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

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

Disclaimer

The present document has been produced and approved by the Mobile Edge Computing (MEC) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

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

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

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.

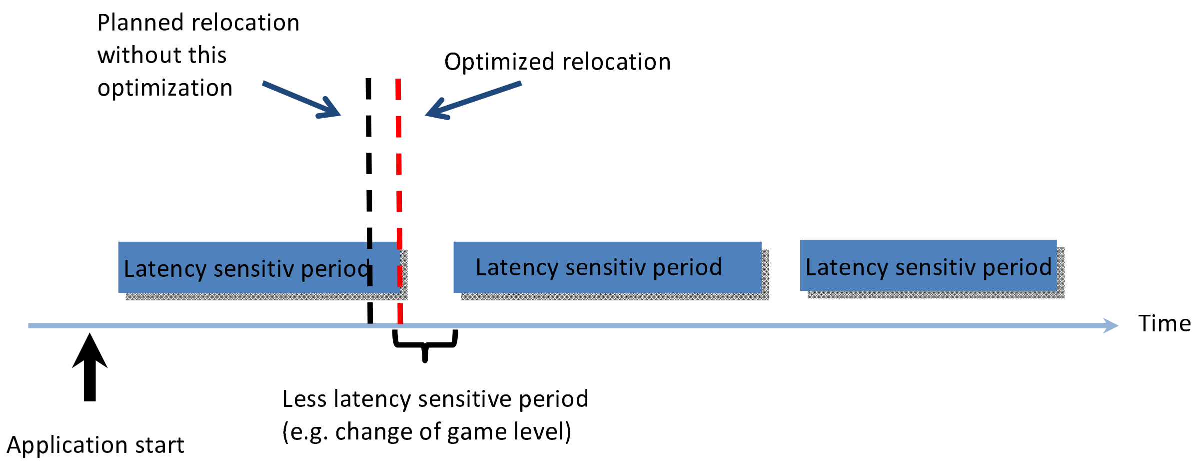

Figure 4.2-1: Optimal time window

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]).

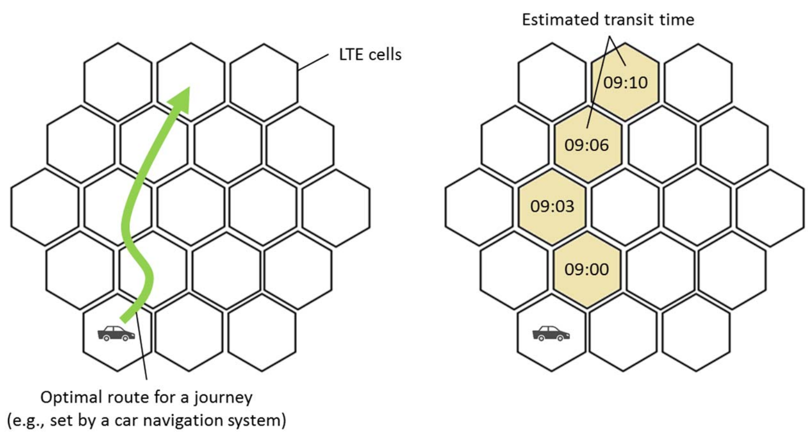

Figure 4.3-1: The prediction of handover timing for a connected car use case

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.

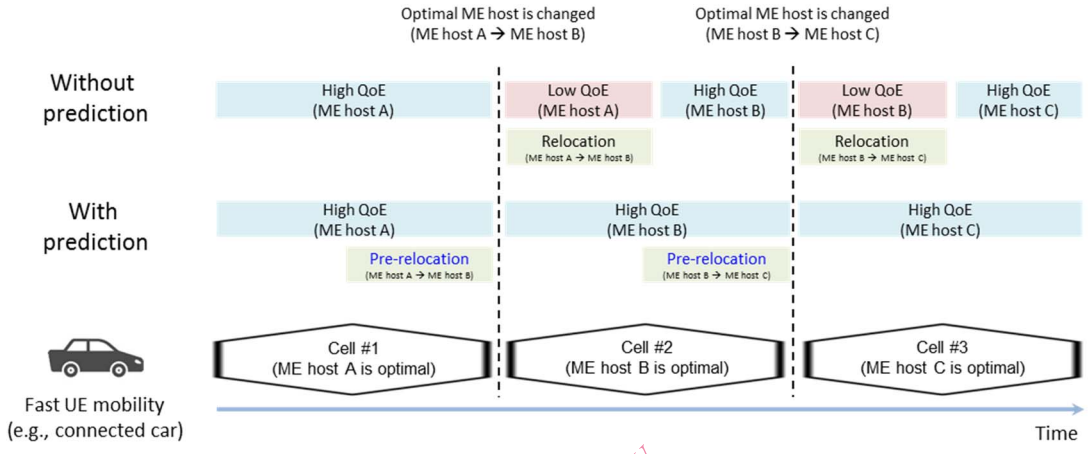

Figure 4.3-2: Pre-relocation of application state information

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.

