

INTERNET-DRAFT
Intended Status: Informational
Expires: August 07, 2017

J. Mueller
AT&T Foundry
T. Herbert
Facebook
February 07, 2017

**Mobility Management Using Identifier Locator Addressing
draft-mueller-ila-mobility-03**

Abstract

This specification describes a new mobile network architecture which improves mobility management using Identifier Locator Addressing (ILA) in IPv6 for next generation mobile telecommunication networks such as 5G and Mobile Edge Clouds. Identifier-locator addressing differentiates between location and identity of an addressable network element, which can be a mobile device or a data center task. The approach presented in this draft enables mobility management on Layer 3, and provides a simplified, GTP-tunnel-free and more efficient architecture with less core network utilization compared to traditional architecture.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/1id-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1. Introduction and Problem Statement](#) [4](#)
- [1.1. Terminology](#) [5](#)
- [2. Related Work, Protocols and Concepts](#) [6](#)
- [2.1. Mobile IPv6](#) [6](#)
- [2.2. Proxy Mobile IPv6 \(PMIPv6\)](#) [7](#)
- [2.3. Host Identity Protocol \(HIP\)](#) [7](#)
- [2.4. Locator/ID Separation Protocol \(LISP\)](#) [7](#)
- [2.5. Identifier-Locator Addressing \(ILA\)](#) [8](#)
- [2.6. Comparison of ILA to alternative approaches](#) [8](#)
- [2.6.1. Identifier Locator Network Protocol \(ILNP\)](#) [8](#)
- [2.6.2. Locator Identifier Separation Protocol](#) [8](#)
- [3. Mobility Management Architectures Using ILA](#) [10](#)
- [3.1. Address format for ILA mobile](#) [10](#)
- 3.2. Architecture with Functional Elements and Reference Points [10](#)
- [3.3. Functional Elements](#) [12](#)
- [3.4. Signaling and Data Flows](#) [14](#)
- [3.5.1. Provisioning](#) [14](#)
- [3.5.2. Attachment](#) [14](#)
- 3.5.3. Communication Scenarios for End-to-End Data Transport Sessions [15](#)
- [3.5.4. Homogeneous Handover](#) [20](#)
- [3.5.5. Heterogeneous Handover](#) [21](#)
- [3.5.6. Detachment](#) [22](#)
- [3.5.6. Idle-mode and paging](#) [22](#)
- [4. Discussion, Evaluation and Summary](#) [22](#)
- [5. References](#) [23](#)
- [5.1. Normative References](#) [23](#)
- [5.2. Informative References](#) [23](#)
- Authors' Addresses [24](#)

Mueller, Herbert

Expires August 07. 2017

[Page 2]

INTERNET DRAFT

<[draft-mueller-ila-mobility](#)>

1. Introduction and Problem Statement

The Internet Protocol (IP) has been overloaded in its functionality in the sense that it has been used as service locator and service identifier at the same time. Since changes of the associated IP address during a connection-oriented TCP session causes service interruptions, mobility has become a challenge. Mobility has been a challenge for IP based network since the area of smart phones began and has been addressed with Layer 2 and Layer 3 tunneling solutions. More specifically, uplink and downlink data traffic is encapsulated with GPRS-Tunneling-Protocol (GTP) headers, which are routed through the core network between the service and the base stations. One big challenge of client mobility is to ensure seamless and transparent mobility (e.g. IP address preservation) for mobile devices among different geographical locations and in between several Radio Access Technologies. Due to the deployment of micro-service architectures, another dimension in the complexity of mobility occurs. Single IP addressable tasks might change their physical location within a (virtualized) data center architecture as well. Therefore, mobility on both ends of the End-to-End (E2E) connection can be observed. Hereby, mobility requires a large number of service registry (e.g. DNS) updates and the state synchronizations between registries perhaps located in different (geographical) locations. In regard to current research and development on next generation mobile broadband networks (such as Mobile Edge Cloud and 5G), key requirements such as high-availability, low-latency and ultra-high-bandwidth are required to ensure the reachability of the massive number of communicating instances including from cellular communications, high-definition multimedia streaming, Internet-of-Things (IoT), critical infrastructures, etc.

This specification describes a new mobile network architecture which improves mobility management using Identifier Locator Addressing ILA) ([nvo3]) in IPv6 for next generation (virtualized) fixed and mobile telecommunication networks such as 5G and Mobile Edge Clouds. ILA shares many properties of the Identifier-Locator Network Protocol (ILNP) ([RFC6740], [RFC6741]) which is also a protocol that perform identifier locator split in IPv6 addresses. Identifier-locator addressing differentiates between location and identity of a addressable network element, which can be a mobile device or a data center task. A network architecture aligned on 3GPP is presented which evolves existing policy, charging and mobility concepts and substitutes GTP tunnels with ILA and therefore, improves mobility, latency, and service placement closer to the edge of the network.

The key advantages of the presented ILA mobility solution are:

- 1) Backwards-compatibility within existing IPv6 network

Mueller, Herbert

Expires August 07. 2017

[Page 4]

- architectures such as the AT&T network,
- 2) Enablement of ultra-low-latency Mobile-Edge-Cloud (MEC) services by locating services closer at the network edge,
 - 3) GTP-Tunnel-free and flatter architecture with less protocol overhead and less hops on the end-to-end path,
 - 4) Reduce complexity by merging data plane gateways into a single gateway,
 - 5) Proven applicability of ILA within the Facebook data centers and related networks at scale.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

The following terminology will be referred to in the document.

