ADVANCED SCHEMES FOR EMERGING MOBILITY SCENARIOS IN THE ALL-IP WORLD

Ph.D. Dissertation summary
of
László Bokor

Supervisors:
Sándor Imre Sc.D.
Gábor Jeney Ph.D.

BUDAPEST, 2014
1. Introduction

Telecommunication systems are converging into a synergistic union of different wired and wireless technologies, where integrated, multimedia services are provided on a universal Internet Protocol (IP) based infrastructure [J1], [C7]. The Internet itself is turning into a fully pervasive and ubiquitous multimedia communication system in which users are expected to use remote resources anytime and anywhere [C8]. This evolution recently made mobile Internet a reality for both users and operators thanks to the success of novel smartphones, portable computers with 3G/4G USB modems and attractive business models. Based on actual trends, vendors prognosticate that mobile networks will suffer an immense traffic explosion in the packet switched domain up to year 2020 [1]. In order to accommodate current systems to the anticipated traffic demands and user requirements, technologies applied in the access, backhaul and core networks must become appropriate to advanced use cases and scenarios. Within these technologies, mobility management protocols and schemes play an essential role when it comes to future mobile Internet architectures [J9].

Legacy IP mobility management solutions like Mobile IPv6 [2] provide transparent session continuity and global handover management for heterogeneous all-IP mobile architectures but could suffer from several well known problems (increased delay, packet loss, and signaling) that have led to the distinction of macro- and micromobility scenarios. Macromobility focuses on mobility management between distant wireless domains and across the Internet [2], [C16], [J4], while protocols designed for micromobility scenarios (e.g., HMIPv6 [3]) reduce the number of network elements that process the signaling information by managing movement inside a specific wireless domain locally. Due to their performance and scalability during handovers within localized areas, optimization, development and integration of micromobility schemes are research topics that live their renaissance nowadays. The optimal design of micromobility domains is also an open issue when deploying these protocols in next generation mobile environments.

Trends clearly show that IP-based mobile and wireless networks will not only support mobility for the widest range of single end terminals, but even for Personal Area Networks (PANs), Vehicle Area Networks (VANs), complex groups of nodes in Intelligent Transportation Systems (ITSs) and Cooperative ITS (C-ITS) architectures [C5], [C10], [C15]. It means that not only single mobile entities with permanent Internet connectivity have to be managed, but also entire mobile networks (i.e., NEMOs) need to be maintained as a whole. The currently standardized NEMO protocol [4] only offers basic solution for this complex problem, thus leaving space for researches on further enhancement and optimization.

The growing number of mobile users and the complexity of emerging mobility scenarios require architectures able to handle the foreseen traffic explosion and assure end-to-end Quality of Service. However, the strongly centralized nature of current and planned mobile Internet standards by the IETF or 3GPP prevents cost effective system scaling for the novel demands. Micromobility protocols try to ease the above issues but, doesn’t find the root of the problem. Aiming to solve the burning questions of scalability from an architectural point of view, distributed [J11] and flat [5] mobile architectures with enhanced, proactive and cross-layer optimized techniques (e.g., [C23], [C30]) are gaining more and more attention today.

However IPv6 shows word-wide proliferation and will play an essential role in the future, it is also anticipated that IP addresses will not continue to remain both locators and identifiers: the semantically overloaded nature of the IP will be obviated by identifierlocator (ID/Loc) separation schemes [6]. The Host Identity Protocol (HIP) family [7]–[10] is one of the most promising ID/Loc separation techniques, which guided me to develop both HIP and pure IPv6 based solutions for the identified problems of emerging mobility scenarios.
2. Research Objectives

The above introduced trends and use-cases pose serious challenges to existing mobile Internet architectures and require special support to efficiently cope with the raised problems and questions. My essential aim was to develop advanced protocols and schemes supporting these emerging mobility scenarios of the all-IP world. By investigating new mobility management techniques, localized mobility solutions, micromobility domain planning algorithms and proactive, cross-layer optimized handover mechanisms, I could also ensure scalability, seamless handover, enhanced network design, and eventually better Quality of Service (QoS), Quality of Experience (QoE) and increased user privacy. Regarding to the previously summarized broad research areas I have grouped my researches into four main topics:

