ETSI GR MEC 018 V1.1.1 (2017-10)

GROUP REPORT

Mobile Edge Computing (MEC); End to End Mobility Aspects

Reference
DGR/MEC-0018E2EMobility

Keywords
MEC, mobility

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Mobile Edge Computing (MEC).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document focuses on mobility support provided by Mobile Edge Computing. It documents mobility use cases and end to end information flows to support UE and Application mobility for Mobile Edge Computing. When necessary, the present document describes new mobile edge services or interfaces, as well as changes to existing mobile edge services or interfaces, data models, application rules and requirements. The present document identifies gaps to support mobility that are not covered by existing WIs, documents these gaps and recommends the necessary normative work to close these gaps.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI GS MEC 001: "Mobile Edge Computing (MEC) Terminology".

[i.2] ETSI GS MEC 003: "Mobile Edge Computing (MEC); Framework and Reference Architecture".

[i.3] ETSI GS MEC 013: "Mobile Edge Computing (MEC); Location API".

[i.4] ETSI GS MEC 010-2: "Mobile Edge Computing (MEC); Mobile Edge Management; Part 2: Application lifecycle, rules and requirements management".

[i.5] ETSI GS MEC 011: "Mobile Edge Computing (MEC); Mobile Edge Platform Application Enablement".

[i.6] ETSI GS MEC 012: "Mobile Edge Computing (MEC); Radio Network Information API".

[i.7] ETSI GS MEC 002: "Mobile Edge Computing (MEC); Technical Requirements".

[i.8] 3GPP TS 23.501: "System Architecture for the 5G System; Stage 2".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in ETSI GS MEC 001 [i.1] and the following apply:

application instance: realized software program executed in mobile edge host, which can provide service to serve consumer(s)

application instance relocation: procedure of moving an application instance running on a mobile edge host to another mobile edge host, to support service continuity over underlying network

application instance state transfer: procedure of transferring the operational state of application instance from the source mobile edge host to the instance of the same application in the target host

application mobility: part of mobility procedure for mobile edge system

NOTE: It relocates an application dedicated to a service consumer or shared by multiple service consumers from one mobile edge host to another. Application mobility may include application instance relocation and/or application instance state transfer from one mobile edge host to another.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI GS MEC 001 [i.1] and the following apply:

App	Application
CPN	Connectivity Provider Network
DP	Data Plane
GW	Gateway
FFS	For Further Study
MAMS	ME Application Mobility Service
ME	Mobile Edge
MEH	Mobile Edge Host
MEO	Mobile Edge Orchestrator
MEP	Mobile Edge Platform
MEPM	Mobile Edge Platform Manager
PDN	Packet Data Network
RNIS	Radio Network Information Service
S-DP	Date Plane of Source MEH
S-MEH	Source MEH
S-MEP	MEP of Source MEH
T-DP	Data Plane of Target MEH
T-MEH	Target MEH
T-MEP	MEP of Target MEH
UE	User Equipment
VIM	Virtualization Infrastructure Manager
VM	Virtual Machine

4 Mobility requirements and use cases

4.1 Requirements and scenarios for mobility in ME system

4.1.1 Requirements for mobility in ME system

ME mobility is an important Mobile Edge Computing feature in a mobile environment, since UE mobility supported by the underlying network can result in UE moving to a network entity associated with a different ME host from the current serving ME host. ME system needs to support the following:

- continuity of the service;
- mobility of application (VM), i.e. relocation of application instance; and
- mobility of application-specific user-related information, i.e. transfer application instance state related to UE.

To support ME service continuity, ME application mobility may involve multiple ME functional entities in order to relocate application instances and transfer user and application-specific information within the ME system. Relocation decisions may be based on UE mobility, customer profiles and/or ME infrastructure capability. The requirements related to ME application mobility are listed as Mobility-01, 02, 03, Connectivity-01, 02, 03, 04, Routing-01, 02, 03, 04, 06, 10, 14, SmartReloc-01, 02, 03, 04,05, 06 in ETSI GS MEC 002 [i.7]. Table 4.1.1-1 summarizes the architecture requirements related to ME mobility in ETSI GS MEC 003 [i.2].

Table 4.1.1-1: Requirements for ME mobility

Numbering	Functional requirement description
Arch-01	The mobile edge system should support the mobility of application: • as a consequence of UE moving within the ME system; or • at a certain condition that requires to move the ME applications to different ME host.
Arch-02	The mobile edge system should support the ME service continuity of UE as the consequence of the application movement within the ME system, and support the mobility of application-specific and user-related information.
Arch-03	The mobile edge system should support application mobility for ME applications not sensitive to UE mobility. See note 1.
Arch-04	The mobile edge system should support application mobility for the ME applications sensitive to UE mobility: • Maintaining connectivity between UE and mobile edge application instance. • Application state relocation. • Application instance relocation within the mobile edge system.
Arch-05	The mobile edge system should support: • application instance or state relocation in MEC system. • application instance relocation between the mobile edge system and/or an external cloud environment. See note 2.
NOTE 1: UE mobility means IP session mobility supported by underlying network. NOTE 2: This requires FFS.	

4.1.2 Mobility scenarios in ME system

Mobility in ME system is concerned with service continuity when the service to UE is relocated to another ME host within the ME system. ME service relocation may be triggered by UE's bearer path change in underlying network, or MEC system optimization to reduce service latency to UE, and procedure of relocation depends on:

- topology of ME host deployment in underlying network;
- scope (to be defined below) of application instances being served to UE(s); and
- aspects of applications.

Scenario 1:

A UE moves in underlying network, but is still in the coverage of serving ME host, i.e. intra ME host mobility. In this scenario, the ME system does not need to relocate service (i.e. application instance being served to UE, and/or UE context) to keep service continuity.

Scenario 2:

A UE moves out of coverage area of source ME host to the coverage area of another ME host (target), i.e. inter ME host mobility. This scenario may result in interrupt of service to the UE. In order to provide service continuity to UE, the ME system needs to relocate the service to UE from source ME host to target ME host.

Relocation of service to UE may need to further consider application scope:

- Dedicated application: An ME application instance is dedicated to serving a specific UE. When the UE moves to another entity of underlying network which is associated to different ME host from the serving ME host, the ME application instance being served to the UE should be relocated to the new ME host from the current serving ME host.

- Shared application: An ME application instance at the serving ME host may not be dedicated to serving a specific UE. Instead, it may serve multiple UEs (such as multi-cast), or all UEs associated with the ME host (broadcast). When a UE moves to another entity of underlying network which is associated to a different ME host from the serving ME host, the ME application may not need to be instantiated at the target ME host, but require transfer the UE context to the application instance if it has been instantiated at the target ME host already. For example, a broadcast service may be provided by the shared application instance. When a UE subscribing to the broadcast service at the serving ME host moves to a new location in the underlying network, UE context related to the broadcast service is transferred from the serving ME host to the broadcast service at the new ME host so that the UE can continue being served by the broadcast service at the new location of underlying network. As the application instance at the source ME host may still be required to serve other UEs, the application instance at the source ME host is not torn down after the UE is served at the target ME host.

In addition, ME mobility also needs to consider aspects of the application instance being served to a UE:

- Stateless: A stateless application is an application that does not memorize the service state or recorded data about UE for use in the next service session; or

- Stateful: A stateful application is an application that can record the information about service state during a session change. The state information may be stored in the UE app or ME app instance in the serving ME host, which can be used to facilitate service continuity during the session transition.

ME mobility for stateless application does not require transferring UE state information to the application instance at target ME host, while ME mobility for stateful application does need to transfer UE context to support service continuity to UE.

Table 4.1.2-1 summarizes relocation of application instance and state information involved in different service mobility, application scopes and aspects.

Table 4.1.2-1: Application instance relocation and UE state information transfer for service continuity

Service Mobility	Scope	State	Application instance relocation for high service continuity	Application instance relocation for low service continuity
Intra ME host	Any	Any	No	No
Inter ME host	Dedicated	Stateless	App instance relocation	FFS
		Stateful	State transfer; and/or App instance relocation	FFS
	Shared	Stateless	App instance relocation (conditional)	FFS
		Stateful	State transfer; and/or App instance relocation (conditional)	FFS

4.2 Use case for optimization of application state relocation

Optimization based on user group

In a virtual reality multiplayer game, the mobile players are one by one moving away for current serving ME host. Ideally it would be to have all players on the same game level hosting on the same ME host. However this may impact latency. The players should be distributed in such a way that as many players as possible experience the required latency. The system may then group the players on to as few ME hosts as possible as long as the latency requirements are still fulfilled.

Optimal time window

A user playing a virtual reality game on a train is moving further away from the current ME host. The user context of the ME application needs to be relocated to a more optimal ME host. An optimal time window needs to be allocated to make the user context relocation with minimum impact on the QoE. Figure 4.2-1 shows an example of optimal time window for a latency sensitive application e.g. when the game is changing between different levels.

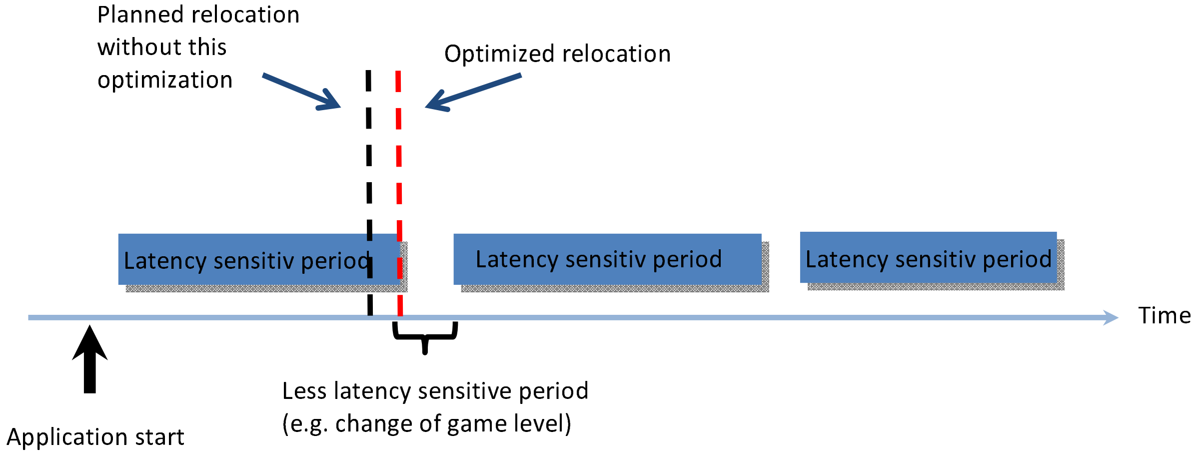

Figure 4.2-1: Optimal time window

4.3 Use case for prediction of relocation timing

Since user mobility in mobile systems is inevitable when a UE moves within a mobile network, the mobile edge host serving to the UE can be changed. If it is foreseen that the application can react to such handover events by application-specific means, or, if the optional SmartRelocation feature is supported, the mobile edge system could relocate the application instance serving the UE to the target host. Reducing the relocation failure rate will be the key to improving the quality of experience (QoE). Relocation failure has three components:

- too late relocation;
- too early relocation; and
- relocation to a wrong ME host.

Therefore, accurate prediction of the handover to the target ME host is an important issue in MEC.

EXAMPLE: If the UE is classified as having high mobility (e.g. connected car), the main concern is about the possibility of too late relocation due to the UE's high velocity. However, if UE mobility information is available, then the ME system can proactively predict the handover timing and guarantee seamless and smooth relocation with optimal ME host selection such that the UE can always receive maximum QoE. Figure 4.3-1 shows an example of the prediction of handover timing for the connected car use case. The transit time in each cell can be estimated by the assistance of the UE application (e.g. the car navigation system) or by a MEC-based solution. The Location Service may also support prediction of the handover timing by retrieving the location information of UEs and radio nodes (see ETSI GS MEC 013 [i.3]).

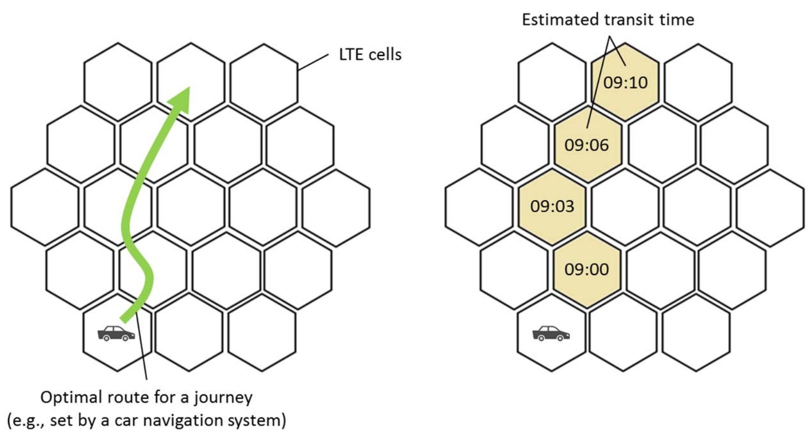

Figure 4.3-1: The prediction of handover timing for a connected car use case

In a make-before-break scenario, the relocation of application state information to another ME host is completed before the optimal ME host is changed, shown in figure 4.3-2. The ME system predicts the handover timing and informs the ME application, which initiates state relocation to the optimal ME host in advance. The aim of pre-relocation is mainly to reduce ME service's end to end delay and relocation delay during high mobility which can severely degrade ME service performance.

NOTE: The actual procedure of application state relocation is application-specific.

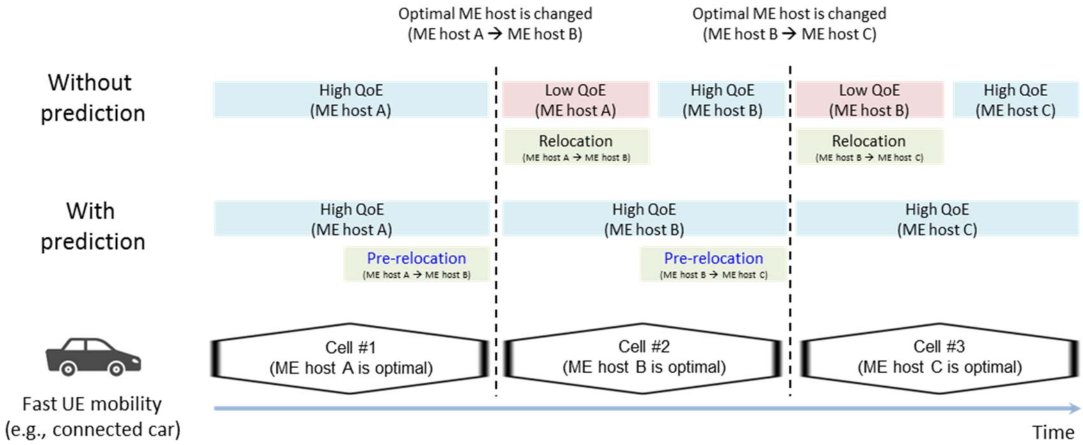

Figure 4.3-2: Pre-relocation of application state information

4.4 Use case for mission critical low latency application relocation

Mission critical low latency applications, such as Industrial IoT, Self-Driving Car, requires communication with very high reliability and availability, as well as very low end toend latency going down to millisecond level. In order to support very low latency, ME application is relocated close to the user as the user moves from one cell to another. The relocation process itself may have a negative effect on application latency, e.g. there may be a "period of time" between the attachment point handover and the ME host relocation, resulting in higher latency. To support critical low latency applications it is therefore necessary for the ME system to support a relocation process that keep handover-induced latency to a minimum.

A relocation process involves the following steps:

- Detect the need for relocation using radio network information.
- Identify target ME host.
- Move the application to the target ME host.
- Setup communication path.

Collecting the radio network information and processing it to complete the relocation process may not only require a considerable amount of time, but will typically not meet the requirements in the worst case scenarios, e.g. when prediction fails due to high UE manoeuvrability with regards to detection accuracy. Thus it becomes unsuitable for mission critical very low latency applications.

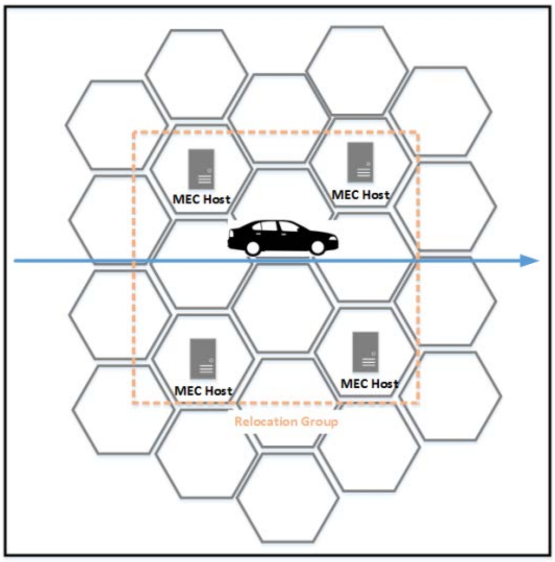

Figure 4.4-1: Preconfigured Relocation Group

In such cases, it may be appropriate to pre-configure a set of ME hosts, where the ME application is allowed to run as the user moves within those ME hosts. This set of ME hosts may be called a "relocation group". The relocation group may be created based on the topological or physical location of ME hosts with regards to the application end users. Users may influence the creation of a relocation group based on their QoE (Quality of Experience) and Security preferences. Otherwise, ME system may also choose the relocation group for a User and ME application based on subscription level and policy.

