MiWEBA

EU Contract No. FP7-ICT-608637

D3.4 Support of handover for multi-layer technologies

Contractual date:	M33
Actual date:	M33
Authors:	**Ingolf Karls**, Hailan Peng, Kazuya Moriwaki, Hidekazu Shimodaira
Participants:	
Work package:	D3.4
Security:	MiWEBA internal only
Nature:	Report
Version:	1.0
Number of pages:	25

Abstract

This report provides analysis results regarding the support of vertical handover for multi-layer technologies in particular for a heterogeneous network architecture comprised of LTE macro and millimetre wave small cells. It is based upon present handover techniques of 3GPP and IEEE as well as interworking ones between 3GPP and non-3GPP networks. Performance metrics are proposed to evaluate handover techniques. Challenges for vertical handover are systematically shown as well as future study items are described. Moreover, achievements on vertical multi-layer handover in the MiWEBA project are also summarized in this deliverable.

Keywords

Mobility management, handoff, vertical hand over, multi-layer networks, heterogeneous networks, next generation networks, IEEE, 3GPP

Contents

Abbreviations

3GPP	3rd Generation Partnership Project
AAA	Authentication, Authorization & Accounting
AP	Access point
ASN	Access service network
BBE	Break before entry
BF	Beam forming
BS	Base station
BSC	Base station controller
CA	Carrier Aggregation
CA	Carrier aggregation
CAC	Call admission control
CoMP	Coordinated Multipoint
CRE	Cell range expansion
DL	Downlink
EBB	Entry Before Break
eNB	Evolved Node B
ePDG	Evolved packet data gateway
EPS	Evolved Packet System
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FD MIMO	Full dimensional MIMO
FeICIC	Further enhanced inter-cell interference coordination
G/U/L	Multi-RAT base station (GSM/UMTS/LTE)
HeNB	Home eNodeB
HetNet	Heterogeneous network
HO	Hand over
HSS	Home Subscriber Server
ICIC	Inter-cell interference coordination
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force

MAC	Media access control
MAHONAHO	Mobile / Network Assisted Handover
MBB	Make before break
MCHO	Mobile Controlled Handover
MDP	Markov Decision Process
MIH	Media Independent Handover
MIHF	Media independent HO functions
MIP	Mobile IP
MME	Mobility management entity
MNO	Mobile network operator
MS	Mobile station, see UE
MSA	Multi-Stream Aggregation
NAR	New Access Router
NAS	Non-Access Stratum
NCHO	Network Controlled, Handover
PAR	Previous Access Router
PCell	Primary cell
PDCCH	Physical Downlink Control Channel
PDN-GW	Packet Data Network Gateway
PHY	Physical layer
QoE	Quality of experience
RACH	Random Access Channel
RAT	Radio access technology
RLC	Radio link control
RNC	Radio network controller
RRC	Radio resource control
RSS	Received signal strength
SCell	Secondary cell
SGSN	Serving GPRS Support Node
S-GW	Serving gateway

SNR	Signal to noise ratio
SON	Self-organizing network
SRC	Single radio controller
UE	User equipment
UL	Uplink
WPL	Work Package Leader(s)

Executive Summary

This report summarizes analysis and exploration results in the area of vertical handover in wireless multi-layer heterogeneous networks in particular being comprised of millimetre wave small cells and LTE macro cells in the environment of the MiWEBA project. A heterogeneous network implementing millimetre wave small cells and LTE macro cells as explored in MiWEBA requires a new and extended mobility management and interworking including new and extended handover techniques to support efficiently all forms of mobility including UE, user and service mobility.

Based on the analysis of handover of IP based data traffic in 3GPP and IEEE networks, it is shown that all handover phases including network discovery and selection, network attachment, configuration, security association, binding update and media re-routing are subject for improvements and extensions to adapt to the characteristics of millimetre wave small cells in a multi-layer environment. It is revealed that millimetre wave small cell heterogeneous networks bring a lot of challenges at PHY, MAC and routing layers for handover. Therefore, the communication architecture and protocols, especially at the layer 1 and 2, need to be revised to adapt signalling and resource allocation and cope with a new level of dynamics in channel attenuation, directionality, and blockage.

The gains of beamforming have to be paid for by far more complicated mobility management and handover techniques. A lot more frequent handover, even for static UEs, could become a potential drawback of millimetre wave small cell systems due to their vulnerability to blockage. The dense deployments of millimetre wave small cell APs intensify frequent handovers between adjacent APs even more. The loss of precise beamforming data will become a further criterion for handover and re-association. To avoid drastically increased handover times, rates and severely decreased handover efficiency some first mitigation options are proposed to reduce the overhead and delay of handover and re-association.

An outlook for many future studies is given to address the optimization and extensions of current 3GPP and IEEE handover solutions. Some of the issues have been well studied in work package 3 and work package 4 of the MiWEBA project.

1 Cross-layer handover techniques for cellular systems

There are three major types of mobility handover (HO) to be supported, which are terminal, personal and session mobility. Here terminal mobility is used. A survey on terminal mobility management in heterogeneous networks (HetNets) is provided in [1] and [2], where HO is defined as a technique by which user equipment (UE) gets handed over from one eNB to another eNB. Handover techniques are needed to allow UEs to move from cell to cell while maintaining service continuity. When the term handoff is used it is synonymously to hand over. HO within one radio access technology (RAT) is defined as horizontal HO and HO across different RATs is defined as vertical HO. HO in 3GPP E-UTRAN and IEEE is carried out differently whereas the differences between E-UTRAN and IEEE 802.11 for example are greater than between E-UTRAN and IEEE 802.16m as seen further below. A UE HO between eNB and access points (AP) in the same subnet can be executed at layer 2 whereas a horizontal and vertical UE HO across several subnets and domains requires layer 3 involvement.

The HO decision at Layer 2 takes into account criteria like the received signal strength (RSS), signal to noise ratio (SNR), and channel availability. In case of vertical HO, the layer 2 HO triggers a layer 3 HO. The timings of the HO preparation and initiation, HO completion and HO exception handling at the two layers should be synchronized efficiently to reduce HO delay. For vertical HO, additional criteria such as network traffic conditions, available bandwidth, velocity of the mobile and mobility pattern, and user preference could be taken into account. An optimization decision process using techniques like pattern recognition and Markov Decision Process (MDP) modelling could be performed across several HO criteria to select the best target network for vertical HO. For example, IEEE 802.21 specifies a framework for exchange of HO messages across different wireless access networks via media independent HO functions (MIHFs). Macro mobility, which refers to mobility across different subnets and domains, is generally handled at Layer 3 by using for example the mobile IP (MIP) IETF protocol whereas micro mobility is handled by retaining the HO decision control in Layer 2 as long as the mobile user does not cross the domain.

