Universidade de Aveiro
Departamento de Electrónica, Telecomunicações e Informática
2014

Tiago Silvestre Condeixa

Descentralização da Gestão da Mobilidade IP nas Redes do Futuro

Decentralizing IP Mobility Management in Future Networks

Tiago Silvestre Condeixa

Descentralização da Gestão da Mobilidade IP nas Redes do Futuro

Decentralizing IP Mobility Management in Future Networks

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Electrotécnica, realizada sobre a orientação científica da Professora Doutora Susana Isabel Barreto de Miranda Sargento, Professora Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

O trabalho desenvolvido no decorrer desta Tese foi financiado pela Bolsa de Doutoramento SFRH/BD/65265/2009, atribuída pela Fundação para a Ciência e Tecnologia (FCT).

o júri / the jury

presidente / president

Professor Doutor António Carlos Mendes de Sousa
Professor Catedrático da Universidade de Aveiro (por delegação do Reitor da Universidade de Aveiro)

vogais / examiners committee

Professora Doutora Susana Isabel Barreto de Miranda Sargento
Professora Auxiliar no Departamente de Electrónica, Telecomunicações e Informática da Universidade de Aveiro (orientador)

Professor Doutor Fernando Pedro Lopes Boavida Fernandes
Professor Catedrático no Departamento de Engenharia Informática da Universidade de Coimbra

Professor Doutor Manuel Alberto Pereira Ricardo
Professor Associado no Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

Professor Doutor Pedro Nuno Miranda de Sousa
Professor Auxiliar no Departamento de Informática da Escola de Engenharia da Universidade do Minho

Professor Doutor António Manuel Duarte Nogueira
Professor Auxiliar no Departamente de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

agradecimentos / acknowledgments

À professora Susana Sargento por toda a sua disponibilidade, apoio e motivação que me permitiu desenvolver o presente trabalho com rigor e profissionalismo e que me ajudou a atingir os objetivos a que me propus.

Aos colegas e amigos do grupo Network Architectures and Protocols (NAP) do Instituto de Telecomunicações por todo o apoio e interesse demonstrado, tendo contribuído amplamente para o meu crescimento pessoal e profissional. Um especial agradecimento ao Ricardo Matos, André Cardote, Nuno Coutinho, Lucas Guardalben e João Soares pela camaradagem e apoio a nível pessoal e profissional. Também um agradecimento especial ao Carlos Frade e Jonathan Carvalho pelo trabalho conjunto que me permitiu obter uma tese mais completa, tornando-a mais coerente e consistente.

À toutes les personnes de Telecom Bretagne et Orange Labs de Rennes qui m'ont soutenu dans mon séjour à Rennes. Un merci tout spécial à Hassan Ali-Ahmad, Pierrick Seite, Philippe Bertin et Xavier Lagrange qui m'ont aidé à grandir professionnellement.

Also to Danny Moses, Alper Yeign and Hassnaa Moustafa for our discussions and collaborative work in the scope of the Distributed Mobility Management (DMM) working group.

À Universidade de Aveiro, ao Departamento de Eletrónica, Telecomunicações e Informática e ao Instituto de Telecomunicações por me terem fornecido todas as condições e apoio necessários ao desenvolvimento desta tese.

À Fundação para a Ciência e a Tecnologia pelo seu contributo para a progressão da ciência em Portugal e pelo financiamento disponibilizado que me permitiu concretizar este trabalho.

À minha família e namorada que sempre me apoiaram ao longo de todo o processo, permitindo-me desenvolver o meu trabalho da melhor forma possível.

Finalmente, aos meus amigos e a todos aqueles que de alguma forma contribuíram para o meu crescimento pessoal e/ou profissional.

Palavras-chave

Gestão de Mobilidade Descentralizada, Mobilidade IP Distribuída, Mobilidade Dinâmica, Ancoramento Dinâmico, Ancoramento baseado em Contexto, Contexto da Mobilidade Distribuído, Gestão do Endereçamento IP Distribuído.

Resumo

Na última década verificou-se uma massificação dos dispositivos móveis e das suas aplicações, o que tem vindo a aumentar o consumo de dados móveis. Este aumento dificulta o planeamento e dimensionamento das redes devido principalmente aos modelos extremamente centralizados adoptados por estas. Os operadores móveis têm vindo a estudar modelos mais *flat* para as redes, os quais distribuem a responsabilidade de fornecer conectividade e mobilidade, no sentido de melhorar a escalabilidade e desempenho da rede. Além disso, de forma a garantir um desempenho elevado na entrega dos conteúdos, os fornecedores de serviço têm vindo a mover os servidores de conteúdos para locais mais próximos do utilizador. Apesar do esforço na procura de soluções para o crescente consumo de dados móveis, os modelos atuais de gestão de mobilidade são demasiado centralizados para conseguir assegurar a continuidade de sessão aos utilizadores conectados à rede. As arquiteturas implementadas têm um número muito reduzido de âncoras móveis centralizadas que gerem todos os dados móveis e a informação de contexto da mobilidade, o que leva a uma diminuição de desempenho e escalabilidade, solucionadas através de mecanismos de rede dispendiosos.

A gestão da mobilidade precisa de ser repensada de forma a poder lidar com arquiteturas de rede *flat* e com a distribuição dos servidores de conteúdos para nós mais próximos dos utilizadores, que é o objectivo principal da Tese apresentada. Primeiro, é apresentada a caracterização da gestão de mobilidade em blocos funcionais, a interação entre eles e potenciais agrupamentos dos mesmos. A gestão da mobilidade descentralizada é estudada através de modelos analíticos e simulações, em que diferentes abordagens distribuem as funcionalidades da mobilidade pela rede. Como resultado deste estudo verificou-se que a descentralização da mobilidade traz vantagens claras. Com base nestes resultados foi proposta uma nova abordagem de gestão de mobilidade distribuída e dinâmica, que é exaustivamente avaliada através de modelos analíticos, simulações e experiências numa bancada de testes. A abordagem proposta é também integrada com mecanismos de *handovers* horizontais transparentes, assim como é avaliada em ambientes veiculares. Os mecanismos de mobilidade da abordagem proposta são também especificados para cenários de *multihoming*, de forma a proporcionar o *offloading* de dados com suporte de mobilidade das redes celulares para outras redes de acesso. Com o objectivo de optimizar o encaminhamento de dados móveis, foi criada uma nova estratégia para o suporte da mobilidade localizada, em que um sistema de replicação de *bindings* é integrado nas âncoras de mobilidade distribuídas através dos *routers* de acesso e dos *gateways*. Finalmente apresenta-se um modelo de ancoramento adaptativo para a mobilidade com base em contexto, o qual dinamicamente determina as âncoras de mobilidade que oferecem a melhor rota para uma dada sessão, baseado na informação do utilizador e da rede.

A integração de conceitos de dinamismo e de distribuição na gestão da mobilidade, como o ancoramento adaptativo e o suporte dinâmico da mobilidade, permitem a optimização dos recursos da rede e uma melhor experiência por parte do utilizador. Os resultados demonstram, de uma forma geral, que a gestão descentralizada da mobilidade é um caminho promissor, logo este deve ser tomado em consideração pelas operadoras móveis aquando do desenvolvimento das redes do futuro.

Keywords Decentralized Mobility Management, Distributed IP Mobility, Dynamic Mobility, Dynamic Anchoring, Context-Aware Anchoring, Distributed Mobility Context, Distributed IP Address Management.

Abstract The massive adoption of sophisticated mobile devices and applications led to the increase of mobile data in the last decade, which it is expected to continue. This increase of mobile data negatively impacts the network planning and dimension, since core networks are heavy centralized. Mobile operators are investigating flatten network architectures that distribute the responsibility of providing connectivity and mobility, in order to improve the network scalability and performance. Moreover, service providers are moving the content servers closer to the user, in order to ensure high availability and performance of content delivery. Besides the efforts to overcome the explosion of mobile data, current mobility management models are heavy centralized to ensure reachability and session continuity to the users connected to the network. Nowadays, deployed architectures have a small number of centralized mobility anchors managing the mobile data and the mobility context of millions of users, which introduces issues related to performance and scalability that require costly network mechanisms.

The mobility management needs to be rethought out-of-the box to cope with flatten network architectures and distributed content servers closer to the user, which is the purpose of the work developed in this Thesis. The Thesis starts with a characterization of mobility management into well-defined functional blocks, their interaction and potential grouping. The decentralized mobility management is studied through analytical models and simulations, in which different mobility approaches distinctly distribute the mobility management functionalities through the network. The outcome of this study showed that decentralized mobility management brings advantages. Hence, it was proposed a novel distributed and dynamic mobility management approach, which is exhaustively evaluated through analytical models, simulations and testbed experiments. The proposed approach is also integrated with seamless horizontal handover mechanisms, as well as evaluated in vehicular environments. The mobility mechanisms are also specified for multihomed scenarios, in order to provide data offloading with IP mobility from cellular to other access networks. In the pursuing of the optimized mobile routing path, a novel network-based strategy for localized mobility is addressed, in which a replication binding system is deployed in the mobility anchors distributed through the access routers and gateways. Finally, we go further in the mobility anchoring subject, presenting a context-aware adaptive IP mobility anchoring model that dynamically assigns the mobility anchors that provide the optimized routing path to a session, based on the user and network context.

The integration of dynamic and distributed concepts in the mobility management, such as context-aware adaptive mobility anchoring and dynamic mobility support, allow the optimization of network resources and the improvement of user experience. The overall outcome demonstrates that decentralized mobility management is a promising direction, hence, its ideas should be taken into account by mobile operators in the deployment of future networks.

Contents

List of Figures

List of Tables

Abbreviations

2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
AAA	Authorization Authentication Accounting
ADSL	Asymmetric Digital Subscriber Line
AHA	Access Home Agent
AMA	Access Mobility Anchor
AN	Access Node
AP	Access Point
AR	Access Router
ASP	Application Service Provider
AuC	Authentication Center
BA	Binding Acknowledge
BS	Base Station
BU	Binding Update
CN	Correspondent Node
CoA	Care-of Address
CP	Costumer Premisses
CPN	Costumer Premisses Network
DAD	Duplicate Address Detection
DDNS	Dynamic Domain Name Server
DMA	Distributed Mobility Anchoring
DMAR	Data Mobility Access Router
DMIPA	Dynamic Mobile IP Anchoring

DMM	Distributed Mobility Management
DNS	Domain Name Server
DTN	Delay Tolerant Network
EN	Edge Node
eNB	evolved Node B
EPC	Evolved Packet Core
ER	Edge Router
FA	Foreign Agent
FBA	Fast Binding Acknowledge
FBU	Fast Binding Update
FMIP	Fast handovers for Mobile IP
FMIPv6	Fast handovers for Mobile IPv6
FN	Foreign Network
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile communication
GTP	Generic Tunneling Protocol
GW	Gateway
HA-C	Home Agent Control
HA-D	Home Agent Data
HA	Home Agent
HAck	Handover Acknowledge
HI	Handover Initiate
HIP	Host Identity Protocol
HIT	Host Identity Tag
HLR	Home Location Register
HMIP	Hierarchical Mobile Internet Protocol
HMIPv6	Hierarchical Mobile Internet Protocol version 6
HN	Home Network
HNP	Home Network Prefix

HoA	Home Address
HSS	Home Subscriber Server
ID	Identification
IMS	IP Multimedia Subsystem
INFO	IP flow mobility and seamless WLAN offload
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISP	Internet Service Provider
LAI	Location Area Identity
LAN	Legacy Access Node
LAR	Legacy Access Router
LCoA	on-Link Care-of Address
LIPA	Local IP Access
LLA	Link Layer Address
LMA	Local Mobility Anchor
LTE	Long Term Evolution
MSCTP	Mobile Stream Control Transport Protocol
MA	Mobility Anchor
MAC	Media Access Control
MAG	Mobile Access Gateway
MAN	Metropolitan Area Network
MAP	Mobility Anchor Point
MAP_C	Mobility Anchor Point Control
MAP_D	Mobility Anchor Point Data
MAR	Mobility-enabled Access Router
MICS	Media Independent Command Service
MIES	Media Independent Event Service
MIH	Media Independent Handover
MIHF	Media Independent Handover Function

MIIS	Media Independent Information Service
MIP	Mobile Internet Protocol
MIPv4	Mobile Internet Protocol version 4
MIPv6	Mobile Internet Protocol version 6
MLMA	Multiple Local Mobility Anchors
MME	Mobility Management Entity
MN	Mobile Node
MNet	Mobile Network
MR	Mobile Router
MSCTP	Mobile Stream Control Transport Protocol
NAI	Network Access Identifier
NAP	Network Access Provider
NAR	New Access Router
NAT	Network Address Translation
NCoA	New Care-of Address
NEMO	Network Mobility
NGN	New Generation Networks
NMAG	New Mobile Access Gateway
NS2	Network Simulator 2
NS3	Network Simulator 3
OSI	Open Systems Interconnection
PAN	Previous Access Network
PAR	Previous Access Router
PBA	Proxy Binding Acknowledge
PBU	Proxy Binding Update
PCoA	Previous Care-of Address
PDN	Packet Data Network
PDP	Packet Data Protocol
PGW	Packet Data Network Gateway
PLMN	Public Land Mobile Network

PMAG	Previous Mobile Access Gateway
PMIP	Proxy Mobile Internet Protocol
PrRtAdv	Proxy Router Advertisement
QoE	Quality of Experience
QoS	Quality of Service
RCoA	Regional Care-of Address
RG	Residential Gateway
RO	Route Optimization
RTP	Real Time Protocol
RtSolPr	Router Solicitation for Proxy Advertisement
SCTP	Stream Control Transport Protocol
SGSN	Serving GPRS Support Node
SGW	Service Gateway
SIP	Session Initiation Protocol
SIPTO	Selected IP Traffic Offload
SMR	Session to Mobility Ratio
TA	Tracking Area
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UNA	Unsolicited Neighbor Advertisement
UPN	User-provided Network
URL	Uniform Resource Locator
VPN	Virtual Private Network
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WMN	Wireless Mesh Network

Chapter 1

Introduction

From recent studies [1], mobile coverage is now taking the fourth place, closely followed by water distribution, both necessary utilities that have to be in place for a good life. Nowadays, having high-quality mobile coverage and internet access are as essential as having clean water and electricity. Thus, it is vital to have satisfied people with their internet connections to guarantee that they are meeting their expectations.

The Net Promoter Score evaluated by Ericsson reveals that network performance is the principal driver of subscriber loyalty to mobile operators, followed by value for money [2]. These two parameters are correlated, since the enhancement of the network quality also leads to a better perceived value for money at a given price level. The outcome from the study demonstrates that addressing network performance has the double of the impact on customer loyalty as customer support, tariff plans offered and account management, as well as four times the impact of loyalty rewards. Network performance should be one of the main concerns of mobile operators, since it is a key factor differentiating promoters from detractors.

This chapter highlights a set of important and innovative Internet evolution trends, introducing the challenges of today's communication environments and the research context that motivates this Thesis. Furthermore, we point out the main objectives pursued in this Thesis, in order to address the identified challenges. This chapter also presents the contributions of the work developed in this Thesis, as well as the resulting scientific publications. Finally, it provides an overview of the Thesis structure.

1.1 Motivation

Mobile data traffic has been experiencing an exponential growth over the last years [3, 2], which doubled between Q1 2012 and Q2 2013, an it is expected to continue in the next years with a growth of 12 times between 2012 and 2018 [2]. Video is now taking the largest segment of data traffic in mobile networks with the popularization of demanding video services, such as Netflix, Vimeo and Youtube. Video is expected to grow by around 60 % annually up until the end of 2018 according to Ericsson [2], which is predicting that video will account for around half of global mobile data traffic by 2018, while Cisco [3] is expecting over 66 % of global mobile data traffic by 2017.

This increasing mobile data traffic is mostly justified by the proliferation of mobile devices, such as smartphones and tablets, which offer the consumer content and applications not supported by previous generations of mobile devices. Telecommunications market receives several new devices with increased capabilities every year, which results in more than 5 billion wirelessly connected mobile devices in service today [3], where most of them

are handheld devices or mobile broadband devices, such as portable computers, tablets and smartphones. It is expected that there will be 8.6 billion handheld or personal mobile devices by 2017. In Q1 2013, the total mobile subscriptions exceeded 6.4 billion, which is expected to reach 9.1 billion by the end of 2018. Regarding the total smartphone subscriptions, it reached 1.2 billion at the end of 2012 and it is expected to grow to 4.5 billion in 2018. In a near future, it is expected that this human-centric connected devices are outweigh from 10 to 100 times by communicating machines, such as surveillance cameras, smart-city and connected sensors. This trend of machine to machine communications introduces a challenging transition from 5 to 50 or perhaps even 500 billion connected devices. These mobile devices are using applications that go beyond traditional mobile data traffic to more bandwidth-demanding and delay-sensitive applications, such as video. Over time, users tend to use more advanced services that put greater demands on device capabilities. Today, smartphone users who subscribe to both music and video streaming services already consume more than 2 GB of data traffic per month on average, which is 4 times the consumption of an average smartphone user. The proportion of mobile users generating more than 2 gigabytes per month has increased significantly over 2012, reaching 18 % of users towards the end of 2012, while users generating more than 5 gigabytes per month reached 1 % in the end of 2012.

The 5G system for 2020 and beyond will meet the long-term vision of unlimited access to information and sharing of data available anywhere and anytime to anyone and anything, according to Ericsson [4]. To cope with this wide vision, it is necessary to address a much wider variety of devices, services and challenges than those accommodated by today's mobile broadband systems. The 5G system will not be a single technology but rather a combination of integrated radio access technologies, including evolved versions of LTE and HSPA, as well as specialized radio access technologies for specific use cases, which will jointly fulfill the requirements in the future.

This growing of mobile devices and mobile data traffic has a serious impact on the dimensioning and planning of mobile networks, which brings two main issues to the mobile network operators in access and core networks, respectively:

Shortage Radio Spectrum: current access network bandwidth cannot be easily increased, since radio spectrum is limited and expensive.

Heavy Hierarchical/Centralized Networks: nowadays mobile core networks are deployed with heavy hierarchical and centralized models, which brings scalability and performance issues.

Mobile operators are addressing the shortage radio spectrum with two complementary strategies: by deploying more spectrum efficient technologies, such as the 3rd Generation Partnership Project (3GPP) and Long Term Evolution (LTE); and by providing strategies to selectively offloading traffic from the cellular access to alternative wireless technologies, such as WiFi or femtocells.

Globally, 33 % of total mobile data traffic was offloaded to the fixed network through Wi-Fi or femtocells in 2012. For users with fixed broadband and Wi-Fi access points at home, or for users served by operator-owned femtocells and picocells, a considerable proportion of data traffic generated by mobile and portable devices is offloaded from the mobile network to the fixed network. In 2017, it is expected an amount of 46 % of offloaded traffic from smartphones and 71 % from tablets.

Mobile operators are addressing the heavy hierarchical/centralized networks by investigating new network architectural models that distribute the responsibility of providing

connectivity and mobility. To solve this issue, it is necessary to look for strategies to alleviate the core network scalability issues in terms of number of users, managed/maintained information, and data traffic load. On the one hand, some short-term solutions are being developed as evolution of currently deployed mobile network architectures to provide relief to current data traffic problems. On the other hand, long-term alternatives capable of coping with the future expected data traffic loads, involving a major redesign of the network architecture, are being researched. The flatten IP networks are considered as the crucial direction in the evolution of the mobile network operator model. The hierarchical topology model currently deployed by mobile operators that has served relatively well the users and operators is being replaced by novel network designs, such as HSPA Evolution, and especially LTE. This flatten network models designed by 3rd Generation Partnership Project (3GPP), are able to provide a much lower cost, a much more flexible network and consequently a superior data traffic delivery performance. The LTE is the pioneer on the evolution of flatten network architectures, where the voice and data cores converge into a single Evolved Packet Core (EPC). It is necessary to continue looking beyond this high-level architectural flattening to examine the case for taking both 3GPP-defined and non-3GPP-defined elements and capabilities that have traditionally resided in the center of the mobile network, and distributing those capabilities closer to the user at the edge of the network, in the pursuit of lower cost and better performance.

There are a considerable number of IP mobility protocols proposed in the last decade, which lead to standardized IP mobility solutions, such as Mobile Internet Protocol version 6 (MIPv6) [5], Hierarchical Mobile Internet Protocol version 6 (HMIPv6) [6] and Proxy Mobile Internet Protocol (PMIP) [7]. These protocols were developed with different functions and messages, but they share three key common features among them.

Centralized Mobility Anchor

The presence of the centralized mobility anchor allows a mobile node to remain reachable after it has moved to a different network. The mobility anchor, among other tasks, also ensures session continuity by routing/forwarding packets destined to, or sent from, the mobile node. In practice, most of the deployed architectures today have a small number of centralized anchors managing the data traffic and the mobility context of millions of mobile nodes. Mobility context is the collection of information required to provide mobility management support for a given mobile node. Compared with a distributed approach, a centralized approach is likely to have several issues or limitations affecting performance and scalability, which require costly network mechanisms to resolve.

Extensions to Optimize Handover Performance

To optimize handovers from the perspective of mobile nodes, the base protocols have been extended to efficiently handle packet forwarding between the previous and new points of attachment. These extensions are necessary when applications have stringent requirements in terms of delay and packet loss. Notions of localization and distribution of local agents have been introduced to reduce signaling overhead at the centralized routing anchor point. Unfortunately, today we witness difficulties in getting such protocols deployed, resulting in sub-optimal choices for the network operators.

Extensions to Enable Multihomed Devices.

Moreover, the availability of multiple-interface devices and the possibility of using several network interfaces simultaneously have motivated the development of even more protocol extensions to add more capabilities to the mobility management protocol. In the

end, deployment is further complicated with the multitude of extensions.

Cellular networks have been hierarchical and content servers have been centralized, so mobility management has been deployed according to a centralized model. Mobility solutions deployed with centralized mobility anchoring in novel flatten mobile networks with distributed content servers are more prone to problems or limitations regarding performance and scalability when compared with distributed and dynamic mobility management [8, 9, 10].

Performance Issues and Limitations

Current mobility models adopt a centralized mobility anchoring, where the routing/forwarding is performed via the centralized Mobility Anchor (MA), which is usually distant from both the content server and the MN. Routing via a centralized anchor is often longer, so that those mobility protocol deployments that lack optimization extensions result in non-optimal routes, affecting performance; whereas routing optimization may be an integral part of a distributed design. As a mobile network becomes less hierarchical, centralized mobility management can become more non-optimal, especially as the content servers in a content delivery network (CDN) are moving closer to the access network. Furthermore, the recent trend in network flattening, with connectivity sharing among users in the same geographical area and direct communications among them, reinforce centralized architectures weaknesses. In contrast, distributed mobility management can support both hierarchical and flat networks as may be needed to support CDNs. Signaling exchanged with closer MAs is able to reduce the handover delay and the associated packet loss.

Mobility patterns indicate that mobile nodes often remain attached to the same point of attachment for considerable periods of time. In general, people in big cities spend half of their time at home, one quarter of it at work or school and the rest in any other location [1]. Specific IP mobility management support is not required for applications that launch and complete their sessions while the mobile node is connected to the same point of attachment. Currently, IP mobility support is designed for always-on operation, maintaining all parameters of the context for each mobile node for as long as they are connected to the network. This can result in a waste of resources and unnecessary costs for the service providers. Infrequent node mobility coupled with application intelligence suggest that mobility support could be provided selectively, thus reducing the amount of context maintained in the network and improving the data delivery performance.

Scalability Issues and Limitations

Centralized route maintenance and context maintenance for a large number of mobile nodes is more difficult to scale. Scalability may worsen if there is no mechanism to determine whether mobility support is needed; dynamic mobility management (i.e., selectively providing mobility support) may be better implemented with distributed mobility management. Excessive signaling overhead should be avoided when end nodes are able to communicate end-to-end, with the capability to selectively turn off signaling not needed by the end hosts. Deployment is complicated with numerous variants and extensions of mobile IP; these variants and extensions may be better integrated in a distributed and dynamic design which can selectively adapt to the needs. Centralized approaches are generally more vulnerable to a single point of failure and attack, often requiring duplication and backups. A distributed approach typically isolates the problem in a single local network, so that the needed protection can be simpler.

IETF recently charted the Distributed Mobility Management (DMM) working group to figure out to distribute the mobility management closer to the user, in order to reduce the network cost and improve the delivery performance. DMM specifies IP mobility, access network and routing solutions, which allow for setting up IP networks so that traffic is distributed in an optimal way and does not rely on centrally deployed anchors to manage IP mobility sessions. The DMM solutions aim for transparency above the IP layer, including maintenance of active transport level sessions as mobile nodes or entire mobile networks change their point of attachment to the Internet. DMM is an alternative to the above centralized deployment. The motivation behind the interests to study DMM are primarily related with the novel distributed trend for networks and content/services that derives from the change of behavior from users and their consumed content. The DMM addresses two complementary aspects of mobility management procedures: the distribution of mobility anchors in the data-plane towards a more flat network and the selective activation/deactivation of mobility protocol support as an enabler to distributed mobility management. The former aims at positioning mobility anchors closer to the user; ideally, mobility agents could be collocated with the first-hop router. The latter, facilitated by the distribution of mobility anchors, identifies when mobility support must be activated and when sessions do not require mobility management support thus reducing the amount of state information that must be maintained in various mobility agents of the mobile network. It can then avoid the unnecessary establishment of mechanisms to forward traffic from an old to a new mobility anchor. Besides of occasional research works on the topic of distributed mobility management, the DMM working group was effectively launched in the beginning of 2012, which means nearly two years latter then the start of this Thesis. Therefore, the motivation of this Thesis was the same as the motivation of the DMM, with several parallel approaches in between this Thesis and DMM, as well as contributions to DMM during this Thesis timeline.

1.2 Objectives

The following research objectives address specific problems from the outlined issues of the current mobility management model, in order to take one more step in the direction of the main purpose of investigating the mobility management approaches in future network.

Research Objective 1: How decentralized should be the mobility management regarding the mobility functionalities to improve the network performance and optimize the network resources?

Current IP mobility management is based on a heavy-centralized model, where all the mobility functionalities are provided by a centralized node, usually called mobility anchor. This centralized model may encounter scalability issues (e.g. network bottlenecks, and single point of failures), security issues (e.g. attacks focused on the centralized mobility anchor), and performance issues (e.g. non-optimized routing). The decentralization of the IP mobility management should be studied and analyzed, in order to understand the mobility functionalities that need to be distributed to improve the network performance and to optimize the network resources. It is also important to go further in the decentralization of the IP mobility management and study the impact of the degree of decentralization/distribution of the mobility functionalities, evaluating both advantages and drawbacks for network and user.

Research Objective 2: How should the mobility anchors be placed and

assigned to reduce the network cost and to improve the user experience?

The mobility anchor is the main entity of current centralized mobility models. The mobility anchor is responsible to route all the mobile data packets to the current location of the user, which brings longer traffic delays and higher network resources consumption. The distribution of mobility anchors through the network nodes should be evaluated, in order to understand the best places to deploy these mobility anchors with the goal to reduce the network cost and to improve the user experience. It is also important to investigate the best way to assign these mobility anchors to users devices, evaluating different models and degrees of personalized anchoring selection (e.g. per network, user or session).

Research Objective 3: How should the mobility context be managed to be scalable and easily-deployed in a large network?

The mobility context is currently managed in a centralized mobility anchor, such as the bindings of all users connected to the network, introducing a single point of failure and higher signaling loads. It is important to evaluate different mobility context management models, where this functionality is distributed through network elements and user devices. The advantages and drawbacks regarding scalability and performance to distribute part or entire mobility context should be analyzed.

Research Objective 4: How should the IP mobility be deployed to exploit the multihoming, allowing IP data offloading schemes with shorter routing paths?

Current centralized mobility anchor is adopted in order to provide IP mobility support to the multihomed devices of the user. The IP session continuity is achieve at the cost of routing all mobile data packets through the centralized mobility anchor, to ensure data offloading through different access networks and user device interfaces when desire. It is important to study how the IP mobility support should be provided to multihomed devices, in order to be able to offload data through different networks/interfaces with the shortest routing path and minimized tunnels, ensuring IP session continuity.

1.3 Contributions

As mentioned in the previous Section 1.2, the work developed in this Thesis addresses the IP mobility management in future networks. As the result of the developed research work, we published the main scientific achievements as summarized in Table 1.1.

As a first step, we analyzed mobility management in several user-centric scenarios [11]. We discussed the efficiency and applicability of current mobility assumptions in user-centric scenarios, discussing their requirements and solutions for several types of networks with user-centric characteristics. We identified the fundamentals of a user-centric mobility management architecture able to efficiently deal with the aforementioned scenarios. Considering that mobility management is a key aspect to consider in future Internet architectures, as these architectures include a highly nomadic end-user which often relies on services provided by multi-access networks, we will have additional requirements, due to the more dynamic behavior in the network, and also a more prominent role from the end-user. Hence, we also proposed a starting point to dismantle current mobility management notions to assist the understanding of such requirements [12]. This contribution was an initial proposal to define mobility management in concrete functional blocks, their interaction, as well as a potential grouping, which later assisted us in deriving more flexible mobility management architectures and protocols. In the scope of the mobility

management analysis, we also collaborated in the IEFT Internet Draft [8], which defines the problem statement for distributed and dynamic mobility management. The draft examines the problems of centralized mobility management, where eight problems affecting performance and scalability are described, and it also points out three requirements for the distributed and dynamic mobility management.

We studied and evaluated decentralized mobility management approaches in [13] and [14], where some of the mobility functionalities are distributed through network. The approach presented in [13] splits the centralized mobility into control and data planes. The control plane remains centralized, while data plane is distributed through the network routers. The studied performed in [14] evaluates the distribution of data management, location management and handover management through the network nodes and mobile devices. This study was extended in [15], which provides a deep study on the distribution of specific mobility functionalities, such as mobility anchoring, mobility context maintenance and IP address management. The work provided by these publications has been important to understand the advantages and drawbacks of distributing the mobility management functionalities through the network, which latter assisted us in developing novel IP mobility management approaches.

From the analysis of the mobility management functionalities and the studies of the distribution of mobility functionalities, we proposed a novel IP mobility management approach for flatten network architectures in [16], called Dynamic Mobile IP Anchoring (DMIPA). DMIPA is a distributed and dynamic IP mobility management approach, where the IP mobility functionalities are distributed through access routers and mobile nodes. The proposed IP mobility approach was extended in [17] with a more detailed description of DMIPA and an exhaustive evaluation through analytical models, simulations and testbed experiments. DMIPA was evaluated in more dynamic environments, such as vehicular scenarios in [18], and it was also integrated with semaless horizontal handovers mechanisms to reduce the handover latency and packet loss during the handover execution in [19]. DMIPA mechanisms were adapted for multihomed scenarios in [20], in order to provide data offloading with IP mobility from cellular to other access networks.

Since DMIPA might not provide the optimized network performance for highly dynamic environments, such as vehicular scenarios, we evaluated a replication strategy of the bindings through mobility anchors spread all over the operator networks, in access routers and gateways, which was compared with centralized and distributed mobility models in [21]. This replication mobility model show us that mobility anchors closer to the user is not always the best location, and forwarding the packets in a mobility anchor closer to the CN or closer to the gateway could be the proper solutions in some scenarios. This assignment of mobility anchors may optimize the routing path of data packets, reducing the network resources consumed and improving the user experience. However, a replication strategy introduces a higher signaling load and a large amount of mobility context in each mobility anchor, which can be avoided through a dynamic selection of the mobility anchor according to the context of the user, the network and the session. Hence, as the last work under the scope of the Thesis, we proposed a context-aware adaptive IP mobility approach in [22], which is able to dynamically assign the mobility anchor that provides the shortest routing path and the minimized tunnel to a session, while the mobile node moves and attaches to different access networks. Through the information about the user, the network and the ongoing sessions, the proposed approach can adapt to the current scenario, improving the performance of the IP mobility management. In the scope of the mobility anchor selection, we also collaborated in the IEFT Internet Draft [23], which presents and discusses different use-case scenarios of the initial mobility anchor selection in

Table 1.1: Scientific publications achieved from the work developed in this Thesis.

Type	**Year**	**Title**	**Venue**
Conferences and Workshops	2010	*A New Perspective on Mobility Management Scenarios and Approaches* [11]	ICST MONAMI
	2011	*A Characterization of Mobility Management in User-centric Networks* [12]	NEW2AN
	2012	*Towards a Distributed Mobility Management Approach Suitable for User-centric Environments* [13]	IEEE ICCCN Workshops
		Decoupling and distribution of mobility management [14]	IEEE GLOBECOM Workshops
	2013	*Dynamic Mobile IP Anchoring* [16]	IEEE ICC
		Make-Without-Break Horizontal IP Handovers for Distributed Mobility Management Schemes [19]	IEEE GLOBECOM Workshops
	2014	*Distributed Mobility Management in Dynamic Environments: V2I Networks* [18]	IEEE ISCC (submitted)
Journals and Letters	2013 and 2014	*Studying the Integration of Distributed and Dynamic Schemes in the Mobility Management* [15]	Elsevier Computer Networks
		Rethinking IP Mobility Management Towards a Distributed and Dynamic Scheme [17]	IEEE/ACM Transactions on Networking (submitted)
		Centralized, Distributed or Replicated IP Mobility? [21]	IEEE Communication Letters
		Dynamic Offload Anchoring with IP Mobility [20]	IEEE Transactions on Network and Service Management (submitted)
		Context-Aware Dynamic IP Mobility [22]	Elsevier Computer Networks (submitted)
IETF Drafts	2011	*Problem statement for distributed and dynamic mobility management (draft-chan-distributed-mobility-ps-05)* [8]	DMM working group
	2013	*Mobility Anchor Selection in DMM: Use-case Scenarios (draft-aliahmad-dmm-anchor-selection-01.txt)* [23]	DMM working group
Book Chapter	2014	*Mobility Estimation in the Context of Distributed Mobility Management* [24]	LNSN User-centric Networking-Future Perspectives
Patent	2013	*Sistema de Gestão da Mobilidade Dinâmica*	National Patent Pending (106.996)

DMM, were several context metrics where considered, such as the MN's mobility context, the application context, and the network context.

In the scope of the distributed mobility management, we also collaborated in a book chapter [24] that addresses the mobility estimation in the context of distributed mobility management. The chapter introduces the need to consider new paradigms to adapt mobility management solutions to future Internet architectures. It provides notions concerning distributed mobility management aspects, and hot mobility estimation can be applied to the current and future mobility management solutions.

The aforementioned work performed in the scope of this Thesis was also an important contribution in the User-centric Mobility Management (UMM) project [25]. UMM is a Portuguese project funded by Fundação para a Ciência e Tecnologia (FCT), being a joint project by SITI, University Lusófona and Instituto de Telecomunicações, University of Aveiro. Following the trend where the end-user has a particular role in controlling content, as well as connectivity based upon cooperation, the main aim of the User-centric Mobility

Management (UMM) project is to propose a new mobility management architecture, better suited to user-centric, spontaneous wireless environments of which a regular hotspot, a mesh network, or a user-provided network (personal hotspot) are examples.

We also submitted a patent in the scope of distributed and dynamic global mobility management for multihomed devices, supported by Institute of Telecommunications and University of Aveiro. The patent pending describes the mobility mechanisms regarding IP mobility management, such as the management of IP addresses, IP routes, IP tunnels and mobility anchors, to provide distributed and dynamic IP mobility in multihomed scenarios.

1.4 Outline

The remainder of the Thesis is organized as follows. Chapter 2 presents the related work on mobility management. Chapter 3 provides an overview of the main work developed in this Thesis, while Chapter 4 introduces the final conclusions and future work.

Chapter 2: presents an overview of the related work within the scope of this Thesis. This chapter begins with base definitions and concepts of mobility, according to different perspectives. Then, current mobility management, according the different OSI layers, such as link layer (e.g. 2G/3G/WiMAX), network layer (e.g. MIP/PMIP), transport layer (e.g. MSCTP) and application layer (e.g. SIP), is also addressed in this chapter. The description of current mobility approaches in network layer is more detailed, since the work developed in the scope of this Thesis is mainly related with IP mobility management. Finally, this chapter describes current related work on specific functionalities of IP mobility which are fundamental to the developed work, such as mobility anchoring and context mobility management.

Chapter 3: describes the work developed on the decentralization of IP mobility management in the scope this Thesis. The chapter starts by characterizing the mobility management in well-defined functional blocks, their interaction, as well as a potential grouping, to later assist in deriving novel mobility management architectures. The chapter also presents the studies performed on decentralized mobility, in which the distribution of the mobility functionalities through network nodes and mobile devices were evaluated through analytical models and simulations. These studies are important to measure the impact of distributing the mobility management functionalities, which help us to understand the guidelines for novel mobility management approaches. Hence, a novel IP mobility management approach for flatten network architectures is proposed, which is described and evaluated through analytical models, simulations and testbed experiments in different scenarios. The proposed approach was also evaluated in vehicular scenarios and integrated with seamless handover techniques. The proposed IP mobility mechanisms were also adapted for multihomed scenarios, in order to provide data offloading with IP mobility from cellular to other access networks. Finally, the chapter presents a context-aware adaptive IP mobility anchoring model, which dynamically assigns the mobility anchors that provides the optimized routing path to a session, based on the context information about the user, the network and the sessions.

Chapter 4: resumes the main contributions and outcomes of this Thesis. After reviewing the research objectives achieved with the developed work, we point out future research guidelines to improve the IP mobility management, as well as its feasibility in real network operator infrastructures.

Appendix: annexes a selection of the most relevant scientific publications, representing the major contributions of the work done in this Thesis.

Chapter 2

State-of-the-Art on Mobility Management

This chapter main goal is to characterize and analyze current mobility management solutions. It starts with an overview of the basic definitions and the characterizations of mobility according to different perspectives. It provides an overview of mobility management of todays' technologies according to the solutions in different layers of the OSI model, such as Second Generation (2G)/Third Generation (3G), Media Independent Handover (MIH), Mobile Internet Protocol (MIP), Host Identity Protocol (HIP), Mobile Stream Control Transport Protocol (MSCTP) and Session Initiation Protocol (SIP). This chapter details the current IP mobility solutions, since mobility at network layer is the main subject of this Thesis. It also introduces the novel mobility proposals in the scope of the distributed mobility management, as well as the mechanisms already developed for mobility anchor selection and IP address management.

2.1 Mobility Management Basics

This section defines the main notions related with mobility management in order to understand the ideas under the scope of this Thesis. This section starts to clarify basic definitions strictly related with mobility, and then it characterizes mobility management by the moving object, service quality and multihoming.

Definitions

In order to better understand the mobility management it is relevant to define its basic concepts [26, 27].

Mobility: The ability for a user or other network mobile element to communicate and access its services regardless of changes in location, network, device or technical environment.

Mobility Management: The set of functions and processes applied to guarantee mobility. These functions and processes include registration, authentication, authorization, accounting, location updating, paging and others.

Home Network: The network to which a mobile user is normally connected, or the service provider with which the mobile user is associated, and where the user's subscription information is managed.

Visited Network: The network outside a home network, also called Foreign Network (FN), that provides service to a mobile user.

Roaming: The ability for the user to maintain minimum requirements to operate in another network, different from the Home Network (HN).

Horizontal Mobility: Mobility on the same layer, referred to as the mobility within the same access technology.

Vertical Mobility: Mobility between different layers, referred to as the mobility between different access technologies.

Characterization by Moving Object

Mobility management embodies different perspectives according to the moving object: personal, terminal, service or network.

Personal Mobility: The ability of a user to reach any of its services based on a personal identifier. This identifier allows the network to bind the user to one (or several) reachability profiles, anywhere, anytime.

Terminal Mobility: The ability for a terminal/device, while in motion, to access the user services independently of the location. Device mobility ensures that packets continue to be delivered to a device as it moves through the network, and as its point of attachment to the network changes.

Service Mobility: The ability of a user to access telecommunication services independently of access type and of the terminal in use, based on a personal identifier. Moreover, it is the capability of the network to provide the acquired services according to the user's service profile.

Network Mobility: The ability of a network to change its point of attachment to the corresponding network upon the network's movement itself. In this context, a network means a group of fixed or Mobile Nodes (MNs) networked to each other.

Characterization by Service Quality

Mobility Management is also classified according to the service quality, being decoupled into Service Continuity and Nomadism.

Service Continuity: Service continuity is provided both in the form of (automatic) IP continuity and of session continuity. In other words, session continuity is kept despite the change of Access Point (AP), or User Equipment (UE). While this characteristic is crucial within the mobile environment, such is not necessarily the case within fixed network environments, given that in the majority of cases, service continuity for fixed environments implies physical reconnection. However, there are cases where session continuity may be required, e.g., between the fixed and the wireless networks. For instance, if a device holds more than one network interface, e.g., a tablet PC or a smartphone equipped with WLAN and 3G interfaces, service continuity should be possible. A specific example for this scenario is a user located in its office, and using its table PC by means of fixed access. It then travels home on a train. While in the train, he might wish to keep his previous service session.

Within the context of MIH [28], session continuity covers adaptation to the new link both on L2 and L3 (address adaptation), as well as session continuity at the application

layer. MIH does not provide a standard in terms of which mobility protocol to use to perform the handover execution, even though the discussion is currently following the direction of MIPv6.

From the 3GPP IP Multimedia Subsystem (IMS) perspective, currently SIP [29] provides some session mobility management to (multimedia) services, when coupled with MIPv6 [5]. In this case, the SIP server is used as a Home Agent (HA), and handoff notification messages are traded via regular SIP messages to the HA (register) and to the Correspondent Node (CN) (re-invite). The problem with the SIP approach is latency (it inherits all the delays of the transport and of the application layers). This aspect is not however crucial to the fixed network mobility, given that the underlying technology always requires the user to physically reconnect.

Nomadism: Nomadism is the ability of the user to change its network AP while moving. When the user moves among networks covered by the same Virtual Private Network (VPN), continuity is not a requirement of this service. It is assumed that the normal usage pattern is that users shut down their service sessions before attaching to a different AP. Behind the nomadic notion, the underlying assumption is that the user will be able to recover its service(s) profile(s) independently of location and device.

This is in fact the model for the current 3GPP roaming service, which allows a user subscribed to a specific service from a specific operator, to obtain that same service by means of another access operator. Therefore, the nomadic perspective relies on a policy-based architecture. MIPv6 [5] is therefore key to achieve nomadism, given that it allows any IP host to be reachable independently of its access network. In this case, a subscriber (and not a device) is assigned to a personal identifier, which would be the key to access the subscriber's profile(s). Whenever the subscriber moves to a new location, a reconnection is triggered and the subscribed set of services can be accessed independently of the access media. However, additionally to the global addressing requirement of nomadism, there is the need to rely on a global Authorization Authentication Accounting (AAA) architecture, capable of providing the service profiles with the lesser disruption and higher security.

Multihoming

Multihoming was firstly introduces as the practice of having a network connected to more than one ISP. A multihomed network increases its fault-tolerance, since there is more than one independent path to access the Internet. The multihomed concept was extended to different themes and it is applied from different perspectives, in network, users, devices and services. A multihomed device has more than one interface attached to different networks or APs. Multihomed devices have been increasing their notoriety due to the widespread of modern mobile terminals (e.g. laptops, smartphones and tablet PCs). The multihoming concept applied to devices is usually assumed as heterogeneity multihoming. So, the result is the capability of having more than one access possibility of different technologies. Multihoming is a feature that may improve mobility, if it is applied with appropriate methods to take advantage of available interfaces. A multihomed user has more than one device that should be seen as belonging to the same entity. A multihomed device has several interfaces and respective IPs according to the number of attached interfaces. However, in order to optimize the potentiality of a multihomed user with multihomed devices, several challenges have to be solved.

MIH [28] is the most adopted approach to deal with heterogeneous multihomed devices. It mainly manages the vertical handovers with the aim of maintaining session continuity. Multihoming enables bandwidth aggregation as a cost effective solution to increase the

overall bandwidth of a network. It bundles multiple connections to combine bandwidth to the network. Multihoming provides certain degree in resilience/redundancy against failures: if one connection becomes unavailable then others can takeover. More reliability is assured when multihoming techniques are implemented in multihoming scenarios. Through redundancy, multihoming reduces downtime during failures. It supports methods to maintain the network connectivity in case of natural disasters or adverse events during a long period of time. Multihoming enables load balancing, increasing throughput and diverts traffic from non-functional links in case of failure. Multihoming provides the substrate to manage network resources, exploiting them dynamically and according to the network behavior.

In the scope of this Thesis proposal, multihoming is usually applied to networks and devices. Multihomed networks enable to define scenarios where the last hop network can be connected to different access networks, technologies and Internet Service Providers (ISPs). Multihomed devices provide the ability to illustrate nowadays devices that can take advantage of multiple interfaces, usually from different technologies and IP addresses.

2.2 Mobility Management from different OSI layers

Today, there are several approaches attempting to provide the best mobility management, from different layers of the Open Systems Interconnection (OSI) stack model, as illustrated in Figure 2.1. The solutions cover several OSI layers, from the data link with cellular technologies to the application layer with SIP [30]. The most relevant proposals for each OSI layer are described along this section, which adopt different perspectives in order to solve mobility management issues.

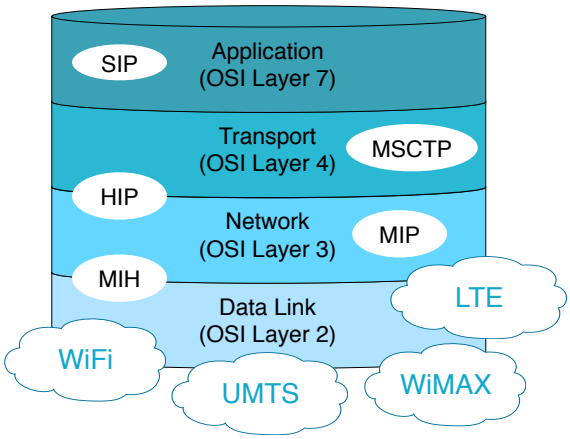


Figure 2.1: Stack perspective for today's mobility management approaches

2.2.1 Data Link Layer

This subsection introduces the mobility management approaches that provide the support at the data link layer of the OSI model.

2G/3G Mobility Management

Mobility management in 2G/3G is performed at the network interface layer, both within the 2G/3G environment and for interconnection with other technologies, e.g.,

WLAN. Mobility management in 3G relates to the tracking and updating of the UE location, and uses mechanisms integrated into Global System for Mobile communication (GSM), General Packet Radio Service (GPRS) or Universal Mobile Telecommunication System (UMTS) [31, 32]. When in idle mode, the UE is not known to the GPRS network and is not able to send or receive any data. First, it needs to attach so that it gets a logical link to the Serving GPRS Support Node (SGSN). There are two types of Public Land Mobile Network (PLMN) backbones, namely, intra-PLMN and inter-PLMN. Intra-PLMN refers to a private IP network which is used for data and signaling transmission within a single PLMN. The inter-PLMN backbone is used to roam between different PLMNs. When roaming, the UE changes its routing location, as illustrated in the left-handed chart of Figure 2.2. If such location is still within the range of the same SGSN, the UE signals the change to the SGSN which simply updates its routing tables. However, if there is a SGSN change involved, then the new SGSN signals the change to the GGSN letting it know about the new Location Area Identity, and sends a message to the old SGSN and to the Home Location Register (HLR). The GGSN then asks the old SGSN to remove the subscriber and to memorize the new SGSN IP address. This is performed simply involving a change of PDP context. Before the UE is ready to send data, PDP context activation is required. This again implies sending the PDP address (IPv4 or IPv6) to the SGSN. Then, the SGSN verifies the subscriber identity and transmits the demand to the GGSN which will memorize the IP address of the SGSN.

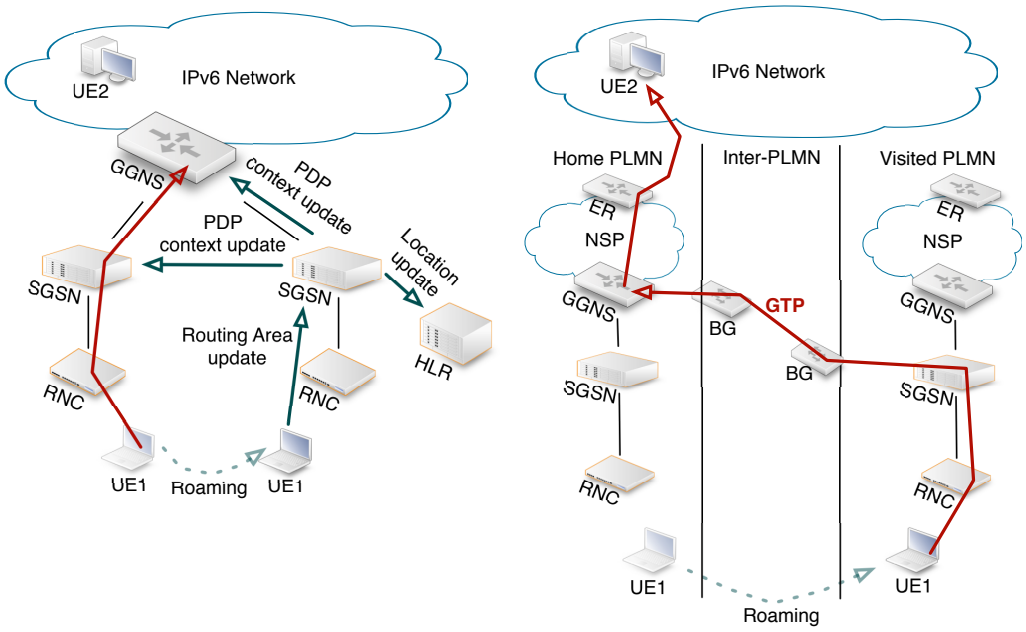


Figure 2.2: 3GPP Intra-PLMN (left) and inter-PLMN (right) roaming

When the UE roams between different PLMNs, as illustrated in the right-handed chart of Figure 2.2, the attachment process is again required. In this process, the UE sends an attachment request to the PLMN SGSN, including identity information, capability, and location. The SGSN then checks the UE identity and performs authentication to secure the transmission path. The SGSN contacts the HLR and obtains information about the roaming subscription and, after finishing the location update procedure, the attachment process is terminated. After the GPRS attach, the UE sends a request for PDP activation providing a reference to the GGSN to be used. This can be either the home GGSN or a visited domain GGSN. While the former implies that IP datagrams are routed to the Home

GGSN being encapsulated in Generic Tunneling Protocol (GTP) over the Inter-PLMN backbone as illustrated, the latter does not require transmission through the inter-PLMN backbone. If the home GGSN is used, then the UE has a network layer identity related to the home network. However, this scenario has the disadvantage of having all traffic flowing through the home GGSN, even if services to be used are locally placed closer to the PLMN. This will strongly impact roundtrip times.

WiMAX Mobility Management

WiMAX [33] incorporates mobility management in the standard 802.16e, in regards to OSI Layer 2 handovers, i.e., when the UE moves between two neighboring Base Stations (BSs). In case of two BSs belonging to the same Network Access Provider (NAP), it is possible that IP connectivity, including IP addresses, is retained after registering to a different BS. In other words, there is no change of IP address. Realizing the need for global mobility, the standard considers MIP [34, 5] or PMIP [7] as a scheme capable of providing OSI network layer mobility management.

The UE retrieves information on possible base stations to perform handover. The UE can perform this either by scanning for available neighbor BSs, or by being informed by its servicing BS of available neighbor BSs. The UE may optionally perform association to these base stations to retrieve more information about their service capabilities.

After this step, either the UE or the BS may trigger a handover decision. The serving BS has the possibility of informing the target BS of the potential handover on behalf of the UE. So, it may include MS related information in order to simplify and shorten handover procedure.

At this stage the UE synchronizes to the downlink transmissions of the target BS and obtains physical layer connectivity. It then performs association to the new BS similar to the network entry procedure. It is possible that the target BS requests to the serving BS for information on the UE. Additionally, the target BS may request additional UE information from an authorizing station (AAA server) via the backbone network.

Depending on the UE information obtained from the serving BS, the target BS may skip some of the network entry steps such as: negotiate basic capabilities, authentication and authorization phase, key exchange, and set up connection phase. The IP connectivity re-establishment phase is optional, so it is not mandatory that a managed UE retrieves new IP address configuration. A UE may retain IP connectivity using, if necessary, L3 protocol exchanges over the secondary management connection. At this state all connectivity has been established and the target BS becomes the new servicing BS. This is followed by the UE terminating services with the original serving BS, where all connections and context belonging to the UE are removed.

In addition to the handover procedure described, two additional handover modes are supported in 802.16: Soft Handover and Fast BS Switching. Soft Handover consists in the ability of a UE to transmit and receive from multiple BSs at the same time. In Fast BS Switching, the UE maintains a list of BSs involved in the process. The UE only communicates with one BS, the anchor BS, in uplink and downlink, for traffic and management information.

2.2.2 Data Link and Network Layers

This subsection introduces the mobility management approaches that provide the support from the data link and network layers of the OSI model.

MIH

MIH [28], also denoted as IEEE 802.21, appears to overcome the heterogeneous mobility among different access technologies. The standard facilitates handovers among heterogeneous networks, such as wireless, cellular and cable, since it provides timely information about link states and available access networks for handover decision makers. It also provides mechanisms to minimize the disturbance of network service during handovers. MIH [35] has link layer intelligence and interacts with upper layers to optimize the handovers, as illustrated in Figure 2.3. So, this solution mainly focuses on Vertical handovers, besides enhancing the Horizontal handovers.

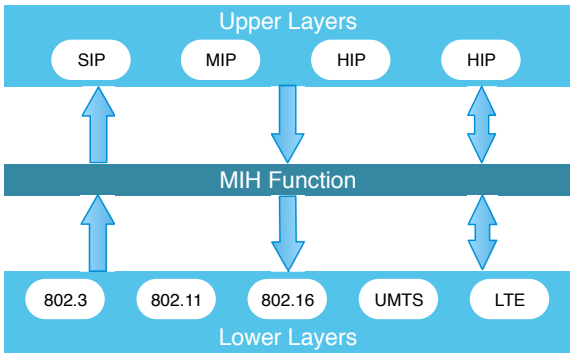


Figure 2.3: Services interaction between MIH components

The IEE 802.21 standard supports cooperative use of information available at the MN within the network infrastructure. In the one hand, the MN is capable of supporting multiple link-layer technologies, which may be wireless or wired, and detect available networks. In the other hand, the network stores network information about lower layers, upper layers and location of MNs. The MIHF is the main structure of the standard, being the logical entity. It is independent of the way that the solution is implemented in the MNs and network. MIHF implements three distinct MIH services: Event Service (MIES), Command Service (MICS), Information Service (MIIS). MIES is responsible to deliver events that may indicate changes in state and transmission behavior of the physical, data link and logical link layer, or predict state changes of these layers. MIES is also used to indicate management or command status, locally or remotely, in the network or specific management entities. MICS enables higher layer to control the physical, data link and logical link layer. The higher layers may control the reconfiguration or selection of an appropriate link through a set of handover commands. The exchanged command can also be local or remote. MIIS provides the mechanisms to obtain network information existing within a geographical area to facilitate handovers, such as access network information, point of attachment information and vendor/network information.

MIHF implements three main phases in the entire process: Discovery, Selection and Completion, illustrated in Figure 2.4. During the initiation phase, both information request and response via neighboring access networks are performed respectively. In the second phase, network selection will be made whereby the decision on the targeted network is chosen. The next phase occurs when the link layer connection from both entities has been established, by releasing unneeded resources from the MN or targeted networks.

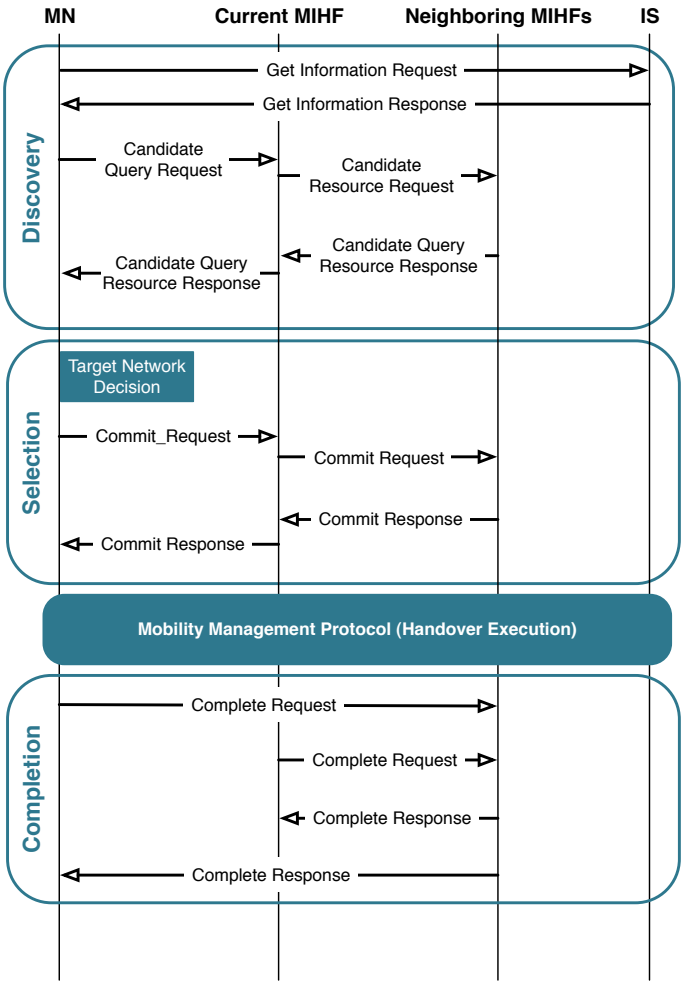


Figure 2.4: Sequence diagram messages of MIH Services

LTE Mobility Management

The Internet Protocol (IP) suite did not originally include any support for endpoint mobility, thus, a whole family of MIP procedures have been introduced over the years in an attempt to provide mobility support in a backward-compatible way. On the other hand, current cellular standards, such as Global System for Mobile Communications (GSM), Evolution Data Optimized (EV-DO), and LTE [36], have all been designed with mobility in mind and integrate the appropriate support in the core network. The cellular control plane includes elements (e.g., the HA of MIP) that store and maintain the state of the terminal for as long as it is associated with the network. It also manages the creation of appropriate bearers to provide seamless access to applications, and provide users with the illusion of a constant connection between the mobile terminal and the network. In the last few years, a powerful impulse toward convergence of "telephone" and "data" networks has led to the progressive standardization of open protocol specifications for telephone functions that were previously implemented by custom interfaces. On one hand, in next-generation LTE cellular networks, the converged protocol of choice is IP, and network interfaces such as Gigabit Ethernet are becoming increasingly common as both local and backhaul interconnects. On the other hand, femtocells are being deployed to the users' premises and communicate with the operator's network over IP through Asymmetric Digital Subscriber

Line (ADSL) or cable modem access connections. An important result of the process of convergence is that, due to the use of a common substrate, telephone control plane functions can be now seen as standard network applications.

The most relevant entities for mobility management are evolved Node B (eNB), Service Gateway (SGW), Packet Data Network Gateway (PGW) and Mobility Management Entity (MME). The messaging sequences are codified by the 3GPP standards as logical interfaces, such as X2 (eNB to eNB), S1 (eNB to SGW or MME), S11 (SGW to MME), S6 (MME to HSS) S5/8 (SGW to PGW), as presented in Figure 2.5. A compact description of the signaling exchanges triggered by call and mobility events can be found in [37].

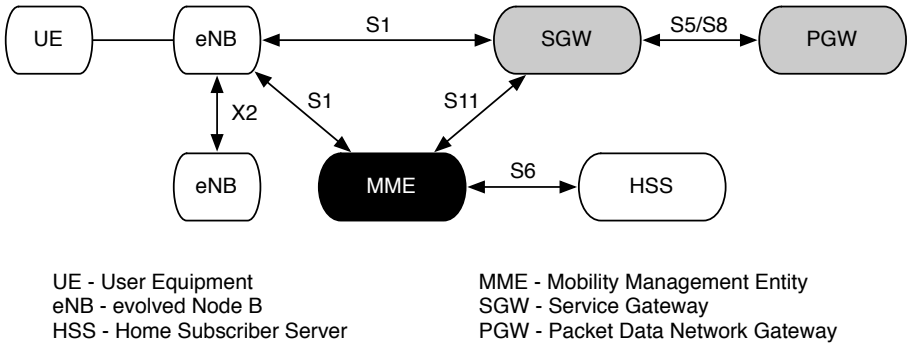


Figure 2.5: Schematic of the main logical interfaces of LTE entities

The functional definition of eNB support all network interface layer features associated to the E-UTRAN OFDM physical interface, and they are directly connected to network routers. There is no more intermediate controlling node (as the 2G/BSC or 3G/ RNC was). This has the advantage of a simpler network architecture (fewer nodes of different types, which means simplified network operation) and allows better performance over the radio interface. The termination of Layer 2 protocols in eNodeB rather than in the RNC helps to decrease data-transmission latency by saving the delay incurred by the transmission of packet repetitions over the hub interface between RNC and eNB. From a functional perspective, the eNB supports a set of legacy features, all related to physical layer procedures for transmission and reception over the radio interface. A new interface (X2) has been defined between eNodeBs, working in a meshed way (meaning that all eNBs may possibly be linked together). The main purpose of this interface is to minimize packet loss due to user mobility. As the terminal moves across the access network, unsent or unacknowledged packets stored in the old eNB queues can be forwarded or tunneled to the new eNB thanks to the X2 interface.

The Home Subscriber Server (HSS) is the concatenation of the HLR and the Authentication Center (AuC), two functions being already present in pre-IMS 2G/GSM and 3G/UMTS networks. The HLR part of the HSS is in charge of storing and updating when necessary the database containing all the user subscription information, including user identification and addressing, as well as user profile information. The AuC part of the HSS is in charge of generating security information from user identity keys. This security information is provided to the HLR and further communicated to other entities in the network.

From a functional perspective, the SGW is the termination point of the packet data interface towards E-UTRAN. When terminals move across eNB in E-UTRAN, the SGW serves as a local mobility anchor, meaning that packets are routed through this point for intra E-UTRAN mobility and mobility with other 3GPP technologies, such as 2G/GSM and 3G/UMTS.

Similarly to the SGW, the PGW is the termination point of the packet data interface towards the Packet Data Network (PDN). As an anchor point for sessions towards the external PDNs, the PGW also supports Policy Enforcement features (which apply operator-defined rules for resource allocation and usage) as well as packet filtering (like deep packet inspection for virus signature detection) and evolved charging support (like per URL charging).

The main mobility-related role of the MME is to keep track of the location or Tracking Area (TA) [38], and the associated state (network identifiers, cryptographic keys) of the user equipment as it moves through the cellular network, thus guaranteeing its reachability in the event of a network-initiated voice or data connection [39]. Because of power management concerns, UEs spend most of the time in low-power mode with their transceiver turned off. Mobile devices listen at regular intervals to the beacons sent by the local eNB. To minimize signaling, UEs only explicitly notify the MME when they detect they have moved to a new TA. Therefore, the MME needs to maintain active records of the state of all UEs when it is not involved in communication, in order to issue wake up calls and notifications of incoming network events. A call directed towards an UE is known as a user-terminated session. When such a call is made, the MME performs paging, that is, it contacts all eNBs in the last known TA in which the UE was detected before widening the scope of the search. If the UE cannot be found, it interrupts the call attempt. When switched on, a piece of user equipment associates with the local eNB and, after successfully completing the authentication procedure, registers with the MME in the network in which it is roaming.

2.2.3 Network Layer

The Mobile Internet Protocol (MIPv4) [34] was first developed in the context of IPv4, and it was presented as the first solution for the global mobility issue, being suitable for large movements. The architecture of this protocol, presented in Figure 2.6, allows terminal mobility, and it is composed by four main entities: MN, CN, HA, Foreign Agent (FA). MN is the terminal that moves through the different networks, changing its access network. CN is the terminal that communicates with the MN. HA is typically a router in the HN of MN, which is responsible for registering the MN location and forwarding it the communications. FA is also a router in the FN, which is visited by the MN. Home Address (HoA) is the IP address of the MN in its HN, and CoA is a temporary IP address acquired when the MN visits the FN.

In this protocol we have to use two different IP addresses, in order to maintain the MN reachable in the FN. The HoA is necessary when other hosts want to communicate with the MN and it is permanently associated to it. The aim of the MIP protocol is to redirect, through IP tunnels, the packets received in the HN, to the FN, where the MN is temporary located.

The IP-in-IP tunnels used by MIPv4 are dynamically managed, in order to allow the MN to be accessible from its HoA. MIPv4 enables that applications designed for the traditional non-mobile Internet will continue to work even in mobile environments. The purpose of the MIPv4 protocol is to allow applications to keep the communications between hosts, while the MN roams between different IP networks. While in standard IPv4 a movement would result in disruption of the ongoing sessions to the MN, with MIPv4 only a short disruption is perceived, but session continuity is assured.

In MIPv4, when a MN moves to a FN, the packets have to continue to be delivered to HN, and the receptive HA is responsible for sending them to the FA through an IP tunnel. When the MN receives the packets, it replies to the CN directly from its new location,

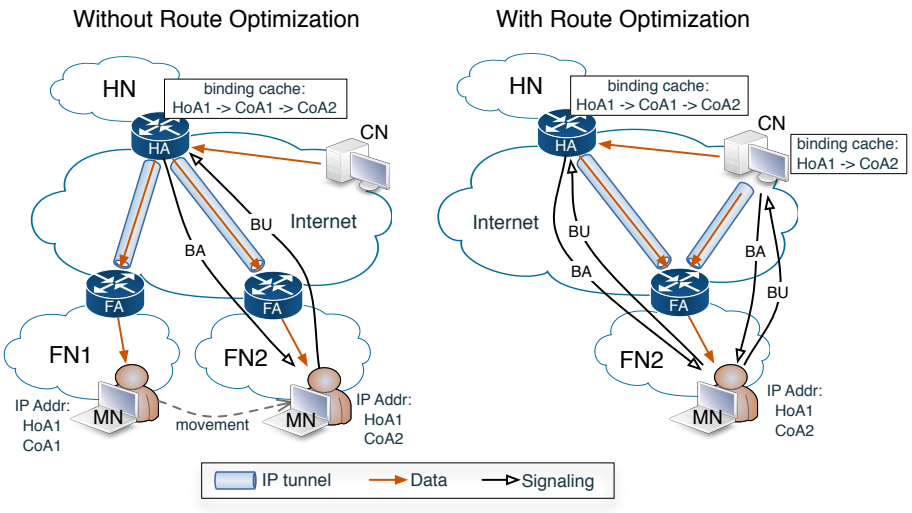


Figure 2.6: MIPv4 scheme

creating a triangle routing. This is a weakness concerning the performance of MIPv4 protocol, which can be solved by introducing a feature in the original protocol, named Route Optimization [40]. The Route Optimization feature eliminates the transparency of mobile process (CN knows current IP address of MN), but it increases the efficiency and reduces delays and resources allocation. In original MIPv4 scenario, the CN never knows when the MN moves from its HN to a FN, but with Route Optimization, the CN gains the capacity of communicating directly with the MN. MN moves to a FN and it sends a Binding Update (BU) message to the CN and the current location to the HA at the same time. In order to know the current location of MNs, CN needs to create a database with all CoAs of MNs, named binding cache. If the MN's CoA does not appear in the CN binding cache, CN uses the original process by sending data packets to the HA of MN.

More mobility protocols for network layer are presented in the next section. However, the main concepts behind all the standardized IP mobility protocols, such as MIPv6 [41] and PMIP [7] based protocols, are the ones already presented before by MIPv4.

2.2.4 Network and Transport Layers

Host Identity Protocol (HIP) [41] provides secure communication with an architecturally sound approach, where it highlights the decoupling between Identification and Location. HIP architecture proposes an alternative for the current scheme of dual use of IP address as locators (routing labels) and identifiers (host identifiers). This separation is explored in [42] to facilitate and support personal mobility and multihoming features. It defines a generalized *Locator* parameter for use in HIP messages. The *Locator* parameter allows a HIP host to notify a peer about alternate addresses at which it is reachable. Regarding terminal mobility, original HIP supports only the function of maintaining sessions active with CNs across subnet changes, since the reachability for incoming requests are not guaranteed by HIP.

However, a HIP extension [43] allows a HIP node to store in the DNS its Host Identity (HI, the public component of the node public-private key pair), Host Identity Tag (HIT, a truncated hash of its HI), and the Domain Names of its rendezvous servers. HIP uses public cryptographic keys (public/private key pair) to identify the host. So, any change of IP addresses can be dynamically authenticated between hosts. HIP introduces a new

namespace, the Host Identity namespace.

HIP has HI and HIT as two different representations of hosts' identification, where the HI is a public key that directly represents the host. Since each public key algorithm has different key lengths, it is not good to use HI as a packet identifier and consequently HIT becomes the operational representation. HIT is a hashed encoding included in the HIP payloads to index the corresponding host state. The HIP base exchange is a two-party cryptographic protocol used to establish communications context between hosts. The base exchange is a Sigma-compliant four-packet exchange, illustrated in Figure 2.7.

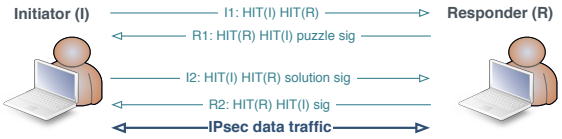


Figure 2.7: HIP base exchange messages

The system initiating a HIP exchange is the Initiator, and the peer is the Responder. These definitions are adopted in the initial base exchange. The Initiator first sends a trigger packet, I1, to the Responder. I1 contains only the HIT of the Initiator and possibly the HIT of the Responder, if it is known. Note that, in some cases, it may be possible to replace this trigger packet by some other form of a trigger, in which case the protocol starts with the Responder. The second packet, R1, starts the actual exchange. It contains a puzzle with a cryptographic challenge that the Initiator must solve before continuing the exchange. The level of difficulty of the puzzle can be adjusted based on the level of trust with the Initiator, current load, or other factors. In addition, the second packet contains the initial Diffie-Hellman parameters and a signature. In the I2 signed packet, the Initiator must display the solution of the received puzzle. Without a correct solution, the I2 message is discarded. The I2 contains a Diffie-Helman parameter that carries needed information for the Responder. The R2 packet is signed and it finalizes the base exchange.

A HIP association between two hosts may need to be updated over the time, due to several purposes, such as rekey expiring user data security association, add new security association and change of hosts' IP address. HIP Update messages carry a monotonically increasing sequence number and are explicitly acknowledged by the peer. Every lost update or acknowledgment message may be recovered via retransmission. Multiple update messages may be outstanding under certain circumstances.

2.2.5 Transport Layer

Mobile Stream Control Transport Protocol (MSCTP) based approaches [44, 45, 46, 47] are alternative solution for IP mobility solved at transport layer in an end-to-end fashion and based on SCTP [48].

SCTP is a transport protocol running on top of IP that encompasses basic functionalities of TCP together with other interesting protocol mechanisms. SCTP performs multihoming, which enables a single SCTP endpoint to support multiple IP addresses with a single association. It is a powerful framework for IP mobility at transport layer, as it already separates the identity of an end system from its current address to which packets are sent. However, the multihoming mechanism's purpose is to increase association reliability in wired networks. So, IP addresses of all end systems are fixed and known in advance. This does not work in a mobile environment as a mobile host does not have a fixed previously known IP address.

An SCTP endpoint can use multiple IP addresses for an association. These are exchanged during the initiation of the association. The multiple addresses of the peer are considered as different paths towards that peer. This means that a server must use multiple IP addresses to provide the mobile client with multiple paths. These will be used while moving between locations. It should be mentioned that this path-concept is used only for redundancy, not for load sharing. Therefore one path is used for normal transmission of user data. It is called the primary path.

Therefore, the basic idea of MSCTP is to let a mobile host have more than one IP address in SCTP association if several access points are available simultaneously. So, MSCTP mobility concerns the terminal mobility through the multihoming concept, where the main scenario exploited by MSCTP is the overlapping of the old AP, in which the MN is currently in use, and the new AP, in which the MN will perform handover. So, the MN can obtain an IP address from the new AP to prepare the handover process, adding the new IP address in SCTP association of the correspondent hosts. However, MSCTP does not support reachability of the MN for incoming requests, since its main goal is to keep sessions across subnet changes. Session continuity performance depends on the distance between the MN and the current CNs, since new IP addresses and ports are directly sent to the CNs.

2.2.6 Application Layer

Following the regular SIP operation [29], SIP-based mobility concerns only nomadism in the form of terminal mobility, given that each time a UE roams, it has to perform re-registration. Terminal mobility refers to support for a device to move between IP subnets while still being reachable for incoming requests, and while keeping sessions across subnet changes. Terminal mobility impacts SIP at three different stages, namely, before a session (pre-call), during a session (mid-call) and to recover from network partitions (network partition).

In the pre-call stage, the MN is assigned a new IP address before establishing any SIP call. In this case, the MN simply has to *re-register* with its SIP registrar every time it receives a new IP address. This means that, in order not to break sessions, the application must be aware of the IP address change, which may be performed either by having OS polls or by having some form of notifying the applications of such changes.

During the mid-call stage, the MN sends a new *"Invite"* request to CNs without going through any SIP proxies. The *Invite* request includes updated session description holding the new IP address. The session is not broken, but a delay directly proportional to the distance between the SIP entities is experienced. This latency can be overcome by having the UE using the address of a SIP proxy (or a RTP translator associated with the SIP proxy) instead of its own. The SIP proxy performs network address translation and the RTP translator intercepts the media packets, redirecting them to the new UE IP address. In this case, the incurred delay equals the handover delay between UE and RTP translator.

For the network partition stage, there are several possibilities. If the stage lasts less than 30s then the regular SIP retransmission operation provides the means to have automatic recovery. However, if the stage lasts longer than that, then it may happen that updates are lost, or a CN moves. This requires each involved peer to perform SIP *Invites*, falling back to the case where *Invites* are sent to the canonical address or to the home proxy of the CN.

SIP can also be engineered to support session mobility using three different methods [49].

A first method assumes that the involved entities hold IP addresses and ports that

are provided and managed by a server (primary end-system). It is up to the primary end-system to convey information about IPs and ports to each party by means of *Invites*.

A second method is the so-called third-party call control where a third element is placed between the participants. It is up to this third-party to control (as the name points out) the mapping between session requests and destination. While providing good session isolation, each of the involved parties performs requests to the third-party. This approach has the disadvantage of keeping the third-party involved in the session, as it will be contacted to change or terminate the session.

The final method is the so-called *Refer* method. This method builds upon the third-party control by having the session transferred to the new destination. The third party is only used during negotiation by regular *Invite* exchange.

Finally, personal mobility allows addressing a single subscriber independently of the terminal by having the same logical address. Different terminals may be used at the same time or in isolation. An example of personal mobility is the case where a subscriber advertises different *urls*, e.g., for private and personal contacts. Personal mobility can only be supported by SIP if the registrars recognize different devices as belonging to the same person.

As described, SIP mobility can only partially replace network-layer mobility, given that it may easily support terminal mobility. Nomadism can also be supported with minor modifications to SIP elements. Moreover, the change of IP addresses strongly impacts SIP in terms of handoff delay. Consequently, if using SIP mobility alone, both session and personal mobility require significant changes either to SIP elements or to the involved applications. On the other hand, SIP mechanisms do not exclude the use of a network-layer mobility management scheme such as MIPv6. When used together, these two mechanisms will profit of their combined strengths. However, given that MIPv6 relies on the use of two different addresses, the HoA and the CoA, problems arise due to the combination of SIP and MIPv6.

2.2.7 Summary

This section presents an overview on how it is provided mobility management from the data link layer to the application layer of the OSI model. We are going towards a global network architecture infrastructure based on the IP technology. Cellular network architectures introduce the recent LTE as the first cellular generation adopting the IP technology as a built-in technology. Hence, we are expecting an Internet access infrastructure connected through IP technology where users freely move, and where each network architecture from a different provider will be configured in a different IP domain, or even different IP domains inside the same ISP. This means that mobility management in future networks cannot be mainly solve in the data link layer, and need to be solved by upper layers. However, the data link layer remains quite interesting for specific scenarios with local mobility or to be able to provide seamless and soft handovers, since it is the first layer to detect a disruption and to initiate the handover.

The application layer can be adopted to solve the mobility management in future network architectures, but it is the one that is more inefficient regarding detection and update time, as well as to maintain all the mobility context. The application layer is not the properly layer for the global mobility management, but it may be quite useful to be integrated in more autonomous and smart applications of the future, since these applications can control the mobility management of its own data sessions, independently of the mobility management of the system and based on privilege information collected/exchanged by the application itself.

The mobility management bring more advantages when provided by network and/or transport layers. The solutions working in these two layers or in a middle layer between the two are able to provide the appropriate delivery of the content while reducing network resources consumed. However, the mobility management in the network layer (IP) is the only one that can be applied in a transparent way to users and service providers. IP mobility management is also able to provide better performance as long as more entities integrate the mobility management process (e.g. users or service providers), being much easier to be implemented in current network architectures and naturally evolving in future network architectures. The mobility management in the transport layer requires changes in both the endpoints (e.g. users and service providers), which is quite difficult to be achieved for the majority of the Internet services (e.g. video streaming), but it is interesting for specific applications between end-users or smaller service providers open to change their systems/networks.

From the reasons pointed out, we decided to study and improve the mobility management in the network layer; thus, the following sections will analyze current centralized IP mobility management standards and introduce the novel distributed IP mobility management trend.

2.3 Centralized IP Mobility Management

The current centralized IP approaches share the same mobility management model, in which a static mobility anchor is responsible to route all mobile packets to the current location of the users, as well as to manage all mobility context of the users connected to the network. All the mobility management process is usually centered in a single node or at maximum in a couple of nodes, serving thousands of users. This section describes the main centralized approaches, which are the current standards in the IP mobility, such as MIPv6, Fast handovers for Mobile IPv6 (FMIPv6), HMIPv6 and PMIP. There are much more MIPv6 or PMIP based approaches, but they introduce minor changes to the standards presented next.

2.3.1 Host-based IP Mobility

We first describe current host-based IP mobility approaches and mechanisms. It is explained MIPv6, which is the main standard host-based IP mobility protocol, and then we give an overview of all the protocols and mechanisms proposing improvement or modification to MIPv6.

MIPv6

MIPv6 [41] is similar to the Mobile Internet Protocol version 4 (MIPv4) protocol, already explained in the mobility management approaches provided in the network layer, thus, we highlight just the main improvements and new features. IPv6 [50, 51] automatically enables every node with mobile functionality, the so-called MIPv6. MIPv6 provides a set of richer features, when compared to MIPv4 [34], namely, better processing of destination options, auto-configuration, routing headers, flow-label and integrated security.

The current MIPv6 specification supports IPv6 only, but the work presented in [52] extends those standards to allow the registration of IPv4 addresses and prefixes, respectively, and the transport of both IPv4 and IPv6 packets over the tunnel to the home agent. It also allows the mobile node to roam over both IPv6 and IPv4, including the case where

Network Address Translation (NAT) is present on the path between the mobile node and its home agent.

The basic operation of MIPv6 is illustrated in Figure 2.8.

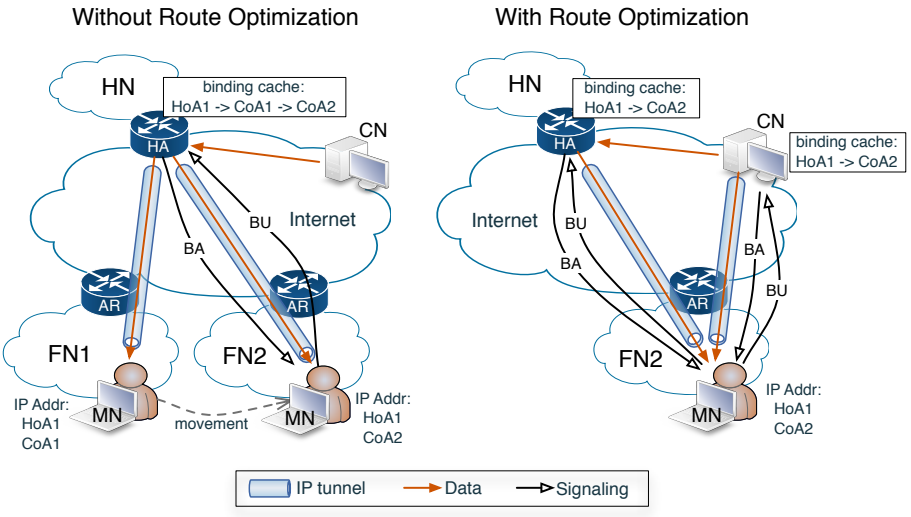


Figure 2.8: MIPv6 example

As illustrated in Figure 2.8, MIPv6 relies on three major components from MIPv4's legacy. The MN, that is an IPv6-enabled node that holds an IPv6 address from its HN range, and CoA provided when it moves to the range of a FN. The CN is another MIPv6 enabled node with whom the MN holds some form of session. When away from its HN, the MN is identified by at least one CoA besides the automatically generated link-local and HN addresses. Each CoA is created using the prefix of its corresponding FN, e.g. DHCPv6 or auto-configuration. This allows packets to be routed to the MN by the regular mechanisms while transparently keeping already established sessions, e.g., TCP sessions. The HA router keeps the list of CoA provided by the MN. This way, when the HA receives packets destined to the MN, it sends them (encapsulating the packets) to the CoA, i.e., the new location of the MN. If a CN is not aware that the MN roamed, then it keeps sending packets to the HA, which on its turn re-directs them to the MN. When the MN receives these encapsulated packets, it sends binding messages to the CN, letting it know its CoA. This way, triangular routing, a major problem in MIPv4, is avoided in MIPv6. When attaching to a new subnet, the MN has to perform a number of steps, e.g., obtain L2 connectivity and then detect the L3 point of attachment, perform link configuration, router discovery, new CoA, bindings to the HA and CNs. This may result in a significant time gap between the moment the MN gets L2 connectivity, and the moment when it can transfer data. This gap may result in significant packet losses.

