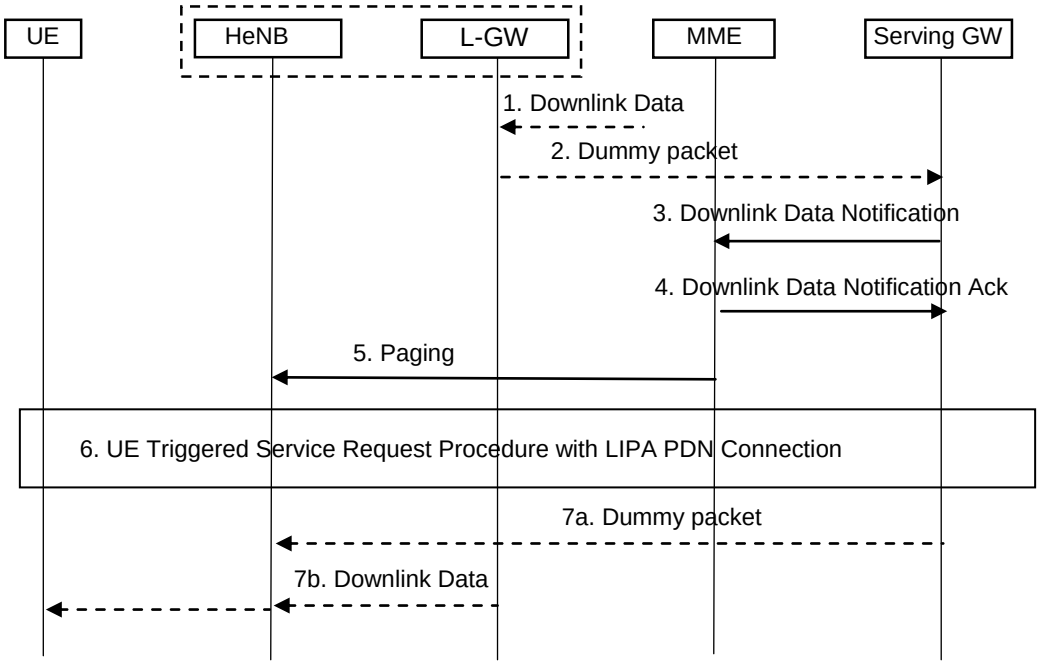

Figure 5.2.3.1.6.1: L-GW Triggered Service request procedure if LIPA PDN connection existed

In comparison with the existing call flow for Network initiated service request procedure in TS 23.401 [6] clause 5.3.4.3, the following steps are worth additional explanation:

1. Downlink packets arriving at the L-GW are buffered in the L-GW.

2. L-GW sends a packet or a "dummy" packet to the Serving GW in order to trigger paging.

Editor's note: It is FFS which of the two paging alternatives should be preferred.

7. Once the UE triggered Service Request procedure with LIPA PDN connection is complete, the Serving GW forwards the packet on S1-U. In the case of a "dummy" packet, the packet is intercepted by the HeNodeB and discarded (step 7a). In parallel, the downlink data that were buffered in the L-GW may start flowing on the direct path (step 7b).

5.2.3.1.7 Standards impacts

The following interface change is needed to support the user subscription enabling LIPA per CSG basis:

- *SGSN/MME/HSS (Gr/S6d/S6a): transmission of the flag from HSS to SGSN/MME:* adding the transmission of the flag from the HSS to the SGSN/MME over the Gr/S6d/S6a interface (Stage 3 only as this is included in the CSG subscription data).

The following functional changes need to be added to the standard to support this solution:

- Including the LIPA_enabled flag (per APN and per CSG) in the user's subscription data stored in the HSS/HLR and transferred to the MME/SGSN;

- (E-)RAB setup messages on Iu-ps/S1: addition of new correlation identifier (user plane L-GW TEID) for each E-RAB in the *E-RAB to be Setup List*;

- Adding the transmission of the IP address of the L-GW in UE-associated signalling on Iu-ps/S1;

- *SGSN/MME: L-GW selection*: algorithm for L-GW (GGSN/S-GW/P-GW) selection is enhanced to take in account the IP address of the L-GW received from the RAN node. Also need to account for the LIPA_enabled flag processing.

Editor's note: the last two bullets apply to the option with L-GW address signalled from the RAN. Alternatively, the L-GW selection is performed with enhancements to the DNS mechanism. It is FFS how this can be achieved and then, which of the two alternatives for L-GW selection should be preferred.

NOTE: For additional impact due to stand-alone L-GW refer to the conclusion section.

There is no impact on S1-U or S5.

There is no impact on the Serving Gateway behaviour.

5.2.3.2 Architecture variant 2 for LIPA

5.2.3.2.1 General

In the EPC architecture defined in TS 23.401 [6], the S11 interface between the MME and the S-GW is the interface that is used to manage the paging and mobility for the PDN connection. Since the S-GW User Plane functions reside in the core, one alternative is to bypass the S-GW for the user plane of the LIPA PDN connection and define a new interface for the control plane comprising a subset of the S11 interface between the L-GW in the HeNodeB Subsystem and the MME to manage the paging and mobility for the LIPA traffic. The same principles apply to the S4 interface.

We define the new interfaces as the L-S11 and L-S4 interfaces.

The figure 5.2.3.2.2.1 and the figure 5.2.3.2.3.1 show the LIPA architecture variants of the Home (e)NodeB subsystem for LTE and S4-UMTS respectively. In this variant, the L-GW can be either collocated with the H(e)NodeB or as a standalone node.

NOTE 1: The difference between residential and enterprise scenarios is that the L-GW needs to support less functionality including no mobility support or the S1-U interface to the HeNodeB.

NOTE 2: In case of stand-alone L-GW a separate IPSec tunnel with the Security GW might be required, whereas a collocated L-GW can reuse the existing IPSec tunnel established by the HeNodeB. The stand-alone L-GW could reuse the same procedures as the HeNodeB for establishing the IPSec tunnel.

5.2.3.2.2 LIPA Architecture for HeNodeB

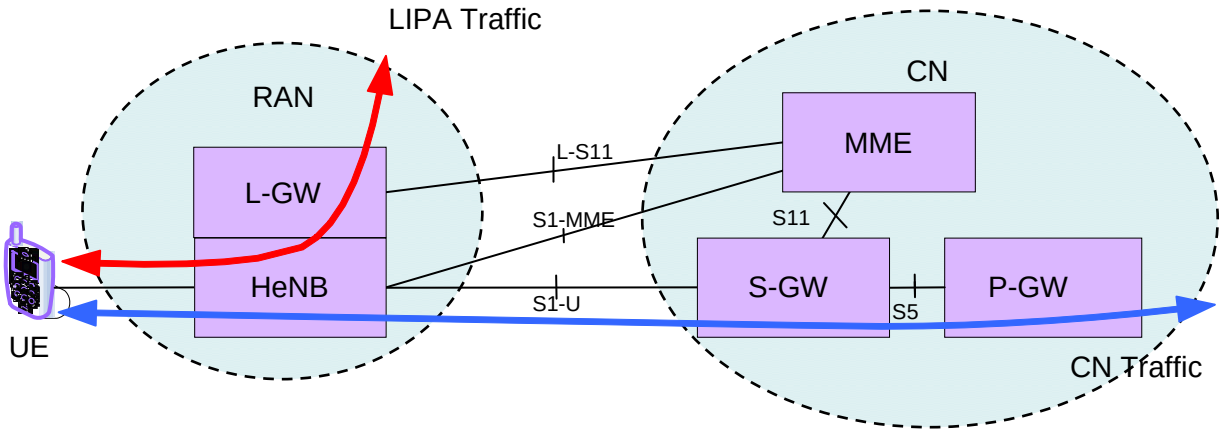

Figure 5.2.3.2.2.1: LIPA solution for HeNodeB

The features of the architecture for LIPA for HeNodeB subsystem are as follows:

- For LIPA traffic, a Local Gateway (L-GW) function for EPS that can be either collocated on the HeNodeB or as a standalone node includes partial P-GW function and S-GW downlink data buffering function; it could also be possible, if required in the future, to be used as anchor point for inter-HeNodeB mobility;
- The S-GW in the core network serves for the CN traffic;
- The L-S11 interface between the L-GW and the MME is used to manage the session for LIPA traffic;
- For the LIPA PDN connection, the L-GW needs to be selected close to the HeNodeB, and establish the connection through L-S11 interface. The L-GW selection mechanism has been specified in clause 6.1;
- When the UE is in idle mode, the LIPA downlink packets are buffered in the L-GW;
- When UE enters connected mode, the packets buffered in L-GW are forwarded on the path between the L-GW and the HeNodeB.

5.2.3.2.3 LIPA Architecture for HNB with S4-SGSN

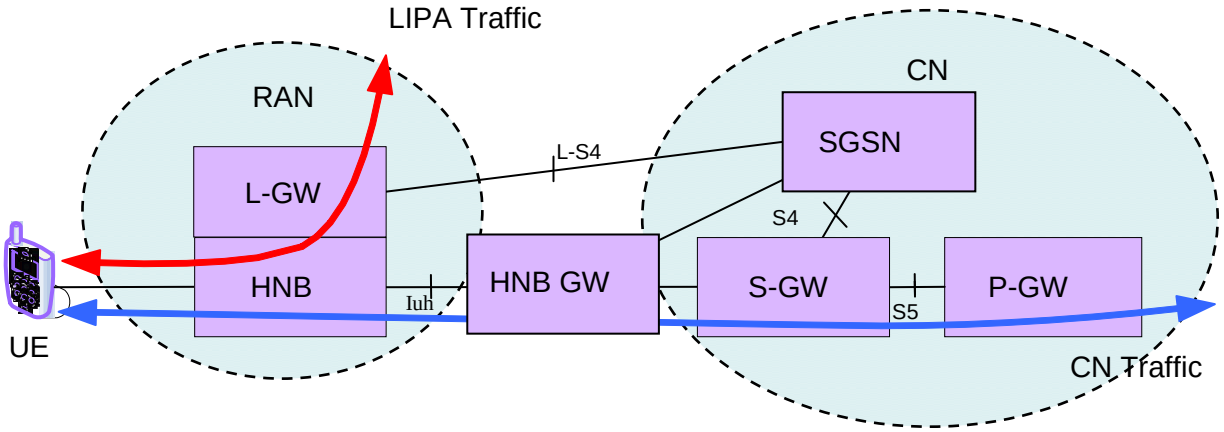

Figure 5.2.3.2.3.1: LIPA and SIPTO solution for HNB with S4-SGSN

The features of the architecture for LIPA for HNB subsystem are as follows:

- For LIPA traffic, a Local Gateway (L-GW) function for S4-UMTS that can be located either on the HNB or on a standalone node includes partial P-GW function and S-GW downlink data buffering function; it could also be possible, if required in the future, to be used as anchor point for inter-HNB mobility;

- The S-GW in the core network serves for the CN traffic;
- The L-S4 interface between the L-GW and the SGSN is used to manage the session for LIPA traffic;
- For the LIPA PDN connection, the L-GW needs to be selected close to the HNB, and establish the connection through L-S4 interface;
- When the UE is in idle mode, the LIPA downlink packets are buffered in the L-GW;
- When UE enters connected mode, the packets buffered in L-GW are forwarded on the path between the L-GW and the HNB.