Current 3GPP E-UTRAN standards support already some of the required cross-layer HO functionalities, such as horizontal intra-system and inter-frequency handovers. 3GPP TS 36.300 [3] specifies handover (HO) as procedure that changes the serving cell of a user equipment (UE) in the state RRC_CONNECTED. 3GPP TS 23.402 "Architecture enhancements for non-3GPP accesses" [4] specifies the stage 2 service description for providing IP connectivity using non-3GPP accesses to the Evolved 3GPP Packet Switched domain where advanced network discovery and selection function (ANDSF) provides information for access networks that are available to the UE. Handover optimizations applicable to all non-3GPP accesses are still void.

HO techniques are one of the crucial components in cellular network mobility management which have severe latency requirements on service interruption time due to its major impact on user quality of experience (QoE). In the past the interworking for example HO between IEEE 802.16m and LTE took different frameworks in different directions due to the relatively separate standardization work done in IEEE and 3GPP. So for example for the HO from IEEE 802.16m to 3GPP LTE, two layer-2 transfer protocols are possible. Firstly using an IEEE 802.16m generic MAC layer-2 transfer tunnel transferring inter-RAT HO signalling data between an MS and corresponding network entities directly for HO preparation and execution. Secondly using the IEEE 802.21 media-independent HO protocol to allow data exchange between an MS and an IEEE 802.21 server, which serves as a proxy for the MS to negotiate inter-RAT HO with other network entities. With these single-radio and dual radio HO from IEEE 802.16m to 3GPP LTE are supported. For HO from 3GPP LTE to IEEE 802.16m only a dual radio procedure via layer 3 transport is supported.

Therefore, state of the art HO techniques for example 3GPP LTE-Advanced and IEEE 802.16m were investigated in MiWEBA usage scenarios regarding their capabilities to be used for cross-layer HO and to minimize and optimize HO latency to fulfil quality of experience requirements. 3GPP LTE–Advanced and IEEE 802.16m were chosen exemplarily as specifications which have got already very stringent requirements on interruption time during HO in order to provide QoE for current and future mobile services.

IEEE 802.16m [5] and 3GPP LTE/LTE-Advanced air interface specifications [3], [6] provide enhanced link layer handover mechanisms with short HO interruption time with backward compatibility, support of multiple deployment scenarios and inter-radio access technology (RAT). Besides clear latency requirements on HO the flexibility to support heterogeneous networks (HetNets) now and in future is likewise significant. The cross-layer techniques for HetNets require good scalability of the HO framework for mobility, dense deployment of small cells, support of many possible HO scenarios and multi-carrier aggregation.

1.1 Handover schemes 3GPP E-UTRAN

3GPP has specified highly optimized service and technology specific mobility protocols to hand over voice sessions. The mobility management entity (MME) and the serving gateways (S-GW) are mostly defined as independent network entities, where the MME hosts MME selection for HO with MME change and serving GPRS support node (SGSN) selection for HO to 2G or 3G access networks. S-GW is the local anchor point for inter-eNB HO. E-UTRAN comprises of enhanced Node Bs (eNBs) and Home Enhanced Node Bs (HeNBs), which are connected by X2 interfaces, and linked to the evolved packet core (EPC), the MMEs and S-GWs, via S1 interfaces. Handover occurs between cells of a single eNB (intra-eNB handover), between eNBs of the same MME/S-GW (inter-eNB intra-MME handover), or between eNBs of different MMEs. Inter-eNB handover without changing MME is typically done over the X2 interface. If there is no direct link, the S1 interface is used (Figure 1-1).

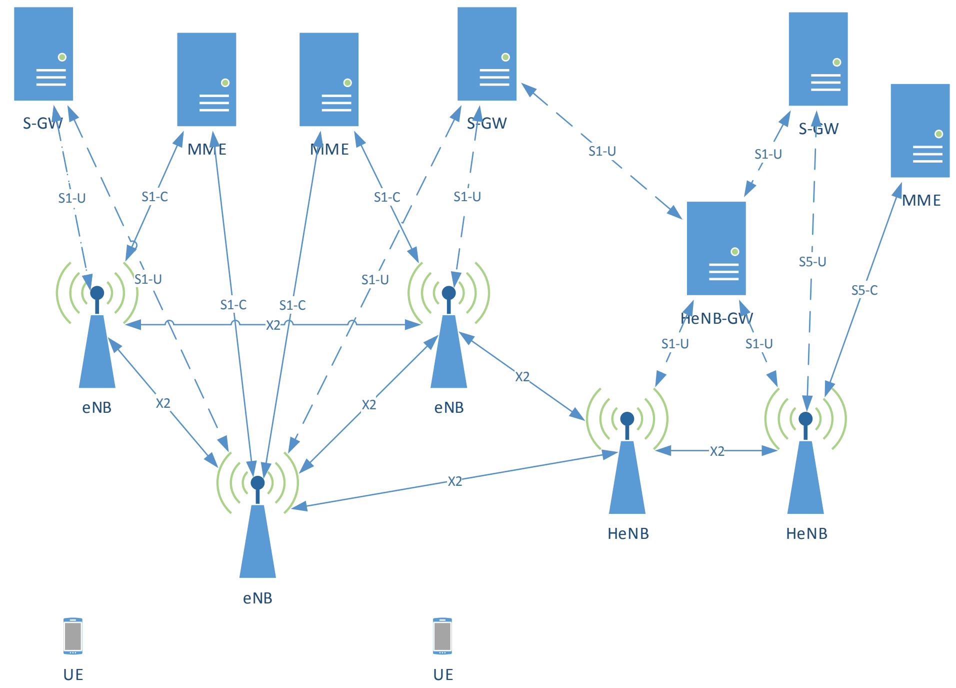

Figure 1-1: Network architecture E-UTRAN

Handover in 3GPP LTE-Advanced systems is a hard handover and network-controlled; so the network decides the target eNB for the UE to hand over to. The handover procure contains 3 parts, firstly the handover preparation and initiation, secondly the handover completion and thirdly the handover exception handling. The HO preparation starts by sending a HO request message from the serving eNB to each candidate eNB which includes a resource reservation at the target eNB (new station identifier, etc.), the setup of a data forwarding path between serving eNB and target eNB, and some optional radio resource reservation such as allocating dedicated random access code for the UE to access the random access channel (RACH) during network re-entry. The target eNB answers with a HO response message to the HO request, which carries the aforementioned data forwarded to the UE via the serving eNB. A simplied message flow is shown in Figure 1-2.