Fast Handovers for MIPv6

Fast Handover for MIPv6 (FMIP) [53] main purpose is to allow a MN to send packets as soon as it detects a new subnet link. It focuses on how to deliver packets to a MN as soon as its attachment is detected by the new access router. FMIP is based on MIP, so its focus is on terminal mobility. It improves the handover management through the introduction of IP messages necessary for its operation regardless of the link technology.

The ability to immediately send packets from a new subnet link depends on the latency associated with the achievement of IP connectivity. In fact, it is strictly related with

latency of movement detection and CoA configuration. FMIP enables a MN to quickly detect a movement to a new subnet by providing the new AP and associated subnet, the prefix information when the MN is still connected to its current subnet. In FMIP, the MN searches for available APs and subnets information at any time while connected to its current router using link-layer specific mechanisms.

The MN resolves the identifiers associated with the APs (ID and respective info), which are used in readily detecting movement. When the MN attaches to an AP (unique ID), it knows the corresponding new router's information, such as its prefix, its IP address and L2 addresses. The Router Solicitation for Proxy Advertisement (RtSolPr) and Proxy Router Advertisement (PrRtAdv) messages are used for aiding movement detection, since the MN uses them to formulate a prospective New Care-of Address (NCoA). The early prefix known eliminates the latency of its discovery. Moreover, this prospective address can be used immediately after the new subnet link attachment, after the MN receives a Fast Binding Acknowledge (FBA) message prior to its movement.

In order to reduce the BU latency, FMIP specifies a tunnel between the Previous Care-of Address (PCoA) and the NCoA. This tunnel is established when the MN sends a Fast Binding Update (FBU) message to its Previous AR Previous Access Router (PAR) and remains active until the MN completes all BUs. It allows to forward packets between PAR and New Access Router (NAR) in order to avoid dropping packets in the MN's handover. So, it improves the session continuity with fast handovers and with the minimum loss packets.

FMIP provides Handover Initiate (HI) and Handover Acknowledge (HAck) messages to allow ARs to transfer network-resident contexts, such as access control, Quality of Service (QoS) and header compression, together with handover.

Hierarchical Mobile Internet Protocol

Hierarchical Mobile Internet Protocol (HMIP) [6] concepts are an extension of the MIP protocol, explained previously. Just like MIP, HMIP is independent of the underlying access technology, allowing mobility within or between different types of access networks. The main goal of HMIP is the same as the one of MIP: to allow nodes to move within the Internet topology while maintaining reachability and on-going connections between MNs and CNs. MIPv6 requires that the MN sends BUs to its HA and CNs, every time it changes its access network. The higher round-trip time due to this BUs disrupt active connections every time a handover to a new AR is performed.

The objective of HMIP is to eliminate this additional delay from the time-critical handover period, in order to improve the performance of MIP. HMIP reduces the number of messages sent over the air interface to all CNs and HA to improve efficiency in wireless environments.

HMIP introduces a new entity, called Mobility Anchor Point (MAP), which introduces a hierarchy of proxy HAs (local anchor point) to reduce the mobility signaling outside the local domain. Unlike FAs in IPv4, a MAP is not required on each subnet and can be located at any level in a hierarchical network of routers, including the AR.

HMIP main goal focuses in the improvement of local mobility. When the MN performs local handovers inside the same MAP domain, the MN sends BUs to the local MAP (Figure 2.9) rather than the HA (typically further away) and all CNs. Besides, only one BU message (Figure 2.9) needs to be transmitted by the MN before the HA or CNs re-route the data traffic to its new location, independently of the number of CNs communicating with the MN. Furthermore, HMIP allows MNs to hide their location from CNs and HAs, while using MIP route optimization. HMIP is capable to support terminal mobility and

it allows the improvement on service continuity performance in local movements, since it reduces the handover signaling and quickly reacts to local movements.

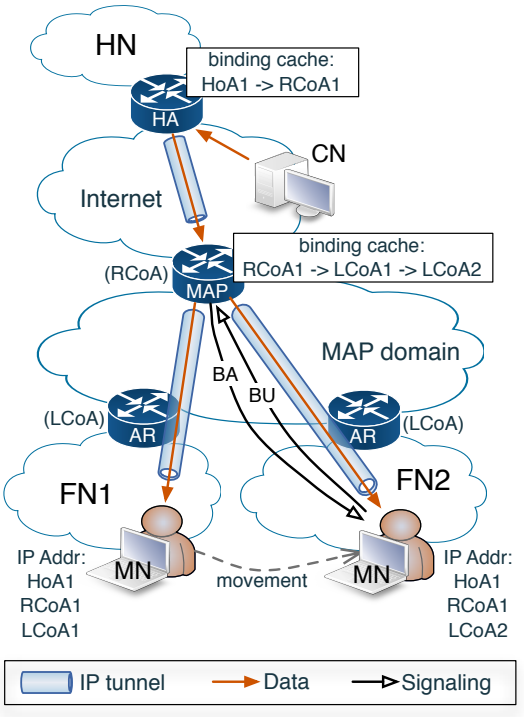


Figure 2.9: HMIPv6 example

HMIP introduces two CoA for the MN, instead of the unique CoA adopted by MIP. A Regional Care-of-Address (RCoA) is obtained by the MN from the visited network. An RCoA is an address on the MAP's subnet that is auto-configured by the MN when receiving the MAP options. The On-link Care-of-Address (LCoA) is the on-link CoA configured on a MN's interface based on the prefix advertised by its default router.

When a MN moves to a new MAP domain, it needs to configure both RCoA and LCoA. A MN entering a MAP domain will receive RAs containing information on one or more local MAPs. The RCoA is formed in a stateless manner. After forming the RCoA based on the prefix received in the MAP option, the MN sends a local BU to the MAP that includes the MN's RCoA in the HA option. The LCoA is used as the source address of the BU, which will bind the MN's RCoA to its LCoA. Then, the MAP performs Duplicate Address Detection (DAD) for the MN's RCoA on its link, and sends a BAck to the MN, informing a success or failure. In a successful response, a bi-directional tunnel between the MN and the MAP is established. MAP will receive all packets on behalf of the MN it is serving, and will encapsulate and forward them directly to the MN's current IP address (LCoA).

An MN may decide to register with more than one MAP simultaneously and to use each MAP address for a specific group of CNs, in order to improve the efficiency of the network bandwidth's use.

Network Mobility Basic Support Protocol

Network Mobility (NEMO) [54] Basis Support protocol ensures session continuity for all nodes in the Mobile Network (MNet), even when the Mobile Router (MR) changes

its point of attachment to the Internet. It also provides connectivity and reachability for all nodes in the MNet as it moves, so its focus is the terminal mobility. A MNet is a network subnet that moves and attaches to arbitrary points in the routing infrastructure. A MNet can only be accessed via specific gateways called MRs (Figure 2.10) that manage its movement, so, each MNet is at least associated with one MR. An MR is like a MN in MIPv6 with routing capability between its point of attachment (CoA) and a subnet that moves with the MR. The MR maintains a bi-directional tunnel to a HA (Figure 2.10) that advertises an aggregation of MNets to the infrastructure. The MR is not responsible to distribute the MNet routes to the infrastructure at its point of attachment.

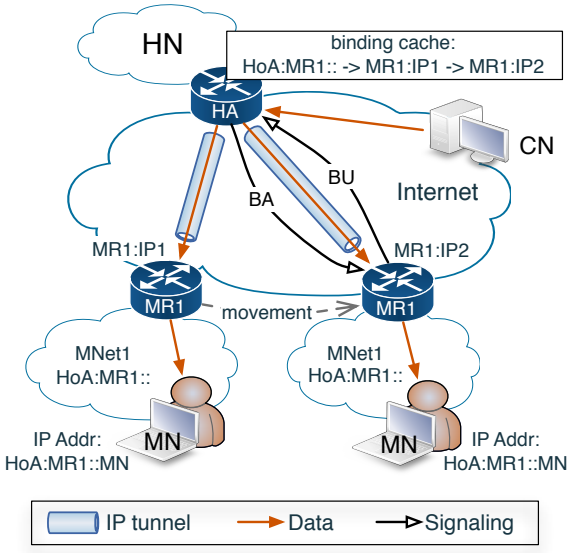


Figure 2.10: NEMO Basic Support Protocol example

An MR has an unique HoA through which it is reachable when it is registered with its HA. The HoA is configured from a prefix aggregated and advertised by its HA. The prefix is advertised on the home link or delegated to the MR. The MR can have more than one HoA if there are multiple prefixes in the home link. The MR advertises one or more prefixes in the MNet attached to it.

When the MR attaches to a new access router different from his home router, it acquires a CoA from the visited router. When the CoA is acquired, the MR sends a BU message to his HA. Then, the HA receives the BU and creates a cache entry binding with MR's HoA and its CoA.

2.3.2 Network-based IP Mobility

It is described the current network-based IP mobility approaches and mechanisms. We start by explaining the PMIP approach, which is the main standard host-based IP mobility protocol, and then we give an overview of all the protocol and mechanisms proposing improvement or modification to PMIP.

Proxy Mobile Internet Protocol

Proxy Mobile Internet Protocol (PMIP) [7] adopts a network-based mobility approach to solve the IP mobility challenge. Network mobility strategy is agnostic to the capability in the IPv6 stack of the nodes that is serves. One of the main advantages of developing a

network-based mobility protocol based on MIPv6 is the reuse of the HA functionality and mobility signaling messages. The other main advantage is that the common HA would serve as the mobility agent for all types of IPv6 nodes. So, PMIP supports mobility for IPv6 nodes, hiding the mobility process from them by extending MIPv6 signaling between a network node and a HA. The mobility entities in the network track the MN's movements, initiating the mobility signaling and establishing the required routing state, allowing terminal mobility for end-devices.

PMIP introduces two new elements as the core functional entities of mobility infrastructure: Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG). LMA is responsible for maintaining the MN's reachability state and is the topological anchor point for the MN's HN prefix(es). MAG is the entity that performs the mobility management on behalf of a MN, and it resides on the access gateway link, where the MN is anchored. MAG is responsible for detecting the MN's movement to and from the access link, and for initiating binding registration to the MN's LMA, using the Proxy Binding Update (PBU) and Proxy Binding Acknowledge (PBA) messages. There can be multiple local mobility anchors in a PMIP domain, each serving a different group of MNs. Figure 2.11 illustrates an example of PMIP architecture scheme.

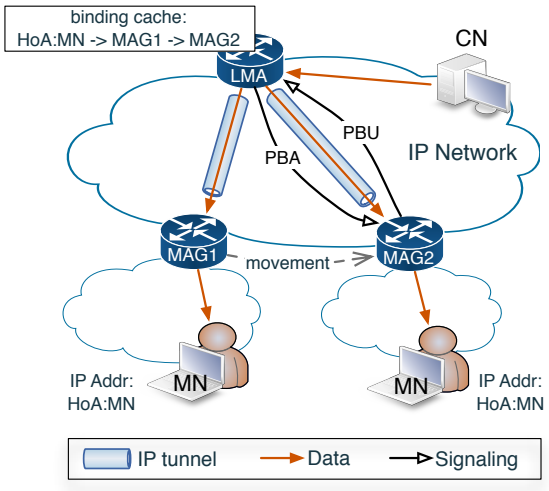


Figure 2.11: Proxy MIP example

When the MN enters in a PMIP domain, it attaches to an access link. The MAG on this access link identifies the MN to determine if the MN is authorized for the network-based mobility management service. An affirmative authorization ensures that the MN using any of the allowed address configuration mechanisms will obtain the address on the configuration interface and move anywhere in that PMIP domain. Since the MN sees the entire PMIP domain as a single link, the network ensures that the MN does not detect any change regarding layer 3 attachment when the MN changes its point of attachments in PMIP domain.

PMIP allows MNs to connect its domain through multiple interfaces and over multiple access networks, but the mobility management is independently provided to each interface; thus, each interface is like a different user. Moreover, personal mobility is not achieved regarding the terminal in use, since different terminals are identified as different users. The network allocates an unique set of home network prefixes for each of the connected interfaces. The MN will be able to configure address(es) on those interfaces from the respective HN prefix(es).

Fast Handovers for Proxy Mobile IPv6

There is a proposal for a fast handover protocol [55] for the network-based mobility management protocol PMIPv6 [7] presented before. In PMIP, a MAG is responsible for detecting the MN's movements to and from the access link, and for initiating binding registrations to the MN's assigned LMA. Fast handover for PMIPv6 assumes that, if MAGs can be informed of the detachment and/or attachment of the MN in a timely manner (e.g. lower-layer signaling), it becomes possible to optimize the handover procedure, which involves establishing a connection on the new link and signaling between mobility agents, compared to the baseline specification of PMIPv6.

The proposed protocol specifies a bidirectional tunnel between the Previous Mobile Access Gateway (PMAG) and the New Mobile Access Gateway (NMAG) to tunnel packets from/to MN. In order to enable the NMAG to send the PBU, the HI and HAck messages in [53] are extended for context transfer, in which parameters such as the MN's Network Access Identifier (NAI),Home Network Prefix (HNP), and IP HoA are transferred from the PMAG.

Since a MN is not directly involved with IP mobility protocol operations, it follows that the MN is not directly involved with fast handover procedures either. Hence, the messages involving the MN in [53] are not used when PMIP is in use. More specifically, the RtSolPr, the PrRtAdv, FBU, FBA, and the Unsolicited Neighbor Advertisement (UNA) messages are not applicable in the PMIP context. A MAG that receives a RtSolPr or FBU message from a MN should behave as if they do not implement FMIPv6 as defined in [53].

PMIPv6 with Bicasting for Soft Handover

In PMIPv6, the mobile agent located in the network will perform the mobility signaling instead of the MN, and will keep track of the movement of the MN. It is noted that PMIPv6 is used mainly for binding update of the location of MNs. There is research work made on the PMIP handover, but there are still several issues that need to be solved in the perspective of seamless handover. There is an approach in [56] that describes a new handover scheme of PMIPv6 with bicasting for seamless IP handover, in which the LMA will bicast the data packets to the PMAG and NMAG towards the MN, when the MN is in the handover region.

First, when the MN moves to the bicasting region, it detects that a handover is imminent and reports its identification and the AP to which the MN is most likely to move. The MN identification could be the NAI or a Link Layer Address (LLA), or any other suitable identifier. This step is access technology specific. In some cases, the Previous Access Network (PAN) will determine which AP the MN is moving to. The PAN, to which the MN is currently attached indicates the handover of the MN to the PMAG.

After the PMAG receives a HI, the PMAG sends HI message to NMAG where the HI message includes MN's IP address that are both Proxy CoA and HoA, LMA address and MN's Identifier. When the NMAG receives HI message, it should examine whether a tunnel to the LMA exists or not. If the tunnel has not been established, it should establish the tunnel from the LMA. To establish the tunnel, the NMAG sends a PBU message to the LMA. It includes the MN Identifier and MN HHoA.

When the LMA receives the PBU message, it creates a new binding entry. If the LMA successfully processes the PBU, it sets the tunnel with NMAG for sending and receiving data packets. After the successful establishment of the tunnel, the LMA sends a PBA message, and it examines whether or not the PBU message was processed successfully. If there is a failure, the PBA message indicates the failure. On the other hand, NMAG creates

a tunnel to the LMA and ensures that the packets with destination address as PCoA are copied and forwarded over the tunnel. It also creates a host route for forwarding packets to the MN. The NMAG sends a HAck message back to the PMAG to indicate whether the handover procedure was successfully performed or not.

2.3.3 Summary

This section describes the main centralized mobility management approaches, which are the ones currently adopted in the network architectures. Besides the differences between these approaches, all share the same mobility management model, where a static and centralized node (e.g. HA or LMA) is responsible to provide mobility support to the users connected to the network, such as routing all data packets, maintain all the mobility context and exchange all the signaling. This centralized mobility management model has limitations and issues regarding scalability, security and performance, when applied to the recent network trends of flatten architectures and distributed content servers. It is necessary to develop end-to-end mobility management schemes more flexible and adapted to the recent network trends, in order to optimize the network resources and to improve the user experience. The following section provides and overview of the novel distributed mobility management concepts and approaches, which fit the main purposes of this Thesis.

2.4 Distributed IP Mobility Management

Current mobility management standards assume a predefined static and centralized mobility management model, where the mobility anchor is responsible to route all mobile data packets and manage all the mobility context of the users connected to the network. This centralized model is quite prone to issues and limitations regarding scalability and performance; thus, IETF charted the Distributed Mobility Management (DMM) [57] working group after the beginning of the work developed in this Thesis, which is focused on a new framework for distributed mobility management in flatten network architectures. Most of the approaches that will be described were introduced during this Thesis; most of the approaches in the scope of DMM were published at the same time or after the research work developed in this Thesis. Hence, this section focuses on the novel mobility management approaches that aim to distribute the mobility management functionalities and elements through the network, where most of the proposed approaches are part of the DMM working group drafts. The approaches are grouped into network and host based approaches to be easily understood, and we also present some assistance mechanisms that can be merged with any of the two models to improve the mobility management performance.

The centralized mobility anchor always provides the anchoring in a static and centralized node, while the majority of the distributed approaches provide the anchoring of a new session in the current AR of the MN or in the CN network. Thus, sessions are anchored in the establishment phase to a node, based on a predefined selection scheme, which remains the mobility anchor for that session until the end of the session. There is some work [58, 8, 9, 10] focused on the problem statement and requirements. This work highlights the main issues and limitations of the centralized mobility model when compared with distributed mobility management. It is presented a characterization of the mobility functionalities and how they are deployed in current centralized approaches. It is identified the limitations in the current practices with respect to providing the expected DMM functionality. The main objective of these approaches is to understand how the mobility management can be enhanced in order to meet the primary goals in the network

evolution. They look for solutions that improve mobility management for future network evolution trends, such as scalability, avoid single points of failure, enable transparent mobility support to upper layers, and provide mobility support just when needed. As referred in the contributions, we collaborate in the definition of the problem statement in the DMM draft [8], since the characterization and analysis of mobility management was one of the first steps of the work developed in this Thesis.

There are some recent scientific publications [59, 60, 61] that provide an overview of the novel distributed mobility approaches in comparison with centralized mobility management. In [59], the authors briefly review the state of mobility management in infrastructure networks, summarize current developments in standardization in general, and introduce the ongoing efforts in the IETF DMM working group. In [60], it is described how current mobility management network architectures, which are being redesigned towards a more distributed operation, are able to mitigate the problems of current centralized mobility models. The authors present an overview of the solutions explored by the two main standardization bodies in the field of mobile communications: the IETF and 3GPP. The authors vision about the potential evolution of these solutions is presented, where they discuss the future evolution of IP mobility management architectures. In [61], the authors discuss motivations and requirements of distributed mobility management, presenting the two different possible approaches: host-based and network-based DMM approaches. It is also carried a comprehensive comparison between existing IP mobility support protocols and the proposed DMM approaches. This work also provides a quantitative analysis regarding registration delay, registration signaling overhead and traffic intensity to a MA. Finally, it points out the challenges in the development of DMM protocols.

2.4.1 Network-based Approaches

One of the first published Internet-Drafts [62] of DMM working group, called Distributed Mobility Anchoring (DMA), proposed the distributed mobility anchoring, which was based on previously scientific publications [63, 64, 65]. This approach was initially proposed in [63, 64] in the initial phase of this Thesis, and before publishing the Dynamic Mobile IP Anchoring (DMIPA) paper [16]. Most existing IP mobility solutions are based on centralized anchoring principals (e.g. MIP and PMIP), where a mobility anchor maintains MN bindings updated, and data traffic is encapsulated between mobility anchor and current MN's AR. These approaches are usually implemented in centralized architectures, so both data and control planes need to be processed by the mobility anchor. Observing the trend of flattening mobile network evolution, the main idea of the proposed approach is the dynamic distribution of mobility functions through the ARs. The author's main idea is to distribute the mobility traffic management with dynamic user's traffic anchoring in Access Nodes (ANs). Therefore, it is introduced a new entity called Mobility capable Access Router (MAR), which is an AR with mobility management functions. When the MN moves to a new MAR, it updates location in the previous MAR. Therefore, current CNs continue sending data traffic to previous MAR that encapsulates packets to new MAR. However, the MN can initiate and maintain new data sessions through the new MAR. The distribution of mobility functions can also be applied in the context of multiple-interfaces terminals with IP flow mobility, but there is no specification in these scenarios.

In [66], it is approached the problem of centralized mobility anchoring in the mobility solutions. It is analyzed the routing and concluded that a centralized anchor is often longer and leads to non-optimal routes that affect performance. Furthermore, mobility management regarding signaling was studied. It was deduced that, if it is implemented with a centralized strategy in recent flattened mobile networks, it becomes non-optimal,

specially if we consider that content servers are moving closer to the access network. So, the author proposes the distribution of mobility functions. It is addressed both partially distributed, where only the data plane is distributed. The idea is to allow a MN in any of theses networks to be served closest by a Mobility Function, bringing mobility anchors closer to the user. Besides distribution of mobility anchor, another main objective of this work is the dynamic activation/deactivation of mobility protocol support. The author improves the proposed ideas in a distributed mobility management framework [67, 68], where it is defined the main functionalities of mobility management. The three main mobility functions are home address allocation, routing management and location management. Based on such functional decomposition, the proposed framework is able to clearly separate data and control plane functionality, and provides the flexibility in an implementation to position logical functions at their most appropriate places in the system design. The proposed work also demonstrates that these logical functions can indeed perform the same functions as the majority of the existing mobility protocols. The author argues that these functions become the foundation for a unified framework upon which different designs of distributed mobility management may be built upon. The proposed approach was partially evaluated in [69] with the distribution of the routing management functionality at the gateways. The scheme performance has been investigated through simulation considering network load and latency. The performance results of packet delivery latency obtained show the benefits of optimizing the route under varying network load conditions and increasing distance from the MN home network to the visited network.

The remainder network-based approaches presented next are based on the concepts already introduced by [62] and [68]. There are minor differences between the approaches regarding messages, entities and classification, but the main concepts behind the approaches are quite similar.

The authors in [70] propose a network-based DMM, where the starting point is making PMIPv6 working in a distributed manner. In the proposed approach, mobility is handled by the network without the MNs involvement, but, differently from PMIP, when the MN moves from one access network to another, it also changes anchor router, hence requiring signaling between the anchors to retrieve the MN's previous location(s). Also, a key-aspect of network-based DMM, is that a prefix pool belongs exclusively to each Mobility Anchor and Access Routers (MAAR), in the sense that those prefixes are assigned by the MAAR to the MNs attached to it, and they are routable at that MAAR. In the proposed approach the authors consider two main schemes to design the DMM solution: a partially distributed scheme, where the data plane is only distributed among access routers similar to MAGs, whereas the control plane is kept centralized towards a cardinal node used as information store, but relieved from any route management and MN's data forwarding task; a fully distributed scheme, where both data and control planes are distributed among the access routers. The authors detail in [71] the introduction of the proposed network-based DMM approach in 3GPP network design. They present an evolution of current 3GPP architecture towards a flat and fully distributed mobility network design. This architecture allows pushing the data anchors towards the edge, alleviating hence the overloaded network core infrastructures of mobile operators. The proposed solution follows the distributed mobility management paradigm, which has been so far mainly discussed at the IETF, but takes into consideration the 3GPP architecture specifics. A new logical entity, called distributed gateway, is located close to the users, anchoring the data communications and supporting mobility when they move to a different gateway. Following the same idea, the authors also propose in [72] a solution that falls in this category, defining a new logical entity,

called Distributed Gateway (D-GW). It basically encompasses the functionalities of plain IPv6 access router, MAG and LMA, on a per-IPv6 prefix basis. The main contribution of this draft is more related with the definition of the mechanisms required to support the operation of such a network-based mobility solution when several flows are simultaneously anchored at different D-GWs, by introducing the concept of Distributed Logical Interface (DLIF). The document also defines the required PMIPv6 signaling extensions. Last, but not least, the solution is also extended to provide session continuity across different domains.

In [73], it is proposed a DMM approach, called enhanced Proxy Mobile IPv6 (ePMIP), which introduces two new PMIP-based logic functions, already described on [74]. The Location Management Function (LMF) maintains the mappings between IP addresses and location information of MNs. Distributed Anchoring Function (DAF) includes Distributed Routing sub-Function (DRF) which enables optimized routing between the MN and its CN, and Distributed Mobility sub-Function (DMF) which guarantees MN's mobility with minimal packet loss when optimized routing is established. DAF can be deployed through PMIP specifying the MAG to constitute an eMAG, and specifying the LMA to constitute an eLMA. The authors reused the mechanism in [75] for delivering mobility management property of IPv6 prefix by the distributed anchor to the MN for distributed mobility management. When distributed anchor detects an initial attachment of a MN, it will send a RA message to that MN. The RA includes IPv6 prefixes, and each prefix is tagged with its properties, which includes its mobility management property. According to the mobility management property, the IPv6 prefix can be distinguished by the MN into two categories: global prefix and local prefix. Global prefix provides IP addresses with fully mobility support that remain valid even when the point of attachment is changed, while local prefix provides IP addresses with no mobility support that are not valid when the point of attachment is changed. Mechanisms used for maintaining mobility for those global IPv6 prefixes are based on [74], including three aspects: initial attachment, data forwarding and handover. In the initial attachment, the distributed anchor sends a RA message with one or more local and global prefixes, which can be distinguished by the MN according to their mobility management property. Distributed anchor only updates the location information to the LMF for those global prefixes. In data forwarding, the MN's applications requiring mobility support will ask an IPv6 address from the global prefix. When the distributed anchor of the CN receives traffic sent to that global address, it will query the LMF for the location information of that global address and forward the traffic based on the location information. In the handover the MN changes its distributed anchor; thus, the new distributed anchor will advertise the same global prefix to the MN on the new link, and update the new location information for that global prefix to the LMF for the purpose of maintaining the reachability of the MN's global prefix.

The work in [76] discusses the 3GPP Evolved Packet Core (EPC) as a deployment of a distributed data plane architecture. Distributed GWs in connection with user mobility may result in the need to optimize data routing by GW relocation. It is proposed to include the GWs itself in the decision on GW relocation, which is currently done in the control plane node (mobility management entity) only. Proposals are made to allow the relocation also for active mode devices (user activity detection), to detect the occurrence of a non optimal routing situation, and to detect situations where a relocation should be suppressed to avoid particularly poor user experience. Solutions discuss how the EPC GWs can decide on a GW relocation or support the MME in such a decision. Fundamental information like user activity or the used services are only available in the data path, and hence, can be provided by the GW only. This allows mitigation of poor user experience

that may result from traffic offload features when data path updates are made by the GW relocation. In addition, the GW can learn the network topology in a self configuration manner by using routing information and avoid additional management effort of providing topology information to the MME. The impact of this on standardization is low: only new cause codes for PDN connection release messages have to be defined for the PGW/SGW to MME interface.

Under the current distributed anchor schemes, when the MN moves to a new anchor, the former flow is forwarded by the former anchor to the new anchor. Therefore, the long routing problem still exists. In the existing work of PMIPv6, the communication between LMAs is not involved since it assumes that the mobility management function should be accomplished under the centralized architecture. However, in the distributed architecture aforementioned, each distributed anchor could be regarded as the LMA for the communication session generated on it. Therefore, it is necessary to have communications between distributed anchors, but current schemes in PMIPv6 do not support such scenario. In [77], the authors propose a routing optimization scheme in distributed anchor scenario, based on PMIPv6 to solve the non-optimization routing problem in the current work. They propose two solutions to optimize the routing. The first solution is the Direct Mode, which means that the routing optimization is set up between the MN's D-MAG and CN's D-MAG by exchanging messages between the two entities directly. The second solution is the Relay Mode, which means that the routing optimization between the MN's D-MAG and CN's D-MAG shall be set up under the assistance of a third D-MAG. There are two stages in the proposed protocol. The first stage is the initiation of the routing optimization, and the second stage is the maintenance of the routing optimization. The initiation is the setup procedure of the routing optimization when the MN moves to a new anchor from the first anchor. The maintenance stage is the maintenance of routing optimization when the MN moves from the previous anchor to the new anchor after the setup of routing optimization.

The network-based distributed mobility management solutions are able to improve the performance of the traffic delivery to the users and reduce the consumption of the network resources. However they require the presence of the mobility anchor in every AR of the network or in nodes close to every CN of the network; otherwise they need to use the LMA of the standard PMIP as the mobility anchor. Most of these network-based distributed mobility management solutions keep the control signaling and mobility context management in a centralized mobility anchor, while the data plane is distributed through the ARs. Although some of the mobility management functionalities (e.g. routing/forwarding and IP address allocation) improve the mobility performance when provided by the network, there are other tasks that are easier implemented and completed, when provided by the MN. For example, the management of mobility support for multihomed devices (e.g. multiple interfaces) and ongoing sessions (e.g. IP addresses management) are mechanisms that can improve scalability and reduce network resources consumption when the MN became part of the process. The host-based mobility model always implies to provide part of the control of mobility management to the user, which network/internet providers do not look favorably. However, both providers and users can have advantages from this shared control schemes.

2.4.2 Host-based Approaches

The adaptation of the DMA [63, 64, 62] approach to a host-based scheme is proposed in [78], where the MN is responsible to register the configured address to the Access Mobility Anchor (AMA) through the BU message. When the MN registers by sending a BU

message at the new access network, it registers not only the newly configured address, but also informs the previous address to the serving AMA. As the serving AMA obtains the previous address information from the BU message, it sends a BU message to the origin AMA that results in updating the MN's mobility context and routing status and establishing a bidirectional tunnel with the origin AMA. From the performance analysis, it has been confirmed that the proposed DMM support approach eliminates the limitations of MIPv6 while taking advantages of the current evolution of mobile network architectures. This work was published at nearly the same time of the first DMIPA publication [16]. Although this approach is host-based, it does not provide any solution when the current AR of the MN does not provide mobility support; thus it requires that all ARs are configured as AMAs. One of the main differences between the host-based approach and the network-based DMA approach is the maintenance of the mobility context in the MN, which eliminates the necessity of a centralized server for this purpose; however the management of the mobility context is not explained in the paper. It also does not provide a description on the protocol operation for multihomed MNs (e.g. multiple interfaces), nor any specifications for a unified mobility management approach that takes into account all interfaces of the MN.

There is a proposal for a distributed mobility management protocol [79], called Distributed Mobile IPv6 (DMIPv6), which is compatible with the standard MIPv6 protocol. DMIPv6 introduces Distributed Home-Proxy (DHP) and Distributed Home Address (DHoA) for a MN, while it also maintains the HAs and HoAs already proposed by MIPv6. MN will use DMIPv6 if the DHP and DHoA are available; otherwise the standard MIPv6 is used. The deployment of the DMIPv6 could be implemented step by step, with the compatibility to the existing MIPv6. Compared to the standard MIPv6 in the management model, DMIPv6 could select different DHPs for a MN's different service flows. The MN takes different management strategies for different service flows according to network conditions and the actual requirements during the move. The introduction of DHP not only reduces the home network congestion and HA load, but also greatly reduces the possible failures in home network and HA, and the bad impacts to the MN. Besides, the MN could achieve optimized transmission path and transmission delay even choosing bidirectional tunnel, because the DHP is located close to the CN. The introduction of DHP makes it possible to enhance the mobility support for the clients without any updates. The authors also developed a mobility API extension for DMM in [80]. This API proposes an extension to [81], which would add more prefix classes so that an application could select prefixes with properties that are important for distributed mobility management. The proposed approach is able to provide the optimizing routing path at the cost of long tunnels between the MN and the DHPs close to the CNs. Hence, it enables the tunnels from the beginning of the sessions without dynamic mobility mechanisms; thus, even when the MN initiates and terminates the session in the same network, the mobility support and tunneling are provided. The proposed approach requires the availability of DHPs close to the CNs; otherwise it adopts the centralized HA from the standard MIPv6, which introduces all the problems already explained in centralized IP mobility management. However, this approach might bring advantages for highly mobile users always requiring most of their content from a small and well defined set of CNs (e.g. Youtube and Facebook). This approach does not provide any solutions for a distributed mobility management for MNs with multiple interfaces.

The solution proposed in [82] utilizes HAs located near CNs (Corresponding Home Agent, CHA) to dynamically allocate a HoA to the MN (Corresponding Home Address, CHoA). Such an address will be used throughout the IP session between the MN and

the CN. Given the topological proximity of the CHA to the direct path between the MN and the CN, it is expected that this solution would not have the negative side effects of providing IP session continuity. CHA may be co-located with the CN (the original content server or the CDN server), or located in the same site as the CN (e.g., on the load balancer, or a dedicated node), or located in an ISP serving the CN site. Not all CNs may be served by a CHA. In case there is no CHA serving the CN, the MN and the CN may communicate using the HoA via the HA. It is expected that CHAs would be deployed for access to heavily-used content on the Internet (e.g., YouTube, Facebook, Netflix, etc.). The CHA deployment is beneficial to the Mobile Network Operator (MNO) of the MN, as the operator offloads the mobility management and data transmission via its core network, and enhances the user experience through transmission latency reduction. The solution does not cover support for IP address reachability. However, this is considered to be acceptable, because only a very small set of applications really need IP address reachability. Those are the applications that are running as servers, which cannot avoid using standard MIP, since they need to accept incoming connections at a specific/published IP address. This solution is quite interesting for some scenarios to optimize the routing path between the CN and the MN; thus, the anchoring model provided close to the CN was considered as one of the possibilities in the context-aware adaptive IP mobility anchoring publication (Paper F) [22]. Furthermore, this DMM internet-draft [82] was elaborated including inputs and comments from the author of this Thesis, as acknowledged in the draft.

2.4.3 Assistance Mechanisms

There are some mechanisms for mobility anchoring and addresses management/selection proposed in DMM, which can be merged with both network and host based DMM approaches. These mechanisms are presented next, since they are interesting proposals to improve the mobility management performance.

In DMM environment, the MN always has more than one IP address to communicate with other ends. There is no problem for the MN to initiate a new IP session with any other CN by using the MN's latest obtained IP address. However, when the CN is initiating an IP session with the MN, the CN does not know how to choose the MN's IP address and which one to choose. In [83], it is proposed two solutions to find and select MN's IP addresses, one is a DDNS-based [84] solution, where the MN registers its new IP address to DDNS server, and the CN obtains the MN's new IP address info from DDNS server. The other is Server Register-based solution, where the MN and CN both register their new IP addresses and ports information to the same server for a given service. There are three methods for the CN to obtain the MN's IP address information and initiate an IP session to the MN, which are peer-to-peer (P2P) mode, server central mode, and combined mode. In P2P mode, the CN directly initiates a new IP session to the MN with the help of the retrieved information from server. In Server central mode, the CN initiates a new IP session to the MN, which has to pass through the server. In Combined mode, for control plane, the CN initiates the connection to the MN by Server central mode, while for data plane, the CN initiates the IP session to the MN by P2P mode.

Mobility anchor selection was introduced in DMM, through the study of use-case scenarios [23]. The work proposed is focused on highlighting the problem of anchoring selection in the scope of DMM. The authors mainly propose an initial anchoring of the sessions to a certain MA and the respective IP address, based on history/statistic about MN, applications and network. The examples are mostly based on the flows duration and mobility pattern of MNs (highly mobile, typical locations or static). This idea also requires that applications provide some input from the sessions to be established. This idea introduces

novel functionalities in the application layer, which can be achieved through APIs. The author of this Thesis is also one of the authors of the proposed DMM draft [23].

It should be noted that in reality not every application may need IP address reachability or IP session continuity. Usually, a client application does not need IP address reachability, and IP session continuity is not required for all types of applications either. Centralized mobility model forces the mobile host's IP traffic to traverse a centrally-located router, which incurs additional transmission latency and network resources, decreasing the reliability of the network with the introduction of a single point of failure. Furthermore, even when an application needs session continuity, it may be able to satisfy that need by using mobility support from mobility solutions in higher layers of the OSI model (e.g. MSCTP). However, this mobility support can be inhibited or ignored by the centralized mobility model, which always force all data traffic to be routed by the centralized mobility anchors (e.g. HA and LMA). Thus, it is proposed a solution where the applications running on the MN can indicate whether they need IP session continuity or IP address reachability [85]. The IP stack on the MN, in conjunction with the network, would provide the required type of IP service. It is for the benefit of both the users and the network operators not to engage an extra level of service unless it is absolutely necessary. So it is expected that applications and networks compliant with this specification would utilize this solution to use network resources more efficiently.

2.4.4 Summary

This section introduces the novel trend of distributed mobility management, which is being supported by the IETF Distributed Mobility Management (DMM) working group. There are novel approaches in this scope of DMM that aim to adapt existing IP mobility protocols, such as MIPv6 and PMIP to the emerging flat network architectures and distribution of the content servers. Most of the proposals aim to confine part of the mobility support functionalities at the access level (e.g. ARs), keeping the rest of the network unaware of the mobility events.

Most of the proposals dynamically activate the mobility support only when it is needed, such as when the MN actually undergoes into an IP handover. New sessions are anchored at the current AR of the MN and initiated using its current IP address. Hence, data packets related to the new sessions are routed via the optimized routing path between the MN and the CN, without tunneling, until the MN undergoes into an IP handover. When the MN undergoes into an IP handover before a session ends, the data related to this session is forwarded to/from the new AR of the MN through tunneling mechanisms. The dynamic mobility support is just provided from the time that a session undergoes into an IP handover; otherwise it is initiated and terminated through the optimized routing path without any mobility support nor any tunneling mechanisms. There is also work on proposing different distributed mobility anchoring models and IP address selection/management, as well as introducing on-demand and selective mobility management according to the application requirements and user profile.

Most of the new DMM proposals were published as long as the work of this Thesis was developed or even after the achieved contributions. DMM proposals already introduce novel concepts and mechanisms that can be applied to the mobility management in future networks, but there are several enhancements to be done, in order to have a standardized distributed mobility management architecture and to prove that it improves the quality experienced by the user, as well as it optimizes the network resources.

Chapter 3

Decentralizing Mobility Management

This chapter describes the work developed on the mobility management decentralization from the scientific publications achieved in the scope of the Thesis. Thus, the chapter provides an overview of the work published in the set of achieved scientific publications, highlighting the purpose and the key ideas behind them, as well as the main contributions and conclusions. Section 3.1 initiates the chapter by understanding and analyzing the current mobility management, in which it is provided the analysis of mobility in user-centric scenarios, as well as the characterization of mobility management in well-defined functional blocks, their interactions, and potential groupings. Section 3.2 presents the studies performed on the decentralized mobility management, in which several approaches distribute the mobility management functionalities through the network in different ways. Theses studies are supported by evaluations through analytical models and simulations, which help us to understand the guidelines for novel IP mobility approaches. Section 3.3 describes novel approaches and mechanisms towards a decentralized mobility management, such as a novel distributed IP mobility approach for flatten network architectures, called DMIPA, which is extensively evaluated through analytical models, simulations and testbed experiments. The proposed DMIPA approach is also integrated with seamless horizontal handover mechanisms, as well as evaluated in vehicular environments. The DMIPA mobility mechanisms are also adapted for multihomed scenarios, in order to provide data offloading with IP mobility from cellular to other access networks. It also presents a study on a novel network-based strategy for localized mobility, in which a replication binding system is integrated in the mobility anchors distributed through the access routers and gateways, being compared with centralized and distributed mobility management. The section concludes with a general perspective that goes further in the mobility anchoring subject, since it presents a context-aware adaptive IP mobility anchoring approach that assigns the MA which provides the optimized routing path and the shortest tunnel to a session, based on the context about the user, the network and the sessions.

3.1 Understanding Mobility Management

This section aims to understand and characterize current mobility management when applied to the recent network trends, such as the user-centric scenarios. As the first step, an analysis on the mobility management for user-centric scenarios is performed in [11] and [12]. The mobility management analysis is applied in user-centric wireless environments, which today correspond to the majority of technical scenarios on the last hop towards the

end-user. Our user-centric environments are located within the customer premises region (where residential households, and enterprise environments reside), while current mobility management relies on functionalities that are on the access or service regions. Hence, the most popular solutions for global mobility management have in common a model where a centralized and static MA is responsible for keeping some form of association between previous and current identities of a MN. There is the need to better understand the roles that a MA can have; the best location for these elements; and efficient ways to select the best MA for a MN. This initial analysis of aspects that have to be considered when attempting to make end-to-end mobility management schemes more flexible and adapted to the recent network trends was performed in [12], contributing to an out-of-the-box notion of mobility management. The mobility management is split into concrete functional blocks, explaining their impact and how to group such blocks. The decoupling of the mobility management is based on the centralized model, which independently of the OSI Layer of the solution, is based in the same principles, roles, and operational behavior. Such splitting and categorization can guide new flexible and user-centric mobility management architectures.

3.1.1 User-centric Scenarios

Out of the several possible user-centric scenarios, we highlight three: a regular hotspot, a User-provided Network (UPN) and a Delay Tolerant Network (DTN). Each scenario is described both from an architectural perspective, as well as from a mobility characterization perspective. The line of thought driving this analysis is that these representative scenarios hold different requirements and are based on specific mobility assumptions. Hence, we provide a mobility characterization for each of the scenarios. A more complete description of these and additional user-centric scenarios can be found in [11] and the annexed paper A [13].

- **Hotspot**: a hotspot scenario corresponds to the regular infrastructure mode in Wireless Fidelity (WiFi) environments. This is currently the most common wireless architecture being deployed around us: each Internet enabled household corresponds to one hotspot. In this scenario, mobility of users is local and confined to small regions, e.g. a room, an apartment, a small office and a mall. Moreover, if the user moves across different APs, then connectivity may be intermittent.

- **User-provided networks**: UPNs [86] have been applied as complement to existing access networks: they allow expansion of infrastructures across one wireless hop. There is usually one individual or entity, called Micro-Provider, which is responsible for sharing its connection with other users from the same community (e.g. FON [87]).

- **Delay tolerant networks**: The DTN scenario relates to the need to establish on-the-fly an autonomous network within a disaster region or a sparse scenario (e.g. after an earthquake) based upon the devices that users in the region control and carry. Hence, such DTN consists of a network composed by users with a common objective (a community), grouped in regions. Some nodes move from region to region, establishing the communication between them.

Table 3.1 summarizes the main characteristics related to the scenarios described, concerning inherent characteristics, and mobility behavior of the users on each of the scenarios presented. There is a detailed analysis of the scenarios in [11] and [12], in which a set of parameters were taken into account when characterizing any mobility management scheme:

i) *identification*, which stands for the device identification both from a user and an access perspective; ii) *network scope*, which relates to the reach of the network; iii) *access control*, which relates to the location of the access control mechanism that is normally applied in each scenario; iv) *mobility patterns*, related to the pattern that nodes are expected to exhibit in each scenario when roaming; v) *pause time*, related to the time that a node exhibits a speed that is zero or close to zero; vi) *handover frequency*, related to the node having to switch between different networks or attachment points; vii) *connectivity sharing*, related to the sharing of Internet access.

Table 3.1: Mobility analysis in user-centric scenarios.

Scenario/ Parameters	Hotspot	UPN	DTN
Identification	MAC address, credentials managed by WISP	Trust management scheme, community credentials	Tokens or certificates; public/private key pair
Network scope	Small environment, e.g. household shops, universities	Small-large, e.g. household to village/city; varies dynamically	Small-large but static does not exhibit a quick growth
Access control	Centralized, on the provider	Decentralized and spontaneous	Decentralized
Node speed	Low	High	Varying
Expected movement frequency	Low	High and global	Low and routine based
Mobility pattern	Local mobility; preferred locations	Human/social patterns; short distance traveling preferred	Local mobility; social patterns
Pause time	Long	Mix, depends on location and user routine	Long
Handover frequency	Low	High	High
Connectivity sharing	None	Yes	Yes

3.1.2 Characterizing Mobility Management

Mobility management is a key aspect to consider in future network architectures, as these architectures include a highly nomadic end-user which often relies on services provided by multi-access networks. In contrast, today's mobility management solutions were designed having in mind simpler scenarios and requirements from the network, and where roaming could often be taken care of with previously established agreements. To assist in understanding such requirements and also how to deal with them, it is proposed in [12] a starting point to dismantle current mobility management notions. This is an initial proposal on defining mobility management in concrete functional blocks, their interaction, as well as a potential grouping, which later assists in deriving novel and flexible mobility management architectures.

In a mobility management system, three elements are considered in the related literature: the MN, an end-user device for which a mobility service is provided; a MA, the

element responsible for providing the mobility management service, it may reside in a server of the network; and the CN, that is any element engaged in active communication with the MN. These are generic roles that are today present in different management solutions, independently of the OSI Layer where the solution resides. For instance, in MIP [5] the MA is the HA, while in SIP [29] the MA is the SIP server. In a 3GPP architecture, the MA is centralized and located in the core network, having all traffic flowing through it, even if services to be used are geographically closer to the MN. The positioning of the MA, as well as the definition of interactions between the different roles of mobility management have been object of heavy analysis. Still, today there is not truly consensus in where MA and additional functionalities should reside. Such positioning depends on the network architecture and requirements; on the OSI Layer being tackled, as well as on the overall complexity from a technical and policing perspective. In order to perform a mobility management characterization, as a result of an initial analysis on current available mobility management approaches and standards, we have identified the following mobility management functionalities:

- **Device identification:** corresponds to the network identification for the MN. Usually the main mechanism for a location management is the association between the device's *known-address* and the device's *real-address*. In MIP, known-address and real-address are IP addresses; in SIP, the known-address is a URI, and the real-address is an IP address. In MIP the device identification control is the HA/CN binding caches. In SIP, it is the user database used by the Proxy server.

- **Identification database control:** corresponds to the mechanism that is applied to control the database identification. This is normally a block relevant from an access perspective, which today follows a centralized approach.

- **Mobility anchor location:** corresponds to the element responsible for supporting all the processes that assist in identifying the MN at any instant in time, as well as processes that support communication to and from the MN.

- **Binding mechanism:** it is the signaling related to the device's registration to the mobility system. It creates/updates a record in the identification database control, associating the known-address to the real-address. In MIP it is the BU message sent to a HA/CN. In SIP it is the REGISTER message sent to the Registrar server.

- **Routing or forwarding:** it is the process of intercepting the packets destined to the known-address, encapsulating them with the real-address, and forwarding them. In MIP this is performed by the HA; in SIP this process is performed by an element named RTP translator (when it is used).

- **Handover negotiation:** the process taken when the device has its real-address changed. It involves negotiation and signaling. The main objective is to guarantee that the user will keep active all its sessions during the handover process. In MIP, the handover negotiation may be anticipated with the Fast Handover extension [53], and the SIP does not implement any anticipation, performing a re-negotiation after the connection between the peers is lost.

- **Resource management:** the resource management is a necessary procedure for the mobility management to guarantee the quality of the connection when the MN changes its point of attachment to the network. However, it is not provided by most of the mobility management approaches. The 802.21 MIH [28, 35] is focused on the

handover process based on a resource management aware negotiation for vertical handovers.

- **Mobility estimation:** it is the procedure of changing the MN point of attachment to the network before its current connection breaks. The extension F̈ast Handovers for MIPänd the MIH provide this functionality.
- **Security/Privacy:** it refers to any security or privacy mechanism used to assure the integrity of the elements and signaling in the mobility management system.

Mobility management usually is mentioned as consisting of two main blocks: location management and handover management. Location management is the block responsible for locating the devices, i.e. for guaranteeing that they are always reachable, independent of their point of attachment to the network. The handover management block is responsible for maintaining active sessions while MNs roam. Therefore, from a high level perspective, mobility management functionality can be split into these two main blocks. Today, these blocks both reside on the mobility anchor point and are based on information provided by the MN. Solutions such as the HIP [41] attempt to provide a decoupling by isolating location management and handover management. Other solutions (e.g. HMIPv6 [6]) optimize handover management by scoping the extent of the impact of such negotiation. Another way to categorize mobility management functionality is to consider a splitting between control and data planes. As part of the control plane we can cite all the procedures related to the signaling, and the data plane is related to the data traffic, routing, forwarding and address translation. Figure 3.1 shows the interactions between the blocks, in order to identify the communication between them. It also shows the classification concerning data and control planes, as well as location and handover management.

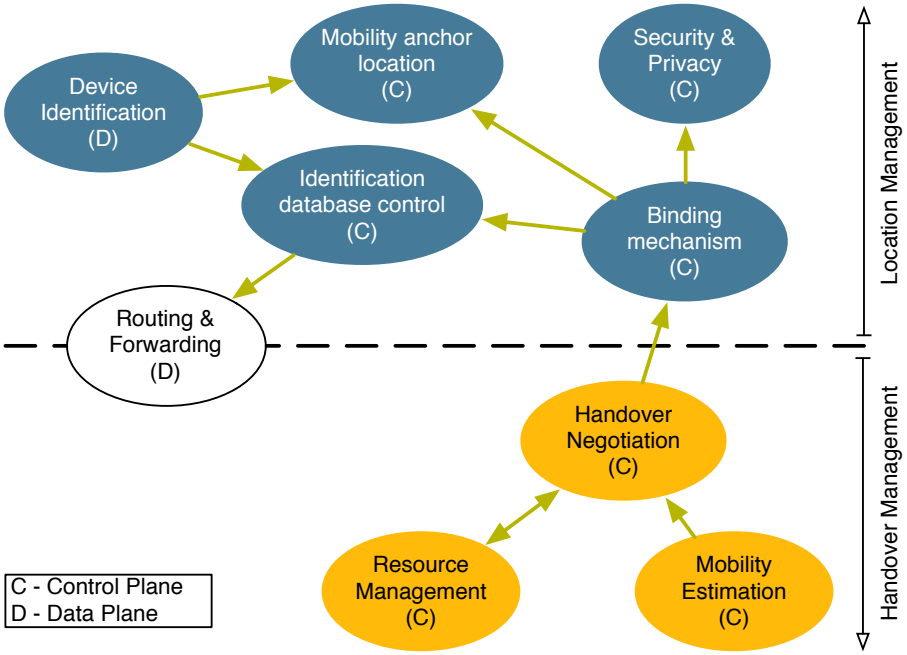


Figure 3.1: Characterization of mobility management into functional blocks.

Usually, all the communication between the blocks of the handover management side of the picture is triggered when a node movement is detected, or predicted. When a handover is detected, the mobility estimation block triggers the handover negotiation, which will take

part in the process. The handover negotiation needs to consult the resource management in order to guarantee that the user will be "always best connected" [88]. For the handover process to complete, the binding mechanism is triggered, so it can update the location information in the identification database control. The identification database control then updates the information in the element responsible for routing/forwarding. The binding mechanism has a periodic communication with the identification database control, because it is the procedure performed to maintain the identification database control updated. It needs to use the security/privacy procedures to guarantee that no third part could take place in the communication. The manner that the decoupled mobility functionalities are currently provided by mobility management approaches of different OSI layers was deeply analyzed, which is resumed in Table 3.2.

Table 3.2: Mobility functional blocks in current solutions.

Parameter	2G/3G	WiMAX	SIP	MIP	HIP	MIH
Device Identification	SIMCard (unique, tied to subscription)	IP address (unique, tied to interface)	SIP URI (unique, tied to user)	IP	Locator ID	MIHF ID
Identification database control	Centralized, controlled by the provider; access through the MA	Centralized, controlled by the provider; access through the MA	Centralized, controlled by the provider; access through the MA	Mobility anchor point	Mobility anchor point	Does not implement an identifier database
Mobility Anchor location	Centralized, located in the provider	Centralized, located in the provider (ASN-GW)	Centralized, located in the provider (Proxy SIP)	Centralized, located in the provider (HA)	Centralized, located in the provider	not implemented
Binding mechanism	REGISTER message, MN to Registrar Server or Outbound Proxy	MIP based and only for vertical handovers	REGISTER message, MN to Registrar Server or Outbound Proxy	Periodic Binding Update message, MN to HA, MAP or CN	Centralized solution, located in the provider premises	Does not implement location management
Routing / Forwarding	Proxy or RTP translator	Point-to-point, Proxy based	Proxy or RTP translator	IP based, regular routing	Dual, based on locator and on IP	Does not implement location management
Handover negotiation	Break-before-make, Register required	Break-before-make, Register required	Break-before-make RE-INVITE message, MN to CN	Make-before-break, with FMIP access routers negotiation	Make-before-break	Make-before-break, attachment points negotiation
Resource management	None	None	None	None	None	Yes
Security/ privacy	Yes, based on 3GPP system requirements	Yes, for the channel established between MN and BS	None	None	Yes, inherent to HIP	Yes, in a security related extension
Mobility estimation	Static roaming (pre-established tunnel)	No	No	Yes (e.g. FMIP)	No	Yes

3.1.3 Summary

The previous work characterized current mobility management when applied to the recent network trends. It introduced an analysis on the mobility management for user-centric

scenarios, which are today the majority of scenarios on the last hop towards the user. The most popular solutions for global mobility management have in common a model where a centralized and static MA is responsible for keeping some form of association between previous and current identities of an MN. Hence, it is presented an initial analysis of aspects that have to be considered when attempting to make end-to-end mobility management schemes more flexible and adapted to the recent network trends, contributing to an out-of-the-box notion of mobility management. It is provided a split of the mobility management into concrete functional blocks, explaining their impact and possible ways to group them. The proposed splitting and categorization can guide new flexible and user-centric mobility management architectures. From the analysis, most of the mobility solutions disregard the resource management, mobility estimation and security/privacy, as well as deploy centralized models for mobility anchor location, routing/forwarding and identification databases control. These centralized models suffer from limitations and issues related with scalability (e.g. network bottlenecks and single point of failure), security (e.g. attacks focused on the centralized mobility anchor) and performance (e.g. non-optimized routing). The following section compliments the performed mobility management analysis, studying different schemes to distribute the mobility management functional blocks, specially the mobility anchor location, the identification database control, the routing/forwarding and the binding mechanism.

3.2 Studying Decentralized Mobility Management

After the characterization of the mobility management into well defined functional blocks and potential grouping them, it is important to evaluate the distribution of the functionality blocks through the network. In order to understand the advantages and drawbacks of distributing the mobility functionalities, the work in this Thesis addresses the decentralization of mobility functionalities in [13], [14] and the annexed paper B [15].

3.2.1 Decoupling Mobility Management into Control and Data Planes

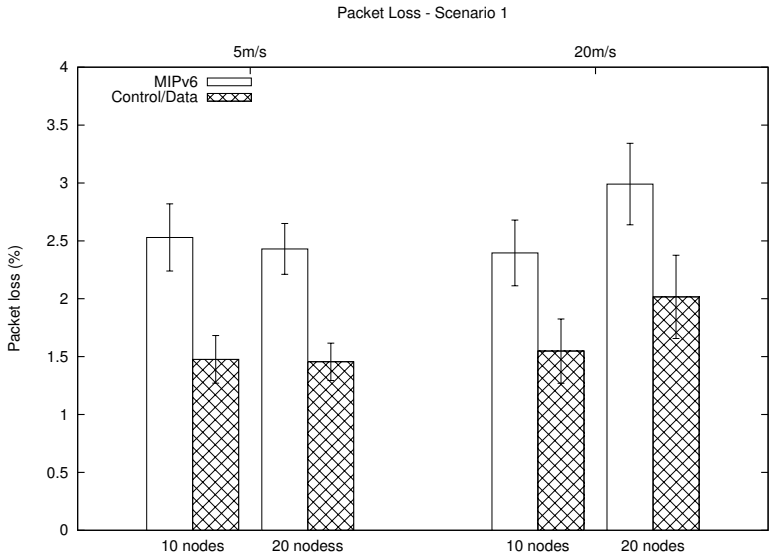
The first study is performed on the decoupling of HA from MIPv6 into Home Agent Data (HA-D) and Home Agent Control (HA-C), which is described and evaluated in [13]. According to the potential way to split data and control planes taking into account the performed characterization, it is considered two different mobility roles placed in two different mobility elements: Mobility Anchor Point Control (MAP_C) and Mobility Anchor Point Data (MAP_D).

- MAP_C**:** is the element responsible for the signaling plane and maintenance of binding cache entries. It is the HA-C that controls the signaling e.g., choice of HA-Ds to consider, or triggers the activation of tunnels between specific HA-D elements. This work considers one HA-C as example, but covers also the notion of several HA-Cs and several HA-Ds involved in the signaling.

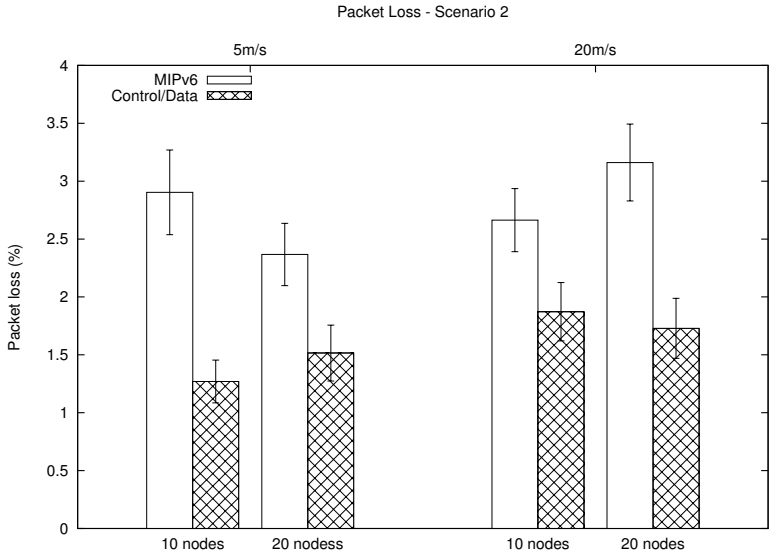
- MAP_D**:** is the element responsible for actions such as encapsulation of data traffic to the new location of a MN. The HA-D role is to activate and deactivate tunnels as well as to encapsulate traffic, whenever an MN performs a handover, and no mobility context status is kept in HA-Ds.

When decoupling control and data planes, the MN shall only exchange signaling information with the HA-C. In the proposal, this signaling corresponds to BU messages which contain information about the current IP addresses (CoA) and original IP address (HoA)

of the MN. A partially distributed approach is implemented, where several MAP_Ds are deployed and controlled by one single MAP_C. The MAP_Ds are placed as close as possible to the MNs, since they are responsible for data forwarding. The evaluation is performed in Network Simulator 2 (NS2) [89] through hierarchical and ring topologies. More information about the simulated scenarios is described in [13].



(a) Hierarchical topology

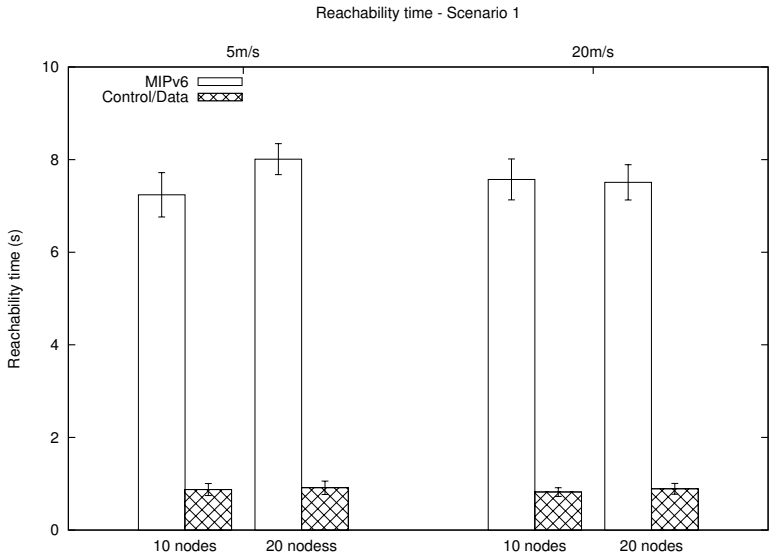


(b) Ring topology

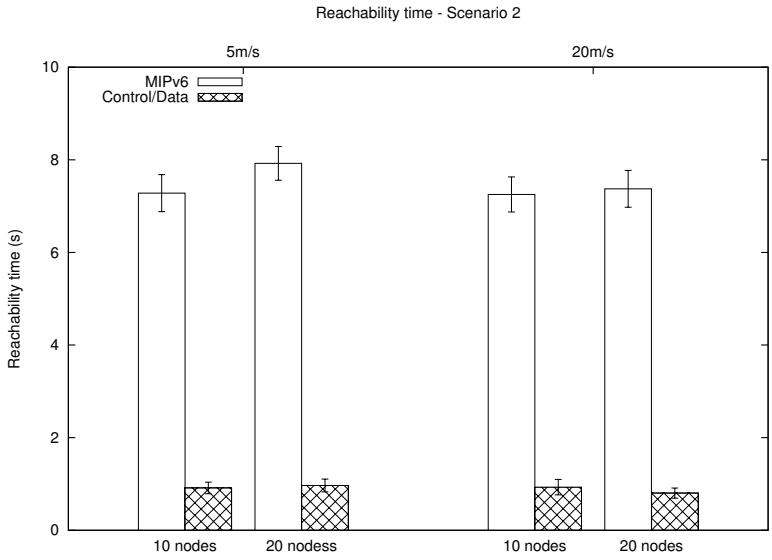
Figure 3.2: Packet Loss.

The proposed approach (Figure 3.2) is less sensitive, regarding packet loss, to the increase in traffic and nodes on the network than MIPv6, across all scenarios. In MIPv6 there is no tunnel between the old and the new location of a MN; thus, the MN will remain unreachable until it finishes the CN binding procedure. Therefore, all packets sent from the CN to the MN will be lost, thus resulting in a higher packet loss.

There is again a significant difference in the time that a node remains unreachable after a handover (Figure 3.3). The MIPv6 approach reachability time varies from 7 to 8



(a) Hierarchical topology



(b) Ring topology

Figure 3.3: Reachability.

seconds, while the reachability time for the proposed approach is lower than 1 second. This significant difference can be justified again by the creation of tunnels between old MAP_D and new MAP_C, so that the time interval a MN is unreachable because it is still performing CN binding update procedure gets considerably reduced. The higher reachability times required by MIPv6 can be justified by the process the MN takes to update its information to the CN.

3.2.2 Decoupling Mobility Management into Location, Handover and Data

A second study on the distribution of the mobility functional blocks went further in the decoupling of mobility management, which splits the mobility management into location

management, handover management and data management. This grouping of mobility management functionalities is initially proposed and evaluated through simulations in [14], which was latter improved, in the annexed paper B [15], through a detailed explanation of the distributed mobility approaches, as well as an exhaustive evaluation using analytical models and simulations. The grouping of mobility management functionalities is the following:

- **Location Management:** maintains the reachability of MNs, independently of their location (network), integrating the identification database that maintains the bindings with MNs' Identification and their IP addresses. It answers to CNs about the current IP address of a MN in the sessions establishment phase, which is updated with the current IP address of a MN through the handover management.

- **Data Management:** is responsible for the encapsulation of data packets through address translation. The data management functionality intercepts data packets, decapsulates them if needed, and then encapsulates them with a new IP header from the address translation rules. Data management does not provide any signaling, since it just receives signaling messages from the handover management.

- **Handover Management:** maintains sessions active when a MN roams between networks, so it provides handover detection and negotiation at IP layer, being responsible for the signaling with data and location management. Handover management also maintains the mobility context and routes of the MN, such as the set of routers serving the MN with the data management and the respective IP addresses of the MN. When an IP handover occurs, the handover management is responsible to detect it, and update the data management with the new MN mobility routes, and update the location management with the new IP address of the MN.

Based on the proposed decoupling of mobility management functionalities, three approaches to distributed data management are compared with centralized mobility management model (e.g. MIP), in order to understand their impact on both the user and the network. In these distributed data management approaches, the handover management is distributed through the MNs, while the location management is maintained centralized in the MA. The combinations of distributing the data management through the ARs and the CNs is performed, placing the data management in the ARs, the CNs and both.

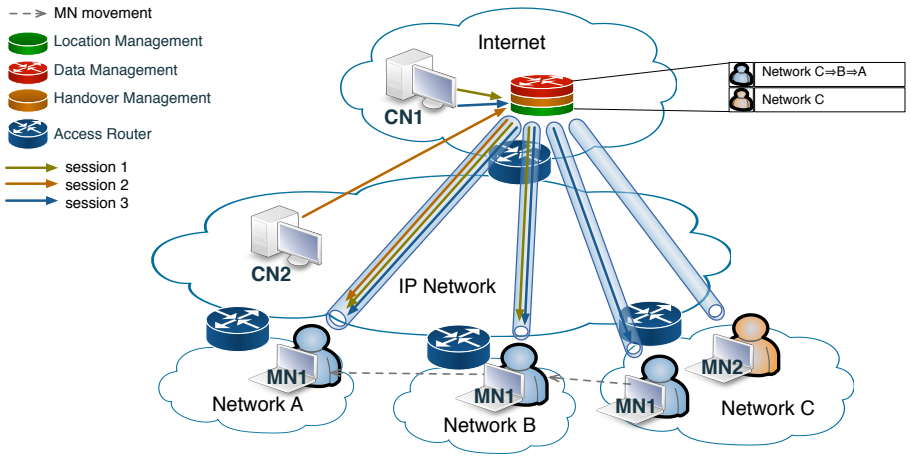


Figure 3.4: Centralized Mobility Management.

Centralized Mobility Management (A): there is an unique element, called MA, responsible for the management of data, handover and location (Figure 3.4). The sessions are always routed via the centralized MA and tunneled to the current IP of the MN. Thus, the end-point of the tunnel is updated in the MA with the IP address obtained from the current MN network. In the centralized model, the MN performs the handover detection and signals the MA with the new IP address, but the handover negotiation and the management of mobility context (e.g. bindings and routes) are maintained centralized in the MA.

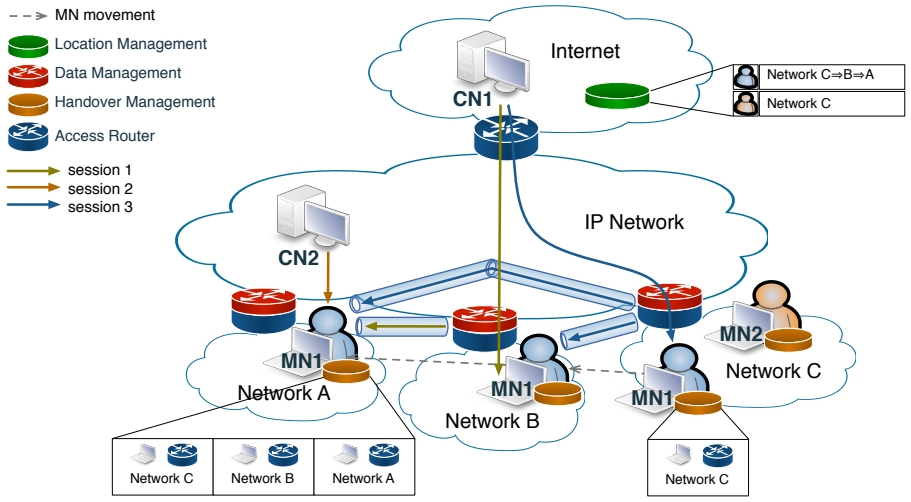


Figure 3.5: Data Management in ARs.

Data Management in ARs (B1): while the MN is attached to a network, the sessions are established and maintained through the current AR in the shortest routing path between the CN and the MN, without any mobility support (Figure 3.5). When the MN attaches to another network, it obtains a new IP address from the new AR, which is detected by the handover management. Then, the handover management establishes tunnels with the previous AR with anchored sessions (sessions initiated by the MN with IP address provided by the AR), in order to maintain the ongoing sessions active. The new sessions of MN are initiated through the current AR with the IP address received, without tunneling or mobility support. The handover management also updates the location management with the current IP address of the MN, received from the currently attached AR in each IP handover performed by the MN. The handover management of a MN maintains its set of ARs with anchored sessions, and the respective mobility context.

Data Management in CNs (B2): the data management is moved from a point close to the user to the sessions source node (CN). The sessions are initially established through the current AR of the MN in the optimized routing path between the CN and the MN (Figure 3.6). However, from the first time that a session suffers an IP handover, the session is maintained through an end-to-end tunnel between the CN and the MN. The handover management of the MN detects the attachment to a new IP network and the new IP address of the MN. Then, the handover management establishes tunnels with CNs involved in the ongoing sessions of the MN, and updates the location management with the new IP address of the MN. The handover management does not maintain the set of ARs with MN anchored sessions, but it maintains the set of CNs of the MN's ongoing sessions. This distribution of the data management in the CNs is able to optimize the routing path to the handover sessions at the cost of introducing longer tunnels and increasing the time

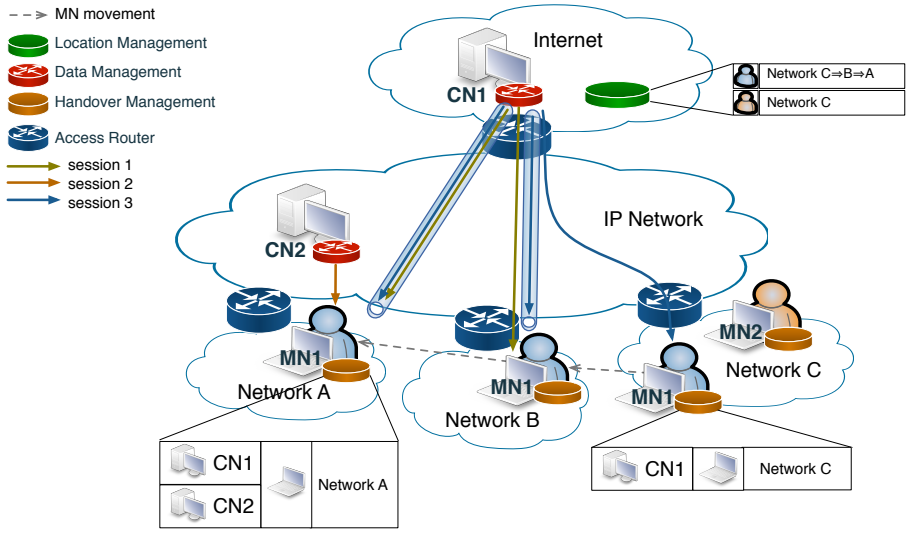


Figure 3.6: Data Management in CNs.

to update the data management.

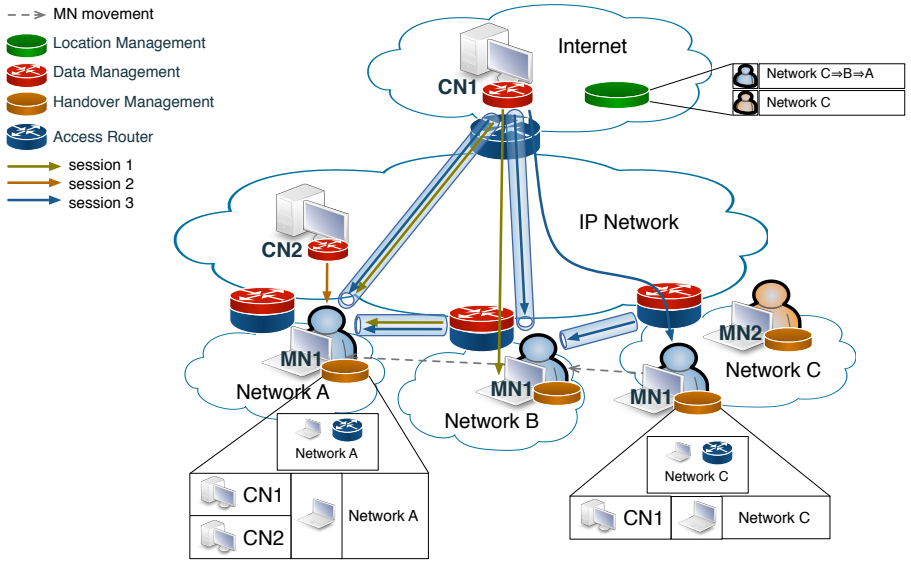


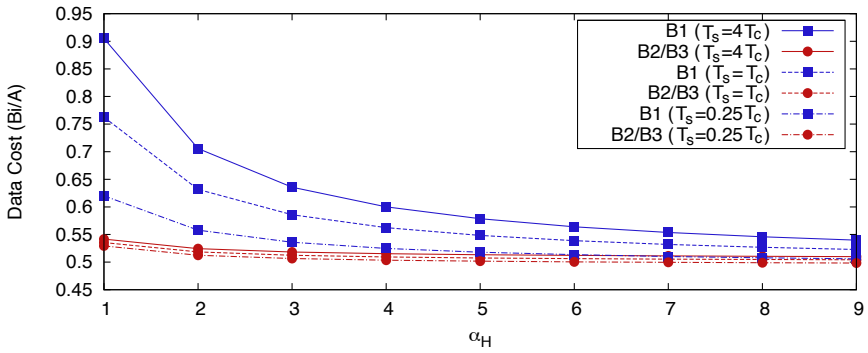
Figure 3.7: Data Management in ARs and CNs.

Data Management in ARs and CNs (B3): The third approach is a mixture of the two previous ones (B1 and B2), since the data management is distributed through the ARs and CNs (Figure 3.7). However, the data management in the ARs is just performed during short periods of time to reduce the handover latency, while the data management in the CNs is not updated. This scheme firstly exploit the MAs closer to the user (AR), in order to improve the handover latency for the ongoing sessions and to initiate the session through the current AR without any mobility support. Then, it exploits the MAs in the CNs to optimize the routing path of the handover sessions. Although there is an expected improvement in the handover latency and data delivery performance, the handover management of the MN has to maintain not only the MN mobility context of the CNs with ongoing sessions, but also the set of ARs with anchored sessions and the

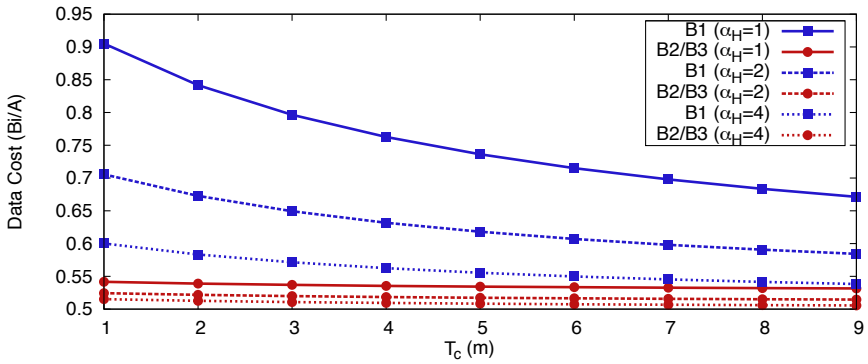
respective mobility context.

We developed analytical models to compare the distributed data management approaches with the centralized one, regarding data and signaling costs metrics. The analytical model is described in Paper B [15]. The distributed approaches ware evaluated according to the analytical model and simulations, for a set of different scenarios, in order to show the data and control performance evolution in comparison with centralized mobility management. The evaluation also includes a validation of the analytical model, based on a comparison with the simulation results for a specific scenario. T_s is the average session service time, T_c is the average cell-residence time and T_a is the average sessions inter-arrival time. α_H is the ratio of 1) the number of hops between nodes in different networks, to 2) the number hops between nodes in the same network: $\alpha_H = H_{HA-AR}/H_{AR-AR}$. The other parameters defined in the analytical evaluation are presented in the annexed Paper B [15].

Analytical Results



(a) Impact of α_H with T_s

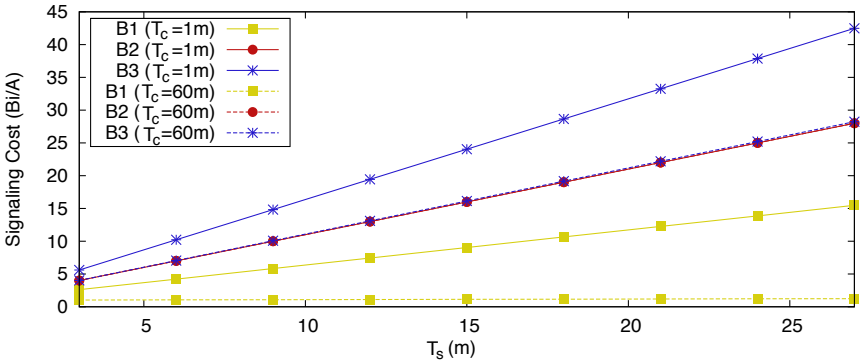


(b) Impact of T_c with α_H

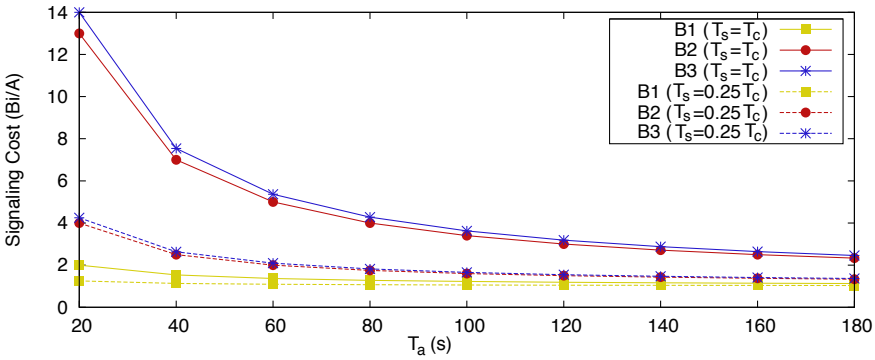
Figure 3.8: Data Cost

The value of α_H in Figure 3.8(a) shows a significant impact in the data cost for approach B1, specially for values of service time higher than the cell-residence time ($T_s > T_c$). However, the data costs of the distributed approaches B are always lower than the ones of the centralized approach for values of α_H higher than 1. For high values of α_H, the distributed approaches B converge to half of the data cost of the centralized approach A.

The influence of α_H and T_c are evaluated in Figure 3.8(b). The increase of the cell-residence time T_c decreases the data cost of approach B1 when compared with approach A, since the probability of a packet to be a handover packet reduces. We can observe a fast decrease of the data cost with the increase of T_c or the decrease of T_s for lower values of α_H.



(a) Impact of T_a with T_s



(b) Impact of T_s with T_c

Figure 3.9: Signaling Cost

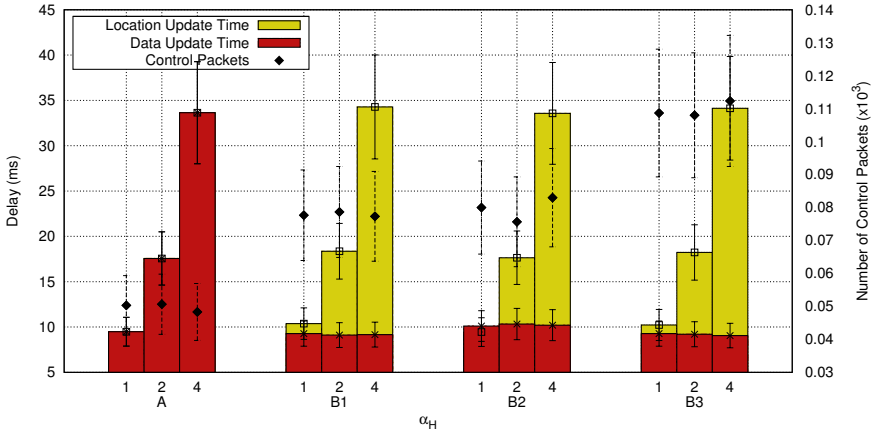
From the evaluation of the signaling cost in Figure 3.9, it is observed that the signaling costs of the distributed approaches B are usually higher than the centralized approach, since there are two separated signaling mechanisms: one for the location management that is maintained in the centralized MA, and another for the data management distributed through the ARs. In Figure 3.9(a), we observe that the increase of α_H presents a similar impact in the signaling cost of distributed and centralized approaches. However, the increase of T_s maintaining the value of T_c, strongly increases the signaling cost of B2 and B3.

The influence of T_s with α_H is evaluated in Figure 3.9(b). The increase of T_s has a strong impact in approaches B2 and B3, since there are more ongoing sessions when the MN roams, which means more CNs to update. However, the increase of α_H does not significantly affect the distributed approaches when compared with the centralized one, since both suffer the variation of α_H in a similar scale. The ratio T_s/T_c is crucial for the signaling cost, which is usually called Session to Mobility Ratio (SMR). When the ratio has a low value, the service time is much lower than the cell-residence time, and the signaling cost of the distributed approaches is close to the signaling cost of the centralized approach. However, as long as the ratio increases, the distributed approaches significantly

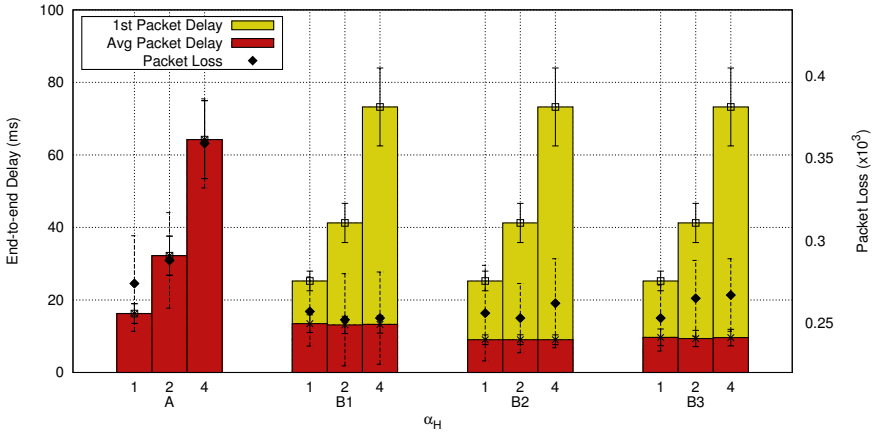
increase the signaling cost compared with the centralized approach, specially approaches B2 and B3.

The distribution of the mobility management functionalities overall increases the signaling cost, specially for approach B2 with the data management in the ARs and CNs, with the combination of the following factors: 1) a very mobile user, 2) high sessions inter-arrival mean rate and 3) long sessions service time. Another main contribution to the increase of the signaling cost of the distributed approaches is the maintenance of the location management centralized in the MA, which is updated every time a MN obtains a new IP address.