5.2.3.2.4 L-GW functions

In addition to the common functions in the clause 5.2.2.2, the L-GW for the architecture variant 2 shall include the following functions:

- DL packet buffering in idle mode;
- Data Routing between L-GW and H(e)NodeB in connected mode.

5.2.3.3 Architecture for LIPA for UMTS

For UMTS the direct tunnel functionality can be used to manage the PDP connection between the L-GW and the HNB.

NOTE 1: It is assumed the U plane does not go through the HNB GW for the direct tunnel.

Figure 5.2.3.3.1 illustrates the LIPA architecture including the L-GW function in UMTS for the HNB, where the L-GW is physically co-located with the HNB and no Gn is supported, i.e. LIPA for the residential or single HNB scenario. In the case of the enterprise network where mobility is supported, the L-GW is located above the HNB but still physically within the enterprise network, and the L-GW now includes the functionality to support a Gn interface to the HNB.

NOTE 2: The Gn interface between the L-GW and the HNB is a U-plane interface only.

NOTE 3: To avoid sending LIPA packets to the SGSN when the UE is idle, paging may be triggered by a "dummy" packet sent across Gn and the downlink packets are buffered in the L-GW.

NOTE 4: In case of stand-alone L-GW, a separate IPSec tunnel with the Security GW might be required, whereas a collocated L-GW can reuse the existing IPSec tunnel established by the HNB. The stand-alone L-GW could reuse the same procedures as the HNB for establishing the IPSec tunnel.

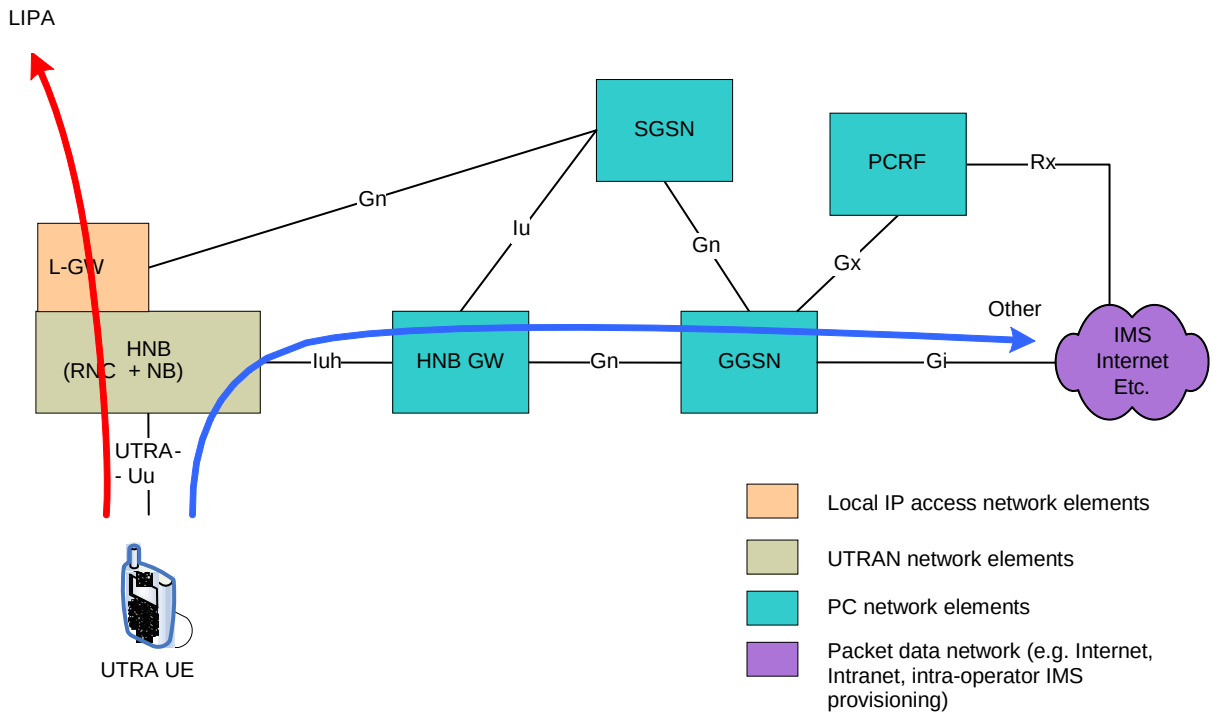

Figure 5.2.3.3.1: UMTS LIPA architecture with mobility support

5.2.4 Open architectural issues

This clause lists the open architectural issues which have been identified for this solution.

Common open issues applying to both UMTS and EPS:

- It is FFS whether Mobility (to macro-network and another H(e)NodeB) is supported/required for LIPA and/or SIPTO traffic;
- It is FFS whether the standalone L-GW architecture is supported for LIPA and SIPTO, and if it is, how.

Open issues applying to UMTS only:

- (none)

Open issues applying to EPS (LTE and S4-based UMTS) only:

- (none)

5.2.5 Evaluation

The following architectures are agreed as the baseline for Solution 1.

Figure 5.2.5.1 illustrates the LIPA architecture for the HeNodeB, where the L-GW is physically co-located with the HeNodeB.

NOTE 1: It is FFS whether mobility within the enterprise is supported for LIPA is defined for this release.

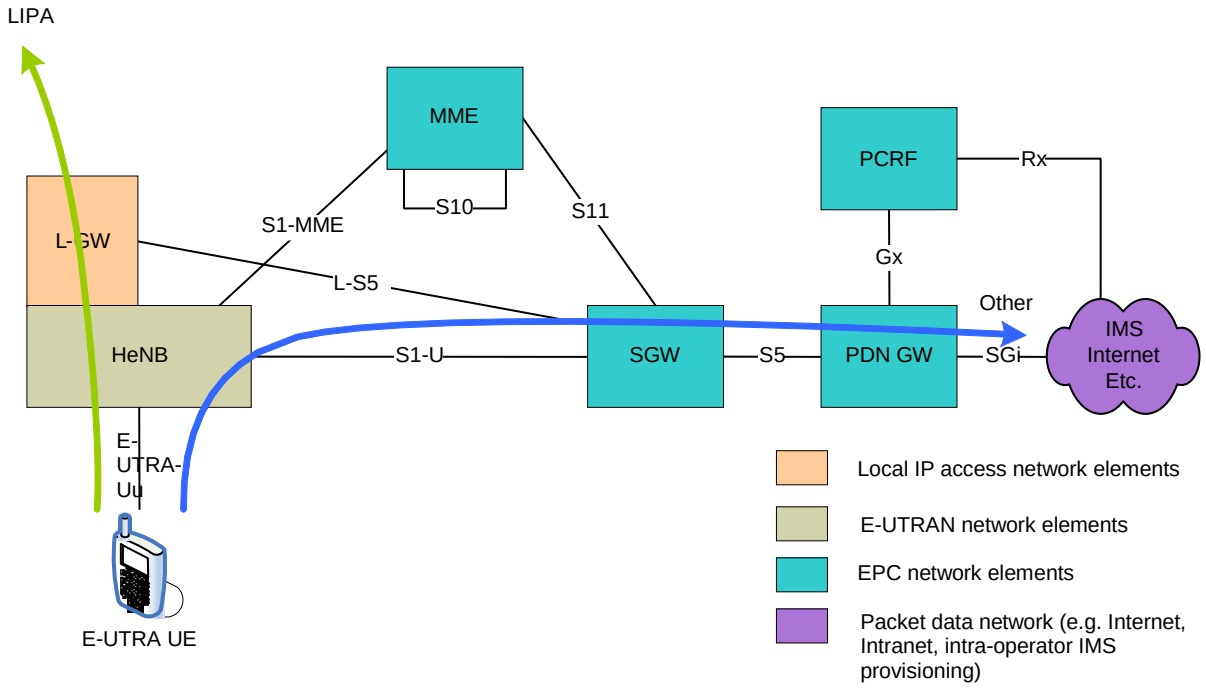

Figure 5.2.5.3: LIPA solution for HNB using local PDN connection for UMTS

5.3 Solution 2 – Local IP Access and Selected IP Traffic Offload at H(e)NodeB by NAT

5.3.1 Applicability

This solution supports the following scenarios:

- Local IP access for HNB and HeNodeB Subsystem
- Selected IP traffic offload for HNB and HeNodeB Subsystem

This solution, implemented by an Offload Processing Module (OPM), is applicable for breakout "in the residential/enterprise IP network".