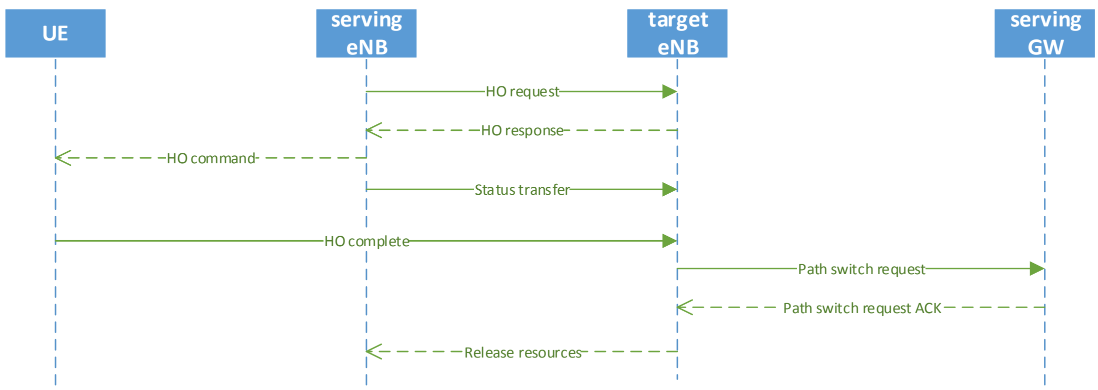

Figure 1-2: Message flow handover within LTE-Advanced

When the HO decision is done, an AAI_HO-CMD message is sent by the serving eNB to the UE to trigger HO execution. Upon reception of the HO command, the UE disconnects from the serving eNB and performs network re-entry to the target eNB specified in the HO command. The network re-entry procedure includes downlink/uplink synchronization, request for uplink grants via RACH and a security key update. The AAI_RNG-REQ/RSP control signalling between UE and target eNB verifies identity and completes this part. Finally, the network data path gets switched from the original eNB to the new eNB as requested by the target eNB (Figure 1-1).

The UE is required to finish the network re-entry within a defined time, which is exchanged during HO preparation as UE timer value. Serving eNB and target eNB retain the UE context for this time. If the UE was not able to do the HO in the time given, the HO is considered as failed and exception handling starts.

3GPP LTE TS 36.300 [3] specifies the HO framework of E-UTRAN (figure 1-2). The single-carrier HO framework of LTE Release 8/9 was very similar compared to IEEE 802.16m but with fewer optimization options trading off complexity versus HO interruption time. The HO framework is a hard HO like IEEE 802.16m, nevertheless it allows only break before entry (BBE) operation and is more tightly network controlled in terms of HO initiation and target selection. 3GPP Release 13 specifies HO in case of carrier aggregation introducing primary cells (PCells) and secondary cells (SCells) serving an UE. PCell can be only changed with HO procedures whereas at intra-LTE HO, radio resource control (RRC) can add, remove, or reconfigure SCells to be used with the target PCell too [7].

1.2 Handover schemes IEEE 802.16m

IETF and IEEE defined mobility protocols for IP traffic in particular for IPv4 and IPv6, which is different compared to 3GPP since IP traffic is radically more diverse than cellular voice. Since it is not easy to design an access specific handover optimization, a not optimized mobile IP HO can take sometimes seconds and degrades the QoE. The Access Service Network (ASN) is a single entity playing the roles of both the MME and S-GW. IEEE 802.16m supports handover between BSs of the same ASN and between BSs of different ASNs (Figure 1-3).

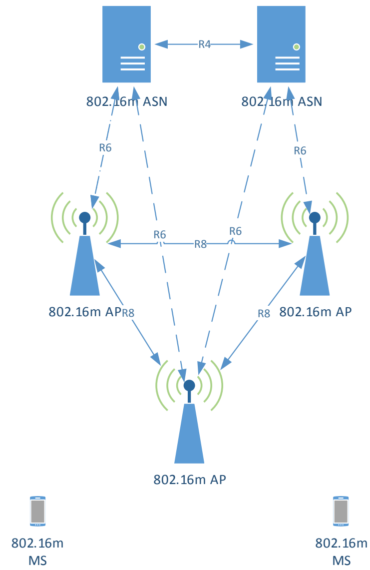

Figure 1-3: Network architecture IEEE 802.16m

Handover in IEEE 802.16m is a hard handover and network-controlled; that is, the network decides the target BS for the MS to hand over to. The handover procure is very similar to 3GPP LTE-Advanced comprising as well handover preparation and initiation, handover completion and handover exception handling. But there are additionally advanced features to improve the HO procedure and to reduce further latency in both PHY and MAC layers.

These features are seamless, entry before break (EBB) and multicarrier HO. Seamless HO is a network re-entry procedure in which the MS is able to send or receive packet data units (PDUs) with the target BS before initiating a network re-entry control message transaction where the interruption time is very low. The baseline HO procedure is BBE, which is the commonly used hard HO procedure where a MS disconnects from the serving BS before executing HO at the target BS in BBE HO with a data interruption. With EBB the MS does a network re-entry at the target BS during the negotiated network re-entry procedure intervals while maintaining links with the serving BS for data exchange. So the total interruption time during handover gets minimized. MSs with multicarrier capability can accomplish HO by keeping links with the serving BS and doing actions like network re-entry at the target BS in a parallel manner giving zero data interruption time during HO. A simplified message flow in IEEE 802.16m for a single and multi-carrier implementation is shown in Figure 1-4

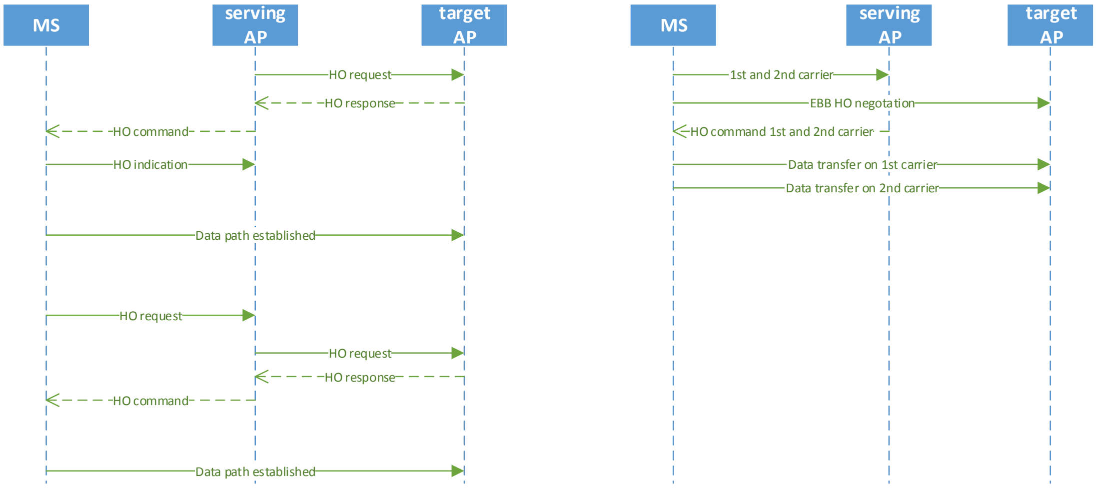

Figure 1-4: HO for single- and multicarrier IEEE 802.16m

.