3.2.2.1 Simulation Results



(a) Signaling

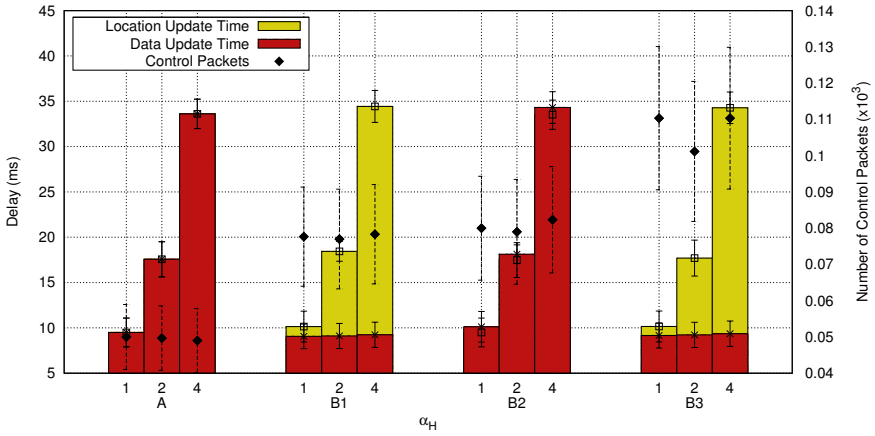


(b) Data

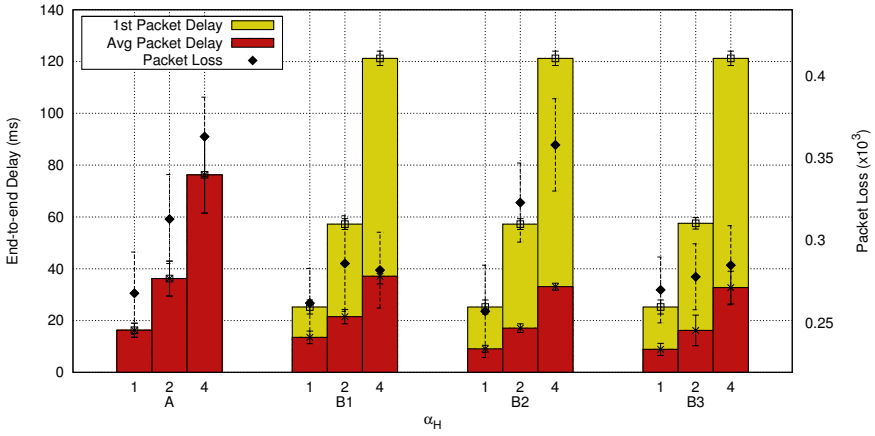
Figure 3.10: Impact of α_H in establishing session with a close CN (CN1)

The platform selected for the evaluation through simulation is the Network Simulator 3 (NS3), where the set-up scenario is configurable, and default configurations are presented in the annexed paper Paper B [15]. The metrics applied in the evaluation of the distributed approaches are the average packet delay, first data packet delay, packet loss, location update time, data update time and overhead. The average packet delay measures the average time that a data packet takes to be transmitted from the CN to the MN, while

the first packet delay measures the time that a CN needs for the location search (MN's IP address), plus the time to deliver the first data packet to the MN. The packet loss is considered the difference between the number of packets that were sent by the CN and the ones received by the MN, not distinguishing the cause of the packets loss, but the majority of the lost packets were due to the handover. The location update time is the average time that a MN needs to update its binding (IP address) in the location management system, while the data update time is the average time that a MN needs to update the mobility routes with the new IP address in the data management system.



(a) Signaling



(b) Data

Figure 3.11: Impact of α_H in establishing session with a far CN (CN2)

In Figure 3.10(b), it is evaluated the impact of α_H in the data delay and loss with sessions with CN1. The increase of α_H benefits the distributed approaches, since the average packet delay increases less than in the distributed approaches. However, the distributed approaches increase the delay to deliver the first data packet, since it is needed a location search mechanism to obtain the current IP address of the MN before the session establishment. The approaches B1 and B3 place the data management in the ARs close to the MN, thus the packet loss does not depend on α_H, only the hop distance among ARs. Thus, the data update time is lower for approach B1 and B3, as illustrated in Figure 3.10(a). The packet loss in B2 depends on the hops distance between an AR and the CN1, while in A it depends on the hop distance between an AR and the MA. Hence, the

increase of α_H increases the packet loss of A and B2 approaches. The location update time is similar for all the studied approaches, since all of them update the IP address of the MN in the centralized MA for location management. Regarding the overhead, the distributed approaches increase the number of control packets in the network, with a higher growth for the approach B3, which updates the data management of ARs and CNs each time the MN roams to another IP network. Considering the number of ARs and CNs used in the evaluation, the number of control packets doubles in the distributed approaches, and it may increase even more in scenarios with more ARs and CNs.

It is evaluated the data and signaling in Figure 3.11 with the impact of α_H in communications with a closer CN, the CN2. In communications with a closer CN, Figure 3.11(b), the distributed approaches reduce the average packet delay, since the data management is placed closer to the MN, in ARs and CN2. Moreover, α_H has no impact in the distributed approaches, since they do not depend on the distance of the centralized MA in communications with a closer CN. The time to deliver the first packet in the distributed approaches is higher than in the centralized approach A, since it is needed the location search signaling to obtain the current IP address of the MN from the centralized MA before the session establishment. The distributed approaches also decrease the packet loss when compared with the centralized approach A, since the time to update the data management system is lower, as observed in Figure 3.11(a). The location update time is similar for all approaches, since all of them place the location management in the centralized MA. The distributed approaches increase the number of control packets in the network, due to the separate signaling messages used to interact with location and data management.

3.2.3 Summary

This section presents the studies performed on the decentralized mobility management, such as the decoupling of IP mobility management into data and control planes, as well as into location management, handover management and data management. Based on the different groupings of functional blocks, we evaluate the distribution of the functionality blocks through the network, in order to understand the advantages and drawbacks of distributing the mobility management functionalities. There is the need to better understand the roles that a MA can have, the best location for these elements, and efficient ways to select the best MA for a MN.

An initial study on the problematic of decentralized mobility management proposes a pragmatic decoupling of the HA entity of MIPv6 into HA control and HA data, in order to separate the mobility management functionalities related with control and data planes. This approach was a first example to understand the trade-off associated to mobility management functionality decoupling. The outcome shows that the decoupling of mobility management can significantly decrease the packet loss and the time a MN remains unreachable during the binding update procedure.

Another study on the decentralized mobility management proposes the decoupling of mobility management into location, handover and data management. The handover management is distributed through the MNs, while location management is maintained centralized. The data management is distributed according to different strategies through ARs and CNs.

The outcome of the evaluation showed that it is difficult to conclude about the best distributed mobility approach in overall, since it depends on the considered scenario. However, the distributed approaches generally improve the mobility management performance, but the best distributed approach strongly depends on the topology of the network and the SMR. The inclusion of data management close to the CNs is advantageous for high

SMRs with more hierarchical networks, since it allows the delivery of sessions through end-to-end tunnels in the optimized routing paths between CNs and MNs. However, for some sessions, the CNs or close network elements may not provide mobility support, and the introduction of the data management in the ARs is a good solution to significantly reduce the data cost and average data packet delay, without compromising the signaling cost. The distribution of the data management through the network overall reduces the network cost to deliver traffic to the users and to improve the traffic delivery performance. The distribution of the handover management through the MNs, such as mobility context maintenance and handover signaling, reduces the signaling cost and handover latency, as well as eliminates the necessity of any bottleneck or single point of failure for mobility management. These studies were important to obtain the guidelines to develop a novel IP mobility management scheme and mechanisms for flat network architectures, which will be presented in the following section.

3.3 Developing Decentralized Mobility Management Solutions

After the study on the decoupling and distribution of the mobility management, this section proposes novel decentralized IP mobility approaches and mechanisms to cope with recent network trends of flatten network architectures, content servers close to the user, and data offloading from cellular to other access networks. This section presents the DMIPA proposal, its extension for seamless handovers and its feasibility in vehicular scenarios, the mechanisms for data offloading, replicated bindings, and the context-aware adaptive IP mobility anchoring.

3.3.1 Dynamic Mobile IP Anchoring (DMIPA)

This section starts by describing a distributed and dynamic IP mobility approach, called DMIPA, which was initially proposed in [16], being evaluated through analytical models in comparison with centralized MIPv6 approach. DMIPA was latter extended in [17], which provides a more detailed description of the proposed mechanisms, as well as an exhaustive evaluation through analytical models, simulations and testbed experiments. DMIPA approach and its evaluation are presented in the annexed paper C [17].

The mobility management deals with the session continuity of the ongoing sessions after a handover of the MN to other IP network, which incorporates two main functionalities addressed here:

- **IP data anchoring:** anchor data sessions at IP layer in the mobility anchor, which is in charge of forwarding these data sessions to the current location of the mobile device, which is done by packets encapsulation/decapsulation according to the address translation rules (e.g. IP tunnels).
- **IP mobility context management:** management of mobility anchors and IP addresses of MNs, as well as the bindings creation/update for routes maintenance through associations between previous and current IP addresses of MNs.

DMIPA distributes the IP mobility functionalities through MNs and Mobility-enabled Access Routers (MARs). While the IP data anchoring is distributed through the MARs, the IP mobility context management is distributed through the MNs, where each one is responsible for the management of its own mobility context.

DMIPA is a host-based distributed and dynamic IP mobility approach based on the following principals:

- **Distributed IP Mobility:** IP mobility functionalities are distributed through MNs and ARs with mobility support, called MARs.
- **Dynamic IP Mobility:** IP mobility support is given for ongoing sessions from the time that they endow an IP handover, otherwise the session are initiated/maintained without any mobility support.
- **Dynamic IP Anchoring:** in the base version of DMIPA, a new session is always anchored to the most recently attached MAR, while ongoing sessions remain anchored to the MARs where the sessions were initiated.
- **Global IP Mobility in Heterogeneous Networks:** when a MN attaches to a MAR, it ensures the forwarding of the ongoing sessions from previous MARs, otherwise these functionalities are provided by the MN.
- **No Centralized Entities:** MNs maintain their mobility context (e.g. set of MARs and bindings) without centralized databases.

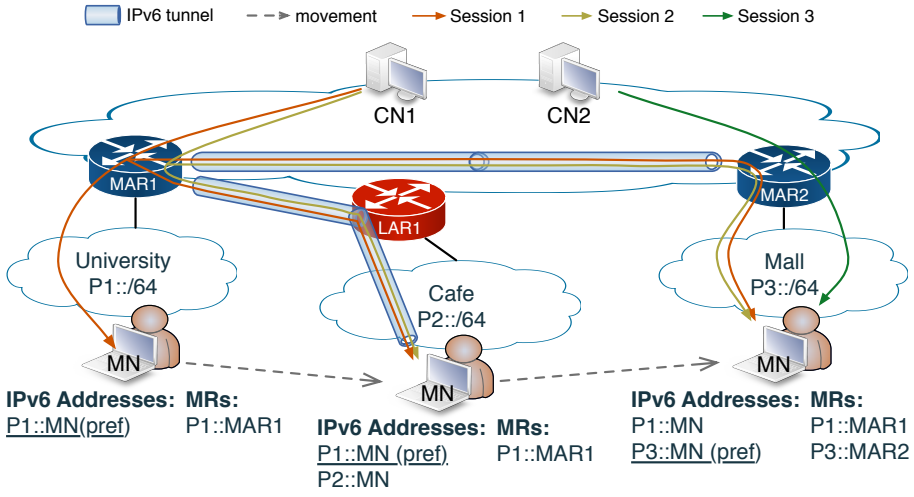


Figure 3.12: DMIPA scheme example.

The DMIPA scheme (Figure 3.12) provides two distinct modes of operation, in order to overcome the scenarios when a MN attaches to a MAR or when a MN attaches to a Legacy Access Router (LAR).

- **MAR:** From the IPv6 prefix and the mobility support indicator received from the attached MAR, the MN configures a new IPv6 address with preferred status, while other addresses received from previous MARs are just maintained for the ongoing sessions. the MN adds the attached MAR to its set of MARs, and sends this set of MARs and the respective MN addresses to the attached MAR. Then, the attached MAR establishes tunnels with previous MARs to maintain the ongoing sessions anchored there, while new sessions are established through the attached MAR without any mobility support.

- **LAR:** From the IPv6 prefix and the absence of mobility support indicator received from the AR, the MN configures a new IP address, and maintains the IPv6 addresses from previous MARs just to maintain ongoing sessions. The IPv6 address obtained from the most recently attached MAR is the preferred one for new sessions. the MN establishes tunnels with previous MARs to maintain the ongoing sessions, while new sessions are established through the tunnel with the most recently attached MAR.

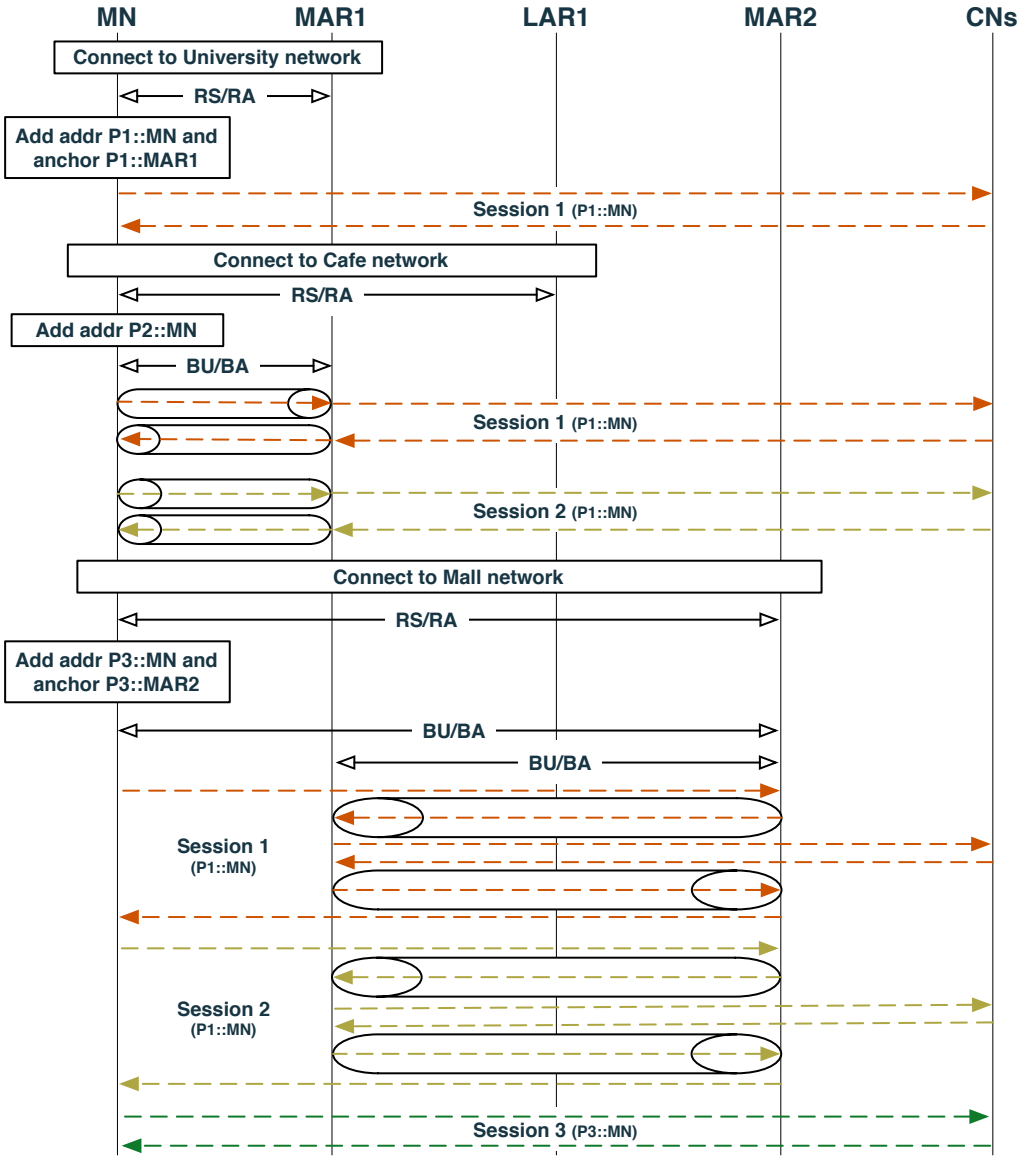


Figure 3.13: DMIPA operational example.

The protocol operation is explained with the example of Figure 3.12 and the support of the diagram of Figure 3.13.

University: After the MN connects to the University network, it exchanges RS and RA messages with MAR1, which provides the IPv6 address of the MAR1 interface and the mobility support indicator. MN configures the new IPv6 address *P1::MN*, based on the information received from the RA message and the MAC address of its interface, which is defined as the *preferred* IPv6 address to establish new sessions. the MN also adds the

MAR1 to its set of MARs, which is the new preferred MAR. MN initiates new sessions, such as Session 1, with the IPv6 address *P1::MN* through MAR1, without any tunneling or mobility support.

Cafe: After the MN connects to the Cafe network, it exchanges RS and RA messages with LAR1, which provides the IPv6 prefix of the network. MN configures the new IPv6 address *P2::MN*, based on the prefix received from the RA message and the MAC address of its interface. The new IPv6 address, based on the prefix received from a LAR, is configured in the *deprecated* state, while the IPv6 address *P1::MN* received from MAR1, remains the *preferred* one to establish new sessions with session continuity support. the MN establishes a tunnel with MAR1 through BU/BA messages, in order to maintain the ongoing sessions anchored to MAR1, as well as to be able to establish new sessions with session continuity support. Thus, packets to *P1::MN* received by MAR1 are tunneled to *P2::MN*. Session 1 is maintained through the configured tunnel between MN and MAR1, as well as the new Session 2 established while the MN is attached to the Cafe network.

Mall: After the MN connects to the Mall network, it exchanges RS and RA messages with MAR2, which provides the IPv6 address of the MAR2 interface and the mobility support indicator. The MN configures the new IPv6 address *P3::MN*, based on the IPv6 prefix received from MAR2 and the MAC address of its interface. The new IPv6 prefix received from a MAR is configured as the *preferred* one for the establishment of new sessions, while the IPv6 address *P1::MN* received from MAR1 is changed to *deprecated* state, just to provide continuity to sessions anchored to MAR1. The MN adds MAR2 to its set of MARs, which is from now on the preferred MAR to anchor new sessions. The MN exchanges BU/BA messages with the attached MAR2 to provide the set of MARs and respective MN addresses. After MAR2 receives this information, it establishes tunnels with previous MARs, thus, MAR2 establishes a tunnel with MAR1 to maintain the ongoing sessions 1 and 2, anchored in MAR1. Packets to *P1::MN* received by MAR1 are tunneled to MAR2 IP address *P3::MAR2*. Sessions 1 and 2 are maintained through the tunnel between MAR1 and MAR2, while new sessions, such as session 3, are established with the new IPv6 address *P3::MN* through MAR2, without any tunneling or mobility support.

The IPv6 addresses from LARs do not need to be maintained when the MN is not attached to them, since these IPv6 addresses are not used to anchor sessions. The IPv6 addresses from MARs are maintained in the MN, as long as the MN uses them for ongoing sessions; otherwise, these IPv6 addresses are removed, as well as the corresponding tunnels. In Figure 3.13, if sessions 1 and 2 are terminated in the end of the sequence, the IPv6 addresses *P1::MN* and MAR1 are removed from the MN interface, and the set of MARs, respectively, as well as the tunnel between MAR1 and MAR2.

We developed analytical models for both MIPv6 and DMIPA. The analytical models presented in appended Paper C [17] are based on the work of [16], where the analytical models were adjusted to fit the envisioned scenarios, and the tunneled packets metric was added. We developed the analytical models to evaluate tunneled packets, signaling cost and data cost. The metric tunneled packets measures the total percentage of data packets being tunneled by the MARs/HA. The signaling cost is the total cost needed for mobility management control messages, defined as the size of the messages multiplied by their remaining time in the network, between source and destination endpoints. The data cost is the end-to-end cost to deliver data packets from the CN to the MN, defined as the data packets size multiplied by the remaining time in the network, between the source and destination.

The evaluation compares DMIPA with MIPv6 through analytical and simulation re-

sults, adopting different metrics. The analytical results compare DMIPA with MIPv6 based on the signaling cost, data cost and tunneled packets for different scenarios, where the parameters of the model are changed. The analytical results are validated through a comparison with simulated results in a specific scenario, where both methods can be directly compared. The evaluation through simulations is performed in scenarios developed in the NS3 environment [90]. Besides the metrics evaluated through the analytical models, the simulation environment also evaluates the average data delay, the number of hops with tunneled packets, the average bindings per MA, and average MAs per MN. Finally, DMIPA is validated in a real environment through a testbed, where the packet delay, bitrate and tunneled packets are evaluated for UDP and TCP sessions. More details on the evaluation and parameters used in the analytical, simulation or testbed evaluations can be found in the annexed Paper C [17].

Analytical Results

DMIPA always presents a lower data cost, when compared with MIPv6 (Figure 3.14(a)). The DMIPA data cost becomes closer to the MIPv6 data cost for low cell-residence times (T_c) and high session service times (T_s), since the the percentage of tunneled packets increases and most of the sessions are tunneled nearly their entire lifetime, from the MAR where the MN was initially connected. Otherwise, if the MN is able to communicate without handover sessions, the data cost always follows the optimized path provided by the routing, reducing the data cost.

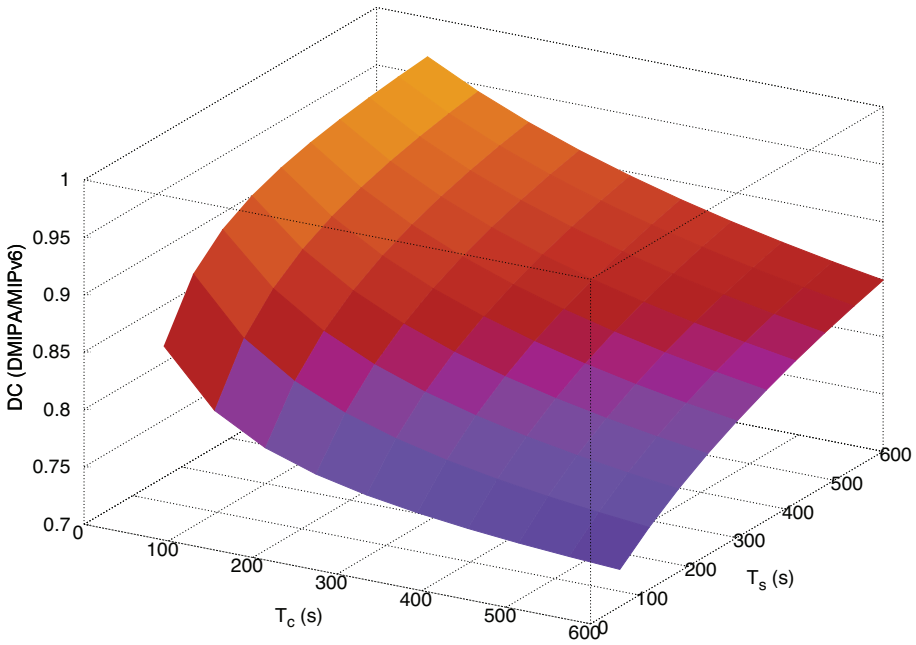
The signaling cost of DMIPA (Figure 3.14(b)) is higher than the signaling cost of MIPv6 for a higher number of active MARs per MN (N_m). Since $H_{a\to a} \approx H_{h\to a}$, DMIPA needs more messages to update the set of MARs of the MN. However, for quite static MNs, such as people that spend long periods at home and work, where the average number of MARs per instant (N_m) is less than one, the signaling cost of DMIPA is reduced when compared with MIPv6. Moreover, the N_m also depends on the SMR, wnich considers the relation between the ARs residence time (T_c) and the sessions service time (T_s). For flatten network architectures, DMIPA reduces the signaling cost when compared with MIPv6.

Simulation Results

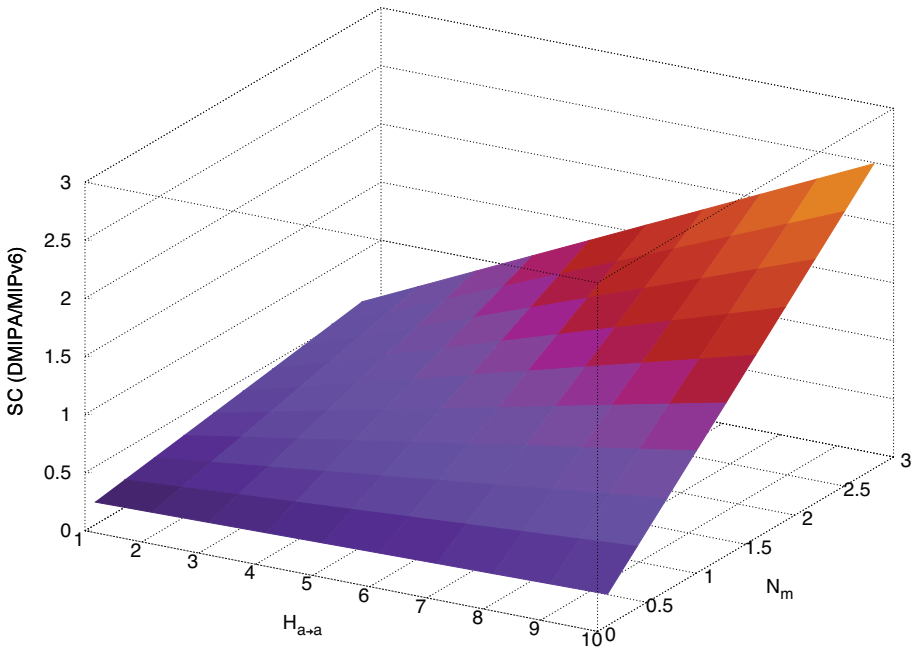
DIMPA reduces the data cost when compared with MIPv6 (Figure 3.15(a)), due to the distributed and dynamic data anchoring support provided. In DMIPA, the data cost decreases even more for longer MN pause times combined with shorter sessions service times, since it reduces the number of handover packets.

DMIPA reduces the signaling cost when compared with MIPv6 (Figure 3.15(b)) for longer pause times and shorter session service times, where the MN has to maintain less MARs per instant (less mobility management messages to be exchanged). The number of messages to be exchanged between the MN and MARs in DMIPA might be higher than in MIPv6, since the MN in DMIPA may have more than one associated MAR, while in MIPv6 the MN is just associated with the centralized HA. However, the signaling cost of DMIPA may be lower than the signaling cost of MIPv6, since the MN exchanges control messages with closer elements (MARs), reducing the time of the messages in the network.

DMIPA reduces the number of tunneled packets, when compared with MIPv6, Figure 3.15(c). In MIPv6, all data packets are tunneled by the HA to the current location of the MN, from the time that the sessions are established. DMIPA is able to reduce the tunneled packets, which is even lower for longer MNs pause times combined with shorter



(a) Data Cost

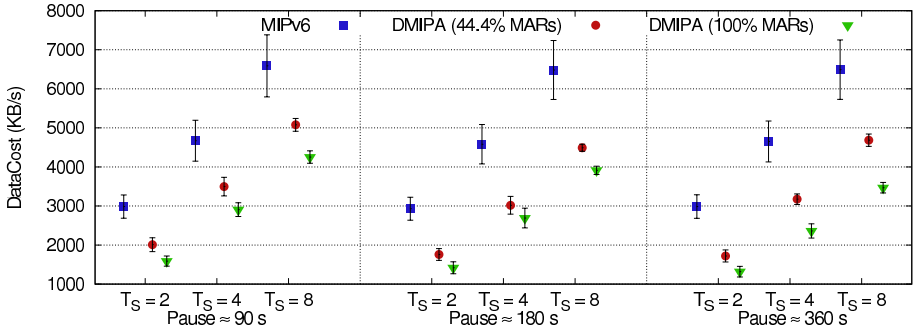


(b) Signaling Cost

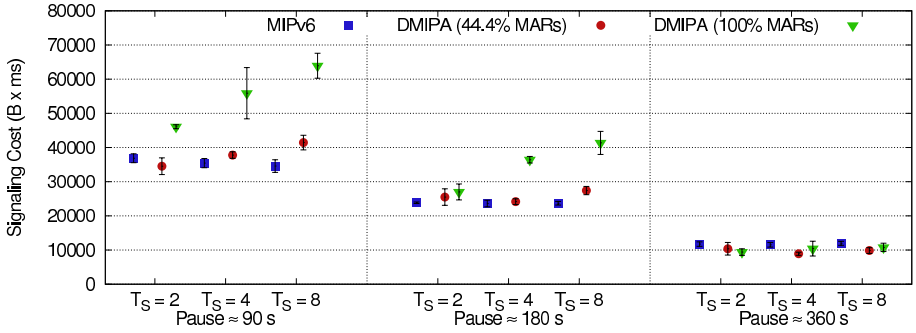
Figure 3.14: Analytical Results.

sessions service times, since it reduces the number/portion of the handover sessions that do not need tunneling.

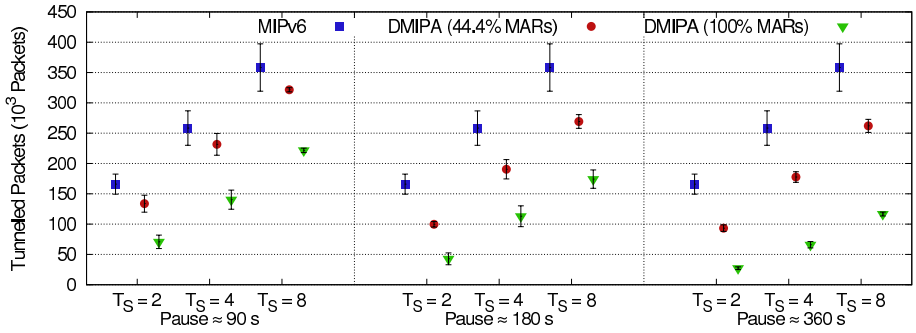
DMIPA reduces the average end-to-end delay when compared with MIPv6, Figure 3.15(d). In MIPv6, the delay is similar for all sessions, since data packets are always routed via the centralized HA, that tunnels these packets to the current location of the MN. In DMIPA, the delay is lower for shorter sessions combined with longer MN pause times, which reduces the number of handover sessions. The availability of more MARs



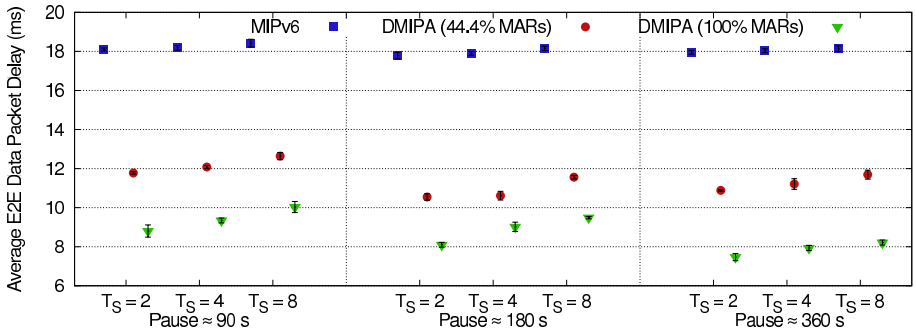
(a) Data Cost



(b) Signaling Cost



(c) Tunneled Packets



(d) Packet Delay

Figure 3.15: Simulation Results.

decreases the data delay, since more sessions are initiated without any mobility support through the optimized path.

Testbed Results

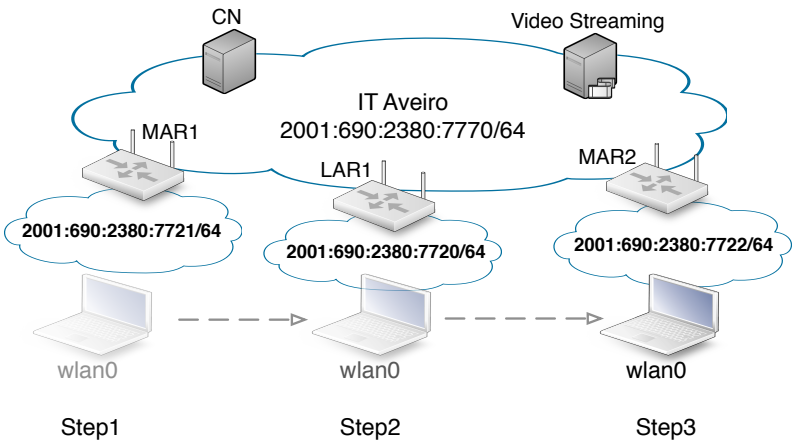
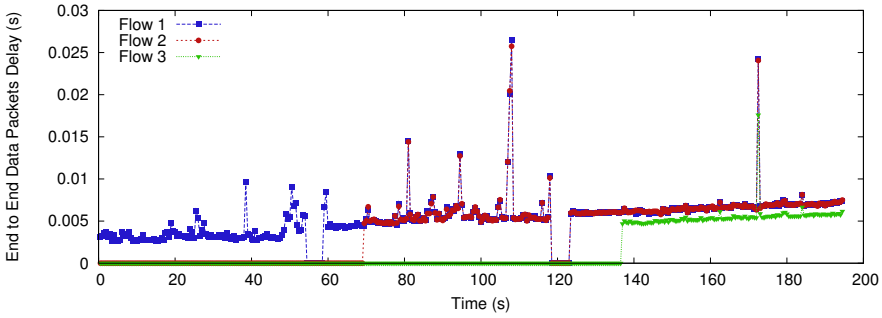
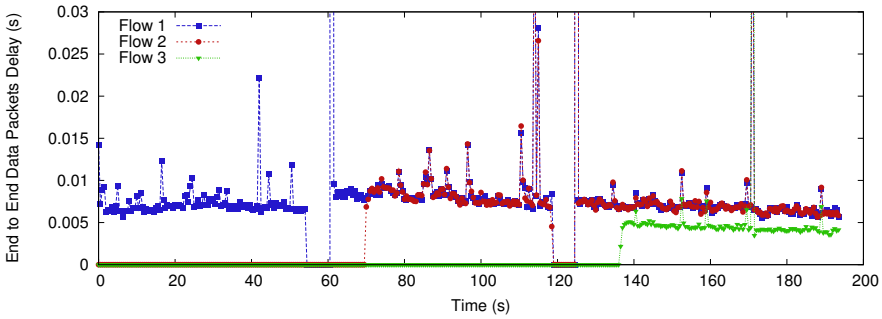


Figure 3.16: Testbed Scenario.

In the last part of the DMIPA evaluation, we perform a validation through the testbed, presented in Figure 3.16, where we configured three wireless routers as MARs and LARs. The MN moves from MAR1 to MAR2 through LAR1. The time that the MN is attached to an AR (MAR or LAR) represents a step in the scenario, where a new UDP/TCP session is initiated, maintaining the previous ones active. Thus, in step 3, the MN has three active sessions, anchored to the 2 available MARs. More details and results of the tested scenario are presented in annexed paper C [17].



(a) UDP



(b) TCP

Figure 3.17: Testbed Results for Data Packet Delay.

As long as new flows are initiated, the packet delay increases, especially due to the

wireless medium access, since we have more traffic being sent from CN to MN (Figure 3.17). The handovers performed at 60 and 120 seconds have a latency around 5 seconds, since there is no mechanism to provide seamless handover. The flow 3 starts around 140 seconds through the optimized path, while other ongoing flows (1 and 2) are maintained through a tunnel between MAR1 and MAR2. The packet delay introduced by the tunnel (encapsulation, forwarding and decapsulation) can be estimated if we compare flow 3 with other 2 flows, which is quite low.

3.3.1.1 DMIPA with seamless handovers

In the scope of DMIPA approach, it is addressed the support of seamless handovers from a distinct perspective, in order to reduce or even eliminate the handover latency, without new entities, signaling messages or even packets buffering/bicasting. We propose in [19] a make-without-break handover approach, which exploits the overlapping regions of APs through two logical connections from the MN during the handover. In several scenarios, the MN may be able to execute the handover to a new IP network, without breaking the previous connection, reducing or eliminating the packet loss. Thus, the ongoing traffic sessions might be maintained through a previous AP, while the signaling for the MN configuration in the new network is performed through the new AP.

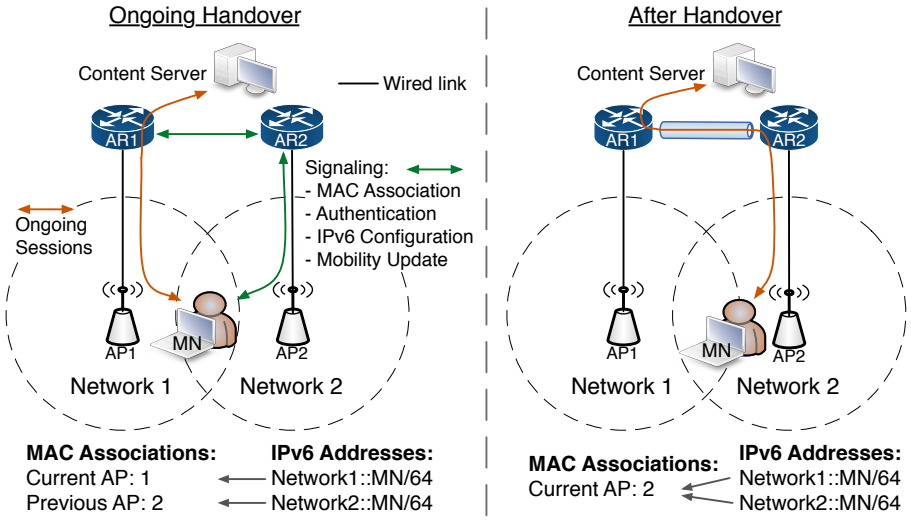
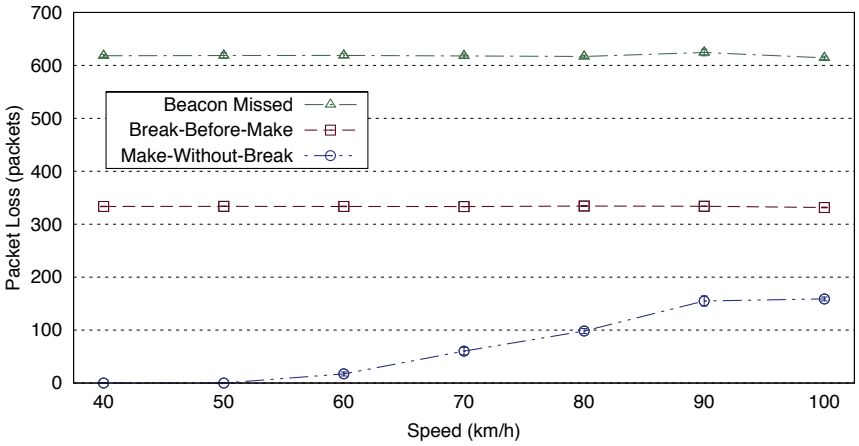


Figure 3.18: Make-without-break double logical connection.

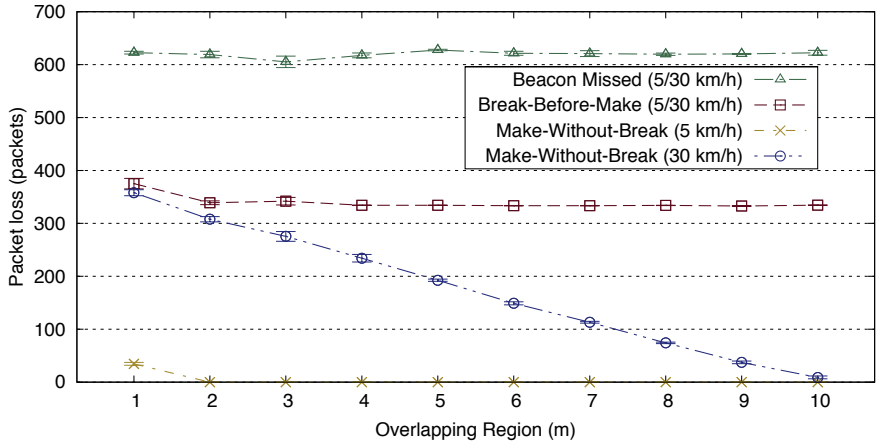
The proposed concept assumes (Figure 3.18), from the physical layer, that the handover is performed between two APs in the same channel, or that a MN physical layer is capable to provide a fast channel switching, or assumes the virtualization of the MN physical interface during the handover. The possibility to maintain the previous IPv6 addresses in DMIPA approach already ensures the continuity of the ongoing sessions at IP layer. In DMIPA, the MN is able to maintain the previous IPv6 address in the preferred status, as well as the respective route, while the new one is being configured, ensuring IPv6 connectivity during the handover execution. However, it is not performed any route/association at the MAC layer during the handover execution, since a new MAC association is usually performed assuming a disassociation from the previous AP. Make-without-break for DMIPA should be able to maintain two MAC associations or/and a simple selection for which packets should use which AP. In order to send packets from the MN to other devices (uplink) during the handover execution, it is maintained a default route to the previous

AP through the link-local IPv6 address of the AP, as well as the respective entry in the neighbor discovery cache. It is also needed a decision function at the MAC layer to enforce the AP (e.g. destination and BSSID MAC addresses in 802.11b/g) according to the IPv6 source address of the packet. From the time the mobility anchor (e.g. AR) is updated, the MN transparently removes the logical connection with the previous AP.

The different handover techniques provide different values for handover latency and packet loss, since they adopt different strategies for the handover. We analyze and compare the different handover strategies: break-before-make, make-before-break and make-without-break in order to understand their impact on handover latency and packet loss due to link disconnection. These strategies are evaluated in NS3 simulation environment, in order to obtain handover latency and packet loss. Handover latency is the time that a MN remains unreachable for the ongoing/new sessions, while roaming between APs from different ARs. A null handover latency represents a MN always reachable and always receiving the required contents, even when executing the handover. Packet loss is related with the packets lost due to the link disconnection, thus it is strictly associated with the handover latency, since no buffering mechanisms are used.



(a) Changing Speed



(b) Changing Overlapping Region

Figure 3.19: Evaluation of the handover techniques

The evaluation of the make-without-break (Figure 3.19) approach shows that it is able to reduce or even eliminate the handover latency and the packet loss during handover. The

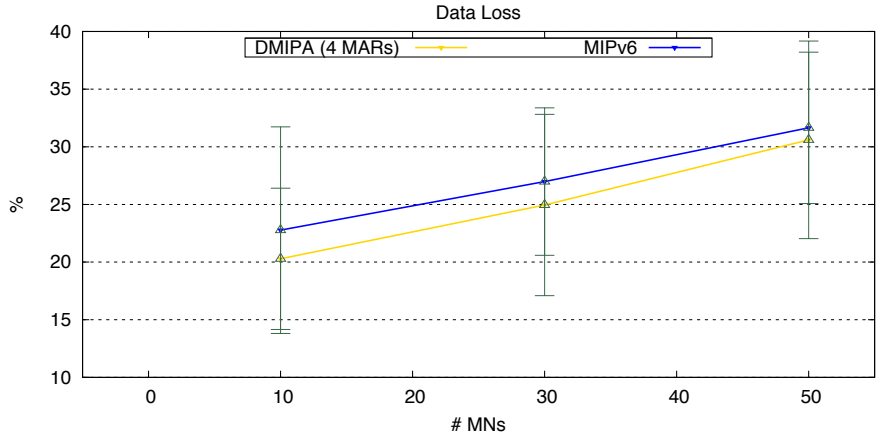
values obtained for make-without-break are enough to provide seamless session continuity for a large set of applications, including real-time applications, such as Voice over IP or Video on Demand. However, the proposed make-without-break approach requires modifications on both MN and AP. More detail on this approach and its results is presented in [19].

3.3.1.2 DMIPA in Vehicular Environments

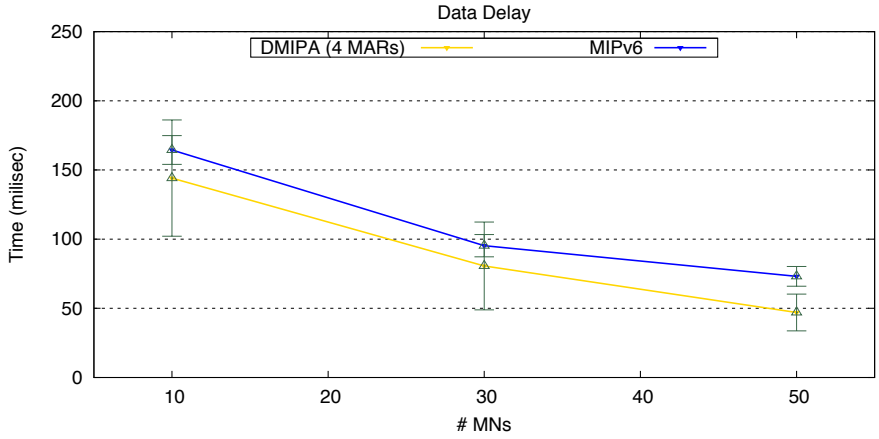
As demonstrated in the previous evaluations of DMIPA, this distributed approach is more prepared to cope with the evolution of mobile network architectures towards flatten architectural models. This way, distributed and dynamic mobility improves scalability and performance of mobility management in flatten networks; however, a distributed mobility approach was never evaluated in vehicular scenarios. Vehicular networks are growing as an important type of access networks. Thus, it is important to understand how distributed mobility models influence the mobility performance in vehicular environments. Hence, we define a set of vehicular scenarios to be evaluated under the centralized MIPv6 and the distributed DMIPA approach in [18].

In the evaluation performed through the NS3 simulator environment, it was considered two vehicular scenarios: Highway and City. In the Highway, the ARs are placed in a line, while in the City scenario they are placed in a grid. Therefore, the mobility patterns of the MN are generated through the Simulator of Urban Mobility (SUMO) [91], which is an open source road traffic simulation package, designed to handle large road networks. The Highway scenario consists of a simple road containing two lanes, in which the vehicles move in only one direction. The City scenario connects nodes with lines creating city blocks. More details on the considered scenarios, as well as a more extensive evaluation, are presented in [18]. We define several metrics to evaluate both MIPv6 and DMIPA in highway and city scenarios: data loss, data delay, BU time and signaling cost. Data loss is the percentage of lost data packets, defined as the total number of sent data packets subtracted by the total number of received data packets, divided by the total number of sent data packets and multiplied by 100. Data delay is the average data packet delay per session, where the average data packet delay per session is previously calculated from the average one way delay of the data packets from that session. BU time is the average time to update the bindings on the respective mobility anchors. Thus, in DMIPA, it is the average time to update the MARs, considering both cases of a MAR and a LAR. In MIPv6, it is the average time to update the HA. Signaling cost is the total cost introduced by the IP mobility to exchange the control messages. It is calculated through the multiplication of the size of the exchanged messages by the time the messages need to be delivered between source and destination nodes.

DMIPA improves the overall performance of mobility in vehicular environments when compared with the MIPv6, regarding the data loss, data delay, binding update time and signaling cost (Figure 3.20). Although DMIPA is a better solution than MIPv6 for IP mobility management in vehicular environments, these protocols face big challenges, due to the high mobility of MNs. Thus, we will continue to investigate the IP mobility in the scope of vehicular networks/environments, such as taking into account the vehicular's and environment behavior to provide the optimized IP mobility management. More detail in the evaluation results is shown in [18].



(a) Packet Loss



(b) Data Delay

Figure 3.20: City Scenario

3.3.2 DMIPA mechanisms for data offloading

The cellular networks have been overloaded with demanding mobile data; thus, mobile networks are adopting data offloading mechanisms through the wireless area networks. However, these data offloading mechanisms need to assure IP session continuity when the sessions are offloaded through the wireless area networks. The current IP offload model is based on a centralized anchor, such as MIPv6 and PMIP, which routes all data packets to one of the networks/interfaces where the user is connected, bringing scalability and performance issues. Oppositely to the current approaches for 3GPP, we propose to extend DMIPA mechanisms in order to provide data offloading with IP mobility support for heterogeneous environments, which are presented in Paper D [20]. The proposed DMIPA-based approach describes the set of mechanisms to provide dynamic offload anchoring, which properly manages IPv6 addresses, offload anchors, IP tunnels, routing rules and interfaces.

Oppositely to MIPv6, which defines a centralized mobility anchor (HA), the DMIPA-based offloading approach also uses the mobility anchors distributed at the access network level, called Access Home Agents (AHAs). We also propose that each MN manages its offload anchors (AHAs), IP tunnels and IP addresses, which completely eliminates the

necessity of a centralized node for offloading management purposes. There are two main mechanisms to offload data traffic: i) offload sessions between interfaces connected to the same AHA, in the same or different IPv6 prefixes. ii) offload sessions between interfaces connected to different access networks, through an interface connected to a AHA or Legacy Access Node (LAN).

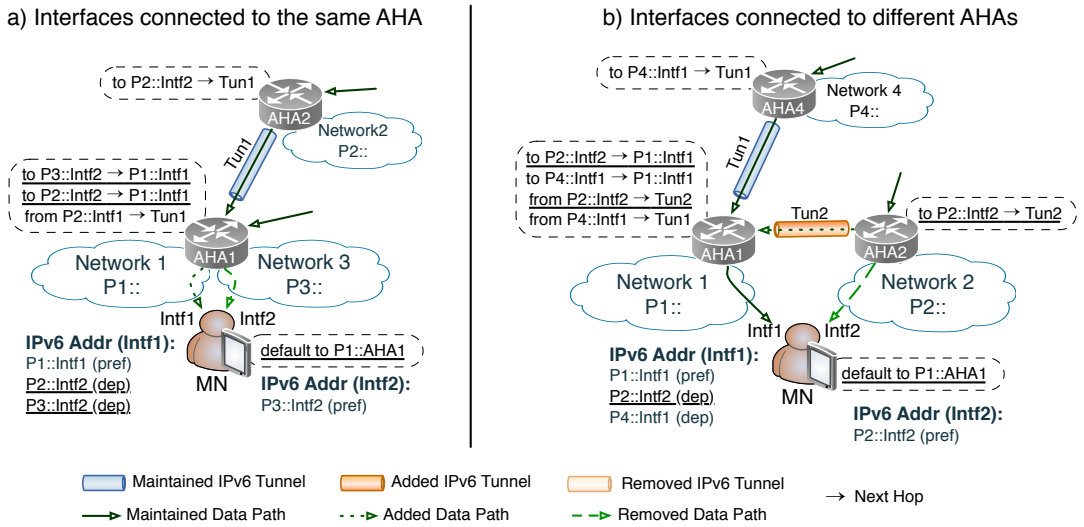


Figure 3.21: Mobility mechanisms for offloading sessions.

Offloading sessions between interfaces connected to the same AHA

The first case (Figure 3.21(a)) provides the mechanisms to offload sessions between interfaces connected to the same AHA, in the same or different IPv6 prefixes. Network architectures might provide one or more IPv6 prefixes using the same or different interfaces from the same AHA, or several access technologies connected through a single AHA sharing one or more IPv6 prefixes. A network operator might use distinct IPv6 prefixes from different access technologies connected to the same AHA, and the user might suffer a disruption in one of the interfaces, or the network operator desires to provide data offloading.

The MN decides to offload sessions from Interface 2 (Intf2) to Interface 1 (Intf1), as illustrated in Figure 3.21(a). It is assumed that the MN has a set of two AHAs (AHA1 and AHA2) with anchored sessions, and it is directly connected to AHA1 through two interfaces (Intf1 and Intf2) from different IPv6 prefixes. There is a tunnel between AHA2 and AHA1 to route the packets from/to *P2::Intf2* of the MN, as well as a rule in AHA1 to locally forward packets destined to *P2::Intf2* to the next hop *P3::Intf2*.

Considering the offloading between interfaces, the AHA1 is responsible to locally forward packets from the IPv6 addresses of Intf2 to the preferred IPv6 address of the Intf1. In the example, the sessions anchored in AHA1 and AHA2 with IPv6 addresses P3::Intf2 and P2::Intf2, respectively, are forwarded to P1::Intf1 in AHA1. Moreover, the MN has also to move/replicate these IPv6 addresses from Intf2 to Intf1, in order to ensure that the Intf1 receives/sends the packets with these IPv6 addresses. More details on the mechanisms are provided in the annexed paper F [20].

Offloading sessions between interfaces connected to different AHAs/LANs

The second case provides the mechanism to offload sessions between interfaces connected to different access networks. While the first case just deals with IPv6 forwarding rules and IPv6 addresses management to offload sessions, this case already introduces

the IP tunnels management. The network operator might decide to offload data through an access network from a different access technology, using the multiple interfaces of the MN, such as from cellular to WiFi access networks. We just provide the example that illustrates (Figure 3.21 b)) the scenario where sessions are offloaded to other interfaces connected to an AHA. The IP mobility might just select part of the AHAs associated to an interface, and consequently the attached sessions, to be forwarded to the other MN's interface through tunnels with the connected AHA.

The MN decides to offload sessions from Interface 2 (Intf2) to Interface 1 (Intf1), where the underlined instructions are the ones changed/added to perform the required offloading between Intf1 and Intf2, while the other ones were previously configured by the offload management. It is assumed that the MN has already a set of 3 AHAs (AHA1, AHA2 and AHA4) with anchored sessions, and that it is directly connected to AHA1 through Intf1 and to AHA2 through Intf2. There is a tunnel between AHA4 and AHA1 to route packets from/to *P4::Intf1*. There is a rule in AHA1 to locally forward packets destined to *P4::Intf1* to the next hop *P1::Intf1*. In the MN, there is a default route to AHA1 for packets from *P1::Intf1*, and another default route to AHA2 for packets from P2::Intf2.

Sessions anchored to AHA2 are now forwarded to AHA1, which is achieved through the configuration of the new Tun2. Hence, packets to the IP address *P2::Intf2* are forwarded to Tun2 in AHA4. There is also a new rule in AHA1 to locally forward packets destined to *P2::Intf2* to the next hop *P1::Intf1*. The MN has to replicate the IP address *P2::Intf2* to Intf1. The IP address *P1::Intf1* assumes the preferred state, while the others are changed to the deprecated state. It is also configured a default route through AHA1 via the Intf1 with the lowest metric, in order to force the packets from ongoing sessions to be transmitted through Intf1, as well as to initiate new sessions.

Evaluation

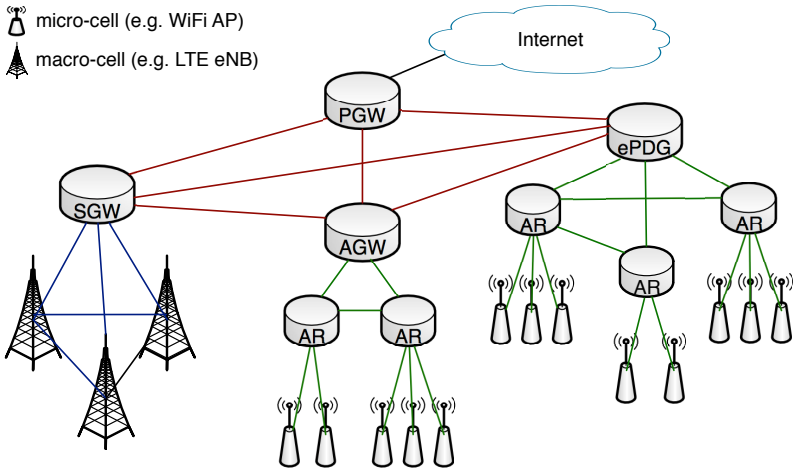


Figure 3.22: An example of an evaluated wired scenario.

We evaluate the two dynamic offload anchoring strategies to place the AHAs at the access network level in comparison with centralized anchoring, through simulations performed in MATLAB [92]. In one of the strategies, called *AHAs in AGs*, the AHAs are distributed through the access gateways of the network (e.g. SGW, AGW, ePDG). In the other strategy, called *AHAs in ARs*, the AHAs are distributed through the access routers of the network, closer to the user (e.g. ARs and eNBs), as illustrated in Figure 3.22. We

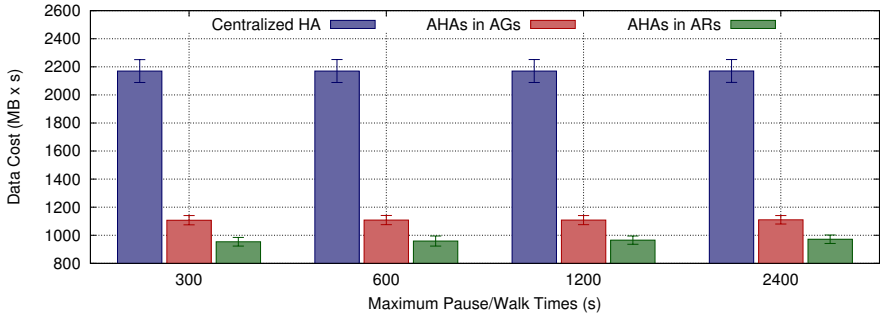
define two types of cells: macro-cells and micro-cells. The macro-cells are distributed in a grid topology to ensure the full coverage in the squared area where MNs move, while micro-cells are uniformly distributed inside of this square. The proposed dynamic offload anchoring was evaluated, based on a well defined offloading strategy. Thus, the offloading management performs the offload of all ongoing and new sessions through the micro-cells, when the MN is in the coverage area of a micro-cell and its speed is nearly null. Otherwise, all sessions will be maintained or initiated through the attached macro-cell. The evaluation is focused on the data cost and data packet delay. The data cost measures the total cost needed to deliver data packets from the CNs to the MNs. It is calculated through the sum of the data packets cost, where each data packet cost is the multiplication of the data packet size, including IPv6 encapsulation if it exists, by the time that the packet spent to cross the network. The data packet delay measures the average time that a data packet experiences to be transmitted from a CN to a MN. We evaluate the influence of the pause and walk intervals of the random waypoint mobility model in the data cost, assuming three types of CNs' placement.

- **CN1:** sessions are uniformly coming from any node of the network.
- **CN2:** sessions are uniformly coming from access gateways or border routers.
- **CN3:** sessions come from a centralized gateway.

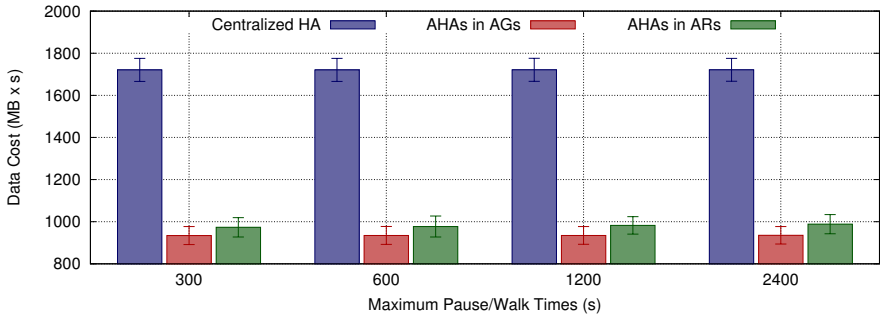
As illustrated in Figure 3.23(a), the centralized HA introduces a higher data cost in the network in user sessions, when compared with distributed AHAs. Sessions are uniformly coming from any node of the network, and a considerable amount of these sessions are initiated and terminated in the same AR/eNB. Thus, the *AHAs in ARs* approach provides the optimized routing path without tunneling to these sessions, while in the centralized HA all packets are firstly routed to the HA, which then tunnels these packets to the current MN location. Moreover, in the approach *AHAs in ARs*, the amount of packets from a session that are offloaded through another AHA different from the anchored one is not very high, and these AHAs can be directly connected, creating short tunnels between them.

When sessions are uniformly coming from access gateways (PGW, SGW, AGW and ePDG) (Figure 3.23(b)), the dynamic offload anchoring approaches continue to reduce both data cost and data packet delay. The main difference is that *AHAs in AGs* is now the approach with the lowest data cost and data packet delay, and the approach *AHAs in ARs* is in the second place. However, both AHA approaches highly reduce the data cost and data packet delay, with both CN1 and CN2 strategies, when compared with the centralized HA. A considerable amount of these sessions are initiated and terminated in the same AR/eNB; thus, both AHA approaches provide the optimized routing path. The amount of packets from a session that are offloaded through another AHA different from the anchored one is not very high. In this case, the *AHAs in AGs* provides the shortest routing paths, through tunnel between AGs or between an AG and an AR. The *AHAs in ARs* approach also provides short routing paths, similarly to the ones of *AHAs in AGs*, since the AHAs can be directly connected, creating short tunnels between them. In the centralized HA, all packets are firstly routed to the HA, which then tunnels these packets to the current MN location (AR or eNB).

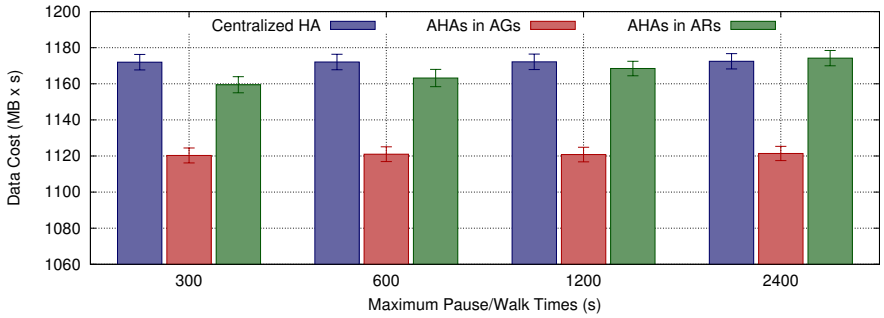
When sessions are coming just from the centralized gateway PGW (Figures 3.23(c)), the centralized HA is the unique approach capable to provide the optimized routing path to all packets at the cost of long tunnels between the HA (PGW) and ARs/eNBs. The centralized HA introduces the higher data cost in the network, since it creates long tunnels



(a) CN1



(b) CN2



(c) CN3

Figure 3.23: Data Cost changing the maximum pause/walk time.

to provide the optimized routing path to user sessions. Hence, the approach *AHAs in AGs* presents the lowest data cost, even when sessions are always coming from the PGW, and the approach *AHAs in ARs* is able to reduce the data cost in the network when compared with centralized anchoring.

We also provide the validation of the dynamic offload anchoring with IP mobility, which aims to demonstrate that it really works in a real testbed. Hence, we used the testbed scenario already presented in DMIPA evaluation (Figure 3.16), configuring AHAs in the three available wireless routers, and using two wireless interfaces (wlan0 and wlan1) and a wired interface in the MN. Three flows are initiated along the experiment in three distinct moments, but all flows are maintained until the end of the experiment. Flow 1 is anchored to AHA1 and it is established through wlan1 interface; flow 2 is anchored to AHA2 and it is established through wlan1 interface; flow 3 is anchored to AHA3 and it is established through wlan0. More details on the testbed experiments can be found in Paper D [20].

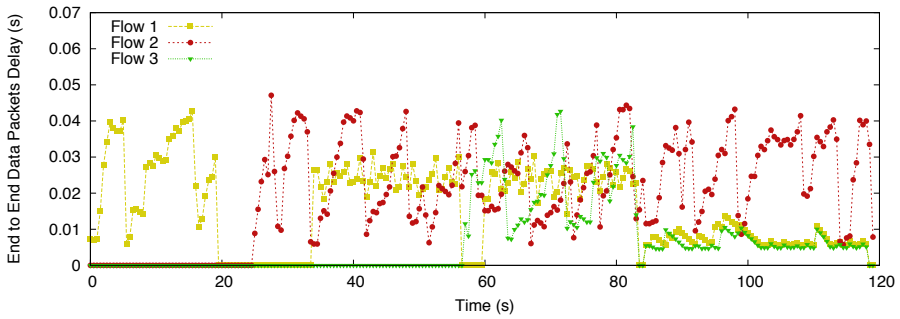


Figure 3.24: Data Delay for TCP Sessions.

The first handover takes place at around 20 seconds, and is the one with higher latency, since it is performed using the same interface; the other two handovers, taking place at around 60 and 80 seconds, are performed using multiple-interfaces to provide data offloading. The first handover might be applied to offload data from a certain eNB to another eNB with more available resources, while the other two handovers might be performed to offload data through a WiFi network or a wired network, respectively. In Figure 3.24, the offloading of flows 1 and 3 through the ethernet cable reduces the average data packet delay and jitter of TCP sessions, since the wireless medium access is usually loaded with several private WiFi networks, while the ethernet cable provides a stable connection with higher bandwidth.

The dynamic offload anchoring is able to reduce the data cost inside the network and the data packet delay of the user's sessions, specially with more distributed content servers and flatten architecture networks, which are the trends for future networks. The validation of dynamic offload mechanisms in a real testbed demonstrated that an appropriate offload management is able to offload sessions with IP mobility, while improving the user experience. More details about this approach and its evaluation are provided in the annexed Paper D [20].

3.3.3 Replicated Bindings for Network-based Localized Mobility

From the previous evaluations, the dynamic mobility anchoring at the ARs level does not always provide the optimized routing path to sessions when sessions are relatively long, or the MN is highly mobile (e.g. in high MN's speed or small cells scenarios). Consequently, the number of simultaneous MAs to be kept for an MN may increase dramatically, as well as the number of sessions being delivered through the non-optimized routing path.

In [21], we propose to study a network-based replication strategy, based on PMIP, in comparison with centralized and distributed mobility management models. Multiple Local Mobility Anchors (MLMA) spread the LMA functionalities through the ARs and the GW of the network, in order to maintain the whole bindings, between the well-known IPv6 address of a MN and its current AR, in each AR/GW of the network. Thus, MLMA is always able to enforce the optimized routing path to data packets, which is already provided by the adopted routing protocol or static routing to the non-mobile IPv6 communications.

MLMA choses an available IPv6 prefix to be used for localized identification of the MNs, while they move inside the operator network, which is used to provide mobility support to the MNs (Figure 3.25). The IPv6 prefix works as an identification inside the operator network, while from the outside, it is seen as another IPv6 prefix that belongs to that operator network. As long as a packet destined to an IP address of this IPv6

prefix enters in the operator network, there is a set of LMAs responsible to forward the packet to the current access network of the MN. The ARs always provide the IPv6 prefix with mobility support, which is the same in all access networks, to maintain the local identification of the MN.

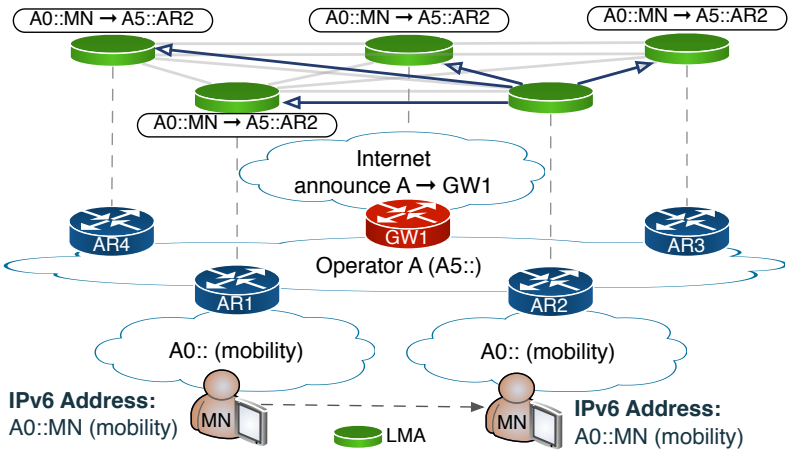
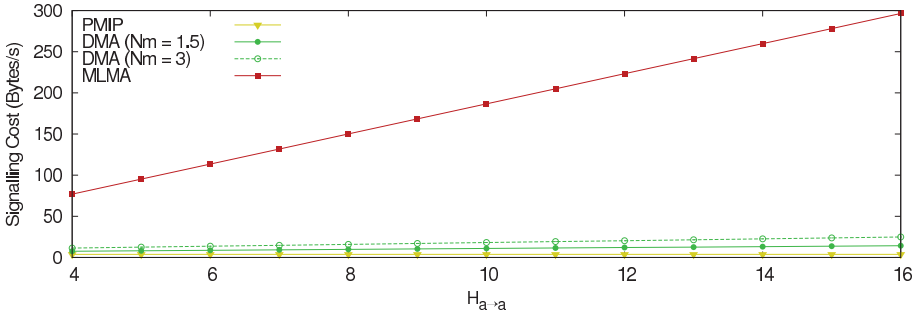


Figure 3.25: Multiple Local Mobility Anchors.

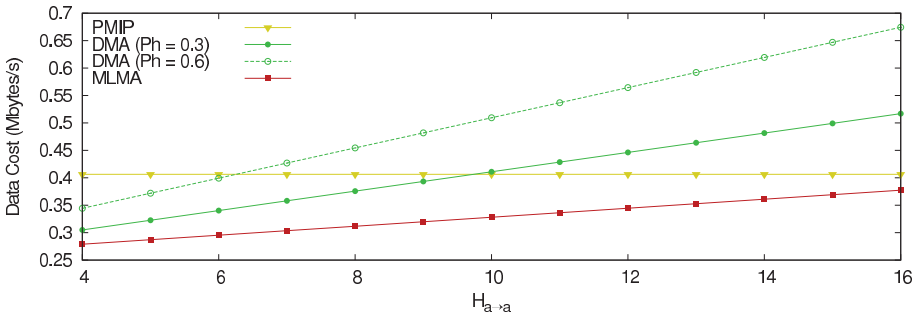
When a LMA detects a MN attachment, it updates all other LMAs of the network through the PBU and PBA messages of PMIP, which contain the well-known IPv6 address of the MN and the IPv6 address of the new AR of the MN. Therefore, it is assured that all LMAs are updated with the IPv6 address of the current AR of the MN, which is used to maintain the ongoing sessions, and in the reachability for new sessions establishments. As it happens in PMIP, each LMA of MLMA may need to maintain a larger binding cache proportional to the number of MNs. This is a disadvantage of these IP mobility models, but it is the price to pay to improve the data routing path optimization while providing session continuity and reachability to the MN. Although the size of the binding cache is larger, it is a simple database system with entries containing only two IPv6 addresses, that can be quickly managed with appropriate algorithms. MLMA is always able to provide the optimized routing path to data sessions in the operator network, at the cost of more signaling messages to update the MN location in the binding caches distributed through the LMAs of the network.

It is also proposed an add-on to remove the tunnels between LMAs, reducing the complexity of the mobility management control and eliminating the extra header introduced by the encapsulation. The idea of MLMA is to move the mobility-awareness of the MNs to the access networks, maintaining the core/backbone network simple and quasi-static. Hence, we propose to deploy the same IPv6 prefix in the entire backbone/core network to connect ARs and the GW. Since the ARs always announce the same IPv6 prefixes and we expect a quite stable core/backbone border elements (e.g. ARs and the GW), the routing inside this core/backbone network will be quite static, even if it includes a large number of routers/switches. This way, instead of tunnels, the forwarding of sessions requiring mobility support will be based on routing rules, where the next hop is the current AR of the MN (obtained from PBU messages), which changes as long as the MN moves between access networks. Each LMA placed in an AR/GW maintains a cache with the bindings of all MNs of the operator network. This add-on is able to eliminate the encapsulated packets in the network, the consumed resources from encapsulation/decapsulation mechanisms, and the number of virtual interfaces per AR/GW for tunneling.

The evaluation of this approach is performed through analytical models for PMIP (centralized), DMA (distributed) and MLMA (replicated), in order to evaluate both signaling and data costs. We evaluate signaling and data costs changing $H_{a\to a}$, and maintaining $H_{g\to a} = 10$, Figure 3.26(a) and 3.26(b) respectively. $H_{a\to a}$ is the average number of hops between two ARs, while $H_{g\to a}$ is the average number of hops between a gateway and an AR. The other parameters of the scenario, as well as other evaluated scenarios are defined in Paper E [21].



(a) Signaling Cost



(b) Data Cost

Figure 3.26: Comparing PMIP, DMA and MLMA changing the $H_{a\to a}$.

The signaling cost of PMIP is the lowest one, and it does not depend on $H_{a\to a}$, since there are no signaling messages exchanged between ARs. The MLMA has a much higher signaling cost when compared with other approaches, and it increases much more with the number of hops between the ARs ($H_{a\to a}$), since the current LMA (AR) of the MN updates all the other LMAs (ARs and GW) of the network per handover. However, this is a small price to pay, when the data cost of MLMA can reach improvements much higher than the increased signaling cost (Figure 3.26(b)). For the smallest $H_{a\to a}$ evaluated, MLMA is able to improve the data cost above 0.124 Mbps when compared with PMIP, while for the highest $H_{a\to a}$ evaluated, MLMA is able to improve the data cost above 0.137 Mbps when compared with DMA. The data cost of PMIP does not depend on the $H_{a\to a}$, since data packets are always routed via the centralized LMA, placed in the GW. For higher values of $H_{a\to a}$, DMA has a higher data cost than PMIP, which fits scenarios with centralized/hierarchical network architectures with highly mobile users.

The replication strategy showed to be better than distributed and centralized models, regarding the in-network communication cost (signaling and data network costs). In scenarios with mobile users consuming more data traffic (e.g. requiring data sessions along the entire day while moving), the replication strategy significantly reduces the in-network communication cost. In scenarios with quite static users that require just few services during the entire day, the replication strategy introduces a higher complexity in order to

maintain the replicated bindings through the entire LMAs of the network, which might not compensate the slightly decrease of the in-network communication cost. However, the MLMA might be adjusted to overcome the different services and user requirements, since it might be integrated with a selective binding update strategy (e.g. choose the LMAs to send the PBU). The MLMA might update just the previous LMAs of the MN or part of them, for a user that stays most of the time at a predefined location (e.g. home or work). Thus, it might increase the data cost of the network, but it happens just for short periods of time when the user is outside of its predefined location. The MLMA might also update a set of LMAs of the network, which are used to establish sessions with the MN. Hence, the other non-updated LMAs of the network maintain the binding to a usual LMA of the MN, which then forwards the packets to the current LMA of the MN. More details about this approach and its evaluation are provided in the annexed Paper E [21].

3.3.4 Context-aware Adaptive IP Mobility Anchoring

The novel distributed mobility management trend is a promising direction to cope with the increasing mobile data traffic and flatten network architectures. Most of the novel mobility approaches distribute the mobility anchors through the access level, such as DMIPA, as opposed to the centralized mobility anchoring model (e.g. MIPv6 and PMIP); but there are other recent approaches argue that mobility anchors closer to the content servers may be the solution to optimize the mobility performance.

Although there are several advances in the mobility management, there is not an IP mobility approach suitable for a large set of scenarios, since its performance depends on several constraints, such as user mobility, network topology/architecture and traffic sessions. Hence, we propose a distributed and dynamic IP mobility model driven by the context information from the user and the network, which means that IP session continuity and IP reachability are assured to the MN sessions according to its individual characteristics/needs. Although there is an initial anchoring of mobility sessions, the proposed approach is able to signal other MAs to optimize the routing path to new and ongoing sessions of the MN. Moreover, the complexity of the bindings management is reduced, since each MN is responsible to maintain its own bindings and the respective signaling.

The context-aware mobility anchoring protocol is developed based on the assumption that the MAs might be placed all over the network, usually at border nodes, such as routers in the MN network, the gateway of the MN network, routers in the CN network, the gateway of the CN network, or the CN itself. The MN can anchor sessions to these distinct MAs, which are able to provide IP address allocation and routing/forwarding, in order to provide the optimized routing with minimum tunneling for new and ongoing sessions. Some of the MAs placed in the corresponding networks are also assigned for routing/forwarding based on bindings provided by the MN. At least the corresponding networks with more established sessions (e.g. youtube, google and Facebook servers) should deploy a MA in the AR or its Content Delivery Networks (CDNs). Using information from user mobility degree, the MAs availability in the network, the network topology and the MN ongoing sessions, we are able to provide the optimized routing path with a minimized tunneling, and without introducing a high complexity to the network.

The proposed approach introduces a flexible and host-based mobility context management, which reduces the complexly to manage the mobility context in a centralized node, since each MN just has to maintain its own set of MAs, the respective bindings and mobility routes. The distribution of the MAs for IP address allocation and routing/forwarding, as well as the distribution of mobility context management through the

MNs, eliminates the necessity of a centralized entity for the IP mobility management, while it adapts the mobility management for flatten network architectures and distributed content servers. The bindings are managed by the MNs, where each MN is responsible to create/update/remove the bindings in its set of MAs. The MN is able to dynamically select its set of MAs as long as it changes access network or initiates/terminates sessions. The MAs allocate Mobile IP Addresses (MoA) to the MN, and the MN collaborates with MAs in the IP addresses management, providing the information about the necessity of the assigned IP addresses. There are also IP addresses assigned by the ARs of the network that do not provide any mobility support. The MAs in the MN network (e.g. in ARs and GWs) may be known through Router Advertisement (RA) or Dynamic Host Configuration Protocol (DHCP) messages, where the IP addresses of the closest MAs have higher metric values. The MAs far from the MN, such as the ones close the CNs, are known through a Mobility Anchor Server (MAS), which contains the IPv6 prefix and the IP address of the MA serving this prefix. The MAS can be deployed using DNS or other similar system.

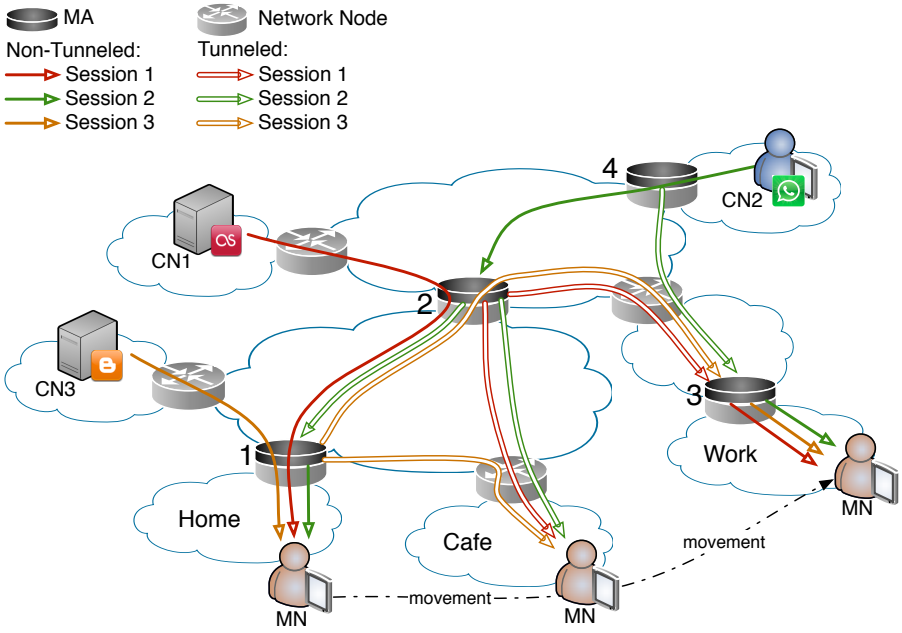


Figure 3.27: An example of routing and anchoring with context-aware adaptive IP mobility anchoring.

The operational example is described based on Figures 3.27 and 3.28, where the signaling messages and routing paths are illustrated. The example addresses the protocol operation based on a MN moving from Home to Work that stops in a Cafe. The MN establishes three sessions with different CNs, while at Home, which will remain active during the movement of the MN to Work. The MAs are distributed over some of the network nodes, as shown in Figure 3.27. More details on the protocol operation can be found in the annexed paper F [22].

Home: The MN receives a normal IPv6 address without mobility support (IPh address) from the Home router. It also provides Mobility IP addresses (MoA), which ensure IP session continuity and IP reachability, such as the MoA1 and the MoA2, from MA1 and MA2 respectively. From the metrics received in the message, the MN is able to distinguish between the closest MA (MA1), usually placed in the AR, and far MAs (MA2). Thus, the MN configures the three IPv6 addresses (IPh, MoA1, MoA2) and adds MA1 and MA2

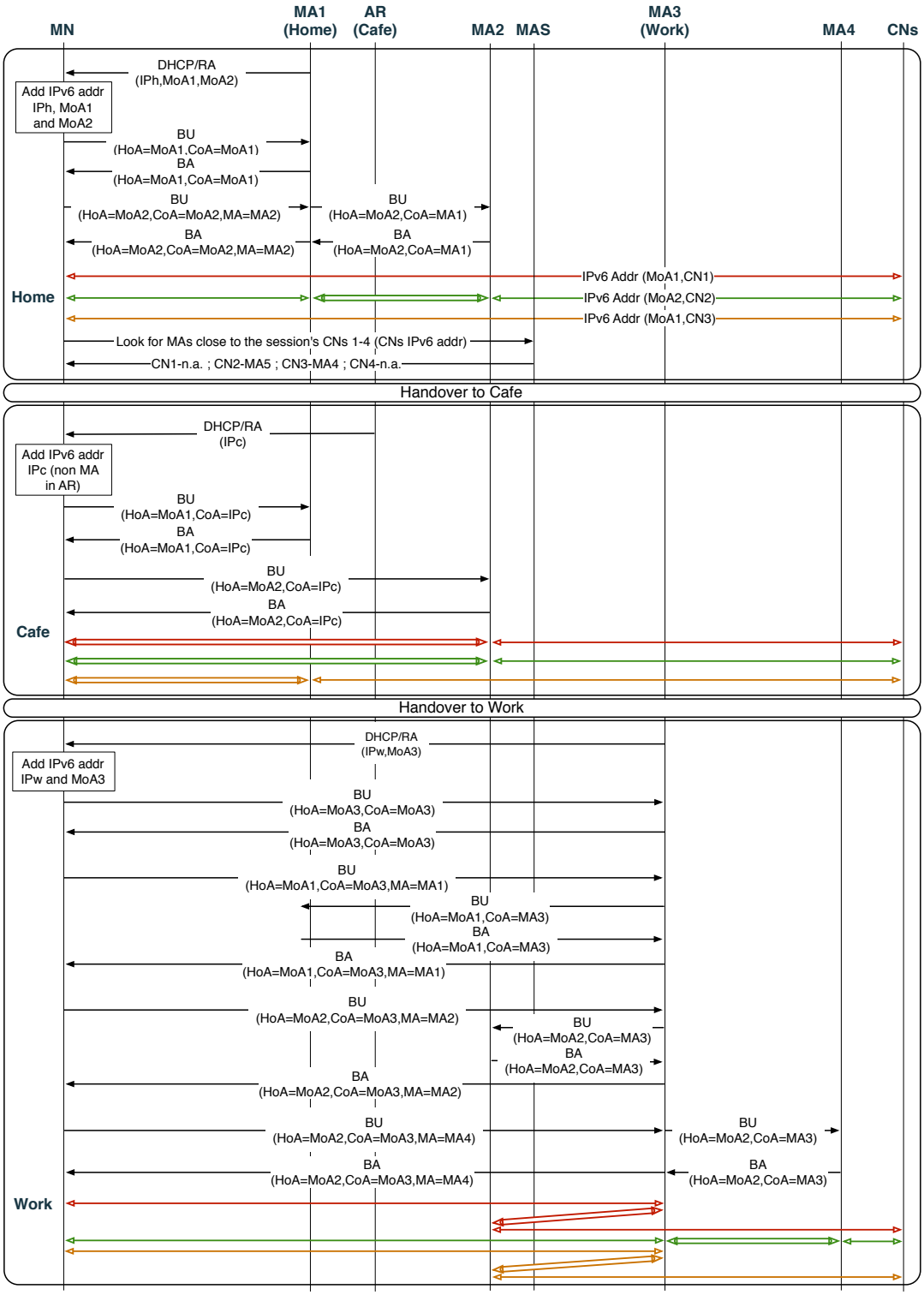


Figure 3.28: An operation example of context-aware adaptive IP mobility anchoring.

to the set of available MAs. The IPh is used for sessions that do not require IP session continuity nor IP reachability support, while MoA1 and MoA2 may be used otherwise. The MN selects the MoA2 from the MA2 to ensure its IP reachability, based on the user profile.