5.3.2 Architectural principles

- UEs are only required to activate one PDN connection for LIPA, SIPTO, and traffic going through the mobile operator's Core Network;
- The OPM has the ability to drag/insert the LIPA and SIPTO traffic from/into PDN connection per operator policies (e.g. destination address, port number, etc.);
- The H(e)NodeB gets offload policies from the H(e)NodeB Management System as H(e)NodeB configuration data;
- There is a NAT inside the OPM to ensure returning LIPA and SIPTO traffic reaches H(e)NodeB despite topologically incorrect source address;
- UEs that support single PDN connection can simultaneously access LIPA, SIPTO and the mobile operator's Core Network;
- For a PDN connection initiated by a UE connected to a H(e)NodeB, the MME/SGSN shall decide whether LIPA or SIPTO is enabled depending on the subscription data and operator policy. If LIPA should be enabled, the MME/SGSN sends an LIPA indication in the S1/RANAP message to the H(e)NodeB;

- For HNB subsystem using IMSI list to control the UE's access, whether LIPA should be enabled can also be configured per list or per IMSI in the list;

- A dedicated APN may be used to indicate that the PDN connection established through this APN is for LIPA or SIPTO. All the traffics associated with this PDN connection are offloaded.

5.3.3 Paging and Mobility Support

- If the LIPA or SIPTO function is enabled by the MME/SGSN for the UE's PDN connection, the MME/SGSN indicates the public IP address of PGW/GGSN to H(e)NodeB;

- When H(e)NodeB receives a downlink LIPA or SIPTO packet it does not know how to deliver (e.g. when the UE is in idle mode, or when the UE has moved out of the H(e)NodeB coverage), it tunnelled the packet to the P-GW/GGSN after NAT ting the packet and the P-GW/GGSN delivers it as non-LIPA or SIPTO traffic;

- The interface needs to be connected via the Internet.

5.3.4 Architecture diagrams

Figure 5.3.4.1 shows the architecture for the case of HeNodeB. Although this figure focuses on HeNodeB, it can be easily mapped to the HNB case.

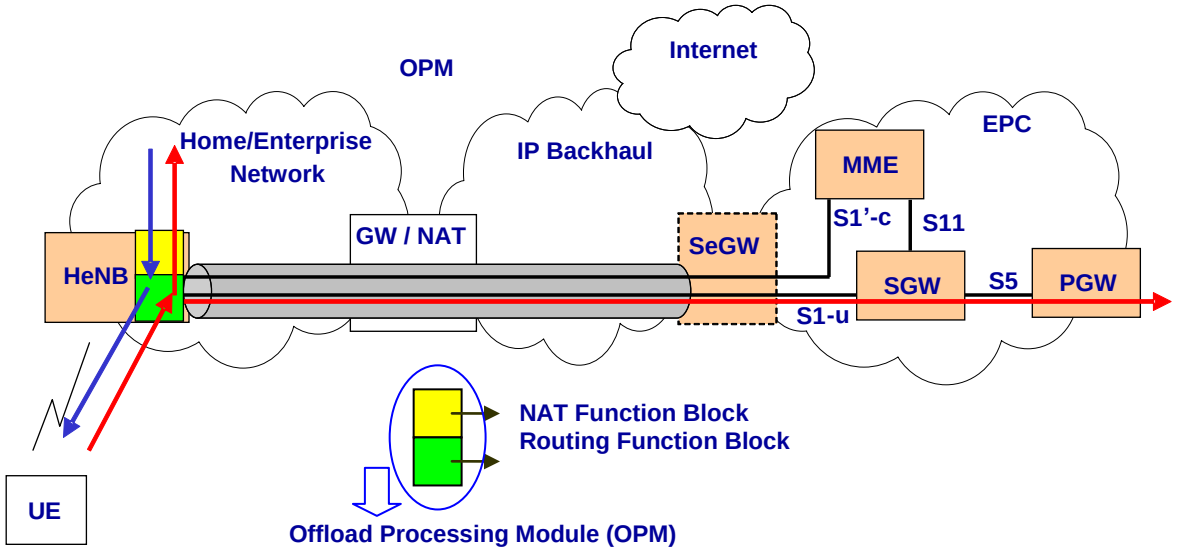

Figure 5.3.4.1: Alternative 1 solution for LIPA / SIPTO with NAT in the HeNodeB

The key features of the architecture shown in Figure 5.3.4.1 are the following:

- The OPM is collocated with the HeNodeB with routing and NATing functions;

- A Security Gateway (SeGW) is located at the edge of the operator's core network. Its role (according to TS 33.320 [7]) is to maintain a secure association with the HeNodeB across an unsecure IP backhaul network;

- A Gateway/NAT device is located at the boundary of the home/enterprise IP network and the IP backhaul network;

- The HeNodeB maintains an S1-u interface with an SGW and an S1-c interface with an MME. The S1-c interface may be enhanced (referred to as S1'-c in Figure 5.3.4.1) in order to support HeNodeB-triggered paging (see further details below);

- The S1-u connection between the HeNodeB and the SGW is maintained even when all traffic is exchanged between the UE and the local home/enterprise network;

- The UE establishes a LIPA enabled PDN connection by using the standard signalling procedures specified in TS 23.401 [6] with no additional requirements. The same PDN connection is used for traffic between the UE and

a packet data network over SGi and for traffic between the UE and local home/enterprise network. The MME indicates to HeNodeB if LIPA is enabled for the established bearer;

- The OPM performs routing enforcement of uplink traffic, i.e. decides if an uplink packet should be routed to SGW through the standard S1-u interface, or if it should be routed to the collocated NAT function. This routing enforcement is transparent to UE and requires no special UE assistance;

- The routing decisions are based on preconfigured LIPA routing rules (either statically configured or downloaded from the HeNodeB management system). Such routing rules can be of the form: "if IP dest_address=192.168.0.0/16, forward traffic to NAT, otherwise forward traffic to the SeGW";

- When a packet arrives at the OPM and the UE is in idle mode, the HeNodeB-triggered paging is invoked by sending a special message to the MME. This is schematically illustrated in Figure 5.3.4.2. The UE identity sent to the MME for initiating paging can be GUTI (this requires changes to specifications).

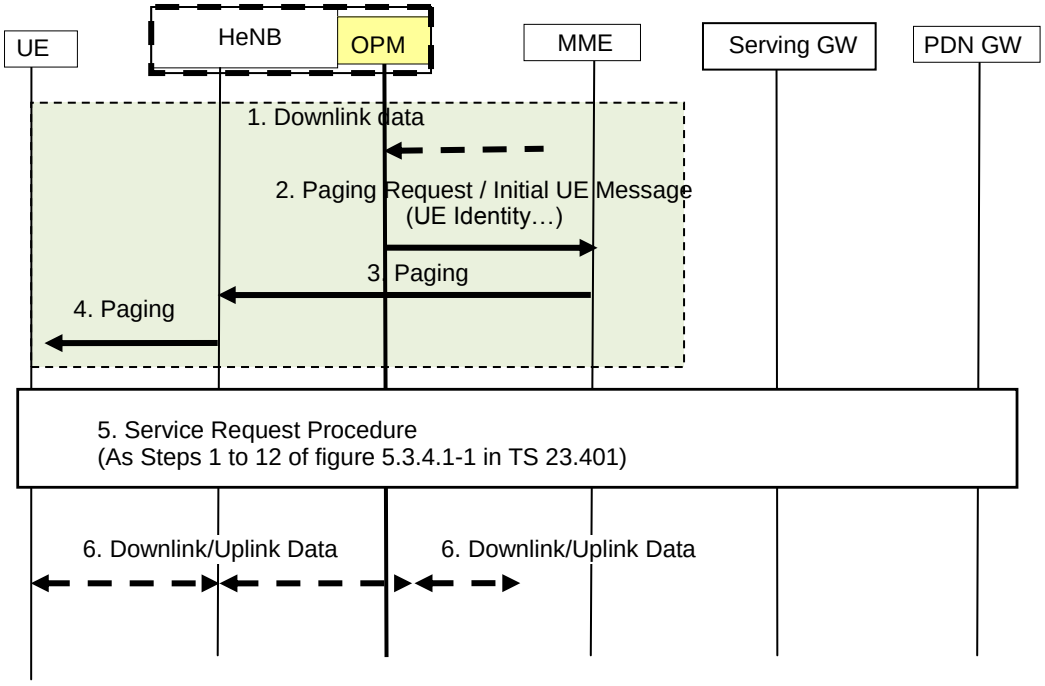

Figure 5.3.4.2: HeNodeB-triggered paging

5.3.5 Deployment requirement and limitations

- The local network manager should be able to configure the local network so that the services that the UE can access have different IP address with the private IPv4 IP address used by operator's services.

- When paging is not supported, the UE is forced to stay in RRC connected mode for as long as it remains within the HeNodeB coverage. The data is discarded/buffered until the IDLE UE returns to connected mode.

- Stand-alone NAT function and session continuity of LIPA traffic during inter-H(e)NodeB handover are not supported in Rel-10.

- IPv6 NAT is not standardized. IPv6 traffic offload is left for implementation which may be custom-built for specific operator.

- If local DNS server is used, the OPM should be able to properly route the DNS query of the local services.

5.3.6 Standard Impacts

1) Add the LIPA_enabled flag (per APN and per CSG) in the user's subscription data stored in the HSS/HLR and transfer to the MME/SGSN;

2) The MME/SGSN indicates the H(e)NodeB to enable LIPA for a UE when a RAB is to be setup;

NOTE 1: The 1) 2) can be avoided in HNB subsystem by configuring LIPA_enabled flag in the IMSI list used to perform access control.

3) The offload policies are sent from the HMS to the H(e)NodeB together with other H(e)NodeB configuration data;

4) Impact if a new S1-MME/RANAP message is introduced for paging:

   - If GUTI is used for UE identity, GUTI is send to RAN in the procedures of PDN connectivity establishment/modification whenever the MME allocates or reallocates a new GUTI to UE, MME/SGSN shall be able to recognize the message.

NOTE 2: The 3) can be avoided if the LIPA routing policy is statically configured.

NOTE 3: The 4) is optional up to whether paging is supported in Solution 2 in Rel-10.

5.4 Solution 3 – GGSN allocation to offload point

5.4.1 Applicability

This solution supports the following scenarios:

- LIPA
- SIPTO from the HNB Subsystem
- SIPTO from the macro network

This solution is applicable for breakout "in the residential/enterprise IP network" and breakout "at or above RAN".

5.4.2 Architectural principles

In this solution, LIPA and SIPTO are enabled by the SGSN selecting a GGSN that provides enhanced (e.g. shorter) traffic routeing capabilities located within the RAN.