1.3 Vertical multi-layer handover between 3GPP and IEEE 802.11

In order to provide high-speed connectivity and a desired QoE, mobile network operators implement interworking between 3GPP LTE and IEEE 802.11 networks (Wi-Fi hotspots), where a trusted Wi-Fi network operated by 3GPP operators and/or a non-trusted Wi-Fi network operated by aggregators, can interoperate with an LTE core network. 3GPP defines two types of IEEE 802.11 access from LTE, one is trusted access (MNO built IEEE 802.11 access with over-the-air encryption and a secure authentication method which provided by the MNO) and untrusted access (any type of IEEE 802.11 access that either is not under control of the MNO or that does not provide sufficient authentication, encryption, etc.). The HO control and data flow is specified in 3GPP TS 23.402 [4]. An example of a message flow is given in Figure 1-5.

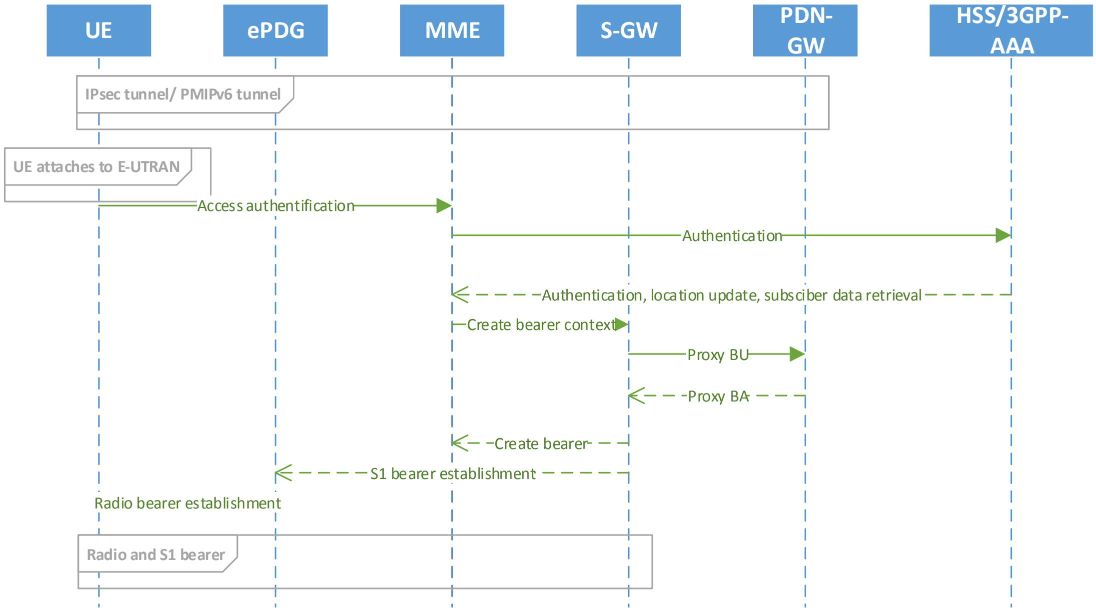

Figure 1-5: Message flow untrusted 802.11 LTE handover

2 Performance requirements of vertical handover

To get a first glimpse into the challenges for integrating millimetre wave small cells with LTE macro cells into a heterogeneous network currently available performance metrics for handover were analysed. In order to specify required QoE performance for IP data traffic, E2E delay, network jitter and packet loss are mainly used. QoS as subset of QoE is defined in 3GPP TS 23.107 [8] and ITU-T G.114 [9]. A vertical multi-layer handover will add definitely to packet loss and network jitter and will sacrifice the overall data throughput because of data interruption and re-transmission. For example, in current 3GPP and IEEE networks users can already experience without any optimization that a UE could get a handover delay of up to 10 seconds depending on the type of mobility protocol, handover and access network, which might be some upper bound of handover time.

Nevertheless, HO events in a small cell millimetre wave environment are caused by fixed and moving obstacles, by user movement and rotations and human being blockage and are reason to assume that the number of HO will drastically increase. The HO between different layers of a heterogeneous LTE-advanced system as a critical parameter to plan the best way of interactive coordination within the network for the proposed HetNet is investigated for example in [10]. The proposed HO algorithms there take into account multiple factors in both HO sensing and decision stages, based on signal power reception, resource availability and HO optimisation, as well as prioritisation among macro and small cell stations, to obtain maximum signal quality while avoiding unnecessary handovers. Figure 2-1 shows exemplarily HO categories being taken into account implementing a heterogeneous network with millimetre wave small cells and LTE macro cells. It is clear that all this makes the HO process even more sophisticated and time consuming.