The MN sends messages to MA1 and MA2 to register the MN, which assures that these MoAs are not allocated to any other MN, while the MN needs them. The MA1 is used as an intermediary in the messages exchanged between the MN and MA2, since the tunnel for the sessions established with MoA2 is configured between the MA1 and MA2.

Sessions 1 and 3, initiated by the MN, are established without tunneling support in the optimized routing path between the CN and the MN, being selected the MoA1 from the MN side. The MN always selects the MoA provided by the closest MA in the establishment of new sessions, if it is available, since it is the one that ensures the optimized routing path with lower tunneling cost. The session 2 initiated by the CN is established with the MoA2 of the MN, since it is the IPv6 address that provides IP reachability. Thus, the packets of session 2 are tunneled from the beginning between MA1 and MA2.

After the sessions are initiated, the MN search for MAs close to the current CNs of the MN (CN1, CN2, CN3) in the MAS. The MAS replies with the MAs close to the CNs, if they have any. While CN2 is close to MA4, CN1 and CN3 do not have any close MA. The set of MAs close to the current CNs is maintained by the MN, which can be latter used to optimize the routing path and/or minimize the tunneling of ongoing sessions.

Cafe: After the handover of the MN to the Cafe network, it receives a normal IPv6 address without mobility support (IPc) from the Cafe router. It is not provided any new MoA nor any indication of the MA from the Cafe router, since there is no MA deployed there. The MN remains with the same set of MAs and MoAs to establish/maintain sessions with mobility support.

In order to provide session continuity to the ongoing sessions and to assure the reachability for new sessions, the MN must send messages to the MAs with anchored sessions. Hence, the MN sends a message to MA2 to update the IP reachability binding and to configure a tunnel between MA2 and the MN (IPc) to maintain the ongoing session 3. The MN also sends a message to MA1 to maintain the ongoing sessions 1 and 3, through a tunnel between the MN and MA1. These two messages are enough to provide IP session continuity and IP reachability, but the routing path of session 1 can be optimized by MA2. Thus, the MN sends a message to MA2 to create a binding there (MoA2-IPc), while the tunnel can be the same already configured for IP reachability. The MA2 has two rules to forward all packets with IP destination MoA2 and MoA1 to IPh.

New sessions requiring IP continuity support are initiated through MoA1 with the tunnel between the MA1 and the MN, while sessions requiring IP reachability are initiated using MoA2 through the tunnel between the MA2 and the MN.

Work: After the handover of the MN to the Work network, the MN receives a normal IPv6 address without mobility support from the Work network (IPw). It is also provided the new MoA3 from the MA3 of the Work router.

The MN assigns the new IPv6 addresses MoA3 and IPw, adding it to the existent ones (MoA1 and MoA2), and also adds the MA3 to the set of MAs. The IPw is used for sessions that do not require IP session continuity nor IP reachability, while MoA2 and MoA3 are used for sessions requiring IP reachability or IP session continuity, respectively. The MN sends messages to MA3 to register the MN, which ensures that MoA3 is not allocated to any other MN. MA3 will be used as an intermediary of the MN to establish tunnels with other MAs of the MN with anchored sessions.

In order to provide IP session continuity and IP reachability, the MN sends messages to the MAs with anchored sessions through the intermediary MA3. Hence, the MN sends a message to MA2 through MA3 to update the IP reachability binding (MoA2-MA3) and the binding (MoA1-MA3), configuring a tunnel between MA2 and MA3 to maintain the ongoing sessions 1 and 2. The MN also sends a message to MA1 through MA3 to

maintain the ongoing session 3, through a tunnel between MA3 and MA1. The routing path of session 2 can be optimized with MA4, close to the CN3. Thus, the MN sends a message to MA4 through MA3, in order to create a binding there (MoA2-MA3) and to configure a tunnel between MA4 and MA3, which forwards all packets with IP destination MoA2 to MA3. From now on, MA4 is included in the set of MAs of the MN.

Evaluation

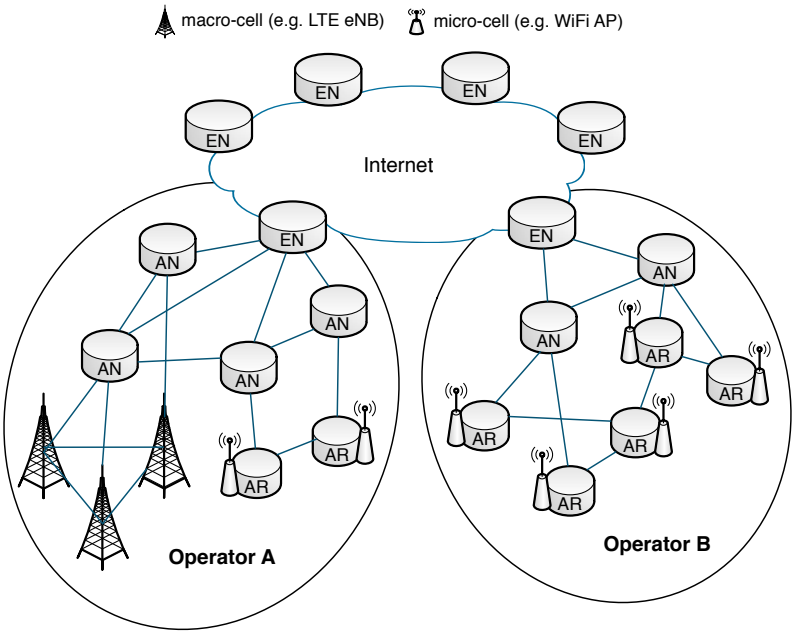


Figure 3.29: An example of an evaluated wired scenario.

We evaluate the proposed mobility anchoring approach in comparison with centralized, AR and CN anchoring models through MATLAB [92]. In the wired topology, we randomly generate flat network topologies, in which we assume a more flatten network topology. A node connects to the higher hierarchy level node, to lower hierarchy level nodes, and it can also connect to nodes of the same hierarchy level, as presented in Figure 3.29. For the wireless part, it was defined two types of cells: macro-cells and micro-cells. The macro-cells mimic the cells currently associated to cellular access networks, while micro-cells represent the cells from the WiFi access networks. The MNs move with a Random Way Point mobility model in this square area with macro and micro cells. We defined four types of scenarios to evaluate the data performance of the considered mobility anchoring models.

Scenario S1: Offload of all ongoing and new sessions through the micro-cells of two operators, when the MN is in the coverage area of these micro-cells and its speed is nearly null. CNs are uniformly distributed all over the network, thus, sessions can come from any node of the network.

Scenario S2: Offload of all ongoing and new sessions through the micro-cells of two operators, when the MN is in the coverage area of these micro-cells and its speed is nearly null. CNs are uniformly distributed just through the edge nodes.

Scenario S3: Offload of all ongoing and new sessions through the micro-cells of one operator, when the MN is in the coverage area of these micro-cells and its speed is nearly null. CNs are uniformly distributed through the network nodes of one of the operators.

Scenario S4: Offload of all ongoing and new sessions through the micro-cells of one operator, when the MN is in the coverage area of these micro-cells and its speed is nearly null. CNs are uniformly distributed through the edge nodes.

More details on the configured scenarios, as well as an exhaustive evaluation can be found in annexed Paper F [22]. The metrics evaluated are the data delay, data cost, tunneled packets and tunnel length, in which data cost and data delay was already defined. Tunneled packets is the ratio of tunneled data packets from the total data packets transmitted between CNs and MNs, while tunnel length is the average tunneled hops per data packet that are tunneled.

The centralized anchoring provides shorter tunnels than AR anchoring and CN anchoring, but it tunnels all packets, as shown in Figure 3.30. The sessions are anchored in a centralized node in the middle of the network, which has a similar distance to the ARs/eNBs of the network. Thus, it does not provide so long routes to offload sessions through the WiFi of the two operators. It has a lower data delay than AR anchoring, but a higher data delay than CN and adaptive anchoring.

The AR anchoring provides the optimized routing path for sessions initiated and terminated in the same AR/eNB (Figure 3.30), but it introduces long tunnels and the highest data delay, due to the offload of sessions through the WiFi from operator A and/or B. Although it introduces long tunnels, the AR anchoring provides the lower number of tunneled packets, independently of the evaluated scenario. However, in the considered four scenarios, the AR anchoring has the highest data cost.

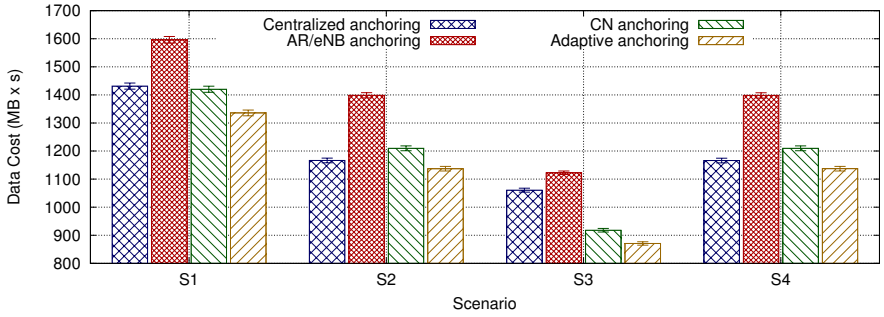
The CN anchoring provides the lowest data delay at the cost of the longest tunnels between the CN and the MN, independently of the scenario evaluated, as shown in Figure 3.30. The CN anchoring tunnels all packets in scenarios S2 and S4, where all packets come from edge nodes, while in scenarios S1 and S3, a small portion of the total packets are not tunneled, when the CN and MN are connected to the same AR/eNB. The CN anchoring shows a data cost similar to the data cost of centralized anchoring, except for scenario S3, where the CN anchoring has a lower data cost than centralized anchoring, since the MN just connects to one operator, in which the CNs are distributed.

The adaptive anchoring dynamically adapts to the evaluated scenarios, since it provides the best MA as long as the context of the user, the session or the network changes. Overall (Figure 3.30), the adaptive anchoring provides the lowest data delay, the shortest tunnel length and the minimum tunneled packets. There is an exception in scenario S3, where the adaptive anchoring has a higher tunnel length than centralized anchoring, but this is necessary to reduce the data delay and to provide the lowest data cost.

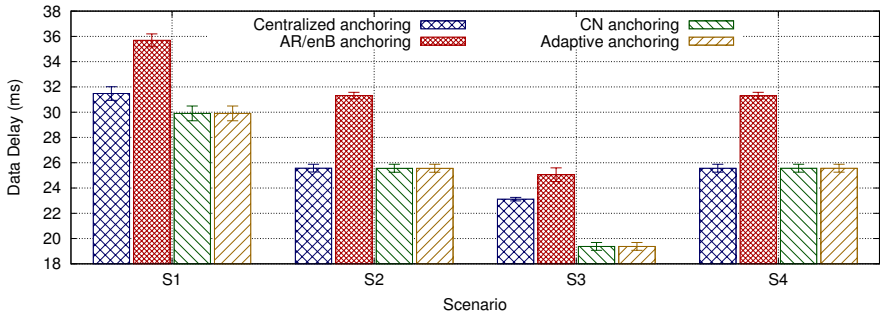
The outcome of the evaluation demonstrates that the adaptive anchoring adapts to the evaluated scenarios, since it always provides the MA that ensures the optimized data performance, when the context of the user, the session or the network changes. Overall, the adaptive anchoring aggregates the advantages of the other three anchoring models, since it provides the lowest data delay as the CN anchoring, the shortest tunnel length as the centralized anchoring, and the minimum tunneled packets as the AR anchoring. There are rare exceptions, where the adaptive anchoring introduces higher tunnels length and tunneled packets, but this is necessary to reduce the data delay and to provide the lowest data cost.

3.3.5 Summary

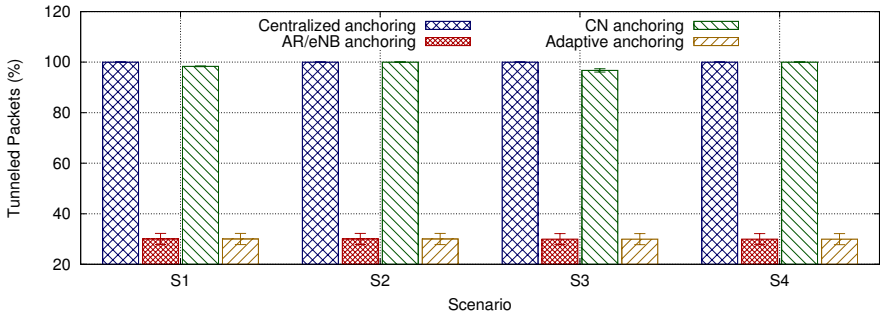
This section describes novel decentralized IP mobility approaches and mechanisms to cope with recent network trends of flatten network architectures, content servers close to



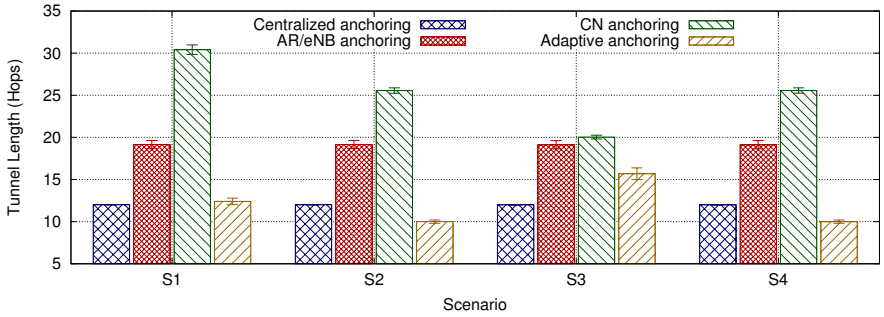
(a) Data Cost



(b) Data Delay



(c) Tunneled Packets



(d) Tunneled Length

Figure 3.30: Mobility anchoring models performance changing the offloading strategy and the CNs location

the users, and data offloading from cellular to other access networks. These approaches and mechanisms were developed based on the outcome provided by the mobility management

analysis and studies of the sections 3.1 and 3.2, respectively. The proposed approaches and mechanisms presented in section 3.3 are mainly focused on mobility anchoring, routing/forwarding, mobility context maintenance/management (e.g. bindings, mobility anchors and mobility routes), and IP address allocation/management. It is proposed a novel mobility management approach, called DMIPA, which distributes the mobility functionalities through Access Routers (ARs) and MNs. DMIPA was exhaustively evaluated through analytical models, simulations and testbed experiments, as well as integrated with seamless horizontal handover mechanisms and evaluated in vehicular environments. The mobility mechanisms from DMIPA were also specified for multihomed scenarios, in order to provide data offloading with IP mobility support from cellular to other access networks. In the pursuing of the optimized routing path, a novel network-based strategy for localized mobility was addressed, in which a replication binding system is deployed in the mobility anchors distributed through the access routers and gateways. Finally, it is presented a context-aware adaptive IP mobility anchoring model that dynamically assigns the mobility anchors that provide the shortest routing path and the minimized tunneling to a session, based on the user and network context. The distribution of mobility management functionalities through the network nodes and user devices showed to be a promising direction to provide more appropriate mobility management in future networks, with flatten architectures and distributed content servers. The integration of novel decentralized concepts, such as adaptive mobility anchoring and dynamic mobility, are able to reduce the network resources consumption and to improve the experience of the required services. More conclusions about the work developed on the distributed mobility management, as well as possible future direction are presented next on the chapter 4.

Chapter 4

Conclusion

The final chapter presents the conclusions of the Thesis, as well as future research directions. This chapter highlights the main contributions from the work developed in the Thesis, answering to the research objectives proposed in the introduction Chapter 1. The major remarks on the decentralized mobility management are presented to provide the guidelines for mobility management in future networks. Finally, the last section point out possible future research directions, proposing future research ideas for the work developed in the Thesis, as well as addressing topics out of the scope of this Thesis that can add value when properly integrated.

4.1 Achievements on the Research Objectives

The main question addressed in this Thesis is *how to addressed the IP mobility management in future networks*? This main problem was decoupled into 4 research objectives in Chapter 1, which were defined from the outlined issues of the current mobility management models applied to the new network evolution trends. In this section, conclusions of those research objectives are presented.

Research Objective 1: How decentralized should be the mobility management regarding the mobility functionalities to improve the network performance and optimize the network resources?

The MN, CN and MA are generic roles that are today present in different mobility management solutions, independently of the OSI layer where the mobility solution resides. Excluding the MN and CN, which are the end-points communicating, current mobility management is heavy-centralized, since all the mobility functionalities are provided by this static and centralized MA. This centralized model brings scalability and performance issues when deployed in novel flatten network architectures (e.g. LTE) prepared for the increasing mobile data. We have characterized mobility management through the identification of the mobility management functional blocks and their interactions. From the identified mobility functionalities, we studied different ways of grouping these functional blocks, in which some functional blocks are distributed through network elements, while other remain centralized in the MA. The outcome from the studies using analytical models and simulations demonstrates that at least part of the mobility functionalities should be distributed, such as the mobility anchor, the binding mechanism, the routing/forwarding and the handover negotiation. The distribution of the mobility anchor and routing/forwarding through network nodes, and the binding mechanism and handover negotiation through the MNs improve the overall mobility management performance. It

reduces the data cost, the signaling cost, the data packet delay, the time to update the bindings and the packet loss. Although the achievement of the research objective is positive, this is not a closed research topic where future directions are presented in the future work section.

Research Objective 2: How should the mobility anchors be placed and assigned to reduce the network cost and to improve the user experience?

In current mobility management models, the mobility anchor is the centralized and static entity that is responsible to manage the mobile data and mobility context of all user connected to the network. This model brings issues related with scalability and performance, such as longer traffic delays and higher network resources consumption. We propose a novel distributed and dynamic host-based IP mobility approach, were the mobility anchors are distributed through the access routers of the network. These mobility anchors are mainly responsible for routing/forwarding, while the handover negotiation and binding mechanisms are assured by the MNs, where each one maintains its mobility context and initiates the binding mechanisms when a handover is detected. This approach completely eliminates the necessity of a centralized entity for mobility management purposes, while ensures IP session continuity to the users. The current AR is the mobility anchor for new sessions, which is maintained while the session remains active, even when the user attaches to another AR. The mobility anchor is dynamically assigned according to the user current ARa and the MA availability. We also studied other approaches to distribute the mobility anchors through the network, such in ARs and GWs of the MN network. or even close to the CN. The outcome demonstrates that the distribution of the mobility anchors reduce the network cost and improve the user experience when compared with centralized mobility anchoring. The proposed approach was not only evaluated through analytical models and simulations, but it was also validated in a real testbed. We showed that the best way to distribute and assign the mobility anchors depends on the user (e.g. mobility degree and patterns), the network (e.g. topology and resources) and the sessions (e.g. CN and lifetime), thus it is difficult to point out one of them to be used in all scenarios. Hence, we defined a context-aware adaptive mobility anchoring approach, which considers the context about the user, the network and the sessions to anchor a new session of the user, as well as to chose another mobility anchor for ongoing sessions in order to optimize the routing path and reduce the tunneling cost. The evaluation of this approach shows that it reduces the data cost, the data delay and tunneling when compared with the predefined anchoring approaches (e.g. Centralized anchoring, AR anchoring and CN anchoring). We not only study different distributed mobility anchoring approaches in comparison with centralized anchoring, but we also point out an interesting approach to place and assign the mobility anchors, which dynamically assigns the proper mobility anchors to new and ongoing sessions, as the scenario changes (e.g. user, network or sessions).

Research Objective 3: How should the mobility context be managed to be scalable and easily-deployed in a large network?

In current mobility management model, the mobility context is maintained and managed in the centralized MA. The mobility context includes all information about the user and the network that is necessary to the mobility management, such as bindings, mobility routes, IP addresses, IP tunnels and mobility anchors. This centralized model to maintain and manage the mobility context introduces performance and scalability problems/limitations, such as single point of failure and higher signaling loads. We proposed a distributed and dynamic IP mobility approach, where the mobility context is maintained

and managed by the MN, thus, each MN is responsible to maintain its own mobility context. The mobility context of a MN is used in the handover negotiation phase to update the mobility anchors with routing/forwarding and a location management system. Oppositely to the current centralized model, where the mobility context of all MNs is maintained by a single and centralized entity, the proposed approach distributes the mobility context through MNs, where each one just keeps a small portion of the entire mobility context, which corresponds to its own mobility context. The proposed approach provides dynamic mobility, which means that mobility support is just provided from the time that a session undergoes into an handover; otherwise the sessions are delivered without any mobility support. Hence, the dynamic mobility reduces the signaling cost, specially for small values of SMR, where the MN stops for longer periods than the sessions duration. The evaluation of the proposed approach shows that it reduces the signaling cost and the binding updating time when compared with the centralized model. A static and centralized entity is usually far from MNs and distributed mobility anchors, while the mobility context carried by the MN can be quickly used, when necessary, to update close mobility anchors.

Research Objective 4: How should the IP mobility be deployed to exploit the multihoming, allowing IP data offloading schemes with shorter routing paths?

Current centralized mobility anchor is adopted by mobile networks, in order to provide IP mobility support to the multihomed devices of the user. The data offloading through a different access network than cellular networks is achieve at the cost of routing all mobile data packets through the centralized mobility anchor, which then forwards the mobile data packets to the selected network/interface of the user. The multihomed support is assured by extensions that enable the use of heterogeneous wireless interfaces for multihomed devices, such as smartphones. The distributed mobility management should be developed considering the multihomed devices to cope with data offloading strategies. From the proposed distributed and dynamic IP mobility approach, we specify the main mechanisms to provide mobility support with multihomed devices, such as the management of mobility anchors, IP tunnels, IP addresses, mobility routes and interfaces. Thus, part of the mobility and offload management is performed by the MN, which should include information from the network in the decision process. The outcome of the evaluation demonstrates that the proposed mechanisms are able to offload new and ongoing sessions with session continuity support through different access networks, in which the packets follow a nearly optimized routing path. The proposed mechanisms were also validated in a real testbed to prove that it properly works on real environments. Data offloading can be used not only to alleviate loaded networks, but also to improve the user experience, when provided by proper management decisions with mobility support.

4.2 Final Remarks

In the scope of this Thesis we defined, studied and evaluated approaches and mechanisms to decentralize the mobility management, in which the main focus was in the distribution of mobility anchors, routing/forwarding, mobility context, handover negotiation and address management. From the overall outcome, the decentralized mobility management is a promising path to cope with flatten network architectures and the distribution of content servers. The decentralized IP mobility brings benefits to the network, such as reducing the network resources consumption and increasing its content delivery performance, as well as assists the user, such as improving the user experience and ex-

ploiting its devices' resources/interfaces. The proposed decentralized mobility approaches and mechanisms should be integrated with proper management decision algorithms, in order to be able to exploit them for the profit of networks and users. Furthermore, and considering the validation provided through a real testbed, the decentralized mobility management concepts should be easily deployable in today's networks, without requiring major modifications to current communication elements. In a glance, this Thesis was able to demonstrate that decentralized mobility management is the future direction for mobility management, hence, its ideas should be seriously taken into account by mobile operators in the deployment of future networks.

4.3 Future Work

Considering the work developed in this Thesis and its broad research scope, there are several topics that may benefit from further development in order to pursue a better mobility management in future networks. These topics are presented next:

Continue the Optimization of the Adaptive Mobility Anchoring

The proposed context-aware dynamic mobility anchoring with on-demand routing optimization demonstrated to be a good solution for mobility anchoring. Besides the advances, we believe that the integration of more context information about the user and the network can improve the mobility management performance. As user context, we consider information on the user history, such as usual networks, mobility pattens and usual required content. As network context, we consider information about the network resources and performance, such as resources available, network congestion, and topology impact. This context information may help both in IP session continuity and IP reachability; thus, its integration fits in the optimization of the global mobility management, including a global location system. The integration of context information may be also interesting for vehicular networks in scenarios of public transportation or GPS-assisted vehicles, in which these vehicles have nearly-predefined routes. The integration of this type of context information should be properly evaluated through simulations and real experiments in distinct scenarios, in order to measure their impact in the mobile data and signaling.

Develop and/or Integrate Seamless and Soft Handovers in Decentralized Mobility

The decentralized mobility management should integrate seamless and soft handover mechanisms, which means that handover between IP networks or a simple data offloading through other interface should not decrease the experience of any user's service. In this sense, the handover should be completely transparent to the user. This subject of seamless and soft handovers is far more important in highly dynamic environments, such as vehicular networks, where the user may be changing between a large set of networks in a short period of time. A decentralized mobility management approach should be able to provide soft and seamless handover to be adopted in most of the future network scenarios. Moreover, this integration should be validated in real environments with available mobile devices (e.g smartphones), since the impact of the handovers is most of the times strictly associated with the technical implementation and optimization of the mechanisms.

Develop and/or Integrate Multihoming/Offloading Management Algorithms

Current distributed mobility management is based on simple decision algorithms for multihomed scenarios. It is important to develop algorithms for the mobility and offloading management decisions. The multihoming concept has a large scope, which may include scenarios where a user can have more than one IP addresses, connections, interfaces, devices and providers. The multihomed scenarios increase the possibilities to establish/maintain communications, but a non-well designed decision algorithm to exploit the multihoming can even be worse than using just one possibility of the multihomed scenario. The mobility/multihoming management algorithms should take into account the context information of user and network to obtain a compromise solution that fits the desire of both users and network operators. The users are always interested in improving their experience at lower costs, while network providers are always looking for strategies to optimize the network resources and increase their profits. Moreover, the data offloading is a use-case of the multihoming that is highly important for the mobile network operators that need to alleviate their cellular networks from the explosion of the mobile data traffic consumption. The data offloading from the cellular networks to WiFi access networks has been increasing over the last years, and it is expected to be applied at large scale for around half of the consumed mobile data traffic, in order to overcome the spectrum limitations of the cellular networks in the next years.

On-Demand and Selective Mobility Management

The on-demand and selective mobility management is crucial for a scalable mobility management in future networks. The mobility management should be provided when needed and to the sessions that really need it. The dynamic mobility support is one of the main topic of DMM, in which approaches have been proposing solutions to provide session continuity support just after a session undergoes into an handover. It is necessary to go further in this topic to provide a more selective mobility management support, and develop solutions that just provide the mobility functionalities that a session requires. The work develop in [85] introduces a proposal, in which applications running on the mobile device can indicate whether they need IP session continuity, IP address reachability or no mobility support at all. The IP stack on the mobile device, in conjunction with the network, would provide the required type of mobility support, indicated by the application. This solution brings benefits for both the users and the network operators, which do not engage an extra level of service unless it is absolutely necessary. It is important to continue developing proposals in this topic to reduce the level of unnecessary mobility support, since a significant part of the sessions do not require any mobility support or it may be already provided at the application layer. It is quite important that applications can provide a high level input from the type of mobility that their sessions need. As the ultimate perspective, the mobility can be seen as a service, which is dynamically and selectively provided according to the session requirements and user agreements.

Distribution of other Mobility Management Functionalities

There has been several advances in the decentralization of the mobility management, especially in the routing optimization and distribution of mobility anchors. However, there are two main mobility functionalities not yet properly studied for distributed mobility management: AAA and billing. In order to deploy distributed mobility management in

real mobile network operators in a near future, it is vital to develop distributed strategies for AAA and billing. The authentication, the authorization and the accounting should be distributed through the network operator to optimize the handover and to cope with the distribution of mobility anchors. Billing is also one of the main problems in the distributed mobility management, which is related with all processes responsible to collect consumed data, calculate charging and billing information, produce bills to customers, process their payments and manage debt collection.

Rethink the Global Location Management

The current IP mobility solutions provide IP reachability for the users based on a conversion between a well-defined IP address and the current IP address of the user. However, most of the internet services are requested based on URL, such as HTTP Web pages, through Domain Name Server (DNS) servers. Thus, current mobility management solutions are not properly integrated with global location systems, since they usually work separated. It is important to study the integration of IP mobility solutions and location systems, in order to assess if the current user location can be directly integrated in the global location system according to the user context (e.g. through Dynamic Domain Name Server (DDNS) or a more dynamic global location systems). It may be possible to develop a single and unique identity per device/user and a respective global location system, which can be used by any application

Multiple OSI layers Supporting Mobility

There are several mobility approaches, which provide mobility support at different OSI layers. There is some work trying to define the best layer to provide mobility support, where some argue that a layer between network and transport is the best [93], while others advocate that mobility should be provided at transport layer [94]. Despite of the advance in the subject, there are no conclusions about the best layer to provide mobility support. From the knowledge gained during the development of the work presented in this Thesis, we believe that mobility management should be solved with a vertical approach of all OSI layers, and not independently solved by each layer. We identified three main issues or limitations coming from the horizontal mobility management support by layers:

- different layers providing mobility support to the same session/application may bring issues and conflicts regarding packet duplication, packet reordering and useless mobility support. The evolution is going towards smart applications and protocol from different OSI layers, which is quite prone to increase the mobility support conflict between different OSI layers. As an example, the IP mobility provides IP session continuity to an ongoing session after the handover, while the application may decide to initiate a new session to continue requiring the same content.

- different layers are able to optimize different aspects of mobility management, depending on the session demands, the mobility support should be provided at different layers, including the collaboration of multiple layers. For example, the achievement of seamless and soft handover is just possible when mobility is provided at data link layer, while routing optimization is provided by network/transport layers, and identification is provided at application layer.

- mobility should be provided from a up-bottom perspective, in the sense that if an upper layer is enough to provide the desired mobility support to a session, the

mobility support should not be provided by the lower layers. For example, if the applications are able to initiate a new session to continue requiring the same content without user notice, the layer 3 or 4 should not provide any mobility support. As another example, if the sessions are using a transport layer with mobility support (e.g. MSCTP) which is enough to provide the desire experience to the user, the network layer should not provide any IP mobility support to that session.

Bibliography

[1] Ericsson, "Voice and internet vital to city life," Ericsson, Report, June 2013.

[2] ——, "Ericsson mobility report: On the pulse of the networked society," Ericsson, Report, June 2013.

[3] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," Cisco, White Paper, February 2013.

[4] Ericsson, "5g radio access," Ericsson, White Paper, June 2013.

[5] C. Perkins, D. Johnson, and J. Arkko, "Mobility support in ipv6," IETF RFC 6275, July 2011.

[6] H. Soliman, C. Castelluccia, K. Elmalki, and L. Bellier, "Hierarchical mobile ipv6 (hmipv6) mobility management," IETF RFC 5380, October 2008.

[7] S. Gundavelli, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy mobile ipv6," IETF RFC 5213, August 2008.

[8] H. Chan, "Problem statement for distributed and dynamic mobility management," IETF, Internet-Draft draft-chan-distributed-mobility-ps-05, October 2011, work in progress.

[9] H. Chan et al., "Requirements for Distributed Mobility Management," IETF, Internet-Draft draft-ietf-dmm-requirements-10, November 2013, work in progress.

[10] D. Liu et al., "Distributed Mobility Management: Current practices and gap analysis," IETF, Internet-Draft draft-ietf-dmm-best-practices-gap-analysis-02, October 2013, work in progress.

[11] T. Condeixa, R. Matos, A. Matos, S. Sargento, and R. Sofia, "A new perspective on mobility management scenarios and approaches," in *Mobile Networks and Management*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg, 2011, vol. 68, pp. 340–353.

[12] A. Nascimento, R. Sofia, T. Condeixa, and S. Sargento, "A characterization of mobility management in user-centric networks," in *Smart Spaces and Next Generation Wired/Wireless Networking*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6869, pp. 314–325.

[13] ——, "Towards a distributed mobility management approach suitable for user-centric environments," in *ICCCN Workshops, 2012 IEEE*, 2012.

[14] T. Condeixa, S. Sargento, A. Nascimento, and R. Sofia, "Decoupling and distribution of mobility management," in *Globecom Workshops (GC Wkshps), 2012 IEEE*, 2012, pp. 1073–1078.

[15] T. Condeixa and S. Sargento, "Studying the integration of distributed and dynamic schemes in the mobility management," *Computer Networks, Elsevier*, 2014.

[16] ——, "Dynamic mobile ip anchoring," in *Communications (ICC), 2013 IEEE International Conference on*, 2013, pp. 3607–3612.

[17] T. Condeixa, J. Carvalho, S. Sargento, and R. Sofia, "Rethinking ip mobility management towards a distributed and dynamic scheme," *Transactions on Networking, IEEE/ACM*, submitted.

[18] J. Carvalho, T. Condeixa, and S. Sargento, "Distributed mobility management in dynamic environments: V2i networks," in *Computer and Communications (ISCC), 2013 IEEE International Symposium on*, 2014, submitted.

[19] T. Condeixa, L. Guardalben, T. Gomes, S. Sargento, and R. Sofia, "Make-without-break horizontal ip handovers for distributed mobility management schemes," in *Globecom Workshops (GC Wkshps), 2013 IEEE*, 2013.

[20] T. Condeixa, J. Carvalho, S. Sargento, and R. Sofia, "Dynamic offload anchoring with ip mobility," *Transactions on Network and Service Management, IEEE*, submitted.

[21] T. Condeixa and S. Sargento, "Centralized, distributed or replicated ip mobility?" *Communication Letters, IEEE*, 2014, accepted.

[22] ——, "Context-aware adaptive ip mobility anchoring," *Computer Communications, Elsevier*, submitted.

[23] H. Ali-Ahmad et al., "Mobility Anchor Selection in DMM: Use-case Scenarios," IETF, Internet-Draft draft-aliahmad-dmm-anchor-selection-01.txt, July 2013, work in progress.

[24] R. Sofia, T. Condeixa, and S. Sargento, *LNSN User-centric Networking - Future Perspectives*. Springer, 2014, ch. Mobility estimation in the Context of Distributed Mobility Management.

[25] UMM. (2013, December) User-centric mobility management. [Online]. Available: http://siti.ulusofona.pt/~umm/

[26] J. Manner and M. Kojo, "Mobility related terminology," IETF RFC 3753, June 2004.

[27] ITU-T, "Mobility management requirements for ngn," International Telecommunication Union, Recommendation Q.1706/Y.2801, November 2006.

[28] A. Vulpe, S. Obreja, and O. Fratu, "A study of mobility management using ieee 802.21," in *Electronics and Telecommunications (ISETC), 2010 9th International Symposium on*, November 2010, pp. 205–208.

[29] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "Sip: Session initiation protocol," IETF RFC 3261, June 2002.

[30] I. F. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 16–28, August 2004.

[31] S.-R. Yang and Y.-B. Lin, "A mobility management strategy for umts." in *ICOIN*, ser. Lecture Notes in Computer Science, vol. 2662. Springer, 2003, pp. 316–325.

[32] Y.-B. Lin, Y.-R. Haung, Y.-K. Chen, and I. Chlamtac, "Mobility management: from gprs to umts." *Wireless Communications and Mobile Computing*, no. 4, pp. 339–359, September 2005.

[33] B. Li, Y. Qin, C. Low, and C. Gwee, "A survey on mobile wimax," *Communications Magazine, IEEE*, vol. 45, no. 12, pp. 70–75, December 2007.

[34] C. Perkins, "Ip mobility support for ipv4, revised," IETF RFC 5944, November 2010.

[35] S. Mansor and T. Wan, "Mobility management in heterogeneous wireless access network with ieee 802.21 services," in *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, April 2010, pp. 110–114.

[36] Motorola, "Long term evolution (lte): A technical overview," 2007.

[37] I. Widjaja, P. Bosch, and H. L. Roche, "Comparison of mme signaling loads for long-term-evolution architectures." in *VTC Fall*. IEEE, 2009.

[38] J. Laganier, T. Higuchi, and K. Nishida, "Network based IP mobility management for the all-IP core network," *NTT DOCOMO Technical Journal*, 2009.

[39] R.-H. Liou, Y.-B. Lin, and S.-C. Tsai, "An investigation on lte mobility management," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 1, pp. 166–176, 2013.

[40] C. Perkins and D. B. Johnson, "Route optimization in mobile ip," draft-ietf-mobileip-optim-11.txt, September 2001.

[41] R. Moskowitz and P. Nikander, "Host identity protocol (hip) architecture," IETF RFC 4423, May 2006.

[42] P. Kikander, T. Henderson, C. Vogt, and J. Arkko, "End-host mobility and multihoming with the host identity protocol," IETF RFC 5206, April 2008.

[43] P. Kikander and J. Laganier, "Host identity protocol (hip) domain name system (dns) extension," IETF RFC 5205, April 2008.

[44] W. Xing, H. Karl, A. Wolisz, and H. Müller, "M-SCTP: Design and prototypical implementation of an end-to-end mobility concept," in *Proc. 5th Intl. Workshop The Internet Challenge: Technology and Applications*, October 2002.

[45] D. Kim, S. Koh, and S. Kim, "msctp-dac: Dynamic address configuration for msctp handover," in *Embedded and Ubiquitous Computing*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, vol. 4096, pp. 244–253.

[46] P. Behbahani, V. Rakocevic, and J. Habermann, "nsctp: A new transport layer tunnelling approach to provide seamless handover for moving networks," in *Mobile Wireless Communications Networks, 2007 9th IFIP International Conference on*, 2007, pp. 71–75.

[47] M. Riegel and M. Tuexen, "Mobile sctp," draft-riegel-tuexen-mobile-sctp-09.txt, November 2007.

[48] R. Stewart, "Stream control transmission protocol," IETF RFC 4960, September 2007.

[49] R. Shacham, H. Schulzrinne, S. Thakolsri, and W. Kellerer, "Session initiation protocol (sip) session mobility," IETF RFC 5631, October 2009.

[50] S. Deering and R. Hinden, "Internet protocol, version 6 (ipv6) specification," IETF RFC 2460, December 1998.

[51] R. Hinden and S. Deering, "Ip version 6 addressing architecture," IETF, RFC 4291, Feb. 2006.

[52] H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers," Internet Engineering Task Force, RFC 5555, 2009.

[53] R. Koodli, "Fast handovers for mobile ipv6," IETF RFC 5568, July 2009.

[54] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "Network mobility (nemo) basic support protocol," IETF RFC 3963, January 2005.

[55] H. Yokota et al., "Fast handovers for proxy mobile ipv6," Sept. 2010.

[56] J. Kim and S. Koh, "PMIPv6 with Bicasting for Soft Handover," Internet-Draft draft-jikim-bpmipv6-00, Sep. 2009.

[57] I. WG. (2013, November) Distributed mobility management. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[58] H. Chan et al., "Distributed and dynamic mobility management in mobile internet: Current approaches and issues," *Journal of Communications*, vol. 6, no. 1, pp. 4–15, Feb. 2011.

[59] K. Pentikousis and P. Bertin, "Mobility management in infrastructure networks," *Internet Computing, IEEE*, vol. 17, no. 5, pp. 74–79, 2013.

[60] J. Zuniga, C. Bernardos, A. de la Oliva, T. Melia, R. Costa, and A. Reznik, "Distributed mobility management: A standards landscape," *Communications Magazine, IEEE*, vol. 51, no. 3, pp. 80–87, 2013.

[61] J. Lee, J. Bonnin, P. Seite, and H. Chan, "Distributed ip mobility management from the perspective of the ietf: motivations, requirements, approaches, comparison, and challenges," *Wireless Communications, IEEE*, vol. 20, no. 5, pp. 159–168, 2013.

[62] P. Seite, P. Bertin, and J. Lee, "Distributed mobility anchoring," IETF, Internet-Draft draft-seite-dmm-dma-06.txt, January 2013, work in progress.

[63] P. Bertin et al., "An evaluation of dynamic mobility anchoring," in *70th IEEE Vehicular Technology Conference Fall*, Sep. 2009, pp. 1–5.

[64] ——, "Distributed or centralized mobility?" in *IEEE Global Telecommunications Conference*, Dec. 2009, pp. 1–6.

[65] H. Ali-Ahmad et al., "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *IEEE WPMC*, Sep. 2012.

[66] H. Chan, "Proxy mobile ip with distributed mobility anchors," in *IEEE GLOBECOM Workshops*, Dec. 2010, pp. 16–20.

[67] ——, "Distributed mobility management with mobile ip," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 6850–6854.

[68] H. Chan et al., "Distributed Mobility Management Framework," IETF, Internet-Draft draft-chan-dmm-framework-03, October 2013, work in progress.

[69] P. Ernest, H. Chan, and O. Falowo, "Distributed mobility management scheme with mobility routing function at the gateways," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, 2012, pp. 5254–5259.

[70] C. Bernardos and J. Zuniga, "A pmipv6-based solution for distributed mobility management," IETF, Internet-Draft draft-bernardos-dmm-pmip-03, July 2013, work in progress.

[71] C. J. Bernardos, J. Zunniga, and A. Reznik, "Towards flat and distributed mobility management: A 3gpp evolved network design," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 6855–6861.

[72] C. Bernardos and J. Zuniga, "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-03, October 2013, work in progress.

[73] W. Luo and Y. Tu, "Distributed mobility management approaches with ipv6 prefix properties," IETF, Internet-Draft draft-luo-dmm-ipv6-prefix-properties-00, July 2013, work in progress.

[74] W. Luo and J. Liu, "Pmip based dmm approaches," IETF, Internet-Draft draft-luo-dmm-pmip-based-dmm-approach-02, July 2013, work in progress.

[75] J. Korhonen, B. Patil, S. Gundavelli, P. Seite, and D.Liu, "Ipv6 prefix properties," draft-korhonen-6man-prefix-properties-02.txt, July 2013.

[76] W. Hahn, "3gpp evolved packet core support for distributed mobility anchors: Control enhancements for gw relocation," in *ITS Telecommunications (ITST), 2011 11th International Conference on*, 2011, pp. 264–267.

[77] K. Xeu et al., "Routing optimization in dmm," IETF, Internet-Draft draft-xue-dmm-routing-optimization-02.txt, June 2013, work in progress.

[78] J. Lee et al., "Host-based distributed mobility management support protocol for ipv6 mobile networks," in *IEEE WiMob*, 2012.

[79] M. Liu and Y. Wang, "Distributed Mobility Management: Service Flows Distribution and Handoff Technique based on MIPv6," IETF, Internet-Draft draft-liu-dmm-flows-distribution-and-handoff-01, September 2013, work in progress.

[80] D. Liu and C. Deng, H. Perkins, "Mobility API Extension for Distributed Mobility Management," IETF, Internet-Draft draft-liu-dmm-mobility-api-02, October 2013, work in progress.

[81] E. Nordmark, S. Chakrabarti, and J. Laganier, "Ipv6 socket api for source address selection," IETF, RFC 5014, September 2007.

[82] A. Yegin et al., "Corresponding Network Homing," IETF, Internet-Draft draft-yegin-dmm-cnet-homing-01, October 2013, work in progress.

[83] C. Xiong and J. Liu, "MN IP reachablility for the DMM," IETF, Internet-Draft draft-xiong-dmm-ip-reachability-00, September 2013, work in progress.

[84] P. Vixie, S. Thomson, Y. Rekhter, and J. Bound, "Dynamic updates in the domain name system (dns update)," Network Working Group, ISI, RFC 2136, April 1997.

[85] A. Yegin et al., "On Demand Mobility Management," IETF, Internet-Draft draft-yegin-dmm-ondemand-mobility-00, July 2013, work in progress.

[86] R. Sofia and P. Mendes, "User-provided networks: consumer as provider," *Communications Magazine, IEEE*, vol. 46, no. 12, pp. 86–91, December 2008.

[87] (2013, December) Fon. [Online]. Available: https://corp.fon.com/en

[88] E. Gustafsson and A. Jonsson, "Always best connected," *Wireless Communications, IEEE*, vol. 10, no. 1, pp. 49–55, 2003.

[89] (2013, December) Network simulator 2 - ns-2. [Online]. Available: http://nsnam.isi.edu/nsnam/

[90] ns 3. (2013, December) ns-3.16. [Online]. Available: http://www.nsnam.org/ns-3-16

[91] SUMO. (2013, December) Simulation of urban mobility. [Online]. Available: http://sumo-sim.org

[92] T. M. Inc. (2013, December) Matlab version 7.10.0 (r2010a). [Online]. Available: http://www.mathworks.com/

[93] M. Ratola, "Which layer for mobility? - comparing mobile ipv6, hip and sctp," *HUT T-110.551 Seminar on Internetworking*, 2004.

[94] W. M. Eddy, "At what layer does mobility belong?" *Comm. Mag.*, vol. 42, no. 10, pp. 155–159, October 2004.

Appendix

The articles included in appendix have been reformatted, in order to increase their readability, while maintaining all the respective content unchanged.

Paper A

A Characterization of Mobility Management in User-centric Networks

Andréa Nascimento, Rute Sofia, Tiago Condeixa and Susana Sargento
in *11th International Conference on Next Generation Wired/Wireless Networking*, St. Petersburg, Russia, August, 2011

A Characterization of Mobility Management in User-centric Networks

Andréa Nascimento, Rute Sofia, Tiago Condeixa and Susana Sargento

Abstract

Mobility management is a key aspect to consider in future Internet architectures, as these architectures include a highly nomadic end-user which often relies on services provided by multi-access networks. In contrast, today's mobility management solutions were designed having in mind simpler scenarios and requirements from the network and where roaming could often be taken care of with previously established agreements. With a more dynamic behavior in the network, and also with a more prominent role from the end-user, mobility management has to deal with additional requirements derived from new Internet paradigms. To assist in understanding such requirements and also how to deal with them, this paper proposes a starting point to dismantle current mobility management notions. Our contribution is an initial proposal on defining mobility management in concrete functional blocks, their interaction, as well as a potential grouping which later can assist in deriving novel and more flexible mobility management architectures.

Keywords: Wireless networks, mobility management, user-centricity

A.1 Introduction

Internet services and models have been going through a paradigm shift, product of three main factors: i) widespread wireless technologies; ii) increasing variety of user-friendly and multimedia-enabled terminals; iii) wider availability of open-source tools for content generation. Together, these three factors are changing the way that Internet services are delivered and consumed as there is a trend where the end-user has a particular role in controlling content as well as connectivity, based upon cooperation. These spontaneous environments, *user-centric* networks, rely on the notion that Internet users carry or own devices that may be part of the network. Hence the human roaming behavior of each user, be it individually or from an aggregate perspective, directly impacts the way the network is operated and perceived.

Human movement patterns in these environments may exhibit high variability as they are based on individual users routines and on users interests towards targets (e.g. locations, other users). Hence, mobility management is required to ensure adequate connectivity models and adequate network operation to support end-user expectations towards his/her roaming services. Considering the dynamics of user-centric networks and its self-organizing nature, it is crucial to attempt to develop end-to-end mobility management solutions more flexible than the ones existing today, as user-centric wireless networks are starting to heavily populate Internet fringes.

Currently, the most popular solutions for global mobility management have in common a model where a centralized and static *mobility anchor point* is responsible for keeping some form of association between previous and current identities of a mobile node. In user-centric environments, as explained, there is the need to better understand the roles that a mobility anchor point can have; the best location for these elements; and efficient ways to select the best anchor point for a mobile node. Moreover, considering that user-centric

environments are heavily based on the users interests on being part of the network, and also assuming that the users might also control management functionalities, the period of time a mobility anchor point may or may not be available is highly variable. This poses extra stress on seamless and centralized mobility mechanisms, which have to manage handovers more often.

The aim of this paper is to provide an initial analysis of aspects that have to be considered when attempting to make end-to-end mobility management schemes more flexible. Our expectations are to contribute to an out-of-the-box notion of mobility management, by splitting mobility management as a whole into concrete functional blocks, and by explaining their impact and how to group such blocks. Our model is based on centralized solutions which, independently of the OSI Layer they tackle, are based in the same principles, roles, as well as similar operational behavior. Such splitting and categorization will give rise, in our opinion, to new mobility management architectures which are user-centric and more flexible.

The paper is organized as follows. In section A.2 we describe related work, explaining the contributions that our work provides. Section A.3 provides a few examples on user-centric networking scenarios, including for each a brief mobility characterization. In sections A.4 and A.5 we describe our study on mobility management, which is a characterization based on the current needs of this emerging user-centric networks, and in section A.6 we conclude this work.

A.2 Related Work

This section provides a brief description on current work related to mobility management proposals which are based on different perspectives than the standardized solutions.

Bolla et al. consider the application of overlays to deal with mobility from a global perspective [1]. They provide a distributed mobility management scheme where mobility anchor points may be located within customer premises. The mobility anchor point itself is still a centralizing element as all the signaling goes through this device. Following the same line of thought, in order to deal with personal mobility and session migration Bolla et al. propose an application layer mobility framework [2] and the usage of a personal address, *"a network identifier dinamically assigned to a specific user for a specific communication session"*. The framework performs functions of personal mobility, terminal handover, session migration, and media adaptation for interactive multimedia applications. Although the authors are focused on addressing specific aspects of environments involving media, they do not attempt to analyze how to globally make mobility management more flexible.

Sofia et al. [3] propose an approach whose main objective is to separate control and data functionalities from the mobility anchor point into two different elements, in order to provide a more flexible mobility management framework, and to assist in developing non-centralized (e.g. distributed or hierarchical) mobility architectures. However, the authors do not present a proposal on how the communication between those separated elements can be performed, nor an analysis on why such splitting was relevant.

Chan [4] proposed the splitting of a mobility system into three logical functions: home network prefixes allocation, location management and mobility routing. The approach is based on the *Proxy Mobile IPv6* [5] extension for *Mobile IPv6* [6], and it is also proposed the usage of two mobility anchor elements, called *Home Mobility Anchor* and *Visited Mobility Anchor*. The main objective is to provide a system with mobility anchors distributed over different networks.

Having in mind the recent trend of flatter mobile network architectures, *Dynamic Mo-*

bility Anchoring [7] [8] addresses the concept of "flattening" by confining mobility support in the access network, e.g. only confining it to access routers through a specific implementation of the application of Proxy Mobile IP. Following the same line of thought, i.e. IP mobility management in flatter mobile networks, Chan [9] describes the differences between centralized and distributed mobility management systems, as well as a list of potential problems and limitations of a centralized approach when compared with a distributed one.

Condeixa et al. [10] analyzed mobility management assumptions and requirements in user-centric scenarios, debating on challenges that need to be addressed to obtain a global mobility management solution considering user-centricity. The authors point out three major concerns for a mobility management system: binding definition, binding maintenance, and forwarding data problem.

Our work has in common with these approaches the motivation that by splitting, decentralizing, or decoupling mobility management functionality into different blocks may assist in better understanding how and where to manage mobility. As described, most of today's attempts of flattening mobility management are being applied in the evolved packet core being the sole reason the urgent need to simplify mobility management. We believe that understanding on how such mechanism may work is key to give rise to new research and business opportunities.

A.3 User-centric Networking Notions

User-centric networks are environments where an Internet end-user owns and often carries devices that can share Internet access. These environments and the amount of end-user devices sharing Internet access are expected to grow, despite the limitations imposed by traditional operator-driven Internet communication models.

In our study, mobility management aspects are addressed from an end-to-end perspective but the analysis is applied in user-centric spontaneous wireless environments, which today correspond to the majority of technical scenarios on the last hop towards the end-user. Our user-centric environments are located within the customer premises region (where residential households, and enterprise environments reside). While in contrast, today's mobility management relies on functional blocks that are on the access or service regions.

Out of the several possible user-centric scenarios, we consider here three: a regular hotspot, a user-provided network (UPN) and a delay tolerant network (DTN). Each scenario is described both from an architectural perspective, as well as from a mobility characterization perspective. The line of thought driving this analysis is that these representative scenarios hold different requirements and are based on specific mobility assumptions. Hence, after providing a mobility characterization for each of the scenarios, the section concludes with a discussion which shall result in the identification of mobility functionality blocks, based on common requirements that each of these scenarios attain. A more complete description of these and of additional user-centric scenarios can be found in [11]:

- **Hotspot**: a hotspot scenario corresponds to the regular infrastructure mode in Wireless Fidelity (Wi-Fi) environments. This is currently the most common wireless architecture being deployed around us: each Internet enabled household corresponds to one hotspot. In this scenario mobility of users is local and confined to small regions, e.g. a room, an apartment, a small office. Moreover, if the user moves across different *Access Points (APs)*, then connectivity is expected to be intermittent. In

a generic hotspot scenario users' mobility speed is low (pedestrian). Mobility inside each hotspot scenario is mostly managed at OSI layer 2; however, the IP address of the active user equipment's interface can change after a break. A key aspect to consider is that if current mobility management solutions are applied to this scenario, despite the fact that most of the movement is local, the mobility anchor point is located on the access or service regions.

- **User-provided networks**: UPNs [12] have been applied as complement to existing access networks: they allow expansion of infrastructures across one wireless hop. There is usually one individual or entity (the *Micro-Provider, MP*) which is responsible for sharing his/her connection with N-1 other users (out of a universe of N users, who today belong to a single community). Moreover, a user is, in a specific community, simply identified by a virtual identifier (usually, a set of credentials username and password) which is stored by a *Virtual Operator (VO)* and relied upon whenever the user decides to access the Internet by means of a specific community hotspot. In these emerging architectures, the nodes that integrate the network are in fact end-user devices which may have additional storage capability and sustain networking services. Such nodes, being carried by end-users, exhibit a highly dynamic behavior. Nodes move frequently following social patterns and based on their carriers interests. The network is also expected to frequently change (and even to experience frequent partitions) due to the fact that such nodes, being portable, are limited in terms of energy resources.

- **Delay tolerant networks**: The DTN scenario relates to the need to establish on-the-fly an autonomous network within a disaster region (e.g. after an earthquake) based upon the devices that users in the region control and carry. Hence, such DTN consists of a network composed by users with a common objective (a community), grouped in regions. Some nodes move from region to region, establishing the communication between them (since gateways are mobile). Considering the main purpose of this kind of network, and the specific type of scenario where it is deployed, it is possible to establish behavior patterns on the mobility of the nodes, making possible to predict their location in a given instant and to schedule the delivery of information. In this case, the mobility pattern may also impact the routing process. Users moving may be good candidates to act as gateways, because they present a higher possibility of reaching other regions. It is important to notice that a region may be composed by only one user.

Table A.1 summarizes the main characteristics related to the scenarios described, concerning inherent characteristics, and mobility behavior of the users on each of the scenarios presented. Based on a detailed analysis of the scenarios described [11] we consider a set of parameters that should be taken into account when characterizing any mobility management scheme: *i) identification*, which stands for the device identification both from a user and an access perspective; ii) *network scope*, which relates to the reach of the network; iii) *access control*, which relates to the location of the access control mechanism that is normally applied in each scenario; iv) *movement patterns*, related to the pattern that nodes are expected to exhibit in each scenario when roaming; v) *pause time behavior*, related to the time that a node exhibits a speed that is zero or close to zero; vi) *handover frequency*, related to the node having to switch between different networks or attachment points; vii) *connectivity sharing*, related to the sharing of Internet access.

In table A.1 we provide a brief analysis on how each of the mentioned parameters relate to the three scenarios described. UPNs stand for a relevant case to address in

terms of mobility management, as this scenario exhibits features that are not available on the hotspot scenario. The same conclusion can be drawn by looking at the DTN characterization. Both UPNs and DTNs exhibit aspects that were not considered when devising the current (centralized) mobility solutions.

Table A.1: Summary of mobility characterization across user-centric scenarios.

Scenario/ Parameters	Hotspot	UPN	DTN
Identification	MAC address, credentials managed by WISP	Trust management scheme community credentials	Tokens or certificates; public/private key pair
Network scope	Small environment, e.g. household shops, universities	Small-large, e.g. household to village/city; varies dynamically	Small-large but static does not exhibit a quick growth
Access control	Centralized, on the provider	Decentralized and spontaneous	Decentralized
Node speed	Low	High	Varying
Expected movement frequency	Low	High and global	Low and routine based
Mobility pattern	Local mobility; preferred locations	Human/social patterns; short distance traveling preferred	Local mobility social patterns
Pause time	Long pause times	Mix, depends on location and user routine	Long
Handover frequency	Low	High	High
Connectivity	None	Yes	Yes

A.4 Defining Mobility Management: A Characterization

This section is dedicated to a proposal on a global architectural definition of mobility management functional blocks, as well as roles based on the scenarios previously described.

A.4.1 Elements and Roles

In a mobility management system, three elements are considered in related literature: the *Mobile Node (MN)*, an end-user device for which a mobility service is provided; a *;Mobility Anchor Point (MAP)*, the element responsible for providing the mobility management service, it may reside in the network (e.g. router or access element) or in a server; and the *Correspondent Node (CN)*, that is any element engaged in active communication with the MN. These are generic roles that are today present in different management solutions, independently of the OSI Layer where the solution resides. For instance, in MIP [6] the MAP is the Home Agent (HA). In the *Session Initiation Protocol (SIP)* [13] the MAP

is the SIP server. In a 3GPP architecture, the mobility anchor is centralized and located in the core network, having all traffic flowing through it, even if services to be used are locally placed closer to the MN.

Towards the idea of making mobility management more flexible (being the aim a reduced operational cost) Seite et al. and Chan et al. suggest to position the mobility anchors closer to the mobile nodes [8], ideally in the first element visible on the path from a MN perspective [9]. Sofia et al. proposed the separation of management functionalities into two elements, attempting to decouple data plane and control plane [3]. In the proposed architecture, the HAC (control plane element) is located in a server, and HADs (data plane elements) are positioned in the access nodes, close to mobile nodes. Chan relies on the Proxy Mobile IP [5], and also splits the mobility anchor functionalities into three logical blocks [9]. Although the author states that those functionalities are placed in the home network, they do not need to be placed in the same physical entity. Those works can be considered as a first step towards an architecture where the management functionalities are splitted and distributed in different places in the network.

Such approaches, the positioning of the MAP as well as the definition of interactions between the different roles of mobility management have been object of heavy analysis. Still, today there is not truly consensus in where MAP and additional functionality should reside. Such positioning depends on the network architecture and requirements; on the OSI Layer being tackled, as well as on the overall complexity from a technical and policing perspective. Considering that user-centric networks present particular characteristics (e.g. there is no clear splitting between network elements and end-devices), the current centralized standards may not be suitable. Thus, a novel mobility management approach should be designed for such networks, considering all its particularities and following this trend of rethinking the mobility anchor point element.

Therefore, thinking about mobility management functioning in a fine-grained way, we have identified a group of functionality blocks. Based on the dynamics of user-centric networks, the first step towards a more suitable mobility management approach is by understanding and further analyzing the basic tasks a mobility management should provide.

A.4.2 Functional Blocks

In order to perform a mobility management characterization, as result of an initial analysis on current available mobility management approaches and standards, we have identified the following mobility management functional blocks:

- **Device identification:** corresponds to the network identification for the MN. Usually the main mechanism for a location management is the association between the device's *known-address* and the device's *real-address*. In MIP, known-address and real-address are IP addresses; in SIP, the known-address is a URI, and the real-address is an IP address. In MIP the device identification control is the Home Agent (HA)/Correspondent Node (CN) cache binding. In SIP, it is the user database used by the Proxy server.

- **Identification database control:** corresponds to the mechanism that is applied to control the database identification. This is normally a block relevant from an access perspective, which today follows a centralized approach.

- **Binding mechanism:** it is the signaling related to the device's register to the mobility system. It creates/updates a record in the identification database control, associating the known-address to the real-address. In MIP it is the Binding Update

message sent to a HA/CN. In SIP it is the REGISTER message sent to the Registrar server.

- **Routing or forwarding:** it is the process of intercepting the packets destined to the known-address, encapsulating them with the real-address, and forwarding them. In MIP this is performed by the HA; in SIP this process is performed by an element named RTP translator (when it is used).
- **Handover negotiation:** the process taken when the device has its real-address changed. It involves negotiation and signaling. The main objective is to guarantee that the user will keep active all its sessions during the handover process. In MIP, the handover negotiation may be anticipated with the Fast Handover extension [14], and the SIP does not implement any anticipation, performing a re-negotiation after the connection between the peers is lost.
- **Resource management:** the resource management is a necessary procedure for the mobility management to guarantee the quality of the connection when the MN changes its point of attachment to the network. However, it is not provided by most of the mobility management approaches. The 802.21 Media Independent Handover (MIH) [15] standard is focused on the handover process based on a resource management aware negotiation for vertical handovers.
- **Mobility estimation:** it is the procedure of changing the MN point of attachment to the network before its current connection breaks. The extension Fast Handovers for MIP, and the 802.21 MIH provide this functionality.
- **Security/privacy:** it refers to any security or privacy mechanism used to assure the integrity of the elements and signaling in the mobility management system.

A.4.3 Discussion on Mobility Characterization

Based on the block characterization there are a few aspects worth to highlight. Firstly, today's mobility management solutions completely ignore the need for adequate resource management. However, this is a crucial aspect for cellular or wireless networks, in particular for session continuity. Database control is normally centralized, an aspect which may not be compatible with the notion of communities that user-centric networks embody. Routing and forwarding is also based on mechanisms (e.g. proxy mechanisms) which may not be completely compatible with the fact that users in our scenarios are expected to roam frequently. This is an aspect that can be improved by integrating mobility estimation mechanisms. Security and privacy aspects are also often disregarded.

Moreover, analyzing the identified blocks, one can notice that there are a few categories onto which they seem to be naturally grouped. Firstly, they can be grouped into *data plane* and/or *control plane*. It is also possible to group the functionality blocks into *location management* and/or *handover management* procedures.

These are aspects that we debate on the next section in an attempt to raise awareness to new and more flexible mobility management schemes.

A.5 Deconstructing Mobility Management Centralized Approaches

This section delves into the potential development of a mobility management architecture that is more adequate to the emerging wireless scenarios described in section 3.

As of today the functional blocks described reside both on the MN and mobility anchor point, being the functionality fully controlled in the later one, which is physically located in the access or service regions. Our aim in analyzing initial forms of deconstructing the need for a centralized mobility management scheme is motivated by the need to find simple and operational ways to split such functionality, as well as ways to "push" such functionality closer to the end-user, having in mind an optimization of mobility management in the context of the scenarios described.

A.5.1 Location and Handover Management Categorization

Mobility management usually is mentioned as consisting of two main blocks: location management and handover management. Location management is the block responsible for locating the devices, i.e. for guaranteeing that they are always reachable, independent of their point of attachment to the network. The handover management block is responsible for maintaining active sessions while MNs roam. Therefore, from a high level perspective, mobility management functionality can be split into these two main blocks. Today, these blocks both reside on the mobility anchor point and are based on information provided by the MN. Solutions such as the *Host Identity Protocol (HIP)* [16] attempt to provide a decoupling by isolating location management and handover management. Other solutions (e.g. Hierarchical Mobile IP [17]) optimize handover management by scoping the extent of the impact of such negotiation.

A.5.2 Control and Data Plane Categorization

Another way to categorize mobility management functionality is to consider a splitting between control and data planes. As part of the control plane we can cite all the procedures related to the signaling, and the data plane is related to the data traffic, routing, forwarding and address translation. Figure A.1 shows the relationship between the blocks, in order to identify the communication between them. It shows also the classification concerning data and control planes, and location and handover management.

Between the functional blocks, it is possible to identify two types of communication, in regards to its periodicity. *Periodic communication* is related to procedures that need to be performed in a regular basis, in order to maintain the system updated. The *occasional communication* is related to the procedures performed only as result of a change in the system, for instance, when a MN performs a handover from one point of attachment to another.

Usually, all the communication between the blocks of the handover management side of the picture is triggered when a node movement is detected, or predicted. When a handover is detected, the mobility estimation block triggers the handover negotiation, which will take part in the process. The handover negotiation needs to consult the resource management in order to guarantee that the user will be "always best connected". For the handover process to complete, the binding mechanism is triggered, so it can update the location information in the identification database control. The identification database control then updates the information in the element responsible for routing/forwarding.

The binding mechanism has a periodic communication with the ID database control, because it is the procedure performed to maintain the ID database control updated. It needs to use the security/privacy procedures to guarantee that no third part could take place in the communication.

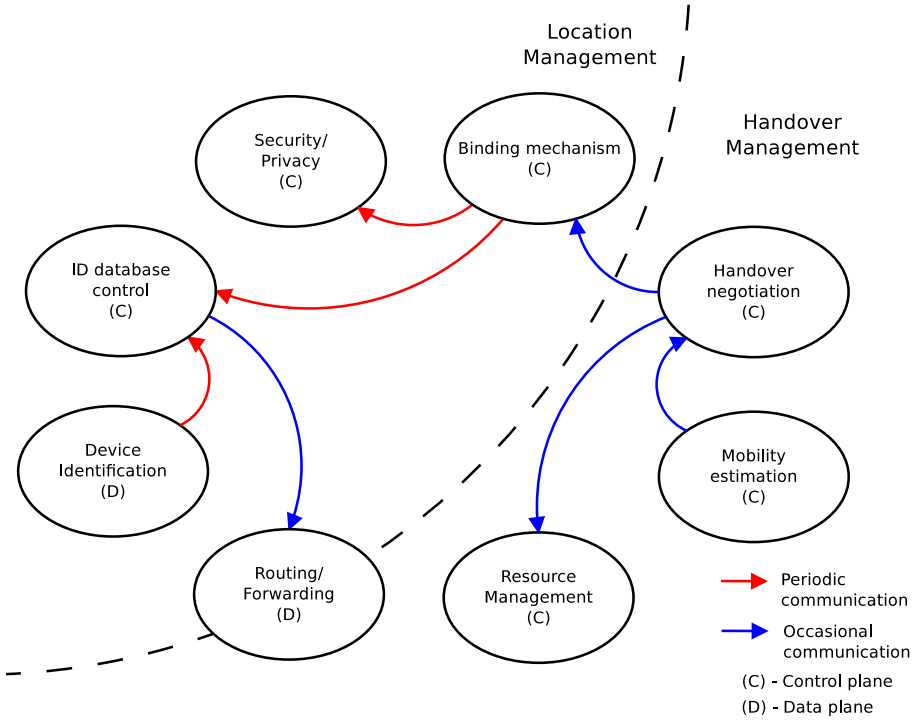


Figure A.1: Mobility management functional blocks.

Table A.2: Location of mobility management functional blocks.

Parameter	Access and user perspective categorization
Device Identification	User
ID database control	Access
Binding mechanism	User and access
Routing / Forwarding	Access
Handover negotiation	User
Resource management	Access and user
Security/privacy	User
Mobility estimation	Access and user

A.5.3 User Perspective and Access Perspective Categorization

Currently, the available mobility management approaches offer most of the functionalities described here, but none of those approaches offer all of the functionalities. Those functionalities are placed in different locations in the network and customer premises, and most of them are centralized in one unique element (usually the mobility anchor point). By taking this perspective, we can categorize the blocks into two groups, blocks located in the user perspective and in the access perspective as provided in table A.2.

Table A.2 shows the current location of each block. It is important to notice that this location is based on current mobility management approaches functioning.

A.6 Conclusion

This paper provides a study and a new perspective on ways to make end-to-end mobility management schemes more flexible, being the motivation the fact that user-centricity

and in particular user-centric environments are a crucial part of the future of the Internet. We went over three different cases of spontaneous wireless deployments abounding around us, and characterized each from a mobility perspective. Based on such characterization we have derived a set of parameters and functional blocks, and discussed ways to attempt to de-construct the need for centralized architectures, starting by proposing concrete categories to tackle.

As follow-up of this work we intend to take advantage on the blocks identification and data/control planes and location/handover management categorizations to evaluate what is the best location for each of the identified functional blocks. Focusing on the user-centricity, the objective is to perform a deeper study on each of those functionality blocks, in order to identify which of them could be placed into customer premises equipment. Placing mobility management functionalities in the customer premises could provide a mobility system user-centric and independent of the access network. A deeper study should clarify if that is possible, and what is the cost to maintain such approach. Hence, as next steps we intend to address ways to bring mobility management closer to the customer premises in a way that is adequate for the network, while keeping the end-user agnostic in regards to the complexity. A second step to be considered is to analyze such splitting based on the potential impact that it may have both from an end-user and from an access perspective.

A.7 References

[1] R. Bolla, A. Ranieri, R. Rapuzzi, and M. Repetto, "Moving towards user-centric paradigms for internet mobility," in *International InterMedia Summer School*, June 2009.

[2] R. Bolla, R. Rapuzzi, and M. Repetto, "A user-centric mobility framework for multimedia interactive applications," in *Wireless Communication Systems, 2009. ISWCS 2009. 6th International Symposium on*, September 2009, pp. 293–297.

[3] R. Sofia, A. Hof, and S. Wevering, "Method for packet-based data transmission in a networkhaving mobility functionality," Patent No. EP1 883 196, Siemens AG., January, 2008.

[4] H. Chan, "Proxy Mobile IP with Distributed Mobility Anchors," *IEEE Globecom Workshop on Seamless Wireless Mobility*, 2010.

[5] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. (2008, August) Proxy Mobile IPv6 - RFC5213. [Online]. Available: http://www.ietf.org/rfc/rfc5213.txt

[6] D. Johnson, C. Perkins, and J. Arkko. (2004, June) Mobility Support in IPv6 - RFC3775. [Online]. Available: http://www.ietf.org/rfc/rfc3775.txt

[7] P. Bertin, S. Bonjour, and J. M. Bonnin, "An evaluation of dynamic mobility anchoring," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, September 2009, pp. 1–5.

[8] P. Seite and P. Bertin, "Dynamic mobility anchoring," draft-seite-netext-dma-00.txt, May 2010.

[9] H. Chan, "Problem statement for distributed and dynamic mobility management," draft-chan-distributed-mobility-ps-02, March 2011, (Work in Progress).

[10] T. Condeixa, R. Matos, A. Matos, S. Sargento, and R. Sofia, "A new perspective on mobility management: Scenarios and approaches," in *Second International ICST Conference on Mobile Networks and Management*, September 2010.

[11] R. Sofia, A. Nascimento, S. Sargento, T. Condeixa, and R. Matos, "User-centric mobility management - D1: Use-cases," Lusófona University, Tech. Rep., March 2011.

[12] R. Sofia and P. Mendes, "User-provided networks: Consumer as provider," *IEEE Communications Magazine*, vol. 46, pp. 86–91, 2008.

[13] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. (2002, June) Session Initiation Protocol - RFC3261. [Online]. Available: http://www.ietf.org/rfc/rfc3261.txt

[14] R. Koodli. (2009, July) Mobile IPv6 Fast Handovers - RFC5568. [Online]. Available: http://www.ietf.org/rfc/rfc5568.txt

[15] IEEE 802.21 Working Group. [Online]. Available: http://ieee802.org/21/

[16] R. Moskowitz and P. Nikander. (2006, May) Host Identity Protocol (HIP) Architecture - RFC4423. [Online]. Available: http://www.ietf.org/rfc/rfc4423.txt

[17] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier. (2008, October) Hierarchical Mobile IPv6 (HMIPv6) Mobility Management - RFC5380. [Online]. Available: http://www.ietf.org/rfc/rfc5380.txt

Paper B

Studying the Integration of Distributed and Dynamic Schemes in the Mobility Management

Tiago Condeixa and Susana Sargento
in *Elsevier Computer Networks*

Studying the Integration of Distributed and Dynamic Schemes in the Mobility Management

Tiago Condeixa and Susana Sargento

Abstract

The raise of demanding multimedia content and the increasing number of mobile devices originated a rapid growth of mobile Internet traffic, which is expected to continue increasing with an exponential behavior in the next years. In order to cope with this rapid increase, service providers are already developing new strategies, such as the selective traffic offloading through the wireless local area networks. Moreover, a new trend is to flatten the network architectures for mobile Internet, and hence, IP mobility management protocols need to be adapted for such evolution. However, current mobility management models rely on a centralized entity, called mobility anchor, which routes the whole data traffic and manages all bindings of its users. With the increase of the mobile Internet traffic and the number of users' devices, such centralized models encounter several barriers for scalability, security and performance, such as a single point of failure, longer traffic delays and higher signaling loads. Hence, we study the distribution of mobility management based on the decoupling of functionalities into: handover management, location management and data management. We evaluate distinct approaches to distribute the mobility functionalities closer to the end-user. We demonstrate, through analytical and simulation results, that distributed mobility management approaches improve the data delivery when compared with current centralized models.

Keywords: Distributed Mobility Management Decoupling of Mobility Functionalities Handover Management Data Management

B.1 Introduction

With the evolution of the society towards a mobile environment, the importance of the mobility management in the network has been increased. Mobility management is responsible to maintain the ongoing communications while the user roams among distinct networks, and to provide reachability of the mobile device in new communications. While moving and attaching to heterogeneous networks, the user desires to maintain the quality of the required services. Users are requiring more demanding mobile multimedia services everywhere and anytime, which consume a great part of available network resources and poses an extra stress in the mobility management. Operators statistics show that the usage of mobile data traffic has doubled during the last year, and this is expected to continue in this decade [1] [2], resulting in an explosion in mobile Internet traffic. Thus, service providers are already developing new strategies, such as selective traffic offloading through wireless local area networks, in order to deal with traffic that exceeds the available capacity. However, the mobility support needs also to be guaranteed to the mobile device when the user communicates through different wireless local area networks.

Moreover, users have been playing a more active role in communications, controlling connectivity and content in cooperative environments. They develop spontaneous wireless networks, simply based on cooperation and access sharing on particular communities. Such user-centric environments raise new challenges to the traditional and tightly controlled

mobility management schemes. Moreover, a more flattened network architecture for mobile Internet is anticipated to meet the needs of increasing traffic from the mobile users and to reduce costs in the core network. To accomplish these trends, there has been a paradigm shift in users traffic behavior with the increase of communication between devices in the same geographical area due to the migration of content servers closer to the end-user, such as Content Delivery Networks. However, current mobility management models have been developed for centralized networks, such as Mobile IPv6 (MIPv6) [3] and Proxy MIPv6 (PMIPv6) [4], which brings several limitations when applied to recent trends [5][6]. In current centralized models there is a central and static entity, called Mobility Anchor (MA), which is in charge of the mobility management functionalities of a large number of Mobile Nodes (MNs), regarding data, context information and signaling. All data traffic traverses the centralized MA, such as the Home Agent (HA) in MIPv6 and the Local Mobility Anchor (LMA) in PMIPv6, and all bindings are managed at this MA as well. As the number of MNs increases and the mobile data traffic explodes, such centralized architectures may encounter several problems. First, the routing is performed via the centralized mobility anchor, which is often longer. This increases the operational cost of the network, the consumption of core network resources, and the end-to-end delay of applications. Moreover, the centralized mobility anchor manages the mobility context and mobility routes of all MNs, which may increase the mobility signaling and handover latency. The adoption of current centralized models forces a static mobility support, independently of the mobility requirement in the MN. Current mobility models always provide mobility support to MNs' sessions while the MN is connected to the network, even for a session initiated and terminated in the same network. Furthermore, a centralized point is commonly more vulnerable to failure or attack.

It is therefore of major importance to re-think mobility management from an out-of-the-box perspective, and in particular, to consider the distribution of mobility management and how it can assist the individual user and the provider in terms of mobility coupled to the day-to-day living of Internet users. Accordingly, the IETF charted recently the Distributed Mobility Management (DMM) working group [7], where various efforts from both industry and academia are being performed on specifying DMM schemes.

In this article we start by studying the impact of distributing the mobility management functionalities. Our aim is to assess the main guidelines for a distributed mobility management architecture. Our first approach, presented in [8], compares different distributed mobility management schemes based on the decoupling of the mobility functionalities into data management, handover management and location management. We extend the presented study on improving the description of the evaluated approaches with examples, and on proposing an analytical model to improve the evaluation of the approaches with data and signaling costs. Moreover, we perform a more exhaustive evaluation through simulations with different scenarios, comparing both analytical and simulation results.

The article is organized as follows. The related work is briefly presented in Section II. Section III explains the decoupling of mobility management functionalities, while Section IV describes the distributed mobility management approaches. Section V describes analytical models to evaluate the distributed approaches and the centralized one. Section VI evaluates them through the analytical models and simulation. Finally, Section VI concludes the paper and introduces the future work.

B.2 Related Work

The currently adopted IPv6 host-based mobility protocol, called MIPv6 [3], was envisioned to provide global IPv6 mobility to the user's device without any support from the network infrastructure. In MIPv6, the data packets are always routed via the HA, which encapsulates them to the current IPv6 address of the MN, in its current location. The HA maintains a binding between the well-known IP address of the MN and the IP address obtained from the current IP network.

The currently adopted IPv6 network-based mobility protocol, called PMIPv6 [4], was developed with a different idea than MIPv6: the network is responsible to provide IP mobility support transparent to the user. The user's mobile device does not require any mobility support, since the Mobile Access Gateway (MAG) implements the same functionalities as the MN in MIPv6. The mobility functionalities are moved from the MN to MAG, which are usually implemented in the Access Routers (ARs). The MAG is responsible to detect the MN handover and signal the centralized mobility management entity, called LMA, with the new network of the MN. The tunnels are established between the LMA and MAGs, being the MN agnostic to the movement.

Hierarchical Mobile IPv6 (HMIPv6) [9] is an attempt to optimize the micro-mobility of MIPv6 through the introduction of a hierarchical level for the management of the bindings and traffic forwarding. It introduces a new entity placed closer to the user's mobile device, called Mobility Anchor Point (MAP), that creates an IP abstraction level in its coverage domain with two IPv6 addresses. The MAP maintains the updated binding of the MN, forwarding data packets to MN while it remains in its coverage domain, reducing signaling cost and handover latency for micro-mobility. The HA and MN remains with the same mobility management functionalities.

There is an increasing interest in the topic of distributed and dynamic mobility management, specially in the problem statement and definition of guidelines for a base distributed mobility management. This increasing research on the distribution of mobility management led to the creation of the IETF Distributed Mobility Management (DMM) working group [7], where several authors are contributing to provide novel work on the topic. Following the idea of splitting the mobility management, the authors in [10][5][6] analyze the issues of the centralized approaches when integrated in flatten network architectures. They highlight the main problems of current mobility management approaches, and propose the initial requirements for a distributed and dynamic mobility management.

A decoupling of mobility management into well-defined functional blocks was proposed in [11], which identified the interactions between blocks, as well as a potential grouping, which later can assist in deriving more flexible mobility management architectures.

Having in mind the idea of dynamic mobility anchoring, [12] and [13] proposed a partial-distributed mobility management approach for flat IP architectures, where mobility management is split into data and control planes. While control plane is maintained in a centralized element, data plane was placed closer to the end-user, in the access nodes. The benefits of this design over centralized mobility management were shown in [14] and [15].

Another approach for distributed mobility anchoring, based on PMIP, is proposed in [16]. Mobility sessions are anchored at the last IP hop router (distributed gateway) in order to allow session continuity in inter-domain scenarios.

Two other distributed mobility control schemes were proposed in [17], based on PMIP protocol: Signal-driven PMIP (S-PMIP) and Signal-driven Distributed PMIP (SD-PMIP). S-PMIP is a partially distributed scheme, in which the control plane is separated from

the data plane, and the optimized data path is direct between the Correspondent Node (CN) and the MN. SD-PMIP is a fully distributed scheme, where each MAG will multicast a Proxy Binding Query message to all MAGs in its local PMIP domain to retrieve the Proxy-CoA of the MN.

A host-based distributed mobility management approach is proposed in [18]. The approach distributes the mobility anchors at the access network level and a MN is served by a close-by mobility anchor. The MN changes its point of attachment to the Internet, maintaining the previous IP addresses used in the ongoing communications, which were established with the previous IP addresses. The performance of the proposed approach is compared through analytical models with MIPv6 in terms of handover latency and throughput. It is also showed in [19] how the proposed host-based distributed mobility management protocol might alleviate the traffic loads in the evolution of the 3GPP network, and accelerate the transmission speed of certain flows by offloading them through WiFi networks.

Although there are some approaches for the distribution of mobility management, there is no study on the performance of distributed mobility according to different schemes. The focus of this article is on the study of distinct schemes to distribute the mobility management functionalities, based on the decoupling of the mobility management functionalities into Data Management, Handover Management and Location Management.

B.3 Decouple of Mobility Management

In a recent vision [7], mobility management has been split according to the control and data planes. As part of the control plane, we can consider all procedures related to the signaling; the data plane is related to the data traffic forwarding through encapsulation/decapsulation and address translation. Another possibility to categorize mobility management functionalities is to consider a decoupling between location and handover management. Location management is responsible to guarantee that MNs are always reachable, independently of their points of attachment in the network. The handover management is responsible for maintaining ongoing sessions while MNs roam.

After a careful analysis of mobility management blocks and their interactions [11], we propose a decoupling of mobility management into the following functionality blocks (Figure B.1): Handover Management, Data Management and Location Management. This approach goes further the division of mobility management into control and data planes, splitting the control plane into location and handover management. Separating the handover management brings a new level of flexibility to deal with a MN's roaming, since different control signaling/decisions are performed at different elements.

Location Management: maintains the reachability of MNs independent of the MN location or connected network. It has associated an identification database, containing bindings with MNs' Identification and its current IP address. An application should be able to interact with location management to request the IP address of a MN based on his Identification. Location management has to be updated with current IP address of a MN through the handover management functionality.