It is enabled by:

- the RAN providing the SGSN with the IP address(es) of one or more GGSNs that the RAN believes offers good traffic routeing capabilities. The RAN provides this information to the SGSN at every RAN initiated Iu-ps connection establishment and, from the target RNS, at every SRNS relocation;
- the SGSN using the information from the RAN and HSS to potentially override the normal GGSN selection algorithm; and
- the SGSN using the permitted CSG/APN information and information supplied by the RAN to cause the release of a PDP context, if required by the service continuity restrictions, when the mobile leaves the CSG.

The SGSN reuses the Direct Tunnel functionality (from TS 23.060 [5] clause 15.6) to establish and maintain user plane connectivity. The conditions restricting the use of Direct Tunnel defined in TS 23.060 [5] also apply when determining whether local breakout can be applied.

The subscription data stored in the HSS indicate which CSGs are permitted to perform local breakout via LIPA for each of the APN subscribed, along with service continuity restrictions/permissions, the type of breakout permitted. The subscription data also indicates which CSGs are permitted to perform Internet Breakout for each of the subscribed APNs along with service continuity restrictions/permissions, the type of breakout permitted.

The RAN shall report to the SGSN the level of support for local breakout to the SGSN. The SGSN decides what level of local breakout to perform based on information received from the RAN and the whether local breakout is permitted by the subscription data. The SGSN operator shall be able to configure the Emergency APN such that Local Breakout does not endanger PS domain Emergency calls.

For dual stack PDP contexts (PDP type = IPv4v6), the assigned GGSN function shall select an IP version appropriate for the breakout connection.

5.4.3 Location of breakout point

Editor's Note: It is FFS whether the same principles can be applied for LTE / S4 breakout.

5.4.2.1 RNC Breakout

To enable RNC breakout, the SGSN allocates a GGSN function located e.g. in or near the RNC, based on the IP address for GTP control plane received for breakout from the RNC during the Iu connection establishment. The presence at the SGSN of an IP address for GTP control plane for Internet Breakout is used as an indicator of support for breakout in the RAN. This location enables SIPTO from the macro network.

5.4.2.2 HNB-GW Breakout

To enable HNB-GW breakout, the SGSN allocates a GGSN function located e.g. in or near the HNB-GW. The HNB-GW appears as an RNC to the SGSN and as such the operation is the same as for RNC breakout 5.4.2.1. This location enables SIPTO from the HNB Subsystem.

5.4.2.3 HNB Breakout

To enable LIPA, the SGSN allocates a GGSN function located at the HNB based on the IP address for GTP control plane for LIPA received from the RAN. The presence at the SGSN of an IP address for GTP control plane for LIPA is used as an indicator of support for LIPA in the RAN. This location enables LIPA and SIPTO from the HNB.

When performing breakout at the HNB, the PCC architecture specified in TS 23.203 [4] is not supported. This requires the operator to correctly configure the HSS such that APNs which require the PCC architecture are not selected for breakout at the HNB.

5.4.4 Mobility aspects of LIPA and SIPTO

When Local Breakout is active, at all mobility events involving the core network the [source] SGSN shall re-evaluate the eligibility of Local Breakout and disconnect any PDP contexts for which the specific breakout point is no longer allowed. If one or more PDP contexts are no longer allowed for the current breakout point:

- at intra-SGSN mobility, the SGSN shall trigger the SGSN initiated PDP context deactivation procedure; otherwise;
- at mobility to a new SGSN/MME, the target SGSN deactivate the PDP contexts determined as not applicable for local breakout.

NOTE: The behaviour should be the same as for PS domain emergency handling for HO into a restricted area when ordinary contexts/bearers are active.

During mobility between different CSGs, the SGSN shall determine whether the CSG-ID has changed, if it has changed the SGSN shall re-evaluate whether the existing local breakout can still be applied for the target CSG.

Editor's Note: Aspects related to handover into a HNB and the need for GGSN function relocation are FFS.

5.5 Solution 4 – Selected IP Traffic Offload at Iu-PS

5.5.1 Applicability

This solution supports the following scenario:

- Selected IP Traffic Offload for UMTS macro;
- Selected IP Traffic Offload for HNB subsystem.

This solution is applicable for breakout "at or above RAN".

5.5.2 Architectural principles

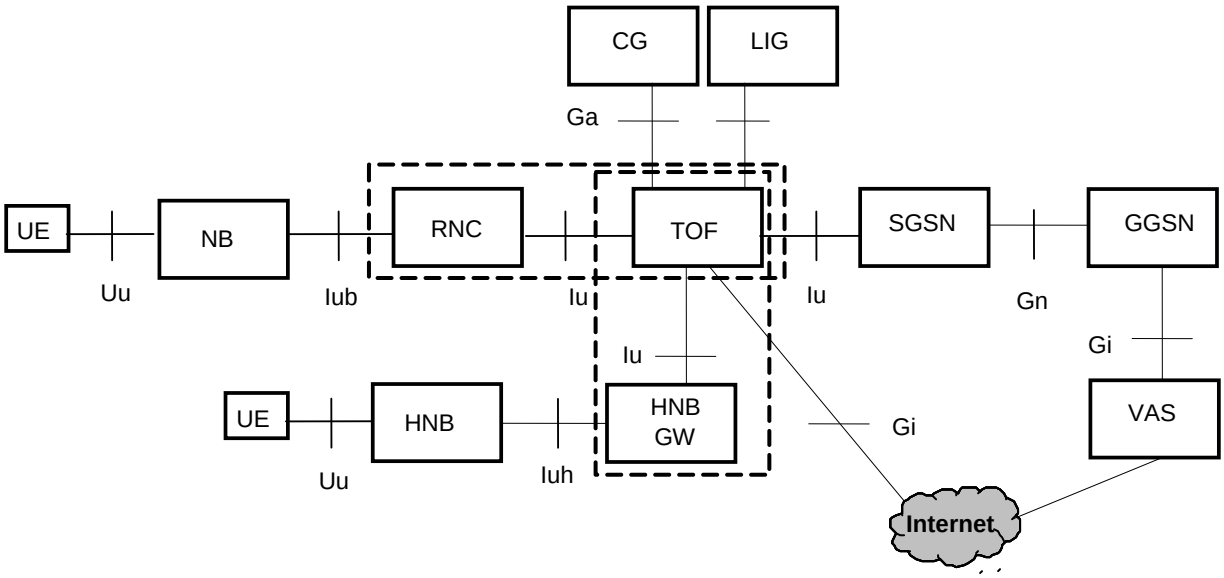

Figure 5.5.2.1: Selected IP Traffic Offload from Traffic Offload Function (TOF) deployed at Iu-PS

NOTE 1: TOF can be a separate entity, or collocated with RNC/HNB GW.

NOTE 2: The interface from TOF to Internet may be a subset of Gi.

The following architecture principles apply to this solution:

- The TOF is located at Iu-PS and provides standard Iu-PS interface to the RNC and the SGSN;
- Selected IP Traffic Offload is enabled by NAT and packet inspection based on operator policies at different levels (e.g. per user, per APN, per service type, per IP address, etc). The policies may be configured via e.g. OAM.

NOTE 3: DPI can be used when operators want to offload the traffic based on e.g. application level information. When offload is performed based on e.g. the port number, there is no need to inspect the detailed content of the packet beyond identifying the port number.

- Enabling/Disabling SIPTO based on user information (e.g. age) can be achieved by interaction with the operator's user data repository.
- One PDN connection or PDP context for both offload traffic and non-offload traffic is supported, while it also allows using different PDN connections or PDP contexts for offload traffic and non-offload traffic (e.g. by selecting the traffic based on APN);
- No impact on the quality of service continuity provided for non-offload traffic during mobility;
- The quality of service continuity provided for offload traffic is same as it is for non-offload traffic during intra TOF mobility.

5.5.3 Traffic Offload Function

TOF includes the following functions:

TOF includes the following functions:

- NAS and RANAP message inspection to build/remove local UE offload context;
- NAS and RANAP message inspection to build/Remove local session off load context;
- Packet inspection and Selected IP Traffic Offload policy enforcement;

- Uplink traffic offload by removing GTP-U header and NAT;
- Downlink traffic offload by reverse NAT and adding GTP-U header;
- Charging for offloaded traffic;
- Lawful Interception for offloaded traffic;
- Offload traffic service continuity during intra-TOF mobility.
- When TOF is configured to perform paging, TOF pages idle mode UE for downlink offload traffic, and when the UE responds TOF modifies the Service type IE in the Service Request message to indicate Data.

NOTE 1: Most Internet applications is in MS originated service mode, or have a heart-beat mechanism to keep the UE in connected mode, so it is considered that paging function is not needed in most cases.

NOTE 2: DPI can be used when operators want to offload the traffic based on e.g. application level information. When offload is performed based on e.g. the port number, there is no need to inspect the detailed content of the packet beyond identifying the port number.

5.5.4 Offload procedure

- TOF inspects both NAS and RANAP messages to get subscriber information and establish local UE offload context.
- TOF inspects both NAS and RANAP messages to get PDP context information and establish local session offload context.
- TOF decides the offload policy to be applied based on above information during e.g. attach and PDP context activation procedures.
- During the data transfer procedure, TOF performs necessary packet inspection to uplink traffic.

NOTE 1: DPI can be used when operators want to offload the traffic based on e.g. application level information. When offload is performed based on e.g. the port number, there is no need to inspect the detailed content of the packet beyond identifying the port number.

- TOF drags the uplink traffic out from the GTP-U tunnel and performs NAT to offload the traffic if offload policy is matched.
- TOF performs reverse NAT to the received downlink offload traffic and inserts it back to the right GTP-U tunnel.
- The IPv6 NAT mechanism is not standardized. If IPv6 NAT is not implemented, IPv6 traffic is not offloaded.
- TOF removes session offload context when the APN which the IP traffic associated with it is to be offloaded and all the PDP contexts associate with that APN are deactivated.
- TOF removes session offload context when the APN which the IP traffic associated with it is to be offloaded and all the PDP contexts associate with that APN are deactivated.
    - The inactivity timer is reset and started with its initial value when the PS signalling connection between the MS and the network is released. The inactivity timer is stopped when the PS signalling connection is established between the MS and the network.
    - When inactivity timer expires, the TOF removes the offload contexts including the UE offload context and the session offload context.
- When TOF is configured not to perform paging, TOF discards received downlink packets.
- When TOF is configured to perform paging and when TOF receives downlink offload packets for idle mode UE, it constructs and sends paging message to connected RNC(s). When TOF receives Service Request message from the UE as paging response, it modifies the service type IE to Data, then forwards it to SGSN. TOF discards the downlink offload packets if there is no corresponding local context.