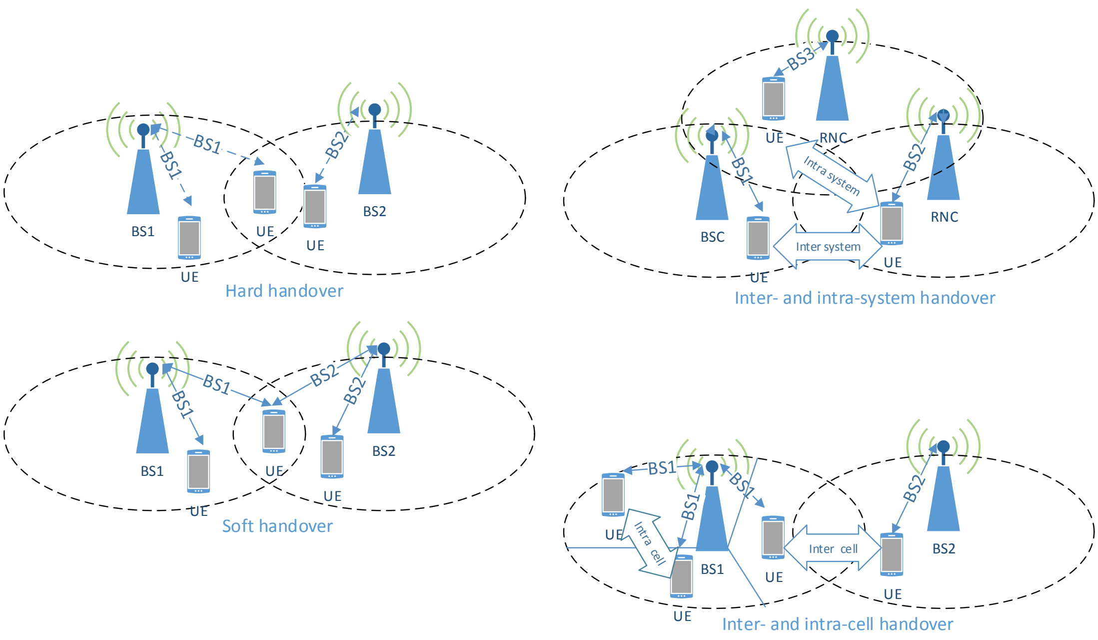

Figure 2-1: HO types

The HO performance could be estimated by looking at performance metrics as follows (this is not a comprehensive list, but shall be used as a starting point):

- HO time
- HO rate
- HO packet loss

Additionally, the work at IETF around IP mobility was used for definitions and terminology of vertical handover topics [11]. Here handover is the process by which an active MS changes its point of attachment to the network in a minimal interruption time whereas the access points are of different radio access technology type. The HO time or latency is defined as the difference between the time a MS is last able to send and/or receive an IP packet by way of the previous access router (PAR), and the time the MS is able to send and/or receive an IP packet through the new access router (NAR).

2.1 Handover time

To estimate the performance of HO techniques, the HO time is very meaningful metric since unnecessary handovers in heterogeneous networks are considered as a critical cause for network degradation in systems and difficulties in capacity and QoE for the integration of various sub-networks. The HO rate or time is typically the time during which the UE cannot exchange packets with any eNB or millimeter wave small cell AP since it is in the transfer phase from one cell to another processing HO preparation and initiation, HO completion and HO exception handling. It is the time interval during which an UE cannot send or receive any data packets while it performs HO between different eNBs or APs. For example, typical HO interruption times in E-UTRAN are 27.5 ms for intra-frequency HO, 40 ms for inter-frequency HO within a spectrum band and 50 ms for inter-frequency HO between spectrum bands [12]. So a typical average handoff interval could add up to several seconds, although for certain type of user actions the average handoff interval can be as low as 0.75 seconds for intra-system HO.

Challenges, despite its advantages in improved capacity and QoE are definitely the time and conditions for handover. To achieve end-to-end QoE fast HO schemes to reduce the HO latency within a single domain and across heterogeneous access domains are required. This will be get particularly challenging for small cell millimetre wave systems asking for a handover time in the ballpark of less than 20 ms.

2.2 Handover rate

The HO rate is determined by the HO blocking probability as the probability that an UE cannot complete the HO within the network dwell time and the HO probability including unnecessary handover probability, missing handover probability and wrong decision probability.

2.3 Handover packet loss

The HO packet loss is the sum of all lost packets destined for an UE during the UE's HO.

3 Challenges of vertical multi-layer handover

The exploration shows that the HO of a session across heterogeneous wireless networks using RATs like LTE and millimetre wave affects significantly the user QoE in particular due to the HO time (latency, delay). The HO time is produced by message exchanges, multiple database access, service negotiation and renegotiation, different service definitions and provisions due to different service providers, UE mobility, and resource availability. Therefore, a heterogeneous network comprised of millimetre wave small cells and LTE macro cells requires a new and extended mobility management and interworking to cope with these challenges including handover techniques to support efficiently all forms of mobility including UE, user and service mobility. Based on the analysis of HO of IP based traffic in 3GPP and IEEE networks, all HO phases including network discovery and selection, network attachment, configuration, security association, binding update and media re-routing need further studies. The challenge is to adopt each of these phases to the characteristics of millimetre wave small cells in a multi-layer environment. Parts of these challenges and issues have been studied in MiWEBA WP3 and WP4 to support seamless handover for multi-layer technologies. The details can refer to the deliverables D3.1 [13], D3.2 [14] D3.5 [15] and D4.3 [16].

3.1 Deployment and usage scenarios

Firstly, a challenge is the millimetre wave small cell deployment with scenarios like street canyon, open space, indoor. Users will have frequently to hand over from one small cell millimetre wave access point to another AP whenever its radio link is disrupted by obstacles [17] contributing to an increase in the HO rate. For example, in the event of shadowing, one small cell millimetre wave AP might have rapidly hand over the UE to another AP in the small cell cluster. Moving obstacles, users' hand motion and changes in UE orientation could all contribute to multiple, successive hand overs. This kind of HO depends heavily on deployment scenario amongst other impacts.

3.2 Quality of experience

Secondly maintaining service connectivity with best user QoE at any kind of UE, user and service mobility across several types of RATs is a major challenge of heterogeneous networks. This is the process by which a UE gets handed over seamlessly or with the least disruption to the service from one RAT to another. It gets initiated either by radio link issues e.g. the link to its current eNB and/or small cell AP either fails or is too weak to provide minimum data rate or by the network. In MiWEBA HOs from millimetre wave to wide area systems and vice-versa like LTE/LTE-A are considered.

It was observed that the mobility performance in a heterogeneous network with LTE macro and millimetre wave small cells might get inferior to homogeneous networks since the number of handovers and handover failures is going to increase and the small cell to macro cell handover might show worse performance. So the impact of UE speed on mobility is significant e.g. fast moving UEs should not be transferred to and kept at small cell layer. So for example, the priority definition in the HO algorithm according to the different mobility patterns and dynamic network conditions as well as the backhaul and front haul realisation and characteristics will get strongly challenged.