- * SIR: As defined in ([\[nvo3ila\]](#)): "In order to maintain compatibility with existing networking stacks and applications, identifiers are encoded in IPv6 addresses using a standard identifier representation (SIR) address. A SIR address is a combination of a prefix which occupies what would be the locator portion of an ILA address, and the identifier in its usual location."
- * SIR: Standard Interface Representation.
- * SIR Prefix: A sixty-four bit network prefix used to identify a SIR address.
- * SIR Address: An IPv6 address composed of a SIR prefix (upper sixty-four bits) and an identifier (lower sixty-four bits). SIR addresses are visible to applications and provide a means to address nodes independent of their location.
- * ILA Identifier (ID): An identifier that identifies an addressable element in the network independent of its location and type. ILA identifiers are sixty-four bit values. An ID can be generated on a per session basis with the effect to secure the privacy of the end point. The International Mobile Subscriber Identity (IMSI) or the International Mobile Equipment Identity (IMEI) can be used for ID generation. The ID is comparable with the Globally Unique Temporary UE Identity (GUTI) in the mobile space.
- * ILA Locator (LOC): A network prefix that routes to a physical host. Locators provide the topological location of an addressed node. The locator is represented in the prefix of the SIR address.

- * ILA Host: An end host that is capable of performing ILA translations for both sending and receiving. An ILA host uses the ILA resolver protocol to get identifier to locator mappings for destinations in communication.
- * ILA Router: A network device that performs ILA translations and packet forwarding. ILA router participate in distribution protocol mapping.
- * ILA Node: A network node capable of performing ILA translations. This can be an ILA router or ILA host.
- * User Equipment (UE): A device with an identifier such as a mobile phone, IoT gateway or another SIM equipped mobile device.
- * Access Point (AP) and Base Station (BS): A edge network element such as evolved-NodeB (eNB) in 4G that bridges the radio network and fixed network.
- * Gateway (GW): A central network element such as Serving-Gateway (SWG) or Packet-Data-Network-Gateway (PGW) in 4G.
- * Application Function (AF): The AF refers to the 3GPP terminology and stands for any IP addressable endpoint such as service or task.
- * EXA: An Internet routable prefix, may be use as a SIR

2. Related Work, Protocols and Concepts

This section provides an overview on the state-of-the-art on protocols and concepts for mobility management on mobile networks. In particular the 4th Generation (4G) of mobile telecommunication networks has been taken into account for this draft on functional and conceptual comparison.

2.1. Mobile IPv6

The IETF specified the Mobile IPv6 ([MIPv6]) protocol to ensure connectivity and reachability in case of client mobility within an IPv6 network. Mobility within a MIPv6 network is solved by assigning an additional IPv6 address - the Care-of-Address (CoA) - next to the current IPv6 address that as been assigned in the home network. Therefore a UE is equipped with a home address together with a primary CoA in case of foreign network attachment. IPv6 is classified as host-based mobility protocol, due to the fact that the UE is in charge of announcing its mobility to the network. In particular it is the client's responsibility for sending binding updates to the Home Agent (HA). In order to ensure reachability, the UE communicates its

new assigned CoA(s) to the HA, which acts as a router and registrar for UEs. Connection requests are intercepted and re-routed in case CoA entries for a UE exists. A tunnel is established between the UE at the CoA and the HA for securely exchanging packets. By default, the first packet is routed from the correspondent UE towards the CoA of the UE via the HA. This route is not always the shortest path. All consecutive packets of the same data stream will follow on the same path, which might include a detour, but hides the new location of the UE for privacy reasons. The feature of route optimization allows the UE to directly contact the correspondent UE, therefore cuts out the HA from the communication path and forwards packets on a shorter route. Security of the Mobile IPv6 is enhanced through IPsec for binding updates to avoid spoofing of CoA for a UE.

2.2. Proxy Mobile IPv6 (PMIPv6)

The IETF specified PMIPv6 ([\[PMIPv6\]](#)) provides network-based mobility management for UEs and extends the Mobile IPv6 in the way, that host-based mobility management functionalities in Mobile IPv6 are excluded from the client into the network. Hereby, the Local Mobility Anchor (LMA) acts as topological anchor point and manages the UE's binding state. The Mobile Access Gateway (MAG) manages mobility-related signaling on behalf of the UE at the access router. It is responsible for tracking the UE's movements to and from the access link for signaling the UE's local mobility anchor.

2.3. Host Identity Protocol (HIP)

HIP ([\[hip\]](#)) is providing a secure solution for identifier/locator-split by adding a new host identity layer into the protocol stack. A cryptographic namespace build upon a host identity as public key allows scalability and multi-homing within the network. An extension of DNS supports rendezvous server functionality for secure host identity lookup. A secure channel is establishment over Diffie-Hellmann-key exchange between two communicating entities. The communication setup is considered as robust against DOS, due to a riddle solved at the requestor side. On the other side a high overhead for the secure communication establishment due to key exchange has to be taken into consideration. HIP requires an additional protocol layer between L2 and L3 for encapsulation.

2.4. Locator/ID Separation Protocol (LISP)

LISP is a network-layer-based protocol that enables separation of IP addresses into two new numbering spaces: Endpoint Identifiers (EIDs) and Routing Locators (RLOCs). Tunnel router encapsulates and encapsulates packets.

2.5. Identifier-Locator Addressing (ILA)

ILA is outlined in detail in ([nvo3]), ([nvo3ila]) as well as in this document. In a nutshell, the concept of ILA splits IPv6 addresses into a locator and an identifier, eliminates the need for tunneling and therefore reduces the header size. ILA routers create and maintain a mapping table of identifiers of locators.

2.6. Comparison of ILA to alternative approaches

This section compares the ILA approach to some alternatives that have been discussed in 5gangip list.