5.5.5 Impacts on specification

Impacts common to both solution 4 and solution 5:

- *HSS: flag for SIPTO per APN:* adding to the HSS a flag per user per APN indicating whether the APN can be offloaded.
- *SGSN/HSS (Gr): transmission of the flag from HSS to SGSN:* adding the transmission of the flag from the HSS to the SGSN over the Gr interface.

Impacts due to solution 4:

- *SGSN: per RAB offload flag on RANAP, based on the APN information:* adding the transmission of the offload flag over the Iu-PS to be captured by the TOF to perform the offload decision.
- *SGSN: Charging parameters including APN on RANAP:* adding the transmission of Charging parameters (e.g. Charging characteristics only) including the APN over the Iu-PS to be captured by the TOF for charging purposes.

The additional information elements should be encoded so that the RNC is not impacted.

5.6 Solution 5 - Selected IP Traffic Offload solution based on local PDN GW selection

5.6.1 Applicability

This solution supports the following scenarios:

- Selected IP traffic offload for macro network
- Selected IP traffic offload for Home (e)NodeB subsystem

This solution is applicable for breakout "in the residential/enterprise IP network" and breakout "at or above RAN".

5.6.2 Architectural principles

Common principles applying to both GPRS and EPS:

- The GW selection mechanism in the MME/SGSN takes into account the location of the user for the PDN connection/PDP context activation, and selects a GW that is geographically/topologically close. As described in clause 6.1, this solution proposes to use a DNS based mechanism to perform GW selection: either the Rel-8 DNS based mechanism or the DNS based alternative for 3G GPRS provided in clause 6.1.
- Selected IP traffic is offloaded at the local gateway using external IP connectivity.
- Multiple solutions exist for determining whether SIPTO is applicable on a per UE per PDN basis, such as: SIPTO enabled flag in HSS; local configuration; or dedicated APN. The choice of the solution may depend on operator deployment preferences.
  - **Alt. A.** A new SIPTO_enabled flag is defined associated with each APN in the user's subscription, where the flag indicates whether the connection to that APN is enabled/disabled for SIPTO. The SGSN/MME can then use this flag to determine whether or not to use a SIPTO PDN for that APN, based on whether a SIPTO GGSN/P-GW is available when the UE establishes the PDP context/PDN connection.

    This option gives flexibility for the operator to be able to configure SIPTO on a per UE and per APN basis together with other subscription data. On the other hand, it relies on an upgraded HSS node which might not be always easy or cost-effective in all deployments. Furthermore it also requires that the subscription data be populated with per APN information which also requires extra attention from the operator, especially if the operator uses wildcard APNs.
  - **Alt B.** MME/SGSN acquires per UE per PDN specific SIPTO information using some kind of local configuration. This may take the form of a DNS query including the APN and some user identifier, with the

appropriate information configured into the DNS system. Or the information can be configured into the MME/SGSN nodes using O&M; other methods may also be used. This approach has the benefit of not requiring an HSS upgrade, and consequently it might be easier to deploy. On the other hand to support this solution for the roaming case, the roaming partners need to implement the same method and configuration (such as DNS) for this to work.

- **Alt. C.** A dedicated APN may be used for SIPTO. The operators can use the already available automatic terminal configuration methods to provide a dedicated APN for SIPTO users. This approach has the benefit of not requiring any upgrades to MME/SGSN or HSS nodes for determining SIPTO eligibility, but the disadvantage is the required terminal APN configuration.

- The existing procedures for PDP context/PDN connectivity activation can also be used to establish SIPTO.

- The existing procedures for PDP context/PDN connectivity deactivation can be used to relocate the SIPTO PDN connections.

- The UE may attempt to re-establish the PDN based on the cause code sent by the SGSN/MME to deactivate the PDN or due to a request to re-establish the PDN from an application. Details of how the cause code is used are FFS.

- Source MME/SGSN may indicate the SIPTO status during SGSN/MME relocation procedure.

NOTE: It is possible to provide a limited SIPTO solution with the existing architecture/system without requiring any normative changes to the specifications. This could be enabled by applying SIPTO to certain APNs in the user's profile and/or some configuration in MME/SGSN (as described in Alt. B and Alt. C above).

5.6.3 Architecture Diagrams

In the following diagrams, the L-PGW/GGSN refers to the P-GW/GGSN closer to RAN.

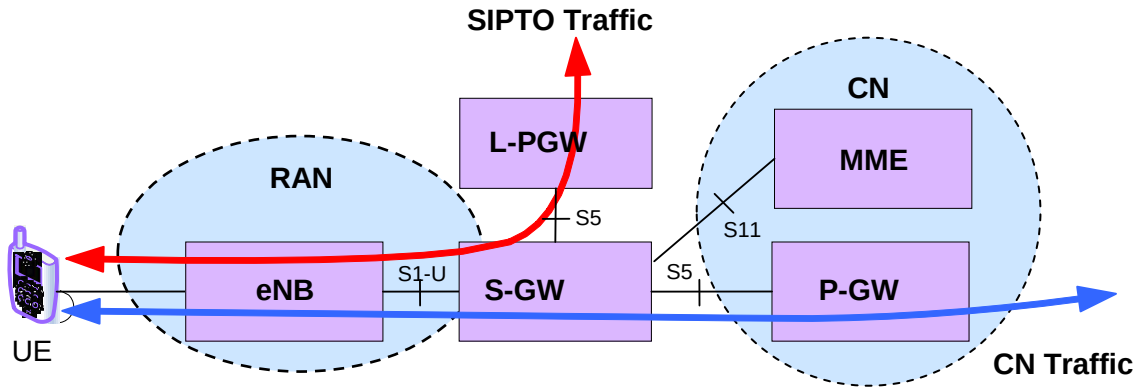

Figure 5.6.3.2: SIPTO for UMTS/EPC macro network with S4 SGSN

SIPTO for UMTS macro network with S4 SGSN is depicted in Figure 5.6.3.2. It is similar to LTE system. The additional architecture principles are listed below:

- The user plane for CN traffic is direct tunnel (RNC-SGW-PGW);

- The PGW can be collocated with SGW which is deployed above the RNC(s).

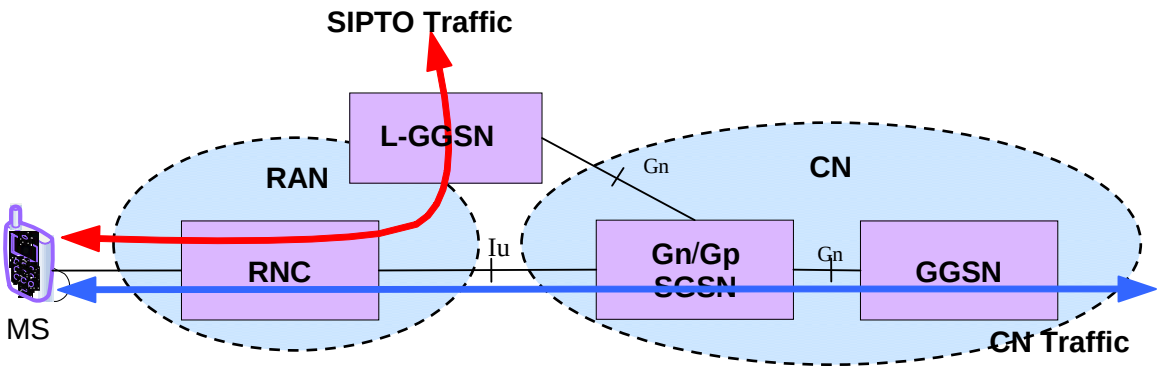

Figure 5.6.3.3: SIPTO for UMTS/GPRS macro network with Gn/Gp SGSN with Direct Tunnel

SIPTO for UMTS macro network with Gn/Gp SGSN is depicted in Figure 5.6.3.3. In this architecture, there is no influence on RAN nodes and core network entities. The features of this architecture are the following:

- The GGSN which belongs to core network is deployed above RNC(s), and closer to RAN;
- Direct tunnel between RNC and GGSN also can be used for CN traffic.

If the operator uses the same single APN for both internet and operator services and wants to perform SIPTO for internet traffic, then the operator can use a local GW for both types of traffic. Internet traffic is offloaded locally, while traffic for operator services also use the same GW but is routed within the operator's network.

NOTE: Additional functionality (e.g. adult content filters) may also be required to be deployed near the GGSN/P-GW in the case of a single APN.

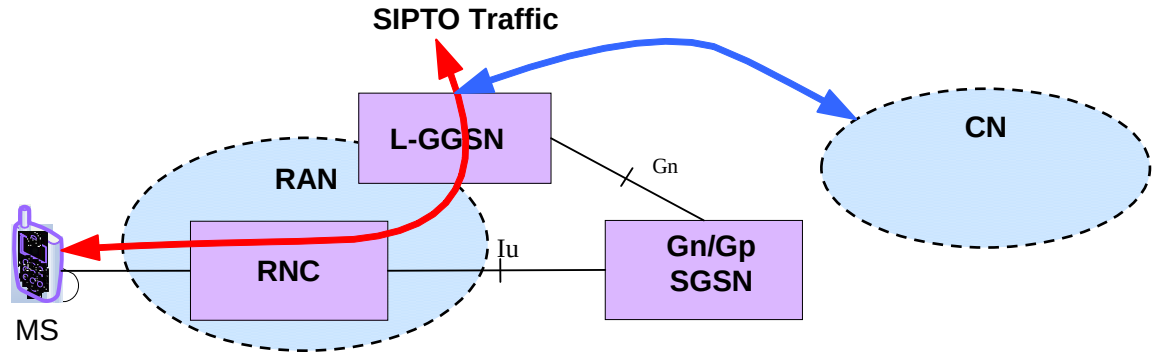

Figure 5.6.3.4: Single APN based solution with GPRS architecture where operator services traffic terminate within CN

5.6.4 Standards impacts

The following interface change needed to support the user subscription enabling SIPTO per APN basis:

- *SGSN/MME/HSS (Gr/S6d/S6a): transmission of the flag from HSS to SGSN/MME:* adding the transmission of the flag from the HSS to the SGSN/MME over the Gr/S6d/S6a interface.