1. In order to enhance macromobility management solutions by increasing their handover performance and scalability, I have followed two separate approaches. On the one hand I was induced to investigate possibilities to enhance the Internet Protocol and design a novel micromobility extension for Mobile IPv6 (Thesis I.1 and I.2). Aiming at a transparent and distributed support of micromobility scenarios my goal was to propose a purely IPv6-based, and transparent micromobility framework, which doesn’t require additional network entities, provides highly decentralized operation, and ensures optimal routes inside the domains without introducing extra signaling load on the wireless interface. In order to support deployment by keeping the scalability and efficiently controlling the size of the micromobility routing domain in the network design phase, the development of a special subnet optimization algorithm for my framework was also an objective within this approach. On the other hand I have decided to exploit a candidate future Internet scheme built upon IP called the Host Identity Protocol, by designing and evaluating a novel HIP-based micromobility protocol (Thesis I.3) naturally relying on the advanced, cryptographic ID/Loc separation scheme of HIP.

2. As mobility becomes one of the most unique characteristics of future’s convergent architectures, more attention must be paid to the problems of location information leakage (i.e., location privacy issues of all-IP mobile communication caused by easy estimation possibilities from IP addresses to precise geographical positions of users), even at the earliest phases of design: at the network planning level. This motivated me to develop mobile network planning tools and algorithms that exploit inherent location privacy support of micromobility protocols (Thesis II.1, II.2, II.3, and II.4). Existing network planning algorithms (e.g., [11]–[13]) are mainly focusing on the trade-off between the paging cost and the registration cost and – to the best of my knowledge – none have introduced privacy awareness in network planning methodologies before my work.

3. For network mobility scenarios several improvements exist to overcome the limitations of the already standardized NEMO Basic Support protocol [4]. NEMO BS operates in the IP layer and inherits the benefits of Mobile IPv6 [2] by extending the binding mechanism of the ancestor, but keeps all the problems of the main approach such as protocol overhead, inefficient routing, security and lack of multihoming support. All of these issues are under examination at the IETF, but this work has not been completed yet. However, there are several extensions of NEMO BS in order to allow multihoming [14], route optimization [15], security problems [16], and handover optimization [17]. Despite the fact that several novel real-life demonstrations [C10] and testbeds [C5] started to prove
the feasibility of NEMO BS and its extensions, the searching for further optimization possibilities and novel solutions like [18] has not stopped. In order to enhance current NEMO schemes, I have followed two approaches. On the one hand I was aiming at improving standard IPv6-based network mobility by forming a framework based on a special handover solution (Thesis III.1 and III.2) using cross-layer optimization and continuous network discovery. On the other hand my goal was to extend the Host Identity layer by developing and evaluating a novel, HIP-based NEMO protocol (Thesis III.3).

4. It is highly expected that due to their centralized design, mobile Internet architectures currently being under deployment or standardization will not scale particularly well to efficiently handle the challenges [19], [J9]. To enhance scalability of mobile Internet architectures and support distributed mobility management scenarios with decentralized, proactive, self-configuring and self-optimizing network structures, the Ultra Flat Architecture (UFA) was proposed as one of the first solutions [5]. The main characteristic of this proposal is that the execution of handovers is managed by the network via the Session Initiation Protocol (SIP). Even though SIP is a very powerful signaling solution for UFA, it is not applicable for non-SIP (i.e., legacy Internet) applications and the published SIP-based UFA scheme also does not comply with ITU-T’s recommendation of requirements for ID/Loc separation in future networks [6]. In order to overcome these issues, my research objective was to develop a Host Identity Protocol based system framework for the Ultra Flat Architecture (Thesis IV.1), and also to design and evaluate a proactive, distributed handover preparation and execution protocol for this framework, supporting complete elimination of centralized IP anchors between Point of Access (PoA) nodes and correspondent nodes, and placing network functions at the edge of the transit and access networks (Thesis IV.2 and IV.3).

3. Research Methodology

In my Thesis I have relied on two classical research approaches: analytical considerations and simulation studies. During the development phase of novel protocols, schemes or algorithms for the identified problems of emerging mobility scenarios, analytical considerations could not be ignored. My work on special network planning solutions in Thesis groups I and II is based on graph models, cost structures, and theory of algorithms (i.e., simulated annealing), while the analysis of my special NEMO optimization framework in Thesis group III relied on probability theory.