Data Management: is responsible for encapsulation of data packets through address translation. The data management functionality intercepts data packets, decapsulates them if needed, and then encapsulates them with a new IP header from the address translation rules. Data management does not provide any signaling; it only receives signaling from handover management.

Handover Management: maintains sessions active when a MN roams between net-

works, so it provides handover detection and negotiation at IP layer, being responsible for the signaling that communicates with data and location management after an IP handover. Another function of handover management is to maintain the mobility context and routes, such as the routers with the data management and respective IPs of the MN.

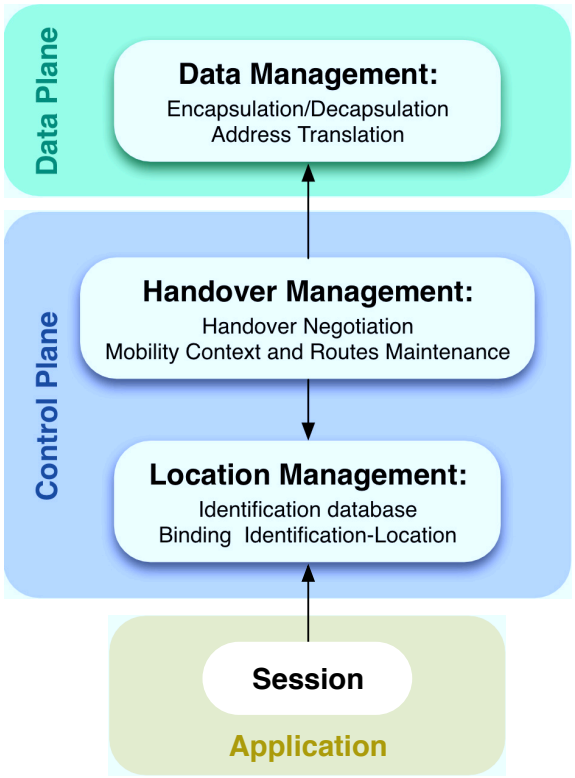


Figure B.1: Mobility Management Functional Blocks

The decoupling of the mobility management functionalities enables the support of three distinct independent mechanisms to deal with the mobility management. However, it is necessary to provide the interactions between the mobility management blocks in order to guarantee the proper behavior of the mobility management. As illustrated in Figure B.1, when an IP handover occurs, the handover management is responsible to update the data management with the new mobility routes, and update the location management with the new IP address of the MN. If an application from the CN initiates a session with the MN, it should interact with location management to obtain the current IP address of the MN.

B.4 Mobility Management Approaches

After the decoupling of mobility management functionalities, it is important to understand the impact, on both the user and the network, of distributing the mobility management functionalities through different network elements. We consider three main approaches for this distribute. In these approaches, we assume the distribution of the handover management through the MNs, and the Location management is maintained centralized in the MA. For the data management, we will propose different distribution approaches.

B.4.1 Centralized Mobility Management

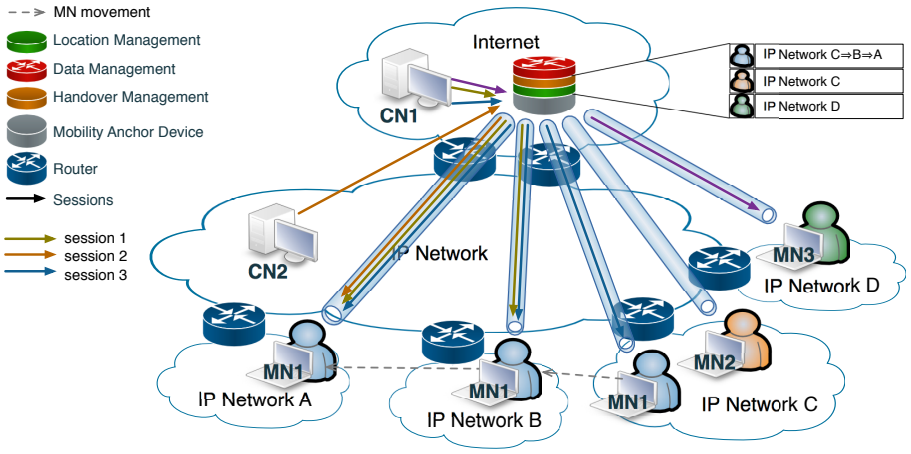


Figure B.2: Centralized Mobility Management Approach

The centralized mobility management approach in Figure B.2 illustrates current centralized models, where a unique element, called Mobility Anchor, is responsible for the management of data, handover and location. The sessions are always routed via the centralized MA and tunneled to the updated IP of the MN. Thus, the end-point of the tunnel is updated with the IP address obtained from the new MN's network after an IP handover. The centralized MA maintains a tunnel with each MN in order to forward the traffic packets to the current network of the MN while it roams among different IP networks. In the centralized model, the MN's device performs the handover detection and signals the MA with the new IP address, but the handover negotiation and the management of mobility context and routes are maintained centralized in the MA. When the MN1 is in network A, all sessions are routed via the MA device, independently of the location of the CN. Even the session 2 established with the closer CN2, initiated in the current network of the MN1 (Network A), is tunneled from the beginning through the MA device.

B.4.2 Distributed Data Management

In this section we present three approaches to distribute the data management. We study the set of combinations of distributing the data management through the ARs and the CNs, placing the data management in the ARs, the CNs and both. We assume the distribution of the handover management functionality through the MNs, since they are responsible for the handover detection, as well as the maintenance of mobility context and routes. We also consider that the location management is maintained centralized in the MA.

B.4.2.1 Data Management in ARs

In the first approach, the data management is distributed through the ARs of the network. In an initial phase, while the MN does not roam to another IP network, the session is maintained through the optimized path between the CN and the MN, such as the MN1 in the Network C of Figure B.3. When the MN1 roams to the Network B, it obtains a new IP address, IP@B. In order to maintain the MN1's ongoing session 3, it is established a tunnel between the AR of Network C and the IP@B of MN1. The MN1's new session 1 is initiated while MN is in Network B without any tunnel or other mobility

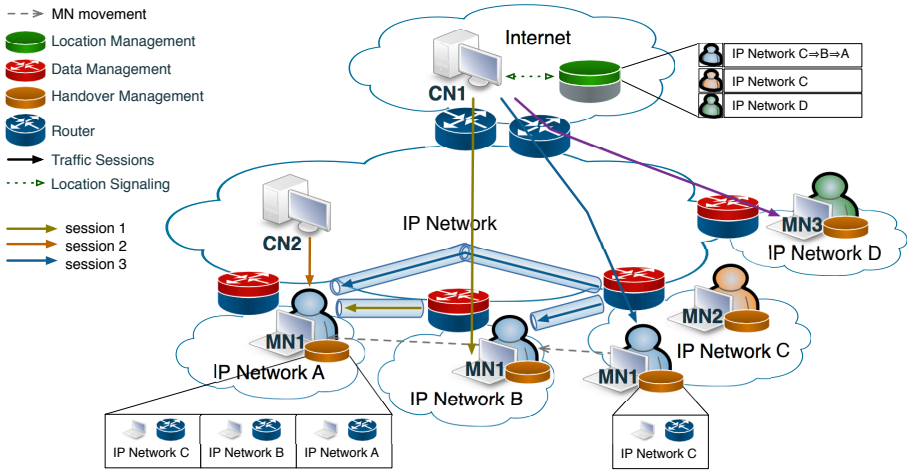


Figure B.3: Data Management in ARs

support. When the MN1 roams to Network A, it maintains the ongoing sessions 1 and 3, through two distinct tunnels with the ARs of Networks C and B respectively. The handover management placed in the MN maintains the set of ARs and respective mobility context updated.

B.4.2.2 Data Management in CNs

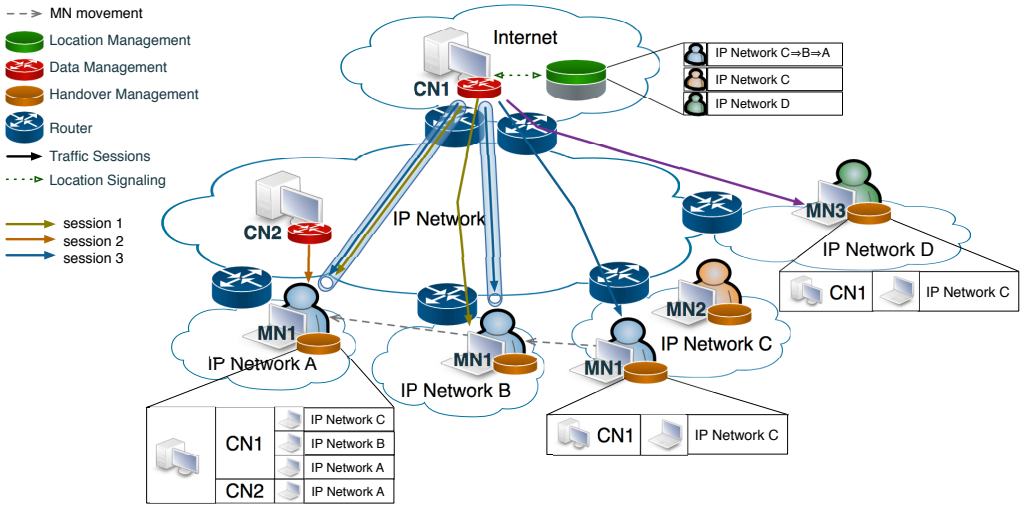


Figure B.4: Data Management in CNs

In the second approach, the data management is distributed through the CNs of the network. Thus, the forwarding of traffic is moved from a point close to the destination to the traffic source point. The sessions follow the optimized path between the CN and the MN, such as illustrated in Figure B.4 with MN1. However, from the first handover of the MN1 to Network B, the MN1's session 3 with CN1 is tunneled in the entire path from the CN1 to the MN1. The MN1's session 2, initiated while the MN1 is connected through the Network B, is tunneled from the time that the MN1 roams to Network A. Moreover, the MN1's orange session established with CN2, initiated while MN1 is connected to Network A, follows the optimized path between the CN2 and the MN without any tunnel or other

mobility support. The handover management does not need to maintain the set of ARs of the MN, but it has to maintain the set of CNs involved in the sessions with MN and all the IPs used during the session with the CNs. This distribution of the data management optimizes the routing path of the session, but it introduces longer tunnels for handover traffic.

B.4.2.3 Data Management in ARs and CNs

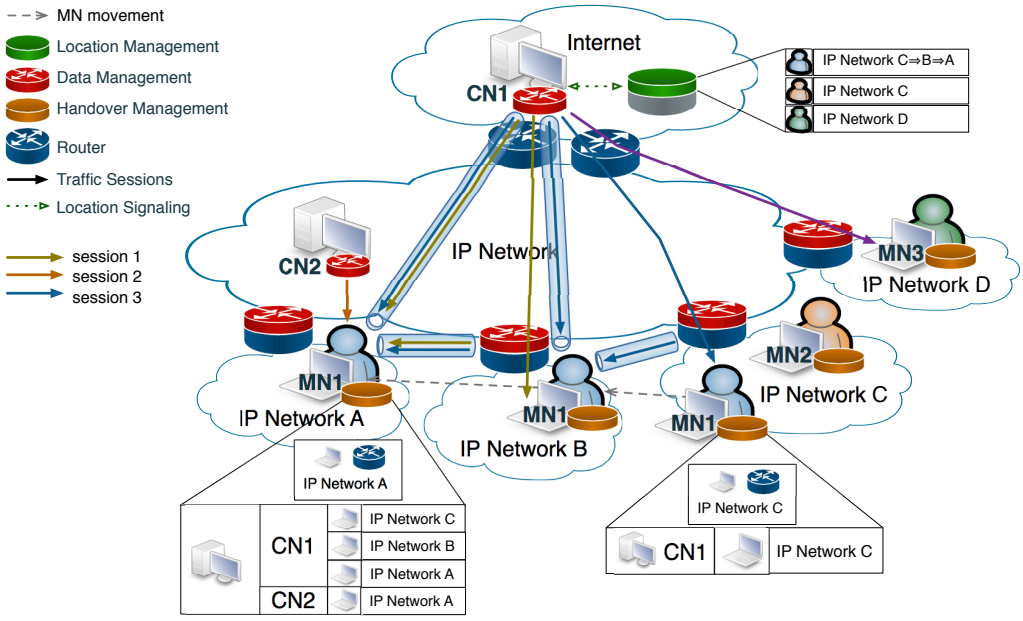


Figure B.5: Data Management in ARs and CNS

The third approach is a mixture of the two previous ones, since the data management is distributed through the ARs and CNs. However, the data management in the ARs is used during short periods of time to optimize the handover. In Figure B.5, when the MN1 roams from Network C to Network B, it updates the mobility routes of the data management of both CN1 and AR of Network C. While the CN1 does not tunnel the traffic to the IP@B of MN1, the AR of Network C tunnels the traffic of the MN1's ongoing session 3 to IP@B of MN1. This scheme uses a MA closer to the user, such as the AR, in order to improve the handover performance regarding the ongoing sessions, and another MA in the CN to optimize the routing path. In this approach, the MN has to maintain the mobility context about the CNs of the ongoing session, as well as the mobility context about the previous network.

B.5 Analytical Modeling

In this section, we develop an analytical model to compare the distributed data management approaches with the centralized one. The analytical models evaluate the data and signaling cost of the approaches. The notation used in the analytical model is described in Table B.1. We adapt the performance analysis methods proposed in [20] to be able to evaluate the distributed approaches.

The wireless transmission cost $\boldsymbol{TC_X}$ is calculated in equation (B.1). $\boldsymbol{X}$ is the type of the packet which is sent/received by the MN and $\boldsymbol{S_X}$ is the size of a packet of type $\boldsymbol{X}$.

Table B.1: Notations and values used in the analytical model

Notation	Definition	Default Value
S_C	Size of a Control Message	56 bytes
S_D	Size of a Data packet	1024 bytes
S_{DT}	Size of a Tunneled Data packet	S_D+40 bytes
pf	Wireless link failure probability	0.2
$1/Ts = \mu_S$	Session service mean rate	$1/240s^{-1}$
$1/Tc = \mu_C$	Cell-residence/handover mean rate	$1/240s^{-1}$
$1/Ta = \mu_A$	Sessions inter-arrival mean rate	$1/60s^{-1}$
H_{X-Y}	Number of IP hops between X and Y	-

H_{MN-AR} is the number of wireless hops between the MN and the AR or vice-versa, and n_f is the number of transmission failures.

$$TC_X = H_{MN-AR} \times S_X + \sum_{n_f}^{\infty} n_f \times Pr\{n_f \text{ failures and 1 success}\} \times H_{MN-AR} \times S_X = \frac{H_{MN-AR}}{1-pf} \times S_X \quad \text{(B.1)}$$

B.5.1 Data Delivery Cost

It is defined as the size of the data payload with encapsulation, and multiplied by the number of IP hops that data packet crosses from CN to MN.

The probability of a data packet to be a handover packet (P_H) is defined in [15] and it is given by (B.2). We assume that the sessions arrival time follows a Poisson process, and their service time is exponentially distributed with a mean expected value μ_S. The cell residence time also follows an exponential distribution with an average μ_C.

$$P_H = \frac{\mu_C}{\mu_C + \mu_S} \quad \text{(B.2)}$$

Centralized (A)

In the centralized approach the data sessions are routed via the centralized MA, which tunnels the session packet to the current network of the MN. As expressed in Equation (B.3), the packets follow the path from CN to MA without any tunnel, and then they are tunneled from the MA to the MN.

$$DC_A = H_{CN-MA} \times S_D + H_{MA-AR} \times S_{DT} + \frac{H_{AR-MN}}{1-pf} S_{DT} \quad \text{(B.3)}$$

Data Management in ARs (B1)

Distributing the data management through the ARs enables the anchoring of the data sessions to the ARs of the network. As considered in (B.4), the sessions are initiated through the optimized path between the CN and the MN without tunneling. After the handover of MN to other IP network, the sessions already established are handover sessions, and the AR of the MN's previous network tunnels the packet to the current location of the MN, expressed in (B.5).

$$DC_{B1c1} = H_{CN-AR} \times S_D + \frac{H_{AR-MN}}{1-pf} S_D \quad \text{(B.4)}$$

$$DC_{B1c2} = H_{CN-AR} \times S_D + H_{AR-AR} \times S_{DT} + \frac{H_{AR-MN}}{1-pf} S_{DT} \quad \text{(B.5)}$$

The total data cost of scheme B1, described in (B.6), is a weighted sum of the two cases: handover packets (B.5) and non-handover packets (B.4). The weight factor $\boldsymbol{P_H}$, expressed through the probability of a packet to be a handover packet, is given by (B.2).

$$\boldsymbol{DC_{B1} = P_H \times DC_{B1c2} + (1 - P_H) \times DC_{B1c1}} \tag{B.6}$$

Data Management in CNs (B2)

Distributing the data management through the CNs enables the data session anchoring to any AR, since from the first handover, the session is tunneled in the entire path from the CN to MN. A session is initiated through the optimized path between the CN and the MN without any tunnel, such as the scheme B1 in (B.4). However, instead of providing a tunnel between the MN and the AR of the previous network for the ongoing session, it is maintained with an end-to-end tunnel between the CN and the MN, expressed in (B.7).

$$\boldsymbol{DC_{B2c2} = H_{CN-AR} \times S_{DT} + \frac{H_{AR-MN}}{1 - pf} S_{DT}} \tag{B.7}$$

The data cost of scheme B2 in (B.8) is a weighted sum of the cases (B.4) and (B.7), where the weight factor $\boldsymbol{P_H}$ is defined in (B.2).

$$\boldsymbol{DC_{B2} = P_H \times DC_{B2c2} + (1 - P_H) \times DC_{B1c1}} \tag{B.8}$$

Data Management in ARs and CNs (B3)

The third scheme is the integration of B1 and B2. A new session is established without tunneling between the CN and MN, such as in B1 and B2 (B.4). After the IP handover of MN, the data management in the CN and in the AR of the previous MN's network are updated. Thus, there will be a tunnel between the AR of the previous and current location of MN, while the CN does not configure the tunnel to the current location of MN, in order to improve the handover performance. However, assuming that the time receiving tunneled packets from previous AR is negligible when compared with the session service time, the data cost of handover packets can be just expressed by (B.7). As a result, the data cost of scheme B3 is equal to the data cost of B2, which is given by (B.8).

B.5.2 Signaling Cost

It is defined as the size of the control messages multiplied by the number of IP hops those messages cross in the network. The location management is maintained centralized in the MA, which has to be updated with the new IP address of the MN after an IP handover.

Centralized (A)

In the centralized approach, the mobility management functionalities are placed in the centralized MA. Therefore, when the MN roams to another IP network, it just needs to update the centralized MA with the obtained IP address from the new network. Then, the MA is able to renew the association between the well-known address of the MN and its IP address in the current network. The signaling cost in (B.9) considers the cost of transmitting a control message from MN to MA.

$$\boldsymbol{SC_A = S_C \times \left(\frac{H_{MN-AR}}{1 - pf} + H_{AR-MA} \right)} \tag{B.9}$$

Data Management in ARs (B1)

The distribution of the data management provides separated signaling mechanisms, one for the data management and another for location management. The signaling cost in (B.10) takes into account the control messages transmitted to the ARs that are anchoring ongoing sessions of the MN, and the control message to update the location management. The average number of ARs anchoring ongoing sessions of the MN is calculated by (B.11), according to the cell-residence mean rate and session service mean rate.

$$SC_{B1} = N_{AR} \times S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-AR}\right) + S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-MA}\right) \tag{B.10}$$

$$N_{AR} = \frac{\mu_C}{\mu_S} \tag{B.11}$$

Data Management in CNs (B2)

The decoupling of data and location management requires that a MN updates both through distinct signaling mechanisms. The signaling mechanism for the location management only requires a control message to update the IP of the MN in the current network. However, to update the data management distributed through the CNs, the MN has to update all CNs involved in ongoing sessions. It is assumed that each new session of the MN is established with a distinct CN, thus the number of CNs involved in the ongoing sessions is equal to the number of ongoing sessions, given by (B.13). The total signaling cost in (B.12) is the addition of the control messages for the data management signaling, and the control messages for location management signaling.

$$SC_{B2} = N_{OS} \times S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-CN}\right) + S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-MA}\right) \tag{B.12}$$

$$N_{OS} = \frac{\mu_A}{\mu_S} \tag{B.13}$$

Data Management in ARs and CNs (B3)

The distribution of data management through ARs and CNs increases the number of control messages transmitted by the signaling mechanism of data management. The signaling cost (B.14) considers the number of ARs anchoring ongoing sessions of the MN, the number of CNs involved in the ongoing sessions of the MN, and the update of the location management placed in the MA.

$$\begin{aligned} SC_{B3} = N_{AR} \times S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-AR}\right) + N_{OS} \times S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-CN}\right) \\ + S_C \times \left(\frac{H_{MN-AR}}{1-pf} + H_{AR-MA}\right) \end{aligned} \tag{B.14}$$

B.6 Evaluation

In this section, we evaluate the distributed mobility management approaches according to the analytical model and simulations. The evaluation measures the impact of distributing mobility management functionalities through network elements closer to the user, in comparison with a centralized model. The objective is to assess the benefits and drawbacks

of the distributed mobility management, regarding data (e.g. data cost, data end-to-end delay and packet loss) and control (e.g. signaling cost and update times). We start the evaluation through the validation of the analytical model, based on a comparison with the simulation results for a specific scenario. Then, we evaluate the distributed approaches according to the analytical model and simulations, for a set of different scenarios, in order to show the data and control performance evolution.

B.6.1 Comparing Analytical Model with Simulations

In this section, we compare the results from the analytical model with the results from simulation, in a specific scenario. We evaluate the data and the signaling cost, measured through the analytical model and simulations, where the parameters are configured with the same values for both methods.

We created a network topology with 8 wireless ARs distributed in a grid topology, in order to ensure a full coverage area. The MNs move in the wireless ARs coverage area using a RandomWayPoint mobility model, with an average pause time of 240s and a speed of 2.5m/s.

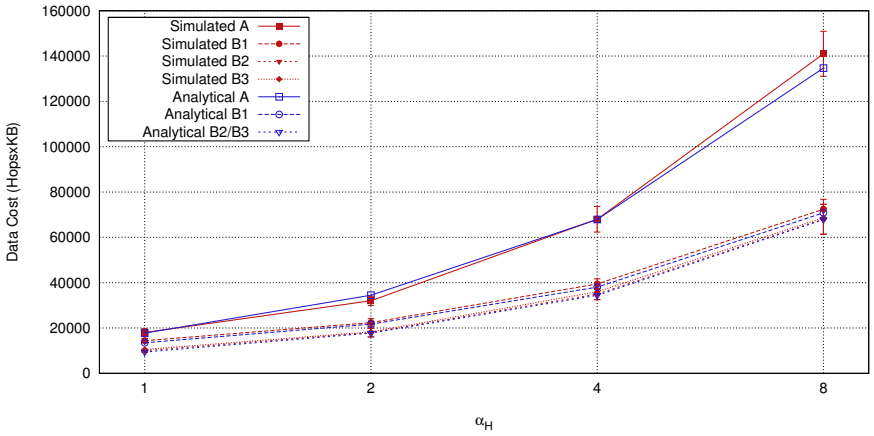
The sessions are configured according to the Table B.1, based on a constant-bit-rate of 128 Kbps and a packet size of 1B, under the Real Time Protocol, from the CN to the MN. The sessions inter-arrival time is exponential distributed with an average of 1 min, as well as the session service time and cell-residence time with an average of 4 min. We define a wireless link failure probability of zero to provide an easy comparison, and a Friis propagation model with a range of 100 meters.

The number of hops among the wireless ARs ($\boldsymbol{H_{AR-AR}}$) is eight, an each hop introduces a delay around 1ms. It is defined an equal number of hops between the pairs: CN-HA, a CN-AR, and HA-AR, thus $\boldsymbol{H_{CN-HA} = H_{CN-AR} = H_{HA-AR}}$. Moreover, the number of hops between these pairs is $\boldsymbol{\alpha_H}$ times higher than the $\boldsymbol{H_{AR-AR}}$, where $\boldsymbol{\alpha_H}$ is tested with different values. The metrics evaluated in this section are the data and signaling cost explained in (B.5).

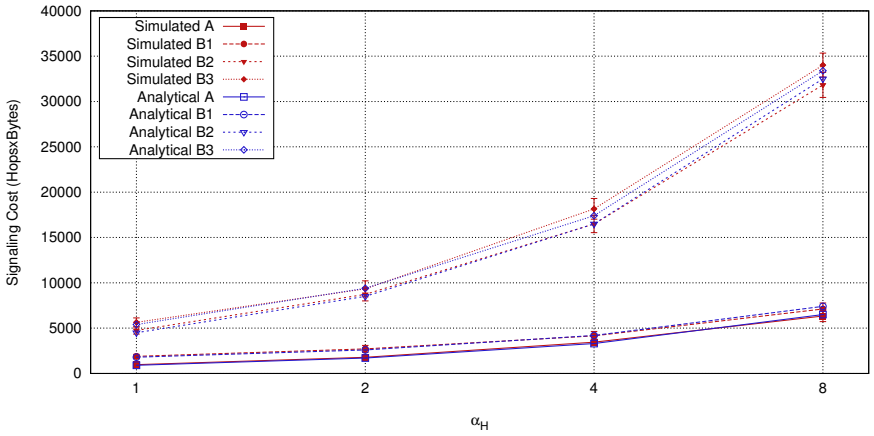
The results presented in Figure B.6 prove the accuracy of the proposed analytical model in analyzing the different approaches, where both the analytical and simulation results follow the same tendency. Figure B.6(a) shows the decrease of data cost of the distributed approaches when compared with the centralized approach A. Regarding the Signaling Cost, as can be observed in Figure B.6(b), the centralized approach has the lowest value. The approaches B2 and B3 significantly increase the signaling cost, since it is related with the number of ongoing sessions per instant and the hop distance between the MN and the CN. Approach B1 slightly increases the signaling cost when compared with approach A, since it is related with the ratio between the handover rate and session service time, as well as the hop distance between the MN and an AR.

B.6.2 Analytical Results

This section evaluates the presented approaches according to the analytical model presented in the previous section. The values used in the evaluation are presented in Table B.1. The sessions inter-arrival time is exponentially distributed with an average of 1 min, as well as the session service time and cell-residence time with an average of 4 min. The data packet size is equal to 1B, and we define a wireless link failure probability of 0.2. In the evaluation we assume that, in average, any two nodes of different networks are distanced through the same number of hops, thus $\boldsymbol{H_{CN-HA} = H_{CN-AR} = H_{HA-AR}}$, and two nodes of the same network ($\boldsymbol{H_{AR-AR}}$) are distanced through the same number



(a) Data Cost



(b) Signaling Cost

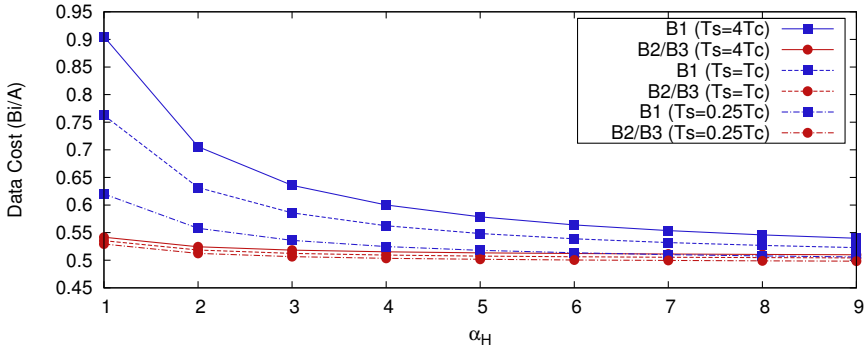
Figure B.6: Analytical Model Validation

of hops. The number of hops in the wireless part is one: $\boldsymbol{H_{MN-AR} = 1}$. We define $\boldsymbol{\alpha_H}$ as the ratio of 1) the number of hops between nodes in different networks to 2) the number hops between nodes in the same network: $\boldsymbol{\alpha_H = H_{CN-HA}/H_{AR-AR}}$. The results to be presented in this section are the data and signaling costs, which are always expressed as the ratio ($\boldsymbol{Bi/A}$) between a distributed approach ($\boldsymbol{Bi}$) and the centralized one ($\boldsymbol{A}$). A cost value closer to one means that both centralized and distributed approaches have a similar cost. Cost values higher than one means that centralized approach is better, while cost values lower than one gives advantage to the distributed approach.

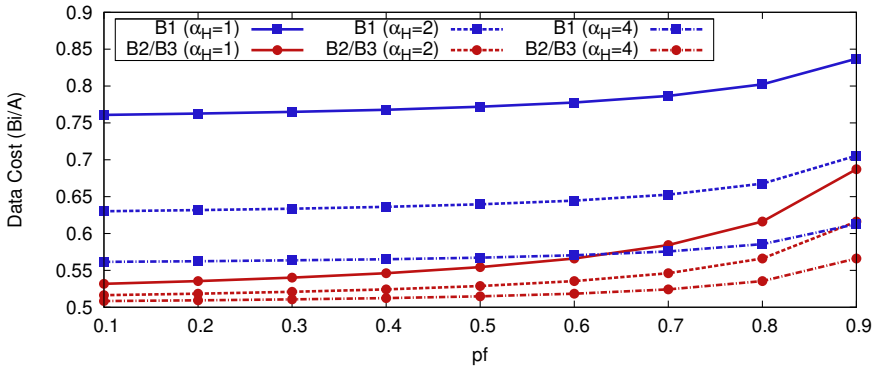
B.6.2.1 Data Cost

The Data Cost is evaluated in Figures B.7 and B.8, according to the analytical model previously presented, varying the most significant input parameters, such as $\boldsymbol{\alpha_H}$, $\boldsymbol{Ts}$, $\boldsymbol{Tc}$, $\boldsymbol{pf}$ and Packet Size. This evaluation aims to study the impact of distributing the mobility management on data cost with distinct scenarios.

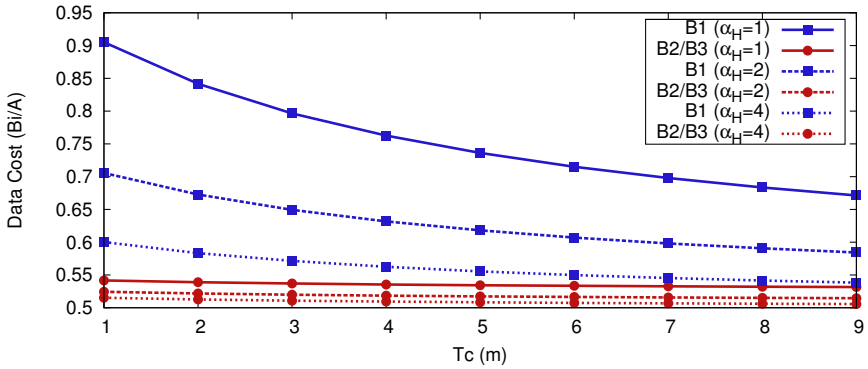
The value of $\boldsymbol{\alpha_H}$ in Figure B.7(a) shows a significant impact in the data cost for approach B1, specially for values of service time higher than the cell-residence time. However, the data cost of the distributed approaches (B1, B2, B3) is always lower than the



(a) Impact of $\boldsymbol{\alpha_H}$ with $\boldsymbol{Ts}$



(b) Impact of $\boldsymbol{pf}$ and $\boldsymbol{\alpha_H}$



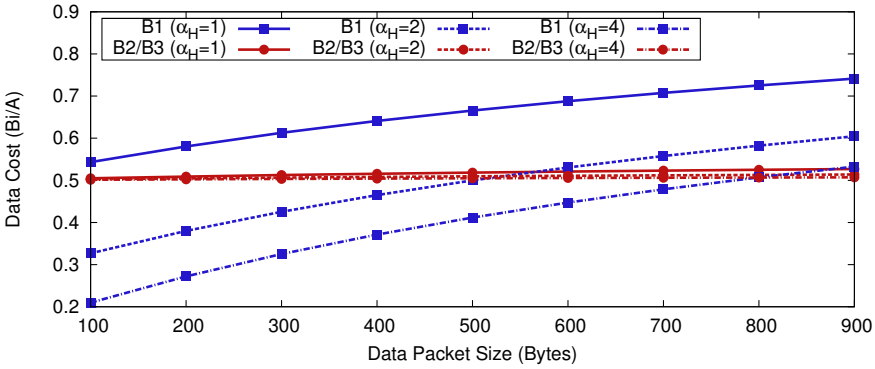
(c) Impact of $\boldsymbol{Tc}$ with $\boldsymbol{\alpha_H}$

Figure B.7: Data Cost

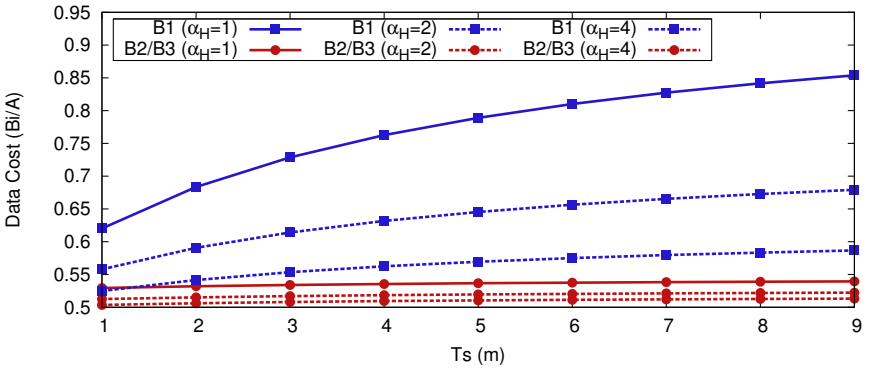
one of the centralized approach for values of $\boldsymbol{\alpha_H}$ higher than 1. For high values of $\boldsymbol{\alpha_H}$, the evaluated approach with distinct service times converge to half of the data cost of the centralized approach A.

In Figure B.7(b), we observe that the growth of the wireless transmission failure probability $\boldsymbol{pf}$ increases the data cost of the distributed approaches when compared with the centralized A, since with a transmission failure probability $\boldsymbol{pf}$ close to 1, any approach works properly. However, even for high values of $\boldsymbol{pf}$, the distributed approaches decrease the data cost when compared with the centralized approach A.

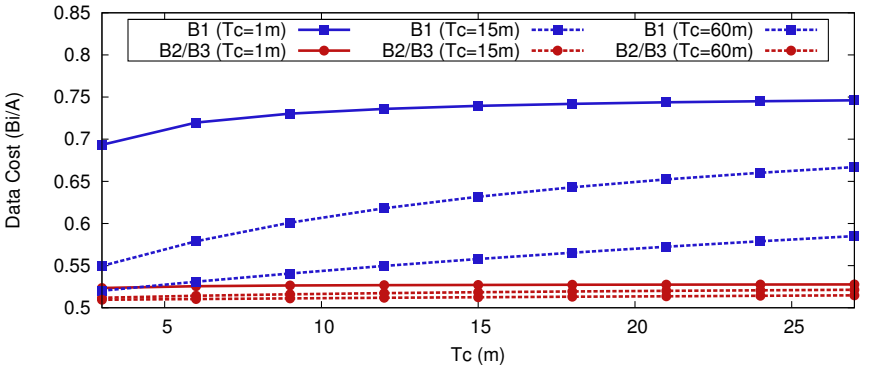
Figure B.7(c) demonstrates that the increase of cell-residence time $\boldsymbol{Tc}$ overall decreases



(a) Impact of Data Packet Size with $\boldsymbol{\alpha_H}$



(b) Impact of $\boldsymbol{Ts}$ with $\boldsymbol{\alpha_H}$



(c) Impact of $\boldsymbol{Ts}$ with $\boldsymbol{Tc}$

Figure B.8: Data Cost *continuation*

the data cost of distributed approaches when compared with centralized A. The cell-residence time $\boldsymbol{Tc}$ has more impact on data cost of approach B1 for low values of $\boldsymbol{\alpha_H}$. The cell-residence time slightly influences the data cost of approach B2/B3, since the packet always follows the optimized path between the CN and the MN, and only the portion of tunneled packets increases.

The size of the data packets, in Figure B.8(a), has a strong influence in the data cost of approach B1. For high values of $\boldsymbol{\alpha_H}$ and low values of packet size, the data cost of approach B1 is less than half of the data cost of the centralized approach A. The approach B2/B3 slightly depends on the packet size, thus its data cost is almost half of the data

cost of the centralized approach independently of the data packet size.

The influence of the service time Ts with different values of $\boldsymbol{\alpha_H}$ and $\boldsymbol{Tc}$ is evaluated in Figures B.8(b) and B.8(c). The decrease of the service time Ts decreases the data cost of approach B1 when compared with approach A, since the probability of a packet to be a handover packet reduces. We can observe a fast decrease of the data cost with the decrease of Ts for lower values of $\boldsymbol{\alpha_H}$. The impact of Ts with $\boldsymbol{\alpha_H}$ and $\boldsymbol{Tc}$ is negligible when compared with the centralized approach, since B2/B3 approach just significantly depends on the distance between the CN and the MN.

B.6.2.2 Signaling Cost

The Signaling Cost is evaluated in Figures B.9 according to the analytical model previously presented, varying the most significant input parameters, such as $\boldsymbol{\alpha_H}$, $\boldsymbol{Ts}$, $\boldsymbol{Tc}$ and $\boldsymbol{Ta}$. This evaluation aims to study the impact of distributing some mobility management functionalities on the signaling cost with distinct scenarios.

From the evaluation of the signaling cost in Figure B.9, it is observed that the signaling cost of the distributed approaches is always higher than the centralized approach, since there are two separated signaling mechanisms: one for the location management that is maintained in the centralized MA, and another for the data management distributed through the ARs.

In Figure B.9(a) we observe that the increase of $\boldsymbol{\alpha_H}$ presents a similar impact in the signaling cost of distributed and centralized approaches. However, the increase of $\boldsymbol{Ts}$ maintaining the value of $\boldsymbol{Tc}$, strongly increases the signaling cost of B2 and B3.

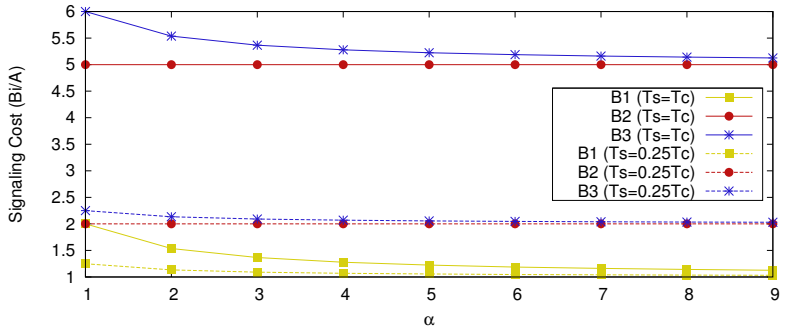
The influence of $\boldsymbol{Ts}$ with $\boldsymbol{\alpha_H}$ and $\boldsymbol{Tc}$ is evaluated in B.9(b) and B.9(c) respectively. The increase of $\boldsymbol{Ts}$ has a strong impact in approaches B2 and B3, since there are more ongoing sessions when MN roams, which means more CNs to update. However, the increase of $\boldsymbol{\alpha_H}$ does not significantly affect the distributed approaches when compared with the centralized one, since both suffer the variation of $\boldsymbol{\alpha_H}$ in similar scale. The ratio $\boldsymbol{Ts/Tc}$ is crucial for the signaling cost. When the ratio has a low value, the service time is much lower than the cell-residence time, and the signaling cost of the distributed approaches is close to the signaling cost of the centralized approach. However, as long as the ratio increases, the distributed approaches significantly increase the signaling cost compared with the centralized approach, specially approaches B2 and B3.

The inter-arrival sessions time in Figure B.9(d) strongly influences the signaling cost of approaches B2 and B3, since a lower value of $\boldsymbol{Ta}$ means more ongoing sessions per instant in the MN, thus more CNs to update when the MN roams. The combination of a low value of $\boldsymbol{Ta}$ with a high value of $\boldsymbol{Ts}$ significantly increases the signaling cost of approaches B2 and B3.

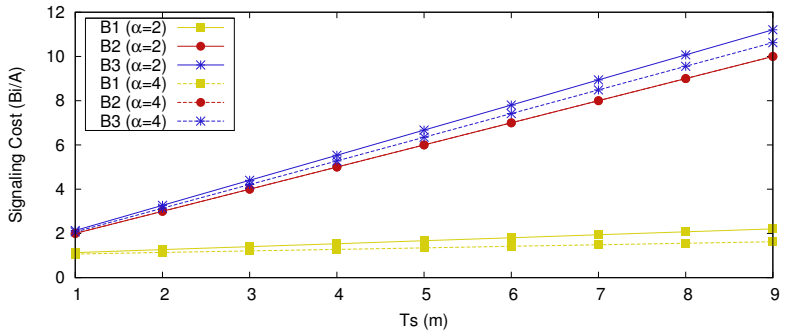
The distribution of the mobility management functionalities overall increases the signaling cost, specially for approach B2 with data management in the ARs and CNs, with the combination of the following factors: 1) a very mobile user, 2) high sessions inter-arrival mean rate and 3) long sessions service time. Another main contribution to the increase of the signaling cost of the distributed approaches is the maintenance of the location management centralized in the MA, which is updated every time a MN obtains a new IP address.

B.6.3 Simulation Results

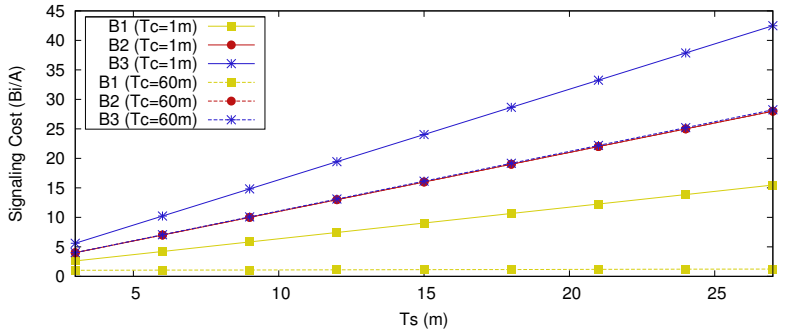
The evaluation through simulations aims to study the distributed approaches in a more realistic environment than the one characterized by the analytical model. The platform



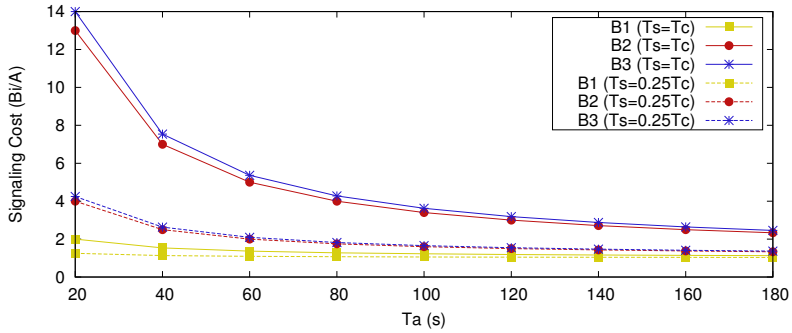
(a) Impact of $\boldsymbol{\alpha_H}$ with $\boldsymbol{Ts}$



(b) Impact of $\boldsymbol{Ts}$ with $\boldsymbol{\alpha_H}$



(c) Impact of $\boldsymbol{Ts}$ with $\boldsymbol{Tc}$



(d) Impact of $\boldsymbol{Ta}$ with $\boldsymbol{Ts}$

Figure B.9: Signaling Cost

selected for the evaluation through simulation was the network simulator 3.12 [21].

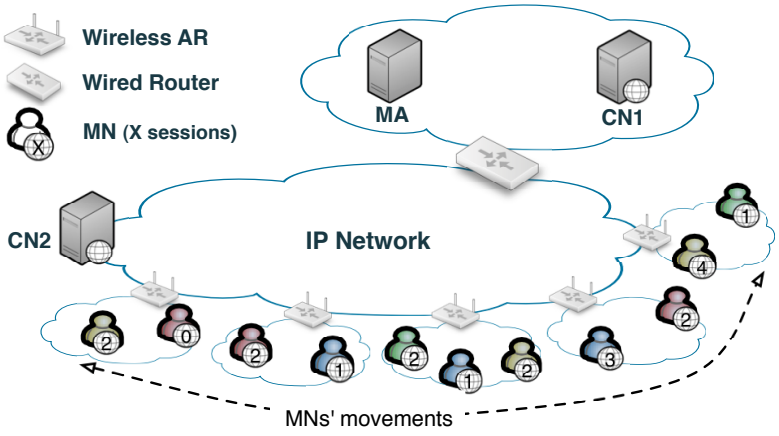


Figure B.10: Simulated Scenario

The scenario defined for the evaluation is illustrated in Figure B.10. The wireless ARs and MNs are configured with 802.11a and the Friis propagation model, with a range of 100 meters. In the defined scenario, MNs move in the wireless ARs coverage region with a RandomWayPoint model provided by the BonnMotion tool [22], initiating and terminating traffic sessions with CNs. RandomWayPoint model causes MNs movement to the area around interesting points with attraction capabilities, after the MNs pause times. The set-up scenario is configurable, and default configurations are the following:

- 8 wireless ARs in a grid to ensure full connectivity coverage.
- $\boldsymbol{H_{AR-AR}}$ is 8, with a hop delay around 1ms.
- The hop distance CN2-AR is equal to the hop distance among the ARs, $\boldsymbol{H_{CN2-AR} = 8}$.
- Equal number of hops between the pairs: $\boldsymbol{H_{CN1-MA}}$, $\boldsymbol{H_{CN1-AR}}$ and $\boldsymbol{H_{MA-AR}}$.
- The number of hops between these pairs is $\boldsymbol{\alpha_H}$ times higher than the $\boldsymbol{H_{AR-AR}}$.
- MNs follow a RandomWayPoint mobility model with a changeable maximum pause time and a speed of 2.5 m/s.
- Traffic Application with a constant-bit rate under Real-Time Protocol, from CN to MN.
- Packet Size and Rate of 1024 bytes and 128Kbps respectively.
- Sessions inter-arrival time per MN exponentially distributed with an average of 60 seconds.
- Sessions service time exponentially distributed with an average of 240 seconds.
- Results from independent simulations with a confidence interval of 95%.

We define several metrics for the evaluation of the proposed approaches. These metrics are evaluated according to distinct values of $\boldsymbol{\alpha_H}$, and movement pause times with communications with CN1 and CN2:

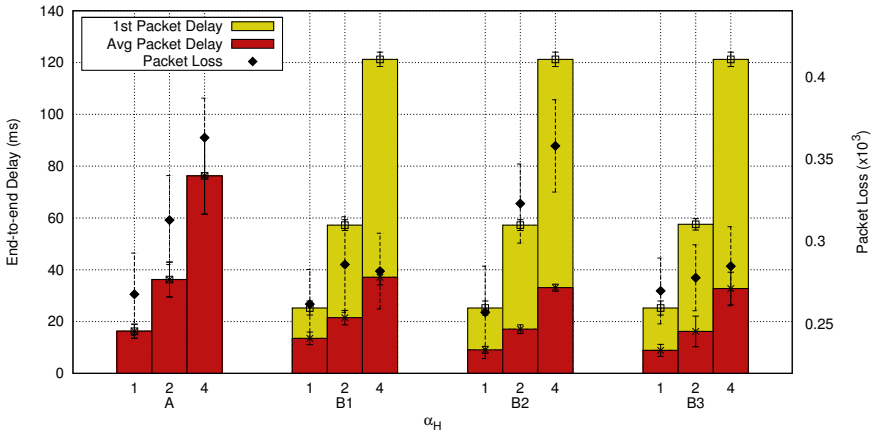
- **Average Packet Delay:** the average time that a data packet takes to be transmitted from the CN to the MN.
- **First Packet Delay:** the time that a CN needs for the location search (MN's IP address), plus the time to deliver the first data packet to the MN.
- **Packet loss:** the difference between the number of packets that were sent by the CN and the ones received by the MN, not distinguishing the cause of the packets loss. The majority of the lost packets were due to the handover, but a slightly part of the packet loss might be caused by wireless interference or collisions.
- **Location Update time:** the average time that a MN needs to update its binding (IP address) in the location management system.
- **Data Update time:** the average time that a MN needs to update the mobility routes with new IP address in the data management system.
- **Overhead:** the number of control packets transmitted.

B.6.3.1 Changing $\boldsymbol{\alpha_H}$

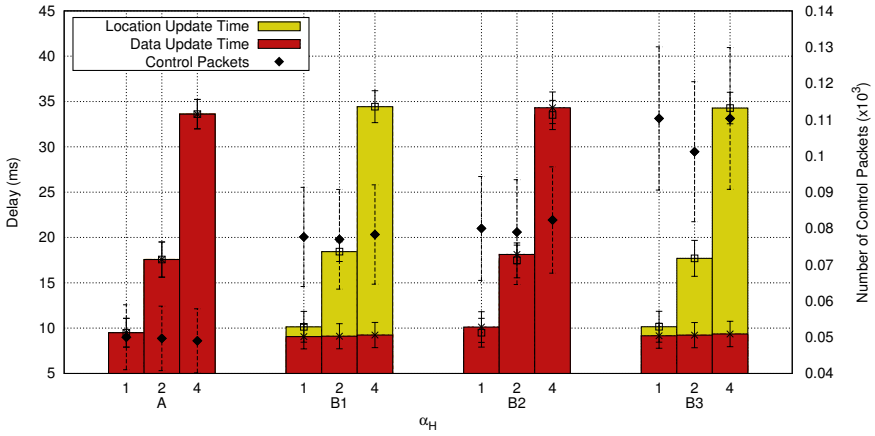
We evaluate the approaches changing the parameter $\boldsymbol{\alpha_H}$, previously defined in the analytical model evaluation. The MNs' RandomWayPoint model is defined with a maximum pause time of 300ms, which is equivalent to a measured handover rate of $\mathbf{160 \pm 17}$ ms. An $\boldsymbol{\alpha_H}$ value of 1 means the same wired distance in hops among all network elements involved: CNs, ARs and MA.

In Figure B.11(a), it is evaluated the impact of $\boldsymbol{\alpha_H}$ in the data delay and loss with sessions with CN1. The increase of $\boldsymbol{\alpha_H}$ benefits the distributed approaches, since the average packet delay increases less than in the distributed approaches. However, the distributed approaches increase the delay to deliver the first data packet, since it is needed a location search mechanism to obtain the current IP address of the MN before the session establishment. The approaches B1 and B3 place the data management in the ARs close to the MN, thus the packet loss does not depend on the $\boldsymbol{\alpha_H}$, only the hop distance among ARs. Thus, the data update time is lower for approach B1 and B3, as illustrated in Figure B.11(b). The packet loss in B2 depends on the hops distance between an AR and the CN1, while in A it depends on the hop distance between an AR and the MA. Hence, the increase of $\boldsymbol{\alpha_H}$ increases the packet loss of A and B2 approaches. The location update time is similar for the four studied approaches, since all of them update the IP address of the MN in the centralized MA for location management. Regarding the overhead, the distributed approaches increase the number of control packets in the network, with a higher growth for the approach B3, which updates the data management of ARs and CNs each time the MN roams to another IP network. Considering the number of ARs and CNs used in the evaluation, the number of control packets doubles in the distributed approaches, and it may increase even more in scenarios with more ARs and CNs.

It is evaluated the data and signaling in Figure B.12 with the impact of $\boldsymbol{\alpha_H}$ in communications with a closer CN, the CN2. In communications with a closer CN, Figure B.12(a), the distributed approaches reduce the average packet delay, since the data management is placed closer to the MN, in ARs and CN2. Moreover, $\boldsymbol{\alpha_H}$ has no impact in the distributed approaches, since they do not depend on the distance of the centralized MA in communications with a closer CN. The time to deliver the first packet in the distributed approaches is higher than in the centralized approach A, since it is needed the location search signaling to obtain the current IP address of the MN from the centralized MA before



(a) Data



(b) Signaling

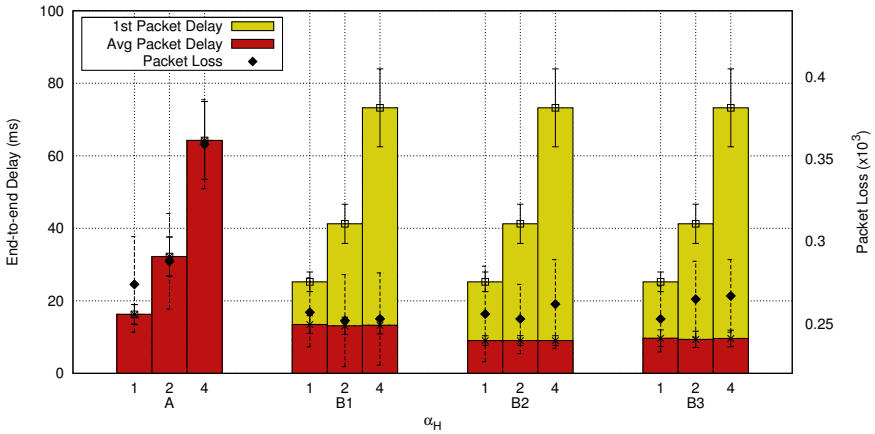
Figure B.11: Impact of $\boldsymbol{\alpha_H}$ in establishing sessions with CN1

the session establishment. The distributed approaches also decrease the packet loss when compared with the centralized approach A, since the time to update the data management system is lower, as observed in Figure B.12(b). The location update time is similar for all approaches, since all of them place the location management in the centralized MA. The distributed approaches increase the number of control packets in the network, due to the separate signaling messages used to interact with location and data management.

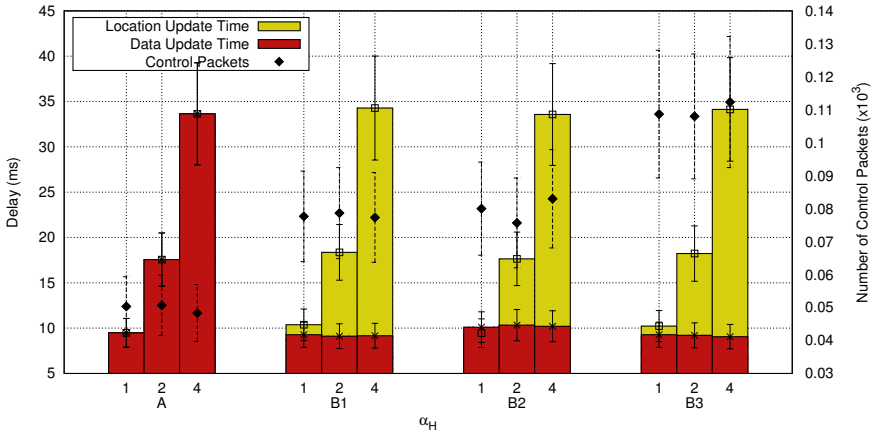
B.6.3.2 Changing Pause Time

We evaluate the approaches changing the maximum pause time of the RandomWayPoint mobility model, which is strictly related with the handover rate or cell-residence time. For these tests, $\boldsymbol{\alpha_H}$ is configured with the value 2, which means a double wired hop distance of CN1-MA/CN1-AR/CN1-AR when compared with the wired hop distance of CN2-AR/AR-AR.

The impact of the maximum pause time on the packet delay and loss with communications with CN2 is illustrated in Figure B.13(a). The values presented for the maximum pause times of 150, 300 and 600 ms are equivalent to a measured handover rates of $\mathbf{115 \pm 15}$, $\mathbf{160 \pm 17}$ and $\mathbf{280 \pm 20}$ ms, respectively. The distributed approaches reduce



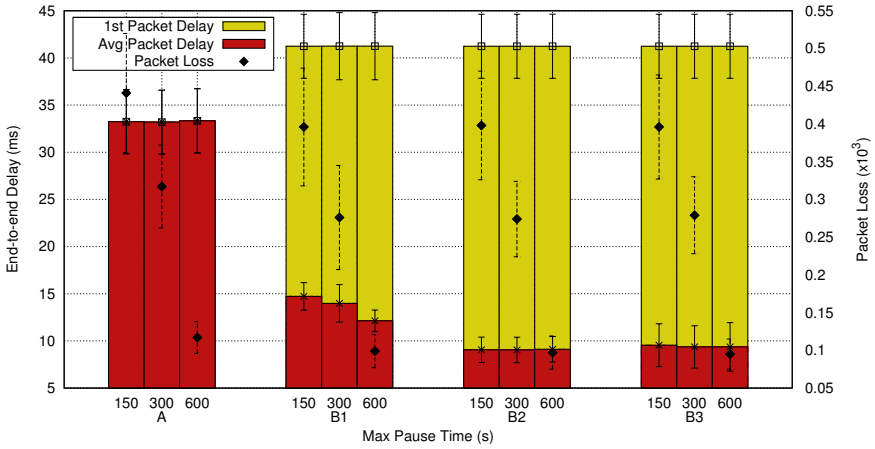
(a) Data



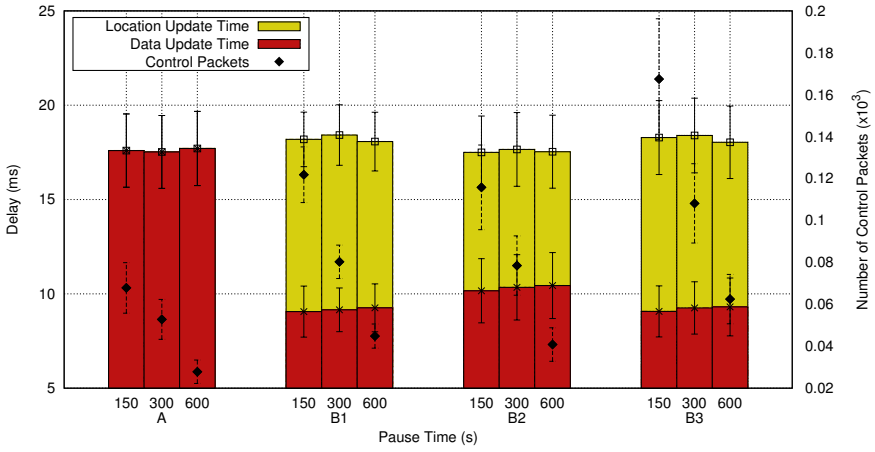
(b) Signaling

Figure B.12: Impact of $\boldsymbol{\alpha_H}$ in establishing sessions with CN2

the average packet delay, independently of the pause time. The approach B2 does not depend on the pause time, since it always uses the optimized routing path between the CN and the MN, through an end-to-end tunnel between the CN and the MN, applied to the handover packets. The average packet delay of approach B1 decreases with the increase of the pause time, since the portion of handover packet is reduced. Besides the decrease of the average packet delay of the distributed approach, the delay to deliver the first data packet increases, since it is needed a location search mechanism to obtain the current IP address of the MN before the session establishment. In the distributed approaches and considering the communications with CN2, the packet loss does not depend on $\boldsymbol{\alpha_H}$, only the hop distance among ARs or between the CN2 and an AR. Thus, the data update time is lower for these approaches, as illustrated in Figure B.13(b). The centralized location management in the MA has similar values for the location update. As observed in the previous evaluation results for the overhead, the number of control packets of the distributed approaches are higher than the ones of the centralized approach A, since the distributed approaches adopt two isolated mechanisms with separated messages to update both data and location management systems.



(a) Data



(b) Signaling

Figure B.13: Impact of Pause Time establishing sessions with CN2

B.6.3.3 Data Management Forwarding

We evaluate the approaches regarding the number of forwarded packets by the DM entity, which represents the packets that have to be forwarded to the current location of the MN, introducing an extra traffic exchange cost due to the non optimized routing path. This metric is indicated in percentage, considering the total amount of exchanged packets between CNs and MNs as the reference. We evaluate the approaches for $\boldsymbol{\alpha_H}$ equal to 1. The MNs' RandomWayPoint model is defined with a pause time of 240 ms, while the session service time is defined for 120, 240 and 480 ms.

The centralized approach A forces the forwarding of all data packets, since all data packets are firstly forwarded to the centralized DM, and then to the current location of the MN. The approach B1 just forwards part of the data packets, and the enforcement of the DM functionality is distributed through the several ARs of the network. Changing the service duration from 120 to 480 seconds, the percentage of packets forwarded by each DM is always less than 60 %. The approach B2 provides the forwarding of data packets from CNs through the encapsulation, thus it does not add any extra forwarding to the normal IPv6 routing of the network. The approach B3 just uses the DM entities of the ARs while the CNs are not updated with the new location of the MN; otherwise, the CNs

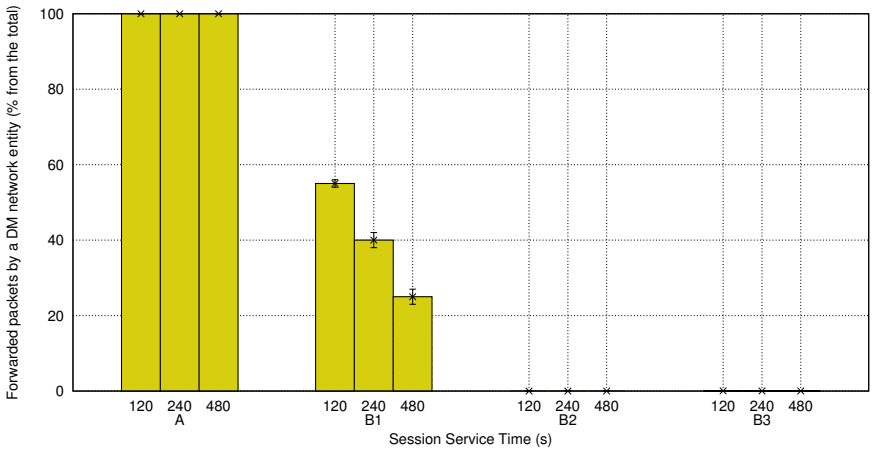


Figure B.14: Percentage of forwarded packets by each DM network entity

forward the data packets to the current location of MN, such as in approach B2. Thus, it introduces a nearly null packets forwarding from the DM entities in the ARs.

B.7 Conclusion

Service and network providers are evolving to flatten network architectures for mobile Internet and developing selective traffic offloading through the wireless local area networks. However, current IP mobility management protocols need to be adapted for such evolution, since they rely on a centralized entity, called mobility anchor, which routes the whole traffic and manages all the bindings of a large amount of users. These centralized models bring several problems of scalability, security and performance, when compared with distributed schemes.

This article proposed to study the different distribution approaches for the mobility management, considering the decoupling into location management, handover management and data management. The handover management is distributed through the MNs, while the location management is maintained centralized in the centralized mobility anchor. It is studied the distribution of data management through the ARs and CNs.

We developed analytical models to obtain the signaling and data cost of the distributed approaches, as well as a model for the centralized approach. We evaluated the approaches based on the developed analytical modes, as well as through the network simulator 3 with different scenarios.

From the summarized evaluation of the approaches, it is difficult to conclude about the overall optimized approach, since it depends on the considered scenario. It is evident from the evaluation that distributed approaches optimize the overall mobility management when compared with the centralized one, such as the data delivery cost and average data delay. Moreover, the best distributed approach strongly depends on $\boldsymbol{\alpha_H}$, the relation between $\boldsymbol{Ts}$ and $\boldsymbol{Tc}$, and the data packet sizes. Independently of the distributed approach, the dynamic mobility (providing mobility just when needed) brings benefits for the data. When the CN supports mobility, it seems advantageous to adopt an end-to-end tunnel between the CN and the MN for handover packets. However, for communications with a CN without any mobility support, the introduction of the data management functionality (e.g. encapsulation based on address translation) in the ARs significantly improves the data performance.

We believe this work is one more step towards a novel distributed and dynamic mobility management scheme, which anticipates the changes of behavior by the network, users and services. From these conclusions, we will define a novel distributed and dynamic mobility management scheme for flat network architectures.

B.8 Acknowledgments

This work has been supported by User-centric Mobility Management (UMM) project (PTDC/EEA-TEL/105709/2008), sponsored by Funda para a Ciia e Tecnologia (FCT). Tiago Condeixa is also supported by a FCT scholarship with the reference SFRH/BD/65265/2009.

B.9 References

[1] S. I. ULC, "Sandvine global internet phenomena report - 2h 2012," Report, Nov. 2012.

[2] C. S. Inc, "Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016," Report, Feb. 2012.

[3] C. Perkins et al., "Mobility support in ipv6," IETF, RFC 6275, Jul. 2011.

[4] S. Gundavelli et al., "Proxy mobile ipv6," IETF, RFC 5213, Aug. 2008.

[5] H. Chan, "Problem statement for distributed and dynamic mobility management," IETF, Internet-Draft draft-chan-distributed-mobility-ps-05, Oct. 2011, work in progress.

[6] H. Chan et al., "Requirements for Distributed Mobility Management," IETF, Internet-Draft draft-ietf-dmm-requirements-07, Aug. 2013, work in progress.

[7] DMM. (2013, Jan) Distributed mobility management ietf workgroup. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[8] T. Condeixa et al., "Decoupling and distribution of mobility management," in *Globecom Workshop MobiWorld*, Dec. 2012.

[9] H. Soliman et al., "Hierarchical mobile ipv6 (hmipv6) mobility management," IETF, RFC 5380, Oct. 2008.

[10] H. Chan et al., "Distributed and dynamic mobility management in mobile internet: Current approaches and issues," *Journal of Communications*, vol. 6, no. 1, pp. 4–15, Feb. 2011.

[11] A. Nascimento et al., "A characterization of mobility management in user-centric networks," in *Smart Spaces and Next Generation Wired/Wireless Networking*, vol. 6869. Springer, 2011, pp. 314–325.

[12] P. Bertin et al., "An evaluation of dynamic mobility anchoring," in *70th IEEE Vehicular Technology Conference Fall*, Sep. 2009, pp. 1–5.

[13] P. Seite and P. Bertin, "Distributed mobility anchoring," IETF, Internet-Draft draft-seite-dmm-dma-00.txt, Feb. 2012, work in progress.

[14] P. Bertin et al., "Distributed or centralized mobility?" in *IEEE Global Telecommunications Conference*, Dec. 2009, pp. 1–6.

[15] H. Ali-Ahmad et al., "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *IEEE WPMC*, Sep. 2012.

[16] C. Bernardos and J. Zuniga, "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-00, Sep. 2012, work in progress.

[17] J. Kim et al., "Use of proxy mobile ipv6 for distributed mobility control," IETF, Internet-Draft draft-sjkoh-mext-pmip-dmc-03, Jun. 2011, work in progress.

[18] J.-H. Lee, J. Bonnin, and X. Lagrange, "Host-based distributed mobility management support protocol for ipv6 mobile networks," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, 2012, pp. 61–68.

[19] ——, "Host-based distributed mobility management: Example of traffic offloading," in *Consumer Communications and Networking Conference (CCNC), 2013 IEEE*, 2013, pp. 637–640.

[20] J. Lee et al., "Enhancing qos of mobile devices by a new handover process in pmipv6 networks," *WPC Journal*, vol. 61, pp. 591–602, 2011.

[21] ns 3. (2013, Jan) ns-3.12. [Online]. Available: http://www.nsnam.org/ns-3-12

[22] BonnMotion. (2013, Jan) bonnmotion 2.0. [Online]. Available: http://net.cs.uni-bonn.de/wg/cs/applications/bonnmotion

Paper C

Rethinking IP Mobility Management Towards a Distributed and Dynamic Scheme

Tiago Condeixa, Jonathan Carvalho, Susana Sargento and Rute Sofia
in *IEEE/ACM Transactions on Networking* (submitted)

Rethinking IP Mobility Management Towards a Distributed and Dynamic Scheme

Tiago Condeixa, Jonathan Carvalho, Susana Sargento and Rute Sofia

Abstract

Currently, there has been a significant increase of mobile traffic and mobile devices in the network. The growing mobile traffic is being partially generated by geographical communication, due to the migration of the content servers closer to the user. Integrating these trends with the heterogeneity of the network and users' routines brings novel challenges to the current IP mobility management schemes, which were developed assuming static and centralized models. This article presents a novel trend for rethinking IP mobility management that distributes the IP mobility management functionalities through the network nodes and the users devices, to cope with the increased flatten nature of the mobile networks. The proposed distributed and dynamic scheme, called Dynamic Mobile IP Anchoring (DMIPA), distributes the mobility context management through the MNs, and data anchoring through the Access Routers (ARs), without centralized entities. DMIPA also provides mobility support when the current access network of the user does no support mobility. The outcome of the entire evaluation through analytical models and simulations shows that DMIPA is able to optimize the network resources and improve the user experience when compared with Mobile IPv6 (MIPv6), by reducing the data delivery cost, the tunneled packets, the hops with tunneled packets and the end-to-end data delay. Moreover, DMIPA was properly validated in a real testbed, being able to provide session continuity support just by introducing a slight delay to data packets, and without impairing the bitrate.

Keywords: Mobility Management, Dynamic Mobility, Dynamic Data Anchoring, Distributed Mobility Context Management, Flat Networks, Heterogeneity

C.1 Introduction

In a modern society, the value of mobility management in the network has been enhanced to provide continuous communications to the demanding mobile users. The mobility management is responsible for the reachability of the mobile devices in the ongoing communications, and to provide reachability in new communications with other devices. There is an increase of users's mobility, and users are requiring more mobile multimedia services with increasing demands, which consume a significant part of available network resources [1] [2].

Hence, there is a paradigm shift in the network architectures with the introduction of flat models to deal with this growth of mobile traffic. Service providers are already implementing selective traffic offload strategies through the wireless local area networks. However, sessions continuity of the user's services shall also be guaranteed independently of the network technology, including the communication through these wireless local area networks. Moreover, users are already developing spontaneous wireless networks, simply based on cooperation and Internet access sharing, becoming Internet micro-providers. Furthermore, there has been a paradigm shift in users' traffic behavior with the increase

of communications between devices in the same geographical area, and the migration of content servers closer to the user, such as the Content Delivery Networks (CDNs).

Current IP mobility models have been developed for centralized network architectures; thus, they are based on a single entity, called Mobility Anchor (MA), that handles the mobility management of a large amount of Mobile Nodes (MNs). These mobility models bring several limitations when applied to the recent network trends [3] [4], such as a single point of failure, longer traffic delays and higher signaling loads. Hence, there is an increasing research on the distribution of mobility management, with novel approaches being proposed to clearly define the problem statement and proposing novel mechanisms. The majority of these novel approaches distribute the data anchoring functionality through the Access Routers (ARs), while the mobility context is maintained centralized.

Although much effort has been performed by some stakeholders, the introduction of dynamic and distributed schemes as a key concept to optimize the overall mobility management is not consensual in the research community. Therefore, this article introduces and studies this novel trend through the proposal of a distributed and dynamic host-based mobility scheme, called Dynamic Mobile IP Anchoring (DMIPA), prepared for flat and heterogeneous networks. DMIPA fully distributes the IP mobility management functionalities through MNs and Mobility-enabled Access Routers (MARs). It enables a dynamic data anchoring strategy that allows the session anchoring to any MAR. It also supports dynamic mobility, which means that it provides mobility support only when sessions really need it.

DMIPA main ideas were initially proposed in [5] and evaluated through an analytical model regarding data delivery cost, signaling cost and handover latency. This article goes further in providing a better problem statement definition through a comparison between centralized and distributed models, as well as a detailed description of DMIPA operation mode. The validation corroborates the analytical model, with network simulations, which compare DMIPA with MIPv6 [6], and a real testbed scenario to validate DMIPA. The obtained overall results show that DMIPA overcomes the centralized MIPv6 model, reducing the average end-to-end data packet delay, the total data cost, the total tunneled packets and hops, the average bindings per mobility anchor, and the IP mobility handover latency (update of mobility anchors). The best scheme regarding the signaling cost depends on the AR pause times and average sessions service times. The testbed results validate DMIPA in a real environment, proving the concept of distributed and dynamic IP mobility.

The reminder of this article is organized as follows. Related work is presented in Section II. Section III analyses current centralized models, and distributed and dynamic mobility trends in the envisioned scenario. Section IV presents a more detailed description of DMIPA, while Section V presents analytical models for both DMIPA and MIPv6. Section VI contains an extensive evaluation/comparison of DMIPA with MIPv6 through analytical and simulation results, as well as DMIPA validation in a real environment. Finally, Section VI concludes the article.

C.2 Related Work

Mobile IPv6 (MIPv6) [6] is one of the most adopted host-based IP mobility protocol, which provides global IP mobility support to the user in IPv6 networks. Mobile IPv6 is based on a centralized mobility entity, called Home Agent (HA), and another mobility entity called MN in the user's mobile device. The MN encapsulates/decapsulates packets to/from the HA, and updates the binding between its well-known IP address and the IP address obtained from the current IP network. The HA maintains these updated bindings

of all MNs in order to apply the rules to encapsulate/decapsulate packets to/from the MN.

MIPv6 introduces Route Optimization (RO) [6] that enables MN and Correspondent Node (CN) to communicate via a tunnel established in the direct routing path, despite the changes in IP connectivity on the MN side. However, RO demands for mobility support from the CN, which might not be prepared to provide the required mobility functionalities, since it can be any device in the network. Furthermore, the RO requires the *return routability procedure* for the correspondent registration/update through end-to-end messages exchange between the MN, the HA and the CN, which has an adverse impact on handover latency. The *return routability procedure* must be repeated in short periodic intervals, even in the absence of changes in IP connectivity on the MN side, which comes at the cost of an increased signaling overhead.

Proxy Mobile (PMIPv6) [7] is currently the most adopted network-based IP mobility protocol, which provides localized mobility support to the user in IPv6 networks. PMIPv6 was developed based on the idea of providing IP mobility to any mobile user without requiring mobility support from the users' devices, as opposed to MIPv6. Thus, PMIPv6 introduces an entity called Mobile Access Gateway (MAG) to hide the mobility from the MN. The mobility management functionalities are moved from the MN to the MAGs, which are usually implemented in the ARs. The tunnels are established between the centralized mobility anchor, called Local Mobility Anchor (LMA) and the MAGs. A MAG is also in charge of detecting an IP handover of the MN to enable the IP mobility management protocol.

There is an increasing interest in the topic of distributed and dynamic mobility management, specially in the problem statement and definition of guidelines for a common distributed mobility management framework. This increasing research led to the creation of the IETF Distributed Mobility Management (DMM) workgroup [8]. The workgroup has been focused on the identification of the centralized mobility limitations and the problem statement [4].

One of the first approaches in the scope of distributed mobility management was proposed in [9] and [10]. It was proposed and evaluated a network-based dynamic mobility anchoring solution for IPv6 mobile networks, where the PMIPv6 was used as the substrate. It distributes the traffic anchoring functionality of mobility management through all ARs of the network, while the mobility context management is maintained centralized. It has been shown through analytical models in [11] that the proposed dynamic mobility anchor solution outperforms PMIPv6.

The proposal in [12] is a conceptual approach for distributing mobility management based on PMIPv6. It distributes the mobility anchors, splitting the mobility functionalities into home network prefixes allocation, location management and routing of mobile traffic. The main focus of the work is to optimize the routing of mobile traffic. Indeed, the main objective is to discuss the overall design of a distributed mobility anchor approach, and to address its performance issues based on the available evaluations.

The approach in [13] provides an overview for two conceptual approaches: a partially and a fully-distributed one. The partially distributed approach splits the control and data planes, distributing only the data plane, while the fully distributed approach distributes both planes. The proposed approach assumes a CN in the same PMIPv6 domain of the MN, which may not be always true, and there is no evaluation nor any proof of concept.

There is a former approach [14] based on PMIPv6, which anchors the mobile traffic sessions at the last IP hop router, the AR. Besides the traditional AR functionalities, it also operates as a LMA or a MAG on a per MN and IPv6 prefix basis. It is a distributed PMIPv6 approach, similar to [9], which provides a detailed description of the required

signaling extensions. The authors suggest extensions to support inter-domain operation, based on a centralized LMA as top-level anchor to guarantee session continuity when crossing operator borders.

The study performed in [15] compares different distributed mobility management schemes based on a decoupling of the mobility functionalities into data management, handover management and location management. The evaluation showed that a distributed scheme may overall optimize the mobility management regarding data delivery optimization, while the signaling might increase for distributed schemes.

From the summarized works, it is observed a novel trend from the research community to develop distributed and dynamic schemes for the mobility management, based on the existing IPv6 mobility models. Distributed and dynamic mobility management is a promising research direction for flat network architectures, but novel distributed mobility schemes maintain the mobility context management centralized, and do not provide any solution when user roams to an access network without mobility support. Furthermore they do not provide a detailed description, a proper evaluation neither a validation in a real environment. This article goes further, providing a detailed description of a novel fully-distributed and dynamic scheme to provide global IP mobility management, DMIPA. DMIPA is compared with the current centralized IP mobility models (e.g. MIPv6), being evaluated through analytical models and simulations, as well as a validation through a real testbed.

C.3 New Trends in IP Mobility Management

The mobility management deals with the location functionalities for the establishment of new sessions, and with the session continuity of the ongoing sessions after a handover. Mobility management incorporates several functionalities, and we highlight the two main ones addressed in this article:

- **IP data anchoring:** anchor data sessions at IP layer in the mobility anchor, which is in charge of forwarding these data sessions to the current location of the mobile device, which is done by packets encapsulation/decapsulation according to the address translation rules (e.g. IP tunnels).

- **IP mobility context management:** management of mobility anchors and MNs IP addresses, as well as the bindings creation/update for routes maintenance through associations between previous and current MNs IP addresses.

The distribution of the location management functionality provides the discovery of the current location (IP address) of the MN. The location management is essential for communications among MNs, such as Voice over IP, where the IP address of the MN changes along the time. However, in applications/services hosted by static nodes (e.g. servers hosted within data centers of enterprise premises), the current location system, based on a conversion between a well know identification (e.g. URL) and a long lifetime IP address (e.g. Domain Name System (DNS)), is enough to cope with the distribution of data anchoring and mobility context. The distribution of the location management has been investigated through the adoption of dynamic/distributed DNS and Distributed Hash Table concepts, such as provided by Skype [16]. The location management is an important subject to achieve a full distributed mobility management solution. Approaches for distributed location management have been addressed in [17], and they can work in parallel with the envisioned dynamic and distributed IP mobility scheme.

There is a considerable change in the assumptions and challenges for the IP mobility management, due to the new behavior of the user (inherently mobile), and the way it consumes the required services. The impact of these changes on the existing IP mobility protocols shall be understood, in order to ensure a better performance of the IP mobility management, which influences the network performance and the user perceived experience.

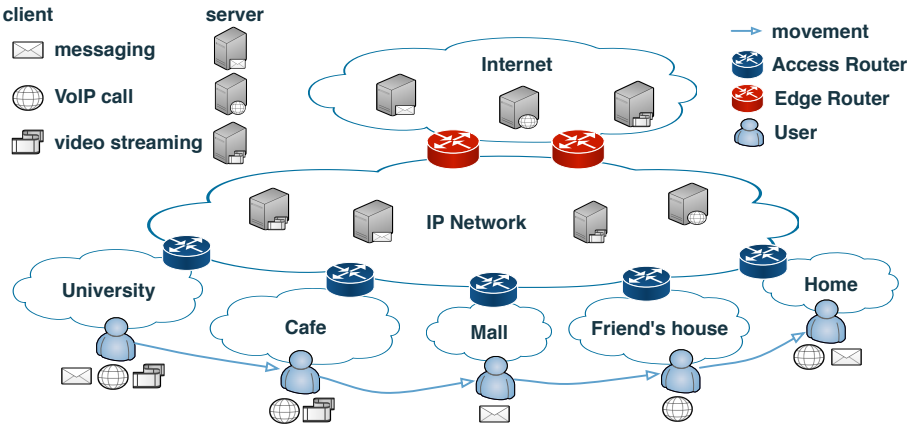


Figure C.1: Envisioned Scenario.

First, we describe the envisioned scenario (Figure C.1) for the IP mobility management, with its respective challenges and assumptions. As illustrated in Figure C.1, the user connects to different IP networks in its daily routine (e.g. University and Home), which might be provided from the same of different access technologies. Furthermore, the users already develop shared wireless networks, simply based on cooperation, becoming Internet micro-providers (e.g. Cafe and Friend's house). In these heterogeneous networks, the IP mobility management should be thought assuming different configurations from the access network, as well as the absence of mobility support.

Summarizing, the user is expecting to be supplied with an optimized IP mobility management model, which may be able to provide session continuity and reachability, independently of the heterogeneity of the network (e.g. technology and mobility support), and the heterogeneity of the required services (e.g. resources and location).

Current IP mobility management models are based on a static and centralized MA, which is responsible for IP data anchoring, mobility context management and signaling of a large number of MNs. Any type of mobile data or signaling packet from/to the MN is routed via the MA. Thus, the centralized MA, such as HA in MIPv6 or LMA in PMIPv6, brings problems and limitations when compared with a distributed and dynamic mobility perspective [4] [3].

We highlight the performance and scalability issues of maintaining the IP mobility management functionalities centralized in the MA, in comparison with a distributed and dynamic scheme. Figure C.2 illustrates an example of the behavior of current host-based IP mobility models (e.g. MIPv6) according to the envisioned scenario, and compare it with a distributed and dynamic host-based IP mobility scheme (e.g DMIPA).