The following functional changes need to be added to the standard to support this solution:

- Including the SIPTO_enabled flag (per APN) in the user's subscription data stored in the HSS/HLR and transferred to the MME/SGSN;
- Indicating how the MME/SGSN processes the SIPTO_enabled flag in order to decide whether to offload the traffic for this APN;
- GW Selection mechanism enhanced to take into account the user's location related information;

- Indicating how the MME/SGSN triggers relocation of the SIPTO PDP context/PDN connection, e.g., by deactivating the PDN connection with reactivation (e.g. guidance as to when to trigger the relocation of the GGSN/P-GW).

5.6.5 Open architectural issues

This clause lists the open architectural issues which have been identified for this solution.

- Whether existing GW selection mechanisms need to be improved for selected IP traffic offload for the case that the GW is co-located with HeNodeB or HNB.

5.7 Solution 6 - Local Gateway based Architecture

5.7.1 Applicability

This solution supports the following scenarios:

- Local IP Access for H(e)NodeB subsystem
- Selected IP Traffic Offload for H(e)NodeB subsystem
- Selected IP Traffic Offload for macro network

The solution applies to both types of approaches: with separate APNs for SIPTO and non-SIPTO traffic, and also with common APN(s) for SIPTO and non-SIPTO traffic.

This solution is applicable for breakout "in the residential/enterprise IP network".

5.7.2 Architectural principles

Figure 5.7.2.1 shows the architectural extension proposed by this solution for the case of E-UTRAN and (macro) eNB; analogous extensions apply for the HeNodeB case and the UTRAN case (both NB and HNB). In the latter case, the GGSN maps onto PDN GW, and SGSN maps onto Serving GW (user plane part) and MME (control plane part).

A Local Gateway (L-GW) is co-located with home or macro cells in support of LIPA or SIPTO. Between the L-GW and the PDN GW (for EPS) or the GGSN (for GPRS), a "L-GW extension tunnel" is configured. The functions of L-GW include:

- gateway and routing to/from external PDN (e.g. internet, enterprise or home NW), equivalent to SGi;
- tunnelling of IP packets through the extension tunnel to/from P-GW/GGSN (e.g. based on GTP, PMIP, IP in IP or other); when the UE is in idle mode, the extension tunnel is used only in the L-GW to P-GW/GGSN direction; while when the UE is in active mode and connected via a different cell from where the L-GW is located, the extension tunnel is used in both directions;
- IP address handling (either IP address allocation and conveyance to P-GW/GGSN, or alternatively reception of IP address from P-GW/GGSN and NATing);
- minimal state maintenance for mapping of traffic onto tunnels (from external PDN onto extension tunnel);
- coordination with (e)NB on usage of local breakout (trigger (e)NB for local traffic handling);
- decision function on usage of local breakout for uplink traffic (optionally it can be part of the (e)NB);
- decision function on routing for downlink traffic (directly to (e)NB versus via extension tunnel);
- traffic monitoring and reporting function (optional): required only as a means to limit the principally assumed flat rate charging.

The basic L-GW functionality is radio and core network technology agnostic, and thus allows that the same L-GW function to be adopted for both 3G and LTE radio cells as well as to GPRS and EPS core networks.

As visible from this list, the L-GW is not a PDN GW or GGSN shifted to eNB/E-UTRAN, but encompasses only minimal functionality.

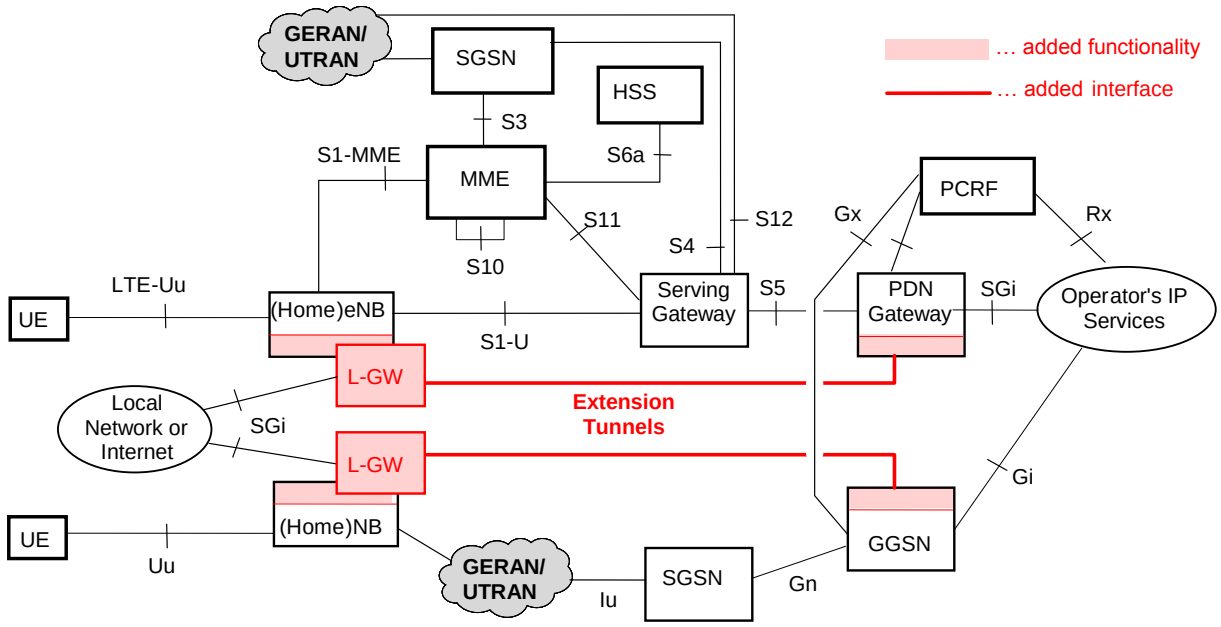

Figure 5.7.2.1: Proposed extension of non-roaming architecture for 3GPP accesses for SIPTO and LIPA

NOTE: As the extension tunnel between the L-GW and P-GW/GGSN is terminated within the Home (e)NB, the security measures already defined for Home (e)NBs (i.e. IPSec on Iuh) apply.

P-GW and/or GGSN functionality is enhanced by:

- establishment of extension tunnel (upon PDN connection or PDP context establishment for APNs matching the criteria for local traffic),
- traffic forwarding through extension tunnel and to/from S5/S8 or Gn tunnel,
- IP address handling (either obtain of IP address from L-GW, or alternatively conveyance to L-GW).

Enhancements of (e)NB are the following:

- provision of UE's access state for the cell(s) served by the (e)NB to the L-GW, and
- (optionally) the decision function on usage of local breakout for uplink traffic (based on APN).

The established 3GPP architectures (GPRS, EPS) and signalling procedures are re-used to the maximum extent possible. Specifically, the paging and mobility signalling procedures are used unchanged.

The enhanced EPS architecture also enables mobility management between 3GPP and non-3GPP accesses: since the PDN GW is always in the path when the UE leaves the eNB, the mobility support function of handover towards non-3GPP accesses can be handled by the PDN GW as usual. Such functionality does not need to be provided as part of the L-GW (or within the eNB).

Regarding charging and policy-control, it is handled by the P-GW in EPS and GGSN in GPRS as usual for non-SIPTO traffic, and for SIPTO/LIPA traffic if the UE moves away from the eNB with the anchoring L-GW. This is inline with the needs for differentiated charging/policing.

Dynamic control for LIPA/SIPTO handling in the PDN-GW is possible (it is switched on only after the extension tunnel is set up).

The architecture solves the configuration problem associated with the masses of (home)(e)NBs in a similar manner as S1-flex.

5.7.3 Open issues

- If one and the same APN is used for SIPTO traffic and non-SIPTO traffic, the technical limitations of NAT apply;
- It is FFS whether the standalone L-GW architecture is supported for LIPA and SIPTO, and if it is, how;
- Details on simultaneous use of LIPA and SIPTO for the same UE;
- How the (Home) (e)NodeB maps the downlink packets received from the L-GW on the appropriate radio bearers;
- Disconnection of the extension tunnel.

5.7.4 Establishment of PDN connectivity subject to LIPA or SIPTO

The signalling flow for selection of the PDN GW, L-GW, IP address allocation and establishment of the extension tunnel is shown in figure 5.7.4.1 for the EPC case. It re-uses the procedure for UE requested PDN connectivity (as specified in clause 5.10.2 of TS 23.401 [6]). Enhancements are necessary for the signalling of the request for "local" connectivity (by a special APN) and (local) IP address allocation.

IP address allocation is done differently for single and multiple APN PDN connection cases:

- Case A: for multiple APN case: the UE is assigned a IP address by the L-GW;
- Case B: for single APN case: the UE is assigned an IP address from the operator's IP address space, which is translated in the L-GW by means of NAT; in this way the UE is unaware of the offloading, and is not involved with IP address handling if it changes the L-GW.