2.6.1. Identifier Locator Network Protocol (ILNP)

ILNP ([rfc6741]) is an experimental protocol that splits and IPv6 address into a locator and identifier. ILA is fundamentally based on ILNP.

The key differences between ILA and ILNP are:

- * ILNP requires changes to the transport layer. This limits ILNP to be used only on hosts and every transport protocol implementation would need to be modified to use ILNP. Presumably to overcome the limitation above, some sort of ILNP proxy could be defined to perform ILNP in a middle-box.
- * ILA does not require changes to the transport layer.
- * Checksum neutral translation means that transport layer does not need to be parsed to perform ILA. This also ensures that existing device offloads (like checksum offload) work seamlessly.
- * ILNP employs IPv6 extension headers which are mostly considered non-deployable. ILA does not use these.
- * Core support for ILA is in upstream Linux, to date there is no publicly available source code for ILNP.
- * ILNP involves DNS to distribute mapping information, ILA assumes mapping information is not part of naming.

2.6.2. Locator Identifier Separation Protocol

Locator Identifier Separation Protocol (LISP ([rfc6830])) is an IP encapsulation protocol where the destination address in the outer IP header is a locator and the destination address in the inner header

is an identifier.

The key differences between ILA and LISP are:

- * ILA is not encapsulation so there is no associated encapsulation overhead. For instance IPv6/IPv6 in LISP would have fifty-six bytes of overhead whereas ILA translation has zero.
- * LISP may not work with some network device offloads whereas ILA works with all stateless offloads (ILA is transparent to the network so that it would just see TCP/IP packets for instance).
- * ILA has been accepted into Linux, LISP has not been accepted.
- * ILA can run either on end hosts (ILA hosts) or in the network (ILA routers). ILA maintain a cache of identifier to locator mappings.
- * ILA defines locators and identifiers to be sixty-four bits whereas LISP allows them to be full 128 bit addresses increasing the memory needed in mapping table.
- * The process of ILA translation is much more efficient than performing LISP. The translation path is:
 - 1) Parse IP header and extract the destination address
 - 2) Lookup destination in a hash table (obviated with cached route for ILA hosts)
 - 3) Write new destination address (16 byte copy)
 - 4) Forward to new destination (or receive at final destination).
 - 5) At the final destination, a reverse translation is performed to restore the originally sent address.

LISP processing is more involved. To do encapsulation:

- 1) An outer IP header, UDP header and LISP header need to be inserted in the packet. Tunnel fragmentation and MTU need to be considered [RFCXXXX] (i.e. increasing the size of a packet may exceed tunnel MTU).
- 2) At the remote tunnel end point, the outer IP header must be validated and a lookup done on the destination address to see if it is a local address. A lookup must be done on the destination UDP port to find that it is a LISP port. If the

UDP checksum is not zero that must also be validated.

- 3) The LISP header must is processed and validated.
- 4) Once the encapsulation is verified, the headers are removed and the inner packet is either forwarded or received.

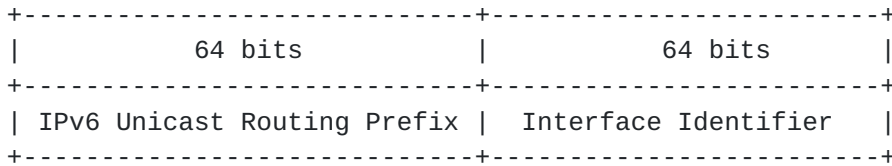
3. Mobility Management Architectures Using ILA

This section outlines the ILA protocol structure and architecture supporting ILA in mobile networks. The main functional blocks for connectivity, mobility support, security and charging are presented. Message flows for basic use cases executed by the mobile UE such as attachment, data transport with session handover and detachment are outlined.

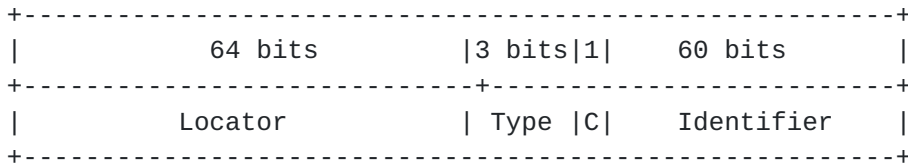
3.1. Address format for ILA mobile

The address format is derived out of the ILA draft in ([nvo3]) and is used without modifications in this draft.

The IPv6 canonical address format is:



The address format using ILA is:



The C bit is used to indicate that checksum-neutral mapping has been performed ([nvo3]).

3.2. Architecture with Functional Elements and Reference Points

The presented architecture in Fig 1 is aligned on the 3GPP Evolved Packet System (EPS) ([23401], [23402]) following the separation of control plane and data plane. Whereas 3GPP EPS addresses mobility through Layer 3 tunneling with GTP, this approach provides a Layer 3 mobility approach utilizing the ILA concepts for mobility without tunnels.