C.3.1 IP Data Anchoring

Current mobility models adopt a centralized data anchoring, where the routing is performed via the centralized MA, which is usually distant from both the content server and the MN. Routing via a centralized point is usually longer and the centralized MA

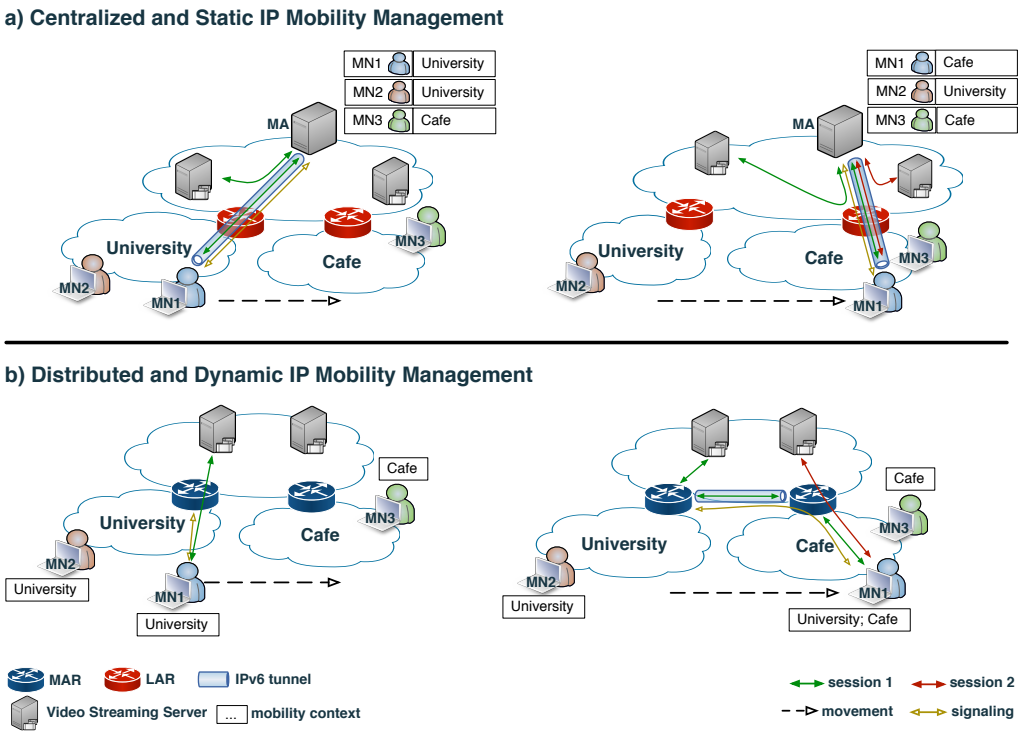


Figure C.2: Mobility Management Schemes.

may be the bottleneck of the network. It also increases both the operational cost of the network and the end-to-end delay of applications, and reduces the resources available in the core/backbone network. The distribution of the data anchoring functionality closer to the user, such as in the MARs of University and Cafe, optimizes the routing path and relieves the core/backbone network resources.

Centralized data anchoring also leads to a constant enforcement of encapsulation/decapsulation by the MA according to a long list of routing rules. This may introduce an extra delay in the MA, added to the non-optimized routing delay, and may require higher computational resources on the MA device. Furthermore, centralized mobility management models were envisioned to always provide mobility support to all sessions while the MN is connected to the network, even if a session is started and terminated in the same network. A dynamic data anchoring scheme provides dynamic provision of mobility support according to the MNs' sessions needs. The mobility support can be activated only for the ongoing sessions from the time that they endow an IP handover; otherwise, the session follows the optimized path without tunneling or any other mobility functionality.

Furthermore, a centralized point is commonly more vulnerable to failures or attacks, since a failure in the MA causes the loss of all sessions of all MNs. Distributing the data anchoring reduces the scope of a failure or an attack, since each MAR is responsible to anchor the data packets for a set of sessions of a confined group of MNs.

C.3.2 IP Mobility Context Management

The centralized MA needs to manage the mobility context and the mobility routes of all MNs to maintain the sessions continuity after a handover. These routes are usually maintained through bindings and enforced through tunnels. Maintaining the mobility context and routes centralized in the MA increases the signaling required for the mobility

management, since all updates and requests are exchanged with the MA, that usually is distant from the current location of the MN. Distributing the mobility context management through the MNs, being each MN responsible to maintain its mobility context and routes, improves the mobility management performance. When the MN roams between IP networks, it carries its own updated mobility context and routes, which can be immediately used to maintain ongoing sessions active through the fast update of the set of MARs of the MN.

Centralized mobility context and routes inherently increases the handover latency, since to maintain the previous sessions of the MN, the MN has to retrieve the mobility context and routes from the centralized MA. Similarly to the case of data anchoring, a centralized point for context and routes management is commonly more vulnerable to failures or attacks. Distributing mobility context and routes' management reduces the scope of a failure or an attack, since a failure or attack in the MN only affects its mobility context and routes, which at maximum results in the loss of the handover sessions of the MN.

C.4 Towards a Distributed and Dynamic IP Mobility Scheme

This section details DMIPA, a distributed and dynamic host-based mobility scheme prepared for heterogeneous and flat network architectures.

C.4.1 DMIPA Overview

DMIPA distributes the IP mobility functionalities through MNs and MARs. While the IP data anchoring is distributed through the MARs, the IP mobility context management is distributed through the MNs, where each one is responsible for the management of its own IP mobility context. Thus, DMIPA completely eliminates the necessity of centralized entities to forward data or to maintain the mobility context of the MNs.

DMIPA also provides dynamic IP mobility, which means that IP mobility support is offered to a session when it is really needed. Thus, the IP mobility support is just provided for ongoing sessions from the time they endow an IP handover, or if the current AR of the MN does not support mobility; otherwise, they are maintained without any mobility support or tunneling mechanism. DMIPA provides dynamic IP anchoring for traffic sessions, whereas new sessions are always anchored to the most recently attached MAR, while ongoing sessions remain anchored to the previous MARs.

DMIPA was envisioned to support global IP mobility in heterogeneous networks. Thus, when a MN attaches to a MAR, it ensures the forwarding of the ongoing sessions from previous MARs; otherwise, these functionalities are provided by the MN. Thus, the scheme provides two modes to overcome the scenarios when a MN attaches to a MAR or when a MN attaches to a Legacy AR (LAR).

MAR

From the IP network prefix and the mobility support indicator received from the attached MAR, MN configures a new IP address with preferred status, while other addresses received from previous MARs are maintained just for the ongoing sessions. MN adds the attached MAR to its list of MARs, and sends this information and the respective MN IP addresses to the attached MAR. Then, the attached MAR establishes tunnels with previous MARs to maintain the ongoing sessions anchored there, while new sessions are established through the attached MAR.

LAR

From the IP network prefix and the absence of mobility support indicator received from the AR, MN configures a new IP address, and maintains the IP addresses obtained from previous MARs for ongoing sessions. The IP address configured from the most recently attached MAR is the preferred one. MN establishes tunnels with previous MARs to maintain the ongoing sessions. New sessions are established through the tunnel with the most recently attached MAR, that is the preferred one to anchor new sessions.

C.4.2 Protocol Operation

The protocol operation is explained with the example of Figure C.3 and the support of the diagram of Figure C.4.

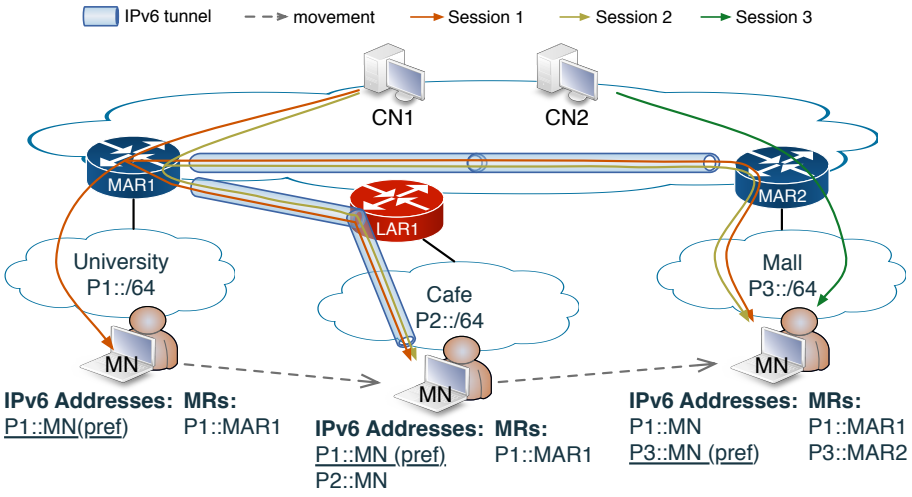


Figure C.3: DMIPA overview.

It is assumed a stateless IPv6 configuration, where the MN is intrinsically aware of a new IPv6 prefix address, from the Router Advertisement (RA) messages received from any AR of the network.

1. After a new MAC association with an AP/BS connected to MAR1, MN sends a Router Solicitation (RS) message to the network, in order to request the IPv6 network prefix and related information. MAR1 answers to the RS with a RA messages, which is configured to provide the IPv6 address of the MAR1 interface and the mobility support indicator.

2. MN configures a new IPv6 address (P1::MN) based on the information received from the RA message and the MAC address of the MN interface, which is defined as the *preferred* IPv6 address to establish new sessions.

3. MN adds the MAR1 IPv6 address (P1::MAR1) to the set of its current MARs, and it is the new preferred MAR to anchor new sessions.

4. MN initiates new sessions, such as Session 1, with the IPv6 address P1::MN through MAR1, without any tunneling or mobility support.

5. After a new MAC association with an AP/BS connected to LAR1, MN sends a RS message to the network. LAR1 answers to the RS with a default RA message, containing the IPv6 prefix of the network.

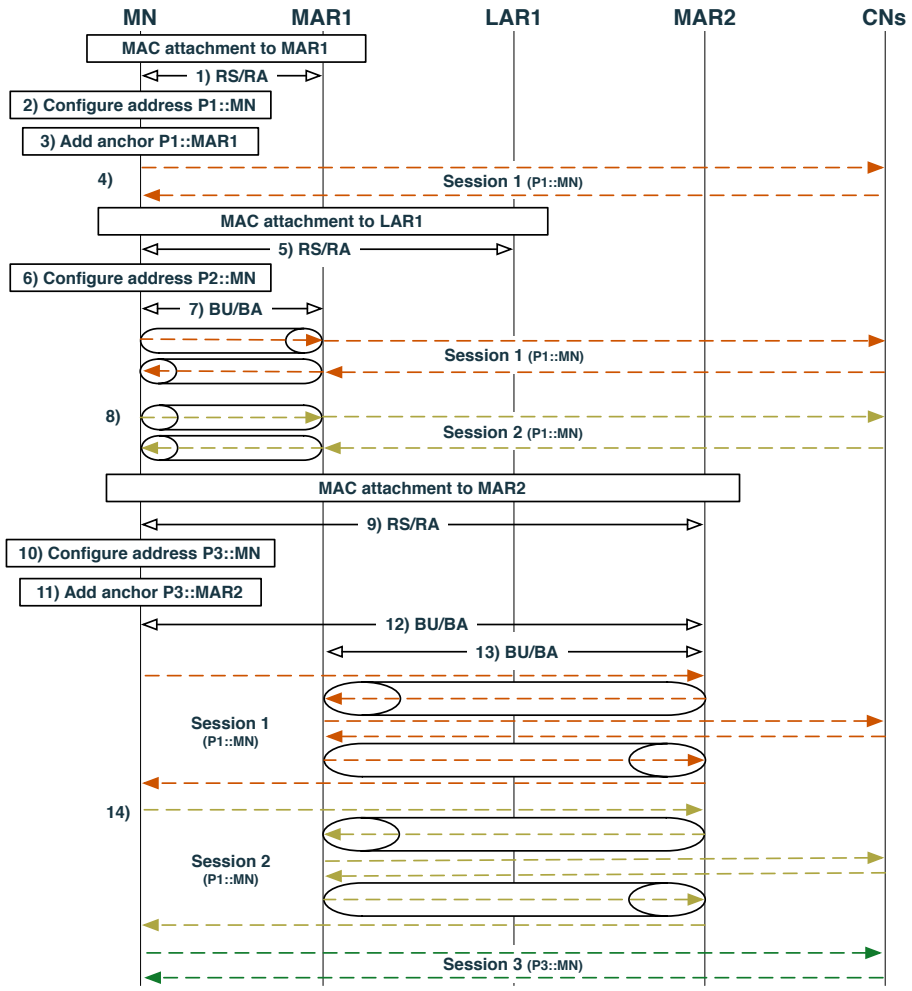


Figure C.4: DMIPA Operation.

6. MN configures a new IPv6 address (P2::MN) based on the prefix received from the RA message and the MAC address of the MN interface. The new IPv6 address based on the prefix received from a LAR is configured in the *deprecated* state, while the IPv6 address (P1::MN) received from MAR1 remains the *preferred* one to establish new session with session continuity support.

7. MN establishes a tunnel with MAR1 through Binding Update (BU) / Binding Acknowledge (BA) messages, in order to maintain the ongoing sessions anchored to MAR1, as well as to be able to establish new sessions with session continuity support. Thus, packets to P1::MN received by MAR1 are tunneled to P2::MN, and packets from P1:MN sent by MN are tunneled to P1::MAR1.

8. Session 1 is maintained through the configured tunnel between MN and MAR1, as well as the new Session 2 established while MN is attached to LAR1.

9. After a new MAC association with an AP/BS connected to MAR2, MN sends a RS message to the network. MAR2 answers to the RS with a RA message, which is configured to provide the IPv6 address of the MAR2 interface and the mobility support indicator.

10. MN configures a new IPv6 address (P3::MN) based on the IPv6 prefix received from the RA message and the MAC address of the MN interface. The new IPv6 prefix received from a MAR is configured as the *preferred* one for the establishment of new sessions, while the IPv6 address (P1::MN) received from MAR1 is maintained in the *deprecated* state just to provide session continuity to sessions anchored to MAR1.

11. MN adds the MAR2 IPv6 address (P3::MAR2) to the set of its current MARs, and it is the new preferred MAR to anchor new sessions.

12. MN exchanges BU/BA messages with the attached MAR2 to provide the information about current MARs and respective MN IPv6 addresses. The BU message contains the IPv6 address of MAR1 (P1::MAR1) and the respective MN IPv6 address (P1::MN), which is needed by MAR2 to establish the tunnels with other MARs of the MN (MAR1).

13. After receiving the pairs of MARs IPv6 addresses and respective MNs IPv6 addresses from MN, MAR2 establishes tunnels with previous MARs. MAR2 establishes a tunnel with MAR1 to maintain the ongoing sessions of MN anchored in MAR1. Thus, packets to P1::MN received by MAR1 are tunneled to P3::MAR2, and packets from P1:MN received by MAR2 are tunneled to P1::MAR1.

14. Sessions 1 and 2 are maintained through the configured tunnel between MAR1 and MAR2, while new sessions, such as Session 3, are established with the new IPv6 address P3::MN through MAR2, without any tunneling or mobility support.

The IPv6 addresses from LARs do not need to be maintained when the MN is not attached to them, since these IPv6 addresses are not used to anchor sessions that need session continuity support. The IPv6 addresses from MARs are maintained in the MN, as long as the MN is using them for ongoing sessions; otherwise, these IPv6 addresses are removed, as well as the corresponding bindings/tunnels. In step 14) of Figure C.4, if sessions 1 and 2 are terminated, the IPv6 addresses P1::MN and P1::MAR1 are removed from the interface and the set of MARs respectively, as well as the tunnel between MAR1 and MAR2, since there are no more sessions anchored to MAR1.

C.5 Analytical Modeling

In this section we develop the analytical models for both MIPv6 and DMIPA. The analytical models presented are based on the work of [5], where the analytical models were adjusted to fit the envisioned scenarios, and the tunneled packets metric was added. The adopted notations and their default values can be found in Table C.1. We develop the equations to evaluate Tunneled Packets, Signaling Cost and Data Cost:

Tunneled Packets: the total percentage of data packets being tunneled by the MARs/HA.

Signaling Cost: the total cost needed for mobility management control messages, defined as the size of the messages multiplied by their remaining time in the network, between source and destination endpoints.

Data Cost: the end-to-end cost to deliver data packets from the CN to the MN, defined as the data packets size multiplied by the remaining time in the network, between the source and destination.

We adopt the model proposed in [18] to calculate the wireless transmission time $\boldsymbol{T_x}$, where $\boldsymbol{x}$ is the type of the packet which will be sent between the MN and the AR or

Table C.1: Notations and values used in the analytical results.

Notation	Definition	Default Value
N_m	Average Number of associated MARs per MN per second	2
S_{bu},S_{ba},S_{bu2},S_{ba2}	Size of the BU/BA/BU_2/BA_2 message	56, 56, 56+32N_m, 56 bytes
S_d, S_{dt}	Size of a data/tunneled data packet	1024, 1024+40 bytes
B_{wl},B_{wr}	Bandwidth of a wireless/wired link	10, 100 Mbps
L_{wl},L_{wr}	Latency of a wireless/wired link	2, 1 ms
N_{ar}	Number of ARs in the network	20
P_f	Wireless link failure probability	0.1
T_s	Session service mean time	240 s
μ_s	Session service mean rate	1/240 s^{-1}
T_c	AR residence/handover mean period	300 s
μ_c	AR residence/handover mean rate	1/300 s^{-1}
P_m	Probability of an AR configured as a MAR	1
$D_{wl}(x)$	Delay of a wireless hop for a packet of type x	-
$D_{wr}(x)$	Delay of a wired hop for a packet of type x	-
$H_{c\to h}$	Average number of hops between a CN and the HA	10
$H_{c\to a}$	Average number of hops between a CN and a AR	10
$H_{h\to a}$	Average number of hops between the HA and a AR	5
$H_{a\to a}$	Average number of hops between two ARs	5
$H_{m\to a}$	Average number of wireless hops between a MN and the connected AR	1

vice-versa, $D_{wl}(x)$ is the wireless delay to send a packet of type x, and n_f is the number of transmission failures.

$$\begin{aligned} T_x &= H_{m\to a} \times D_{wl}(x) \\ &+ \sum_{n_f}^{\infty} n_f \times P\{n_f \text{ failures and 1 success}\} \times H_{m\to a} \times D_{wl}(x) \\ &= \frac{H_{m\to a}}{\overline{P_f}} \times D_{wl}(x) \end{aligned} \tag{C.1}$$

The delay added by each hop to the packet x is defined in (C.2), where wr/wl is the wired or wireless medium access, and S_x is the size of the packet x. $B_{wr/wl}$ and $L_{wr/wl}$ are the bandwidth and latency, respectively, of a wired/wireless link.

$$D_{wr/wl}(x) = \frac{S_x}{B_{wr/wl}} + L_{wr/wl} \tag{C.2}$$

C.5.1 Tunneled Packets

The Tunneled Packet (TP) models the percentage of tunneled packet needed to ensure session continuity support.

MIPv6

In MIPv6 all data packet are tunneled by the HA from the establishment of the session, thus $TP_{MIPv6} = 100\%$.

DMIPA

The probability of a data packet to be a handover packet (P_h) is defined by [11] and it is given by (C.3). It is calculated assuming that the sessions arrival is a Poisson process and that the service time is exponentially distributed with the average μ_s, as well as the cell residence time with the average μ_c.

$$P_h = \frac{\mu_c}{\mu_c + \mu_s} \tag{C.3}$$

The probability of a data packet to be a tunneled packet for DMIPA should consider not only the handover packet probability, but also the probability to be connected to a MAR (P_m), which depends on the number of ARs configured as MARs, and the probability of the CN and MN to be connected to the same AR. The conditional probability of a MN

to be connected or perform a handover to a AR, where the CN is connected, is shown in (C.4). It is considered an equal probability of a CN to be connected to any AR of the network. If it is assumed that the CNs are not communicating through the ARs, but they can be reached through an edge router, the $\boldsymbol{P_s}$ is zero.

$$\boldsymbol{P_s = \frac{1}{N_{ar}} \times \frac{N_{ar}-1}{N_{ar}-1} + \frac{N_{ar}-1}{N_{ar}} \times \frac{1}{N_{ar}-1} = \frac{2N_{ar}-2}{N_{ar} \times (N_{ar}-1)}} \tag{C.4}$$

The probability of a data packet be a tunneled packet is given by the expression (C.5).

$$\boldsymbol{TP_{DMIPA} = \overline{P_s} \times P_m \times P_h + \overline{P_m}} \tag{C.5}$$

C.5.2 Signaling Cost

The Signaling Cost (SC) models the cost of exchanging control messages to ensure the mobility management.

MIPv6

There are two control messages exchanged to update the HA: BU and BA. The total signaling cost is the addition of the costs of the two messages as defined in (C.6), considering the number of hops and the delay added by each hop. The costs of the BU and BA consider the wireless hops between the MN and the attached AR, and the wired hops between the AR and HA.

$$\begin{aligned} \boldsymbol{SC_{MIPv6}} = \boldsymbol{S_{bu}}\Big(\boldsymbol{\frac{H_{m\to a}D_{wl}(bu)}{\overline{P_f}} + H_{h\to a}D_{wr}(bu)}\Big) \\ \boldsymbol{+S_{ba}}\Big(\boldsymbol{H_{h\to a}D_{wr}(ba) + \frac{H_{m\to a}D_{wl}(ba)}{\overline{P_f}}}\Big) \end{aligned} \tag{C.6}$$

DMIPA

The DMIPA signaling cost is composed by two main cases, as presented in (C.7). The first case ($\boldsymbol{\beta_1}$) represents the MARs update from a MN connected to an AR, while the seconds case ($\boldsymbol{\beta_2}$) is the MARs update from a MN connected to a MAR.

$\boldsymbol{\beta_1}$: MN sends BU messages directly to its current set of MARs to create tunnels, since the MN is connected to an AR, and it receives the respective BAs.

$\boldsymbol{\beta_2}$: MN sends a BU message to its attached MAR with its set of MARs, and then it receives the BA message from the MAR. Then, the attached MAR establishes tunnels with other MARs of the MN through BU and BA messages.

$$\boldsymbol{SC_{DMIPA} = \overline{P_m} \times SC(\beta_1) + P_m \times SC(\beta_2)} \tag{C.7}$$

The $\boldsymbol{SC(\beta_1)}$ in (C.8) is obtained multiplying the average number of active MARs per MN per second, the size of BU and BA messages, the hops in wireless and wired, and the average wireless and wired hop delay. In the wireless hops we also consider the transmission failure probability $\boldsymbol{P_f}$.

$$\begin{aligned} \boldsymbol{SC(\beta_1)} = \boldsymbol{N_m}\Big(\boldsymbol{S_{bu}\frac{H_{m\to a} \times D_{wl}(bu)}{\overline{P_f}} + S_{bu}H_{a\to a} \times D_{wr}(bu)} \\ \boldsymbol{+S_{ba}\frac{H_{m\to a} \times D_{wl}(ba)}{\overline{P_f}} + S_{ba}H_{a\to a} \times D_{wr}(ba)}\Big) \end{aligned} \tag{C.8}$$

In the first part of $\boldsymbol{SC(\beta_2)}$, depicted in (C.9), it is represented the cost of transmitting $\boldsymbol{BU_2}$ and $\boldsymbol{BA_2}$ messages in the wireless network, considering a transmission failure probability $\boldsymbol{P_f}$. In the second part, it is added the cost of transmitting BU and BA messages

between the attached MAR of the MN and the other MARs of the MN.

$$\begin{aligned} SC(\beta_2) = \frac{H_{m\to a}}{\overline{P_f}}\Big(S_{bu2}D_{wl}(bu2) + S_{ba2}D_{wl}(ba2)\Big) \\ +N_m H_{a\to a}\Big(S_{bu}D_{wr}(bu) + S_{ba}D_{wr}(ba)\Big) \end{aligned} \tag{C.9}$$

C.5.3 Data Cost

In IPv6 networks, the Data Cost (DC) is proportional to the delay between the source and the destination, as well as the size of the data packet payload.

MIPv6

In MIPv6, the DC depicted in (C.10) is always the same, since it follows the same path; from the CN to HA and then from HA to MN. The first part corresponds to the first term of the equation, in which the packet is a normal IPv6 data packet from CN to HA, while from HA to MN the data packets are tunneled, traversing both wired and wireless hops.

$$DC_{MIPv6} = S_d H_{c\to h} D_{wr}(d) + S_{dt} H_{h\to a} D_{wr}(dt) + S_{dt}\frac{H_{m\to a} D_{wl}(dt)}{\overline{P_f}} \tag{C.10}$$

DMIPA

In DMIPA, we can distinguish 3 main cases regarding the data cost:

γ_1**:** MN directly communicates in normal IPv6 with CN through its current MAR; it is the optimized routing path.

γ_2**:** MN communicates through direct tunnels with the previous MAR, while it is attached to a LAR.

γ_3**:** MN communicates through tunnels established between the previous MARs and the attached MAR.

The DC_{DMIPA} in (C.11) gives the data delivery cost based on the probabilities distribution of the previous three cases.

$$DC_{DMIPA} = P_m\Big(\overline{P_h} \times DC(\gamma_1) + P_h \times DC(\gamma_3)\Big) + \overline{P_m} \times DC(\gamma_2) \tag{C.11}$$

In $DC(\gamma_1)$ presented in (C.12), the packets do not require handover, so the cost contains the wired and wireless hops of the optimized path multiplied by the size of the data packet payload.

$$DC(\gamma_1) = S_d\Big(H_{c\to a} D_{wr}(d) + \frac{H_{m\to a}}{\overline{P_f}} D_{wl}(d)\Big) \tag{C.12}$$

In $DC(\gamma_2)$ presented in (C.13), the first part of the equation represents the cost of a data packet between the CN and a previous MAR of MN, while the second part considers a tunneled data packet from a previous MAR to MN, including both the wired and wireless hops.

$$DC(\gamma_2) = S_d H_{c\to a} D_{wr}(d) + S_{dt}\Big(H_{a\to a} D_{wr}(dt) + \frac{H_{m\to a}}{\overline{P_f}} D_{wl}(dt)\Big) \tag{C.13}$$

In $DC(\gamma_3)$ in (C.14), the first part of the equation is the cost of a original data packet, from a CN to a previous MAR of MN plus the wireless hops between the attached MAR and MN. The second part represents the cost of a tunneled data packet between a previous MARs and the MAR where MN is attached to.

$$DC(\gamma_3) = S_d\Big(H_{c\to a} D_{wr}(d) + \frac{H_{m\to a}}{\overline{P_f} D_{wl}(d)}\Big) + S_{dt} H_{a\to a} \times D_{wr}(dt) \tag{C.14}$$

C.6 Evaluation

This section evaluates and compares DMIPA with MIPv6 through analytical and simulation results, adopting different metrics. The analytical results compare DMIPA with MIPv6 based on the signaling cost, data cost and tunneled packets for different scenarios, where the parameters of the model are changed. The analytical results are validated through a comparison with simulated results in a specific scenario, where both methods can be directly compared. The evaluation is also performed through a scenario developed in the ns-3 environment [19]. Besides the metrics evaluated through the analytical models, the simulation environment also evaluates the average data delay, the number of hops with tunneled packets, the average bindings per MA per instant, and average MAs per MN per instant. Finally, the evaluation section validates DMIPA in a real environment through a testbed, where the packets delay, bitrate and tunneled packets are evaluated for UDP and TCP sessions.

C.6.1 Analytical Results

This section evaluates DMIPA and MIPv6 according to the analytical model presented in the previous section. The signal cost evaluated is the average signaling per handover, while the data cost evaluated is the average data per data packet exchanged between the CN and the MN. The other values used in the evaluation of the scenarios are the ones presented in Table C.1. The analytical signaling cost is measured in a handover of a single MN, which is the cost of signaling when a MN moves to another AR; thus, the number of MNs is not relevant to the analytical model. The analytical data cost is measured per average data packet, which is the average cost to exchange a data packet between CN and MN. The analytical model depends on the hops distance between the network elements, and does not depend on the topology nor number of routers and ARs, where this information is integrated in the values defined for the hops distance.

In the evaluation, we assume that, in average, $\boldsymbol{H_{c\to h} = H_{c\to a}}$, and $\boldsymbol{H_{a\to a} = H_{h\to a}}$. The number of hops in the wireless part is one: $\boldsymbol{H_{m\to a} = 1}$. We also define $\boldsymbol{\alpha}$ according to the equation (C.15):

$$\boldsymbol{\alpha = \frac{H_{c\to h}||H_{c\to a}}{H_{a\to a}||H_{h\to a}}} \tag{C.15}$$

All results presented are calculated considering the ratio between DMIPA and MIPv6: any $\boldsymbol{Metric(m) = DMIPA(m)/MIPv6(m)}$, where $\boldsymbol{m}$ is the evaluated metric.

C.6.1.1 Tunneled Packets

In Figure C.5, we consider a mean AR residence time ($\boldsymbol{T_c}$) of 300s and the number of ARs ($\boldsymbol{N_{ar}}$) as 20. The mean session service time ($\boldsymbol{T_s}$) and the probability of an AR to be a MAR ($\boldsymbol{P_m}$) are changed in order to understand their impact on the tunneled packets in DMIPA, when compared with MIPv6. The number of tunneled packets in DMIPA is strongly reduced for low values of $\boldsymbol{T_s}$ and high values of $\boldsymbol{P_m}$, when compared with MIPv6. For scenarios where all ARs are mobility-enabled (MARs) and the MNs are quite static ($\boldsymbol{T_C >> T_S}$), DMIPA might achieve huge reductions in tunneled packets, with values around 10% of MIPv6. Lower values of $\boldsymbol{T_s}$ decrease the probability of having handover packets, since most of the sessions are terminated before the handover of the MN, or they are in the handover situation for short periods. The increase of ARs configured as MARs in the network, increase the number of packets delivered through the optimized path, without any mobility support, which is provided for ongoing sessions just after the handover of the MN.

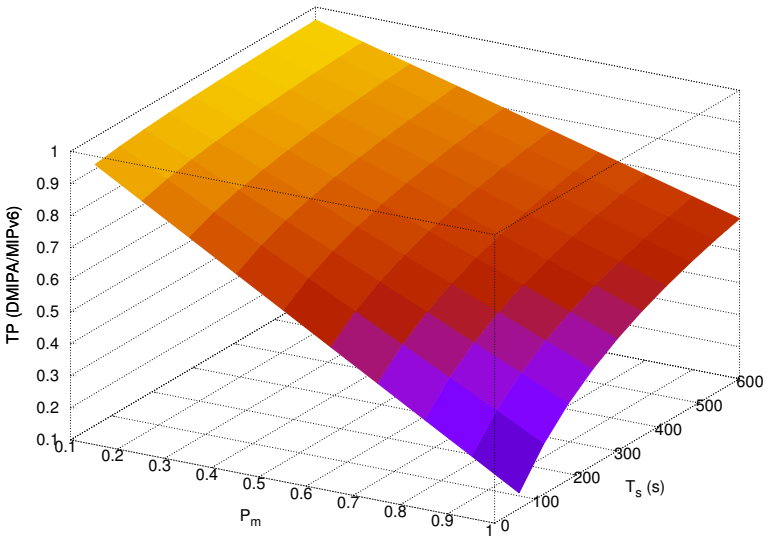


Figure C.5: Tunneled Packets (DMIPA/MIPv6) varying $\boldsymbol{P_m}$ and $\boldsymbol{T_s}$.

In Figure C.6, the influence of $\boldsymbol{T_s}$ and $\boldsymbol{T_c}$ in the tunneled packets is evaluated, for a $\boldsymbol{P_m}$ of 1 and $\boldsymbol{N_{ar}}$ of 20. The number of tunneled packets in DMIPA strongly reduces with the increase of $\boldsymbol{T_c}$ and the decrease of $\boldsymbol{T_s}$, where the probability of having more packets following the optimized routing path increases. The residence time of a MN per AR ($\boldsymbol{T_c}$) is longer, reducing the frequency of IP handovers, or the duration of the sessions are shorter, reducing the number of handover packets per session. Even in the worst case for DMIPA in the evaluated scenario (high $\boldsymbol{T_s}$ and low $\boldsymbol{T_c}$), the number of tunneled packets is lower than in MIPv6. In the scenario of MNs without handover sessions, where $\boldsymbol{T_c} \rightarrow \boldsymbol{\infty}$ (static/quasi-static MNs or MNs turned off in movement), the number of tunneled packets is null, since the dynamic mobility support for session continuity does not need to be enabled.

C.6.1.2 Signaling Cost

In Figure C.7, the $\boldsymbol{H_{c\rightarrow a}} = \boldsymbol{H_{c\rightarrow h}} = \boldsymbol{H_{h\rightarrow a}} = 10$, and the $\boldsymbol{H_{a\rightarrow a}}$ is changed from 1 to 10, while the other values are the default ones from Table C.1. The signaling cost of DMIPA (Figure C.7) is higher than the signaling cost of MIPv6 for a higher number of active MARs per MN ($\boldsymbol{N_m}$). Since $\boldsymbol{H_{a\rightarrow a}} \approx \boldsymbol{H_{h\rightarrow a}}$, DMIPA needs more messages to update the set of MARs of the MN. However, for quite static MNs (and considering $\boldsymbol{H_{a\rightarrow a}} \approx \boldsymbol{H_{h\rightarrow a}}$), such as people that spend long periods at home and work, where the average number of MARs per instant ($\boldsymbol{N_m}$) is less than one, the signaling cost of DMIPA is reduced when compared with MIPv6. In scenarios of MNs without handover sessions, the signaling cost is null, since the dynamic mobility support for session continuity does not need to be provided. The evaluated $\boldsymbol{P_m}$ of the analytical modeling does not reflect its direct influence in the $\boldsymbol{N_m}$, but there is a proportional relation between $\boldsymbol{P_m}$ and $\boldsymbol{N_m}$, since more MARs in the network ($\boldsymbol{P_m}$) increase the average number of active MARs per MN, leading to more control messages to be exchanged when an IP handover occurs. Moreover, the $\boldsymbol{N_m}$ also depends on the relation between the ARs residence time ($\boldsymbol{T_c}$) and the sessions

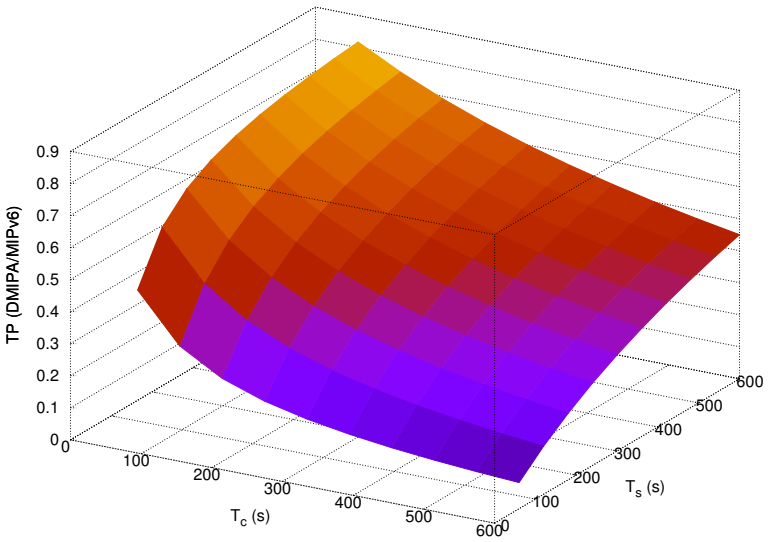


Figure C.6: Tunneled Packets (DMIPA/MIPv6) varying $\boldsymbol{T_c}$ and $\boldsymbol{T_s}$.

service time ($\boldsymbol{T_s}$). The $\boldsymbol{P_m}$ evaluated in this model just impacts the signaling strategy used to exchange the control messages, $\boldsymbol{\beta_1}$ for an AR or $\boldsymbol{\beta_2}$ for a MAR. The increase of $\boldsymbol{P_m}$ slightly reduces the signaling cost of DMIPA for a higher number of $\boldsymbol{N_m}$, since the MN just exchanges two control messages with the attached MAR through the wireless medium, containing the active MARs and respective bindings, and then the attached MAR exchanges BU/BA messages with others to establish the tunnels. Otherwise, the MN directly exchanges BU/BA messages with active MARs, increasing the number of control messages in the wireless access medium.

The decrease of $\boldsymbol{H_{a\rightarrow a}}$, in comparison with $\boldsymbol{H_{h\rightarrow a}}$, reduces the DMIPA signaling cost when compared with MIPv6. For a $\boldsymbol{H_{a\rightarrow a}}$ equal to half of $\boldsymbol{H_{h\rightarrow a}}$, DMIPA signaling cost is lower than MIPv6 for a $\boldsymbol{N_m}$ less than 1.9, which enables a MN to have around two MARs in average along the time, for the same cost of MIPv6 with its single mobility anchor (HA).

C.6.1.3 Data Cost

The data cost to deliver the packets from traffic session is evaluated in Figure C.8, where $\boldsymbol{T_c}$ and $\boldsymbol{T_s}$ are changed, and the other configured parameters are the ones defined in Table C.1. Considering the set of values tested for $\boldsymbol{T_c}$ and $\boldsymbol{T_s}$, DMIPA always presents a lower data cost, when compared with MIPv6. The DMIPA data cost becomes closer to the MIPv6 data cost for low values of $\boldsymbol{T_c}$ and high values of $\boldsymbol{T_s}$. These merged conditions increase the percentage of tunneled packets from the handover sessions, where most of the sessions are tunneled nearly their lifetime, from a MAR where the MN is not connected. Considering a quite static MN (e.g. people spending long periods at home and work), the data cost of DMIPA is strongly reduced when compared with MIPv6, since most of the sessions are terminated without the need for tunneling mechanisms. If the MN is able to communicate without handover sessions, the data cost always follows the optimized path provided by the routing, reducing the data cost to 70% of MIPv6 for the considered

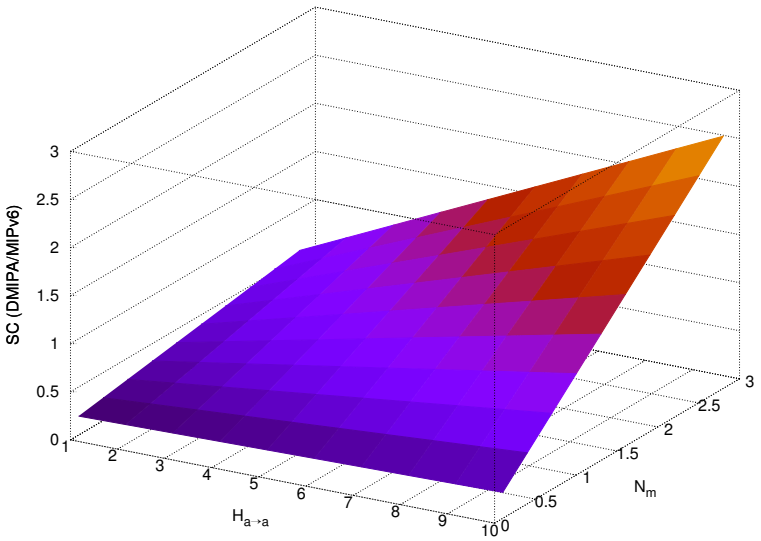


Figure C.7: Signaling Cost (DMIPA/MIPv6) varying $\boldsymbol{H_{a\to a}}$ and $\boldsymbol{N_m}$.

scenario.

In Figure C.9, $\boldsymbol{T_s}$ and $\boldsymbol{T_c}$ are defined with 240 and 300 seconds respectively, changing $\boldsymbol{\alpha}$ and $\boldsymbol{P_m}$. The data cost of DMIPA reduces with the increase of $\boldsymbol{\alpha}$, since the weight in hops inside the operator network tends to achieve the same value of the path between CNs and operator network. One of DMIPA advantages is to reduce the tunneled/forwarded packets inside the operator network; thus, the data cost is lower for higher values of $\boldsymbol{\alpha}$ when compared with MIPv6. The probability to connect to a MAR ($\boldsymbol{P_m}$) does not have a great impact on the data cost of DMIPA, for lower values of $\boldsymbol{\alpha}$, when compared with MIPv6, since the number of hops between ARs ($\boldsymbol{H_{a\to a}}$) is negligible when compared with $\boldsymbol{H_{c\to a}}$ and $\boldsymbol{H_{c\to h}}$. Thus, the impact of having more packets forwarding/tunneling inside the operator network is nearly insignificant. Otherwise, for higher values of $\boldsymbol{\alpha}$, the enabling of more ARs with mobility (more MARs) strongly decreases the data cost of DMIPA, when compared with MIPv6. More MARs introduce the possibility to initiate sessions through the current AR through the optimize path, while when the MN is connected to a LAR, the sessions are initiated through the previous MAR (MN is not directly connected), which forwards the packets to the current network of the MN.

C.6.2 Simulation

This section evaluates DMIPA, which has been implemented considering the assumptions and guidelines previously defined. The evaluation is achieved through a case study that represents IEEE 802.11 wireless networks with different MN mobility patterns, traffic characteristics and availability of MARs. The evaluation is performed in the ns-3 simulation environment [19]. We also compare DMIPA against the centralized MIPv6. Figure C.10 illustrates the topology of the evaluated network.

In the evaluated scenario, MNs move with a Steady State Random Way Point model among the wireless networks with different IPv6 domains, initiating and terminating traffic sessions with the CNs. This mobility model provides MNs movement to the area around

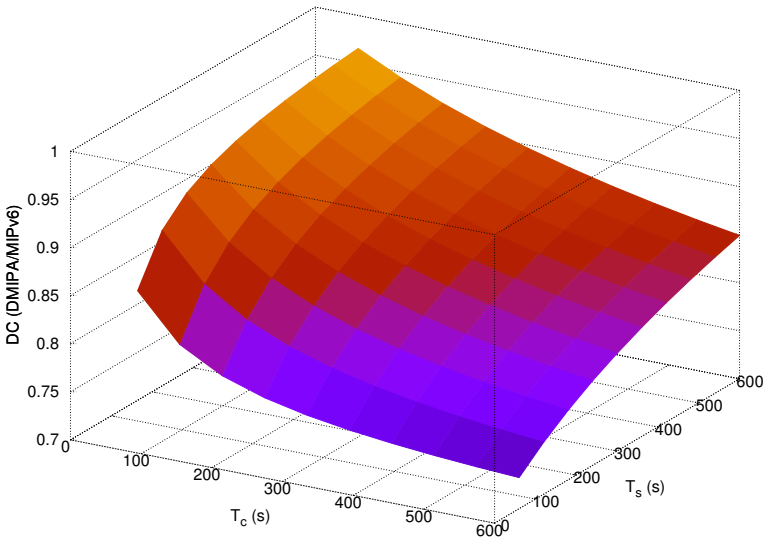


Figure C.8: Data Cost (DMIPA/MIPv6) varying $\boldsymbol{T_c}$ and $\boldsymbol{T_s}$.

interesting points with attraction capabilities, after the MNs pause times. In the simulated scenario, each AR is collocated with a single AP to simplify the simulations, since our goal is to study the movement of the MN among ARs. Thus, a handover between two APs in the simulation is the same as a handover between two ARs, which is usually attached to several APs in a real environment, such as an university, a mall or a neighborhood. The MN pause time used in the evaluation is considered inside an AR domain, even if MNs attach to several APs connected to the same AR during the pause time.

The set-up scenario is configurable, and these are the main settings:

- Grid topology with 9 wireless APs collocated with ARs, ensuring full connectivity coverage.
- Wireless configured with Friis Propagation Model with a range of 100 meters (e.g. delay depends on the load).
- Steady-State Random Way Point model with pause times around 90, 180 and 360 seconds, and speed of 3 ± 1 m/s.
- Stateless IPv6 configuration with Router Advertisement Daemon installed in ARs.
- Portion of Mobility-enabled ARs: nearly half of the ARs with 44.4% (4 MARs in 9 ARs) and total of the ARs with 100%.
- Wired links configured with a delay of 2 ms.
- Video/Audio Streaming sessions with a constant-bit rate of 128 Kbps and a packet payload size of 1Kbyte, under UDP from CN to MN.
- Streaming sessions inter-arrival time per MN is exponentially distributed with an average of 60 seconds.

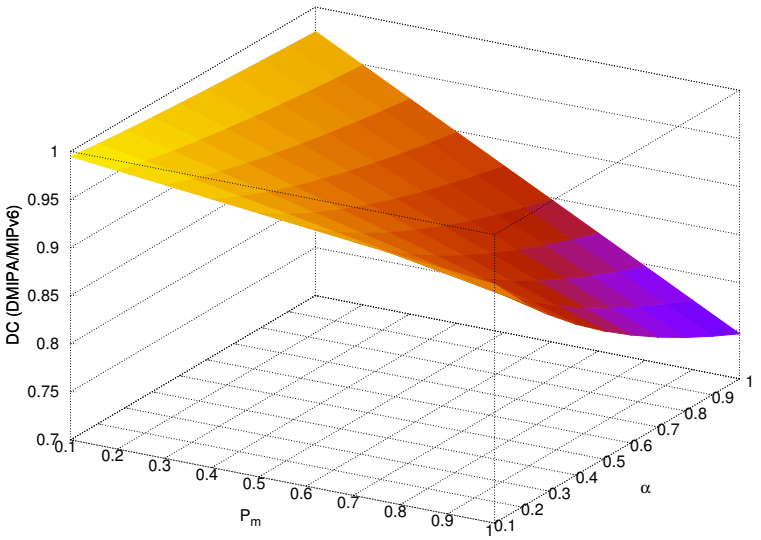


Figure C.9: Data Cost (DMIPA/MIPv6) varying $\boldsymbol{P_m}$ and $\boldsymbol{\alpha}$.

- Streaming sessions service time ($\boldsymbol{T_s}$) is exponentially distributed with an average of 2, 4 or 8 minutes.
- Results obtained from 10 independent simulations of 1200s with a confidence interval of 95%.

C.6.2.1 Analytical Validation

Before evaluating DMIPA through the ns-3 simulation environment [19], the analytical model is validated through a comparison with the simulation in a specific scenario. The setup previously defined is the platform used to validate the analytical models, where 30 MNs are used with a traffic rate of 16Kbps, both $\boldsymbol{T_c}$ and $\boldsymbol{T_s}$ are varied during the evaluation. In these tests with more MNs, the traffic rate is reduced, in order to obtain the results in a reasonable time; however, the session inter-arrival rate is maintained, which is the important requirement to maintain ongoing sessions in the MN. The values used to validate the analytical model are presented in Table C.2, where some of them are defined according to the respective average values from the simulated scenarios. The analytical signaling cost evaluates the cost per handover; thus, it is multiplied by the number of MNs and the number of handovers during the simulated scenario. The analytical model for data cost measures the average data packet cost; thus, it is multiplied by the number of data packets exchanged during the simulated scenarios to match the simulated condtions.

The tunneled packets validation, presented in Figure C.11, is performed for a $\boldsymbol{T_s}$ of 240 seconds. The results of both the analytical model and simulation show the same behavior of tunneled packets for 44.4% and 100% of ARs configured as MARs. The percentage of tunneled packets decreases with the increase of pause time, since more packets are optimally delivered without tunneling mechanisms. The higher confidence interval for 44.4% of MARs is due to the random placement of these MARs in the ARs of the network, as well as the exponential distribution of the sessions service and inter-arrival times, which introduces a higher variability of the tunneled packets. The analytical results are a little

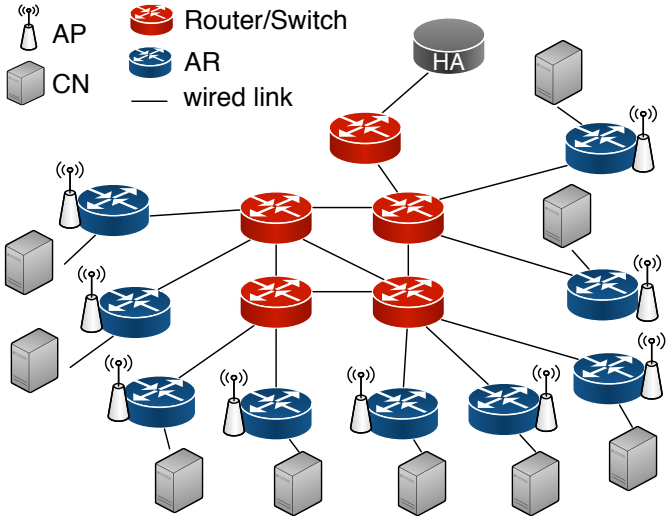


Figure C.10: Simulated Scenario.

Table C.2: Values used in the scenario for analytical model validation.

Notation	Value
N_m	average from this simulated scenario
S_{bu},S_{ba},S_{bu2},S_{ba2}	58, 16, 28 + 32N_m, 12 bytes
S_d,S_{dt}	1024, 1064 bytes
B_{wl}, B_{wr}	54, 100 Mbps
L_{wl}	0.2 ms (average from this simulated scenario)
L_{wr}	2 ms
N_{ar}	9
P_f	0.05 (average from this simulated scenario)
μ_s, μ_c	$1/T_s$, $1/T_c$ s^{-1}
P_m	0.444 or 1
$H_{c\rightarrow h}$	5 (average from simulated scenario)
$H_{c\rightarrow a}$	3.7 (average from simulated scenario)
$H_{h\rightarrow a}$	4 (average from simulated scenario)
$H_{a\rightarrow a}$	2.9 (average from simulated scenario)
$H_{m\rightarrow a}$	1 (average from simulated scenario)

higher in some cases, since in the simulation the MN can return to the previous AR, where some sessions were initiated, reducing the tunneled packets; in the analytical scenario this factor was not considered.

The signaling and data costs are validated for a fixed pause time (T_c) of 240 seconds, and the average sessions service time (T_s) is changed during the evaluation. The analytical model for signaling cost is defined per MN and per handover, which means that the analytical model results are multiplied by the number of MNs and the number of IP handovers (simulation time divided by pause time). The signaling costs of both analytical and simulation results are similar, following the same trend, where the MIPv6 signaling cost is equal for different values of T_s, while the DMIPA signaling cost increases with the increase of T_s. The set of N_m values used in the analytical model to achieve the results for signaling cost were measured from the simulations. The higher confidence interval for 44.4% of MARs is due to the random placement of these MARs in the ARs, as well as the exponential distribution of the sessions service and inter-arrival times, which introduces a higher variability of the signaling cost.

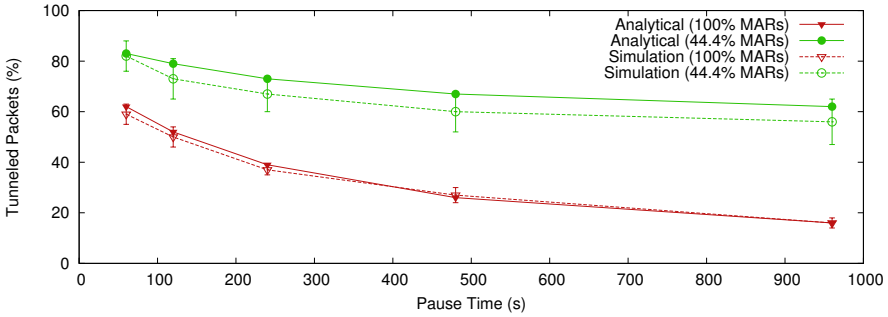


Figure C.11: Comparing Analytical and Simulated Results of Tunneled Packets.

This higher variability is not noticed in the data cost, since the confidence interval is small (Figure C.12). The data costs of the analytical models for both MIPv6 and DMIPA are calculated per data packet, thus the analytical results are multiplied by the total number of data packets measured from the simulation, which are the same for MIPv6 and DMIPA. The analytical values follow the same trend of the simulation ones, where signaling cost increases for higher values of $\boldsymbol{T_s}$. The analytical values are slightly higher than the simulation values, specially for DMIPA, since in the simulations the MN can again return to the previous AR, with ongoing MNs anchored session, reducing the number of tunneled/forwarding packets without the optimal path.

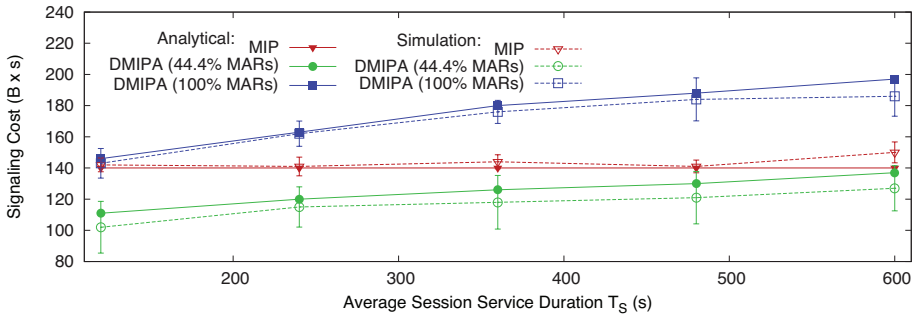


Figure C.12: Comparing Analytical and Simulated Results of Signaling Cost.

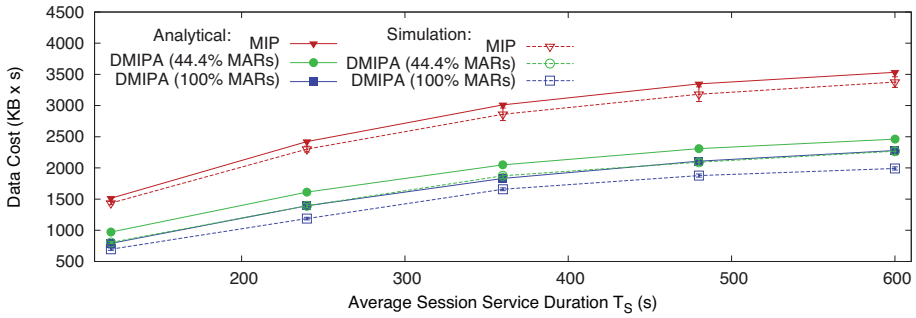


Figure C.13: Comparing Analytical and Simulated Results of Data Cost.

The metrics evaluated are very similar for both analytical and simulation results, which validates the analytical tunneled packets, signaling cost and data cost. The analyt-

ical model is accurate enough to estimate the behavior of DMIPA in different scenarios, according to different network performances and users' behaviors. The following sections evaluate DMIPA and MIPv6 in a larger set of metrics, using the simulation environment.

C.6.2.2 End-to-end Data Packet Delay

The end-to-end data packet delay measures the average time between the transmission of data packets by the CN and the reception of the respective packets by the MN.

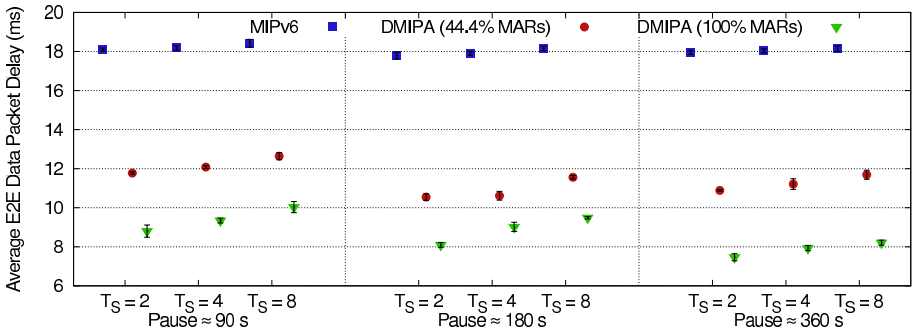


Figure C.14: Average E2E Data Packet Delay.

DMIPA reduces the average end-to-end delay when compared with MIPv6, Figure C.14. In MIPv6, the delay is similar for different session service and MN pause times, since data packets are always routed via the centralized HA, that tunnels these packets to the current location of the MN. In DMIPA, the delay is lower for shorter sessions combined with longer MN pause times, which reduces the number of handover sessions. The availability of more MARs decreases the data delay, since more sessions are initiated without any mobility support through the optimized path. In the DMIPA without fully MARs availability (e.g. 44.4% MARs), the data delay tends to a minimum value strictly associated with the percentage of MARs. Even for longer pause times, the probability of a MN to be attached to a MAR is just 44.4%, thus in the remaining percentage (55.6%) the MN establishes data sessions through a tunnel with the latest attached MAR, from the beginning of the sessions (non-optimized routing path).

C.6.2.3 Tunneled Packets/Hops

This section measures the number of packets being encapsulated by IP tunnels, and the total number of IP hops with tunneled packets.

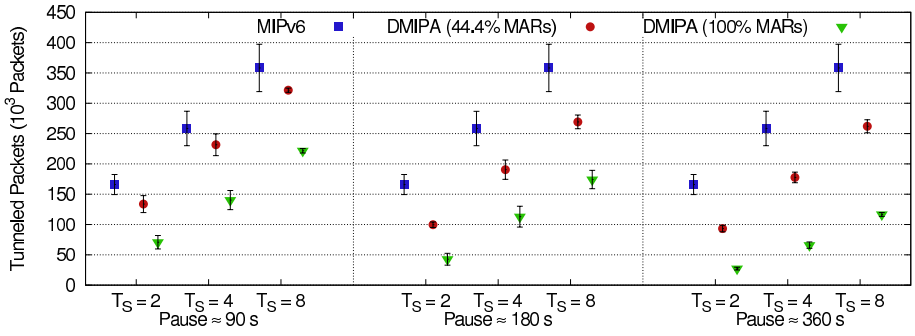


Figure C.15: Total number of tunneled data packets.

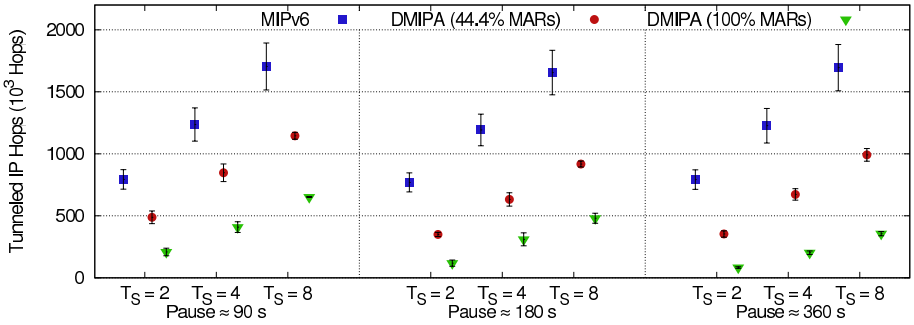


Figure C.16: Total number of tunneled IP hops of data packets.

DMIPA reduces both the number of tunneled packets and the number of tunneled IP hops, when compared with MIPv6, Figures C.15 and C.16 respectively. In MIPv6, all data packets are tunneled by the HA to the current location of the MN, from the time that the sessions are established. DMIPA is able to reduce the tunneled packets and tunneled IP hops for different session service and MN pause times. The tunneled packets and IP hops are lower for longer MNs pause times combined with shorter sessions service times, since it reduces the number/portion of the handover sessions without any tunneling support. Even for short MN pause times (e.g. 90 s) combined with long sessions service times (e.g. 8 min), the number of data packets is reduced when compared with MIPv6. It was performed another test for traffic sessions with a rate of 16Kbps and a $\boldsymbol{T_s}$ of 240s, changing the number of MNs. The results for DMIPA with 100% of MARs, regarding the percentage of tunneled data packet is presented in Figure C.17. The number of MNs does not have impact on the percentage of tunneled data packets, while the increasing of MNs pause time reduces the percentage of tunneled data packets. Even for a short MN pause time of 60s (pause time is per AR domain), which is four times higher than $\boldsymbol{T_s}$, the number of tunneled data packets is around 40% lower than in MIPv6.

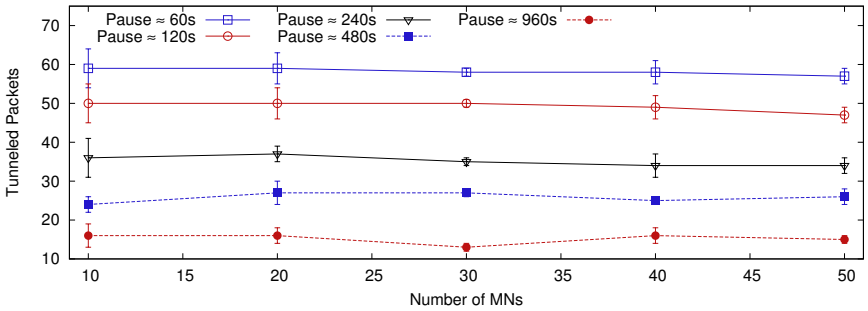


Figure C.17: DMIPA Tunneled Packets (100% MARs and $\boldsymbol{T_s}$ = 240s).

C.6.2.4 Active Bindings and MAs

This metric measures the average number of bindings (tunneling/translation rules) per second per MA (MAR/HA), as well as the average number of active MAs (MARs/HA) per second per MN.

DMIPA reduces the average instant number of bindings per MA when compared with MIPv6, Figure C.18. In MIPv6, the bindings of the five MNs are maintained in the HA, during the entire simulation, which just needs to maintain a single binding per MN. The

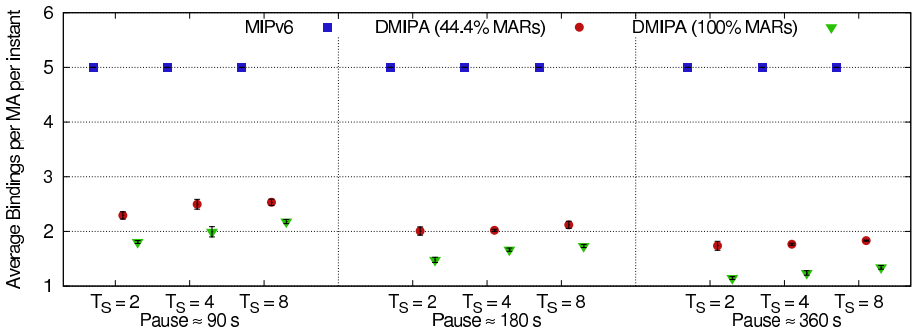


Figure C.18: Average Bindings per MA per instant.

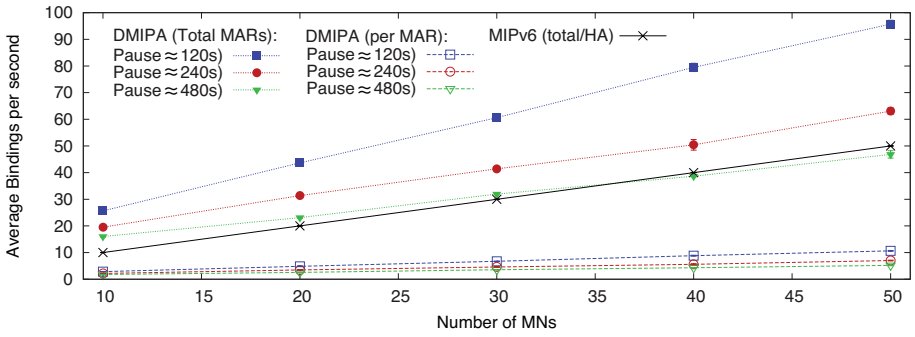


Figure C.19: Average Bindings per instant with number of MNs.

distribution of the data anchoring functionality through the MARs reduces the number of bindings in each one. However, the total number of bindings in the mobility management system might increase in DMIPA, since each MN might have more than one binding at a time. Thus, the bindings in the entire system are four or nine times the bindings per MAR, for 44.4% of MARs and 100% of MARs respectively. The number of bindings per MA is lower for shorter session service times combined with longer pause times, since the number of handover session is reduced and the MN has to maintain less bindings for shorter periods of time. It was performed another test to measure the average bindings per second for traffic sessions with a rate of 16Kbps and a $\boldsymbol{T_s}$ of 240s, changing the number of MNs (Figure C.19). Although there is a higher number of bindings per instant in the entire system for shorter pause times, the average bindings maintained per second by each MAR is much less than the one maintained by HA in MIPv6. The increase of the number of MNs has more impact in MIPv6 than in DMIPA, due to the distributed and dynamic data anchoring provided. Moreover, for long pause times (e.g. 480s) and a high number of MNs, the total bindings in the DMIPA mobility system is even lower than in MIPv6, since MNs are attached to the same MARs for long periods.

DMIPA increases the average number of MAs per MN per instant when compared with MIPv6 (Figure C.20), due to the dynamic anchoring feature introduced, where the MN might be associated to more than one MAR at a time. In MIPv6 each MN just needs to maintain the HA as the centralized mobility anchor. Scenarios with shorter pause times together with longer sessions service times increase the average number of MARs per MN, since the MN performs several handovers among ARs in a short time period. Thus, sessions remain anchored to the initial MAR, which need to be maintained in the set of MARs by the MN. Even in scenarios with short MN pause times (e.g. 90s) and long sessions service times (e.g. 8 min), the average number of MARs per MN per instant is

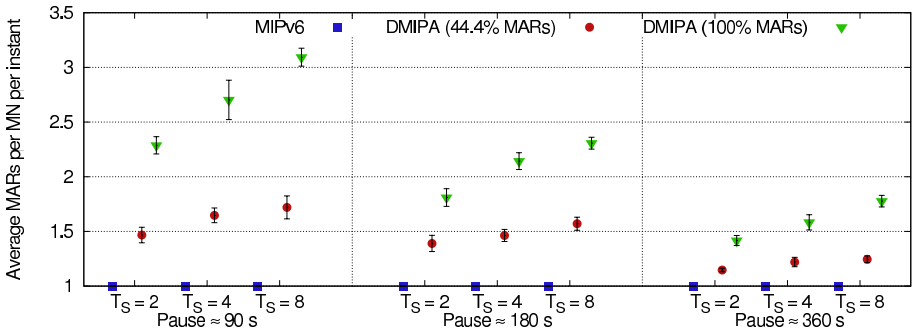


Figure C.20: Average MA per MN per Second.

approximately 3.

C.6.2.5 Data Cost

The data cost measures the cost to exchange the data in the network. It is defined as the multiplication of the data packets payload size by their remaining time in the network, between a CN and a MN.

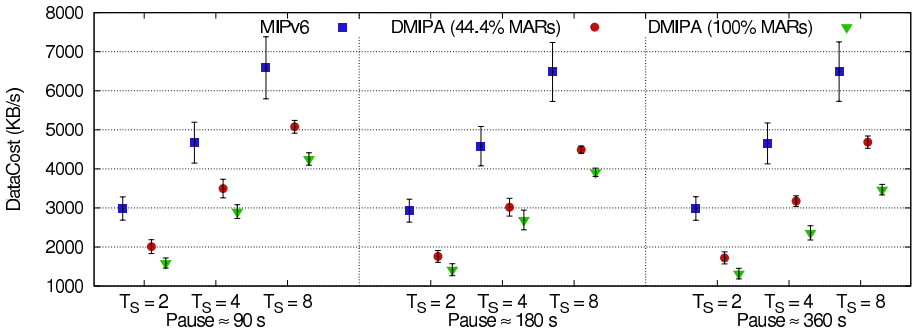


Figure C.21: Data Cost.

DIMPA reduces the data cost when compared with MIPv6 (Figure C.21), due to the distributed and dynamic data anchoring support provided. In DMIPA, the data cost decreases for longer MN pause times combined with shorter sessions service times, since it reduces the number of handover packets. The data cost has a minimum value strictly associated with the percentage of MARs. Even for longer MN pause times, the probability of a MN maintaining data sessions while attached to a MAR is just 44.4%, while the remaining data sessions are established through the non-optimized routing paths.

C.6.2.6 Signaling Cost

The signaling cost measures the cost to exchange the control messages of the IP mobility management. It is defined as the multiplication of the IP mobility message sizes by their remaining time in the network, from the source to the destination.

DMIPA reduces the signaling cost when compared with MIPv6 (Figure C.22) for longer pause times and shorter session service times, where the MN has to maintain less MARs per instant (less messages to be exchanged when the MN connects to a new AR (LAR/MAR)). The number of messages to be exchanged between the MN and MARs in DMIPA might be higher than in MIPv6, since the MN in DMIPA may have more than one associated MAR,

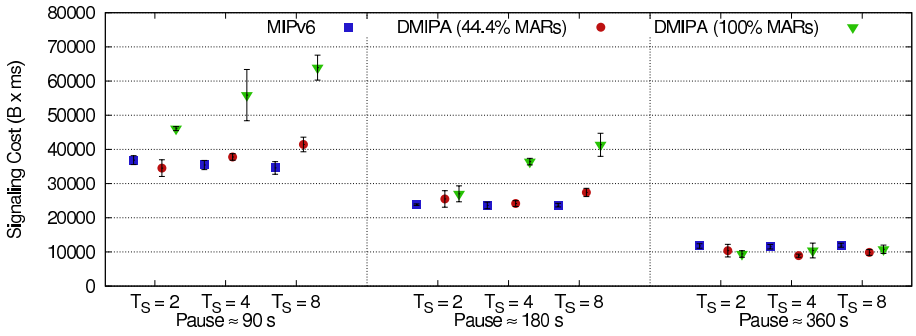


Figure C.22: Signaling Cost.

while in MIPv6 the MN has just associated with the centralized HA. However, the signaling cost of DMIPA may be lower than the signaling cost of MIPv6, since the MN exchanges control messages with closer elements (MARs), reducing the time of the messages in the network. Considering DMIPA with 100% of MARs, shorter pause times with longer sessions service times increases the signaling cost when compared with MIPv6. However, the slightly increase of signaling cost in these scenarios enables a great improvement in the data delivery performance.

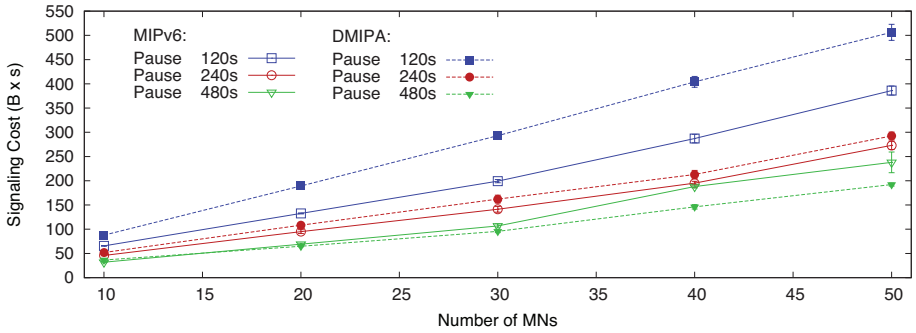


Figure C.23: Signaling Cost with Number of MNs.

It was performed another test to measure the signaling cost for traffic sessions with a rate of 16Kbps and a T_s of 240s, changing the number of MNs (Figure C.23). For pause times lower or equal to sessions service times, DMIPA has a higher signaling cost than MIPv6, otherwise DMIPA outperforms MIPv6 regarding the signaling cost. The increase of the number of MNs highlights the difference of signaling cost between DMIPA and MIPv6.

C.6.3 Testbed

This section aims to validate and evaluate a testbed with DMIPA working in a real environment. The tested scenario is presented in Figure C.24. The small scale of the testbed does not allow us to provide the same kind of evaluation as the one perform in ns-3 simulator. Thus, the evaluation performed through the testbed is focused on metrics related with handover and data delivery for a small set of sessions.

The Network of Instituto de Telecomunicações (IT) in Aveiro already provides IPv6 through the Neighbor Discovery Protocol messages. We connected three Single Board Computers (SBCs) with OpenWrt [20] through ethernet cable to the IT Aveiro network. The SBC used is a GW2358-4, which is a member of the Gateworks Cambria Network

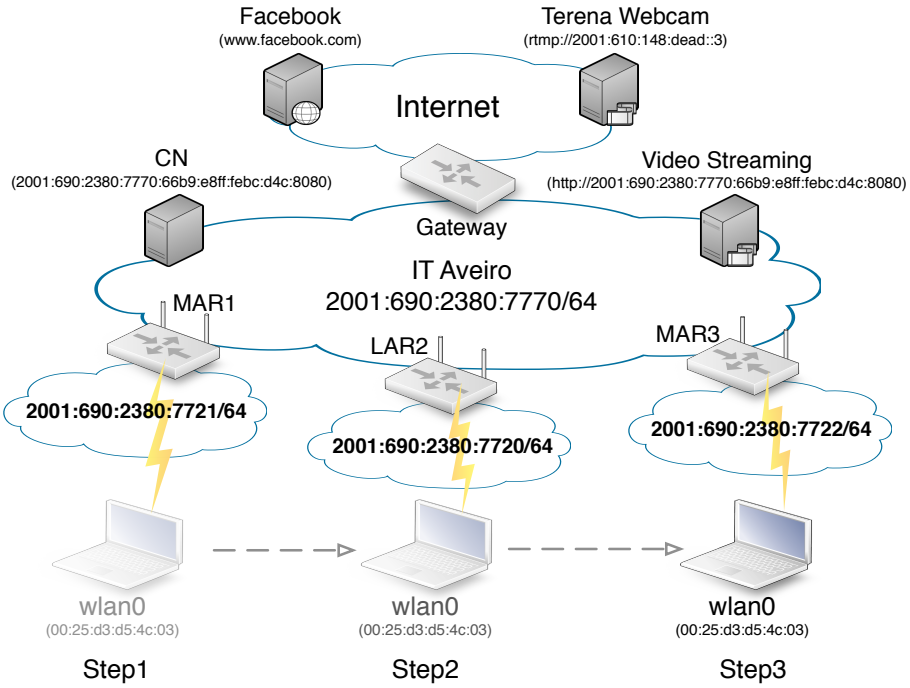


Figure C.24: Testbed.

Processor family. This network processor consists of an Intel® IXP435 XScale® operating at 667MHz, 128Mbytes of DDRII-400 SDRAM, and 32Mbytes of Flash. Peripherals include four Type III Mini-PCI sockets, two 10/100 Base-TX Ethernet ports. In one of the Mini-PCI sockets it is installed an Atheros WiFi 802.11 b/g card. Each SBC is configured to provide a different IPv6 prefix through the Wifi interface, where the emphradvd daemon [21] announces the respective IPv6 prefix of the network. One SBC is configured as a LAR and the others as MARs. In MARs, the emphradvd is configured with the HA flag equal to 1, and to announce the IPv6 address of the interface, instead of the IPv6 network prefix.

The MN is a laptop with a built-in WiFi 802.11b/g interface (wlan0). The laptop is an EEEPC Intel® Atom ™CPU N450 operating at 1.66GHz x 2, with 1 Gbyte of DDRII-400 SDRAM, where Ubuntu 12.10 32-bits operative system is installed.

We experimented our testbed with IPv6 services provided from the Internet and from the IT Aveiro network. We used the Ping6 application with Facebook, and the Video LAN Client (VLC) to receive video streams: one from the Terena Webcam available in the Internet; and another provided by a laptop connected to the IT Aveiro network through the HTTP protocol. Although all services are available for the MN, the evaluation was performed with Distributed Internet Traffic Generator (D-ITG) [22] installed at CN and MN endpoints, in order to establish UDP and TCP flows between CN and MN, initiated at different time instants.

In the scenario of Figure C.24, the MN moves from MAR1 to MAR3 through LAR2. The time that the MN is attached to an AR (MAR or LAR) represents a step in the scenario, where a new UDP/TCP session is initiated, maintaining the previous ones active. In Step 3, the MN has three active sessions initiated while connected to three different ARs. It is defined a residence time around 1 min in each step to repeat the scenario 10 times and obtain results with 95% of confidence interval.

DMIPA main focus is on distributed and dynamic mobility to optimize the session de-

livery for both user and network, thus the seamless handover subject was not improved. In the testbed, it was used the connection manager of ubuntu to perform the handovers, without any seamless technique. DMIPA can be integrated with seamless handover techniques, such as the ones provided by [23] [24] [25] already working with MIPv6 and PMIPv6. In these techniques, part of the configurations for the new network are performed during the preparation phase, reducing the time of the execution phase. The testbed of DMIPA just initiates the procedures for session continuity after the conclusion of the Duplicated Address Detection (DAD) mechanisms for the new obtained IPv6 global address.

In the testbed evaluation, we provide results along the tested time for data packet delay and bitrate of UDP and TCP sessions. Finally, we provide average UDP/TCP results of several metrics from 10 repetitions of the scenario. Note that DMIPA testbed is connected through the IT Aveiro network, which is used for all collaborators to access the Internet, thus we do not have the control of the traffic being exchanged. In the first tests, UDP flows are configured with a payload size of 1024 bytes, and a rate of 250 and 1000 packets per second. It is initiated a new UDP flow in each network from CN to MN, which is maintained until the end of the test. Figures C.25 and C.26 show the data packets delay along the time for rates of 250 and 1000 packets per second, respectively.

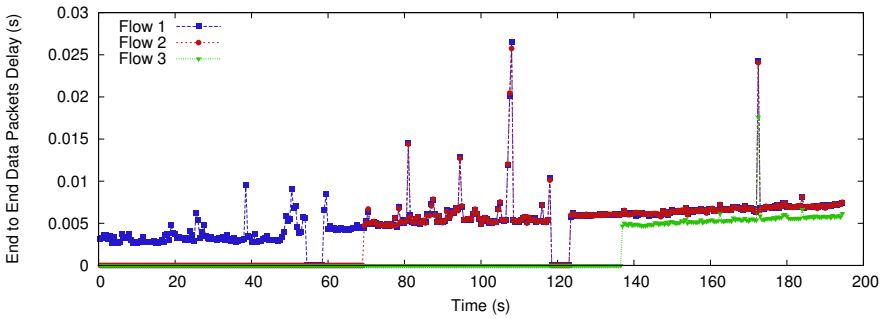


Figure C.25: UDP Packets Delays: 250 pkts/s, packet size of 1KB.

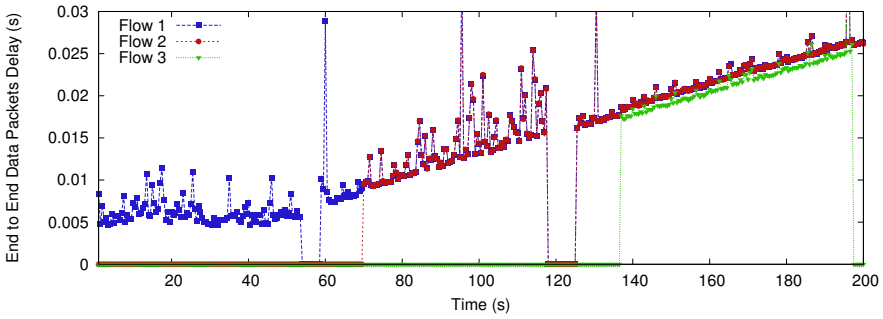


Figure C.26: UDP Packets Delays: 1000 pkts/s, packet size of 1KB.

As long as new flows are initiated, the packet delay increases, especially due to the wireless medium access, since we have more traffic being sent from CN to MN. The handovers performed at 60 and 120 seconds have a latency around 5 seconds, since there is no mechanism to provide seamless handover. The flow 3 starts around 140 seconds through the optimized path, while other ongoing flows (1 and 2) are maintained through a tunnel between MAR1 and MAR2. The packet delay introduced by the tunnel (encapsulation, forwarding and decapsulation) can be estimated if we compare flow 3 with other 2 flows,

which is around 1.2 ms. Even with the increase of the UDP flows rate to 1000 pkts/s (Figure C.26), the delay introduced by the tunnel is similar, providing a similar experience of data packet delay to the MN, for handover and non handover traffic.

The same test is repeated for TCP sessions, but using a lower rate of 100 packets per second, Figure C.27. The packet delay slightly increases in TCP test, since the rate of the flows are lower. However, the difference between flow 3 and the other two flows is quite similar.

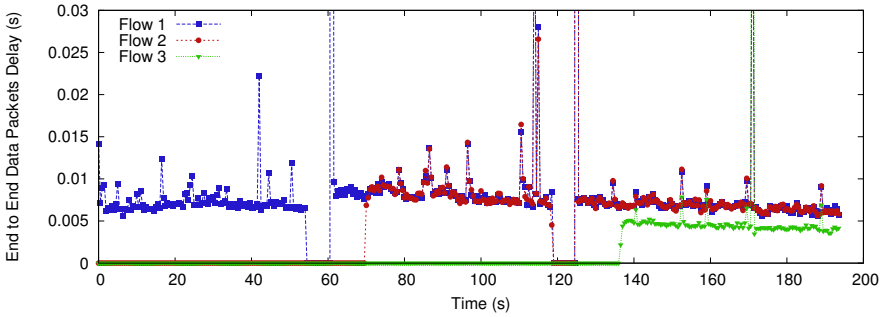


Figure C.27: TCP Packets Delays: 125 pkts/s, packet size of 1KB.

For the same scenarios, it is also measured the bitrate along the time, for UDP flows with a rate of 250 and 1000 pkts/s. For a rate of 250 pkts/s (Figure C.28), the bitrate along the time is quite similar, even when the three sessions are active at the end of the test. The bitrate is not influenced by the tunnel between MAR1 and MAR2 for handover traffic, since the flow 3 and the other two flows have a similar bitrate. For a rate of 1000 pkts/s (Figure C.29), the bitrate decreases with the establishment of new flows, especially in the last step, where 3 flows are active at the same time. Besides the decrease of the bitrate, due to the increase of data traffic in the network, the bitrate of handover and non handover traffic is similar, since non handover flow 3 and handover flows 1 and 2 are affected in the same way.

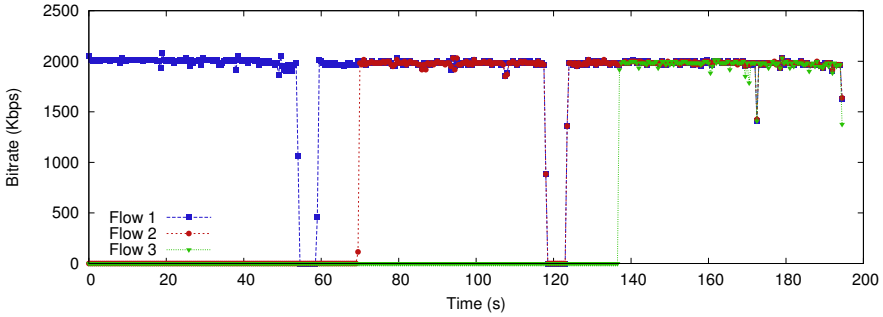


Figure C.28: Received UDP bitrate: 250 pkts/s, packet size of 1KB.

The same test with TCP sessions is also evaluated, using a rate of 125 packets per second, Figure C.27. The bitrate remains constant with the establishment of new TCP flows, where handover and non handover traffic presents the same bitrate values.

From the average results of UDP flows (Table C.3) and TCP flows (Table C.4), we observe that the average packet delay is similar for the 3 flows. However, the flows maintained during more time through the non-optimized routing path, while the MN is not connected to the MAR where they started, have a slight increased in the average packet

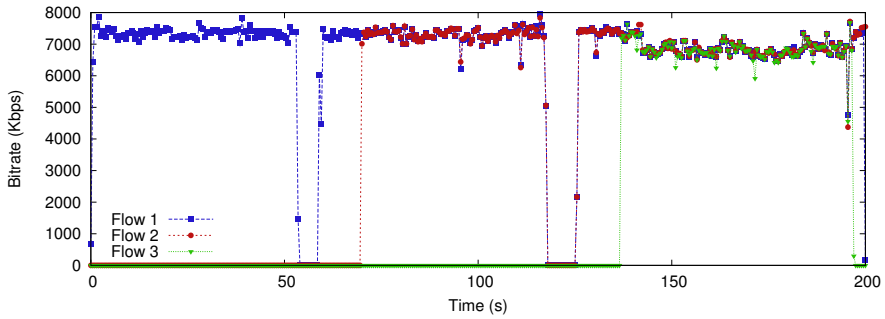


Figure C.29: Received UDP bitrate: 1000 pkts/s, packet size of 1KB.