All ME hosts in the group may share user and application information, so that when user changes its attachment point, communications among ME hosts of the relocation group may be setup quickly. Additionally, the ME application may even start communicating with the UE before actual handover. As the user moves, ME system knows the list of target hosts. Depending on the latency and criticality requirement, it may relocate ME application instance or application state in advance to one or more hosts.

4.5 Use case for service continuity with the UE moves in/out of ME host serving area

When a UE (in active mode) moves out of the ME host serving area and no other ME host could provide the service, the UE should be provided with the service from the application server in SGi to maintain the service continuity and keep the UE IP address the same. When the UE moves out of the ME host serving area, the user plane could have the capability to retrieve the UE context from the ME host. When the UE moves in the ME host serving area, the control plane could determine whether the current service from SGi should be terminated and the UE context is retrieved from user plane to MEC host.

EXAMPLE: A user is watching videos on the mobile phone in a driving car and is moving out of the current ME host serving area. There is no other ME host providing video service outside of the current ME host. The video service should be terminated and the user context of the ME application needs to be relocated to user plane. Figure 4.5-1 shows an example of this use case.

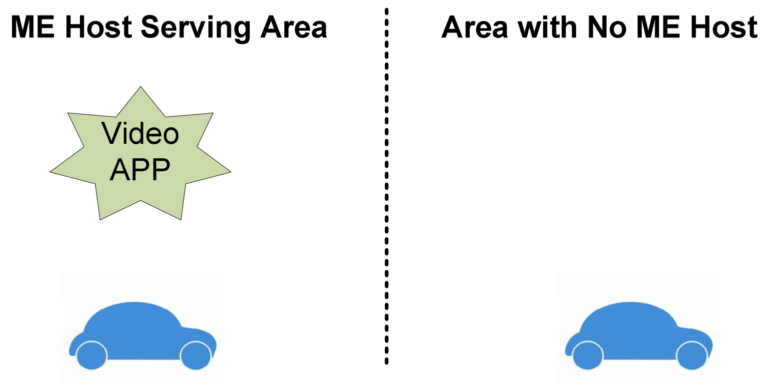

Figure 4.5-1: Use case of service continuity with the UE moves in/out of ME host serving area

4.6 Use case for initial or/and simple deployment

A budget constrained 3rd party MEC operator starts to roll out a new MEC network consisting of a few servers. The operator does not have an RNIS available. What this operator has is only the servers connected to several gateways of a mobile network or of a few networks of different technologies, based on a simple Service Level Agreement (SLA). The application services provided by this operator have a QoS target that deviates from that of other operators. It is expected that the design of MEC system provides means and options for such an operator to conduct business, without additional technical limitations.

4.7 GW based use case

In this scenario, the mobile edge host is co-located with gateway. See figure 4.7-1.

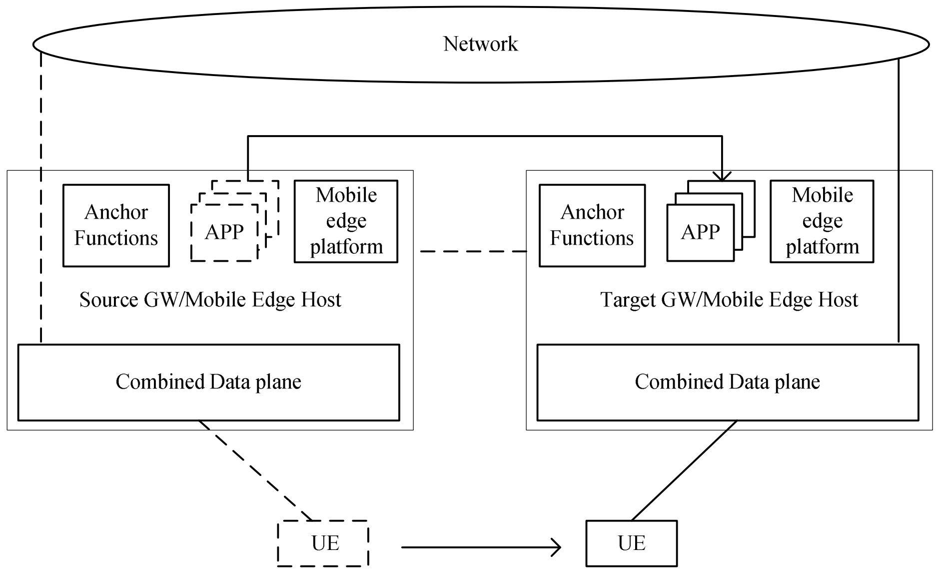

Figure 4.7-1: Mobility in GW-based Use Case

When a UE moves to the serving area of target mobile edge host, the relocation of the mobile edge applications will be performed for a guaranteed performance. As a consequence, the target mobile edge host will continually serve the UE instead of the source mobile edge host. Potential issues caused in this situation are listed below:

- How the GWs (i.e. mobile edge host) keep session continuity for the PDN sessions of the moving UE whose IP address changes after UE handover procedure.
- How to assist the mobile edge system to select the target mobile edge host with the information provided by GWs.

- How to optimize the application relocation with the mobile edge services, as well as the local context in GWs.
- How to maintain service continuity with upper layer mechanism if session continuity is not supported.
- How to reuse the Anchor functions provided by the GWs in mobile edge host.
- What information, favourable for (re)selection of user plane path and UE handover, can be fed back to GWs by the mobile edge system.

5 End to end information flows to support mobility

5.1 An overview

ME mobility may involve multiple network entities, functional modules and services in ME system. The overview of end to end message flows for support mobility provides a high level picture of relocating the service offered by an ME application instance to another ME host of ME system. It aims to decouple the complete message flow of ME mobility into one or more basic procedures which are defined in the present document or other MEC documents. Therefore ME mobility can reuse those existing procedures, APIs, message flow and data models for relocating ME applications or services as much as possible.

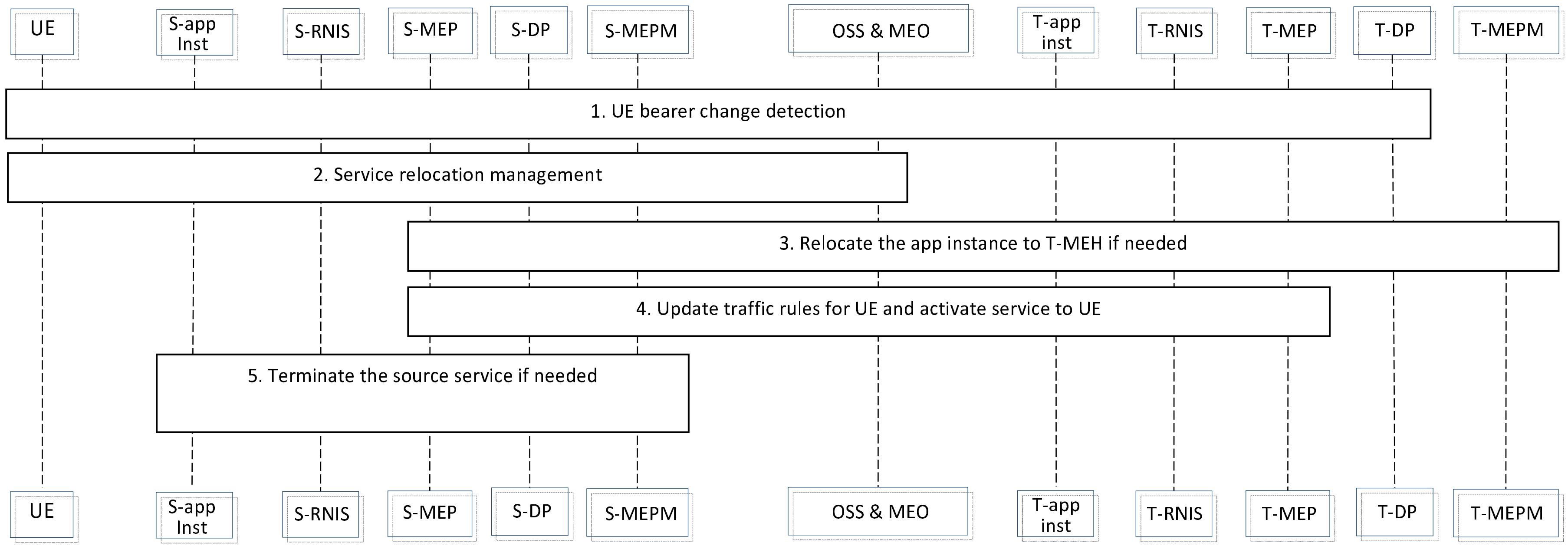

Figure 5.1-1: An overview of ME mobility flow

An ME mobility can start from an UE movement in underlying network. Figure 5.1-1 shows an example of ME mobility flow triggered by UE movement in underlying network, with separated sub-procedures:

1) UE bearer change detection. An UE moves in underlying network and results in its bearer path change. The bearer path change could cause:

   a) UE is still in the coverage of serving ME host; or

   b) UE moves out the coverage of serving ME host to the coverage of another ME host (target).

   The UE bearer change could be detected by:

   a) UE if its IP address is changed; or

   b) RNIS if the MEH is associated with the signalling of underlying network, like S1 interface;

   c) Data plane (DP) if a special traffic filter is applied to DP such as "end mark", source and/or destination IP address of data packet which belongs to other MEH, etc.

2) Service relocation management. In this step, the detector of UE bearer path change may:

   a) notifies MEO for UE assisted application instance relocation if UE detects its bearer path change;

   b) notify S-MEP if S-DP or S-RNIS detects the UE bearer path. S-MEP may communicate with MEO to relocate the application instance if it is required;

   c) notifies T-MEP if T-DP detects IP address of received data packet that belongs to other MEH. T-MEH may communicate with other MEH to get UE context related to received data packet. T-MEH may communicate with MEO for relocation of application instance associated to the UE if it is required.

   Service relocation management decides what kind of service relocation is required and how sub-procedures of service relocation is involved.

3) Application instance relocation. In this step, the application instance being served to UE is relocated from S-MEH to T-MEH, if needed.

4) Traffic rule update. This sub-procedure includes updating the traffic rules for the UE and activating the new service for UE at target ME host.

5) Termination of the service at S-MEH. This sub-procedure is to terminate the service and release resource associated to the application instance being served to UE at S-MEH if needed, depending on the application scope.

5.2 A flow for intra-ME host mobility

Intra-ME host application mobility is triggered by a UE moving in underlying network, but not resulting in a change and relocation of ME application instance that is serving to the UE, as the serving ME host may connect to multiple network entities (like eNBs) in underlying network.

Intra-ME host mobility does not require relocation of application instance or transfer UE state.

5.3 A flow for inter-ME host mobility

Inter-ME host mobility is caused by a UE moving to another network entity of underlying network that is not associated to the ME host being served to the UE. Therefore inter-ME host mobility may trigger:

- relocate the application instance being served to the UE from the serving ME host (source) to another ME host (target), depending the application scope;
- update traffic rules related to the UE;
- activate the service at target ME host to continue the service;
- terminate the service associated to UE at S-MEH, depending on the application scope.

Solutions for inter-ME host application mobility scenario may involve:

- update traffic rules at source and target ME host (refer to clause 6.2.6 Key Issue: Update traffic routing rules);
- application instance relocation (refer to clause 6.2.5 Key Issue: Application instance relocation);
- activate the service at T-MEH (to be investigated later);
- terminate the service produced by application instance at S-MEH if needed (refer to clause 5.4).

5.4 A flow for terminating an application instance

5.4.1 Terminate a dedicated service produced by application instance

A service produced by an application instance can be dedicated to a service consumer. In that case when the service has been moved to a new MEH and is running there, the service at the original MEH needs to be terminated and its associated resources will be released.

Figure 5.4.1-1 shows an example of message flow for terminating an application instance of dedicated service at S-MEH and releasing its associated resource.

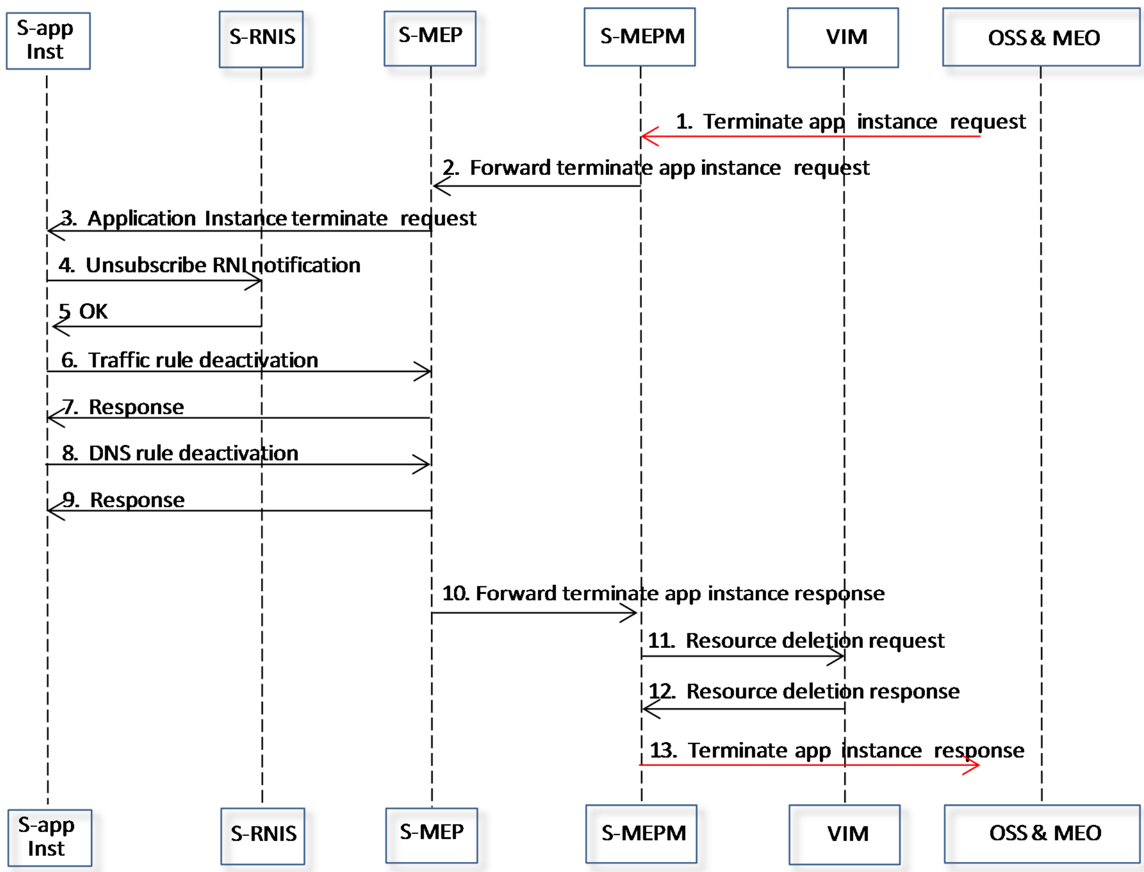

Figure 5.4.1-1: A message flow for terminating an application of dedicated service

1) The S-MEPM receives a terminate app instance request.

   The terminate app instance request can be from the MEO, or from T-MEP when the application instance being served to the UE has been relocated to T-MEH and is running there.

2) S-MEPM forwards terminate app instance request to S-MEP to start the procedure of terminating the application instance.

NOTE 1: The MEO can query the application instance information to get appInstanceId from MEPM. Refer to clause 5.4 of ETSI GS MEC 010-2 [i.4].

3) S-MEP sends an application instance terminate request to S-App for terminating the instance.

NOTE 2: Refer to clause 5.2.2 of ETSI GS MEC 011 [i.5].

4) S-App instance sends a unsubscribe RNI notification to S-RNIS.

NOTE 3: Refer to clause 6.2.5.4 of ETSI GS MEC 012 [i.6].

5) After successful unsubscribe the S-App instance, S-RNIS acknowledges to S-App instance.

6) S-App sends a traffic rule deactivation to S-MEP for deactivating traffic filtering service associated with the application instance.

NOTE 4: Refer to clause 5.2.6 of ETSI GS MEC 011 [i.5].

7) Once the S-MEP successfully deactivates the traffic rules for the application instance, it sends a response back to S-App instance.

8) S-App sends a DNS rule deactivation to S-MEP for deactivating DNS service associated with the application instance.

NOTE 5: Refer to clause 5.2.7 of ETSI GS MEC 011 [i.5].

9) Once S-MEP successfully deactivates DNS service for the application instance, it sends a response back to S-App instance.

10) Once the graceful termination timer expires, S-MEP can trigger the procedure of release resources associated with the application instance via sending a forwarding terminate app instance response to S-MEPM.