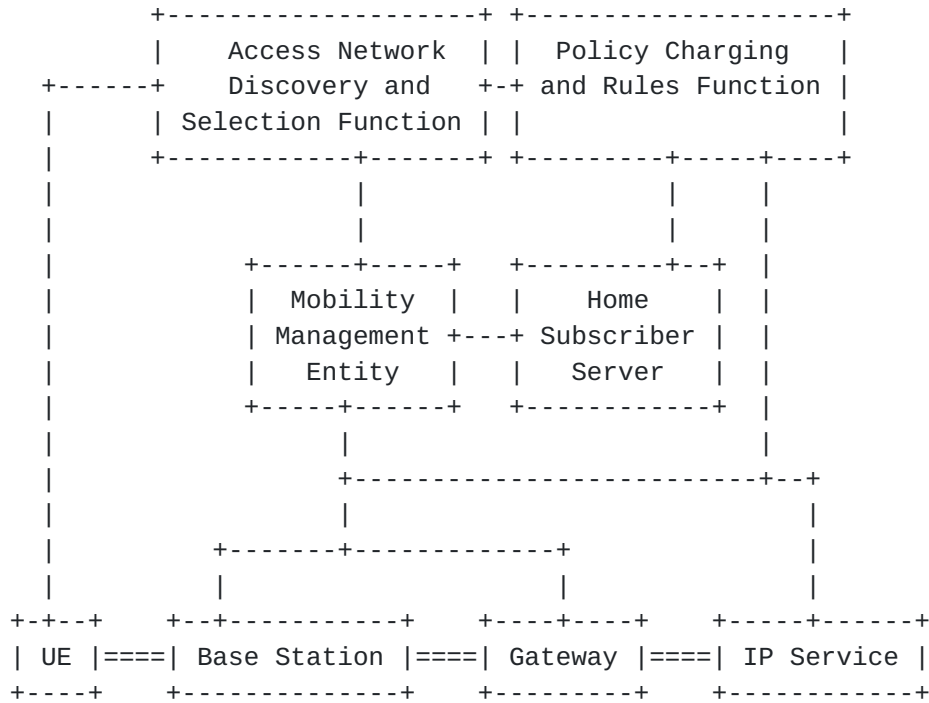


Fig 1: ILA-Based Architecture for Improved Mobility

Similarities between this new mobile network architecture and the 3GPP EPS are:

- 1) Split control and data plane.
- 2) Policy Control and Charging Rules Function (PCRF) architecture and interfaces.
- 3) Network-based mobility management using Access Network Discovery and Selection Function (ANDSF).
- 4) Home Subscriber Service (HSS) for managing and control dynamic and static user profile information.

Differences between this new mobile network architecture and the 3GPP EPS are:

- 1) Combined gateways and aggregated data plane functions within a single gateway.
- 2) Interfaces between the PCRF and ANDSF, HSS.
- 3) Enhanced mobility support as further outlined in 3.5.3. Communication Scenarios for End-to-End Data Transport Sessions.
- 4) Localized traffic handling within the Base Station to transport traffic among associated clients without core network involvement.
- 5) The associated network address is an ILA IPv6 address



Mueller, Herbert

Expires August 07. 2017

[Page 11]

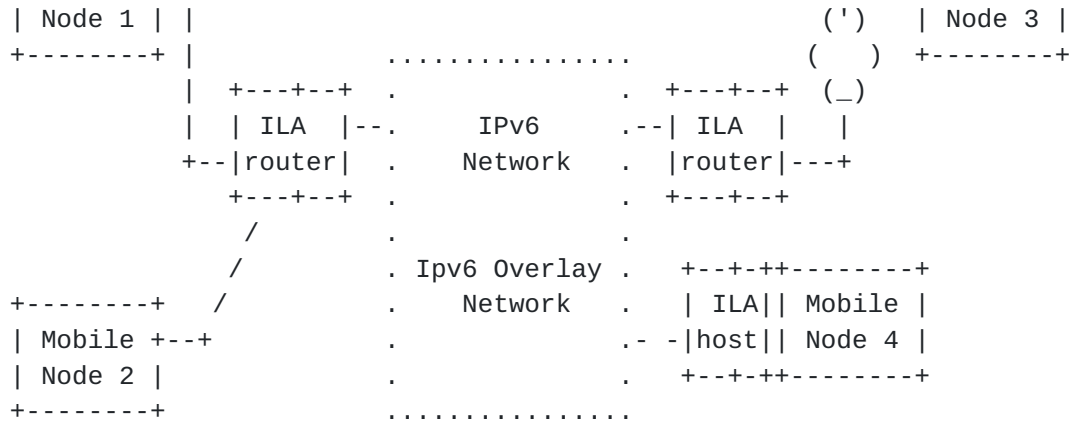


Fig 2: Distributed ILA Network Architecture [[nvo3ila](#)]

3.3. Functional Elements

This subsection summarizes the key functional elements of the ILA based mobility architecture.

* The User Equipment (UE) is the SIM enabled mobile device (cellular, gateway, etc.) executing services such as apps on the device, binding apps to the ID as communication endpoints, handling the bindings of all associated LOC/ID's and performing mobility as described below. The UE performs security related functions via its (embedded) SIM handling at least one or multiple identifiers provisioned by one or multiple network operators. Security related functions include authentication of the UE towards the network (more specifically the BS) and certificate management for establishing secure transport connections. Either the UE supports IPv6 or ILA for handling locator and ID bindings and updates or the network is handling ILA functionality on behalf of the UE. Storage and management of multiple locators for multi-path and multi-homing is supported by the UE.

* The Base Station (BS) or Access Point (AP) are the first point of contact from the UE when attaching over radio to the network. Its main purpose are routing, gating and forwarding data and control packets. The Radio Access Technology (RAT) is independent of the proposed concept and therefore out of scope of this document. 3G, 4G, 5G or WiFi are applicable RATs for the presented architecture. The BS is also capable of caching of content and state as close as possible to the user at the edge of the network. Another aspect of the cache is to support transparent handovers, during which buffering of packets at the target BS is required. Therefore, an X2-like connection between BSs is required. The BS supports a support a policy enforcement function (PEF) as well as a Event Reporting Function (ERF) aligned on the 3GPP defined Policy Control and Charging (PCC) functionality for the EPS in ([[23203](#)], [[29212](#)]).