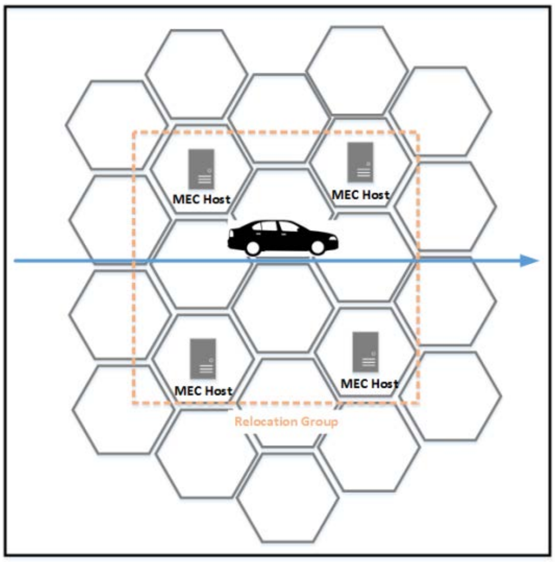

Figure 4.4-1: Preconfigured Relocation Group

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.

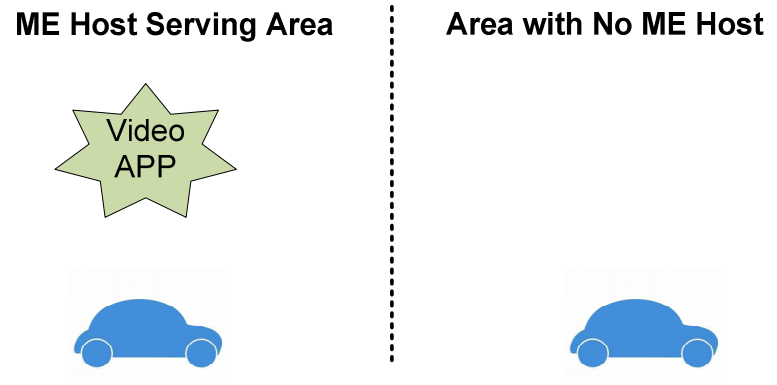

Figure 4.5-1: Use case of service continuity with the UE moves in/out of ME host serving area

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.

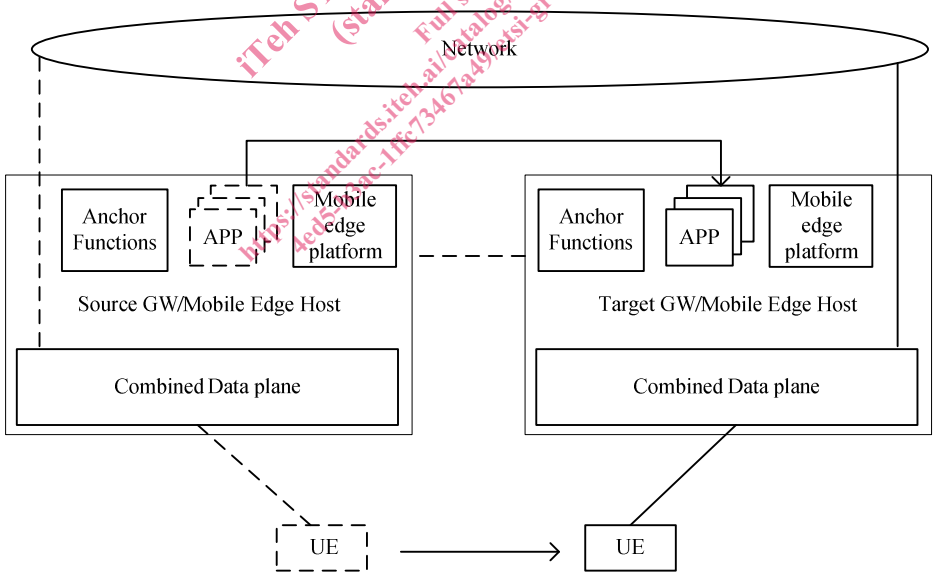

Figure 4.7-1: Mobility in GW-based Use Case

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.