NOTE 6: Refer to clause 5.4 of ETSI GS MEC 010-2 [i.4].

11) S-MEPM then send a resource deletion request to VIM to remove the resource allocated to the application instance which was just deactivated.

12) VIM successfully releases the resource allocated to the deactivated application instance and sends a resource deletion response to S-MEPM.

13) S-MEPM sends a terminate app instance response back to the initiator of terminating application instance.

5.4.2 Terminate a shared service produced by application instance

A service produced by an application instance can be shared by multiple service consumers. In that case when the application instance being served to a consumer has been moved to a new MEH and is running there, it may need to decide whether or not to terminate the service at the source MEH and release its associated resources, depending on whether the service is still consumed by other consumers.

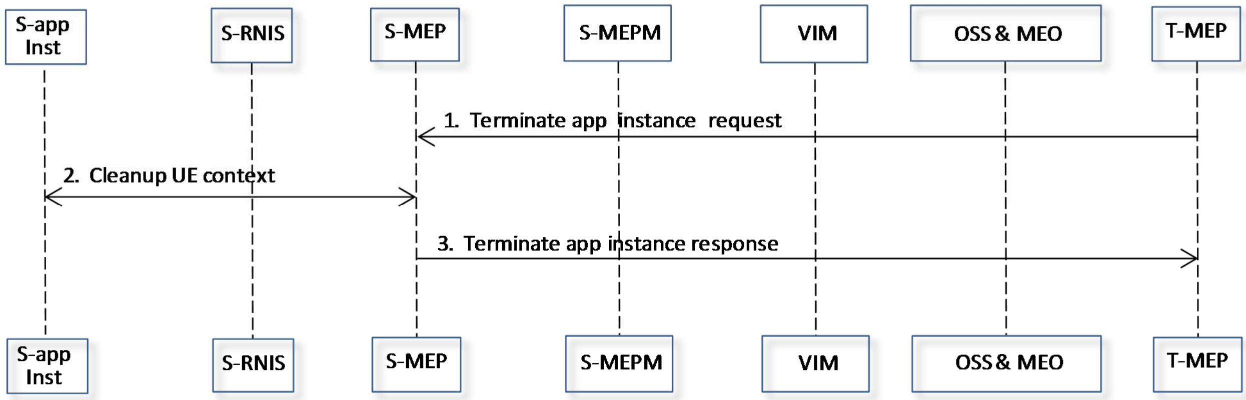

Figure 5.4.2-1: A message flow for terminating application instance of shared service triggered by T-MEP

Figure 5.4.2-1 shows the message flow for terminating application instance of shared service initiated by T-MEP. Assuming a UE as service consumer moves to the coverage area of T-MEH which triggers application instance relocation to T-MEH. T-MEP may send a terminate app instance request to S-MEP.

1) S-MEP receives a request from T-MEP for terminating app instance running at S-MEH. S-MEP checks whether the application instance is shared by multiple service consumers.

NOTE 1: An appInstanceId is allocated by the mobile edge platform manager during the creation.

NOTE 2: Refer to clause 7.10 of ETSI GS MEC 011 [i.5] and clause 6.3.1.2 of ETSI GS MEC 010-2 [i.4].

2) If there are more than one service consumer subscribing to this application instance, S-MEP may not terminate the application instance. Instead, it may send a message of cleanup UE context to S-App instance for deleting the context related to UE in the application instance.

   Otherwise, if S-MEP does not find any service consumer other than the UE subscribing the application instance to be terminated, it then triggers the procedure of terminating application instance defined in clause 5.4.1.

3) After the UE context is cleaned up by S-App instance, S-MEP cleans up the UE context at MEP and sends a terminate app instance response to T-MEP, indicating the UE context has been cleaned up.

   Otherwise, the S-MEP sends a terminate app instance response to T-MEP indicating the app instance has been terminated.

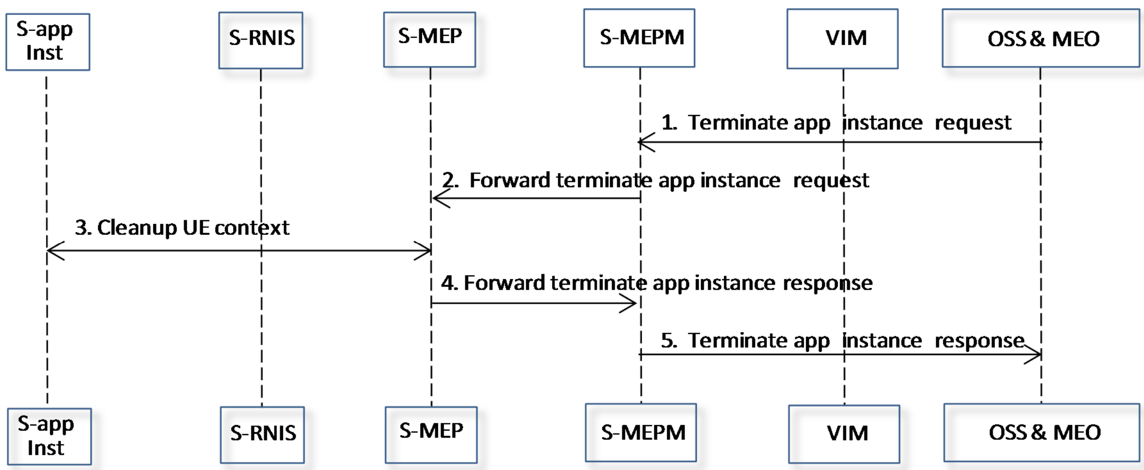

Figure 5.4.2-2: A message flow for terminating a shared service of application instance triggered by MEO

Figure 5.4.2-2 shows the message flow for terminating application instance of shared service initiated by MEO. Assuming a UE as service consumer moves out of the coverage area of S-MEH which triggers application instance relocation to the cloud data centre, for example. If the application instance is successfully relocated to the cloud data centre, MEO may send a terminate app instance request to S-MEP.

In the case that MEO wants to terminate application instance, it can send a (forced) terminate app instance.

NOTE 3: It is FFS in the normative specification stage how to specify the information model and/or APIs to distinguish the forced terminate application instance invoked by MEO, or the non-forced terminate shared application instance when it is still serving other UEs.

1) MEO sends a (non-forced) terminate application instance request to S-MEPM.

NOTE 4: Refer to figure 5.4-1 of ETSI GS MEC 010-2 [i.4].

2) S-MEPM forwards the terminate application instance request to S-MEP. Once S-MEP receives a request for terminating the application instance running at S-MEH, it checks whether the application instance is shared by multiple service consumers.

NOTE 5: An appInstanceId is allocated by the mobile edge platform manager during the creation. Refer to clause 7.10 of ETSI GS MEC 011 [i.5] and clause 6.3.1.2 of ETSI GS MEC 010-2 [i.4].

3) If there are more than one service consumer subscribing to this application instance, S-MEP may not terminate the application instance. Instead, it may send a message of cleanup UE context to S-App instance for deleting the context related to UE in the application instance.

   Otherwise, if S-MEP does not find any service consumer other than the UE subscribing the application instance to be terminated, it then triggers the procedure of terminating application instance defined in clause 5.4.1.

4) If the UE context is cleaned up by the S-App instance, S-MEP cleans up the UE context at MEP and sends a terminate app instance response to S-MEPM, indicating the UE context has been cleaned up.

   Otherwise, the S-MEP sends a terminate app instance response to S-MEPM indicating the app instance has been terminated.

5) S-MEPM sends a terminate application instance response to MEO.

5.4.3 Gaps identified by information flow

5.4.3.1 App instance life cycle change/creation/deletion notification

Clause 5.4.1 provides a message flow of terminating a dedicated application instance at S-MEH. The termination of ME application instance operation could be initiated by T-MEP or MEO. According to ETSI GS MEC 010-2 [i.4], the termination of application instance operation is allowed to be initiated by MEO or OSS. If the termination of ME application instance is initiated by T-MEP, the T-MEP may need to send an application instance termination request to MEPM to start the termination procedure. Therefore it needs to enhance the termination of application instantiation operation in ETSI GS MEC 010-2 [i.4] to support termination initiated by T-MEP over reference point Mm5, and enhance the notification of termination of application instance to the MEO over the reference point Mm3.

5.4.3.2 Shared application instance creation and deletion

Clause 5.4.2 describes the termination of a shared service produced by an application instance. This operation is already supported in the present release of Platform Application Enablement Service specified in ETSI GS MEC 011 [i.5]. Therefore no further enhancement of platform application enablement service is needed.

5.5 Evaluation of application instance relocation

5.5.1 General

On one hand, the end to end communication is supported between two mobile edge applications, on the other hand, it also can occur between an external application and a mobile edge application. In the former situation, the typical use case is the API invocation between the applications consuming mobile edge services and the applications providing mobile edge services. In the latter situation, the typical use case is the user traffic routing between applications on the UE and mobile edge applications near the attachment point of the UE. For the mobility of external applications, it is beyond the scope of the present document. This clause will focus on the mobility of the mobile edge applications in mobile edge system.

5.5.2 Relocation notations

Mobile edge application relocation procedure is a process that solves the identified key mobility issues and aims at keeping the end to end communication between the peer applications, regardless of whether the communication is new or re-established. The means to achieve this goal calls application relocation method. For example, the instance-based relocation and the state-based relocation method. The exact application relocation procedure varies with the application relocation methods. In general, three relocation roles are indispensable in the application relocation procedure. They are depicted in figure 5.5.2-1.

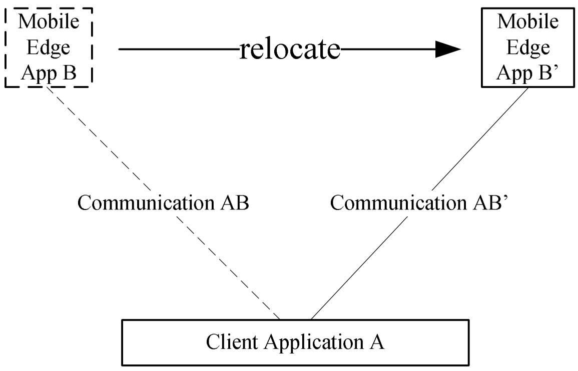

Figure 5.5.2-1: General illustration for mobility of mobile edge application

- Client Application A: the application, utilizing the communication with mobile edge application B before the application relocation, and the communication with mobile edge application B' after the application relocation.
- Mobile Edge App B: the source mobile edge application communicating with the client application A before the application relocation.
- Mobile Edge App B': the target mobile edge application communicating with the client application A after the application relocation.

Both the mobile edge applications and the management system of mobile edge computing involve in application relocation procedure. There can be differences in case of entities involved in different application relocation procedures i.e. state based and instance based. In case of state based procedure, the client application is associated to a new instance of the mobile edge application after the state information has been synchronized. In case of instance based relocation, the serving application instance is moved to a different host.

In order to understand the relocation procedure; the following terms are defined in order to describe the end to end information flows in high-level.

- Relocation requirement: requirement related to mobility of the mobile edge application, that can be configured by the mobile edge operator, application developer and the mobile subscriber, or collected from the UE applications and CFS portal on-demand.

- Relocation indicator: the criteria to evaluate the relocation method. For example, service continuity level supported, service interruption time, application IP address change supported or not, assisted by the mobile edge application or not, relocation time, transferred data size.
- Relocation trigger: the event that initiates the relocation procedure. It possibly includes load balancing, performance optimization, policy consistency, and relocation request on-demand.
- Relocation phase: logically the application relocation procedure can be separated into several phases: relocation initiation phase, relocation validation phase, relocation preparation phase, relocation execution phase, relocation completion phase. Normally, these relocation phases should be performed in order. If some of them can be skipped or performed in parallel, an explicit annotation is needed.

5.5.3 High-level information flows for relocation procedure

A high-level information flow for mobile edge application relocation is illustrated below. In order to keep the communication for client application available after application relocation procedure, 5 relocation phases are drawn in sequence.

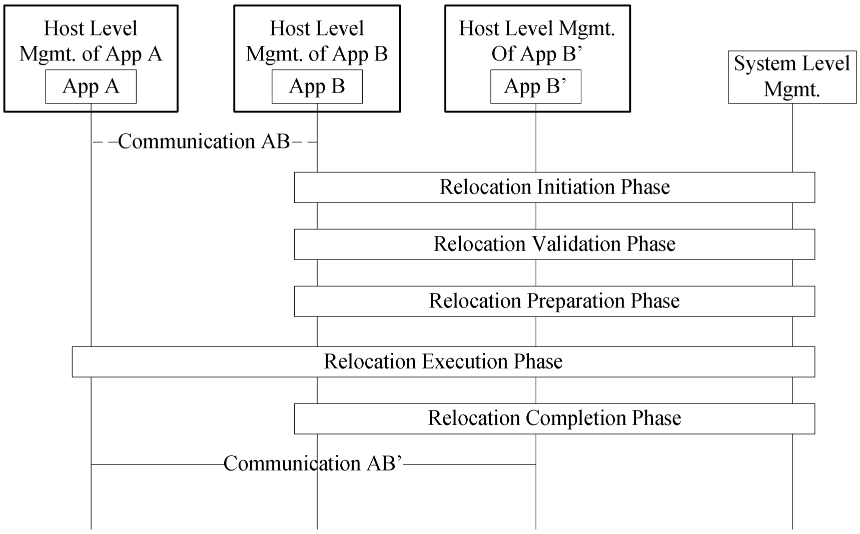

Figure 5.5.3-1: High-level information flow for relocation procedure

- Relocation Initiation Phase: At this phase, all the relocation triggers that may cause the relocation procedure should be considered. A preliminary relocation decision can be made by the App B, App B', Environment of App B and Environment of App B'; or a final relocation decision made by the Management.
- Relocation Validation Phase: The preliminary relocation decision is collected by the Management for validation based on the relocation requirements (e.g. required resources) and for relocation method pre-evaluation based on the relocation indicators. If all satisfied, a final relocation decision including the relocation method identifier is made by the Management.
- Relocation Preparation Phase: Synchronize the application related context between Environment of App B and Environment of App B'.
- Relocation Execution Phase: Execute the relocation procedure in practice based on the relocation method indicated in the final relocation decision.
- Relocation Completion Phase: Release the old and temporary resources. Change the Communication AB to Communication AB'.

5.5.4 Overview for a specific relocation method

Table 5.5.4-1 provides an overview for a specific application relocation method.

Table 5.5.4-1: Overview for a specific relocation method

Relocation Roles	Relocation Method (State based or instance based)				
	Initiation Phase	Validation Phase	Preparation Phase	Execution Phase	Completion Phase
client Application A	∨			∨	∨
App B	∨	∨		×	∨
App B'	∨	∨		×	∨
Host level mgmt. of App B	∨	∨	×	∨	×
Host level Mgmt. of App B'	∨	∨	×	∨	×
System level Mgmt.	∨	×	∨	∨	∨
NOTE 1: For the mark "∨", it means this relocation role is probably involved in this phase. It is up to the specific relocation method. NOTE 2: For the mark "×", it means this relocation role will be involved in this phase regardless of the relocation methods.					

5.5.5 Evaluation for a specific relocation method

Table 5.5.5-1 provides an evaluation for a specific application relocation method based on the relocation indicators.

Table 5.5.5-1

Relocation Indicators	Relocation Method	
	Satisfied	Value
Indicator 1		M
Indicator 2	×	
…	×	
Indicator N		K
NOTE: For the mark "×", it means this measurement indicator is satisfied. For the M and K, they give the exact values for the measurement indicators.		

5.5.6 Gaps identified by information flow

The information flows of allocation relocation described by clause 5.5 include two stages:

- Stage 1: configuration for support of application relocation.
- Stage 2: performing the application relocation.

In the stage of configuration for support application, the related requirement and parameters may need to be provided in AppD of the application package. This may require enhancement the on-boarding procedure of ETSI GS MEC 010-2 [i.4].

In the stage of performing the application relocation, the described procedures are similar to the clauses 5.1, 5.2, 5.3 and 5.5 with additional requirements.

- Relocation Initiation Phase: this phase is the similar to the step 1 of figure 5.1-1.
- Relocation Validation Phase: this phase is similar to the step 2 of figure 5.1-1. It may involve MEPM and MEO to validate the application relocation request based on the provided relocation requirements. This may need to enhance the instantiation procedure of present specifications ETSI GS MEC 010-2 [i.4] and ETSI GS MEC 011 [i.5], if the application relocation reuses the procedure of application instantiation.
- Relocation Preparation and Execution Phases: those phases are similar to the step 3 of figure 5.1-1.

NOTE 1: Refer to the analysis in clauses 5.2 and 5.3.

- Relocation Completion Phase: this phase is equivalent to the step 5 in figure 5.1-1.

NOTE 2: Refer to the analysis in clause 5.4.

6 Key issues and potential solutions

6.1 Introduction

6.1.1 Roles and functions in service provisioning

In a broad context, service provisioning in MEC system involves three mutual connected components. They offer three individual connections:

- between MEHs, that hosts the application server among other things, and UE (User Equipment);
- between UE and CPN (Connectivity Provider Network); and
- between MEH and CPN, respectively.