Mueller, Herbert

Expires August 07. 2017

[Page 12]

Uplink QoS management is handled by the BS, too. In order to differentiate between multiple types of data traffic, signaling, high-priority, real-time and non-real-time connections can be distinguished and the order of packet processing in the BS can be influenced for uplink. The same concept applies for downlink in the GW. Forward Error Correction (FEC), IP header compression, encryption of user data stream are supported by the BS, too. Traffic filtering, gating, legal interception on the BS, to include the case, in which traffic re-routed only by the BS and is not traversing the GW.

* The Gateway (GW) encompasses data and control traffic related functions. Its main purpose is routing, gating and forwarding data and control packets. Therefore functionalities such as downlink QoS enforcement, APN management and charging is performed by each GW.

* The Application Function (AF) or IP Service are an examples of any IP addressable service in network. Other than in the 3GPP defined architecture, the IP service does not need to reside in the SGi LAN reachable only after terminating the GTP tunnel in the PGW. Furthermore services can be reached directly after the RAN connection is terminated within the BS.

* The Mobility Management Entity (MME) handles the initial authentication, authorization and mobility management of UE's over the control plane. The MME is responsible for tracking the UE's mobility and is in charge for updating the registries with near real-time status updates for LOC/ID mapping. ID and LOC assignment are performed by the MME.

* The Home Subscriber Server (HSS) stores and manages user profile information. These include the static information such as the assigned ID, security credentials as well as dynamic information LOC and the current Tracking Area.

* The Policy Charging and Rules Function (PCRF) controls data flows in the network architecture according to pre-defined rules. Such rules can be created by the network operator such as rate limiting or traffic shaping. Other rules differentiate between class of services for various traffic flow types identified on their Traffic Flow Template (TFT) characteristics such as source, destination, port and protocol information. The PCRF is handling charging for traffic flows using online (pre-paid) and offline (post-paid) charging. Both charging modes include a modes of operations with metrics such as service invocations, online time, data transferred, or no-charging. Out of credit events may influence the current connectivity for online charging, whereas offline charging is accumulating charging records which are usually processed in a monthly period.

Mueller, Herbert

Expires August 07. 2017

[Page 13]

* The Access Network Discovery and Selection Function (ANDSF) is an operator controlled database used for mapping the user location with available (radio) access networks. With this information, the ANDSF is capable of signaling suggestions for handovers to UE's. A UE is therefore able to operate only on one interface at a time to save resources. In case of the availability of adjacent RAT and after reception of a handover suggestion from the ANDSF, the UE is able to enable the suggested interface, perform a scan and finally decide whether or not to attach to the new targeted RAT. The database can be filled using device monitoring/telemetry statistics signaled from the UE to the network or by active measurements of the environment.

3.4. Signaling and Data Flows

3.5.1. Provisioning

A Subscriber Identity Module (SIM)-card is provisioned by the network operator with a unique and secure identifier that is comparable to the IMSI in 3GPP telco architectures (2G, 3G and 4G). This draft does not differentiate between a physical or an embedded SIM. In addition, security credentials and preferred network identifier are provisioned for authentication as well as network selection are provisioned. The matching information to the SIM card is stored in the HSS.

3.5.2. Attachment

After powering on the device, a scan for available radio networks is performed on the device, which selects the initial network (e.g. with the strongest signal) and performs a network attachment procedure aligned on ([[23401](#)], [[23401](#)]) towards the BS using security parameters, ID, last MME associated with (GUMMEI) and last GUTI assigned by MME with ID GUMMEI - the Packet Temporary International Mobile Subscriber Identity (M-TSMI). A secure identifier on the SIM is used to generate a temporary ID (the ILA ID), which is only valid for one session, hides the privacy of the UE in the network and unambiguously identifies the UE within the global network. This ID is used for identification, authentication, authorization and charging purposes.

For each network attachment, and due to privacy concerns for not revealing the identify of the UE towards the public, a new unique and temporary ID is generated. This ID is valid for a single session and is renewed afterwards.

The BS derives the last MME association out of the network attachment request sent by the UE and queries the last or a new MME based on availability of information for UE authentication. The MME performs a lookup in the user database of the network operator, which is the

Home Subscriber Server (HSS) and/or Home Location Register (HLR) and receives a profile in return. Hereby the MME is able to query the mapping database for existing mappings or to retrieve a unique ID for the UE.

In the following, the MME selects and configures the BS and GW according to the profile received and signals the profile including the ID towards the BSs of a certain tracking area and GW.

The BS allocates a LOC for the UE, binds the ID-LOC combination locally in a cache, publishes its binding in the MME/NVA and signals the ID-LOC towards the client.

Quality of Service (QoS) and charging related policies are installed in the BS and GW. The BS handled uplink and the GW downlink related traffic shaping functions. Charging can be performed in both functional elements (BS or GW), whereas a centralized charging in case of multi-path streaming is preferred.

After the successful attachment, a service can be invoked.

3.5.3. Communication Scenarios for End-to-End Data Transport Sessions

After the successful attachment, applications can start communicating in the network using its assigned ILA by constructing IPv6 packets with the SIR source information (ID+LOC) and the mandatory target ID. The target LOC can be either set directly or can be defined as a broadcast message, in which the target LOC will be determined at the edge of the target.

The following main high level use cases have been defined. The use cases can be distinguished into the following cases:

- 1) UE accessing a service in the AF,
- 2) UE is communicating with another UE attached to a different base station via a gateway,
- 3) UE is communicating with another UE attached to a different base station,
- 4) UE is communicating with another UE attached to the same base station,
- 5) Mobile Edge Cloud for localized traffic handling and low-latency communication,
- 6) Gateway mobility for enhanced mobility use cases

The example use cases are outlined below in details and the differences compared to today's networks are discussed. Fig. 3 depicts the network elements and control and data flows related to those use cases. The communication form can be multicast, broadcast, anycast or unicast.