3.3 Multiple RATs

Thirdly the HO management among macro and small cell millimetre wave parts of the network presents another key challenge, based upon creation of a neighbour cell list, HO procedure and traffic model for macro and small cell millimetre wave networks. The challenge is the variety of RATs itself, e.g. measuring reference signal received power (RSRP) and quality (RSRQ) values by the UE and reported via measurement report (MR) for all RATS involved, already seen in the interworking of 3GPP E-UTRAN and IEEE 802.11 and IEEE 802.16m. Whereas HO in 3GPP E-UTRAN is very mature this challenge arise from the necessity of interworking of layer 2 and layer 3 for vertical HO which requires a fine-tuned timing of HO preparation and initiation, HO completion and HO exception handling. And there is the challenge of macro and micro-mobility for the UE, user and services which gets handled primarily at layer 3 involving mobile IP concepts from IETF. It can be seen that vertical and horizontal HOs are very different from the 3GPP E-UTRAN perspective while micro and macro mobility have to be distinguished from an administrative domain perspective.

Just to highlight some of the issues which are up for further studies in relationship with the implementation of heterogeneous networks based upon millimetre wave small cells and LTE macro cells further research emphasis is needed on the multi-layer handover in case of carrier aggregation of contiguous and non- contiguous carriers aggregated from multiple different RATs, spectrum sharing of licensed, light-licensed and un-licensed bands and C-plane U-plane split architectures.

3.4 Small cell energy efficiency

Fourthly the strive for energy efficiency in heterogeneous networks gets challenged by switching on and off of millimetre wave access points and base stations envisaged as an effective and efficient solution to save energy in next generation networks. But this behaviour will have a strong impact on the availability of small cell millimetre wave APs and therefore on HO and is up for further study as well.

Another issue are the DL and UL power relationships at the cell edge. Heterogeneous networks perform worse compared to a homogeneous network due to more cell borders and smaller handover regions [18]. The most critical part is the small cell to macro HO due the fact that the rapidly changing signal strength within a small cell and the strong macro cell interference at the cell edge, particularly when using CRE, therefore SON Mobility Robustness Optimization is essential and needs to be investigated further.

3.5 Security

Fifthly, another topic for further studies is the security key management in the handover between macro cell eNBs and small cell millimeter wave APs (example is given in figure 3-1). Currently the Evolved Packet System (EPS) supports intra- and inter-MME handovers according to the anchor points involved. Since the vulnerable period is the time difference between a de-synchronisation attack on a eNB or AP and the time of updating the root key as shown in [19], it is challenging to determine the optimal interval of security key updates in particular implementing small cell millimeter wave systems.

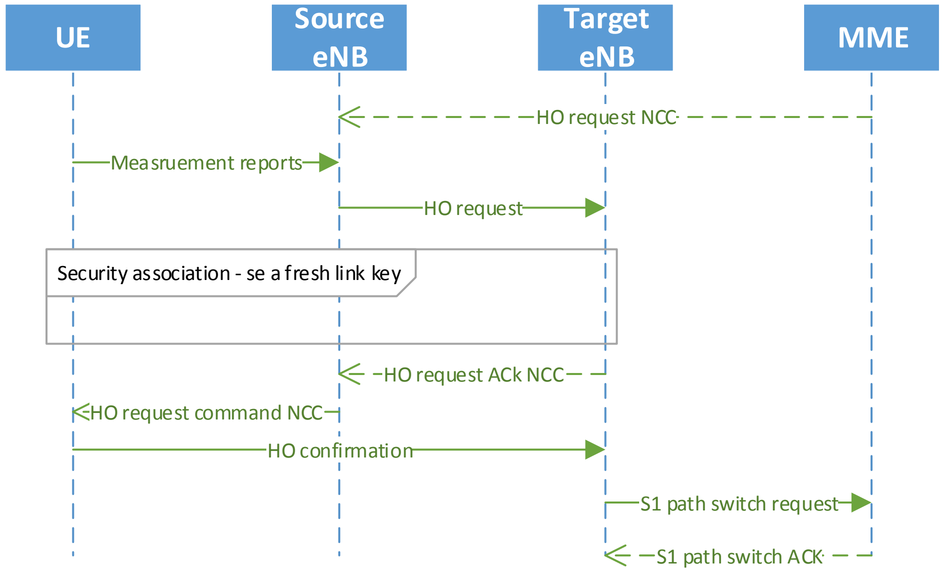

Figure 3-1: Message flow inter-cell HO with security key management in E-UTRAN

4 Vertical multi-layer handover

Based on the analysis of current hand over (HO) techniques and performance criteria, lessons learned gets adopted for the MiWEBA heterogeneous network case with emphasis on proposals to solve the challenges of HO time reduction, HO rate decrease and HO efficiency improvement. Handover control mechanism foreseen are i.e. mobile controlled handover (MCHO), network controlled handover (NCHO) and mobile / network assisted handover (MAHO/NAHO).

4.1 Handover criteria

The HO process is contained of HO preparation and initiation, handover completion and handover exception handling, where the decision to hand over a connection is based primarily on RSS, SNR and channel availability as in horizontal HO. In case of vertical multi-layer HO several other criteria might be added. These are for example radio resource management context related topics like power consumption, spectrums availability, network load, battery life time, network congestion, mobility of UE, user and service as well as user preferences and cost. Doing so the UE context including security (e.g. authentication, authorization and accounting (AAA)) and QoE set-ups needs to be transferred between source and target eNBs and/or millimetre wave small cell APs too.

Some examples using a subset of criteria above are given as follows:

- Small cell HO: HO is initiated only if the signal strength of the current small cell AP drops below a threshold value, using a RSS threshold and a hysteresis avoids handing over a UE too quickly to the neighbouring small cell AP and the ping pong effect at the border of the small cells. This is adapted from inter-system HO of LTE.

- Macro cell small cell HO: mobility is the important criteria, where UE, user, service with low mobility would be handed over to the small cell, while UE; user, service with high mobility will be supported preferably under the macro cell coverage.
- Bandwidth, latency, network jitter, packet loss and network costs: depending on user context these parameters could change across macro and small cell networks and could be combined with RSS, SNR and channel availability for vertical HO
- Bandwidth, UE, user and service mobility: a HO into the target network could impact the user QoE in a positive direction
- QoE, session duration and a signalling cost function: this could be considered from a service provider and network operator point of view
- Energy efficiency including battery life of UEs and load balancing in RAN: for example the IEEE 802.21 media independent handoff function (MIHF) takes care of the MS battery life and the AP load

In the MiWEBA project, indeed, the novel cell association method was proposed. The method can utilize the millimetre wave resources efficiently by considering the user context parameters, e.g. traffic demand, SINR, and channel availability [16], [20]. Since the expecting performance gain of millimetre wave heterogeneous network can be achieved by introducing this cell association method, these additional criteria should be added in the case of vertical multi-layer HO.