The three connections are of different natures and purposes and, as such, deserve different treatments. In order to avoid semantic confusion in elaborating solutions, it is beneficial to have more accurate definition of these relations, in terms of roles they play in the service provisioning, and to introduce qualified notations within this context to facilitate clarity:

a) The user equipment (UE) has at least two different roles, or "personalities":

   - UE as part of the connectivity provider network that provides connectivity for the service provisioning by MEC system. In this role it does not have to be a consumer of the application service offered by MEC system, or not a consumer of any application service beyond the connectivity provided by the CPN. To emphasize this role, notation ctUE can be used to qualify UE as "connectivity terminal" hosted by UE.

   - UE as the peer application that maintains information exchange over application specific transport layer, e.g. HTTP, to communicate with the application server, independent of the underlying CPN. To emphasize this role, notation paUE can be used to qualify UE as "peer application" hosted by UE.

   Service provisioning by MEC system is only possible when UE assumes both roles described above.

b) The CPN provides connectivity between the MEH and UE, where ctUE provides network bearer for paUE. No restriction is imposed on the way, how CPN is providing the connectivity, as long as transport service between paUE and application server (AS) is maintained. This role can be characterized in two aspects:

   - Different application platforms/systems can be supported by the same CPN, while the MEC system is only one of them. In this sense, CPN is transparent to the applications. But integration of CPN into the service provisioning is possible.

   - Being a mediator between the application server and application client, CPN has the capability of authenticating both the MEH and the UE, i.e. ctUE, on its own right, independent of the business relation between the application provider and the application consumer. This capability can also be utilized by the applications.

   CPN can be based on any access/networking technology. For paUE, a CPN is nothing but a medium that provides the necessary connectivity.

c) The Application Server (AS) is the core of application service provisioning and it shares application specific interfaces with the paUE. Within the context of MEC, the application server is hosted by MEH and provides services generated by the hosted applications. To improve the quality of service, different MEHs can be inter-connected into a network - a network that is dedicated to supporting the communications between application servers. This network, can be seen as ASN (application server network), is conceptually independent of the underlying CPN. To emphasize this difference, the following notations may be helpful, whenever confusion could occur:

   - MEH Network: network that connects ME hosts, versus.

   - CPN: Communications network provided by the connectivity provider.

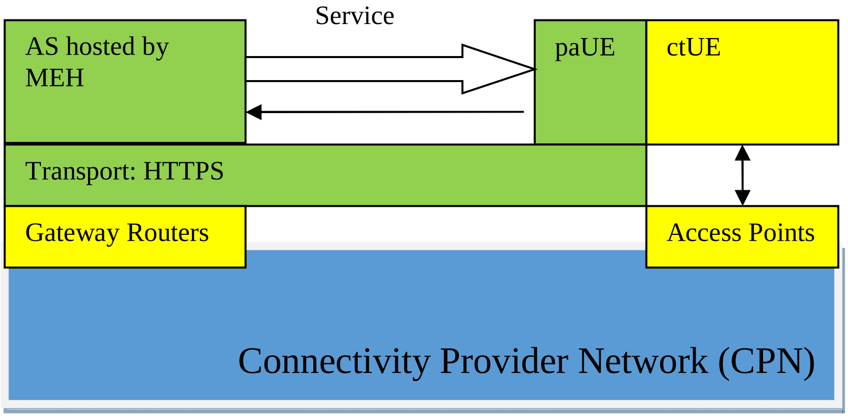

Figure 6.1.1-1: Illustration of relations of components in MEC for service provisioning

6.1.2 Problems and Assumptions

The purpose of the clause is to have an overview of possible solutions to problems identified so far. The solutions are to be judged by feasibility only. The evaluation for other purpose and selection of the solution for MEC remains work of Phase 2 work and, as such, will not be considered here. For the same reason, the terminologies used in the following are tentative and subject to change later.

Seen from architectural point of view, there are three major players considered:

- Mobile Edge Host;
- User Equipment; and
- Communications Network.

The service provisioning is based on a client-server relation between UE and MEH, the connectivity between them is provided by Communication Network.

One of objectives of MEC design is to allow 3rd party provider of MEC service, hence it is necessary to classify the solutions in "network independent" and "network dependent". The network independent solutions are generic solutions, that apply to any network and any technology, while the network dependent solutions are those making use of the network specific features and technology specific information to improve the performance.

An important service quality measure for MEC is the service continuity, which can be assured only when the service instance state is conserved during the operation and, in particular, before and after service interruptions. Hence, the faithful state relocation is one of major objectives of the present document and the other objective is the instance relocation. Purpose of looking at different ways of providing service continuity is to achieve a better understanding of the issues, so that optimal design can be obtained for different requirements. Of course, there are many criteria that can qualify the optimality depending on the business objective of MEC operator, e.g. low cost, protocol simplicity, network dependency, network independency, etc. As long as the solution is feasible and applies to MEC, no restriction will be made on the criteria and hence the solution in the present document.

As is known in mathematics, optimization is meaningless without conditions. Hence, a set of working assumptions are necessary. Given that different solutions are to be considered, the working assumptions will be presented in two parts:

- common working assumptions, that apply to all solutions; and
- specific working assumptions, that apply to individual solutions.

The following are the Common Working Assumptions:

- MEP is responsible for launch and release of the service, as well as UE to ME application association.
- MEPM is responsible for the resource allocation as well as the coordination with the ME orchestrator.
- MEH is responsible for application control and service data encapsulation (a.k.a. Data Plane).

- AP is the access point that provides connectivity between UE and MEH through a (wireless) communications network and AP is part of this network. MEP obtains the mobility related information about UE through RNIS.
- Acronym: Prefix S and T are used to refer to source and target, respectively. AP stands for access point.

To complete the introduction, a formal Problem Statement is given below.

A user is consuming a service, while moving out of the coverage area of S_MEH. Later he/she enters the coverage area of T_MEH and expects to resume the same service. This requires a relocation of service instance state from S_MEH to T_MEH such that the service delivered is continuous, with respect to the following:

- Pre-Condition:
  - UE is consuming a service offered of an application offered by S-MEH. The communications channel is established, passing through S_AP.
  - S-MEP subscribes to the mobility information service (for this UE) by S-RNIS.
- Post-Condition:
  - UE is consuming the same service by the same application offered by T-MEH. The communications channel is established, passing through T-AP.
  - T-MEP subscribes to the mobility information service (for this UE) by T-RNIS.

An illustration of the problem is given in figure 6.1.2-1.

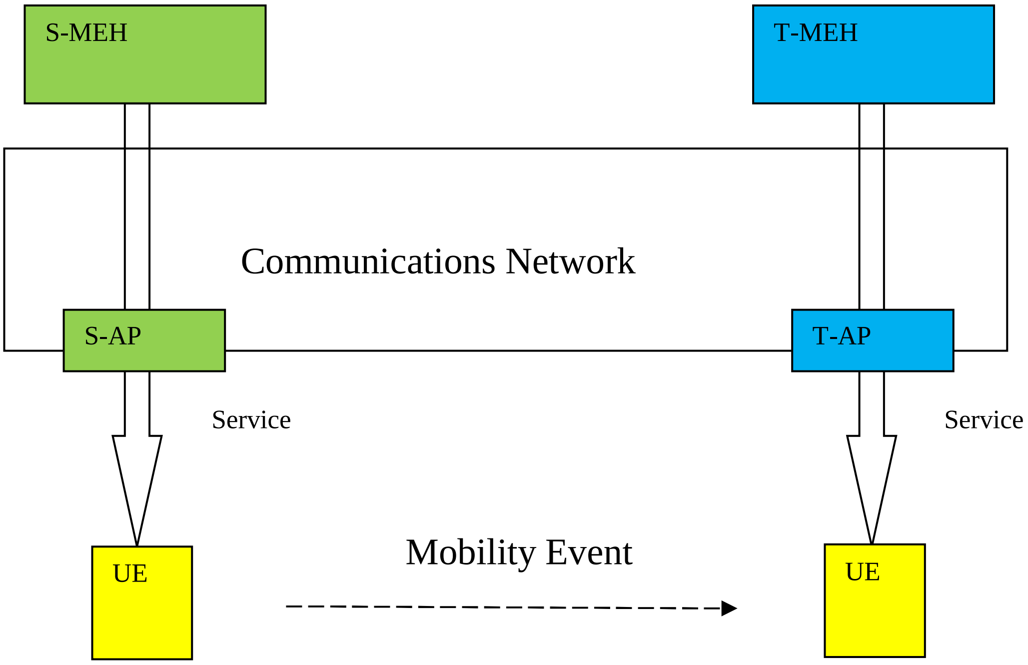

Figure 6.1.2-1: Illustration of the general problem

6.1.3 Characterization of Service Continuity

The services provided by MEC to the users can experience interruption, caused by application software failure, malfunction of MEC system, loss of connectivity due to network failure and user's voluntary or involuntary action. If the interruption is due to user's voluntary action, the user perhaps does not expect continuity. Otherwise, the user expects certain continuity and that is better quantifiable. The continuity of the service is the perception of the out-of-service time and can be measured by the latency between the terminated instance and the resumed instance of the same service, maintaining the instance state.

Clearly, service continuity is an important performance metric for MEC system. A quantitative characterization of this metric is, however, not easy, as it depends on the deployment scenarios and offered applications. Since in both aspects there are unknowns and uncertainty, only a loose characterization can be taken into the vocabulary of the present document:

- No continuity: This applies to the case of stateless transition of service instance, characterized by an independence of the service states before and after the service interruption.
- Low continuity: This applies to the case when the instance state is preserved when the service is resumed after an interruption. The time of the service interruption can be minutes, hours, days or even longer. Normally, an upper bound is given.
- Soft continuity: This applies to the case when the instance is preserved for resuming the service after a short interruption with a variation that is either tolerable by the user, or by the application. The duration of the interruption is usually less than minutes, but is allowed in a rather large range given by an appropriate probability distribution.
- High continuity: This applies to the so-called "seamless" transition of service instances. The requirement is most likely posed by the time sensitive applications, such as ITS or gaming, and the allowed service interruption is given in milliseconds. Normally, an upper bound is given.

6.2 Key Issues

6.2.1 Key Issue: The optimization of application state relocation

6.2.1.1 Description

This key issue should study how the ME system optimizes the relocation of application state information based on application related information.

6.2.1.2 Solution 1: MEH assisted state relocation without RNIS

As a MEH assisted state relocation further reduces the threshold for MEC provisioning, this solution does not rely on RNIS. Instead it makes use of the location information available on UE. How this information is present on UE and in which format/granularity it is available depends on the CPN and UE capability and will be specified in MEC Phase 2. Target MEP is required to resolve the source MEH identity/address through the received information provided by the peer application running on UE.

Special Assumption:

- S-MEH and T-MEH are connected through a network.
- MEH can identify service interruption, based on application specific means.
- T-MEP is capable of deriving identity/address of S-MEH based on location information provided by UE and UE has means to obtain information of its own location and to prepare it in an adequate format.

Description:

1) S-MEP observes service interruption. In response to this event, S-MEP requests "service instance freeze" by S-MEPM, and stores the instance state.

2) UE, having established network connectivity through T-AP, requests service by the application hosted by T-MEH . UE provides information of its last location before service interruption when requesting the service resumption. Using this information, T-MEP locates, and sets up connection with, S-MEP. Over this connection, T-MEP sends "state relocation" request to S-MEP.

3) S-MEP, upon receiving a "state relocation" request, sends "instance state" to T-MEP, together with UE's temporary identifier.

4) T-MEP, upon receiving "instance state" from S-MEP, sends "resource allocation" request to T-MEPM. Upon receiving confirmation, T-MEP either initiates the service instance and provides the requested service, or rejects the requested service, depending on whether T-MEPM has confirmed the "resource allocation" or not.

5) S-MEP, having received "state instance" acknowledgement from T-MEP, deletes service instance and service context that belongs to UE.

Analysis:

- Pros:
  - Network agnostic.
  - Independent of RNIS.
- Cons:
  - MEHs are required to resolve MEH identity/address from location information provided by UE.

6.2.1.3 Solution 2: MEH assisted proactive state relocation for UE in connected mode

This solution is a special case of MEH assisted state relation, aiming at reducing the latency based on mobility events and CPN technologies. To assist the elucidation, mobility events of handover in LTE/EPS based CPN is given as example.

The improvement is articulated by the fact that latency is reciprocal to the success probability of the handover operation. Hence, the decision taken by ME system is a probabilistic and subject to performance trade-off. It serves as an example for technology specific optimized MEH assisted state relocation.

Special Assumption:

- Same as the solution of MEH assisted state relocation, in addition to.
- UE is in connected mode and performing an handover based on S1-AP.
- Core network does not change gateways being used during the handover.

Description:

Based on solution of MEH assisted state relocation, technology specific values to the "mobility event start" (MoE start) and "mobility event end" (MoE end) can be made by the MEC system based on the in-depth knowledge about the CPN with assistance of RNIS. Consider the example when the CPN is LTE/EPC and the UE is undergoing S1 based handover. There is a sequence of messages exchanged between SeNB and TeNB. To reduce the latency of ME application state relocation, different time points in the RRC protocol sequence can be used as MoE start/end, so that more time can be gained for S-MEH and T-MEH to prepare the MEH assisted state relocation. Some of the time point pairs (MoE start, MoE end) applicable for this purpose are shown in table 6.2.1.3-1 for LTE, where the given list, however, is not meant to be exhaustive.

The improved latency is accomplished at the price of relocation success, because failed handover operation may nullify the state transition preparation made by MEHs. This is the trade-off between latency and reliability. Therefore, the ME system performance depends on the performance and implementations of the CPN For the same reason, the decision on which set points for MoE start/end to choose, is up to the MEC operator.

Table 6.2.1.3-1: Possible set points for "mobility event start" and "mobility event end"

Index	MoE start	MoE end	Comment
1	Handover Ack (sent by TeNB to SeNB)	SN State Transfer Request (sent by SeNB to TeNB)	RRC level
2	RECONFIGURATION (sent by SeNB to UE)	SN State Transfer Request (sent by SeNB to TeNB)	RRC level
3	RECONFIGURATION (sent by SeNB to UE)	RRC_CONNECTION COMPLETION (sent by UE to TeNB)	RRC level
4	RECONFIGURATION (sent by SeNB to UE)	End Marker Response (sent by TeNB to SeNB)	RRC level
5	SN State Transfer Request (sent by SeNB to TeNB)	End Marker Response (sent by TeNB to SeNB)	RRC level
6	TBD	TBD	Upper Layer triggers

Table 6.2.1.3-1 shows an example of LTE for possible set points for "mobility event start" and "mobility event end". The latency increases with the number of the index in the first column.

Table 6.2.1.3-1 provides an LTE based example that shows the possible set points from "mobility event start" to "mobility event end". Note that ETSI GS MEC 012 [i.6], table 7.4.2-1 has specified handover status value for E-UTRAN, to allow different MoE start/end set points. Using those handover status values the interpretation of the set points lies with RNIS. As a result, the performance in terms of the trade-off between latency and reliability depends on RNIS implementation, The solution described here requires that the ME system decides the MoE start/end, based on the situation at individual location and time, to better achieve adaptive performance. It does not only allow MEC system, or even application, to improve the overall performance, but also make the same procedure applicable to other access technologies including future technologies.

Analysis:

- TBD.

6.2.1.4 Solution 3: UE assisted state relocation

Example scenario: A consumer is watching a movie on his/her handset at home. Then he/she walks from home to office, without explicitly turning off the service. Later, he/she wants to continue to watch the same movie in a later time. Generally speaking, the solution applies to either latency tolerant service or latency tolerant consumer.

Special Assumptions:

- MEHs are independent of each other and there is no connection between them.
- Information is added as parameter sent from the ME application to the peer application in order to identify the packet containing the relevant state information.
- RNIS is available and is subscribed to by the ME application.

Description:

1) The ME application receives from S-RNIS the "mobility event" notification indicating that UE is about to leave the serving area of S-MEH.

2) The ME application includes "instance state" information in the last packet before service interruption and sends it to the peer client application on UE.

   (some time later)

3) When the UE is served in the new cell the peer client application will forward the same information towards the network. Depending upon whether the T-MEH has the correct traffic rule settings the T-MEH may detect the inserted information which may trigger follow-up actions by T-MEH:

   a) If the ME application is running, forward the information.

   b) If the ME application is not running, trigger "resource allocation" via T-MEPM.

4) The ME application then inspects the first packet received from the peer client application hosted by UE, to obtain the "instance state".

5) T-MEP, being confirmed by T-MEPM of the "resource allocation", initiates the application instance with the received "instance state" and provides the requested service. If T-MEPM did not confirm the "resource allocation", T-MEP rejects the requested service.

Analysis:

a) Pros: Network agnostic.

b) Cons: One bit information to cache by UE.

6.2.1.5 Solution 4: User plane optimization function

This solution enables an application state relocation to, also consider application preferences, and not only based on UE mobility.

The solution introduces a User plane Optimization Function (UOF) as part of the MEPM that has the role of collecting, aggregating, processing and analysing relevant information from ME applications.