My proposed schemes were implemented in two different simulators. On the one hand I modified and extended an existing, proprietary Java-based mobility simulator [20], [J3], producing realistic cell boundary crossing (i.e., inter-cell movement rate) values and incoming call database in the particular (micro)mobility system under evaluation in Thesis group I and II. This simulator provided a realistic representation of the mobility patterns and was prepared to execute the different algorithm variants over an initial domain structure. On the other hand I have modified and extended an existing C++ model package for a general purpose open-source, component-based, discreet event simulation environment called OMNeT++ [21]. Thesis groups I, III and IV rely on the extensive evaluations performed with the help of my contributions to this powerful environment [C17].

I have strongly relied on statistics and probability theory also within my simulation analysis when handling large amount of measurement data came into picture.
4. New Results

4.1. Micromobility Management Protocols

Rapid evolution of wireless networking has provided wide-scale of different wireless access technologies with motivation of operators to integrate them in a supplementary and overlapping manner. To provide ubiquitous mobility between these technologies, Internet Protocol v4 and v6 emerged as the common technology platform [15], [B6]. Although macromobility management protocols like Mobile IPv6 [2] are capable of handling global mobility of users, they introduce low scalability, significant signaling overhead, and increased delay and packet loss when mobile terminals change their Internet point of attachment (PoA) frequently within geographically small areas (i.e., micromobility domains). In order to overcome these performance deficiencies, several micromobility approaches (e.g., [3]) attempt to offer faster and more seamless handover management while also enable more scalable operation and resource utilization. However these approaches usually suffer from lack of robustness, inefficient handling of intra-domain traffic and added complexity, furthermore they often require employing of new protocol stacks, and in general do not offer optimal performance in several scenarios. In order to enhance macromobility solutions by increasing their handover performance and scalability, I have followed a purely IPv6-based (Thesis I.1 and I.2) and a HIP-based approach (Thesis I.3).

In my IPv6-based proposal, the main goals were to relay on the characteristics and latest results of the IPv6 anycasting [22], and also to extend its possible application use-cases like in [J10]. In the proposed IPv6 mobility management framework the anycast addresses are identifying the mobile nodes (MNs) entering a micromobility domain. In the micromobility domains the registering and the membership management of the mobile anycast nodes is done by anycast group membership management protocols like [23]. The location- and handover management of mobile nodes within a given micromobility domain (i.e., intra-domain communication of a given anycast subnet) is based on the underlying anycast routing protocol (e.g., [24]). Inter-domain handovers are managed with the well-known Mobile IPv6 macromobility protocol.

Thesis I.1 [C1],[C2],[C3],[B1] I have proposed an anycast based micromobility framework (ABMF), which provides completely distributed, highly decentralized operation and optimal routes inside the micromobility domains without introducing extra signaling load on the wireless interface.

In ABMF, when a mobile node enters a micromobility domain, the Care-of-Address (CoA) obtained is a unique anycast address (aCoA), thus an anycast address identifies a single mobile node. Therefore the packets sent to the aCoA of the mobile terminal have no chance of reaching another mobile node, since in this sense the anycast addresses assigned to the mobile nodes are unique. The assigned anycast address has a validity area or region – an Anycast Subnet (AS) defined by the P prefix and the scope – where the anycast address might be located. As a result the mobile node in the validity area of the anycast address can move without being forced to change its anycast Care-of-Address. In my scheme the validity area determined by the length of the P prefix of the anycast address equals a micromobility domain. As a result the movements within the micromobility domain (i.e., anycast subnet) are handled locally decreasing the signalling overhead of MIPv6 as the corresponding macromobility protocol.

The mobile node after entering a micromobility domain and getting an aCoA becomes a member of a Virtual Anycast Group (VAG). The VAG size depends on the size of the
micromobility area (or anycast subnet) since the anycast address is valid in the whole micromobility domain. The members of the VAG are the virtual (possible) locations of the mobile node (Fig. 1). However the mobile node’s actual position is the only one that has a valid routing entry. The underlying anycast routing algorithms are supposed to find out the appropriate destination for a packet destined to a VAG member.