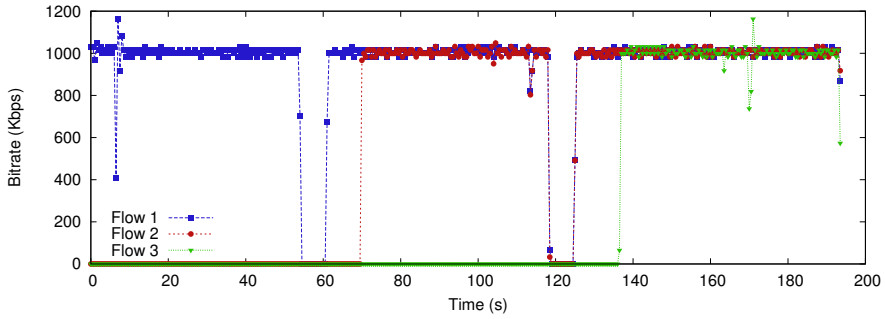


Figure C.30: Received TCP bitrate: 125 pkts/s, packet size of 1KB.

delay (between 0.5 and 1.5 ms). The sequence of flows, from the higher to the lower packet delay is: flow 2, flow 1 and flow 3. This order matches the order of the percentage of tunneled packets, since more tunneled packets means more packets not following the optimized path (added delay of the tunneling). Flow 2 is connected through LAR2 and maintained by MAR1 from the beginning through a tunnel (100% tunneled packet), while the MN is outside of MAR1 network during the entire session (always non-optimized routing path). Flow 1 is initially established through MAR1 when MN is connected to MAR1, thus, it has around 77% of tunneled packets in UDP and 79.7% of tunneled packets in TCP, which are tunneled after the first handover around the 60s. Flow 3 is established through MAR3, which is always maintained through the optimized routing path, since the MN remains connected in MAR3 during the entire flow duration (0% tunneled packets). The bitrate is lower for the first flows initiated, since they suffer more handovers where the bitrate is zero. The average handover latency for UDP flows 1 and 2 are similar, with 4.8 and 5.3 seconds, respectively, and TCP flows 1 and 2 of 6.4 and 6.5 seconds. As previously explained, DMIPA was implemented in a real environment without any seamless handover technique, based on the current connection manager of Ubuntu.

The DMIPA testbed demonstrates that it is able to provide an IP session continuity in heterogeneous environments, distributing the mobility anchors (MARs) and using a dynamic anchoring selection. This is achieved without introducing a significant delay to data packets of UDP/TCP, and without impairing the bitrate received by the user device.

C.7 Conclusion

This article presented the trend of distributed mobility management, describing and studying the behavior of a distributed and dynamic IP mobility scheme, called Dynamic

Table C.3: DMIPA testbed result for UDP with 125 pkts/s and a packet size of 1KB.

	Flow 1	Flow 2	Flow 3
Duration (s)	193.5 ± 18.7	123.4 ± 11.9	56.5 ± 5.5
Delay (ms)	5.9 ± 0.7	6.3 ± 0.8	5.5 ± 0.3
Jitter (ms)	1 ± 0.4	1 ± 0.4	1 ± 0.4
Bitrate (Kbps)	948 ± 92	957 ± 92	1004 ± 97
Packet Loss (%)	5.9 ± 0.6	5 ± 0.5	0.2 ± 0.17
Handover Latency (s)	4.8 ± 0.5	5.3 ± 0.5	-
Tunneled Packets (%)	77 ± 7.0	100 ± 0.0	0 ± 0.0

Table C.4: DMIPA testbed result for TCP with 125 pkts/s and a packet size of 1KB.

	Flow 1	Flow 2	Flow 3
Duration (s)	193.5 ± 18.7	123.7 ± 12.0	56.7 ± 5.5
Delay (ms)	5.6 ± 0.7	6 ± 0.6	4.9 ± 0.5
Jitter (ms)	1 ± 0.4	1 ± 0.4	1 ± 0.4
Bitrate (Kbps)	926 ± 90	943 ± 91	1002 ± 97
Handover Latency (s)	5.4 ± 0.8	5.5 ± 0.9	-
Tunneled Packets (%)	79.7 ± 8.7	100 ± 0.0	0 ± 0.0

Mobile IP Anchoring (DMIPA). DMIPA provides IP mobility support in heterogeneous environments under flat network architectures, and considering communications with any device in the network. The IP mobility management scheme distributes the IP data anchoring functionality through the Mobility-enabled Access Routers (MARs), and the IP mobility context management through the users' devices (Mobile Nodes). It also integrates the concept of dynamic mobility, which means that mobility is only provided when sessions really need it.

The outcome of the evaluation from the analytical model and simulation shows that DMIPA overcomes the centralized MIPv6 model, reducing the average end-to-end data packet delay, the total data cost, the total tunneled packets and hops, the average bindings per mobility anchor, and the IP mobility handover latency (update of mobility anchors). The best scheme regarding the signaling cost depends on the AR pause times and average sessions service times. The testbed validated DMIPA in a real environment, proving the concept of distributed and dynamic IP mobility.

As future work, DMIPA concept will be validated in multihomed scenarios with ARs from different access technologies. Moreover, the selection and management of the MARs in multihomed environments is a direction to be exploited in the next steps of this research work.

C.8 Acknowledgments

This work has been supported by the UMM FCT project (PTDC/EEA-TEL/105709/2008). Tiago Condeixa is also supported by the FCT scholarships SFRH/BD/65265/2009.

C.9 References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," Cisco, White Paper, Feb. 2013.

[2] Ericsson, "Ericsson mobility report: On the pulse of the networked society," Ericsson, Report, Jun. 2013.

[3] H. Chan et al., "Distributed and dynamic mobility management in mobile internet: Current approaches and issues," *Journal of Communications*, vol. 6, no. 1, pp. 4–15, Feb. 2011.

[4] H. C. (Ed.), "Problem statement for distributed and dynamic mobility management," IETF, Internet-Draft draft-chan-distributed-mobility-ps-05, Oct. 2011, work in progress.

[5] T. Condeixa and S. Sargento, "Dynamic mobile ip anchoring," in *IEEE ICC*, Jun. 2013.

[6] C. Perkins et al., "Mobility support in ipv6," IETF, RFC 6275, Jul. 2011.

[7] S. Gundavelli et al., "Proxy mobile ipv6," IETF, RFC 5213, Aug. 2008.

[8] I. WG. (2013, Aug) Distributed mobility management. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[9] P. Bertin et al., "Distributed or centralized mobility?" in *IEEE Global Telecommunications Conference*, Dec. 2009, pp. 1–6.

[10] ——, "An evaluation of dynamic mobility anchoring," in *70th IEEE Vehicular Technology Conference Fall*, Sep. 2009, pp. 1–5.

[11] H. Ali-Ahmad et al., "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *IEEE WPMC*, Sep. 2012.

[12] H. Chan, "Proxy mobile ip with distributed mobility anchors," in *IEEE GLOBECOM Workshops*, Dec. 2010, pp. 16–20.

[13] J. Kim et al., "Use of proxy mobile ipv6 for distributed mobility control," IETF, Internet-Draft draft-sjkoh-mext-pmip-dmc-03, Jun. 2011, work in progress.

[14] C. Bernardos and J. Zuniga, "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-00, Sep. 2012, work in progress.

[15] T. Condeixa et al., "Decoupling and distribution of mobility management," in *IEEE Globecom Workshop MobiWorld*, Dec. 2012.

[16] S. A. Baset and H. Schulzrinne, "An analysis of the Skype peer-to-peer Internet telephony protocol," in *Proc. of the INFOCOM '06*, April 2006.

[17] S. Cirani and L. Veltri, "Implementation of a framework for a dht-based distributed location service," in *SoftCOM*, Sep 2008, pp. 279–283.

[18] J. Lee et al., "Enhancing qos of mobile devices by a new handover process in pmipv6 networks," *WPC Journal*, vol. 61, pp. 591–602, 2011.

[19] ns 3. (2013, Aug) ns-3.14. [Online]. Available: http://www.nsnam.org/ns-3-14/documentation

[20] OpenWrt. (2013, Aug). [Online]. Available: http://www.openwrt.org/

[21] radvdv. (2013, Aug) Router advertisement configuration. [Online]. Available: http://wiki.openwrt.org/doc/uci/radvd

[22] D-ITG. (2013, Jun) Distributed internet traffic generator. [Online]. Available: http://traffic.comics.unina.it/software/ITG/index.php

[23] R. Koodli, "Fast handovers for mobile ipv6," RFC 5568, Jul. 2009.

[24] H. Yokota et al., "Fast handovers for proxy mobile ipv6," IETF, RFC 5949, Sept. 2010.

[25] J. Kim and S. Koh, "PMIPv6 with Bicasting for Soft Handover," Internet-Draft draft-jikim-bpmipv6-00, Sep. 2009.

Paper D

Dynamic Offload Anchoring with IP Mobility

Tiago Condeixa, Jonathan Carvalho, Susana Sargento and Rute Sofia
in *IEEE Transactions on Network and Service Management* (submitted)

Dynamic Offload Anchoring with IP Mobility

Tiago Condeixa, Jonathan Carvalho, Susana Sargento and Rute Sofia

Abstract

The massive adoption of mobile applications for portable devices has been overloading cellular networks with mobile data. Thus, mobile networks are adopting data offloading mechanisms, often via wireless local networks. The current offloading models are centralized in nature, forcing data to be routed via the mobility anchor point, thus introducing scalability and performance issues. This article introduces dynamic anchoring mechanisms to support data offloading. In contrast to 3GPP models, we propose a distribution of offloading nodes closer to the end-user. We also assume that these nodes are fully or partially controlled by the end-user. We compare the distributed anchoring approach with centralized anchoring model through simulations, and validate it in a real scenario. The results demonstrate that distributed offload anchoring decreases the data cost and the data packet delay.

Keywords: data offloading, dynamic anchoring, multihoming, IP mobility.

D.1 Introduction

There has been a rapid increase of mobile devices and mobile data [1, 2], and current cellular networks are not prepared to deal with such surge. Thus, operators are seeking solutions to accommodate this growth of data consumption. There is an introduction of flatten models in cellular networks (e.g. LTE) to deal with this evolution of the mobile data behavior. Moreover, data offloading is considered as one of the key strategies to accommodate such amount of mobile data, where the wireless area networks are the preferred access networks to offload this growing data. Session continuity support needs to be ensured to the user's ongoing sessions, even for the data offloaded through these wireless access networks [3]. There are several daily routine scenarios where the user's device is connected to the network simultaneously through different interfaces, or when it disconnects one of the interfaces with ongoing sessions. In theses cases, the IP mobility should be able to provide session continuity through the offloading of sessions to other networks and/or interfaces.

The new cellular technologies (e.g. LTE) were envisioned to increase the data rates, and to support multihomed devices with both 4G and WiFi connections simultaneously. These mobile devices are able to connect through multiple interfaces to heterogenous access networks. The new 4G cellular networks are already IP-based, thus it is required an IP mobility protocol to provide mobility to the mobile devices, while they move between different IP networks. However, current IP mobility protocols need to be adapted for such evolution. Current IP mobility models, such as Dual Stack Mobile IPv6 (DSMIPv6) [4], Mobile IPv6 (MIPv6) [5] and Proxy Mobile IP (PMIP) [6], have been developed for centralized network architectures; thus, they are based on a single entity, called mobility anchor (e.g. Home Agent (HA) in MIPv6/DSMIPv6 and Local Mobility Anchor (LMA) in PMIP), that handles the mobility management of a large amount of Mobile Nodes (MNs). These mobility models bring several limitations when applied to the recent network trends [7], such as a single point of failure, longer traffic delays and higher signaling loads. The

data offload anchor matches the mobility anchor, since the sessions are offloaded through the IP layer (e.g. IP addresses), and they need IP session continuity support. In the scope of this paper, a data offload anchor is the same as a mobility anchor, since it is the node responsible to route packets of the user's ongoing sessions to one of its device's interfaces. The assignment of the proper mobility anchor to a session is fundamental to provide the optimized routing path to that session, before and after the offloading, which impacts the network resources consumption and the user experience.

We propose to extend the Dynamic Mobile IP Anchoring (DMIPA) [8], which already distributes the mobility anchor functionalities through the access network level, to provide dynamic offloading in multihomed environments. In DMIPA, the MN selects its current access router/gateway for a new session, which is maintained for the whole session duration, and when it roams to another IP network, the new router/gateway establishes a tunnel with the previous one to maintain the ongoing sessions of the MN. This concept, called dynamic anchoring, aims at optimizing the routing path assuming that most of the sessions are short enough, such that a session is terminated before experiencing several IP handovers. The extended approach of DMIPA describes the set of mechanisms to provide dynamic offload anchoring for multihomed devices, which properly manage interfaces, IPv6 addresses, offload anchors, IP tunnels and routing rules. The proposed offloading mechanisms are evaluated through simulations and compared with centralized mobility anchoring, regarding data cost and data packet delay, and are also validated in a real testbed.

The reminder of this article is organized as follows. An overview of current data offloading with IP mobility is presented in Section II. Section III provides an overview of DMIPA and introduces the mechanisms to provide dynamic offload anchoring. Section IV evaluates the dynamic offload anchoring in comparison with centralized anchoring. Section V presents a proof of concept of the proposed approach in a real environment. Finally, Section V concludes the article.

D.2 Overview on Data Offloading with IP Mobility

Data offloading from cellular to wireless networks is being considered as a cost effective solution to cope with the explosion of mobile internet data [9, 3, 10]. In a study performed in [11] using real mobility traces from the city of San Francisco, it is shown that only few hundreds of WiFi Access Points (APs) deployed there can offload half of the mobile data from the 3G network. Hence, data offloading is one of the major work items in 3GPP Release 10 (LTE-Advanced) [12, 13] to support the mobile Internet traffic growth by lowering the network congestion, achieving the network load balancing, and optimizing the network resources. The user experience can be maximized while network operators enable a selective offload of certain types of data sessions, if IP session continuity is assured to the offloaded sessions. IP session continuity is related with the ability to maintain the ongoing IP sessions that are offloaded by keeping the same end-point IP addresses, despite being maintained through a different IP network than the initial one.

There are IP flow mobility approaches being standardized in the IETF and 3GPP to provide data offloading, which are both host-based and network-based [14]. Current flow mobility approaches introduce extensions to MIPv6 or PMIP mobility management solutions to provide data offloading. The extensions to MIPv6 need to provide multiple CoA registration support [15], flow bindings support [16], and traffic selectors definition [17]. The extensions to PMIP [18] span one mobility session across multiple MN interfaces, configure the same HNPs on multiple interfaces of MN, and transfer the policies between

the MN and the network to install the required filters in the LMA/MAG for flow routing. Besides the achievements, additional modifications on the client side may be required to fully exploit the mobility and traffic management enhancements performed on the network counterpart, and to benefit from simultaneous connectivity from heterogeneous accesses [14].

Currently, 3GGP standards are considering three data offloading mechanisms: Local IP Access (LIPA), Selected IP Traffic Offload (SIPTO), and IP Flow Mobility (IFOM).

The main purpose behind LIPA [12] is to enable an MN to directly transfer data to a local network, without the data traversing the macro cellular network. LIPA can be used by an MN to offload data through local networks, such as when a femtocell is deployed as a local network, directly accessing any Internet services (e.g. video streaming) and bypassing the macro cellular network. The gateway of the local network is the offloading point of LIPA, which is a network-based offloading mechanism.

The main purpose behind SIPTO [12] is to extend LIPA to support data offloading also through cellular networks. SIPTO helps reducing the network load at the macro cellular network, by breaking out selected data closer to the edge of the network where the MN is currently attached. Besides the gateway of the local network, SIPTO also introduces the eNodeB (eNB) gateway of the cellular network as a data offloading point. SIPTO is also a network-based offloading mechanism.

IFOM [13] has a different purpose, since it allows an MN to switch specific flows between different access networks. When an MN is connected to cellular and WiFi networks simultaneously, it can offload selected flows from the cellular to the WiFi network by sending a signaling message to the HA, usually placed in the Packet Data Network Gateway (PGW) of the 3GPP Evolved Packet Core (EPC). INFO is an MN-based offloading mechanism that has to be integrated with an IP mobility management protocol. Initially, it was defined for DSMIPv6/MIPv6, but there are other proposals to be integrated with PMIP [19]. Thus, the HA of MIPv6/DSMIPv6 or the LMA of PMIP are used as the offloading anchor of IFOM.

Current centralized approaches, such as DSMIPv6 [4], MIPv6 [5] and PMIP[6], are IP mobility protocols used in 3GPP to manage mobility between 3GPP and non-3GPP networks [20]. In 3GPP, any connection over a 3GPP access network is considered as the home link, and a non-3GPP network as the foreign link. The protocols are based on a centralized mobility entity, the HA in MIP-based approach and LMA in PMIP-based approaches, which is usually placed in the PGW, and another mobility entity in the MN/AR. The PMIPv6-based approaches introduce an entity, called Mobile Access Gateway (MAG), to hide the mobility from the MN. The mobility management functionalities are moved from the MN to the MAGs, which are implemented closer to the user. The LMA forwards the data packets to the current MAG of the MN, which then delivers the packets to the MN. A MAG is also in charge of detecting an IP handover of the MN to update the LMA with the new MAG of the MN.

The IETF charted a working group, called Distributed Mobility Management (DMM) [21], which focuses on developing a new framework for distributed mobility management in flatten network architectures. There were already defined the limitations and problems of centralized mobility management when compared with distributed mobility management [7], as well as the requirements for a proper distributed mobility management solutions [22]. Most of the novel DMM schemes [23, 8, 24] introduce the idea of distributing the MA through the Access Routers (ARs) [25]. The MN selects the MA located at its current AR for a new session, which is maintained for the whole session duration, even if the MN moves among ARs. The routing path may be optimized, assuming that most of the sessions are

short enough to be terminated before experiencing several IP handovers. There is also a recent DMM scheme [26] that proposes to anchor session in mobility anchors close to the CNs. The mobility anchor in the CN network forwards the ongoing sessions to the current location of the MN through tunnels from the beginning of the session, thus the optimized routing path is always assured at the cost of longer tunnels. These novel approaches to distribute the mobility management were evaluated, and they improve the network and user performance when compared with centralized approaches [27, 28, 29].

Summarizing, the data offload mechanisms need to be intrinsically associated with IP mobility protocols, in order to provide IP session continuity to the offloaded sessions. Current data offloading mechanisms (e.g. IFOM and Flow Mobility) use the HA/LMA as the offload anchor, which is also the centralized mobility anchor of the IP mobility protocols (e.g. MIPv6, DSMIPv6 and PMIP). In these IP mobility protocols the centralized offload anchor (e.g. HA or LMA) routes the mobile data and manages the mobility context of all MNs of the network. As the mobile data traffic increases, such centralized anchoring model may encounter scalability and performance issues. The DMM working group has been introducing novel approaches to distribute the mobility management that are able to improve user and network performance when compared with centralized approaches. However, there are no specifications in DMM to provide data offloading with session continuity support for multihomed devices. Thus, we propose to extend the host-based distributed mobility approach Dynamic Mobile IP Anchoring (DMIPA) [8], in order to provide an improved data offloading with session continuity support. This extended approach will optimize the network resources (reducing the data cost) and improve the user experience (reducing the data packet delay).

D.3 Dynamic Offload Anchoring

We propose to extend the DMIPA mechanisms [8] to provide dynamic offload anchoring for multihomed devices. Oppositely to MIPv6, which defines a centralized mobility anchor (HA), DMIPA introduces several mobility anchors distributed at the access network level, denoted here as Access Home Agents (AHA). DMIPA already ensures that each MN manages its offload anchors (AHAs), IP tunnels and IP addresses, which completely eliminates the necessity of a centralized node for offload management purposes.

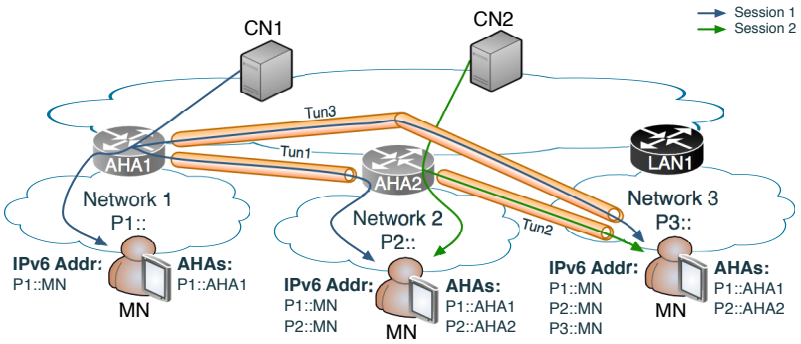


Figure D.1: Overview on the Dynamic Mobile IP Anchoring (DMIPA).

An overview of DMIPA for an MN with a single interface is explained with the example of Figure D.1. The AHA helps in the assignment of IPv6 prefix/addresses to the MNs (e.g. AHA1 assigns IPv6 addresses from prefix P1::). The MN establishes new sessions through the AHA where the MN is currently attached to, which will be the AHA for that

session during the sessions lifetime (e.g. session1 is anchored to AHA1). When the MN attaches to another AHA, the new AHA establishes bidirectional tunnels with previous AHAs, which have anchored sessions of the MN (e.g. tunnel1 between AHA1 and AHA2). This information about AHAs with anchored sessions is maintained by the MN, and thus, the MN is responsible to manage its set of AHAs and its respective IPv6 addresses, which are delivered to the MN's new AHA through a BU-based message. This BU-based message contains the IPv6 addresses of the AHAs and the IPv6 addresses that the MN receives from these AHAs. For instance, when the MN attaches to AHA2, it sends a BU-based message containing the pair: P1::AHA1 - P1::MN. The established bidirectional tunnels are used for packets forwarding, thus, packets associated to the MN's IPv6 address allocated by the initial AHA are tunneled to the current MN AHA. When there is no AHA in the current network of the MN, being it attached to a Legacy Access Node (LegAN), the MN establishes bidirectional tunnels with previous AHAs with anchored sessions (e.g. tunnel3 between AHA1 and MN, and tunnel2 between AHA2 and MN).

The mechanisms presented in this article are focused on the resources management optimization and data delivery performance, and not on the improvement of handover execution time nor the handover latency. There are two main mechanisms to offload data traffic, which will be described in more detail in the following sub-sections.

D.3.1 Offloading sessions between interfaces connected to the same AHA

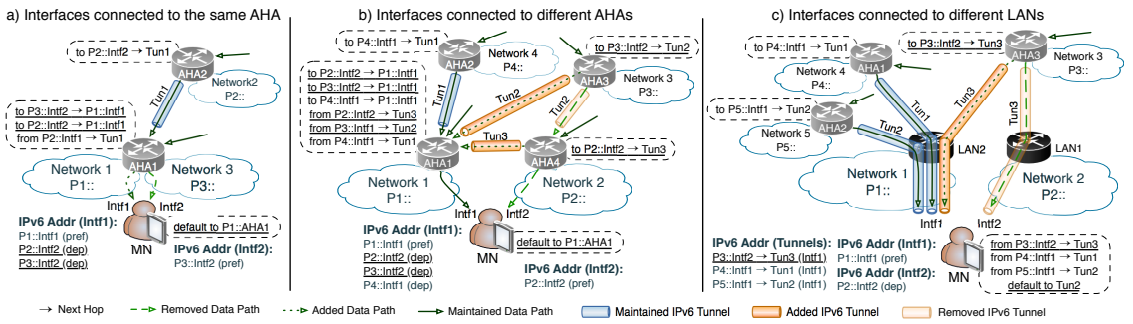


Figure D.2: Distributed mobility mechanisms for offloading sessions.

The Case A (Figure D.2 a)) provides the mechanisms to offload sessions between interfaces connected to the same AHA, in the same or different IPv6 prefixes.

Some network topologies might provide one or more IPv6 prefixes using the same or different interfaces from the same AHA, or several access technologies connected through a single AHA sharing one or more IPv6 prefixes. There are some use-case scenarios from the users' daily routine where it might be advantageous to enable this mechanism:

- an MN connected through multiple interfaces to distinct medium access technologies (e.g. 802.3, 802.11b/g and 802.11p) sharing the same IPv6 prefix (same AHA) suffers a disruption in one of the interfaces or one of the medium access is overloaded.
- the network operator might use distinct IPv6 prefixes from the same or different access technologies (e.g. LTE and UMTS or 802.11g and 802.11p) connected to the same AHA, and the user might suffer a disruption in one of the interfaces, or the network operator desires to maintain load balancing in the medium accesses.

In the case A, the MN decides to offload sessions from Interface 2 (Intf2) to Interface 1 (Intf1), as illustrated in Figure D.2 a). The underlined instructions are the ones

changed/added to perform the required offloading between Intf1 and Intf2, while the other ones were previously configured by the offload management. It is assumed that the MN has a set of two AHAs (AHA1 and AHA2) with anchored sessions, and it is directly connected to AHA1 through two interfaces (Intf1 and Intf2) from different IPv6 prefixes. There is a tunnel between AHA2 and AHA1 to route the packets from/to P2::Intf2 of the MN, as well as a rule in AHA1 to locally forward packets destined to P2::Intf2 to the next hop P3::Intf2.

Considering the offloading between interfaces, the AHA1 is responsible to locally forward packets from the IPv6 addresses of Intf2 to the preferred IPv6 address of the Intf1. In the example, the sessions anchored in AHA1 and AHA2 with IPv6 addresses P3::Intf2 and P2::Intf2, respectively, are forwarded to P1::Intf1 in AHA1. Moreover, the MN has also to move/replicate these IPv6 addresses from Intf2 to Intf1, in order to ensure that the Intf1 receives/sends the packets with these IPv6 addresses. According to the offload management strategy, one of the IPv6 addresses of Intf1 is configured with preferred state, while the others are configured in the deprecated state, just to maintain the ongoing sessions using these IPv6 addresses. The offload management decides to anchor new sessions to AHA1, through Intf1; thus, the IPv6 addresses replicated/moved from Intf2 to Intf1 (P2::Intf2 and P3::Intf2) are configured with the deprecated state, while the IPv6 address received in Intf1 from the Network 1 (P1::Intf1) is configured with the preferred state. The default route through AHA1 via the Intf1 is also configured with the lowest metric, in order to force the packets from ongoing sessions to be transmitted through Intf1, as well as to initiate new sessions.

D.3.2 Offloading sessions between interfaces connected to different access networks

The Case B provides the mechanism to offload sessions between interfaces connected to different access networks. While Case A just deals with IPv6 forwarding rules and IPv6 addresses management to offload sessions, Case B introduces the IPv6 tunnels management. There are some use-case scenarios from users' daily routine where it might be advantageous to enable this mechanism:

- offload data from an access network to another one of the same technology, using the same MN interface.
- offload data through an access network from a different access technology, using the MN multiple interfaces, such as from cellular to WiFi access networks.

The sub-case of Figure D.2 b) illustrates the scenario where sessions are offloaded to other interfaces connected to an AHA, while the sub-case of Figure D.2 c) illustrates the scenario where sessions are offloaded to other interfaces connected to a LegAN. The IP mobility might just select part of the AHAs associated to an interface, and consequently the attached sessions, to be forwarded to the other MN's interface, through direct tunnels with MN, when it is connected to a LegAN, or through tunnels with the connected AHA otherwise.

Offloading through an interface connected to a AHA:

In the sub-case of Figure D.2 b), the MN also decides to offload sessions from Interface 2 (Intf2) to Interface 1 (Intf1). The underlined instructions are the ones changed/added to perform the required offloading between Intf1 and Intf2, while the other ones were previously configured by the offload management. It is assumed that the MN has already

a set of four AHAs (AHA1, AHA2, AHA3 and AHA4) with anchored sessions, and that it is directly connected to AHA1 through Intf1 and to AHA2 through Intf2. There is a tunnel between AHA2 and AHA1 to route packets from/to P4::Intf1, as well as a tunnel between AHA3 and AHA4 to route packets from/to P3:Intf2. There is a rule in AHA1 to locally forward packets destined to P4::Intf1 to the next hop P1::Intf1, as well as a rule in AHA4 to locally forward packets destined to P3::Intf2 to the next hop P2::Intf2. In the MN, there is a default route to AHA1 for packets from P1::Intf1, and another default route to AHA4 for packets from P2::Intf2.

Considering the decision to offload sessions between interfaces, sessions anchored to AHA3 and AHA4 are forwarded to AHA1. This is partially achieved through the update of Tun2 and the configuration of the new Tun3. Hence, packets to IP address P2::Intf2 are forwarded to Tun3 in AHA4, while packets to P3::Intf2 are forwarded to Tun2 in AHA3. There is a new rule in AHA1 to locally forward packets destined to P2::Intf2, P3::Intf1 and P4::Intf1 to the next hop P1::Intf1. The MN has to replicate the IP address P2::Intf2 to Intf1, and has to move the IP address P3:::Intf2 also to Intf1, where IP addresses are configured in the deprecated state to maintain the ongoing sessions. From the proposed offloading decision, the IP address P1::Intf1 assumes the preferred state, while the others are changed to the deprecated state. It is also configured a default route through AHA1 via the Intf1 with the lowest metric, in order to force the packets from ongoing sessions to be transmitted through Intf1, as well as to initiate new sessions.

Offloading through an interfaces connected to a LegAN:

In Figure D.2 c), the MN also decides to offload sessions from Intf2 to Intf1. The underlined instructions are the ones changed/added to perform the required sessions offloading from Intf2 to Intf1, while the other ones were previously configured by the offload management. It is assumed that the MN has already a set of three AHAs (AHA1, AHA2 and AHA3) with anchored sessions, and it is directly connected to LAR1 through Intf2 and connected to LAR2 through Intf1. The IPv6 addresses P1::Intf and P2::Intf2, received from LAN2 and LAN1 respectively, cannot be assigned to sessions requiring IP session continuity. Thus, the Tun2 interface, configured with IP address P5::Intf1, is the preferred one to use the Intf1, while the Tun3 interface, configured with IP address P3::Intf2, is the preferred one to use the Intf2. The sessions anchored to AHA1 (P4::Intf1) and AHA2 (P5::Intf1) are tunneled to P1::Intf1 through tunnels Tun1 and Tun2 (attached to Intf1), respectively, while sessions anchored to AHA3 (P1::Intf2) are tunneled to P2::Intf2, through the tunnel Tun3 (attached to Intf2). In the MN, there is a default route to LAN2 via Tun2 for packets from P5::Intf1, and another default route to LAN1 via Tun3 for packets from P2::Intf2.

Considering the decision to offload sessions between interfaces, sessions being received by Intf2 are offloaded to Intf1. Hence, sessions anchored to AHA3 are offloaded through Tun3, which is attached to Intf1, instead of Intf2, in the MN. In the AHA3, the tunnel end-point is changed from the IPv6 address P2::Intf2 to the IPv6 address P1::Intf1. Thus, packets to IPv6 address P3::Intf2 are forwarded to Tun3 in AHA3. Since each IPv6 address is configured in a different interface (physical or tunnel interface), all IPv6 addresses remain in the preferred state, and the preference of the interface is the crucial metric. A default route to LAN2 via Tun2 is configured with the lowest metric; thus, new sessions are initiated through Tun2 with IPv6 address P5::Intf1. The MN also configures three routes to forward packets from IPv6 addresses P3::Intf2, P4::Intf1 and P5::Intf1 to Tun3, Tun1 and Tun2, respectively.

Table D.1: Parameters changed in each scenario

Scenario	Max value of Pause and Walk Intervals (s)	Avg Sessions Arrival Interval (s)	Avg Session Duration (s)	Connection Probability (ARs and eNBs)
A/B	300, 600, 1200, 2400	120	240	0.5
C	480	120	240	0, 0.5, 1
D	480	120	240, 480, 960	0.5
E	480	120, 240, 480	240	0.5

D.4 Evaluation

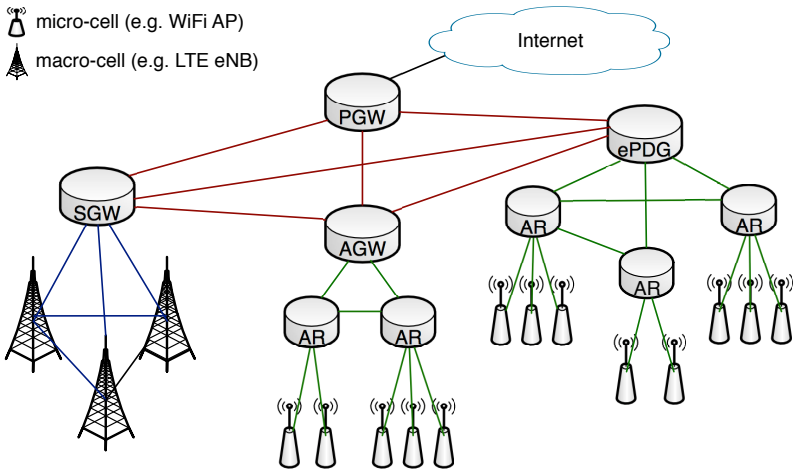


Figure D.3: Example of a wired topology.

We evaluate the dynamic offload anchoring in comparison with centralized anchoring through MATLAB [30], where the simulation time is 1 hour and the simulation step is 100ms, which is a sufficient value to accurately integrate user movement and ongoing sessions in the evaluation of the data cost and the data packet delay. In the wired topology, we randomly generate flat network topologies, where Figure D.3 illustrates an example. We assume a more flatten network topology, which is a mixture between a hierarchical network and a flat network. A node connects to the higher hierarchy level node, to lower hierarchy level nodes, and it can also connect to nodes of the same hierarchy level. The wired links are defined with the following delays and connection probabilities, in order to achieve a real network.

- **Among PGW, SGW, ePDG and AGW:** delay of 16 ms and a connection probability of 100%.
- **eNB-SGW or AR-AGW or AR-ePDG:** delay of 8 ms and a connection probability of 100%.
- **eNB-eNB or AR-AR:** delay of 4 ms and a variable connection probability (Table D.1).
- **AR-AP:** delay of 1 ms and a connection probability of 100%.

In the wireless part of the simulated scenario, illustrated in Figure D.4, we define two types of cells: macro-cells and micro-cells. The macro-cells mimic the cells currently

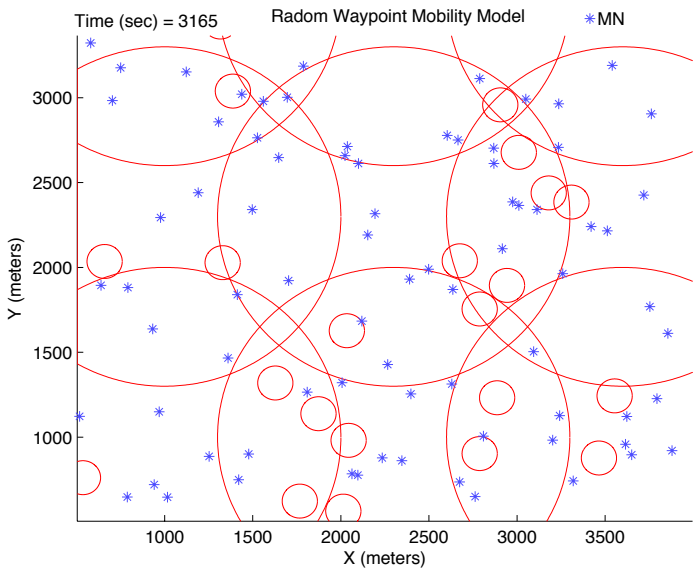


Figure D.4: Snapshot from an example of a wireless scenario.

associated to cellular access networks (e.g. Node from UMTS and eNB from LTE), while micro-cells represent the cells from the WiFi access networks (e.g APs from 802.11b/g). We distribute 9 macro-cells in a 3x3 grid to ensure the full coverage in a squared area of 4Km x 4Km where MNs move. We also place 27 micro-cells uniformly distributed in the squared area (4Km x 4Km) where MNs move. The wireless scenario is just used to know the current cells where the MN is connected to, in order to calculate the data cost and data packet delay of the ongoing sessions in the network, according to the offloading strategy.

MNs move inside the coverage area of the set of macro-cells (4x4 $\boldsymbol{km^2}$) with a Random Way Point mobility model with the following parameters:

- **Speed:** uniformly distributed between 1 and 10 m/s.
- **Pause Interval:** uniformly distributed between 1 min and a maximum variable value (Table D.1).
- **Walk Interval:** uniformly distributed between 1 min and a maximum variable value (Table D.1).

We assume that CNs can be attached to any node of the network; thus, sessions can be established with any network node, which are randomly selected. The sessions for each MN are established between the CNs (network nodes) and the MN itself, with the following characteristics:

- **Sessions arrival interval:** exponentially distributed with a variable average (Table D.1).
- **Session duration:** exponentially distributed with a variable average (Table D.1).
- **Data packet rate:** uniformly distributed between 128 Kbps and 2 Mbps.
- **Data packet size:** uniformly distributed between 128 bytes and 1 Kbytes.

In the evaluation of the data cost and the data packet delay, the simulations are repeated 10 times to improve the accuracy of the results with a confidence interval of 95%. The evaluated metrics are defined as following:

Data Cost: measures the total cost needed to deliver data packets from the CNs to the MNs. It is calculated through the sum of the data packets cost, where each data packet cost is the multiplication of the data packet size, including IPv6 encapsulation if it exists, by the time that the packet spends to traverse the network.

Data Packet Delay: measures the average time that a data packet experiences to be transmitted from a CN to an MN.

Our main goal is to evaluate the dynamic mobility anchoring, based on a well defined offloading strategy; thus, we define the offloading strategy to perform the offload of all ongoing and new sessions through the micro-cells, when the MN is in the coverage area of a micro-cell and its speed is approximately zero. Otherwise, all sessions will be maintained or initiated through the attached macro-cell.

We evaluate two different strategies to place the AHAs at the access network level. In one of the strategies, called '*AHAs in AGs*', the AHAs are distributed through the access gateways of the network (SGW, AGW and ePDG). In the other strategy, called '*AHAs in ARs*', the AHAs are distributed through the access routers of the network (ARs and eNBs). In the strategy to distribute the AHAs through the access gateways, we consider that an AHA may assign IPv6 prefixes/addresses to the MNs, but it is needed an IPv6 tunnel between the access gateway and the access router (e.g. between SGWs and eNBs or between ePDGs/AGWs and ARs) to ensure IP packet forwarding until the attachment point of the MN. Both strategies are compared with the centralized HA from the current centralized offload anchoring model.

D.4.1 Pause and Walk Intervals

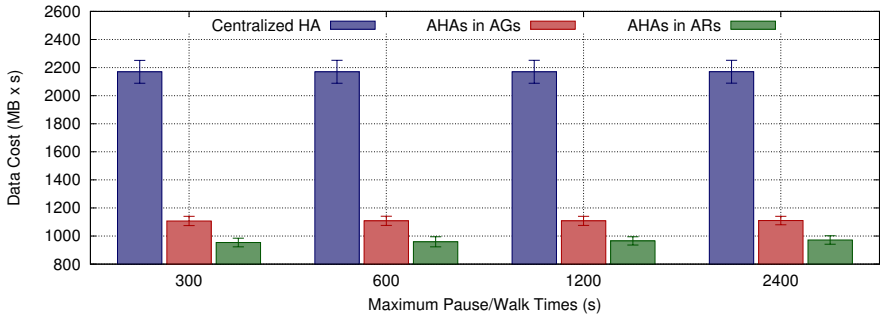


Figure D.5: Data Cost changing the maximum pause/walk time.

We start by evaluating the influence of the pause and walk intervals of the random waypoint mobility model, in the data cost and data packet delay, defining the variable values according to the Scenario A/B of Table D.1.

As illustrated in Figures D.5 and D.6, the centralized HA introduces a higher data cost in the network and a higher data packet delay in user sessions, when compared with distributed AHAs (AHAs in AGs or ARs). Sessions are uniformly coming from any node of the network, and a considerable amount of these sessions are initiated and terminated in the same AR/eNB. Thus, the '*AHAs in ARs*' approach provides the optimized routing path without tunneling to these sessions, while in the centralized HA all packets are

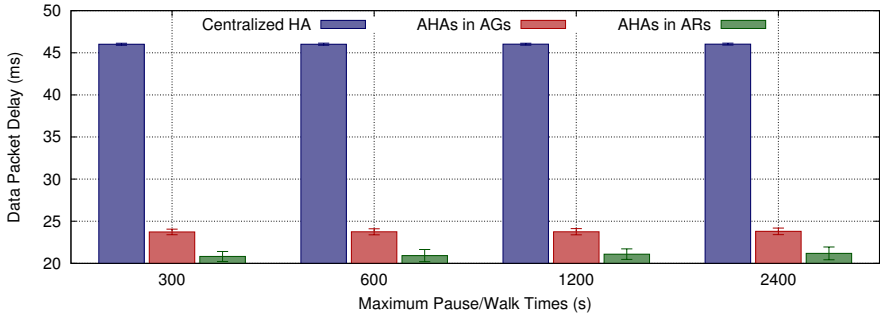


Figure D.6: Data Packet Delay changing the maximum pause/walk time.

firstly routed to the HA, which then tunnels these packets to the current MN location (AR or eNB). Moreover, in the approach '*AHAs in ARs*', the amount of packets from a session that are offloaded through another AHA (eNB or AR) different from the anchored one is not very high, and these AHAs can be directly connected, creating short tunnels between them. The increase of the maximum value of pause and walk intervals does not significantly affect any of the approaches, since it is higher than the average session duration, and sessions are coming from all nodes of the network.

D.4.2 CN Location

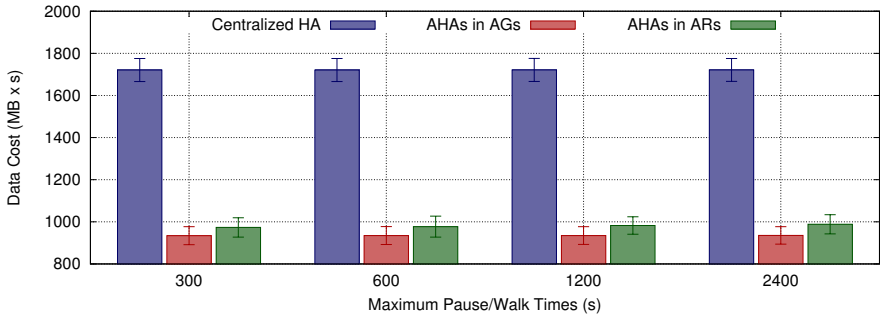


Figure D.7: Data Cost changing the maximum pause/walk time for CN2.

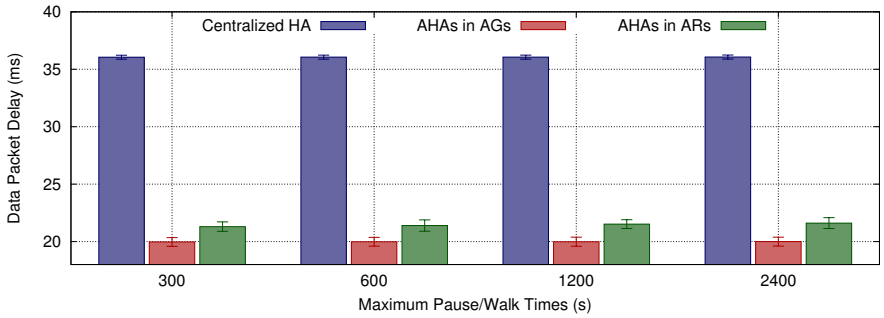


Figure D.8: Data Packet Delay changing the maximum pause/walk time for CN2.

In the previous test, we evaluated the influence of pause and walk intervals assuming that CNs are attached to any node of the network (strategy **CN1**), thus, sessions are

uniformly coming from any network node. However, in this set of simulations, we test two different ways to receive sessions:

- **CN2** sessions are uniformly coming from PGW, SGW, AGW and ePDG.
- **CN3** sessions come from the PGW.

We evaluate the two new CN placement strategies to receive sessions, repeating the same conditions of the previous evaluation, presented in Scenario A/B of Table D.1.

When sessions are uniformly coming from PGW, SGW, AGW and ePDG (Figures D.7 and D.8), the dynamic offload anchoring approaches continue to reduce both data cost and data packet delay. The main difference is that '*AHAs in AGs*' is now the approach with the lowest data cost and data packet delay, and the approach '*AHAs in ARs*' is in the second place. However, both AHA approaches highly reduce the data cost and data packet delay, with both CN1 and CN2 strategies, when compared with the centralized HA. A considerable amount of these sessions are initiated and terminated in the same AR/eNB; thus, both AHA approaches provide the optimized routing path. However, the approach with '*AHAs in AGs*' introduces short tunnels between AGs and ARs, increasing the data cost. The amount of packets from a session that are offloaded through another AHA (eNB or AR) different from the anchored one is not very high. In this case, the '*AHAs in AGs*' provides the shortest routing paths, through tunnel between AGs or between an AG and an AR. The '*AHAs in ARs*' approach also provides short routing paths, similarly to the ones of '*AHAs in AGs*', since the AHAs can be directly connected, creating short tunnels between them. In the centralized HA, all packets are firstly routed to the HA, except for sessions initiated with PGW, which then tunnels these packets to the current MN location (AR or eNB).

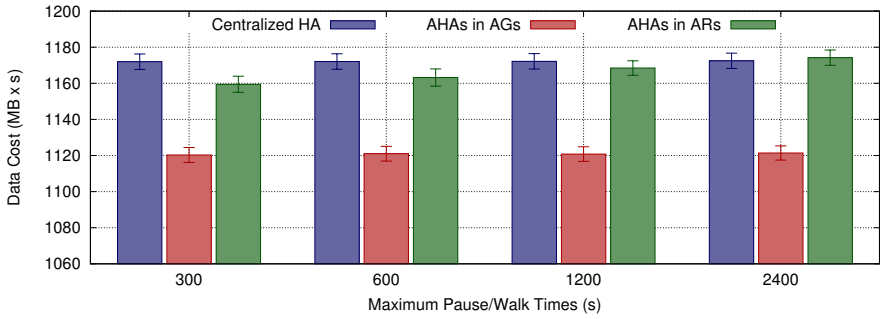


Figure D.9: Data Cost changing the maximum pause/walk time for CN3.

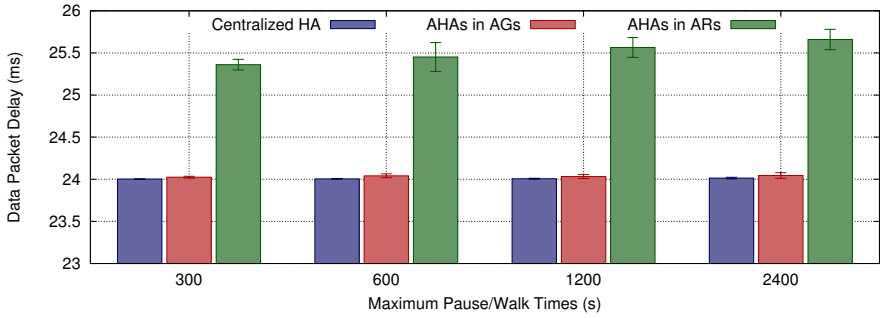


Figure D.10: Data Packet Delay changing the maximum pause/walk time for CN3.

When sessions are coming just from PGW (Figures D.9 and D.10), the centralized HA is the unique approach capable to provide the optimized routing path to all packets at the cost of long tunnels between the HA (PGW) and the ARs (ARs or eNBs). The centralized HA is the approach with the lowest data packet delay (Figure D.10), but the approach with '*AHAs in AGs*' has a quite similar performance regarding the data packet delay, since a considerable amount of these sessions are initiated and terminated in the same AG (SGW, AGW and ePDG). The approach '*AHAs in ARs*' increases the data packet delay of user sessions, but the average added delay is not significant (less than 2 ms). Regarding the data cost, the centralized HA introduces the higher data cost in the network, since it creates long tunnels to provide the optimized routing path to user sessions. Hence, the approach '*AHAs in AGs*' presents the lowest data cost, even when sessions are always coming from the PGW. Besides the fact that the approach '*AHAs in ARs*' slightly increases the data packet delay when compared with centralized HA (less than 8%), it is able to reduce the data cost in the network. The increase of the maximum value of pause and walk intervals slightly increases the data packet delay and data cost of the approach '*AHAs in ARs*', but it does not significantly affect the evaluation, since the maximum value of pause and walk intervals is higher than the average session duration.

D.4.3 Connection Probability

We change the probability of connection between the network nodes in the same hierarchical level, such as between ARs or between eNBs. A lower connection probability represents a more hierarchical network, while a higher connection probability leads to a flatten network. We assume that CNs are distributed through the network nodes, according to the previously defined strategies CN1 and CN2. The variable values are defined according to the Scenario C of Table D.1.

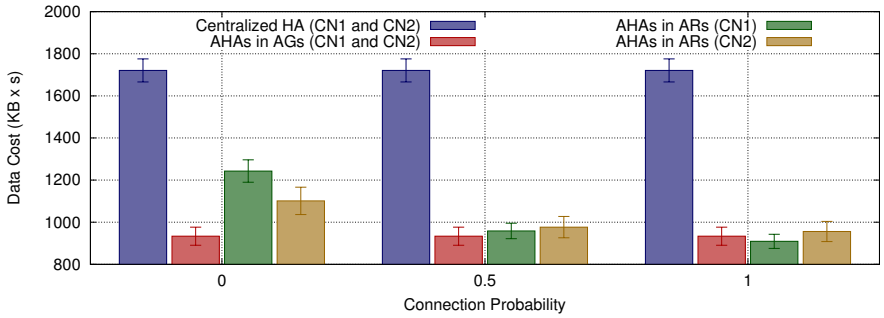


Figure D.11: Data Cost changing the connection probability.

As can be seen in Figures D.11 and D.12, with CNs placement strategies CN1 and CN2, the AHA approaches always present a lower data cost and data packet delay than the centralized HA, independently of the connection probability between ARs or eNBs. The approach '*AHAs in AGs*' is the one with the lowest data cost and data packet delay, except for the combination of the connection probability of 1 and placement strategy CN1, where the approach '*AHAs in ARs*' presents the lowest data cost and data packet delay. The approach '*AHAs in ARs*' is the only one that is influenced by the connection probability of ARs and eNBs, since the AHAs are placed in these nodes, while other approaches place the AHAs/HA at higher hierarchical levels, where packets are always firstly routed. The increase of the connection probability decreases both the data cost and the data packet delay, but this decrease is more notorious with the CNs placement

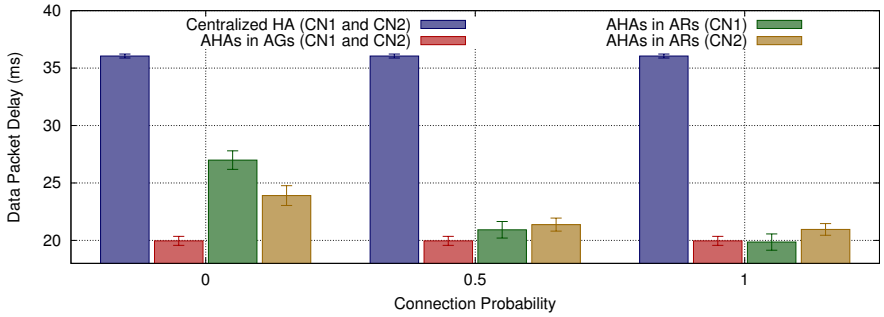


Figure D.12: Data Packet Delay changing the connection probability.

strategy CN1.

D.4.4 Session Duration

The impact of the average session duration in the data cost and data packet delay is evaluated, assigning the variable values according to the Scenario D of Table D.1.

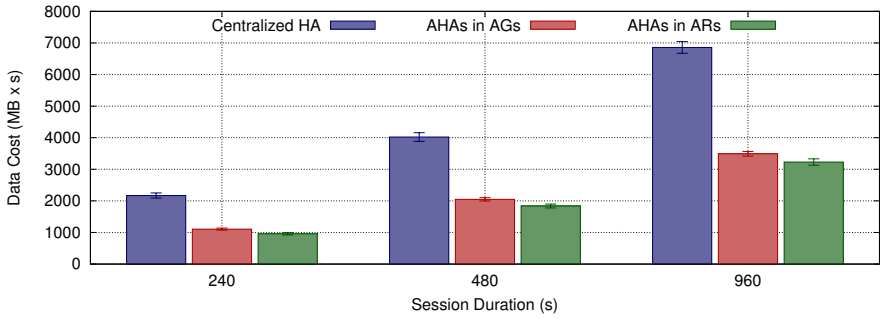


Figure D.13: Data Cost changing the session duration.

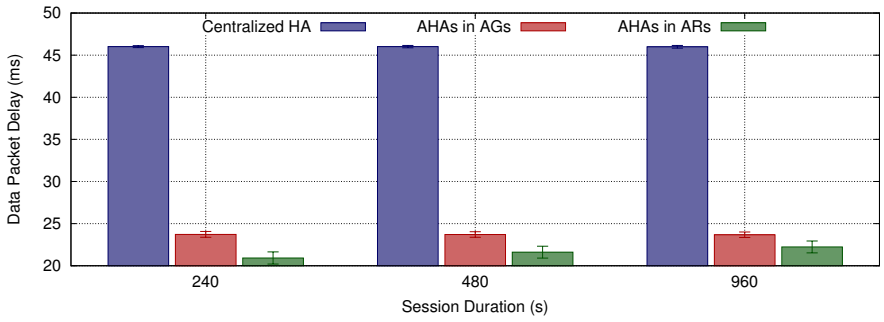


Figure D.14: Data Packet Delay changing the session duration.

The session duration has a high impact on the data cost (Figure D.13), where longer sessions increase the data cost in the network for all the three approaches. However, the AHA approaches have a lower data cost than the centralized HA, which are nearly half of the data cost of centralized HA, due to the dynamic offload anchoring. The increase of the average session duration (Figure D.14) slightly increases the data packet delay for the '*AHAs in ARs*' approach, which achieves values similar to the '*AHAs in AGs*' approach

for a session duration of 960 seconds. However, the AHA approaches have a much lower data packet delay than the centralized HA, independently of the session duration.

D.4.5 Sessions Arrival Interval

The influence of the sessions arrival interval in the data cost and data packet delay is evaluated, defining the variable values according to the Scenario E of Table D.1.

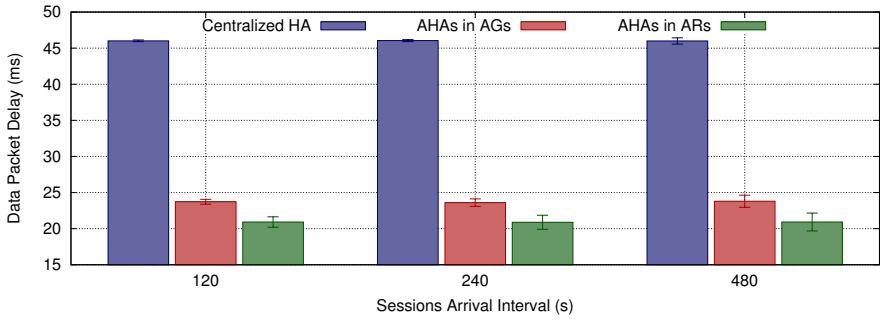


Figure D.15: Data Packet Delay changing the sessions arrival interval.

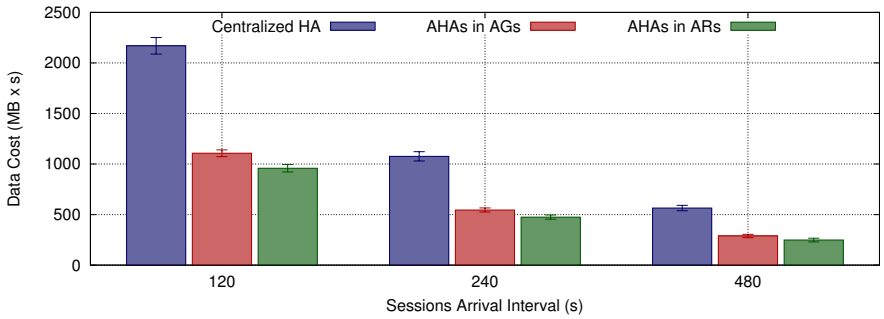


Figure D.16: Data Cost changing the sessions arrival interval.

The data packet delay is not significantly influenced by the sessions arrival interval (Figure D.15), since it just interferes with the average number of active sessions per instant, but the average packet delay remains quite similar. The data packet delay is lower for AHA approaches than for the centralized HA. The sessions arrival interval has a high impact on the data cost (Figure D.16), where a shorter sessions arrival interval increases the data cost in the network. The AHA approaches have a lower data cost than the centralized HA, which are nearly half of the data cost of centralized HA, due to the dynamic offload anchoring.

D.5 Proof of Concept

The previous section compared the dynamic offload anchoring with centralized anchoring through simulations, while this section aims to demonstrate that it really works in a real testbed. Hence, the testbed deployed to validate dynamic offload anchoring is presented in Figure D.17.

The Network of the Institute of Telecommunications (IT) of Aveiro, Portugal, was used as the core network, since it already provides IPv6 through the Neighbor Discovery

Protocol messages. We connected three Single Board Computers (SBCs) with OpenWRT through ethernet cable to the IT Aveiro network. Each SBC was configured to provide a different IPv6 prefix through the WiFi 802.11g and one of the ethernet interfaces, where the *radvd* daemon announces the respective IPv6 prefixes. The SBCs can be configured as legacy access routers or as AHAs. In the AHAs, the *radvd* is configured with the HA flag equal to 1, and to announce the IPv6 address of the interface instead of the network prefix. The MN is a laptop with three interfaces: a built-in WiFi 802.11b/g (wlan0), an usb WiFi 802.11b/g (wlan1) and an ethernet (eth0) interface.

The testbed was experimented with IPv6 services from the Internet, but to evaluate the presented mechanisms, it was used the Distributed Internet Traffic Generator (D-ITG) [31], installed at both the CN and the MN endpoints, where UDP and TCP sessions are established between the CN and the MN at different time instants. In the testbed evaluation, we provide results along the tested time for data packet delay and bitrate of UDP and TCP sessions.

The main focus of the proposed mechanisms is on the dynamic mobility anchoring to optimize the session delivery for both user and network; thus, the seamless handover time is not improved. In the testbed, it is used the connection manager of *Ubuntu* to perform the handovers, without any seamless technique. However, the proposed mechanisms can be integrated with seamless handover techniques, such as the ones provided by [32] [33] [34] already working with MIPv6 and PMIP. In these techniques, part of the configurations for the new network are performed during the preparation phase, reducing the time of the execution phase. The implemented testbed just initiates the procedures for IP session continuity after the conclusion of the Duplicated Address Detection (DAD) mechanisms for the new obtained IPv6 global address, which introduces a significant handover latency in single-interface data offloading scenarios.

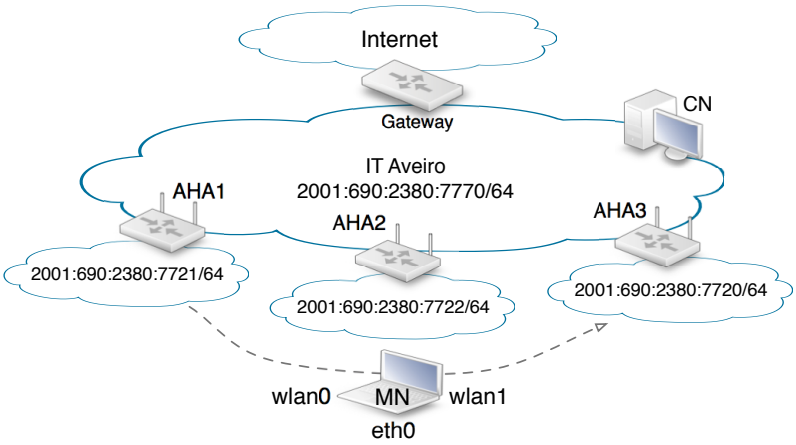
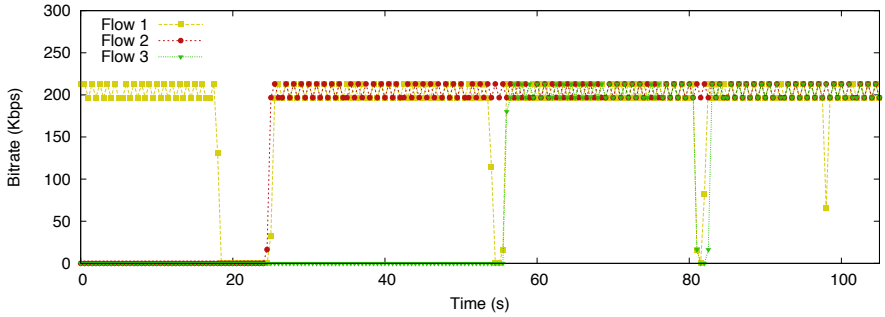
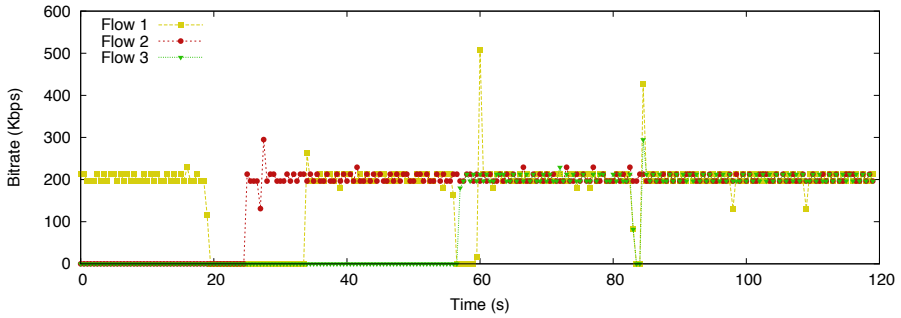


Figure D.17: Testbed Scenario

In the first test, UDP and TCP flows are configured with a payload size of 1024 bytes, and a rate of 200 Kbps. Three flows are initiated along the experiment, which are maintained until the end of the experiment. Flow 1 is anchored to AHA1 and it is established through wlan1 interface; flow 2 is anchored to AHA2 and it is established through wlan1 interface; flow 3 is anchored to AHA3 and it is established through wlan0. As can be seen from Figure D.18, the bitrate is maintained even when the MN performs handovers or when the data is offloaded. The first handover takes place at around 20 seconds, and is the one with higher latency, since it is performed using the same interface; the other two handovers, taking place at around 60 and 80 seconds, are performed using

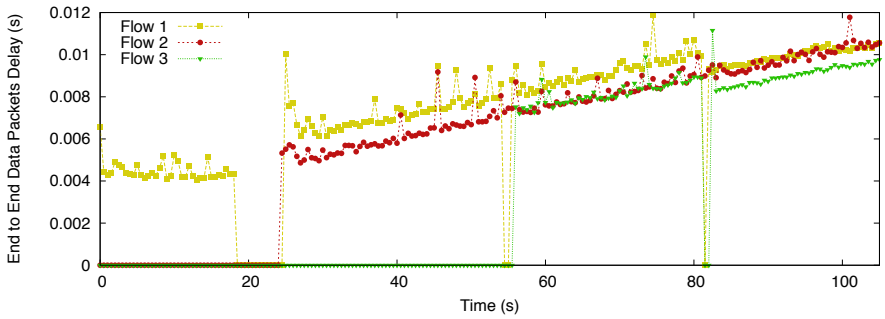


(a) UDP Sessions

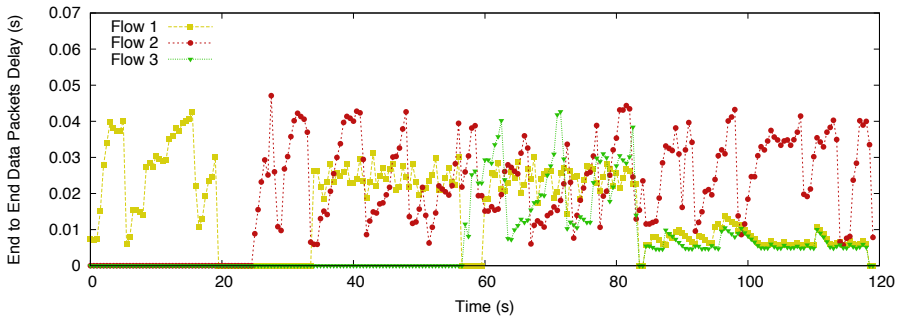


(b) TCP Sessions

Figure D.18: Evaluation of Bitrate



(a) UDP Sessions



(b) TCP Sessions

Figure D.19: Evaluation of Data Packets Delay

multiple-interfaces to provide data offloading. The first handover might be applied to offload data from a certain eNB to another eNB with more available resources, while the other two handovers might be performed to offload data through a WiFi network or a wired network, respectively. From Figure D.19 (a), we can observe that flows 1 and 3 are offloaded through the ethernet interface of the MN connected to AHA3 (at 83 seconds of experiment), which reduces the data packet delay. The delay of flow 1 is higher than the one of flow 3, since flow 1 is anchored to AHA1, which needs to forward the packets of flow 1 to AHA3, while flow 3 is anchored to AHA3 in the establishment of the session. In Figure D.19 (b), the offloading of flows 1 and 3 through the ethernet cable reduces the average data packet delay and jitter of TCP sessions, since the wireless medium access is usually loaded with several private WiFi networks, while the ethernet cable provides a stable connection at 100 Mbps.

D.6 Conclusion

Current offload models are based on a centralized mobility anchor, which routes all mobile packets to the preferred network/interface of the user, as well as it manages all bindings of the users. These heavy-centralized models bring scalability and performance issues.

This article extends DMIPA to provide dynamic offload anchoring, where the anchors are distributed through the access network nodes closer to the user. Moreover, the MN manages its offload anchors, IPv6 addresses and IPv6 tunnels, eliminating the necessity of a centralized node for offload control. The dynamic offload anchoring is compared with centralized anchoring through simulations, being able to reduce the data cost inside the network and the data packet delay of the user's sessions, specially with more distributed content servers and flatten architecture networks. The offload mechanisms with IP mobility are also validated in a real testbed, in order to demonstrate that an appropriate offload management is able to offload sessions with session continuity support while improving the user experience.

D.7 References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," Cisco, White Paper, Feb. 2013.

[2] Ericsson, "Ericsson mobility report: On the pulse of the networked society," Ericsson, Report, Jun. 2013.

[3] C. B. Sankaran, "Data offloading techniques in 3gpp rel-10 networks: A tutorial." *IEEE Communications Magazine*, vol. 50, no. 6, pp. 46–53, 2012.

[4] H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers," Internet Engineering Task Force, RFC 5555, Jun. 2009.

[5] C. Perkins et al., "Mobility support in ipv6," IETF, RFC 6275, Jul. 2011.

[6] S. Gundavelli et al., "Proxy mobile ipv6," IETF, RFC 5213, Aug. 2008.

[7] H. Chan, "Problem statement for distributed and dynamic mobility management," IETF, Internet-Draft draft-chan-distributed-mobility-ps-05, October 2011, work in progress.

[8] T. Condeixa and S. Sargento, "Dynamic mobile ip anchoring," in *IEEE ICC*, Jun. 2013.

[9] Y. Choi, H. W. Ji, J. yoon Park, H. chul Kim, and J. Silvester, "A 3w network strategy for mobile data traffic offloading," *Communications Magazine, IEEE*, vol. 49, no. 10, pp. 118–123, 2011.

[10] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *Wireless Communications, IEEE*, vol. 20, no. 2, pp. 104–112, 2013.

[11] S. Dimatteo, P. Hui, B. Han, and V. Li, "Cellular traffic offloading through wifi networks," in *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, 2011, pp. 192–201.

[12] 3GPP, "Technical Specification Group Services and System Aspects; LOcal IP Access and Selected IP Traffic Offload (LIPA-SIPTO)," 3rd Generation Partnership Project (3GPP), TR 23.829, Oct. 2011.

[13] ——, "Technical Specification Group Services and System Aspects; IP flow mobility and seamless Wireless Local Are Network (WLAN) offload," 3rd Generation Partnership Project (3GPP), TS 22.101, Mar. 2012.

[14] A. De La Oliva, C. Bernardos, M. Calderon, T. Melia, and J. Zuniga, "Ip flow mobility: smart traffic offload for future wireless networks," *Communications Magazine, IEEE*, vol. 49, no. 10, pp. 124–132, 2011.

[15] R. Wakikawa et al., "Multiple Care-of Addresses Registration," Internet Engineering Task Force, RFC 5648, Oct. 2009.

[16] G. Tsirtsis et al., "Traffic Selectors for Flow Bindings," Internet Engineering Task Force, RFC 6088, Jan. 2011.

[17] ——, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support," Internet Engineering Task Force, RFC 6089, Jan. 2011.

[18] C. Bernardos, "Proxy mobile ipv6 extensions to support flow mobility," IETF, Internet-Draft draft-ietf-netext-pmipv6-flowmob-08, October 2013, work in progress.

[19] T. Melia, C. J. Bernardos, A. de la Oliva, F. Giust, and M. Calderón, "Ip flow mobility in pmipv6 based networks: Solution design and experimental evaluation," *Wireless Personal Communications*, vol. 61, no. 4, pp. 603–627, 2011.

[20] I. Ali, A. Casati, K. Chowdhury, K. Nishida, E. Parsons, S. Schmid, and R. Vaidya, "Network-based mobility management in the evolved 3gpp core network," *Communications Magazine, IEEE*, vol. 47, no. 2, pp. 58–66, 2009.

[21] I. WG. (2013, Aug) Distributed mobility management. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[22] H. Chan et al., "Requirements for Distributed Mobility Management," IETF, Internet-Draft draft-ietf-dmm-requirements-10, November 2013, work in progress.

[23] P. Seite, P. Bertin, and J. Lee, "Distributed mobility anchoring," IETF, Internet-Draft draft-seite-dmm-dma-06.txt, January 2013, work in progress.

[24] C. Bernardos and J. Zuniga, "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-02, Apr. 2013, work in progress.

[25] J. Lee, J. Bonnin, P. Seite, and H. Chan, "Distributed ip mobility management from the perspective of the ietf: motivations, requirements, approaches, comparison, and challenges," *Wireless Communications, IEEE*, vol. 20, no. 5, pp. 159–168, 2013.

[26] A. Yegin et al., "Corresponding network homing," IETF, Internet-Draft draft-yegin-dmm-cnet-homing-00, Jul. 2013, work in progress.

[27] P. Ernest, H. Chan, and O. Falowo, "Distributed mobility management scheme with mobility routing function at the gateways," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, 2012, pp. 5254–5259.

[28] H. Ali-Ahmad et al., "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *IEEE WPMC*, Sep. 2012.

[29] T. Condeixa and S. Sargento, "Studying the integration of distributed and dynamic schemes in the mobility management," *Computer Networks*, vol. 60, no. 0, pp. 46–59, 2014.

[30] T. M. Inc. (2013, Nov.) Matlab version 7.10.0 (r2010a). [Online]. Available: http://www.mathworks.com/

[31] D-ITG. (2013, Jun) Distributed internet traffic generator. [Online]. Available: http://traffic.comics.unina.it/software/ITG/index.php

[32] R. Koodli, "Fast handovers for mobile ipv6," RFC 5568, Jul. 2009.

[33] H. Yokota et al., "Fast handovers for proxy mobile ipv6," IETF, RFC 5949, Sept. 2010.

[34] J. Kim and S. Koh, "PMIPv6 with Bicasting for Soft Handover," Internet-Draft draft-jikim-bpmipv6-00, Sep. 2009.

Paper E

Centralized, Distributed or Replicated IP Mobility?

Tiago Condeixa and Susana Sargento
in *IEEE Communications Letters*

Centralized, Distributed or Replicated IP Mobility?

Tiago Condeixa and Susana Sargento

Abstract

This letter studies a network-based IPv6 mobility model based on Proxy Mobile IP, called Multiple Local Mobility Anchors. It replicates the Local Mobility Anchors of Proxy Mobile IP through access routers and gateways of the network. The Local Mobility Anchors maintain the binding between the well-known IPv6 address of the mobile node and its current access router, in order to achieve an optimized routing path to data sessions that require mobility support. Multiple Local Mobility Anchors approach is compared with Proxy Mobile IP and Distributed Mobility Anchoring through analytical models. The outcome demonstrates that it significantly improves the data cost at the expense of a slight increase in the signaling cost.

Keywords: IP mobility, proxy mobile IP, distributed mobility anchoring, local mobility anchor.

E.1 Introduction

There has been an increase of mobile data traffic and mobile devices, which is expected to continue in this decade [1]. Moreover, there has also been a paradigm shift in users' data behavior with the increase of communications between devices in the same geographical area, as well as the migration of content servers closer to the user. This increase in demand is having a serious impact on the dimensioning and planning of mobile networks, since mobile core networks are highly hierarchical/centralized, which introduces serious performance, scalability and reliability issues. Hence, there is a paradigm shift in the network architectures with the introduction of flat architectural models to deal with this growth of mobile data.

However, current IP mobility schemes are not prepared for such trends, since they are deployed in a centralized manner, relying on a centralized Gateway (GW) to manage all mobile data and mobility context. Thus, in IP mobility protocols, such as Proxy MIP (PMIP) [2], Mobile IPv6 (MIPv6) [3] and Hierarchical MIPv6 (HMIPv6) [4], the whole data is routed via a centralized Mobility Anchor (MA), such as the Local Mobility Anchor (LMA) in PMIP-based approaches and Home Agent (HA) in MIP-based approaches. Moreover, the bindings between the well-known IPv6 addresses of the Mobile Nodes (MNs) and their current IP addresses or their ARs are managed at this MA as well. The MIP-based approaches are host-based, since they introduce mobility functionalities in the MNs, such as handover detection and signaling, as well as to the tunnel end-points. Oppositely, PMIP-based approaches introduce the Mobility Access Gateway (MAG) in the ARs to hide the mobility from the MN. The mobility functionalities are moved from the MN to the MAG, which detects the handover of the MN and communicates this information to the LMA. MAG maintains also the tunnel end-points with LMA to deliver the data in the MN network. HMIPv6 introduces a new entity, called Mobility Anchor Point (MAP), placed closer to the user, in order to locally manage the bindings and traffic redirection,

while a MN remains in the MAP's area. However, HMIPv6 continues to adopt the centralized mobility anchoring from MIPv6, where traffic is firstly routed via the HA and then via the MAP. HMIP main focus in on reducing the handover latency and binding update cost in micro-mobility, through a hierarchical system of bindings, thus, it has the same problems of the centralized architectures MIPv6 and PMIP. As the mobile data increases, such centralized architectures may encounter scalability issues (e.g. network bottlenecks), security issues (e.g. attacks focused on the centralized MA), and performance issues (e.g. non-optimized routing).

The IETF recently charted the Distributed Mobility Management (DMM) working group [5], which has been focused on the requirements for a distributed framework [6]. Novel DMM schemes, such as Distributed Mobility Anchoring (DMA) [7], Dynamic Mobile IP Anchoring [8] and PMIPv6-based distributed anchoring [9], introduce the idea of distributing the MA through the Access Routers (ARs). The MN selects the MA located at its current AR for a new session, which is maintained for the whole session service time, even if the MN moves among ARs. This concept, called dynamic mobility anchoring, aims at optimizing the routing path, assuming that most of the sessions are short enough, such that a session is terminated before experiencing several IP handovers.

In this letter, we propose to deploy PMIP adopting a replication strategy, instead of using centralized or distributed mobility management models. Multiple Local Mobility Anchors (MLMAs) spreads the LMA functionalities through the ARs and the GW of the network, in order to maintain the whole bindings, between the well-known IPv6 address of a MN and its current AR, in each AR/GW of the network. Thus, MLMA is always able to enforce the optimized routing path to data packets, which is already provided by the adopted routing protocol or static routing to the non-mobile IPv6 communications. The goal of this letter is to evaluate the replication strategy of MLMA, in comparison with the centralized approach PMIP and the distributed approach DMA. It aims to answer the following question: what is the model with the lowest in-network communication cost (data and signaling network cost) to deal with the localized IP mobility management in flat network architectures: centralized, distributed or replicated? This comparison is performed through analytical models evaluating signaling and data costs.

The letter is organized as follows. Section II details the MLMA, comparing it with PMIP and DMA. Section III provides the analytical models of the signaling and data costs for the three approaches, while section IV performs the analytical evaluation. Finally, Section V concludes the letter.

E.2 Multiple Local Mobility Anchors

This letter describes the MLMA, which is a network-based approach to provide localized IP mobility support (e.g. operator network). MLMA introduces the idea of replicating the mobility functionalities of the LMA of PMIP through the ARs and the GW of the network, while in PMIP there is an unique and centralized LMA with the mobility functionalities. In the proposed approach, a LMA has both the functionalities of LMA and MAG from PMIP. Moreover, the MLMA does not need both traditional Home Address (HoA) and Care of Address (CoA), since each MN just uses one IP address with mobility support to establish/maintain the sessions requiring mobility support.

MLMA choses an available IPv6 prefix to be used for localized identification of the MNs, while they move inside the operator network, which is used to provide mobility support to the MNs. The IPv6 prefix works as an identification inside the operator network, while from the outside, it is seen as another IPv6 prefix that belongs to that operator

network. As long as a packet destined to an IP address of this IPv6 prefix enters in the operator network, there is a set of LMAs responsible to forward the packet to the current access network of the MN. The ARs always provide the IPv6 prefix with mobility support, which is the same in all access networks, to maintain the local identification of the MN. As illustrated in Figure E.1, there is a set of LMAs attached to the ARs and the GW of the network. The MN moves between the different access networks, and the ARs always provide the prefix *A0* with mobility support. Thus, the MN always maintains the IPv6 address *A0::MN* configured in its interface, and the LMAs are responsible to forward the data packets to this interface, in order to provide session continuity support and reachability. The network operator A announces to the Internet that IPv6 addresses starting with *A* belong to it, thus other operator networks forward packets to these IP addresses to GW1.

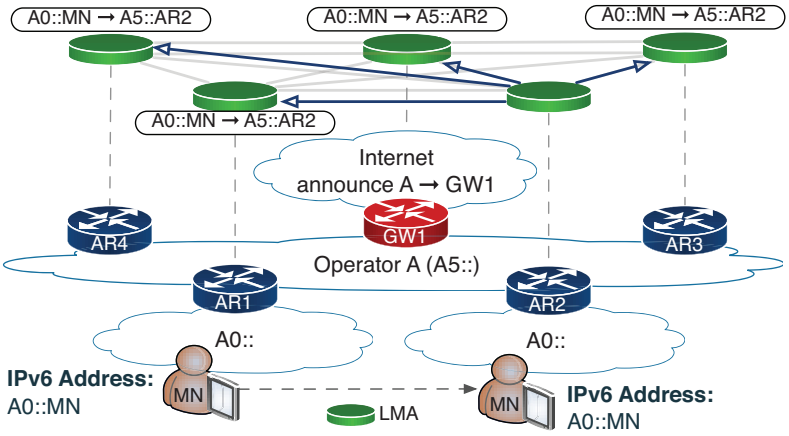


Figure E.1: Multiple Local Mobility Anchors

When a LMA detects a MN attachment through the Router Solicitation (RS) message, it updates all other LMAs of the network through the Proxy Binding Update (PBU) and Proxy Binding Acknowledgment (PBA) messages of PMIP, which contain the well-known IPv6 address of MN (mobility IPv6 address *A0::MN*) and the IPv6 address of the new AR of the MN. Therefore, it is assured that all LMAs are updated with the IPv6 address of the current AR of the MN, which is used to maintain the ongoing sessions, and in the reachability for new sessions establishments. As it happens in PMIP, each LMA of MLMA may need to maintain a larger binding cache proportional to the number of MNs. This is a disadvantage of these IP mobility models, but it is the price to pay to improve the data routing path optimization while providing session continuity and reachability to the MN. Although the size of the binding cache is larger, it is a simple database system with entries containing only two IPv6 addresses, that can be quickly managed with appropriate algorithms. MLMA is always able to provide the optimized routing path to data sessions, which is given by the routing protocol or the static routing, at the cost of more signaling messages to update the MN location in the binding caches distributed through the LMAs of the network.

The localized location management functionality of MLMA provides the discovery of the current AR (IPv6 address) of the MN inside the operator network, but it can be integrated, as well as PMIP, in a global location system (e.g. Domain Name System (DNS), Session initiation Protocol (SIP) or Location Identification Separation (LISP)) for communications between MNs, such as Voice over IP.

Table E.1: Notations and values for the analytical model

Notation	Definition	Default
N_m	Average number of MAs per MN per time instant	1.5/3
S_{bu},S_{ba}	Size of the PBU/PBA message	56 bytes
S_{ip}	Size of an IPv6 address	16 bytes
S_d	Size of a data packet	800 bytes
S_{dt}	Size of a tunneled data packet	840 bytes
μ_r	Average MN exchanged data packets rate	41 pkts/s
T_h	Average MN handover period/rate between ARs	300 s
μ_h	Average MN handover rate between ARs	1/300 s^{-1}
N_{ar}	Number of ARs in the network	50
P_o, P_i	Probability of an outside/inside data packet	0.75, 0.25
$H_{g\to a}$	Average number of hops between a GW and an AR	10
$H_{a\to a}$	Average number of hops between two ARs	5

E.3 Analytical Model

This section provides analytical models for PMIP, DMA and MLMA, to evaluate both signaling and data costs. The adopted notations and their values can be found in Table E.1.

E.3.1 Signaling Cost

The Signaling Cost (SC) is the cost introduced by the exchange of the mobility management messages per MN, defined as the size of the messages, multiplied by the hops between source and destination endpoints, and multiplied by the MN IP handover rate.