UOF and MEO exchange information to decide whether to perform application state relocation or application relocation. If any relocation has been decided then UOF needs to update the CPN (Connectivity Provider Network) on the new user plane path. In case the CPN is 5GC this is the interface to the SMF (See 3GPP TS 23.501 [i.8], clause 5.6.7 on "Application influence on traffic routing").

NOTE 1: It is FFS which interface to use in EPC.

NOTE 2: ME application should be able to provide application related information to the UOF relaying through the MEP via Mp1.

The User plane Optimization Function is responsible for informing the CPN about the user plane traffic to be routed and where to route the traffic.

6.2.1.6 Gaps identified by the key issue

6.2.1.6.1 ME application relocation detection

Clause 6.2.1 analyses the optimization of application state relocation in following aspects:

- Trigger of the application relocation:
  - Solution 1 (MEH assisted application relocation without RNIS) relies on the location information of the UE to start the application relocation. It implies that the T-MEH needs to identify UE movement based on the UE location information and identity/IP address of the S-MEH.
  - Solution 2 (MEH assisted proactive state relocation for UE in connected mode) relies on the special event (such as the End Marker) detected by the MEH.
  - Solution 3 (UE assisted state relocation) requires UE to send an indication to trigger the application relocation. This would be UE or UE application client dependant, and may not be directly related to MEH.
- The procedure of application relocation. This part is similar to the analysis in clause 6.2.6.3.
- Optimize the application relocation.

Therefore, the solutions for application relocation in clause 6.2.1 would require MEH to detect application traffic path change, which might need to define a new service or enhance the existing service like RNIS (ETSI GS MEC 012 [i.6]) in MEH to perform such traffic path change detection, incorporating with the platform application enablement function (see ETSI GS MEC 011 [i.5]) and notify a function entity in MEPM.

6.2.1.6.2 New function of user plane optimization function in MEPM

Clause 6.2.1.5 provides a solution of optimizing application relocation via a new function in MEPM: User plane Optimization Function (UOF) for collecting, aggregating, processing and analysing relevant information for ME applications over existing reference point Mm5 or other (TBD). UOF can use the existing reference point Mm3 to exchange information with MEO to decide application state relocation. ME application may provide application related information to the UOF relaying through the MEP via Mp1. But there is no direct reference point specified in the MEC network reference architecture for UOF to updates the CPN on the new user plane path, like interface to SMF in 5G. So it may need to define a reference point to the functional entity of underlying network.

6.2.2 Key Issue: UE IP address change

6.2.2.1 Description

In 3GPP EPC (and earlier) environment, if the mobile edge host is located on SGi reference point, the GWs and the mobile edge host are most likely located close to the radio access point. After a route optimization, which causes IP anchoring point relocation (e.g. PGW relocation), the UE IP address will be changed.

In 3GPP 5G environment, the change of the IP anchor GW/user plane function/UPF for a PDU session in session and service continuity/SSC mode 2 and 3 results in the assignment of a new UE IP address by the network, in which situation the new UE IP address will be applied to the IP connection instead of the old UE IP address. SSC mode 2 and mode 3 definitions as per 3GPP TS 23.501 [i.8]:

- SSC Mode 2: For PDU session of SSC mode 2, the network may trigger the release of the PDU Session and instruct the UE to establish a new PDU session to the same data network immediately. At establishment of the new PDU Session, a new UPF acting as PDU session anchor can be selected.
- SSC Mode 3: For PDU session of SSC mode 3, the network allows the establishment of UE connectivity via a new PDU session anchor to the same data network before connectivity between the UE and the previous PDU session anchor is released. When trigger conditions apply, the network decides whether to select a PDU session anchor UPF suitable for the UE's new conditions (e.g. point of attachment to the network).

For further details, refer to 3GPP TS 23.501 [i.8].

The key issue studies how the mobile edge system keeps the connectivity between the UE App and the application instance located in the original PGW or behind the old UPF (user plane function in 5G environment) after the UE IP address changes (i.e. the UE is served by a new PGW or a new UPF).

6.2.2.2 Solution 1

The MEP is made aware of the User or UE ID (e.g. IMSI if the MEP belongs to the trusted area of the operator's network, or the device identifier IMEI) and UE IP address, and possibly of further parameters like e.g. bearer/tunnel ID, upon a UE attaching or registering to the network. Depending on the access and core network (e.g. LTE or 5G), different implementation specific or to-be-standardized ways may be used to access the information.

The MEP/ME Host is made aware of a new application session related to the User/UE ID and IP address. Different implementation specific or to-be-standardized ways, e.g. traffic detection by the MEC data plane, may be used to access the information.

When a handover between gateways/user plane functions occurs and a new IP context with a new IP address is established between the UE and the new gateway/user plane function, the MEP may use the User/UE ID to bind the ongoing application session to the new IP context.

The MEP may create or apply application specific rules or measures to maintain the session and service continuity. For example, MEP may be aware of the application not tolerating the UE IP address change during an ongoing session, and may apply measures, e.g. address translation, to let the session and service continue.

NOTE: Details may possibly be solved in the normative standardization phase, if needed or applicable, or may be left up to the implementers/implementations to solve.

6.2.2.3 Gaps identified by the key issue

Depending on the origin of the possible application specific information to be used for creating or applying application specific rules or measures to maintain the session and service continuity, there may be a need to add related parameters to Mm3/Mm5 or Mp1. To be defined, if the solution (with further details) is introduced in normative standards.

How RNIS acquires information from the underlying access or core network is not specified for LTE, but 5G may grant an opportunity to standardize an interface between the ME Host and access/core network. Possible related parameters to be defined:

1) if the solution (with further details) is introduced in normative standards; and

2) if a relevant interface is available in 5G.

Possible application specific rules or measures to maintain the session and service continuity may create a need to add parameters to the Mp2 interface. However, the current assumption in ETSI GS MEC 003 [i.2] is that "this reference point is not further specified".

6.2.3 Key Issue: Updating downlink traffic rules

6.2.3.1 Description

When the mobile edge host is located on S1 reference point, it is possible that the mobile edge host serves more than one radio access nodes (e.g. eNB). The UE may move among the radio access nodes in connected mode or idle mode. And the UE may turn from connected mode to idle mode, and vice versa.

After the UE moves from one radio access node to another, or after UE state changes, the mobile edge host should be able to route the traffic to the UE via the correct radio access node and correct tunnel; and the mobile edge host should not transfer the traffic to wrong UE in any case.

6.2.3.2 Solution 1: Update bearer information between different RNIS within one host

Source Radio Network Information Service (S-RNIS) is the service providing information of the source radio node whereas Target RNIS (T-RNIS) is the service providing information of the target radio node. Data plane supports the encapsulation of IP traffic before passing it to the network and de-capsulation of encapsulated IP traffic received from the network before routing it to the authorized mobile edge applications. The mobile edge platform (MEP) is responsible for controlling traffic rules. S-RNIS, T-RNIS and MEP reside in the same mobile edge host connected to source and target radio nodes correspondingly. Figure 6.2.3.2-1 is the description of the information flows.

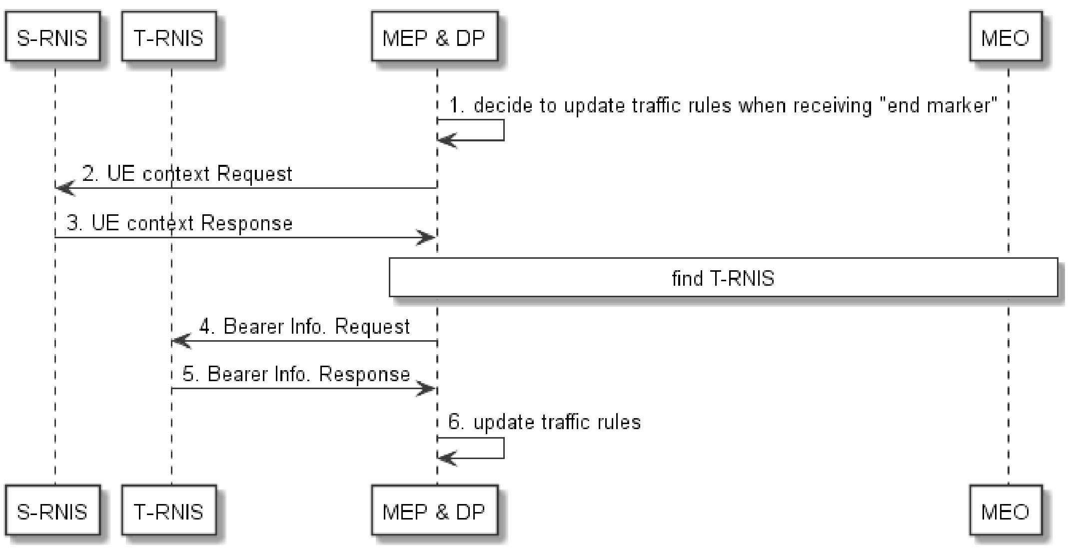

Figure 6.2.3.2-1: Information flow for updating bearer information between different RNIS within one host

1) MEP decides to update the traffic rules when receiving the traffic "end marker", from which the ME platform can extract bearer information (source eNodeB TEID, source eNodeB IP, SGW IP, source eNodeB port, SGW port).

NOTE 1: The traffic "end marker" inserted in the related data flow indicates that the UE handover procedure is in progress.

2) MEP sends a UE context Request message (TEIDs) to S-RNIS which provides information of the source radio node. TEID included in this message indicates the bearer carrying the "end marker" received in step 1. Meanwhile, it is also used to identify UE context (UE ID and UE Bearer ID) in S-RNIS.

3) Based on the TEIDs received in UE context Request, S-RNIS finds the corresponding UE ID and related Bearer ID, and the target radio node ID, and returns them to the MEP in the UE context Response message.

   The MEP then queries the MEO for the target RNIS (T-RNIS) associated with the target radio node ID received in the Response. The MEO determines the T-RNIS and returns the information to the MEP.

NOTE 2: UE ID can be S-TMSI, a temporary UE identifier in 3GPP network. Bearer ID can be EPS bearer ID/ERAB Bearer ID/S1 Bearer ID in 3GPP network.

4) MEP sends Bearer Information Request (UE ID, Bearer ID) to the T-RNIS. UE ID and Bearer ID allow the T-RNIS to find the specific bearer, and acquire the corresponding new bearer information (target eNodeB TEID, SGW TEID, target eNodeB IP, SGW IP, target eNodeB port, SGW port).

5) T-RNIS returns the new bearer information to MEP in the Bearer Information Response.

6) MEP updates traffic rules; especially the bearer information received in previous step and save them in MEP.

6.2.3.3 Solution 2: Update bearer information by RNIS within one host

RNIS is the service exposing radio network information from the source and the target radio nodes. Data plane supports the encapsulation of IP traffic before passing it to the network and de-capsulation of the encapsulated IP traffic received from the network before routing it to the authorized mobile edge applications. Mobile edge platform is responsible for controlling traffic rules. RNIS, ME platform and Data plane reside in the same mobile edge host connected to the source and the target radio nodes. Figure 6.2.3.3-1 is the description of the information flow.

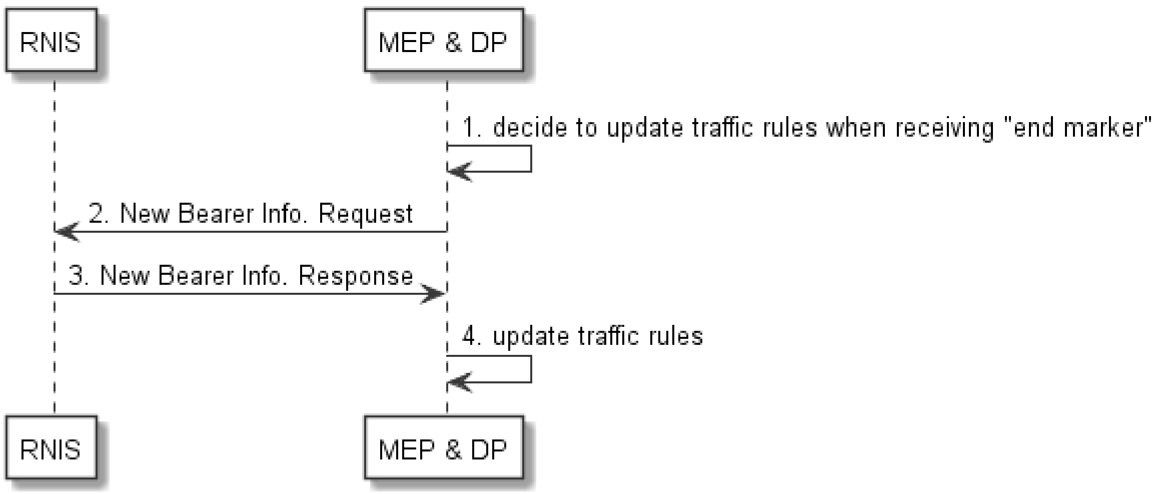

Figure 6.2.3.3-1: Information flow for updating bearer information by RNIS within one host

1) MEP decides to update the traffic rules when receiving the traffic "end marker"; from which, MEP can extract bearer information (source eNodeB TEID, source eNodeB IP, SGW IP, source eNodeB port, SGW port).

NOTE 1: The traffic "end marker" indicates that the UE handover procedure is in progress.

2) MEP sends a New Bearer Information Request message (TEIDs) to RNIS which provides information of the source and the target radio nodes. TEID included in this message indicates the bearer carrying the "end marker" received in step 1. Meanwhile, it is also used to map to new bearer through identifying the UE context (UE ID and UE Bearer ID) in RNIS.

3) RNIS finds the new bearer information based on TEIDs received in step 2, and returns it to the MEP in the New Bearer Information Response.

NOTE 2: The RNIS can map TEIDs to UE context (UE ID, Bearer ID), where UE ID is e.g. S-TMSI, a temporary UE identifier in 3GPP network. Bearer ID can be EPS bearer ID/E-RAB Bearer ID/S1 Bearer ID in 3GPP network. The RNIS acquires the specific bearer information (target eNodeB TEID, SGW TEID, target eNodeB IP, SGW IP, target eNodeB port, SGW port) of the UE context.

4) MEP updates the traffic rules based on the received information from the RNIS.

6.2.3.4 Solution 3: Update bearer information between different ME hosts

RNIS is the service exposing radio network information from the underlying radio nodes. Data plane supports the encapsulation of IP traffic before routing it to the network and de-capsulation of the encapsulated IP traffic received from the network before routing to the authorized mobile edge applications. Mobile edge platform is responsible for controlling the traffic rules. S-RNIS, S-DP, S-MEP all reside in the source mobile edge host connected to the source radio node while T-RNIS, T-DP, T-MEP are in the target mobile edge host connected to the target radio node. Figure 6.2.3.4-1 is the description of the information flow.

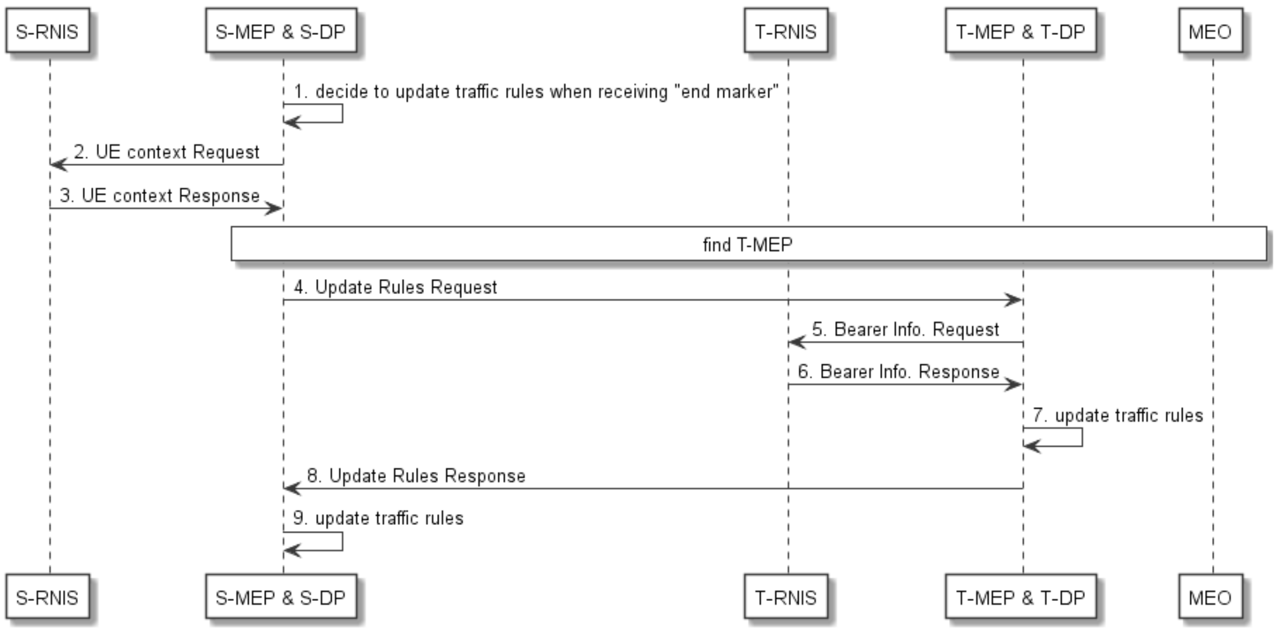

Figure 6.2.3.4-1: Information flow for updating bearer information between different ME hosts

1) S-MEP decides to update the traffic rules when receiving the traffic "end marker" from which it can extract the bearer information (source eNodeB TEID, source eNodeB IP, SGW IP, source eNodeB port, SGW port).