![Diagram showing micromobility area with anycast subnet]

**Figure 1:** Anycast-based Mobility Framework (left) and details of AOSPFv3 applicability for ABMF (right)

It is obvious that one of the most important questions regarding any anycast based application is the underlying routing protocol. In case of ABMF I have presented the applicability of both the Anycast Extension to OSPFv3 [C9] and the ARIP [C3]. However, the biggest concern when introducing anycast routing in ABMF (an in case of any hop-by-hop micromobility solution) is the large number of routing entries in the routing domain, since mobile nodes must be maintained as separate routing entries. In order to control the size of the routing domain, keep the scalability and help the design and formation of micromobility domains in ABMF, I have proposed a special subnet optimization algorithm also handling the tradeoff between the paging cost and the registration cost.

**Thesis I.2** [C9], [C12], [J3] I have developed a two-phase anycast subnet forming algorithm where firstly a greedy grouping is adopted to form a basic partition of wireless attachment points into anycast subnets (ASs), and then simulated annealing is applied to provide the final partitioning. I have shown that the proposed two-phase Simulated Annealing Based Anycast Subnet forming algorithm (SABAS), which is an improvement of the SABLAF scheme, reduces the registration cost by an average 35% compared to the reference forming scheme.

In ABMF, at each AS boundary crossing, the mobile nodes register their new locations through signalling messages of MIPv6 in order to update the location management database of the Home Agent. In this way the system is able to maintain the current location of each user, but this will produce a registration cost in the network. Therefore the question arises, what size the AS should be for reducing the cost of paging, maintaining routing tables (intra-domain handovers) and registration signalling (inter-domain handovers).

I qualified the paging cost together with the maximal routing table size as a constraint: therefore the registration cost was left alone in the objective function. Hence I defined and formulated a problem in which the final goal is the determination of optimum number of wireless Internet points of attachment per an anycast subnet for which the registration cost is minimal, with the limitations of the paging cost and the routing table sizes as an inequality constraint function. This problem is similar to the well-known Location Area planning problem [13], therefore I have applied the widely used fluid model for calculations about the movement of MNs among the ASs, relied on the results of [25] for the definition of the MIPv6 registration cost and the paging cost, and used the equation of [20] for the calculations of $N_{\text{max}}$ (the maximum possible number of cells in the AS) as a main input for my AS
Another input of SABAS is the boundary crossing database: a handover rate for each cell pair, defined on the border of these cells.

SABAS starts with a greedy solution, which will provide the basic AS partition as an input to the simulated annealing method. The algorithm chooses the cell pair with the biggest handover rate in the given structure of wireless Internet points of attachment \( q_{\text{max}} \) and includes the two PoAs into the \( AS_1 \) set of cells. In the next step, SABAS searches for the second biggest handover rate among the cell pairs for which is true, that one of them belongs to the \( AS_1 \) set of cells. The algorithm checks whether the inequality \( N_k < N_{\text{max}} \) is satisfied, where \( N_{\text{max}} \) is the maximized value of \( N_k \), namely the maximum number of cells in an AS which provides the minimum of the registration cost. If the inequality is satisfied, the cell can be included into \( AS_1 \) set of cells. After the processing of all cell pairs in the above sequential way, there will be cells that are not group of any set of cells. These cells will form another AS, which is not the best solution, but this will be only a basic AS partition which will serve as an input to the simulated annealing based SA forming scheme. The simulated annealing procedure starts with this basic partition, \( s_0 \). A neighbour to this solution \( s_1 \) is then generated as the next solution by simulated annealing, and the change in the registration cost \( \Delta C_{\text{Reg}}(s_0, s_1) \) is evaluated. The acceptance function is \( e^{-\frac{\Delta C_{\text{Reg}}}{T}} \), while the stopping rule is the maximal iteration step number or maximum number of steps when the \( \Delta C_{\text{Reg}} \) do not changes.

I have defined another constraint, the maximum number of MNs in one AS \( K_{\text{max}} \), considering the scalability challenges of the non-aggregatable anycast routing entries in a given anycast subnet. Therefore when the number of the routing entries reaches the \( K_{\text{max}} \) value in one AS (one routing entry for every MN), the value of the \( N_{\text{max}} \) need to be decreased, hence the ASs will consist of less number of cells in average, so the number of entries will be smaller in an AS proportionally. This decreasing should be continued until the number of routing entries goes under the \( K_{\text{max}} \) constraint.