Furthermore, the novel small cell selection method considering traffic load of each small cell was also studied in MiWEBA Deliverable 3.5 [15]. The effectiveness of the proposed small cell selection method is verified by both computer simulations and hardware-based experiment. The results in [15] show that the proposed small cell selection method can improve the average downlink throughput compared to traditional signal strength based small cell selection. The details can refer to [15].

4.2 HO time reduction

Firstly, lean protocols are proposed as option to reduce HO latency. HO in a multi-layer cellular network was investigated in [21], showing that the HO events in next generation network systems will occur at a much higher rate. Cell splitting schemes including cell sectoring, frequency hopping combined with discontinuous transmission and power control and reuse partitioning, FD MIMO and BF will extend these further even only in E-UTRAN. The objective must clearly be to minimise the network control and HO time.

Besides mobility issues, each HO also adds signalling load to the core network (CN) and backhaul and front haul, which is an impact on QoE and network complexity with increased number of small cells and with increased user mobility. So mobility issues are going to lead to more signalling and backhaul load which will increase further with network densification and the growing number of small cells. The signal measurement by UEs in conjunction with the large number of idle small millimetre wave cells are further major challenges implementing HO techniques.

One example, how to solve this, is the call admission control (CAC) mechanism between IEEE 802.16m APs and small cell APs, which uses only a single parameter for resource availability with the signal strength for the HO optimisation processes. Another option could be to consider additionally both the resource availability and user residence duration and small cell priority over the macro cell eNB [22]. Then there would be two options for CAC algorithms - overflow CAC scheme and class based CAC scheme, where a maximum use of multi-layer architecture can be achieved only if the CAC admits new mobile users with high bit rates and low mobility into any lower layer cell available while users with low bit rates and high mobility are given admission to the upper layer.

And, signalling overhead and latency within HO process could be reduced for example as follows. Each RAT is currently E2E coupled, which means PHY to MAC/RLC to non-access stratum (NAS) and UE to RAN to CN, where an inter-RAT HO impacts UE as well as RAN and CN. Thus, this architecture is not optimal supporting a multi-RAT network. An interface proposed in [23] unifies the interface between single radio controller (SRC) and CN, where UE and CN use a unified NAS. The SRC manages multi-RAT radio resources and so the CN is not involved in HO procedure reducing signalling latency.

The HO between LTE macro cells and millimetre wave small cells allows UEs with poor millimetre wave coverage to switch to LTE with better coverage and UEs in LTE can change to millimetre wave RAT to get high throughput if there is coverage. This handover will require rather extensive RRC and CN signalling as well as cell discovery and synchronization, which results in relatively long handover timing. With the increased amount of LTE and millimetre wave frequency bands and the massive introduction of beamforming, it is expected that the cell discovery time and handover time thereof could increase drastically to up to 300 milliseconds without countermeasures.

However, with a U-plane aggregation/dual connectivity of LTE macro cells and millimetre wave small cells using a common C-plane connected to both the LTE eNB and the millimetre wave AP, no signalling or only few signalling would be required when only a U-plane HO needs to be performed. Furthermore, the U-plane HO might be almost instantaneous. Therefore, the handover time/latency would be reduced significantly. Another advantage with a C-plane in dual/multi-connectivity is the increased reliability which needs to be traded versus multiple flows of the C-plane and its overhead. Such a fast switched HO could lead to nearly 0 milliseconds service interruption delay. This C/U plane splitting mechanism for LTE macro cells and millimetre wave small cells has been fully discussed and studied in MiWEBA WP3 Task 3.1. The research results are included in MiWEBA Deliverable 3.1 [13] and Deliverable 3.2 [14], where, the overall architecture and proposed protocol stacks are introduced. The effectiveness of the LTE/millimetre-wave aggregation architecture is verified by both computer simulations and hardware-based experiment. The results in [14] show that only U-plane handover is necessary and the handover time is very small, i.e., less than 50ms. The average downlink throughput of UEs in cellular millimetre wave heterogeneous networks with directional beamforming antennas is 5 times higher compared with traditional CN level interworking solutions. And the maximum instantaneous downlink data rate can be more than 10Gbps even when UE speed is 20km/h. The details can refer to [14].

4.3 HO rate decrease

One option is to predict mobility. A number of schemes have been investigated since years to predict user movement using it for radio resource and mobility management [24]. So setting up and reserving resources along a user's mobile path, and planning quick handovers between the base stations is proposed to minimise the occurrence of location registration and update procedures. Nevertheless, the prediction remains reasonably accurate despite the influence of random movements. The performance of prediction algorithms based on location (based on user's current location), the direction (based on current user's direction), the segment (based on pattern matching), Bayes' rule (based on conditional probability of future direction), and the time (based on direction with additional time constraint). Criteria show a prediction accuracy rate between 10 and 70 % due to challenging characteristics of user mobility models and the not perfect decision making mechanisms of HO in cellular systems. Although the use of movement prediction seems to be a promising approach for improving the efficiency, reliability and adaption of wireless networks, the actual user mobility patterns are not yet well understood and are up for further studies in particular for heterogeneous networks.

Another option is a Markov Decision Process (MDP), which could be used effectively for HO decisions, where vertical HO decisions can be evaluated via a HO cost function and a HO threshold function, which can be adapted to changes in the network environment dynamically.

Furthermore, the small cell deployment with constrained coverage poses challenges on the coordination of radio resource utilizations and mobility management including HO across multiple eNBs, in particular for non-ideal backhaul cases as shown already in [25] for LTE. The macro cell assistant mobility management implementing millimetre wave small cells with a new architecture of the C-plane at LTE macro cells and U-plane at small cells would be also beneficial to avoid high HO failure rate and bad QoE at the edge of small cells and between small cell and macro cell. An analysis in [26] shows, that in heterogeneous networks macro cells and small cells provide essential coverage, whereas the large macro cells are advantageous in supporting high-mobility UE for reduced HO frequency. The above-mentioned LTE/ millimetre-wave aggregation architecture using C/U splitting mechanism can solve this problem well. This technology has been studied well in MiWEBA WP3 Task 3.1. The details are included and explained in [13] and [14]. We defined that one-time HO consists of C-plane HO and U-plane HO. In Sect. 4 of [13], the simulation results also show that when more than 10 small cells are deployed in a macro cell, the number of HOs can be reduced to about 1/2 by introducing the C/U-plane splitting scheme. In Sect. 2 of [14], from the simulation results, we know that the handover rate of the aggregation architecture has a 50% decrease compared to traditional CN level interworking solutions, i.e., only U-plane HO exists. This also decreases the handover delay since C-plane HO causes the main latency. Without frequent C-plane HO, the mobility management of UEs becomes easier and stable communication can be realized.