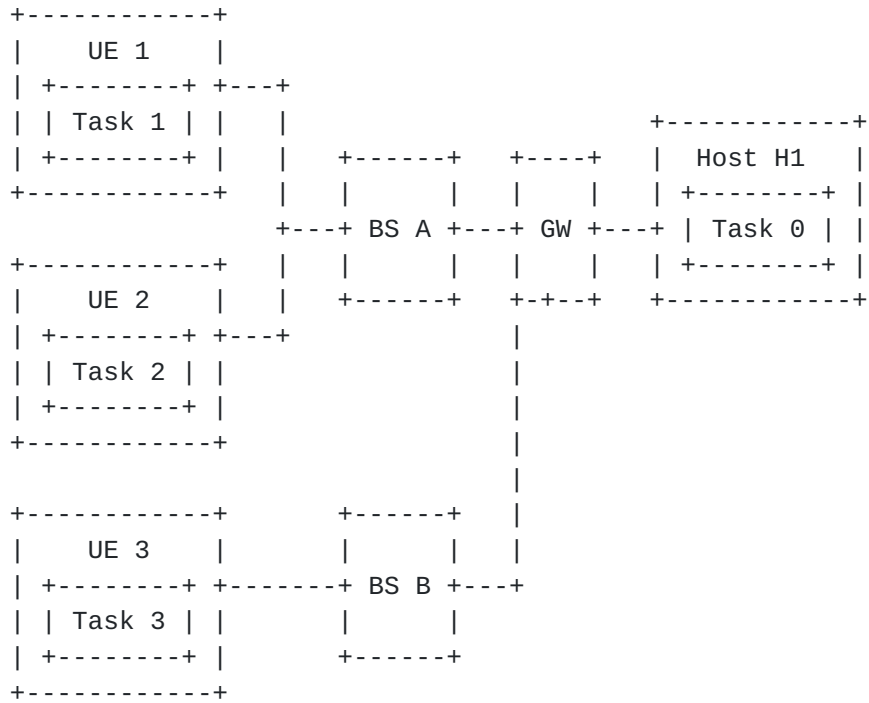


Fig 3: UE attachment over multiple base stations

1) E2E connection between the UE to AF (Task to Internet)

Considering a communication scenario in which a UE (source) queries a web site (target) e.g. "http://about.att.com/innovation/foundry" in a browser represented by T1.

```

SIR:T1,Iaddr -> // Transport endpoints at T1
L1:T1,Iaddr -> // On the wire in data center
EXA:T1,Iaddr // In the Internet

```

The request is forwarded to the BS, which performs ILA router functionality. In case a broadcast address has been selected as a target LOC, a cache lookup in a local lookup table is performed. Depending on finding an entry in the local cached lookup table, the routing is influenced and the packet is redirected. Otherwise the packet is routed on to the destination ILA SIR address (LOC/ID).

2) UE 1 to UE 3 are attached to distinct BSs via gateway

Considering a communication scenario in which one task (T1) mobile device of UE 1 is contacting a second task (T3) on mobile device of UE 3. Both UEs are connected to different BSs. ILA routing is done in the BS.

The transport endpoints for task to task between UE's communication are the SIR addresses for the tasks. When a packet is transported through the ILA network, the locators are set in source and destination addresses of the packet. On reception the source and destination addresses are converted back to SIR representations for processing at the transport layer.

If task T1 on UE 1 is communicating with task T3 on UE 3, the ILA translation sequence would be:

```
SIR:T1,SIR:T3 -> // Transport endpoints on T1
UE1:T1,UE3:T3 -> // ILA used on the wire
SIR:T1,SIR:T3   // Received at T3
```

3) UE 1 to UE 2 are attached to distinct but interconnected BSs An extension of the scenario will happen, when If task T1 on UE1 is communicating with task T2 on UE 2 via interconnected BSs, the ILA translation sequence would be:

```
SIR:T1,SIR:T2 -> // Transport endpoints on T1
UE1:T1,UE2:T2 -> // ILA used on the wire
SIR:T1,SIR:T2   // Received at T2
```

4) UE 1 to UE 2 attached to the same BS

Considering a communication scenario in which two communicating entities are attached to the same BS and therefore are in close proximity. The solution for routing traffic in today's network is the establishment of the data path from the UE over the access network (e.g. eNB) through the core network (e.g. EPC) into the AF (any IP addressable service or task) and backwards to the access network and finally terminated at the UE. Charging needs to be performed in the BS for this data flow. This communication pattern in today's networks creates a delay caused by the bearer concept of 3GPP network, which encapsulate and de-capsulate data in GTP-tunnels between the eNB and the PGW.

A practical use case is the communication between autonomous vehicles (e.g. self-driving cars or self-organized and autonomous drone swarms) through a telecommunication infrastructure. A very low delay is required for the interaction and precise management. In order to reach such a low delay, the communication needs to stay local in order to result in a low delay.

The presented solution on ILA mobility allows to keep traffic local for the case in which the communicating parties attached to the same BS.

Due to a lower amount of hops between UE 1 and UE 2, a lower latency can be achieved which results in a lower delay.

5) Low-latency Mobile Edge Cloud (MEC) Service

Considering a use case in which a UE is accessing a service with ultra-low latency requirements in the network such as image recognition, Augmented Reality (AR), Virtual Reality (VR) or other 5G low-latency and/or high throughput services. Other examples include vehicle control (drone or fleet management, traffic information, robot or power grid). In order to provide a high quality of experience for the user and customer, latency in the communication between the mobile device and the service has to be reduced in order to achieve a lower delay. Where multimedia streaming has an acceptable latency requirement of ~100ms, ultra-low latency services have strict requirements on the communication with under ~10ms or even close to 1ms. Classic cloud approaches that concentrate services centralized in the network are not applicable for ultra-low delay services due to the fact, that E2E latency is even too high. Violations of latency requirements result in motion sickness for VR users, outdated traffic information for autonomous self-driving cars, accidents with robotics in factories and the development of a new type of MEC services is hindered.