A realistic mobile environment simulator capable of providing rural and urban mobile environments [20], [J3] was extended by me in order to generate the input metrics (cell boundaries crossing and incoming session statistics) and execute the algorithm. Then I have compared SABAS with a manual AS grouping solution where the partitions are made intuitively (this reference manual solution should be considered as a planned partition, but likely not the optimal one). I have examined how the registration cost changes by increasing the maximum number of cells in one AS.

![Figure 2: The registration cost in rural (left) and urban (right) environments](image-url)
As my results depicted in Fig. 2 shows, SABAS finds a much better solution for every value of $N_{\text{max}}$ both in rural and urban environments, and decreases the average registration cost by an average 35% compared to the reference algorithm.

The Internet Protocol was not designed with any kind of mobility in mind: the inseparable bond between the locator (Loc) and identifier (ID) functions of IP addresses makes it complicated, inconvenient or in some cases even impossible to design efficient, scalable and secure mobility and multihoming solutions. To get down to the roots of this problem by separating the dual role of IP addresses and provide an extended TCP/IP stack for future mobile Internet, the Host Identity Protocol (HIP) [7], [8] was designed. In this architecture transport level connections are not bound to IP addresses, which are dynamically changeable in several cases, but to permanent identifiers, which remain the same for quite a long time. This property provides sophisticated and secure mobility/multihoming support [9], [10] for standard macromobility scenarios, but further extension of the base protocol is needed for micromobility scenarios. The original idea of integrating micromobility with HIP was presented in [26] but their solution was not built on an effective and intact micromobility model as the focus was on the security issues, and the authors did not consider protocol details regarding the operation and the mobility support. Moreover, in their method MNs still need to update their location information at the RVS during the handover, therefore the scheme cannot fulfill the requirements of micromobility architecture: it is only a partial answer for the complex problem. This motivated me to develop an enhanced micromobility solution based on HIP, and by this way to highlight the emerging mobility applications of this promising ID/Loc separation protocol family.

Thesis I.3 [C4], [C11], [C17], [C21], [B3], [J14], [J20] I have developed a Host Identity Protocol based micromobility solution (µHIP) that makes HIP able to efficiently serve frequently moving mobile users while preserving all the advantages of the standard HIP protocol suite. I have also introduced a paging method fitting into the proposed µHIP architecture. I have shown by extensive simulations built on complex protocol models that my proposed µHIP scheme outperforms the standard HIP mobility management solution in micromobility environments by providing an average TCP performance gain of 20%, while introducing only a 9% decrease during the much less frequent macromobility scenarios.

In order to distribute HIP anchor nodes (Rendezvous Servers – RVSs [10]) and control micromobility domains in the µHIP architecture I have introduced a novel HIP gateway entity called the Local Rendezvous Server (LRVS) which is responsible for managing HIP Mobile Nodes (MNs) in a given domain (Fig. 3). LRVS gateways provide HIP registration service for users in the domain, and also introduce an IP address mapping function which is used to attach the MNs to the µHIP access network by registering the local locators (IP_L) of MNs.
IP_L is valid only in the given domain and the LRVS is responsible for mapping every IP_L to a globally routable address (i.e., global locator, IP_G). IP_G is used to register the MNs at their standard RVSSs and to deliver packets outside the micromobility domain during further communication sessions.

The basic operation of µHIP starts with an initialization mechanism: the MN physically connects to one of the access routers (AR) of the domain, then gets the IP_L based on e.g., IPv6 stateless autoconfiguration. After this, the MN either may actively initiate a HIP service discovery procedure [27] or passively wait for a service announcement in order to detect the LRVS service (HIT_LRVS, IP_LRVS) provided in the visited micromobility area, and will register itself to the LRVS with the standard Base Exchange (BEX) sequence [7]. During this service discovery and registration procedure the LRVS not only registers the MN’s HIT with the new IP_L, but maps IP_L with an assigned IP_G as well. After the MN successfully registered at the LRVS (with the HIT_MN-IP_L-IP_G triplet), it needs to perform the update and/or registration procedures at its RVS and current CNs (with the HIT_MN-IP_G pair). Therefore the MN – strongly relying on the self-certifying cryptographic identifiers provided by HIP and on the mechanisms introduced in [28] [C21] – delegates its signaling rights to the LRVS at which it is registered. The appropriate certificates are sent after the BEX, resulting that the LRVS will own the rights to signal on behalf of all MNs in the micromobility domain under its authority. In possession of these delegated rights the LRVS is able to securely register or update to the RVSSs and CNs on behalf of the MNs with the IP_G global locators assigned to them.