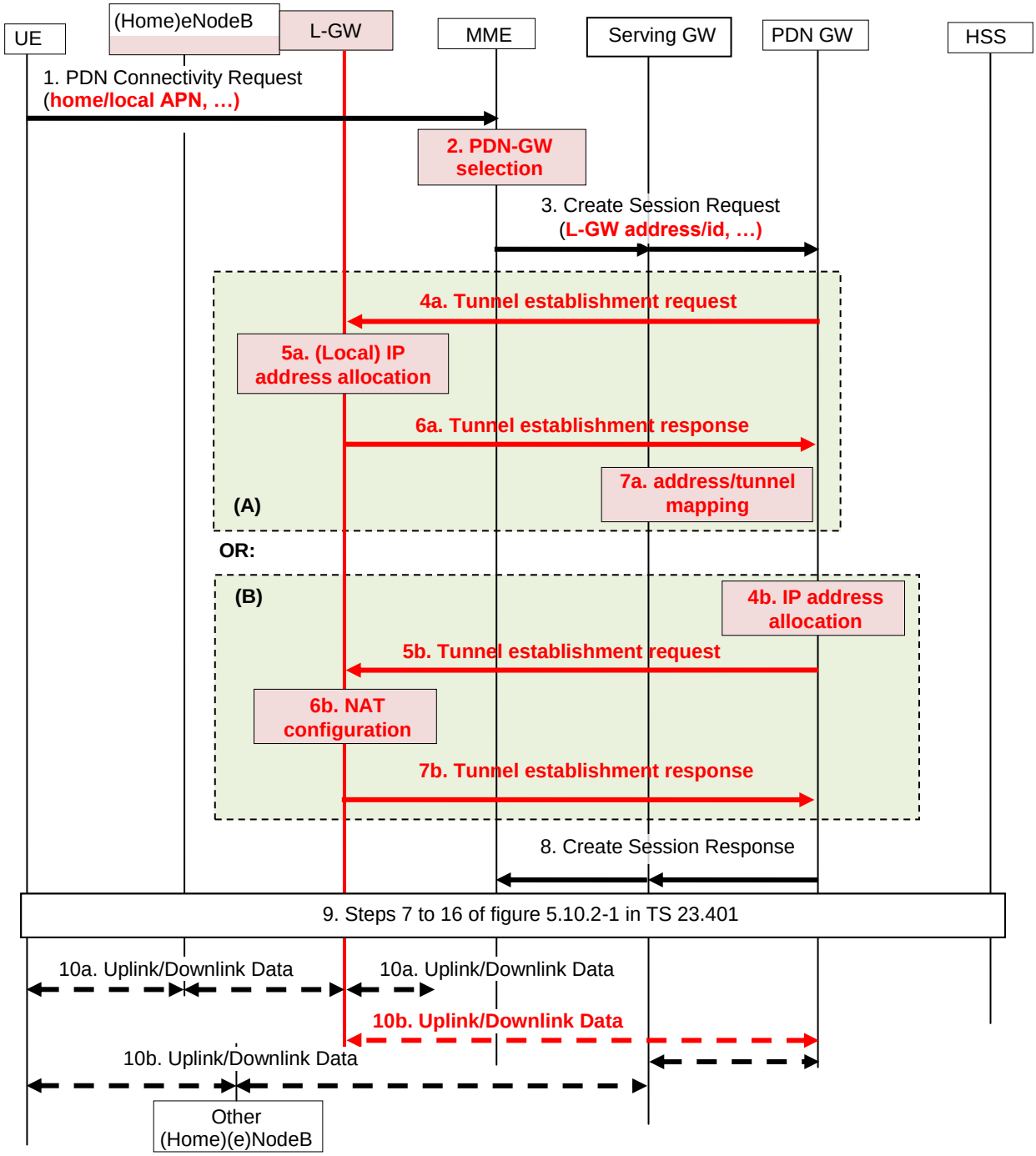

Figure 5.7.4.1: Information flow for establishment of PDN connectivity subject to LIPA/SIPTO with extension tunnel

In detail, the steps are:

1. the UE sends a PDN Connectivity Request message to the MME, conveying a special APN (e.g. "home" for LIPA or "internet" for SIPTO).

2. The MME checks in subscription data if "home" or "internet" access is allowed and configured. In case of "home" access the address of the L-GW is directly stored with subscription data; in case of "internet" access the address is derived from the eNB address by MME. The MME selects a PDN GW, based on the PDN GW selection procedure defined in clause 4.3.8.1 in TS 23.401 [6]; additionally, the location of L-GW may optionally be considered in order to minimize the length of extension tunnels through the network.

3. Signalling for session establishment is performed towards Serving GW and PDN GW, conveying the L-GW address or identity. If only the L-GW identity was provided in the previous step, the PDN GW has to resolve it to an IP address.

- Case A: for multiple APN case, IP address allocation by L-GW:

  4a. PDN GW requests establishment of the extension tunnel

NOTE: the description here remains generic, as long as no decision on a particular user and control plane protocol has been taken; examples of protocols serving this purpose are GTP and PMIP.

  5a. The L-GW allocates IP address information (IPv4 or IPv6 prefix or both), which is local to the L-GW and co-located (Home)eNB.

  6a. Signalling for extension tunnel establishment is completed, and the IP address information is conveyed to PDN GW.

  7a. The PDN GW uses this information to send it to the UE in the subsequent steps and uses it in the mapping of extension and S5/S8 tunnels.

- Case B: for single APN case, IP address allocation by PDN GW:

  4b. PDN GW allocates IP address information (IPv4 address or IPv6 prefix or both).

  5b. PDN GW requests establishment of the extension tunnel from L-GW and conveys the IP address information in this signalling message.

  6b. The L-GW uses IP address information for configuring NAT.

  7b. Signalling for extension tunnel establishment is completed.

8. Response signalling for session establishment is performed from PDN GW towards Serving GW and MME.
9. The remaining steps of a UE requested PDN connectivity procedure are executed as usual.
10. At this point uplink and downlink data may be transferred between a PDN and the UE:

  10a. via L-GW and this (Home)eNodeB directly, if the UE is connected via the (Home)eNodeB supporting the LIPA/SIPTO.

  10b. through the extension tunnel, via PDN GW, Serving GW and any other (Home)eNodeB, if the UE is not connected via the (Home)eNodeB supporting LIPA/SIPTO.

5.7.5 Terminating traffic handling

Three cases need to be considered:

1. The UE is active at the (Home)(e)NodeB where L-GW is co-located: a local shortcut of traffic needs to be applied.
2. The UE is in idle mode: paging has to be performed via the core network (i.e. through the extension tunnel).
3. The UE is active at a (Home)(e)NB different from the one where L-GW is co-located: traffic has to be forwarded through the extension tunnel to the core network.

The decision between case 1 and the other two cases can be taken easily by the (Home)(e)NB. Cases 2 and 3 start with identical handling, and the differentiation occurs at Serving GW; the signalling flows for these are shown in figure 5.7.5.1.

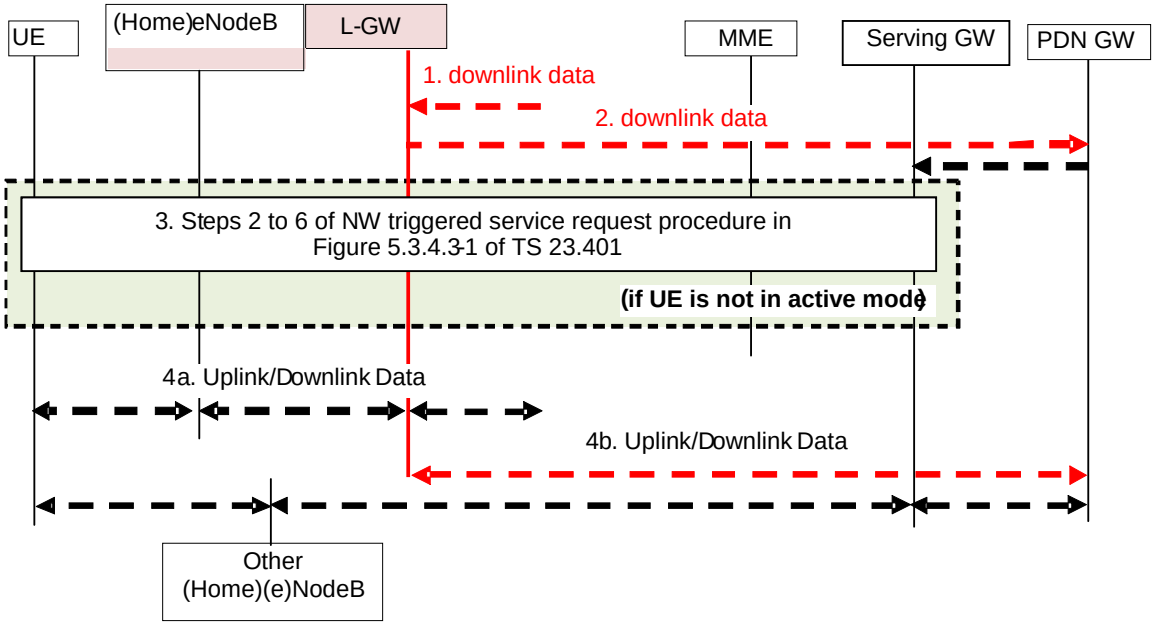

Figure 5.7.5.1: Information flow for terminating traffic with extension tunnel, incl. paging (UE not active at the (Home)eNodeB supporting LIPA/SIPTO)

The following steps:

1. Downlink data arrives at the L-GW; L-GW knows that the user is not locally connected and therefore this traffic is subject to forwarding through the extension tunnel.

NOTE: The UE could be idle or active in another cell.

2. L-GW tunnels and forwards the data to PDN GW; the PDN GW de-tunnels traffic coming from the extension tunnel, maps it to the appropriate S5 tunnel and forwards it to the Serving GW.

3. If the UE is not in active mode: the network triggered Service Request procedure (including paging) is initiated.

4. The UE is in connected mode now, and uplink and downlink data can be transferred between a PDN and the UE:

   4a. via L-GW and this (Home)eNodeB directly if the UE is connected via the (Home)eNodeB supporting the LIPA/SIPTO;

   4b. through the extension tunnel, via PDN GW, Serving GW and any other (Home)eNodeB if the UE is not connected via the (Home)eNodeB supporting LIPA/SIPTO.

6 Evaluation

Editor's Note: This clause is to discuss and evaluate the architecture solutions and key architectural aspects common to different solutions.

6.1 Evaluation of GW Selection mechanism

6.1.1 General

There are so far two main approaches for GW selection, they are described further. These approaches may be applied to multiple of Architecture alternatives described in clause 5.

NOTE: Additional selection mechanisms may be included as work progresses. Applicability for the GW selection mechanism may vary depending on the architecture solution and thus need to be evaluated accordingly.

The GW selection mechanism described here does not apply to TOF based option described in clause 5.5.