Finally, Carrier Aggregation (CA) is an extension to multi-stream aggregation (MSA) and CoMP. These kinds of operations applied to higher layers would allow UEs to receive traffic streams from multiple transmission nodes in particular in case of non-ideal backhaul and front haul. MSA could also ease the routing UE traffic to different nodes based on QoE requirements. MSA could possibly be applied to millimetre wave small cell heterogeneous networks with intra-frequency or inter-frequency HO.

5 Conclusions

Current handover techniques of IP based data traffic in 3GPP and IEEE networks support a very small subset of vertical handover issues in multi-layer heterogeneous networks. Looking at all handover phases including network discovery and selection, network attachment, configuration, security association, binding update and media re-routing from a heterogeneous network built with LTE macro cells and millimetre wave small cells point of view, it becomes clear that a lot of improvements and extensions are required to adopt these available handover techniques to the characteristics of millimetre wave small cells in a multi-layer environment.

Millimetre wave small cell heterogeneous networks bring a lot of challenges at PHY, MAC and routing layers for handover requesting that the communication architecture and protocols, especially at the layer 1 and 2, need to be revised to adapt signalling and resource allocation and cope with a new level of dynamics in channel attenuation, directionality, and blockage. For example, the gains of beamforming have to be paid for by far more complicated mobility management and handover techniques. A lot more frequent handover, even for static UEs, could become a potential drawback of millimetre wave small cell systems due to their vulnerability to blockage. The dense deployments of millimetre wave small cell APs intensify frequent handovers between adjacent APs even more. The loss of precise beamforming data will become a further criterion for handover and re-association. To avoid drastically increased handover times, rates and severely decreased handover efficiency some first mitigation options are proposed to reduce the overhead and delay of handover and re-association.

An outlook for many future studies is given to address the optimization and extensions of current 3GPP and IEEE handover solutions. Some of the issues have been well studied in WP 3 and WP4 of MiWEBA project.

6 References

[1] E. Hossain, Heterogeneous Wireless Access Networks - Architectures and Protocols, New York: Springer, 2008.

[2] H. S. Ashutosh Dutta, Mobility protocols and handover optimization : design, evaluation and application, IEEE Press, 2014.

[3] 3GPP TS 36.300 v13.2.0, “TS Group Radio Access Network E-UTRA and E-UTRAN overall description stage 2,” 3GPP, 2016.

[4] 3GPP TS 23.402 v13.5.0, “Architecture enhancements for non-3GPP accesses,” 3GPP, 2016.

[5] R. Kim, I. Jung, X. Yang and C. C. Chou, “Advanced handover schemes in IMT-advanced systems [WiMAX/LTE Update],” *IEEE Communications Magazine ,* vol. 48, no. 8, pp. 78 - 85, 2010.

[6] 3GPP TR 37.834 v12.0.0, “Study on Wireless Local Area Network (WLAN) - 3GPP radio interworking,” 3GPP, 2014.

[7] J. D. Gibson, Mobile Communications Handbook, CRC Press., 2013.

[8] 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture," 3GPP, 2015.

[9] ITU-T G.114, “Series G: transmission systems and media, digital systems and networks,” ITU-T, 2003.

[10] J. P. C. R. N. G. A. Mohammadreza Behjati, "Self-organising comprehensive handover strategy for multi-tier LTE-Advanced heterogeneous networks," *IET Science, Measurement and Technology,* vol. 8, no. 6, p. 441–451, 2014.

[11] IETF RFC 3753, “Mobility Related Terminology,” IETF, 2004.

[12] A. A. G. B. A. Guillaume de la Roche, LTE – advanced and next generation wireless networks : channel modelling and propagation, Wiley, 2013.

[13] ICT FP7 MiWEBA project No. 608637, Deliverable D3.1, "Design of mm-Wave Access Links and Integration into Cellular Network Infrastructure," Nov. 2014.

[14] ICT FP7 MiWEBA project No. 608637, Deliverable D3.2, "Extension of control plane," Editor: S. Nanba [KDDI] , Feb. 2016.

[15] ICT FP7 MiWEBA project No. 608637, Deliverable D3.5, "Enhancement of mobility robustness", Feb. 2016.

[16] ICT FP7 MiWEBA project No. 608637, Deliverable D4.3, "Dynamic resource

optimization algorithm and control protocols," May 2015.

[17] M. C. A. G. Anup Talukdar, “Handoff Rates for Millimeterwave 5G Systems,” IEEE VTC Spring, Seoul, Korea, 2014.

[18] 3GPP TR 36.839, "Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks," 3GPP, 2013.

[19] C.-K. Han and H.-K. Choi, "Security Analysis of Handover Key Management in 4G LTE/SAE Networks," *IEEE Transactions on Mobile Computing,* vol. 13, no. 2, pp. 457 - 468 , 2013.

[20] K. Sakaguchi, G. K. Tran, H. Shimodaira and etc., “Millimeter-wave Evolution for 5G Cellular Networks,” IEICE Trans. Commun., vol. E98-B, no. 3, pp. 388-402, Mar. 2015.

[21] D. W. A. A. Mahboubeh Lohi, “Trends in multi-layer cellular system design and handover design,” *IEEE ,* pp. 898 - 902, 1999.

[22] 3GPP TR 36.842, “Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects,” 3GPP, 2014.

[23] WWRF WGC, “Communication architectures and technologies,” WWRF, 2013.

[24] J. Chan and A. Seneviratne, “A practical user mobility prediction algorithm for supporting adaptive QoS in wireless networks,” in *IEEE International Conference on Networks*, 1999.

[25] Huawei, "The second phase of LTE-Advanced," Huawei, 2013.

[26] S. P. Yeh, S. Talwar, G. Wu, N. Himayat and K. Johnsson, “Capacity and coverage enhancement in heterogeneous networks,” *IEEE Wireless Communications ,* vol. 18, no. 3, pp. 32 - 38, 2011.

[27] C. F. G. F. ´. P. P. Z. Hossein Shokri-Ghadikolaei, “Millimeter Wave Cellular Networks: A MAC Layer Perspective,” *IEEE Transactions on Communications,* vol. 63, no. 10, pp. 3437 - 3458, 2015.

[28] M. I. K. M. G. Mahardhika, “Multi-Criteria Vertical Handover Decision in Heterogeneous Network,” in *IEEE Symposium on Wireless Technology and Applications*, Bandung, Indonesia, 2012.