In case of intra-domain handovers the MN will receive a new IP_L from the new Access Router belonging to the serving LRVS. In this case the MN – realizing the change of its IP address – updates its registration (and if needed its delegation certificate as well) with its new IP_L at the serving LRVS. It is important to note that neither the CNs of the mobile node nor the RVS has to be informed about the intra-domain movement as the address changes are locally handled by the proposed micromobility extension. The movements of nodes are completely hidden from the outside world resulting in less signaling overhead, packet loss and handover latency.

In order to evaluate my proposed HIP-based micromobility solution and to provide a highly configurable, extensible, and adequate model for HIP, µHIP and other related protocols, I have designed an IPv6-based Host Identity Protocol simulation framework called HIPSIm++ (publicly available under GNU GPLv3 licence) [C17] on the top of the OMNeT++ 4.2 discrete event simulation environment [21].

I have used the standard HIP scenario as a reference, where the mobile HIP host (MN) changed its network point of attachment by connecting to another Wi-Fi access point (AP) due to its movement. As the APs were connected to different access routers advertising different IPv6 prefixes, the IPv6 address of the MN was changed after reattachment. Standard HIP mechanisms were applied to handle this mobility situation by running the HIP UPDATE process [9]. For the µHIP scenario the difference lies in the introduction of micro-mobility domains: two HIP LRVSs replace the access routers and control their Domains (1 and 2), where the first one owns two access points (AP1, AP2) providing possibilities to simulate intra-domain handovers within its LRVS control node. Inter-domain handovers are also implemented in the model: during its movement the MN changes its network point of attachment from AP2 to AP3 (belonging to Domain 1 and 2 respectively).

In the above two main scenarios the MN is able to communicate with UDP/TCP and also to migrate between the different APs such provoking handovers situations. By inducing 100 independent handovers during simulation runs I have measured three key performance indicators in three different sub-scenarios. The simulation results gathered are presented in three different graphs (Fig. 4).
4.2. Location Privacy Aware Micromobility Domain Planning Schemes

Mobile terminals’ location data possess important service-enabler potential, but in wrong hands it can be used to build up private and intimate profile of the mobile user and can pose serious threats to location privacy. In the all-IP world of future mobile Internet, location privacy of users is even harder to protect as the most common parameters in every single packet – i.e., the source and destination IP addresses – can easily be translated to a quite accurate estimation of the peers’ actual geographical location [29], [30] thus making third parties able to track mobiles’ real-life movements [31]. In next generation all-IP heterogeneous wireless communication systems moving across multiple IP subnets will occur more likely, resulting in much frequent IP address changes compared to today’s mainly homogeneous architectures, therefore further aggravate problems of location information leakage. However, micromobility solutions – besides grouping IP subnets into domains and providing near-seamless local handoffs – also include capabilities to support location privacy: localization of mobility events inside a micromobility domain can hide location information easily exposable by IP address changes of handovers [26].

As mobility becomes one of the most unique characteristics of future’s convergent architectures, more attention must be given to the location privacy issues, even at the earliest phases of design: at the network planning level. Existing network planning algorithms (e.g., [11], [25], [32], [J3]) are mainly focusing on the trade-off between the paging cost and the registration cost and – to the best of my knowledge – none of them have introduced privacy awareness in network planning methodologies. Also the potential of micromobility protocols to efficiently support location privacy was never taken into consideration in any domain

Figure 4: Simulation results of the µHIP scheme
planning algorithms available in the literature. This motivated me to develop mobile network planning tools that exploit inherent location privacy support of micromobility protocols while also considering the strict constraints formed by paging and registration costs.

**Thesis II.1** [J8], [B4] I have developed a simple location privacy policy model to provide boundary conditions for location privacy aware domain planning where both static requirements and dynamic demands are to be respected. Based on this model I have proposed a special rate weighting technique for enhanced and privacy aware graph representation of mobile networks. Using this novel toolset I have developed a privacy aware domain planning algorithm called PA-SABLAF (Privacy Aware Simulated Annealing based Location Area Forming) which is an improvement of my SABAS algorithm decreasing the number of inter-domain handovers while also considering the location privacy in the created structure.