NOTE 1: The traffic "end marker" indicates that the UE handover procedure is in progress.

2) S-MEP sends a UE context Request message (TEIDs) to S-RNIS. TEIDs included in the message indicate the bearer carrying the "end marker" received in step 1. Meanwhile, it is also used to identify the UE context (UE ID and UE Bearer ID) in the S-RNIS.

3) Based on the TEIDs received in the UE context Request, the S-RNIS finds the corresponding UE ID and related Bearer ID, and the target radio node ID, and returns them to the mobile edge platform in the UE context Response message.

The S-MEP then queries the MEO for the target mobile edge platform based on the target radio node ID received in the UE context Response. The MEO returns the information to the S-MEP.

NOTE 2: UE ID can be S-TMSI, a temporary UE identifier in 3GPP network. Bearer ID can be EPS bearer ID/E-RAB Bearer ID/S1 Bearer ID in 3GPP network.

4) The S-MEP sends the Update Traffic Rules Request to the T-MEP. This request contains the UE context acquired in step 3, as well as the traffic rules related to the Source ME host.

5) T-MEP sends the Bearer Information Request (UE ID, Bearer ID) to T-RNIS. The UE ID and the Bearer ID can be used by the T-RNIS to acquire the corresponding new bearer information (target eNodeB TEID, SGW TEID, target eNodeB IP, SGW IP, target eNodeB port, SGW port).

6) T-RNIS returns the new bearer information to the T-MEP in the Bearer Information Response.

7) T-MEP updates the traffic rules.

8) T-MEP returns the Update Traffic Rules response to the S-MEP. This response indicates the completion of updating of the traffic rules by the T-MEP.

9) S-MEP updates the traffic rules accordingly.

6.2.3.5 Gaps identified by the key issue

Refer to clause 6.2.6.5 for the gap analysis about the traffic rules update.

6.2.4 Key Issue: Updating forwarding path inter ME host

6.2.4.1 Description

This key issue studies how the ME system keep the IP connectivity after a UE moves to the serving area of a new ME host, while the ME application instance serving the UE is still located in the original ME host.

The key issue considers the use case in which the ME hosts is located between eNB and EPC.

The key issue considers both the idle mode and the connected mode movement of the UE.

The solutions to the key issue should consider the route optimization if it is needed.

6.2.4.2 Solution

The solution to this key issue may be related to the location of MEHs and the UE movement in the underlying network. There are two possible scenarios caused by UE movement:

1) The IP address of UE is not changed after the UE moved to the coverage of new MEH. In this case, the S-MEP may or may not update the traffic routing rule depending on the criteria for the application mobility. The application mobility criteria may take into account of transmission latency, throughput, traffic load in S-MEH and T-MEH, etc. If the application is not necessary to relocate from S-MEH to T-MEH, then the application can still use existing traffic routing rules to forward the application data to the UE, and vice versa. Otherwise it needs to perform relocating the application from S-MEH to T-MEH, and traffic rules update, which is covered by clause 6.2.6.

2) The IP address of UE is changed after the UE moved to the coverage of new MEH. In this case, the S-MEP has to update the traffic rules.

The solution should be independent to the UE mode (ide or connected).

6.2.4.3 Gaps identified by the key issue

The traffic rule update about forwarding path in inter MEH application mobility, if necessary, is covered by clause 6.2.6.

6.2.5 Key Issue: Application instance relocation

6.2.5.1 Description

This key issue studies how the ME system support ME application instance relocation from source ME host to a target ME host.

The application instance relocation should consider the maintenance of the VM state and disk state of the ME application instance.

The packet loss due to application instance relocation should be minimized as much as possible.

The down time due to application instance relocation should be minimized as much as possible.

The duration of application instance relocation should be minimized as much as possible.

6.2.5.2 Solutions

6.2.5.2.1 Assumption

The application instance relocation may depend on the scope of application instance being served to the UE:

- Stateless application: as it does not memorize the service state or record data about the UE, it may not need to synchronize the service state between S-App and T-App for the service continuity when the application instance being served to the UE is relocated to the T-MEH.
- Stateful: the application instance can memorize the service state of application instance being served to the UE. Therefore the service state information is required to transfer to and synchronize with the T-app instance so that the service continuity can be maintained during the service transition.

Therefore it needs to consider solutions of application instance relocations in those two different cases of applications.

Assumption**:**

- An UE moves out from the S-MEH to the coverage of T-MEH, without change its IP address. The UE IP session continuity is maintained by the underlying network.
- RNIS provides the information of UE bearer change over the underlying network.

6.2.5.2.2 Solution for stateless application instance relocation

The stateless application instance relocation is a part of ME mobility procedure for both dedicated and shared applications if the application instance is not running at the target ME host. The stateless application instance relocation between two ME hosts may require to:

- instantiate an application instance at T-MEH;
- activate the new application instance at T-MEH.

Figure 6.2.5.2.2-1 shows an example of message flow for stateless application instance relocation between two ME hosts.

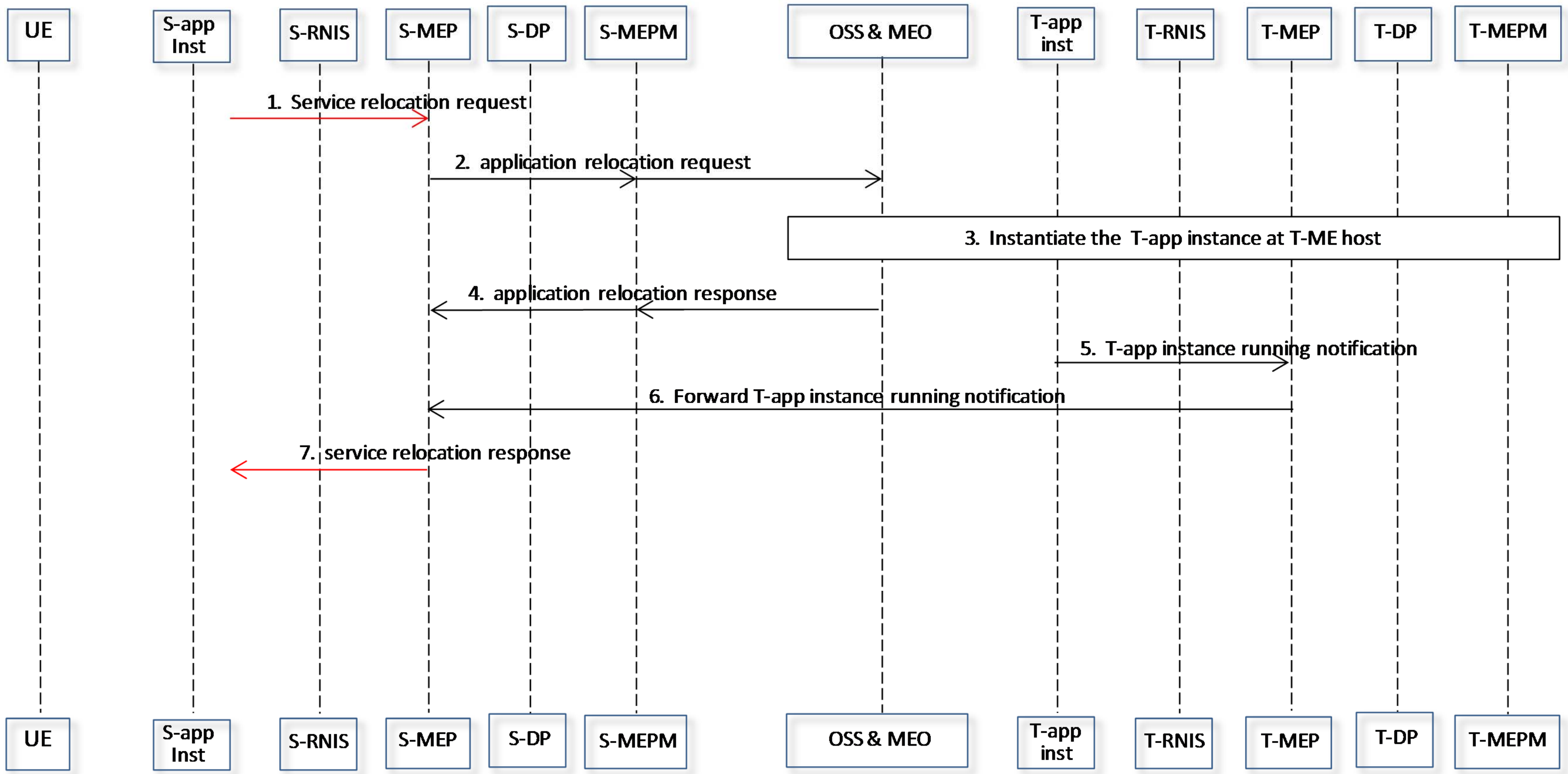

Figure 6.2.5.2.2-1: Example of message flow for stateless application instance relocation

1) The application instance running on the S-MEH provides service to the UE. The application instance consumes S-RNIS. The S-MEP may receive a trigger for service relocation based on the information received from S-RNIS. The service relocation request triggers the application instance relocation between S-MEH and T-MEH.

2) S-MEP sends to MEO through S-MEPM an application instance relocation request including the information of appDId and target radio node ID.

3) Upon receiving the application instance relocation request from S-MEPM, MEO selects a T-MEH associated to target radio node, and sends instantiate application request to the T-MEPM to inform instantiating an application instance using the procedure of "Application instantiation" defined in clause 5.3 of ETSI GS MEC 010-2 [i.4].

4) Once an application instance is created at T-MEH successfully, MEO sends an application relocation response to S-MEP through S-MEPM.

5) The new created T-app instance sends an application instance running notification to T-MEP.

6) T-MEP then forwards the application instance running notification to S-MEP.

7) S-MEP sends a service relocation response to the sender of the service relocation request, which may triggers the next procedure of ME mobility.

6.2.5.2.3 Solution for stateful application instance relocation

A stateful application instance relocation is a part of ME mobility procedure for dedicated or shared applications. The stateful application instance relocation between two ME hosts requires to:

- record the service state of the application instance being served to the UE at S-MEH;
- instantiate an application instance at T-MEH;
- transfer the service state information of the application instance being served to the UE from S-MEH to T-MEH, and synchronize with the new created application instance at T-MEH;
- activate the new application instance at T-MEH.

Figure 6.2.5.2.3-1 shows the message flow of stateful application instance relocation between two ME hosts.

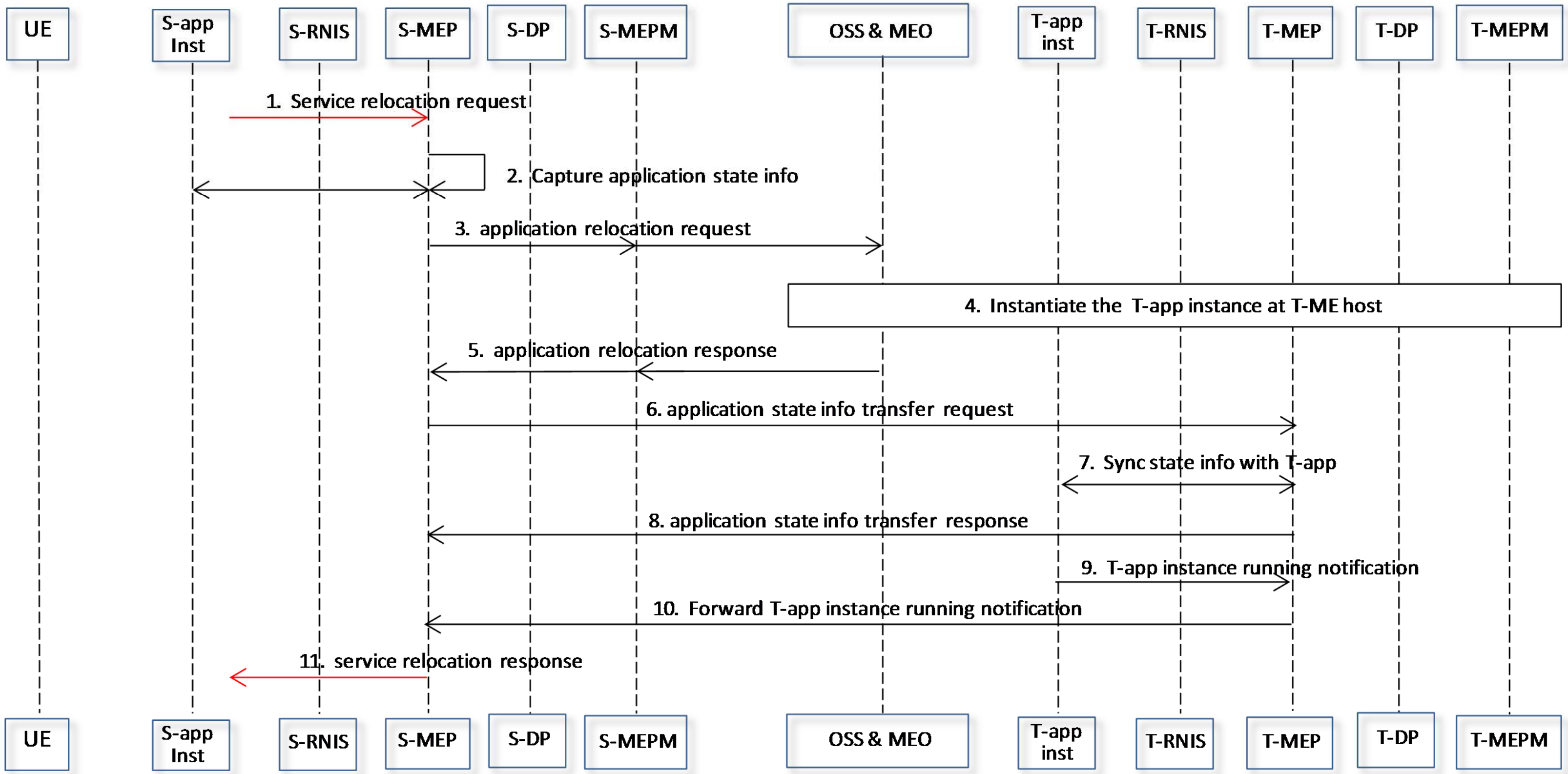

Figure 6.2.5.2.3-1: Message flow for stateful application instance relocation

1) The application instance running on the S-MEH provides service to the UE. The service instance consumes S-RNIS. The S-MEP may receive a trigger for service relocation based on the information received from S-RNIS. The service relocation request triggers the application instance relocation between S-MEH and T-MEH.

2) When receiving the service relocation request for the UE, S-MEP identifies the stateful application instance, and triggers the application instance to capture the service state of application being served to the UE.

NOTE 1: It is FFS in the normative specification stage to define the information model and/or APIs for application instance state transfer

3) S-MEP sends to MEO through S-MEPM an application instance relocation request including the information of appDId and target radio node ID.

4) Upon receiving the application instance relocation request from S-MEPM, MEO selects a T-MEH associated to target radio node, and sends instantiate application request to the T-MEPM to inform instantiating an application instance using the procedure of "Application instantiation" defined in clause 5.3 of ETSI GS MEC 010-2 [i.4].

5) Once the application instance is created at T-MEH successfully, MEO sends an application relocation response to S-MEP through S-MEPM.

6) S-MEP sends an application state info transfer request with the captured service state information for the UE to T-MEP.

   The service state information of application instance being served to the UE is captured at the step 2.

NOTE 2: It is FFS in the normative specification stage to define the information model and /or APIs for application instance state transfer.

7) T-MEP transfers the received service state information to the application instance (T-app) associated to the UE which is running at T-MEH, and activates it.

8) T-MEP sends an application state information transfer response to S-MEP to indicate the success of service state of application instance transfer.

9) T-app instance sends an application instance running notification to T-MEP after synchronization of the service state of application instance with the received service state information.

10) T-MEP forwards the application instance running notification to S-MEP.

11) S-MEP sends a service relocation response to the sender of service relocation request, which may trigger the next procedure in the ME mobility.

6.2.5.3 Gaps identified by the key issue

6.2.5.3.1 ME application mobility service

Clause 6.2.5 studies the key issue of application instance relocations and categorizes them as stateless and stateful application instance relocations. For the stateless application instance relocation, it may follow existing application enablement procedure to instantiate a new application instance at T-MEH. For stateful application relocation, it may require additional procedure to capture the instance service state and transfer it from S-MEH to T-MEH.

To serve this application relocation, it might need to define a new service such as ME application mobility service (MAMS) to handle the application relocation offered by the ME platform. MAMS may support:

- Communicate with other functional entities of MEC for invoking the application instance relocation being served to a UE.
- Inform the Platform Application Enablement Service at S-MEP to capture the service state of application instance.
- Transfer the captured state of stateful application from MAMS of S-MEP to T-MEP.