GW@ suggested by RAN node: This approach is applicable to select either a GW above the RAN node based on the UE's current location, or a GW co-located with the RAN node. The advantage is that it is a simple concept that can cover both usage scenarios with the same solution. The main disadvantages are that this would present a deviation from current (Release 8) DNS based GW selection which might present an operational burden, and would require additional RAN node configuration that also limits its applicability to the cases only when the RAN node is upgraded/new.

DNS based selection: This approach is applicable for selecting a GW above the RAN node based on the UE's current location, or selecting a GW co-located with the RAN node. The advantage is that DNS based selection is aligned with current system behaviour and this approach is compatible with existing S1/Iu/Iuh specifications and hence it can co-exist even with legacy nodes. The DNS system is also very flexible for future enhancements should new requirements emerge.

6.1.2 Scenario 1: GW close to the UE's point of attachment

In this SIPTO for Macro Access networks scenario, the MME/SGSN selects a GW that is geographically (and topologically) close to the UE's point of attachment to the network. This means that the GW selection takes into account the UE's current location.

Base Solution : Release-8 DNS

The DNS based GW selection procedures as defined in TS 29.303 [8] for Release 8 already cater for TAC/RAC based GW selection. If the TAI/RAI granularity is seen as sufficient to base the GW selection on, and then there is no need to extend the selection to an even finer granularity (i.e., cell level) in case the GW is above the RAN node. Hence, Release 8 DNS mechanisms already specified can be used to perform location based GW selection. If a finer granularity is needed, an extension of the mechanism is needed.

There is still a use case though which requires special attention: the case of 3G access when EPC is not yet deployed by the operator. In that case, too, the release 8 DNS mechanisms are applicable. Nevertheless we also look at the case when the release 8 DNS procedures are not deployed. Without using the release 8 DNS mechanisms, there is no way currently to base the selection on the RAC.

Solution 1.A: GW@ suggested by RAN node

As proposed in clause 5.4 (Solution 3 – GGSN allocation to offload point), the RAN node (i.e., RNC or HNB or HNB GW) may suggest a GW address to the SGSN based on some local configuration. The SGSN can then select that address for SIPTO instead of using the regular DNS based GW selection mechanism. The same mechanisms can be applied towards E-UTRAN/EPC.

NOTE: applicability of this mechanism for E-UTRAN/Home eNB subsystem has not been described in the TR yet.

Advantages:

- Simple mechanism in concept.

Disadvantages:

- Extra parameter impacts on Iu/S1;
- Requires an RNC/eNB update;
- Deviation from existing DNS based GW selection scheme, which may pose an additional operational burden for the operator.

Solution 1.B: DNS based selection

With this solution, the SGSN prepends some location based information (e.g., the RAC or the RNC id) to the APN before making the DNS query for the GGSN selection. This would give a solution to make the GGSN selection RAC location dependent similar as for the release 8 DNS scheme, although the format of the DNS string would differ from the Release 8 scheme. The DNS system is configured with the proper mapping of the RAC location information to the

GWs where applicable, as desired by the operator. Based on this configuration, the DNS system provides a GW address to the SGSN taking the RAC location information into account. The same mechanisms are applied towards E-UTRAN/EPC.

Advantages:

- Simplified operation/management as DNS remains the single system for managing GW selection information;
- Similar GW selection handling in SGSN/MME for both SIPTO and for regular connections;
- No impact on Iu/Iuh, hence the solution is compatible with legacy RANs;
- This feature is forward compatible with the enhanced DNS selection mechanisms defined for release 8 and thereby simplifies future migration for the operator.

Disadvantages:

- If small RNCs are being used (e.g. RNC functionality integrated in the NB site) then RAC granularity might be insufficient. In this case SAI might need to be added to the DNS enquiry.

6.1.3 Scenario 2: GW co-located with HeNodeB or HNodeB

Applicability of the GW Selection using DNS for Home (e)NB Subsystem and the evaluation/comparison with GW selection using GW@ from RAN Home eNB Subsystem is FFS.

6.2 Evaluation of GW re-selection mechanism for SIPTO

6.2.1 General

During idle mode mobility procedures, reselection of a UE's offload point for SIPTO traffic that is geographically/topologically close to the user shall be possible. Two approaches are suggested to detect the need for reselection and inform the UE that it should re-initiate the PDN connection to the same APN.

During the re-establishment of the SIPTO traffic PDN connection, the MME/SGSN shall select a GW that is geographically/topologically close to the user.

The approaches listed here can apply to multiple architecture solutions which are multiple-PDN based.

MME/SGSN detecting the reselection: During the idle mode mobility procedures (e.g. TAU/RAU), the MME/SGSN can trigger the release of the SIPTO PDN connection based on certain conditions.

GW detecting the reselection: The GW can trigger the release of the SIPTO PDN connection according to GW configuration.

The UE may re-initiate the SIPTO PDN connection according to the information from the network and/or the UE's preference.

6.2.2 GW re-selection criterion

One alternative way that can be used by MME to decide when to perform GW re-selection is the following: when TAU happens, if the PDN connection is a SIPTO PDN connection, the MME will check whether the new TA and the old TA belong to the same group in the Table 6.2.2.1. If they belong to the same group (the same row in Table 6.2.2.1), there is no need to perform GW reselection. Otherwise, the MME will get the corresponding FQDN and will initiate DNS query to get the IP address of the LGW for GW reselection.

Table 6.2.2.1: The FQDN table for L-GW selection

APN	TAName	TAI List	FQDN
SIPTO_APN	TA_A	TAI_1, TAI_2, TAI_3, TAI_4	<TA_A>.<SIPTO_APN>.epc.<mnc>.<mcc>.3gppnetwork.org
SIPTO_APN	TA_B	TAI_5, TAI_6, TAI_7, TAI_8	<TA_B>.<SIPTO_APN>.epc.<mnc>.<mcc>.3gppnetwork.org
…	…	…	…

NOTE: The above policy table and reselection procedure are illustrated in the LTE case. It can be similarly implemented in UMTS.

7 Conclusions

7.1 Conclusion on SIPTO macro

For the support of SIPTO at or above the RAN, it has been concluded that:

- solution 5 described in clause 5.6 is to be included in normative specifications, with the impacts as described in clause 5.6.4.
- additions to RANAP interface described in clause 5.5.5 to enable solution 4 are to be included to normative specifications. No other normative specification work will be done in this area. Basic informative text to describe the motivation for these additions are expected to be included.

7.2 Conclusion on LIPA

7.2.1 Conclusion on the LIPA architecture

For the support of LIPA, solution 1 variant 1 (described in clause 5.2 and more specifically clause 5.2.3.1) is selected as the basis for LIPA to be included in normative specifications, supporting both a collocated and stand-alone L-GW as well as mobility.

The identified impacts common to collocated and stand-alone L-GW configurations are the following:

- a *LIPA_enabled* flag (per APN and per CSG) in the user's subscription data stored in the HSS/HLR and transferred to the MME/SGSN;
- SGSN/MME/HSS (Gr/S6d/S6a): transmission of the flag from HSS to SGSN/MME: adding the transmission of the *LIPA_enabled* flag from the HSS to the SGSN/MME over the Gr/S6d/S6a interface (Stage 3 only as this is included in the CSG subscription data);
- SGSN/MME: L-GW selection: algorithm for L-GW (GGSN/S-GW/P-GW) selection is enhanced to take in account the *LIPA_enabled* flag;
- (E-)RAB setup messages: addition of new correlation identifier (user plane L-GW TEID) for each (E-)RAB in the (E-)RAB to be Setup List;
- Adding the transmission of the IP address of the L-GW in UE-associated signalling in the uplink, or, alternatively, DNS-based L-GW selection;
- Possible Multicast support in the L-GW.

The impacts to support the stand-alone L-GW configuration are FFS.

7.2.2 Conclusion on the architecture for Rel-10

For the support of LIPA, solution 1 variant 1 (described in clause 5.2 and more specifically clause 5.2.3.1) is selected as the basis for LIPA to be included in normative specifications in Rel-10, supporting only a L-GW collocated with H(e)NodeB without mobility.

The identified impacts are the following:

- LIPA-related settings (per APN and per CSG, as well as per PLMN) in the user's subscription data stored in the HSS/HLR and transferred to the MME/SGSN;
- SGSN/MME/HSS (Gr/S6d/S6a): transmission of the LIPA-related information from HSS to SGSN/MME;
- SGSN/MME: L-GW selection: algorithm for L-GW (GGSN/S-GW/P-GW) selection is enhanced to take in account the LIPA settings;
- (E-)RAB setup messages: addition of new correlation identifier for each (E-)RAB in the (E-)RAB to be Setup List;
- Adding the transmission of the L-GW IP address in UE-associated signalling in the uplink;
- Multicast support in the L-GW.

7.3 Conclusion on SIPTO in the Home (e)NodeB subsystem

The support of SIPTO in the Home (e)NodeB subsystem is FFS.

Annex A:
Evaluation of methods for operator control of SIPTO traffic

A.1 General

There are different views on the following aspects when determining SIPTO handling. This clause documents different aspects on these issues in an attempt to clarify the issues.

- At what granularity shall SIPTO control be performed?
- Where SIPTO routing policies shall be enforced?
- How shall SIPTO routing policies be communicated/configured?

A.2 SIPTO traffic control granularity

Possible levels of SIPTO traffic control granularities are:

- *per APN:* All traffic of a certain APN (e.g. Internet) is subject to offload.
- *per application protocol:* Traffic associated with certain application protocols (e.g. identified based on transport protocol type and port number) is subject to offload.
- *per destination IP address:* Traffic is offloaded based on the destination IP address (range).

A.3 Enforcement of SIPTO Routing Policies

Depending on the SIPTO solution and need for enforcement of routing policies, it may be performed by different entities, for example:

- by the UE;
- by a network entity.

Annex B:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2011-03	SP-51	-	-	-	MCC Update to version 10.0.0 after TSG SA approval	2.0.0	10.0.0
2011-10	-	-	-	-	MCC correction to figure 5.2.3.1.2.1 (renamed Figure 5.2.3.1.2-1 and replaced dashed box between HeNB and L-GW.	10.0.0	10.0.1