PMIP

The new MAG (AR) where the MN is connected sends a PBU message to the LMA, which replies with a PBA message. The total signaling cost is the addition of the two messages as defined in (E.1), considering the MN handover rate (μ_h), the number of hops between the LMA (GW) and MAGs (ARs), and the size of the PBU/PBA messages.

$$SC_{PMIP} = \mu_h\big(H_{g\to a}(S_{bu} + S_{ba})\big) \tag{E.1}$$

DMA

The current Mobility-capable AR (MAR) of the MN updates its previous MARs (ARs) with anchored sessions, after retrieving this information from the centralized mobility management database, as defined in (E.2). First, the current MAR sends a PBU message to the mobility database (GW), which replies with the set of MN's MARs and MN's respective IPv6 addresses. Then, current MAR establishes tunnels with previous MARs through PBU/PBA messages.

$$SC_{DMA} = \mu_h\big(H_{g\to a}(S_{bu} + (S_{ba} + 2N_m S_{ip})) + N_m H_{a\to a}(S_{bu} + S_{ba})\big) \tag{E.2}$$

MLMA

The current LMA (AR) of the MN updates all other LMAs placed in ARs and the GW of the network, with its IPv6 address. Thus, all these LMAs, except itself, are updated

through PBU/PBA messages, according to the MNs handover rate, as defined in (E.3).

$$\boldsymbol{SC_{MLMA} = \mu_h(S_{bu} + S_{ba}) \times \left(H_{g\to a} + H_{a\to a}(N_{ar} - 1)\right)} \tag{E.3}$$

E.3.2 Data Cost

The Data Cost (DC) is the cost to deliver data packets inside the operator network per MN, among ARs or between ARs and the GW. It is defined as the data packet size, multiplied by the hops between the source and destination, and multiplied by the average MN exchanged data packets rate. In the analytical models, we consider part of the data packets coming from the Internet through the GW, called outside data packets, and another part coming from the ARs of the operator network, called inside data packets. The analytical model for data cost evaluation is defined considering IP tunnels to forward data packets with mobility support among ARs, or between the GW and ARs. However, if the add-on for tunneling elimination is adopted, the analytical model can properly accommodate it, just by defining $\boldsymbol{S_{dt}}$ equal to $\boldsymbol{S_d}$.

PMIP

The data packets are always routed via the centralized mobility anchor (LMA), which is placed in the GW of the network, independently of the data being generated inside or outside of the network operator. Thus, while data packets from/to the outside are directly forwarded between MAGs (ARs) and the LMA (GW), the inside traffic is firstly forwarded to the LMA (GW) and then forwarded again to the current MAG (AR) of the MN, as defined in (E.4).

$$\boldsymbol{DC_{PMIP} = \mu_r\left(P_o H_{g\to a} S_{dt} + P_i H_{g\to a}(S_d + S_{dt})\right)} \tag{E.4}$$

DMA

Both inside ($\boldsymbol{P_i}$) and outside ($\boldsymbol{P_o}$) data sessions are always routed to the MAR, where these sessions were initially established. Then, the data sessions are forwarded to the current MAR of the MN, if the MN performed a handover to another MAR before the end of these sessions, as defined in (E.5). The probability of a data packet to be a handover packet ($\boldsymbol{P_h}$) was defined in [10].

$$\boldsymbol{DC_{DMA} = \mu_r\left(P_h H_{a\to a} S_{dt} + P_o H_{g\to a} S_d + P_i H_{a\to a} S_d\right)} \tag{E.5}$$

MLMA

Both inside ($\boldsymbol{P_i}$) and outside ($\boldsymbol{P_o}$) data are always forwarded through the optimized routing path, since the LMAs placed in ARs and the GW are able to forward the data packets to the current LMA (AR) of the MN, as defined in (E.6).

$$\boldsymbol{DC_{MLMA} = \mu_r\left(P_o H_{g\to a} S_{dt} + P_i H_{a\to a} S_{dt}\right)} \tag{E.6}$$

E.4 Evaluation

This section evaluates signaling and data costs of PMIP, DMA and MLMA, according to the analytical models presented before. The analytical signaling and data costs are defined per MN; thus, the number of MNs is not relevant to the analytical model evaluation. The analytical model depends on the hops distance between the network elements,

and does not depend on the topology nor the number of routers, where this information is integrated in the values defined for the hops distance. The other default values are defined in Table E.1.

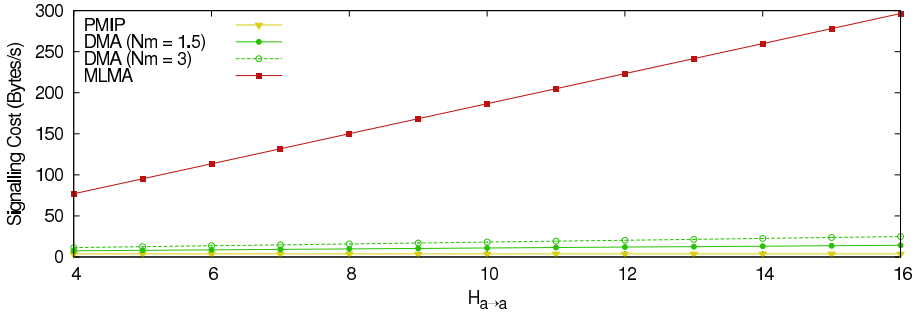


Figure E.2: Signaling Cost varying $\boldsymbol{H_{a \to a}}$

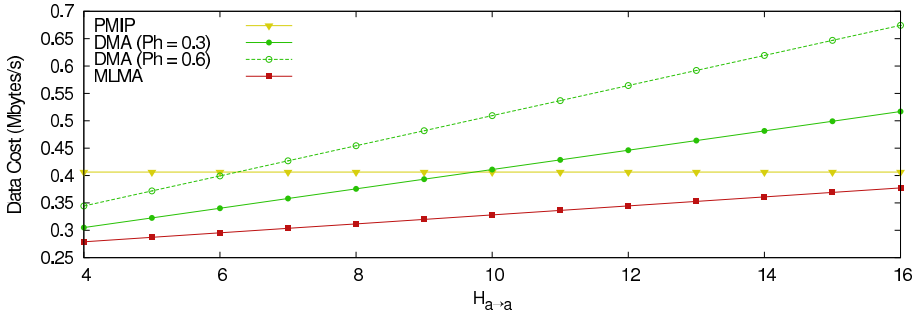


Figure E.3: Data Cost varying $\boldsymbol{H_{a \to a}}$

In the first test, we evaluate signaling and data costs changing $\boldsymbol{H_{a \to a}}$, and maintaining $\boldsymbol{H_{g \to a} = 10}$, Figure E.2 and E.3 respectively. The other parameters were defined according to the default values of Table E.1. The signaling cost of PMIP is the lowest one, and it does not depend on $\boldsymbol{H_{a \to a}}$, since there is no signaling messages exchanged between the ARs. The MLMA has a much higher signaling cost when compared with other approaches, and it increases much more with the number of hops between the ARs ($\boldsymbol{H_{a \to a}}$), since the current LMA (AR) of the MN updates all the other LMAs (ARs and GW) of the network per handover. In the worst case of MLMA ($\boldsymbol{H_{a \to a} = 16}$), the signaling cost is nearly 300 bytes/s higher than PMIP and DMA. However, this is a small price to pay, when the data cost of MLMA can reach improvements always higher than 25 Kbytes/s in the evaluated scenario (Figure E.3). For the smallest $\boldsymbol{H_{a \to a}}$ evaluated, MLMA is able to improve the data cost above 0.124 Mbytes/s when compared with PMIP, while for the highest $\boldsymbol{H_{a \to a}}$ evaluated, MLMA is able to improve the data cost above 0.137 Mbytes/s when compared with DMA. The data cost of PMIP does not depend on the $\boldsymbol{H_{a \to a}}$, since data packets are always routed via the centralized LMA, placed in the GW. For higher values of $\boldsymbol{H_{a \to a}}$ and $\boldsymbol{P_h}$, DMA has a higher data cost than PMIP, which fits scenarios of centralized network architectures with highly mobile users.

We evaluate the signaling cost, changing the MN handover rate between ARs (Figure E.4), and maintaining the other values according to the Table E.1. The decrease of the handover period $\boldsymbol{T_h}$ increases the signaling cost of the three approaches, since there are more messages along the time to update the current location of the MN. However, the decrease of $\boldsymbol{T_h}$ has more impact on the MLMA, since the signaling cost per handover is much higher due to the update of all LMAs (ARs and GW) of the network. The increased

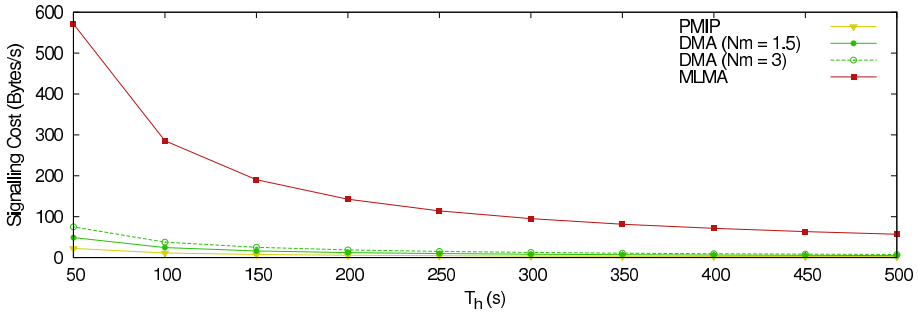


Figure E.4: Signaling Cost varying $\boldsymbol{\mu_h}$

signaling cost of MLMA, when compared with other approaches, is less than 600 bytes/s, which is around 2.3% of the minimum data cost gain (25 Kbytes/s) of MLMA in the previous scenario, per MN.

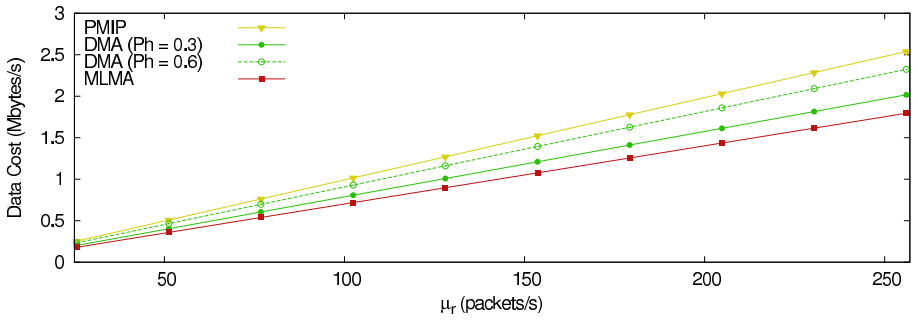


Figure E.5: Data Cost varying $\boldsymbol{\mu_r}$

Figure E.5 evaluates the data cost changing the exchanged data packets rate from 25 to 256 packets/s. The other values are the ones already defined in Table E.1. PMIP is the approach more affected by the increase of the data packets rate ($\boldsymbol{\mu_r}$), while MLMA is the less affected one. The lower data cost calculated per packet in MLMA allows a smooth growing of the data cost with the increase of $\boldsymbol{\mu_r}$. For the maximum evaluated $\boldsymbol{\mu_r}$ of 256 packets/s, there is a gain in MLMA data cost close to 1 Mbytes/s over PMIP, per MN.

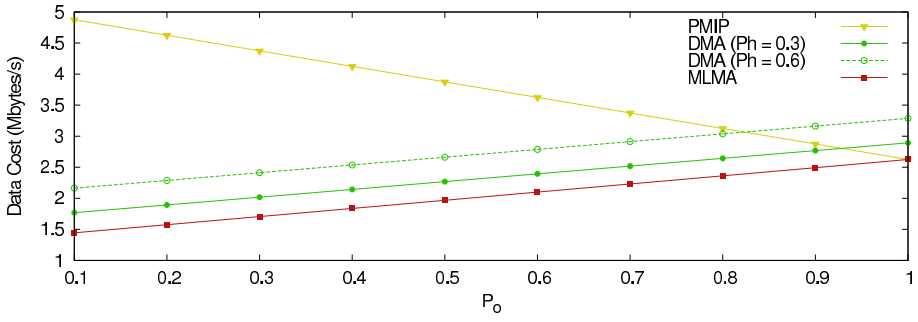


Figure E.6: Data Cost varying $\boldsymbol{P_o}$

Finally, we evaluate the data cost changing the probability of inside ($\boldsymbol{P_i}$) and outside ($\boldsymbol{P_o}$) data packets. The other values used are defined in Table E.1. PMIP was developed for hierarchical/centralized networks, thus it decreases the data cost with the increasing of $\boldsymbol{P_o}$, since the outside data packets are normally routed via the GW (LMA) through the optimized routing path. Oppositely, DMA and MLMA were developed for flatten network

architectures, thus the increase of $\boldsymbol{P_o}$, and consequently the decrease of $\boldsymbol{P_i}$, increases the data cost. However, MLMA has always the lowest data cost of the three approaches (except for $\boldsymbol{P_o = 1}$, which is equal to PMIP). As long as the communications in the same geographical area increase (higher $\boldsymbol{P_i}$), the DMA and MLMA increase the data cost gain over PMIP. MLMA is able to achieve the lowest data cost for $\boldsymbol{P_o = 0.1}$, where the data cost is decreased almost 3.5 Mbytes/s regarding PMIP, per MN.

Discussion

The replication strategy showed to be better than distributed and centralized models, regarding the in-network communication cost (signaling and data network costs). In scenarios with mobile users consuming more data traffic (e.g. requiring data sessions along the entire day while moving), the replication strategy significantly reduces the in-network communication cost. In scenarios with quite static users that require just few services during the entire day, the replication strategy introduces a higher complexity in order to maintain the replicated bindings through the entire LMAs of the network, which might not compensate the decrease of the in-network communication cost. However, the MLMA might be adjusted to overcome the different services and user requirements, since it might be integrated with a selective binding update strategy (e.g. choose the LMAs to send the PBU). The MLMA might update just the previous LMAs of the MN, for a user that stay most of the time at a predefined location (e.g. home or work). Thus, it might increase the data cost of the network, but it happens just for short periods of time when the user is outside of its predefined location. The MLMA might also update a set of LMAs of the network, which are usually used to establish sessions with the MN. Hence, the other non-updated LMAs of the network maintain the binding to the usual LMA of the MN, which then forwards the packets to the current LMA of the MN.

E.5 Conclusion

This letter proposes the network-based IP mobility model MLMA, which deploys the LMAs of PMIP through the ARs and the GW of the network. MLMA is compared against PMIP and DMA, through analytical models regarding signaling and data costs. The outcome shows that a slight increase of the signaling cost, to update the bindings among the replicated LMAs, is able to reduce the data delivery cost, through an optimized routing path. Hence, MLMA provides the lowest in-network communication cost to deal with localized IP mobility management in flat network architectures.

E.6 References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," Cisco, White Paper, Feb. 2013.

[2] S. Gundavelli et al., "Proxy mobile ipv6," IETF, RFC 5213, Aug. 2008.

[3] C. Perkins et al., "Mobility support in ipv6," IETF, RFC 6275, Jul. 2011.

[4] H. Soliman et al., "Hierarchical mobile ipv6 (hmipv6) mobility management," IETF, RFC 5380, Oct. 2008.

[5] IETF-WorkingGroup. (2013, Jul.) Distributed mobility management. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[6] H. Chan et al., "Requirements for Distributed Mobility Management," IETF Internet-Draft draft-ietf-dmm-requirements-05, Jun. 2013.

[7] P. Seite et al., "Distributed mobility anchoring," IETF Internet-Draft draft-seite-dmm-dma-06.txt, Jan. 2013.

[8] T. Condeixa et al., "Dynamic mobile ip anchoring," in *IEEE ICC*, Jun. 2013.

[9] C. Bernardos et al., "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-03, Oct. 2013.

[10] H. Ali-Ahmad et al., "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *IEEE WPMC*, Sep. 2012.

Paper F

Context-Aware Adaptive IP Mobility Anchoring

Tiago Condeixa and Susana Sargento
in *Elsevier Computer Networks*(submitted)

Context-Aware Adaptive IP Mobility Anchoring

Tiago Condeixa and Susana Sargento

Abstract

The novel distributed mobility management trend is a promising direction to cope with the increasing mobile data traffic and flatten network architectures. Most of the novel mobility approaches distribute the mobility anchors through the access level, as opposed to the centralized mobility anchoring model. Other recent approaches argue that mobility anchors closer to the content servers may be the solution to optimize the mobility performance. However, none of the mobility anchoring models is ideal for all scenarios, since it depends on the user, the session and the network. Hence, we propose an IP mobility approach driven by the context of the user, sessions and the network, where the mobility anchors for IP address allocation and for routing/forwarding are distributed through the network nodes, while the mobility context is managed by the mobile devices. Although each session is properly anchored in the establishment phase, the routing/forwarding is adapted over time, according to the user, the session and the network context: the proposed approach is able to signal different mobility anchors to optimize the routing path to new and ongoing sessions of the user. The outcome of the evaluation shows that the proposed approach overall reduces the data cost, the data delay, the tunneled packets and the tunnel length, when compared with other anchoring models.

Keywords: distributed mobility, dynamic anchoring, mobility context, routing/forwarding, reachability, session continuity.

F.1 Introduction

There has been a strong increase of mobile traffic, which is expected to continue in the next years [1] [2], and mobile operators have to adapt their network architectures in a near future. Service providers are already migrating the content servers closer to the user, such as the Content Delivery Networks (CDNs), in order to ensure high availability and delivery performance of this massive required content. Thus, users have been changing their traffic consumption behavior with the increase of demanding communications with content servers in the same geographical area. Moreover, there is a paradigm shift in the network architectures, with the introduction of flat models, and the adoption of data offloading strategies to deal with this growth of mobile traffic and to exploit the distribution of content servers. However, current IP mobility schemes are not prepared for such trends, since they are deployed in a centralized manner, relying on a centralized entity to manage all mobile data and mobility context. Thus, in existing mobility management protocols, such as Mobile IPv6 (MIPv6) [3] and Proxy MIP (PMIP) [4], the whole data is routed via a centralized Mobility Anchor (MA), like the Home Agent (HA) in MIPv6 and the Local Mobility Anchor (LMA) in PMIPv6. Moreover, the bindings between the well-known IPv6 addresses of the Mobile Nodes (MNs) and their current IPv6 addresses are managed at this MA as well. As the mobile data increases, centralized architectures may encounter scalability and performance issues (e.g. network bottlenecks, single point of failures, non-optimized routing).

The IETF recently charted the Distributed Mobility Management (DMM) working group [5], which has been focused on the requirements for a distributed framework [6].

Novel DMM schemes [7] [8] [9] introduce the idea of distributing the MA through the Access Routers (ARs). The MN selects the MA located at its current AR for a new session, which is maintained for the whole session duration, even if the MN moves among ARs. The routing path may be optimized, assuming that most of the sessions are short enough to be terminated before experiencing several IP handovers. Other DMM schemes have been recently proposed, such as the anchoring model close to the CNs [10]. Besides the advances in DMM, there is still no appropriate solution for all possible scenarios, since the protocols performance depends on several constraints, such as user mobility patterns, network topology/architecture and traffic sessions.

In this article we propose a distributed and dynamic IP mobility model driven by the context information from the user and the network, which means that IP session continuity and IP reachability are assured to the MN sessions according to individual characteristics/needs. Thus, two MNs in the same location might have a completely distinct set of MAs at the same time. Although there is an initial anchoring of mobility sessions, the proposed approach is able to signal other MAs to optimize the routing path to new and ongoing sessions of the MN. Moreover, the complexity of the bindings management is reduced, since each MN is responsible to maintain its own bindings and the respective signaling. The proposed mobility protocol is detailed with the initial anchoring, the re-anchoring for routing/forwarding and an operational example. We perform an evaluation of the adaptive anchoring through simulations with different scenarios, in order to obtain the data delay, tunneled packets, tunnel length and data cost, and compare it with other anchoring models.

The article is organized as follows. Section II analyzes the current work on mobility anchoring. Section III introduces the context-aware adaptive mobility anchoring, while Section IV evaluates it through simulations in comparison with other anchoring models. Finally, Section V concludes the article.

F.2 Current IP Mobility Anchoring

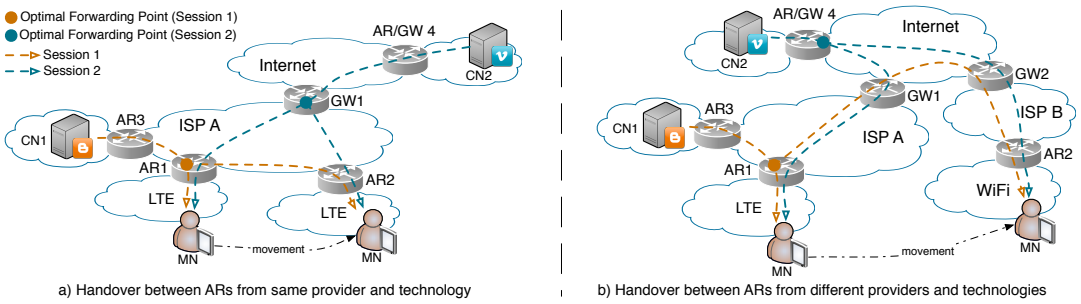


Figure F.1: Example of optimized routing path for two handover use-cases

IP mobility models are designed to provide IP session continuity and IP reachability to MNs. IP session continuity allows the user to maintain the ongoing sessions, while its device changes the attached network. A session is initially anchored to an MA, maintaining the IP address until the end of the session, and the IP mobility protocol provides routing/forwarding mechanisms (e.g. IP tunnels or routing rules) to deliver the packets of that session to the MN's current location. The reachability allows any device to reach an MN, based on a predefined identification (e.g. URL or IP address). The IP reachability is usually provided by centralized mobility models, based on a translation between a well-

known IP address of the MN and its current IP address. The well-known IP address is assigned by the centralized MA, which forwards the packet directed to the well-known IP address to the current IP address of the MN. However, most of the internet services need a global location system (e.g. Domain Name System - DNS), which translates a well-known URL into an IP address, before the IP reachability provided by the mobility management.

In centralized mobility approaches [3] [4], all sessions are anchored in the centralized MA, such as the HA in MIPv6 and the LMA in PMIP. Thus, the MA routes the packets to the current location of the MN, based on the binding between the well-known IP address of the MN and its current IP address. The anchoring model adopted by these solutions is denoted as 'centralized anchoring' in the remainder of the article.

The DMM approaches introduce novel concepts, such as dynamic anchoring and dynamic mobility. Dynamic mobility allows the mobility support to a session when it really needs it. Hence, mobility support is provided from the time that a session undergoes into a handover; otherwise, the session is maintained like a normal IP session. In DMM approaches [7] [8] [9], a new session is anchored in the current AR of the MN, while the previous sessions remain anchored to their initially assigned MAs. The previous MAs forward these sessions through tunnels to the current AR of the MN, in order to maintain the ongoing sessions. These approaches that provide the anchoring at the access router level are focused on improving the mobility performance for sessions that require IP session continuity; thus, IP reachability is provided by the centralized anchoring, where the MN receives a static IPv6 address. The approaches providing the mobility anchoring at the access router level are denoted as 'AR anchoring' in the remainder of the article.

The Corresponding Network Homing [10] introduces the idea of anchoring the sessions to MAs close to the CNs, which means that, for each new session with a CN from a different network, it is assigned an IP address of that network to the MN, which is used in the establishment of the new session. The MA in the CN network forwards the ongoing sessions to the current location of the MN through tunnels from the beginning of the session, thus the optimized routing path is always assured at the cost of longer tunnels. This approach is focused on improving the mobility performance for sessions that require IP session continuity; thus, IP reachability is still provided by the centralized anchoring. The approaches adopting the mobility anchoring close to the CNs are called 'CN anchoring' in the remainder of the article.

The on-demand mobility management [11] provides the mobility support, according to the session continuity and reachability requirements. The sessions with reachability needs use the centralized mobility model, while sessions that just need session continuity can use AR or CN anchoring. The sessions that do not require session continuity nor reachability will not have any mobility support. However, this on-demand idea requires the integration of Application Programming Interfaces (APIs) in the application layer to provide the mobility requirements of the sessions, which is not commonly provided by today's applications.

Mobility anchor selection [12] was introduced in DMM through the study of use-case scenarios. The work focuses on the initial anchoring of the sessions to MAs, based on history about the MN, application and network (e.g. flow duration and MN mobility behavior). This approach also requires some input from applications in the sessions establishment phase, and the estimation of non-easy or even unpredictable metrics (e.g. session duration). Moreover, optimization of the handover time is not in the scope of this work.

Although there are several advances in the mobility management, there is not an IP mobility approach suitable for a large set of scenarios, since its performance depends on several constraints, such as user mobility, network topology/architecture and traffic

sessions. As the last resource, we may decide to deploy the IP mobility protocol that is optimized for the majority of the nowadays scenarios. However, as observed from the last years, there are always new networking paradigms, and the assumptions to provide mobility management might change in a short period of time, such as different platforms, services, network architectures and mobile devices. It is then better to provide a more flexible and adaptive IP mobility anchoring model that can easily adapt to new network trends.

Figure F.1 illustrates two different handover scenarios, which have a high impact on mobility anchoring and traffic forwarding decision. While scenario a) illustrates a normal handover between two ARs in the same ISP/network domain (e.g. handover between LTE eNBs), scenario b) illustrates a handover between two ARs from different ISP/networks and technologies (e.g. traffic offloading through a wireless area network). The optimized routing path for two different sessions, before and after a handover, are highlighted in the scenarios. The MA with forwarding/routing functionalities that provides the optimized routing path for both sessions depends on the CN and MN location, as well as the network topology/architecture. Our aim will be to re-select a new MA with the forwarding/routing functionality for an MN session, after a handover or data offloading decision, to optimize the routing path during the session's lifetime. Hence, we propose a distributed mobility management model driven by the context information, which provides adaptive mobility anchoring for sessions that require IP reachability and/or IP session continuity, based on user and network context. The mobility approach takes advantage of the MAs distributed through the networks to shorten the routing path and to minimize the tunnel for each individual ongoing/new session.

F.3 Context-Aware IP Mobility Anchoring

The novel IP mobility anchoring model is based on the assumption that each MN dynamically selects its set of MAs for routing/forwarding from the available MAs of the network, according to its mobility degree, ongoing sessions and the network topology/architecture. The degree of personalization introduced by this context-aware IP mobility anchoring shall be able to improve the mobility performance, such as shortening routing paths and reducing the tunnels length.

F.3.1 Overview

The context-aware mobility anchoring protocol is developed considering that the MAs might be placed all over the network, usually at border nodes, such as routers in the MN network, the gateway of the MN network, routers in the CN network, the gateway of the CN network, or the CN itself. Hence, the mobility protocol exploits this placement of the MAs in the network, through the utilization of the proper mechanisms and signaling. The MN can anchor sessions to distinct MAs, which are able to provide IP address allocation and routing/forwarding, in order to provide the optimized routing with minimum tunneling for new and ongoing sessions. Some of the MAs placed in the corresponding networks are also assigned for routing/forwarding based on bindings provided by the MN. At least the corresponding networks with more established sessions (e.g. youtube, google and Facebook servers) shall deploy an MA in the AR or its Content Delivery Networks (CDNs). Using information from user mobility degree, the information about MAs availability in the network, the network topology and the MN ongoing sessions, we are able to provide the optimized routing path with a minimized tunneling, and without introducing a high

complexity to the network.

The proposed approach introduces a flexible and host-based mobility context management, which reduces the complexity to manage the mobility context in a centralized node, since each MN just has to maintain its own set of MAs, the respective bindings and mobility routes. The distribution of the MAs for IP address allocation and routing/forwarding, as well as the distribution of mobility context management through the MNs, eliminates the necessity of a centralized entity for the IP mobility management. The bindings are managed by the MNs, where each MN is responsible to create/update/remove the bindings in its set of MAs. The MN is able to dynamically select its set of MAs as long as it changes the access network or initiates/terminates sessions. The MAs allocate Mobile IP Addresses (MoA) to the MN, and the MN collaborates with MAs in the IP addresses management, providing the information about the requirement of the assigned IP addresses. There are also IP addresses assigned by the ARs of the network that do not provide any mobility support.

We consider that each session may be dynamically forwarded by a maximum of three distinct MAs in its lifetime: one in the AR of the MN's network, another in the GW of the MN's operator network, and another close to the CN. The MAs in the MN network (e.g. in ARs and GWs) may be known through Router Advertisement (RA) or Dynamic Host Configuration Protocol (DHCP) messages, where the IP addresses of the closest MAs have higher metric values. The MAs far from the MN, such as the ones close the CNs, are known through a Mobility Anchor Server (MAS), which contains the IPv6 prefix and the MA IP addresses serving this prefix. The MAS can be deployed using DNS or other similar system.

F.3.2 Initial Mobility Anchoring

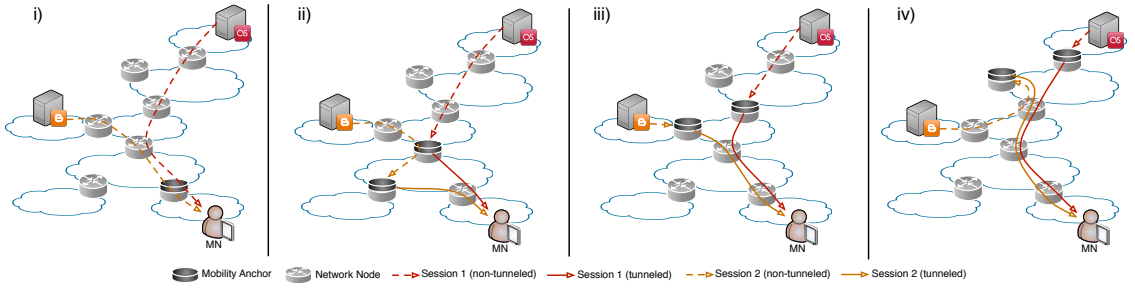


Figure F.2: Example of initial anchoring for sessions requiring IP session continuity

The MA assigned for a new session provides to the MN an IP address ensuring IP reachability and/or IP session continuity, as well as the respective routing/forwarding support from the establishment of the session.

IP Reachability: The anchoring of a session requiring IP reachability is assured by an MA previously selected. There is always just one MA per MN at a given instant for IP reachability support, which may be changed over time, providing both the IP address allocation and routing/forwarding functionalities. The IP address allocated by this MA is updated in the global location system, in order to provide an association between the MN identification and its location. The selection of the MA for IP reachability depends on the user mobility patterns, thus, highly mobile users may use just one static and not so close MA for IP reachability (e.g. centralized MA), while a more stationary user may have a small set of usually close MAs (e.g. in ARs) for IP reachability, selecting one of

them over time according to its context. The sessions that require IP session continuity may be anchored in other MAs distinct from the MA that provides IP reachability. Thus, an MN may manage an MA for IP reachability and a set of MAs for IP session continuity.

IP Session Continuity: The proposed approach follows several steps until it selects the MA that best fits a new session requiring IP session continuity. If the selected MA is already being used by another session, the new session can assign the MA without any operations.

The proposed approach always prefers to anchor a new session in the current AR of the MN (Figure F.2 i)), if the AR is an MA, since it provides the optimized routing path without tunneling or other mobility support. The current AR with MA functionalities provides the IP address with IP session continuity support to the MN.

If there is no MA in the current AR of the MN, the session will be anchored in a close MA, in the AN of the access network or in a close AR, as illustrated in Figure F.2 ii). Since the current AR of the MN is not an MA, the MN establishes a tunnel with this close MA, which provides to the MN an IP address with IP session continuity support.

If there is not a close MA in the access network, the anchoring is provided in two distinct places according to the CN location. For a close CN, the session is anchored in an MA close to the CN (e.g. AR of CN), as presented in Figure F.2 iii) with session 2. The MA close to the CN provides an IP address with IP session continuity support to the MN, and the MN establishes a tunnel with this MA to ensure the routing/forwarding functionality to the session being anchored there. For a far CN, the session is anchored in the centralized MA, as presented in Figure F.2 iii) with session 1. The centralized MA provides an IP address with IP session continuity support to the MN, as well as a tunnel with it to route/forward packets from the being anchored session.

If the current network does not have any MA, the session may be anchored in an MA close to a far CN, or at a last resort, in any centralized MA from other operator network, as presented in Figure F.2 iv) with session 1 and 2, respectively. In both cases, the selected MA provides an IP address with IP session continuity support, and a tunnel with the MN for routing/forwarding of the new sessions.

F.3.3 Mobility Re-Anchoring

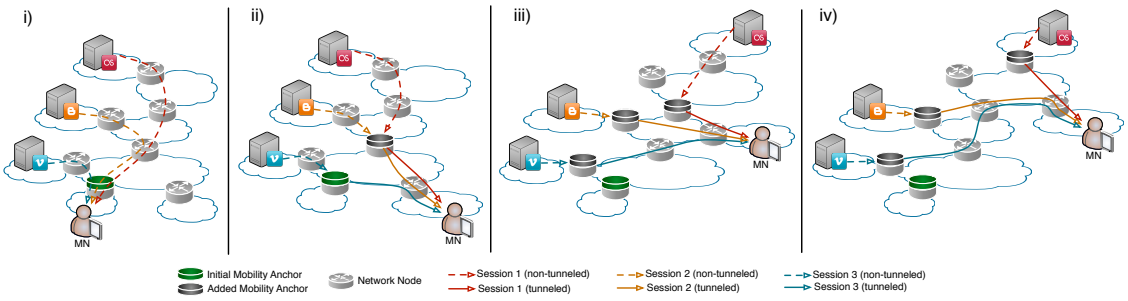


Figure F.3: Example of re-anchoring on routing/forwarding for ongoing sessions

The mobility re-anchoring is just possible for the routing/forwarding functionality, with main focus on the improved delivery of sessions requiring IP session continuity. However, the proposed approach also includes the possibility to change the MA for new sessions requiring IP reachability. It includes the possibility to update the global location system with the IP address of the MN, when it spends long periods in the same network, such as home and work. Although there is the possibility to update the global location system

(e.g. DNS), we are expecting a few daily location updates from the MN (e.g. a maximum of 5 per day), in order to maintain the global location nearly stable.

It is provided a simple mobility re-anchoring algorithm, but the proposed approach can be integrated with more advanced selection algorithms. This algorithm uses context from the user, the network and sessions, in order to shorten the routing path and minimize the tunnels. The MA of the session, the current AR of the MN, the CN location and the network topology are the parameters taken into account in the mobility re-anchoring decision. The algorithm determines the MA to provide an improved IP session continuity of the ongoing session, which provides routing/forwarding functionalities to the session, being enabled by an IP handover or a data offloading decision.

It is assumed that it is selected an initial MA in the establishment phase of the sessions, as illustrated in Figure F.3 i) for new sessions 1, 2 and 3. Thus, the IP address assigned to these three sessions in the beginning is the one provided by the initial MA, which has to be maintained by the MN until the termination of the three sessions. If there is any change, but the AR of the MN remains the MA of the session, the session is maintained active without any tunneling mechanisms. If the MN receives a new IP address or there is a decision to offload data through another MN interface connected to the same MA, there is just an update of the mobility routes/rules in the current MA to forward the packet of that session between IP addresses.

If the new AR and the MA of the session are in the same access network, the re-anchoring depends on the CN location, but the IP address provided by the initial MA needs to be maintained by the MN and assigned in the MA. For a CN located in the same access network, the MA remains the same, or it is selected the MA of the CN, if it is available, as illustrated by session 3 of Figure F.3 ii). If the MA remains the same, the MN or its new AR (if it is an MA) establishes a tunnel with the MA, and all packets to the MN are routed/forwarded to its new location. If it is selected the MA of the CN as the new MA, it is established a tunnel between the MA of the CN, and the MN or its current AR (if it is a MA), to provide the routing/forwarding functionality for IP session continuity. For a CN located outside of the access network, it is selected the MA in the AN of the network for routing/forwarding functionalities, which establishes a tunnel with the MN or its new AR (if it is an MA), as shown through sessions 1 and 2 in Figure F.3 ii).

If the new AR of the MN and the MA of the session are in different access networks of the same operator network, the re-anchoring depends on the location of the CN. For a CN located in the same operator network, it is selected the MA of the CN, as illustrated by sessions 2 and 3 of Figure F.3 iii). It is established a tunnel between the MA of the CN and the MN or its current AR (if it is an MA) to provide the routing/forwarding functionality for the sessions requiring IP session continuity. For a CN located outside of the operator network, it is selected the MA in the EN of the network for routing/forwarding functionalities, which establishes a tunnel with the MN or its new AR (if it is an MA), as presented by session 1 in Figure F.3 iii).

If the new AR of the MN and the MA of the session are in different operator networks, the re-anchoring is provided by the MA of the CN, independently of the CN location, as highlighted in Figure F.3 iv). It is established a tunnel between the MA of the CN and the MN or its current AR (if it is a MA) to provide the routing/forwarding for the session requiring IP session continuity.

F.3.4 Operation

The operation example is described based on Figures F.4 and F.5, where the signaling messages and sessions' routing paths are illustrated. The example addresses the protocol operation based on an MN moving from Home to Work that stops in a Cafe. The MN establishes three sessions with different CNs, while at Home, which will remain active during the movement of the MN to Work. The MAs are distributed over some of the network nodes, as shown in Figure F.4.

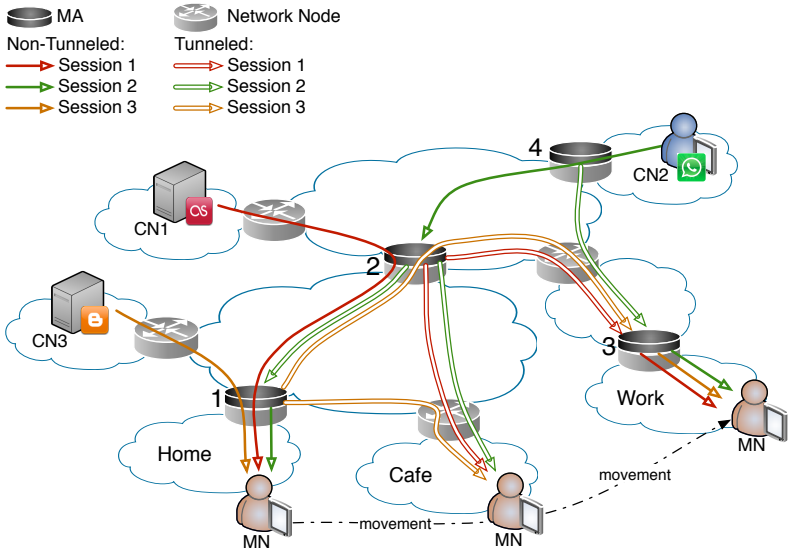


Figure F.4: An example of mobility anchoring with context-aware adaptive mobility anchoring model

Home:

The MN receives a Router Advertisement (RA) or a Dynamic Host Configuration Protocol (DHCP) Reply message through the Home router, which provides a normal IPv6 address without mobility support, called IPh address. It also provides Mobility IP addresses (MoA), which ensure IP session continuity and IP reachability, such as the MoA1 and the MoA2, from MA1 and MA2 respectively. From the metrics received in the message (higher metrics for IPv6 addresses from closer MAs), the MN is able to distinguish between the closest MA (MA1), usually placed in the AR, and far MAs (MA2), usually placed in GWs.

The MN configures the three IPv6 addresses (IPh, MoA1, MoA2) and adds MA1 and MA2 to the set of available MAs. The IPh is used for sessions that do not require IP session continuity nor IP reachability support, while MoA1 and MoA2 may be used otherwise.

Based on the user profile (e.g. a high IP handover rate), the MN selects the MoA2 from the centralized MA2 (always an IP address provided by a MA) to ensure its IP reachability, in new sessions initiated by other devices of the Internet. The association of the MN in the global location systems (e.g. DNS) is updated with the MoA2, if a previous IP address was previously designated. As long as the association with the previous IP address remains valid in the global location system, the respective MA is updated with the current location of the MN.

The MN sends BU messages to the added MA1 and MA2 to register the MN in the MAs. The MAs provide DAD defense to MoA1 and MoA2; thus, these MoAs shall not be allocated to any other MN, while the MN needs them. MA1 and MA2 reply to the MN

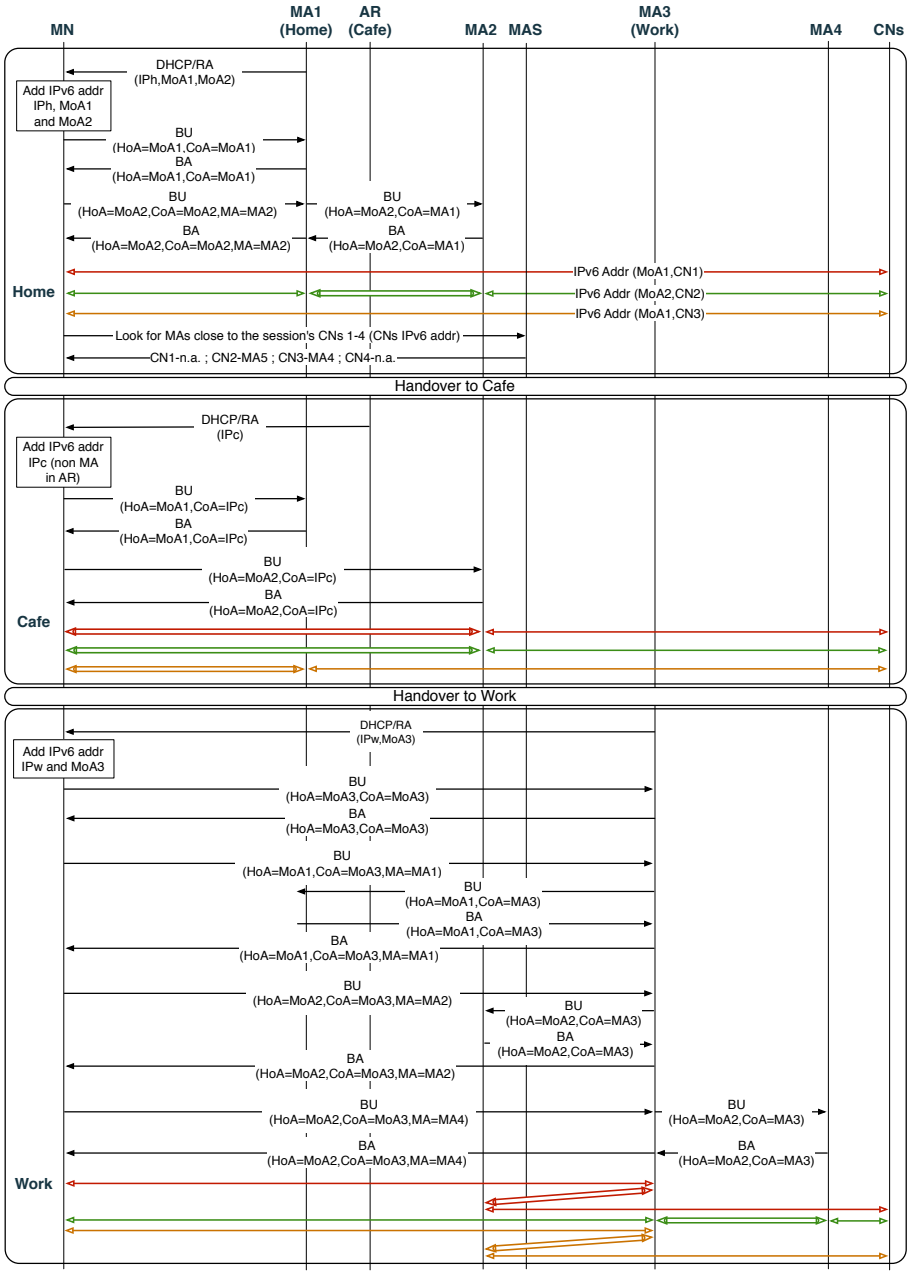


Figure F.5: An operation example of context-aware adaptive mobility anchoring

with BA messages to confirm the registration success. The MA1 is used as an intermediary in the BU/BA messages exchanged between the MN and MA2, since the tunnel for the sessions established with MoA2 is configured between the MA1 and MA2.

Sessions 1 and 3, initiated by the MN, are established without tunneling support in the optimized routing path between the CN and the MN, being selected the MoA1 from the MN side. The MN always selects the MoA provided by the closest MA in the establishment of new sessions, if it is available, since it is the one that ensures the optimized routing path with lower tunneling cost. The session 2 initiated by the CN is established with the MoA2 of the MN, since it is the IPv6 address provided by the global location system when the CN looks for the MN. Thus, the packets of session 2 are tunneled from the beginning between MA1 and MA2.

When the sessions are initiated, the MN searches for MAs close to the current CNs of the MN (CN1, CN2, CN3) in the MAS, which can be achieved using the DNS. The MAS replies with the MAs close to the CNs, if they have any. While CN2 is close to MA4, CN1 and CN3 do not have any close MA. The set of MAs close to the current CNs is maintained by the MN, which can be later used to optimize the routing path and/or minimize the tunneling of ongoing sessions.

Cafe:

After the handover of the MN to the Cafe network, the MN receives a RA or a DHCP Reply message, which provides a normal IPv6 address without mobility support, the IPc address. It is not provided any new MoA nor any indication of the MA from the Cafe router, since there is no MA deployed there. The MN remains with the same set of MAs and MoAs to establish/maintain sessions with mobility support.

In order to provide session continuity to the ongoing sessions and to assure the reachability for new sessions, the MN must send BU messages to the MAs with anchored sessions. Hence, the MN sends a BU to the MA2 to update the IP reachability binding and to configure a tunnel between MA2 and the MN (IPc), in order to maintain the ongoing session 3. The MN also sends a BU message to the MA1 in order to maintain the ongoing sessions 1 and 3, through a tunnel between the MN and MA1. These two BU messages are enough to maintain the ongoing sessions and to ensure IP reachability for new sessions, since the packets to MoA1 and MoA2 are forwarded to the IPc of the MN. However, the routing path of session 1 can be optimized with the help of the MA2. Thus, the MN sends a BU message to MA2 to create a binding there (MoA2-IPc), while the tunnel can be the same already configured for IP reachability. The MA2 has two rules to forward all packets with IP destination MoA2 and MoA1 to IPh. All MAs reply to the BU with a BA message to ensure the reception of the updated bindings in the binding cache.

New sessions requiring IP continuity support are initiated through MoA1 with the tunnel between the MA1 and the MN, while sessions requiring IP reachability are initiated using MoA2 through the tunnel between the MA2 and the MN.

Work:

After the handover of the MN to the Work network, connected to a different operator, the MN receives a RA or a DHCP Reply message, which provides a normal IPv6 address without mobility support from the Work network, the IPw address. It is also provided the new MoA3 from the MA3 installed in the router of the Work network.

The MN assigns the new IPv6 addresses MoA3 and IPw, which are added it to the existent ones (MoA1 and MoA2); it also adds the MA3 to the set of MAs (MA1 and MA2). The IPw is used for sessions that do not require IP session continuity nor IP reachability, while MoA2 and MoA3 are used for sessions requiring IP reachability or IP session continuity, respectively. The MN maintains the MoA2 to ensure the reachability of the MN in new sessions initiated by other devices of the Internet.

The MN sends BU messages to the added MA3 to register the MN in the MA. The MA performs DAD defense to MoA3; thus, this MoA shall not be allocated to any other device until the MN needs it. The MA3 replies to the MN with BA messages to confirm the registration success. The closer MA3 will be used as an intermediary of the MN to establish tunnels with other MAs of the MN.

In order to provide IP continuity to the ongoing sessions and to maintain the IP reachability for new sessions, the MN sends BU messages to the MAs with anchored sessions. Hence, the MN sends a BU to the MA2 through MA3 to update the IP reachability bindings (MoA2-MA3 and MoA1-MA3), configuring a tunnel between MA2 and MA3 to maintain the ongoing sessions 1 and 2. The MN also sends a BU message to MA1 through

Table F.1: Parameters changed in each scenario

Scenario	Max value of Pause and Walk Intervals (s)	Avg Sessions Arrival Interval (s)	Avg Session Duration (s)	Connection Probability (ANs, ARs and eNBs)
A	200, 400, 800, 1600	150	300	0.5
B	600	150	100, 200, 400, 800	0.5
C	600	150	300	0, 0.5, 1
D	600	150	300	0.5

MA3 in order to maintain the ongoing session 3, through a tunnel between the MA3 and MA1. These two BU messages are enough to maintain the ongoing sessions and ensure IP reachability, since packets to MoA1 and MoA2 are forwarded to the MA3 and then to the MN. However, the routing path of session 2 can be optimized with the help of MA4, closer to the CN3. Thus, the MN sends a BU message to MA4 through MA3, in order to create a binding there (MoA2-MA3) and to configure a tunnel between MA4 and MA3, which forwards all packets with IP destination MoA2 to MA3. From now on, MA4 is included in the set of MAs of the MN. All MAs reply to the BUs with BA messages to ensure the reception of the updated bindings in the binding caches.

New sessions requiring IP continuity support are initiated through the MoA3 without tunneling, while sessions requiring IP reachability are initiated using MoA2 through the tunnel between MA3 and MA2.

When there are not more sessions using an MA (which does not provide IP reachability), the MN sends a BU message with the deregistration option to this MA. For instance, if all three sessions are terminated while the MN is connected to the Work network, the following procedures are required:

Session 1: the deregistration of the binding MoA1-MA3 in MA2, as well as the tunnel between MA2 and MA3, is achieved through a BU message between MA3 and MA2, after the MN sends this message to MA3.

Session 2: the deregistration of the binding MoA1-MA3 in MA4, as well as the tunnel between MA4 and MA3, is achieved through a BU message between MA3 and MA4, after the MN sends this message to MA3. The session 3 may also require the deregistration of the binding MoA2-MA3, as well as the tunnel between MA3 and MA2. However, this binding and tunnel cannot be removed, since they are needed to provide IP reachability to the MN for new sessions.

Session 3: the deregistration of the binding MoA1-MA3 in MA1, as well as the tunnel between MA3 and MA1, is achieved through a BU message between MA3 and MA1, after the MN sends this message to the MA3.

F.4 Evaluation

We compare the proposed mobility anchoring with centralized, AR and CN anchoring models, through MATLAB [13], where the simulation time is 1 hour and the simulation step is 100ms, which is a sufficient value to accurately integrate user movement and ongoing sessions in the evaluation of the data performance. In the wired topology, we randomly generate flat network topologies, where Figure F.6 illustrates an example. We assume a more flatten network topology, which is a mixture between a hierarchical network and a flat network. A node connects to the higher hierarchy level node, to lower hierarchy level

nodes, and to nodes in the same hierarchy level. The wired links are defined with the following hops and connection probabilities, where a wired hop has a 1 ms delay:

- **Among ENs:** 16 hops and a connection probability of 1.
- **EN-AN:** 8 hops and a connection probability of 1.
- **AN-AR and AN-eNB:** 4 hops and a connection probability of 1.
- **AN-AN:** 4 hops and a variable connection probability (Table F.1).
- **AR-AR and eNB-eNB:** 2 hops and a variable connection probability (Table F.1).

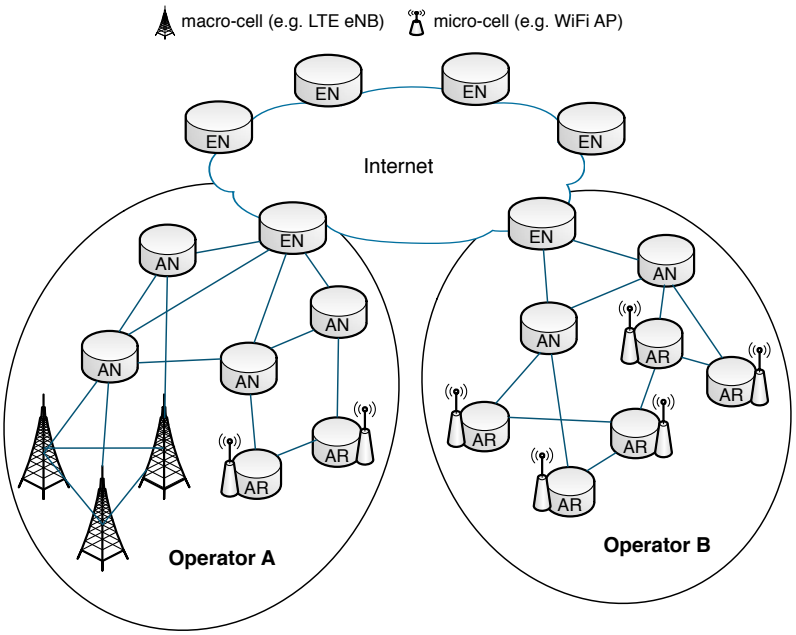


Figure F.6: Example of a wired topology.

In the wireless part of the simulated scenario, illustrated in Figure F.7, we define two types of cells: macro-cells and micro-cells. The macro-cells mimic the cells currently associated to cellular access networks (e.g. Node from UMTS and eNB from LTE), while micro-cells represent the cells from the WiFi access networks (e.g WARs from 802.11b/g). We distribute 16 macro-cells in a 4x4 grid to ensure the full coverage in a square area of 5Km x 5Km where the MNs move. We also place 48 micro-cells with a uniform distribution in the square area (5Km x 5Km) where the MNs move. The 48 macro-cells are randomly connected to ANs of the operators A and B. The wireless scenario is just used to know the current cells where the MN is connected, in order to obtain the path followed by the ongoing sessions inside the network, according to the offloading strategy.

MNs move inside the coverage area of the set of macro-cells (5x5 $\boldsymbol{km^2}$) with a Random Way Point mobility model with the following parameters:

Speed: uniformly distributed between 1 and 10 m/s.

Pause Interval: uniformly distributed between 1 min and a variable value (Table F.1).

Walk Interval: uniformly distributed between 1 min and a variable value (Table F.1).

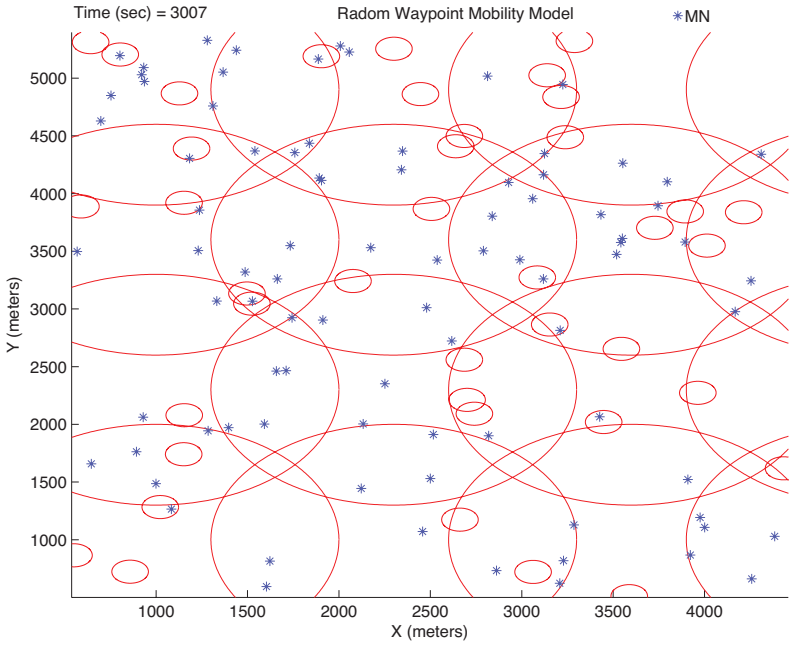


Figure F.7: Snapshot from an example of a wireless scenario.

We assume that CNs can be attached to any node of the network; thus, sessions can be established with any network node, which is randomly selected. We assume that all sessions evaluated are initiated by the MN; thus, sessions just require IP continuity support. The sessions are established between the CNs (network nodes) and the MN itself, with the following characteristics:

Sessions arrival interval: exponentially distributed with a variable average (Table F.1).

Session duration: exponentially distributed with a variable average (Table F.1).

Data packet rate: uniformly distributed between 128 Kbps and 2 Mbps.

Data packet size: uniformly distributed between 128 bytes and 1 Kbyte.

In the evaluation, the simulations are repeated 10 times to improve the accuracy of the results with a confidence interval of 95%. The evaluated metrics are defined as follows:

Data Delay: the average time that a data packet experiences to be transmitted from a CN to an MN.

Tunneled Packets: the ratio of tunneled data packets from the total data packets transmitted between CNs and MNs.

Tunnel Length: the average tunneled hops per data packet that enters in a tunnel.

Data Cost: the total cost needed to deliver data packets from the CNs to the MNs. It is calculated through the sum of the data packets cost, where each data packet cost is the multiplication of the data packet size, including IPv6 encapsulation if it exists, by the time to traverse the network.

Our purpose is to evaluate the context-aware adaptive mobility anchoring in heterogeneous networks. Thus, we define a simple offloading strategy that performs the offload of all ongoing and new sessions through the micro-cells of both operators A and B, when the MN is in the coverage area of a micro-cell and its speed is null. Otherwise, all sessions will be maintained or initiated through the attached macro-cell.

F.4.1 Pause and Walk Intervals

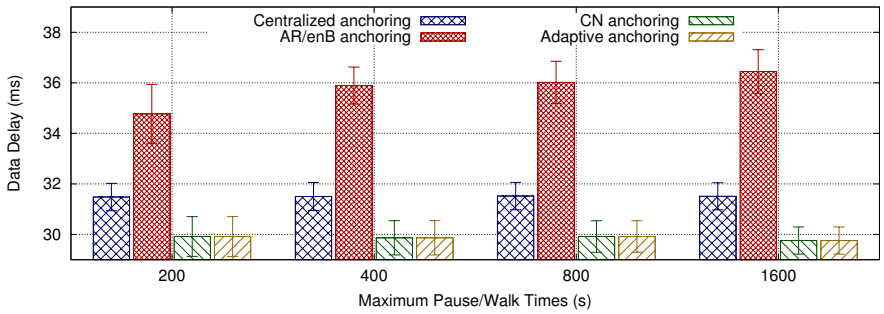


Figure F.8: Data Delay changing the maximum pause and walk times.

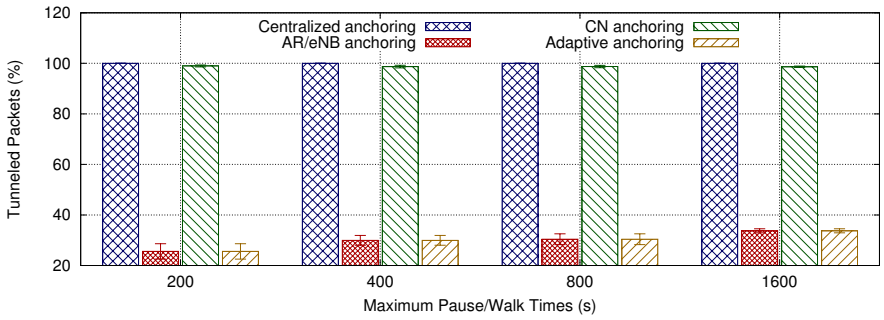


Figure F.9: Tunneled Packets changing the maximum pause and walk times.

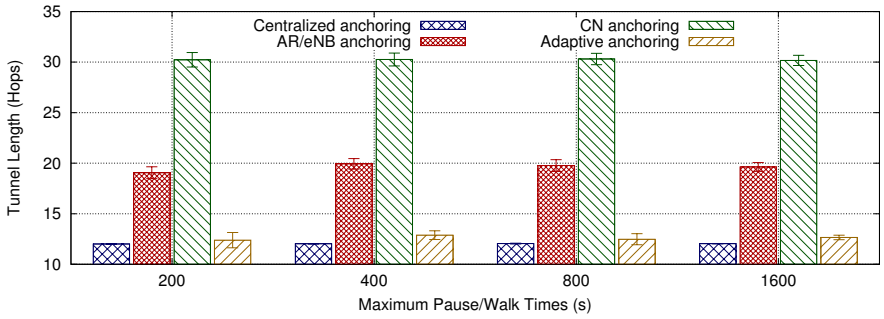


Figure F.10: Tunnels Length changing the maximum pause and walk times.

The first test evaluates the influence of the pause and walk intervals of the random waypoint mobility model, in which the variable values are defined according to the Scenario A of Table F.1. The results are shown in Figures F.8, F.9, F.10 and F.11, which evaluate the data delay, tunneled packets, the tunneled length and the data cost, respectively.

The centralized anchoring provides the shortest tunnels, but it tunnels all packets. The centralized anchoring uses a centralized node for the MA, which is not optimized to offload sessions through the WiFi of operator B. Hence, it has a lower data delay than the AR anchoring, but a higher data delay than the CN and adaptive anchoring. The data cost is able to integrate the impact of the other three metrics in one metric; thus, centralized anchoring has a similar data cost than the one of CN anchoring, which is lower than the

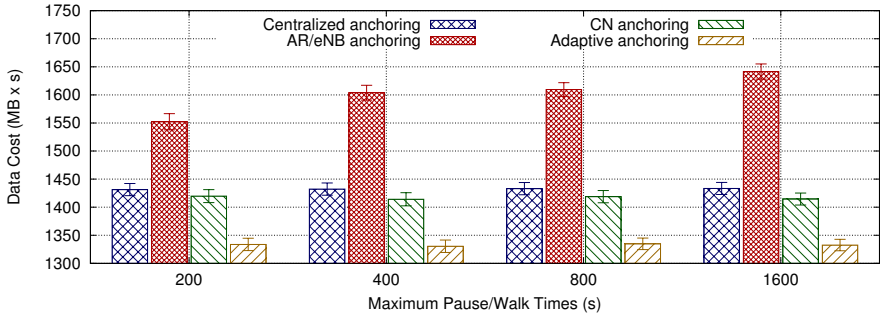


Figure F.11: Data Cost changing the maximum pause and walk times.

AN anchoring and higher than the adaptive anchoring. The pause and walk times do not have impact on the centralized anchoring, since the sessions are anchored in a centralized node in the middle of the network. This centralized node has a similar distance to the ARs/eNBs of the network; thus, the MN attachment to different ARs/eNBs does not affect the centralized anchoring performance.

The AR anchoring provides the optimized routing path for sessions initiated and terminated in the same AR/eNB, but it introduces the longest tunnels and a higher data delay, due to the offload of sessions through the WiFi of operator B. Although it provides long tunnels, the AR anchoring provides the lower number of tunneled packets. The AR anchoring has a higher data cost than the other approaches, which measures the impact of the other three metrics at once. The pause and walk times have impact on the AR anchoring, since MNs with longer walk intervals lead to more sessions not being initiated and terminated in the same AR/eNB; thus, more packets are tunneled in the non-optimized routing path.

The CN anchoring provides the lowest data delay at the cost of the longest tunnels between the CN and the MN. The CN anchoring tunnels almost all packets, except when the CN and MN are connected to the same AR/eNB, which just represents approximately 1% of the total packets in the evaluated scenarios. The data cost of the CN anchoring, which merges the other three metrics in one, is similar to the data cost of the centralized anchoring, lower than the AN anchoring and higher than the adaptive anchoring. The pause and walk times do not have impact on the CN anchoring, since the sessions are anchored close to the CN, not depending on the MN mobility.

The adaptive anchoring dynamically exploits the benefits of the other three approaches, since it provides the selection of the best MA when the context of the user, the session or the network changes. During a session's lifetime, the MA that is used for routing/forwarding can be changed from the AR/eNB until the CN. The adaptive anchoring provides the lowest data delay (as the CN anchoring), the shortest tunnel length (as the centralized anchoring), and the minimum tunneled packets (as the AR anchoring). Hence, the adaptive anchoring has the lowest data cost, which weightly merges the other metrics into a single one.

F.4.2 Session Duration

In this test, we evaluate the influence of the session duration, in which the variable values are defined according to the Scenario B of Table F.1. The results are shown in Figures F.12, F.13 and F.14, which evaluate the data delay, tunneled packets and the data cost, respectively. From the evaluation, the tunnel length is not significantly affected by

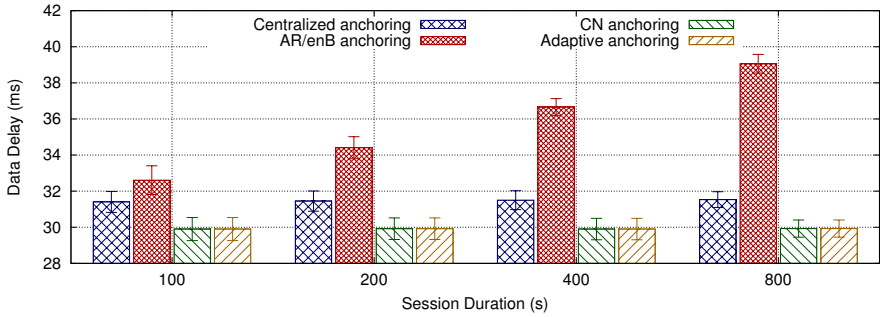


Figure F.12: Data Delay changing the session duration.

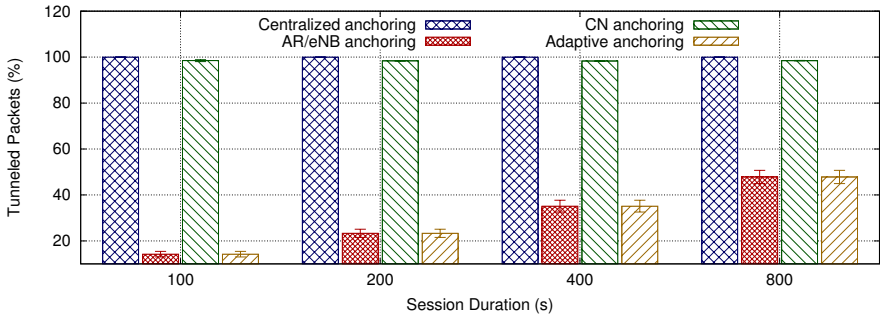


Figure F.13: Tunneled Packets changing the session duration.

the change of session duration in any anchoring scheme, thus, data is not shown.

The session duration has no significant impact on the centralized and CN anchoring. In the centralized anchoring, packets are always tunneled from the centralized anchor to the current MN location, while in the CN anchoring, the packets are always tunneled from the CN to the MN location. The data cost of the centralized anchoring is only increased due to the increase of data packets being received by the MNs.

The increase of the session duration in the AR anchoring leads to more packets being tunneled between ARs of different operators through highly long paths. Thus, higher session duration in the AR anchoring increases the tunneled packets and the data delay. Since the data cost reflects the weighted influence of the other three metrics, the data cost of AR anchoring is the most affected by the increase of the session duration when compared with other approaches.

The tunneled packets of the adaptive anchoring increases with a larger session duration, since more sessions are not able to be initiated and terminated in the same AR/eNB. The data cost of the adaptive anchoring is mainly increased by the data packets being received by the MNs, but it is also affected by the increase of tunneled packets. The adaptive anchoring continues to provide the lowest data delay, the shortest tunnel length, the minimum tunneled packets and the lowest data cost, independently of the session duration.

F.4.3 Connection Probability

In this test, we change the probability of connection between the network nodes in the same hierarchical level, such as between ANs, between ARs or between eNBs. A lower connection probability represents a more hierarchical network, while a higher connection probability leads to a flatten network. We define the variable values according to the Sce-

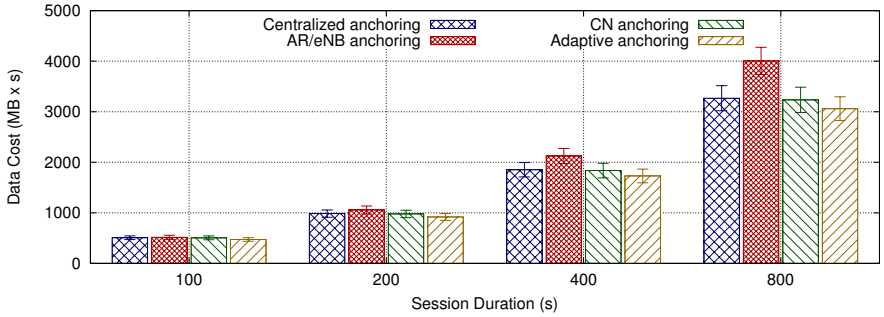


Figure F.14: Data Cost changing the session duration.

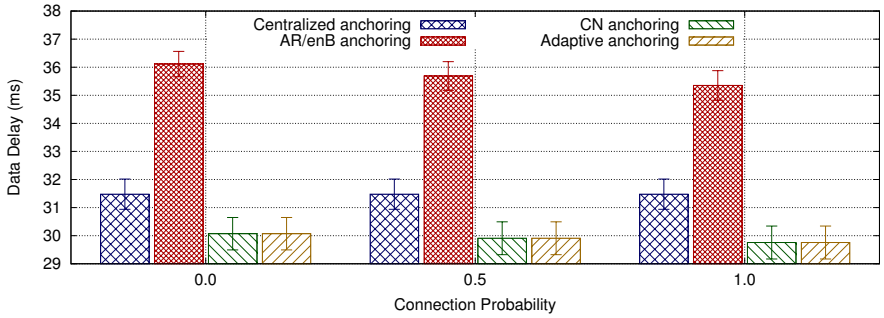


Figure F.15: Data Delay changing the connection probability.

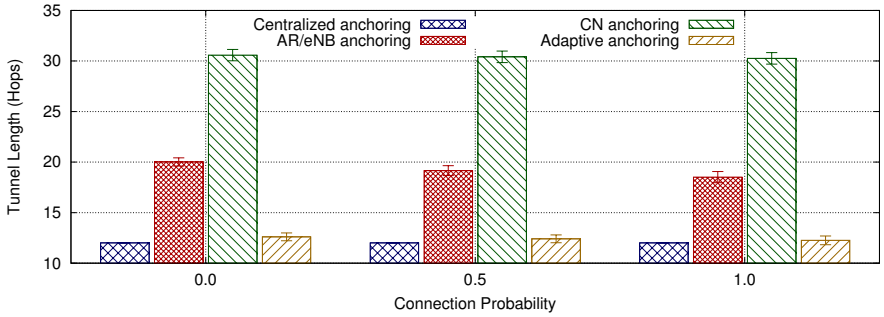


Figure F.16: Tunnel Length changing the connection probability.

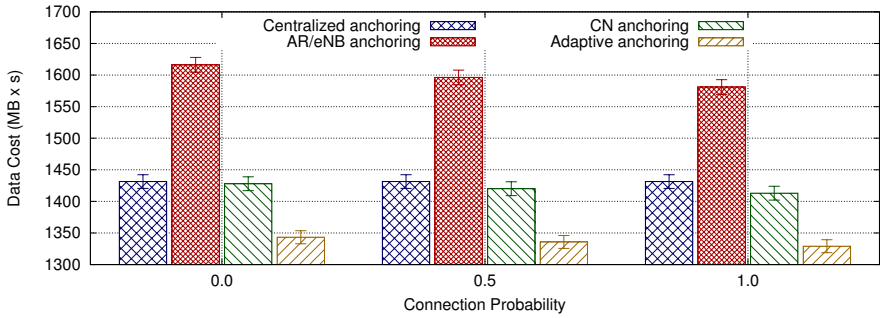


Figure F.17: Data Cost changing the connection probability.

nario C of Table F.1. The results are shown in Figures F.15, F.16 and F.17, which evaluate the data delay, the tunnel length and the data cost, respectively. From the evaluation, the

tunneled packets are not significantly affected by the change of the connection probability in any anchoring scheme, thus, data is not shown.

The connection probability has no significant impact on the centralized anchoring, since the packets are always tunneled from the centralized anchor to the current MN location, which does not exploit the connections between nodes in the same hierarchical level.

The increase of the connection probability in the AR anchoring increases the probability to have shorter tunnels between ARs/eNBs for ongoing sessions, after a handover or a session offloading decision. Thus, a higher connection probability in the AR anchoring decreases the data delay and the tunnel length, and consequently reduces also the data cost. Although there is a decrease of the data delay and data cost for higher connection probabilities, the AR anchoring remains with the highest value when compared with other anchoring approaches.

The increase of the connection probability is able to reduce the data delay of the CN anchoring, since there are shorter paths between the CNs and MNs to be exploited by the end-to-end tunnel, when the CN and the MN are in the same operator network. The adaptive anchoring is also able to exploit the shorter paths between CNs and MNs to reduce the data delay. The data cost of the adaptive anchoring is reduced, and it continues to provide the lowest data delay, the shortest tunnel length, the minimum tunneled packets and the lowest data cost, independently of the connection probability.

F.4.4 Offloading Strategy and CNs Location

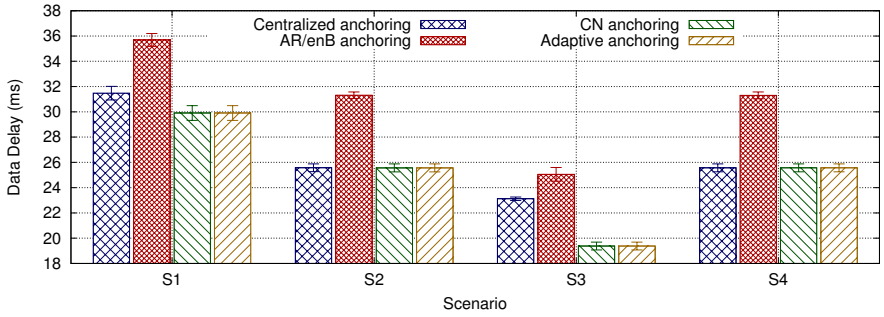


Figure F.18: Data Delay changing the offloading strategy and the CNs location.

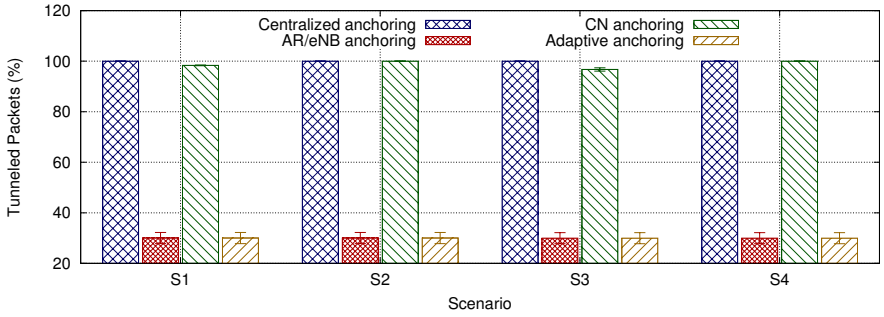


Figure F.19: Tunneled Packets changing the offloading strategy and the CNs location.

In this test we define four scenarios with variations of the offloading strategy and the CNs location. The variable values are defined according to the Scenario D of Table F.1.

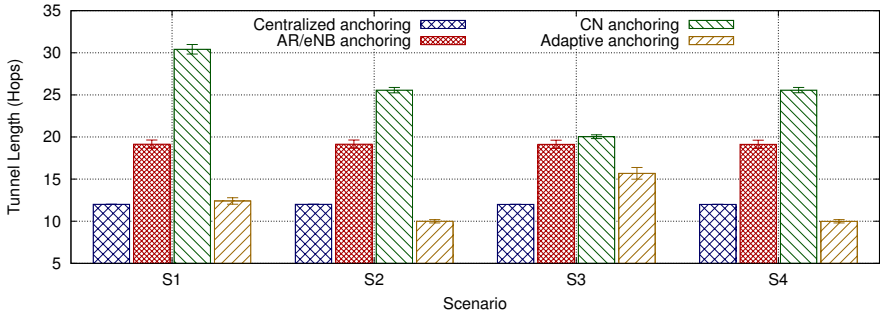


Figure F.20: Tunnel Length changing the offloading strategy and the CNs location.

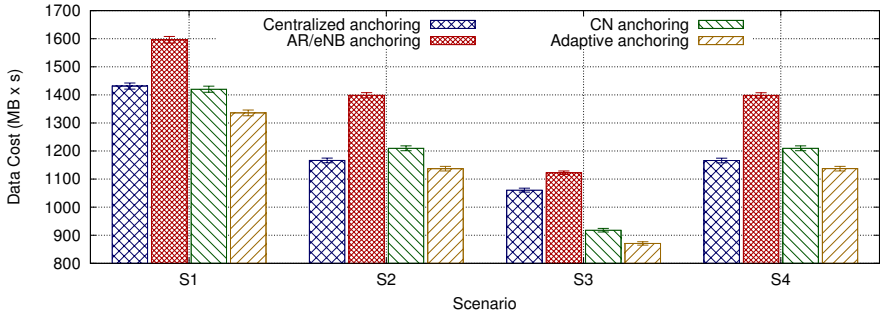


Figure F.21: Data Cost changing the offloading strategy and the CNs location.

Scenario S1: Offload of all ongoing and new sessions through the micro-cells of operators A and B, when the MN is in the coverage area of these micro-cells and its speed is null. CNs are uniformly distributed all over the network; thus, sessions can come from any node of the network.

Scenario S2: Offload of all ongoing and new sessions through the micro-cells of operators A and B, when the MN is in the coverage area of these micro-cells and its speed is null. CNs are uniformly distributed through the ENs.

Scenario S3: Offload of all ongoing and new sessions through the micro-cells of operator A, when the MN is in the coverage area of these micro-cells and its speed is null. CNs are uniformly distributed through the network nodes of operator A and ENs of the Internet.

Scenario S4: Offload of all ongoing and new sessions through the micro-cells of operator A, when the MN is in the coverage area of these micro-cells and its speed is null. CNs are uniformly distributed through the ENs of the Internet.

The results of the different scenarios for the anchoring approaches are shown in Figures F.18, F.19, F.20 and F.21, which evaluate the data delay, tunneled packets, the tunneled length and the data cost, respectively.

The centralized anchoring provides shorter tunnels than AR anchoring and CN anchoring, but it tunnels all packets. The sessions are anchored in a centralized node in the middle of the network, which has a similar distance to the ARs/eNBs of the network. Thus, it does not provide so long routes to offload sessions through the WiFi of operator A and/or B. It has a lower data delay than AR anchoring, but a higher data delay than CN and adaptive anchoring.

The AR anchoring provides the optimized routing path for sessions initiated and terminated in the same AR/eNB, but it introduces long tunnels and the highest data delay,

due to the offload of sessions through the WiFi from operator A and/or B. Although it introduces long tunnels, the AR anchoring provides the lower number of tunneled packets, independently of the evaluated scenario. However, in the considered four scenarios, the AR anchoring has the highest data cost.

The CN anchoring provides the lowest data delay at the cost of the longest tunnels between the CN and the MN, independently of the scenario evaluated. The CN anchoring tunnels all packets in scenarios S2 and S4, where all packets come from ENs, while in scenarios S1 and S3, a small portion of the total packets (less than 4%) are not tunneled, when the CN and the MN are connected to the same AR/eNB. The CN anchoring shows a data cost similar to the data cost of centralized anchoring, except for scenario S3, where the CN anchoring has a lower data cost than centralized anchoring, since the MN just connects to operator A, in which the CNs are distributed.

The adaptive anchoring dynamically adapts to the evaluated scenarios, since it provides the best MA when the context of the user, the session or the network changes. Overall, the adaptive anchoring provides the lowest data delay (as the CN anchoring), the shortest tunnel length (as the centralized anchoring), and the minimum tunneled packets (as the AR anchoring). There is an exception in scenario S3, where the adaptive anchoring has a higher tunnel length than centralized anchoring, but this is necessary to reduce the data delay and to provide the lowest data cost.

F.5 Conclusion

Distributed mobility management has emerged as a promising paradigm to accommodate the exponential growth of mobile traffic, where part of the mobility functionalities are distributed through the network nodes. Despite of the advances in the distributed mobility management, none of the current approaches is suitable for all scenarios, since the mobility handling depends on the user, the sessions and the network. Thus, we propose a context-aware adaptive IP mobility anchoring approach driven by the context of user, sessions and network. It provides a dynamic initial anchoring and the adaptive mobility anchoring regarding routing/forwarding for sessions that require IP reachability and/or IP session continuity, based on the creation/update of the bindings in the distributed MAs. The outcomes of the approach demonstrate that the proposed approach optimizes the routing path and minimizes the tunneling, when compared with other approaches, reducing the data cost, the data delay, tunneled packets and the tunnel length. The context-aware adaptive IP mobility anchoring exploits the benefits of the other anchoring schemes by properly adapting to different scenarios.

F.6 Acknowledgments

The authors would like to acknowledge the support of this work by Fundação para a Ciência e a Tecnologia (FCT) with the MC-WMNS project (PTDC/EEA-TEL/120176/2010) and the PhD grant of T. Condeixa (SFRH/BD/65265/2009).

F.7 References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," Cisco, White Paper, Feb. 2013.

[2] Ericsson, "Ericsson mobility report: On the pulse of the networked society," Ericsson, Report, Jun. 2013.

[3] C. Perkins et al., "Mobility support in ipv6," IETF, RFC 6275, Jul. 2011.

[4] S. Gundavelli et al., "Proxy mobile ipv6," IETF, RFC 5213, Aug. 2008.

[5] I. WG. (2013, Aug) Distributed mobility management. [Online]. Available: http://datatracker.ietf.org/wg/dmm

[6] H. Chan et al., "Requirements for Distributed Mobility Management," IETF, Internet-Draft draft-ietf-dmm-requirements-07, Aug. 2013, work in progress.

[7] P. Seite, P. Bertin, and J. Lee, "Distributed mobility anchoring," IETF, Internet-Draft draft-seite-dmm-dma-06.txt, Jan. 2013, work in progress.

[8] T. Condeixa and S. Sargento, "Dynamic mobile ip anchoring," in *IEEE ICC*, Jun. 2013.

[9] C. Bernardos and J. Zuniga, "Pmipv6-based distributed anchoring," IETF, Internet-Draft draft-bernardos-dmm-distributed-anchoring-02, Apr. 2013, work in progress.

[10] A. Yegin et al., "Corresponding network homing," IETF, Internet-Draft draft-yegin-dmm-cnet-homing-00, Jul. 2013, work in progress.

[11] ——, "On demand mobility management," IETF, Internet-Draft draft-yegin-dmm-ondemand-mobility-00, Jul. 2013, work in progress.

[12] H. Ali-Ahmad et al., "Mobility anchor selection in dmm: Use-case scenarios," IETF, Internet-Draft draft-aliahmad-dmm-anchor-selection-01, Jul. 2013, work in progress.

[13] T. M. Inc. (2013, Nov.) Matlab version 7.10.0 (r2010a). [Online]. Available: http://www.mathworks.com/