- Inform the Platform Application Enablement Service at T-MEP to synchronize the state of new instantiated stateful application.

6.2.5.3.2 Enhancement of platform application enablement

To support stateful application instance relocation, the Platform Application Enablement Service defined in ETSI GS MEC 011 [i.5] may need to enhance, including:

- define a new message interface to communicate with MAMS;
- identifying whether the application to be relocated is the stateful application instance;
- communicate with the application instance to capture the service state of stateful application instance being served to the UE.

6.2.6 Key Issue: Update traffic routing rules

6.2.6.1 Description

When the UE has moved to the area covered by a new MEH the traffic may still be routed to the original application instance in the S-MEH. At some point, when the instance of the same ME application is available at the T-MEH it may be necessary to redirect the UE traffic to this ME application instance.

In such situation, in order to get the related UE traffic routed to the target mobile edge host after the application relocation, the target mobile edge host needs to set the related traffic routing rules, consisting of:

- Traffic filters (at least IP 5-tuples): identifying the UE traffic incoming.
- Rules: containing the actions performing on the UE traffic, and if applicable, the application IP addresses which are used to indicate the destinations of the UE traffic.

This is illustrated in figure 6.2.6.1-1.

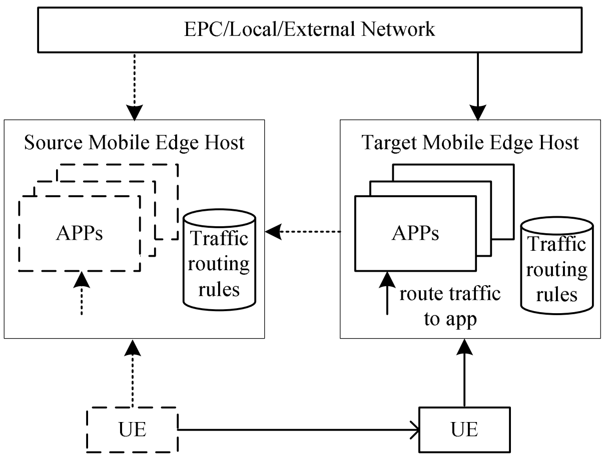

Figure 6.2.6.1-1

6.2.6.2 Solution 1: update traffic routing rules triggered by RNIS

Radio Network Information Service (RNIS) is a service that provides radio network related information to mobile edge applications and to mobile edge platform. Data plane supports the encapsulation of UE traffic before passing it to the network and de-capsulation of encapsulated UE traffic received from network before routing it to the authorized mobile edge applications. Mobile edge platform is responsible for controlling traffic routing rules for mobile edge applications hosted in the same mobile edge host. S-RNIS, S-App, S-ME platform, S-Data Plane reside in the source mobile edge host while T-RNIS, T-Data Plane, T-App, T-ME platform are in the target mobile edge host.

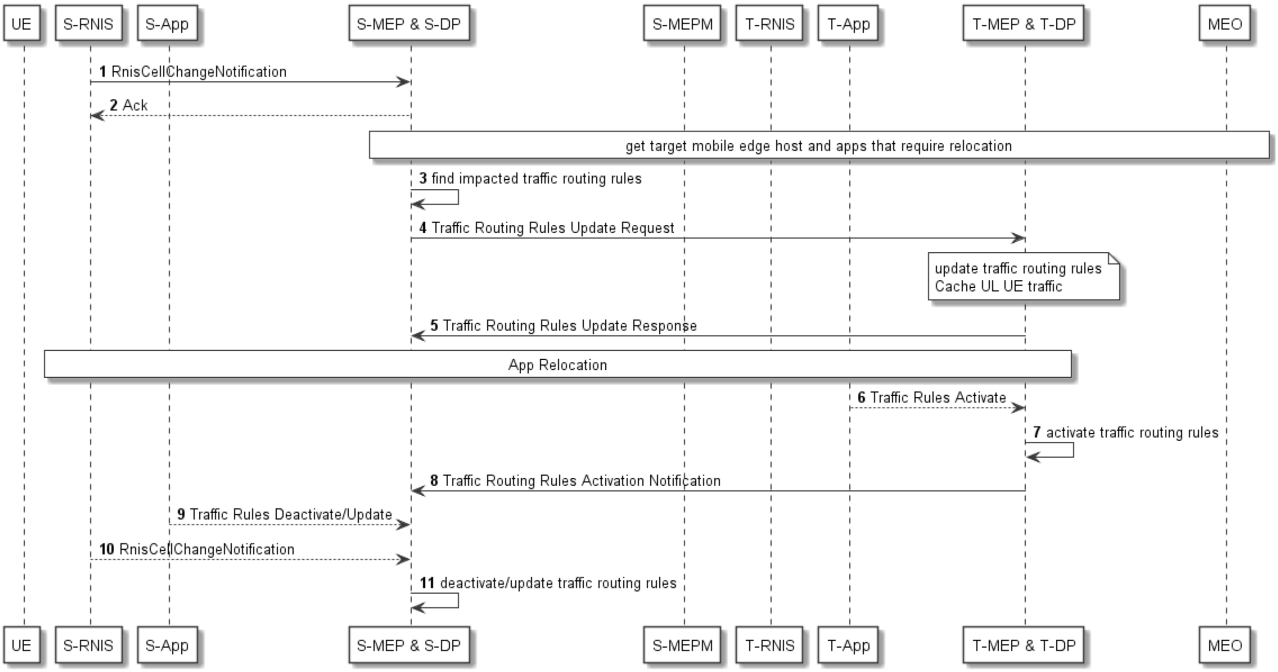

Figure 6.2.6.2-1: Information flow for updating traffic routing rules triggered by RNIS

Logic for this message flow:

- First, notification received from RNIS (step 1, 2).
- Second, Confirm with MEO;.
- Third, Synchronize traffic routing rules (step 3, 4, 5).
- Fourth, Perform app relocation.
- Finally, Activate/Deactivate/Update traffic routing rules (step 6, 7, 8, 9, 10, 11).

The more detailed information is described below:

1) S-MEP receives the RNIS Cell Change Notification (UE IDs, HO Status, target cell ID) from S-RNIS when S-RNIS has detected the UE handover to another radio node. This notification is used to inform the mobile edge platform that the UE handover is in progress. The UE IDs, e.g. IP address or GTP TEIDs, in the notification is used to indicate the UE performing the handover, the HO status indicate the status of the HO procedure, and the target cell ID is used to indicate the handover target.

2) S-MEP returns an acknowledgement of the notification.

   S-MEP then queries the MEPM, which forwards the query to the MEO, for the target mobile edge platform (T-MEP) with the target cell ID received in the notification, which indicates UE location. Taking the operator policy, mobility policy, application requirements, application capabilities, mobile edge host features, the status of mobile edge system, etc., into account, finally mobile edge orchestrator determines the target mobile edge platform and passes the information to the S-MEP for it to contact the T-MEP. The information of applications that require migrating to the target mobile edge host is returned, e.g. application instance ids, application instance IP addresses.

3) With the UE IDs and information of applications that require relocation, S-MEP retrieves the impacted traffic routing rules, which includes the UE IDs as part of the traffic filters, and application information as part of the rules.

4) The S-MEP sends Traffic Routing Rules Update Request (App Information, Impacted traffic routing rules) to the T-MEP. T-MEP updates the traffic routing rules accordingly, and caches the UE UL traffic until the relocation of the application is finished.

NOTE 1: Assume the data plane is able to cache the UL traffic from specified UEs and applications, and the MEP has issued such request to the data plane.

5) S-MEP receives Traffic Routing Rules Update Response from T-MEP. This message denotes that the target mobile edge host is ready to perform application relocation and serve the HO UE instead of source mobile edge host.

   Application relocation procedure is executed between the source and target mobile edge hosts.

NOTE 2: It is FFS in the normative specification stage how to relocate the application state/VM/container, and specify in the information model and APIs.

6) If applicable, target mobile edge application will activate traffic rules via Mp1 interface instantly after finishing the relocation.

7) T-MEP activates the corresponding traffic routing rules via Mp2 interface, which is out of scope of ETSI MEC.

8) S-MEP receives Traffic Rules Activation Notification (App Information, Impacted traffic routing rules) from T-MEP.

9) If applicable, source mobile edge application will deactivate/update traffic rules via Mp1 interface instantly after finishing the relocation.

10) If applicable, source mobile edge platform can receive the RNIS Cell Change Notification from S-RNIS, in order to get the status of UE HO procedure.

11) At receiving Traffic Rules Activation Notification from T-MEP, and possible the Traffic Rules Deactivate message from source mobile edge application, as well as the RNIS Cell Change Notification from S-RNIS, S-MEP deactivates/updates corresponding traffic routing rules.

6.2.6.3 Solution 2: update traffic routing rules triggered by UE UL Traffic

Data plane supports the encapsulation of UE traffic before passing it to the network and de-capsulation of encapsulated UE traffic received from network and its routing it to the authorized mobile edge applications. Mobile edge platform is responsible for controlling traffic routing rules for mobile edge applications hosted in the same mobile edge host. S-App, S-MEP, S-DP reside in the source mobile edge host while T-App, T-RNIS, T-MEP, T-DP are in the target mobile edge host.

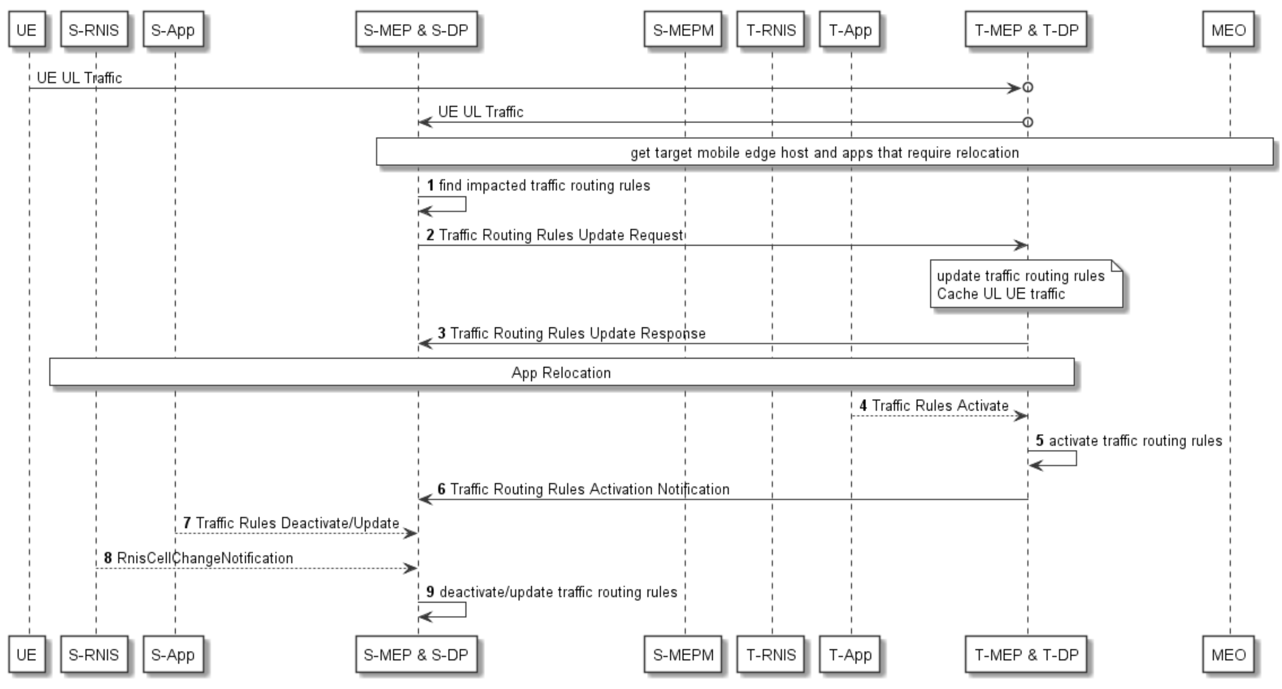

Figure 6.2.6.3-1: Information flow for updating traffic routing rules triggered by UE UL Traffic

Assumption: An IP range is bound to a mobile edge host and the IP address of the mobile edge application instance is within the IP range of the mobile edge host where this mobile edge application is running. At the same time, every mobile edge host is configured with the IP range of the neighbouring mobile edge host. When receiving packets whose destination IP address belongs to the IP range of its neighbouring mobile edge host, the mobile edge host will directly route the IP packets to the corresponding neighbouring mobile edge host.

Logic for this message flow:

- First, Trigger by UE UL Traffic.
- Second, Confirm with MEO.
- Third, Synchronize traffic routing rules (step 1, 2, 3).
- Fourth, Perform app relocation.
- Finally, Activate/Deactivate/Update traffic routing rules (step 4, 5, 6, 7, 8, 9).

The more detailed information is described below:

1) Once the S-MEP detects the change of the path of the UE UL packets, it should obtain the address of the T-MEP through configuration and confirm with the MEPM, which forwards this confirmation to MEO, the resources on the target mobile edge host are available.

   With the UE IDs from the data plane and information of applications that require relocation, S-MEP retrieves the impacted traffic routing rules, which includes the UE IDs as part of the traffic filters, and application information as part of the rules.

NOTE 1: Refer to clause 6.2.5.2 for the specific key issue and solution.

2) The S-MEP sends Traffic Routing Rules Update Request (App Information, Impacted traffic routing rules) to the T-MEP. T-MEP updates the traffic routing rules accordingly, and caches the UE UL traffic until the relocation of the application is finished.

NOTE 2: Assume the data plane is able to cache the UL traffic from specified UEs and applications, and the MEP has issued such request to the data plane.

3) S-MEP receives Traffic Routing Rules Update Response from T-MEP. This message denotes that the target mobile edge host is ready to perform application relocation and serve the HO UE instead of source mobile edge host.

   Application relocation procedure is executed between the source and target mobile edge hosts.

NOTE 3: It is FFS in the normative specification stage how to relocate the application state/VM/container, and specify in the information model and/or APIs.

4) If applicable, target mobile edge application will activate traffic rules via Mp1 interface instantly after finishing the relocation.

5) T-MEP activates the corresponding traffic routing rules via Mp2 interface, which is out of scope of ETSI MEC.

6) S-MEP receives Traffic Rules Activation Notification (App Information, Impacted traffic routing rules) from T-MEP.

7) If applicable, source mobile edge application will deactivate/update traffic rules via Mp1 interface instantly after finishing the relocation.

8) If applicable, source mobile edge platform can receive the RNIS Cell Change Notification from S-RNIS, in order to get the status of UE HO procedure.

9) At receiving Traffic Rules Activation Notification from T-MEP, and possible the Traffic Rules Deactivate message from source mobile edge application, as well as the RNIS Cell Change Notification from S-RNIS, S-MEP deactivates/updates corresponding traffic routing rules.

6.2.6.4 Solution 3: update traffic routing rules triggered by Mobile Edge Orchestrator

The mobile edge orchestrator is able to trigger the application relocation owing to external relocation request, unsatisfied performance requirements, load balancing and disaster recovery. Data plane supports the encapsulation of UE traffic before passing it to the network and de-capsulation of encapsulated UE traffic received from network and its routing it to the authorized mobile edge applications. Mobile edge platform is responsible for controlling traffic routing rules for mobile edge applications hosted in the same mobile edge host. S-App, S-ME platform, S-Data Plane reside in the source mobile edge host while T-App, T-RNIS, T-ME platform, T-Data Plane are in the target mobile edge host.

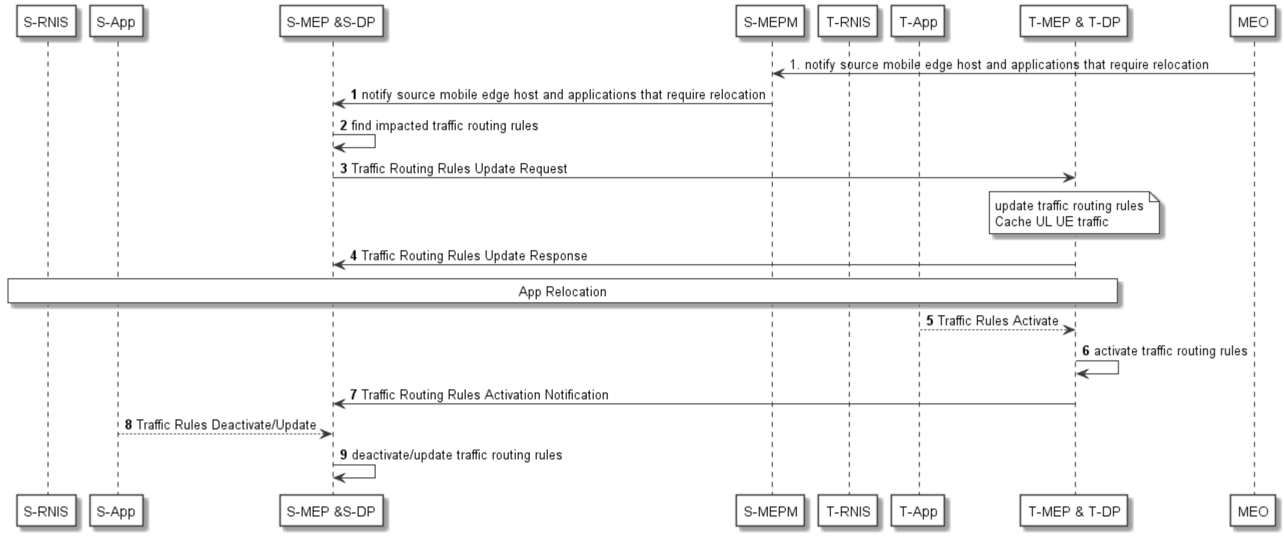

Figure 6.2.6.4-1: Information flow for updating traffic routing rules triggered by MEO

Logic for this message flow:

- First, Trigger by MEO (step 1).
- Second, Synchronize traffic routing rules (step 2, 3, 4).
- Third, Perform app relocation.
- Fourth, Activate/Deactivate/Update traffic routing rules (step 5, 6, 7, 8, 9).