Therefore, a request is created and addressed with the source LOC/ID and targeted towards the destination LOC/ID.

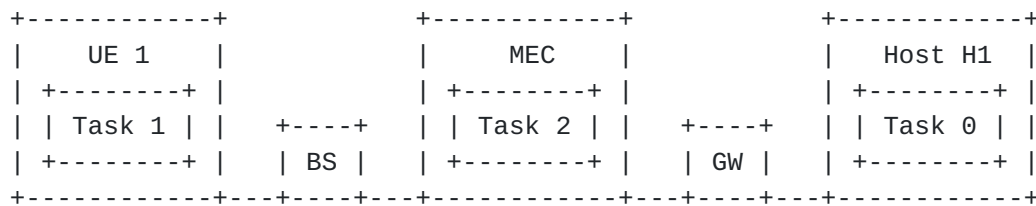


Fig 4: Mobile Edge Computing architecture with ILA

The innovative point in this use case depicted in Fig. 4 is the fact that the URL invocation may result in a redirect to a local service or content rather than a remote object. Hereby, the request may trigger a (third party) service or content deployment at the network edge instructed by the MEC_orchestrator and a policy decision. The policy decision is the outcome of the reasoning process within the MEC_orchestrator which takes context, user behavior, system load (throughput, latency, packet-loss, etc.), network topology map, distance between UE and service measured in hops, and other available metrics into account. Geographical load-balancing is therefore possible and enabled. Even when the first set packets of the connection are exchanged with a remote service over a longer

Mueller, Herbert

Expires August 07. 2017

[Page 18]

geographical network distance, a context handover during the active session away from the remote and towards a closer service instance (out of the same type or load-balancing group) can be applied.

In case service, task or content are available in MEC, ILA translation on the wire at the BS changes the locator to the closest point of presence. The following ILA steps are performed.

```
SIR:T1,Iaddr_T0 -> // Transport endpoints at T1 invokes Task 0 (T0)
L1:T1,Iaddr_T2 -> // Optional ILA translation from T0 to T2
EXA:T1,Iaddr      // In mobile edge cloud data center
```

Finally, the number of hops between UE and AF are reduced and a lower delay on the data path is achieved. Otherwise, in case the Edge Cloud capabilities cannot be utilized, basic routing is applied as outlined example 1) E2E connection between the UE to AF (Task to Internet) above.

6) Base Station or Gateway Mobility This use case covers situations in which the user stays connected to a BS but the core network is mobile and changes its location and connectivity to the service/Application Function. One example would be a BS that is attached to a vehicle (drone, car/bus, train, cargo ship, etc). The user facing side provides cellular service, backhaul is either WLAN, satellite, laser, MMWave or temporary a fixed connection. The AF facing side changes connection with each change of a transport connection such as WiFi, cellular or satellite.

Gateway mobility requires the update of forwarding entries in related BS and AF to continuously forward the packets on the data path.

The network setup is depicted in Fig.5 and Fig. 6 with both gateways (GW 1 and GW 2) both have connections to the same BS and AF.

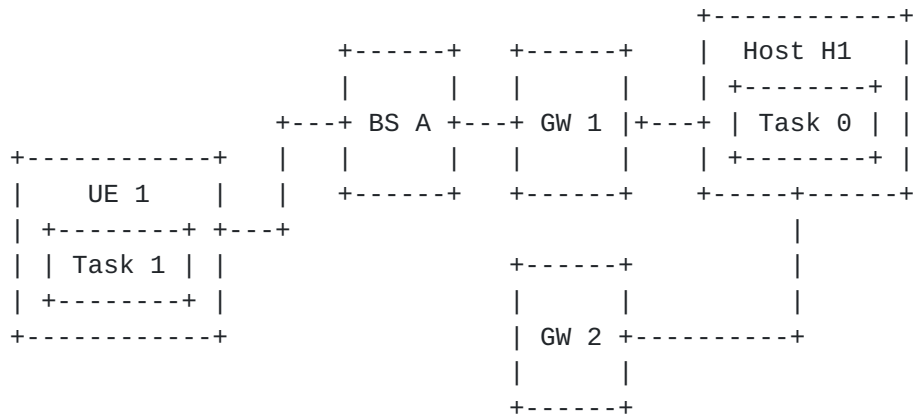


Fig 5: Gateway mobility with attached UE via GW 1

the BS due to emergency scenarios or higher priority calls, UE/BS/service load balancing or physical mobility of the UE among the network. Current resource utilization (e.g. data rates) of the UE or historical traffic pattern may influence the handover and the BS selection process.

Mainly the MME selects a target BS (BS_target) as target for the handover of the UE away from the current BS (BS_source). The decision is signaled to related BS's and the UE. BS_source starts de-allocating resource blocked by the UE and BS_target blocks resources required by the UE. Since most UE's are considered to have only a single RAT of each type (one WLAN or one LTE interface) an interruption in the connection while handover is to be expected. In order to avoid packet loss at the UE, buffering at the BS_target as well as packet forwarding from BS_source to BS_target are supported. Only after UE successfully establishes connectivity at the BS_target, previously blocked resources at BS_source are freed up, which are used as handover role-back in case of failure. Finally the MME announces the new ILA ID (BS_target_LOC)/ID for the UE as an update at GW and in the ILA registry.