The more detailed information is described below:

1) In case of external relocation requests, unsatisfied performance requirements, load balancing and disaster recovery and in consideration of the operator policy, mobility policy, application requirements, application capabilities, mobile edge host features, the status of mobile edge system, etc., MEO decides to trigger the relocation for specific application between the source mobile edge host and target mobile edge host. It should pass the address of the T-MEP to the S-MEP via MEPM, as well as the information of applications need to be migrated, e.g. App IP addresses, App instance IDs.

2) With the information of applications that require relocation, S-MEP retrieves the impacted traffic routing rules, which includes the application information as part of the rules.

3) The S-MEP sends Traffic Routing Rules Update Request (App Information, Impacted traffic routing rules) to the T-MEP. T-MEP updates the traffic routing rules accordingly, and caches the UE UL traffic until the relocation of the application is finished.

NOTE 1: Assume the data plane is able to cache the UL traffic from specified UEs and applications, and the MEP has issued such request to the data plane.

4) S-MEP receives Traffic Routing Rules Update Response from T-MEP. This message denotes that the target mobile edge host is ready to perform application relocation and serve the UE impacted by the application instead of source mobile edge host.

Application relocation procedure is executed between the source and target mobile edge hosts.

NOTE 2: It is FFS in the normative specification stage how to relocate the application state/VM/container, and specify in the information model and/or APIs.

5) If applicable, target mobile edge application will activate traffic rules via Mp1 interface instantly after finishing the relocation.

6) T-MEP activates the corresponding traffic routing rules via Mp2 interface, which is out of scope of ETSI MEC.

7) S-MEP receives Traffic Rules Activation Notification (App Information, Impacted traffic routing rules) from T-MEP.

8) If applicable, source mobile edge application will deactivate/update traffic rules via Mp1 interface instantly after finishing the relocation.

9) At receiving Traffic Rules Activation Notification from T-MEP, and possible the Traffic Rules Deactivate message from source mobile edge application, S-MEP deactivates/updates corresponding traffic routing rules.

6.2.6.5 Gaps identified by the key issue

Clause 6.2.6 describes three different solutions for traffic rules update, which consist following procedures:

- A: Trigger the traffic routing rules update.
- B: Synchronize traffic routing rules.
- C: Perform application relocation.
- D: Activate/Deactivate/Update traffic routing rules.

Procedure A makes the difference among those solutions. The triggering entity can be:

- RNIS: The traffic rules update triggered by RNIS event of Cell Change Notification is already specified in the current release of ETSI GS MEC 012 [i.6].
- UL traffic change: The traffic rules update triggered by detection of UE UL traffic change may need a new special service (or function) for detecting UE UL traffic change on data plane, which is independent from RNIS. The special traffic path change detection service may be through configuration of a special traffic rule over Mp2 to detect the UE traffic on data plane, which is similar to the triggering parts of solutions specified in clauses 6.2.1.2 and 6.2.1.3.
- MEO: If it is MEO that initiates the ME application relocation for purpose of load balance, or optimizing performance, it needs to define a new message (or API) over Mm3 reference point to trigger the traffic rules update on the Data Plane to forward UE's data to the application on the target MEH.

Procedure B and D described by those solutions are same. Procedure D is supported by the existing specifications. For procedure B, there might be two possible ways to synchronize the traffic routing rules related to the application between S-MEP and T-MEP:

1) the traffic rules related to the application could be synchronized between S-MEP and T-MEP via Mp3 directly;

2) the MEO acquires the latest traffic rules on the S-MEP and updates them to the application on T-MEP via MEPM over Mm5.

NOTE: Mm5 is not specified in the present document.

The Procedure C Application relocation described in this clause can refer to the solution in clause 6.2.5.

6.2.7 Key Issue: Connectivity architecture

6.2.7.1 Description

The most important promise made by MEC low is latency for time sensitive applications. To fulfil this promise, a MEC should be able to support high service continuity for some applications and services. Capability of routing path optimization turns out to be essential for purpose: It has clearly two goals, i.e. for UE:

1) to maintain a unique set of IP addresses to assure routability; and

2) to achieve a possibly short latency despite UE's mobility;

3) To assure service continuity.

To achieve these goals interaction with the connectivity provider is indispensable. Given that 3GPP system has the most advanced mobility capability and session management for mobile users, 3GPP as CPN seems to be the technology of choice to support MEC, without excluding other options. Since MEH network and CPN are independent systems, each follow its own performance and optimization targets , the question on how are MEHs connected to CPN is of fundamental importance for the success of MEC. The complexity of this issue can be reduced to the question on attach point to CPN.

Attach Points:

The term "attach point" refers to CPN architectural reference point, where the MEC system receives connectivity. Different attach points result in different connectivity architectures, each with intrinsic implication to the MEC performance. A complete architecture for MEC system includes MEC system internal software architecture and external connectivity architecture. The later comprises the attach points and the related interfaces and entities. Service continuity depends many things and a proper connectivity architecture is one of the necessary conditions.

Within the context of 3GPP system, e.g. LTE/EPC, potential attach points can be identified, shown in figure 6.2.7.1-1.

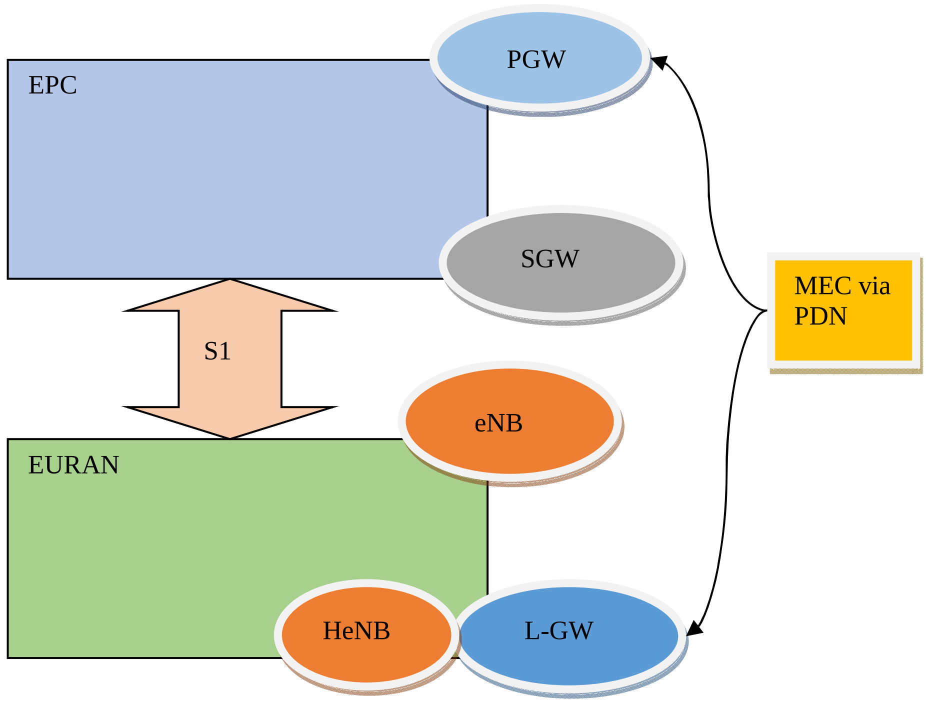

Figure 6.2.7.1-1: Potential attach points of MEH in LTE/EPC network

1) MEH attached at eNB(s):

   - S1-U is the desired reference point. Unfortunately, this interface is not offered as service in the 3GPP architecture yet. The only possible scenario is to utilize the architecture of HeNB collocated with L-GW, a technology primarily designed for CSG and, as such, has limited support for mobility in a broader context that involves non-CSG cells. Regardless whether L-GW is used or not, deployment of this attach point will need additional standardization, if the option is not supposed to be a proprietary solution. A solution of proprietary nature could challenge the 3rd party MEC provision.

2) MEH attached at PGW:

   - SGi is a well-defined reference point. When this attach point is chosen as attach point, a coverage oriented MEC network would require collocation of MEH and PGW, while a performance oriented MEC network would require collocation, or at least closeness, of PGW and SGW (combined architecture). As result, the design is a trade-off between latency and cost, leading to different architectural choices depending on business interest.

To summarize, connectivity provided by CPN needs an adequate architecture that favours adaptive path optimization and service continuity. The necessary interfaces are partially available already in the industry. How and whether those interfaces are to be deployed by MEC are still to be studied and decided. The choice should be based on business model (e.g. 3rd party business, connectivity provider, multi-access scenarios, etc.), relation between ME system components and CPN functions (e.g. security, charging, identity, service vs. MME, AAA, DNS, PCRF, PCC), feasible attach points with associated technical constraints and system evolution path. As bottleneck of the mobility aspects, development of connectivity architecture also depends on interaction with other SDOs. As a starting point, possible attach points are summarized in table 6.2.7.1-1.

Table 6.2.7.1-1: Possible attach points and interfaces for MEH

Network Entities	PGW	SGW	LGW	S1-AP
Interface and conditions	SGi	PGW Combined	HeNB co-located	S1-U
Available	Yes	Yes	Conditional	No
Sub-KI	1	2	3	4

Each of the columns in the table is worth a study as a sub-key issue, each requires an elaborated solution that explains how the associated reference points are used in conjunction with MEC architecture. Moreover, a comparative study of those sub-key issues would shed light on the feasibility of ME system in different business scenarios.

6.2.7.2 Solutions

It is FFS in the normative specification stage.

6.2.8 Key Issue: Session and service continuity mode indication to network

6.2.8.1 Description

In 5G network, there are three types of session and service continuity (SSC) mode. Different SSC mode can guarantee different levels of service continuity.

SSC mode 1: The same UPF is maintained.

SSC mode 2: The network may trigger the release of the PDU Session and instruct the UE to establish a new PDU session to the same data network immediately. At establishment of the new PDU Session, a new UPF acting as PDU session anchor can be selected.

SSC mode 3: The network allows the establishment of UE connectivity via a new UPF to the same application server before connectivity between the UE and the previous UPF is terminated.

Different applications have different service and session continuity requirements. Therefore, in order to achieve efficient control of MEC APPs with different SSC mode requirements in 5G network, some mechanisms should be discussed in this key issue about the coordination between ME APPs on the MEH and the 5G network (e.g. how to indicate the SSC mode requirement of the ME APP to the 5G network).

6.2.8.2 Solutions

The ME Application Mobility Service (MAMS) is a service offered by the MEH to support ME application mobility.

If an ME application includes the SSC mode requirement in the application package, the ME application needs to register with the MAMS for receiving the ME application mobility service. If the MAMS supports the SSC mode requirement, it responds with REGISTERED. Otherwise, it responds with FAILED and an error code.

MAMS may communicate with a ME Application or MEPM to get the SSC mode requirement about ME application, if it is available.

MAMS may communicate to 5G network over MEP/MMEO of MEC system to indicate the SSC mode requirement of ME application, if it is available and supported by the MEC. Otherwise, it relies on the rules of underlying network to handle the SSC mode requirement for forwarding the application data traffic.

MAMS may communicate with the UOF in MEPM to assist application state relocation or application instance relocation.

6.2.8.3 Gaps identified by the key issue

Refer to clause 6.2.5 Application instance relocation. The coordination between ME applications and 5G network is FFS during the normative specification stage.

6.2.9 Key Issue: Application relocations in case of ping-pong handovers

6.2.9.1 Description

In the scenario where the UE moving repeatedly across the ME host serving area's boundary, ME application relocation may also be performed repeatedly between the T-MEH and S-MEH, which cause significant waste of resources and bad user experiences.

In order to avoid repeated application relocations in case of Ping-Pong handovers, some mechanisms should be considered.

Tasks of this key issue will include, but not limited to the following issue:

1) How to avoid Ping-Pong application relocations when UE moving repeatedly across ME host serving area boundary.

2) The detailed information flows between two ME hosts when UE moving repeatedly across ME host serving area boundary.

6.2.9.2 Solutions

It is FFS during the normative specification stage.

7 Gap Analysis and Recommendations

7.1 Gap analysis

Table 7.1-1 summarizes gaps identified in clauses 5 and 6.

Table 7.1-1: Summary of gap analysis

GAP	Clause No	Functional Entity Impacts	Reference Point Impacts
GAP-1	5.4.3.1	Enhancement of ETSI GS MEC 010-2 [i.4] application instance life cycle management.	Mm3 and Mm5
GAP-2	6.2.5.3.1	Add a MAMS in MEP to invoke the application mobility service and transfer the stateful application state between MEPs.	Mp3
GAP-3	6.2.5.3.2	Enhance PAE in ETSI GS MEC 011 [i.5] to identify the stateful application instance, trigger the application instance to capture the service state of stateful application instance being served to the UE, and synchronize the new stateful application instance with the captured state.	Mp1
GAP-4	6.2.2.3	MEO may provide application specific information, via MEPM, that may be taken into account by MEP for defining application specific measures for session and service continuity.	Mm3/Mm5 (may be impacted, depending on the to-be-defined details of the solution) See note
GAP-5	6.2.2.3	The application may provide application specific information that may be taken into account by MEP for defining application specific measures for session and service continuity.	Mp1 (may be impacted, depending on the to-be-defined details of the solution)
GAP-6	6.2.1.2, 6.2.1.3, 6.2.1.6.1, and 6.2.6.5	Defines a new service or enhances the existing service (such as RNIS) to allow the MEP to detect the application traffic path changes.	Mp1, Mp2 and Mp3
GAP-7	6.2.1.6	Coordination for application instance relocation is needed, for example in MEPM.	Mm5 (TBD): for collecting info of application on MEP Mm3: communicate with MEO Mp1: MEP acquires information related to App instance. May need to define a new reference point (TBD): for communication between MEPM and CPN
GAP-8	6.2.6.5	Define a new message(s) or enhance existing message(s) for MEO to trigger the traffic path update, and synchronize traffic routing rules between S-MEP and T-MEP for relocating application.	Mm3 and Mm5 See note
GAP-9	6.2.6.5	Synchronize the traffic routing rules related to the relocating application between S-MEP and T-MEP via Mp3 directly.	Mp3
GAP-10	5.5	May need to enhance MEO and MEPM on-boarding procedure (and parameters) to include the requirement and parameters in ETSI GS MEC 010-2 [i.4].	Mm1 and Mm3
GAP-11	5.5	May need some additional validation procedure for relocating the application instance or state in ETSI GS MEC 010-2 [i.4] and ETSI GS MEC 011 [i.5], if re-using the instantiation procedure.	Mp3, Mm5 See note
NOTE: Mm5 is not specified in the present document.			

7.2 Recommendations

ME application mobility is an optional feature of MEC system, which allows relocating the application from one MEH to another. In order to provide flexibility in application relocation, ME application mobility (feature) could be supported as a service of MEP to offer to applications running in the MEC system.

As ME application mobility requires coordination between MEHs which are either via direct communication or through MEO over different reference points, it is recommended to define a normative stage 2 specification for support of ME application mobility. The specification may include:

- end to end information flows for support of application relocation;
- application mobility service subscription and configuration;
- application mobility initiation including various detection or triggering procedures;
- application relocation validation procedure;
- application relocation execution, such as operational state capture, application instance or state transfer and synchronization procedures;
- application relocation completion, including application termination procedure, traffic routing rule update procedures, etc.

The normative stage 2 specification for support of ME application mobility may contain corresponding information models associated with individual procedures to provide necessary information for development of ME application mobility stage 3 specifications.

It recommends that the development of API specification for support ME application mobility follows existing MEC specification documentation structure to re-use existing APIs to avoid duplication:

- enhance the present specifications, if the support of ME application mobility requires to amend the API specified by the existing specifications, such as ETSI GS MEC 010-2 [i.4], ETSI GS MEC 011 [i.5], etc.;
- develop a new specification, if the support of ME application mobility requires to define API over the reference point which is not specified by existing specifications, like the reference point Mp3.

For the interface to underlying network (i.e. 5G network), it requires co-operation between MEC and 3GPP, and is FFS.

History

Document history		
V1.1.1	October 2017	Publication