New incoming connections are forwarded directly towards the UE over BS_target using the proclaimed ILA ID (LOC/ID).

Homogenous handovers with one radio technology interface supported have interruptions during the handover. Nevertheless those interruptions are relatively small due to techniques such as improved handovers in WiFi (802.11x, 802.11k, 802.11r, and 802.11v) or context handover via X2 in 4G.

3.5.5. Heterogeneous Handover

Client mobility may involve various Radio Access Technologies (RAT), in which the client is handed off from RAT_1 to RAT_2. The client is not required to move physically for heterogeneous mobility. Instead measurements on the UE or suggestion from the network (signaled over the ANDSF) may trigger handovers to alternative networks even when the UE is physically not moving. Such a handover can be done between WiFi, 4G and 5G.

Heterogeneous handovers are motivated for optimizing connectivity between UE and AF/service to move a multimedia connection with high bandwidth requirements from cellular (4G/5G) towards WLAN, a security sensitive bank transaction from WLAN towards cellular or simply a better transport-cost-per-bit-ratio.

Heterogeneous (compared to homogenous) handovers may be performed seamlessly with establishing a second alternative connection in

parallel to the existing active connection and tearing down the old connection only, after successfully establishing the new connection. Usually this is possible, because more than one interface is available on the client, which can be used in parallel to establish connectivity in parallel. In order to provide higher bandwidth over multi-path, both connections may be kept open in parallel. In this regard, the MME adds another LOC'/ID as update to the existing entry LOC/ID in the registry on the gateways and DNS.

3.5.6. Detachment

A detachment from the network can happen gracefully by shutting down the phone and de-registering it from the BS or suddenly due to a loss of connection. In both situations, a de-registration from the UE out of the list of active users attached to the BS is done directly or indirectly (after inactivity for a predefined time). Resource reservations are freed up again after detachment and opened up for other connections.

3.5.6. Idle-mode and paging

Power saving methods are working transparent to the ILA mobility concept such as described in ([[23401](#)], [[23402](#)]). The device toggles from active to inactive mode in idle-mode in order to reduce the communication interval between device and antenna. Resource reservations in the network are kept alive in order to allow a fast wake-up and connection re-establishment caused by paging of the BS towards the device.

4. Discussion, Evaluation and Summary

New low-delay services are appearing with AR/VR, drone communication, self-driving cars and robotic control that have new requirements on the network, which cannot be fulfilled by today's network and cloud architectures. New ultra-low latency is a key requirement on connectivity that is enabling a new services. One way of improving the End-to-End (E2E) connectivity is to substitute the underlying technology with a new generation and to improve the performance. Part of this improvement is described in Moore's-law, which highlights that the number of components per integrated circuit is doubling every 18 month. Another approach is to reduce the E2E latency by reducing the physical distance between device and service measured in number of hops and at the same time provide a backwards compatible solution for WiFi and 4G networks.

This draft is addressing the above mentioned challenges and provides a solution in form of a new mobile network architecture based on Identifier-Location Addressing (ILA) mobility. ILA decouples the

identity from locator within an IPv6 address. Therefore mobility can be achieved by preserving the same ID at the endpoint and only adapting the locator used routing in case of mobility. The implications are improved mobility with less control signaling and a more efficient tunnel-free core network architecture.

ILA applied on mobile networks and the gained improved mobility enables multiple new and innovative use cases compared to legacy telecommunication networks. Summarizing, the above presented Communication Scenarios for data transport for an End-to-End session outlines ways to improve connectivity, optimize routing and enable a new type of service: Mobile Edge Cloud services. Firstly, the improved data path requires less hops to traverse between UE <-> AF enabled by Mobile Edge Computing or locally between UE_1 <-> UE_2 due to the flatter architecture. Secondly, less overhead is created due to the reduction of GTP tunnels between network elements. Thirdly, the presented approach of ILA mobility is backwards compatible with today's IPv6 based fixed and mobile telecommunication networks.

5. References

5.1. Normative References

[rfc6741] Identifier-Locator Network Protocol (ILNP) Engineering Considerations, Jan 2013, <https://tools.ietf.org/html/rfc6741>

[nvo3ila] Identifier-locator addressing for network virtualization, [draft-herbert-nvo3-ila-02](#), Tom Herbert, Mar 2016, <https://tools.ietf.org/html/draft-herbert-nvo3-ila-02#page-17>

5.2. Informative References

[rfc6830] The Locator/ID Separation Protocol (LISP), D. Farinacci, Jan 2013

[MIPv6], Mobility Support in IPv6, C. Perkins, Ed. et al., Jul 2011, <https://tools.ietf.org/html/rfc6275>

[PMIPv6] S. Gundavelli, Ed. et al., Aug 2008, <https://tools.ietf.org/html/rfc5213>

[23401] 3GPP TS 23.401 V13.7.0 (2016-06), General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 13)

[23402] 3GPP TS 23.402 V14.0.0 (2016-06), Architecture enhancements for non-3GPP accesses (Release 14)

INTERNET DRAFT

<[draft-mueller-ila-mobility](#)>

[23203] 3GPP TS 23.203 V14.0.0 (2016-06), Policy and charging control architecture (Release 14)

[29212] 3GPP TS 29.212 V14.0.0 (2016-06), Policy and Charging Control (PCC); Reference points (Release 14)

Authors' Addresses

Dr.-Ing. Julius Mueller
260 Homer Ave
Palo Alto, CA 94301
US

E-Mail: jmu@att.com

and

Tom Herbert
Facebook
1 Hacker Way
Menlo Park, CA 94052
US

Email: tom@herbertland.com