In the location privacy policy model I have proposed, a combination of two substances is used to provide boundary conditions for location privacy aware domain planning. On the one hand I introduced the static location privacy significance level of the cells (SLP$_k$ for cell $k$) which can separate coverage areas inside the operator’s network that are considered to be more sensitive to location privacy than others. On the other hand I defined user’s location privacy profile for different location types (ULP$_{u,t,k}$ for user $u$ and location type $t$ of cell $k$) to describe what level of location privacy protection is required for a mobile user at a given type of location. The incoming dynamic demands are cumulated and the average will be compared with the static location privacy significance level of the issued cell at every announcement. The winner of this comparison – called the cell’s overall location privacy factor – will take over the role of the cell’s static significance level. In this simple way not only operators’ requirements, but also the dynamic demands of mobile users can be respected during the location privacy aware network design.

In order to integrate the effects of the cells’ overall location privacy factor into the boundary crossing rates between neighboring cells, I have created a special rate weighting technique. In the mathematical representation I applied, the cells are the nodes of a graph, the cell border crossing directions are represented by the graph edges and the weights are assigned to the edges based on the cell border crossing rates of every direction. These rates are weighted with the overall location privacy factor of the destination cell:

$$WR_{k|l} = CR_{k|l} \times OLPF_{k|l} + CR_{l|k} \times OLPF_{l|k}$$  \hspace{1cm} (1)

where $WR_{k|l}$ is the weighted rate of edge between cells (graph nodes) $k$ and $l$, notation $CR_{k|l}$ stands for the cell border crossing rate from cell $k$ to $l$, and $OLPF_{k|l}$ is the overall location privacy factor of cell $l$.

Based on the above definition, my proposed PA-SABLAF algorithm starts with a greedy phase by choosing the cell pair with the biggest weighted rate in the cell structure and includes them into domain $D_1$ of cells. In the next step, it searches for the second biggest weighted rate among the cell pairs for which is true, that one of them belongs to domain $D_1$. It checks whether inequality $N_k < N_{max}$ is satisfied, where $N_k$ is the number of cells in the $k^{th}$ domain and $N_{max}$ stands for the maximum number of cells in a single micromobility domain which provides the minimum of the registration cost and the maximum size of the location privacy protective micromobility domain. If the inequality is satisfied, the cell can be included into set $D_1$. If the inequality is not satisfied, the cell cannot be included into this set; a new domain with this cell is to be created in order to prevent exceeding the paging cost constraint, similarly to the operation of the SABAS [C9] and SABLAF [12] algorithms. In this way PA-SABLAF can join the most important cells according to the location privacy policy model.
which are also in the same dominant moving directions (highways, footpaths, etc.). After processing all the cell pairs in the above sequential and greedy way a likely sub-optimal domain structure will be created, which will serve as the initial solution ($s_0$ domain partitioning) for the simulated annealing part of the algorithm.

**Thesis II.2** [J8], [B4] I have proposed a location privacy metric called $LP_{mic}^S$ to express how efficiently a given micromobility domain structure takes into account static location privacy significance of cells and the incoming dynamic location privacy demands of users during operation. I have shown that PA-SABLAF appreciably improves the domain structure compared to its predecessor algorithm with an average of 10% location privacy gain.

I have proposed $LP_{mic}^S$ to show how effective could be the protection of users’ location privacy while keeping paging and registration costs on a bearable level in a given micromobility environment. I have quantified the inability of non inside-domain attackers in tracking mobile users by computing a weighted number of inter-domain changes of mobile nodes in the network. For every inter-domain handover of a mobile node and for the previous and the next cells of such handovers the metric calculation algorithm sums the value of the cells’ static location privacy significance and the squared value of the level of the mobile node’s location privacy profile set for the issued location types. The above calculation is performed for every mobile node, and the sum of these values will stand for the location privacy metric of the whole micromobility domain system:

$$LP_{mic}^S = \sum_u \sum_{h \in IH_u} (ULP_{u}^{[k]} + (ULP_{u}^{[l]})^2 + SLP_{[k]} + SLP_{[l]}$$

(2)

where $IH_u$ means the set of all inter-domain handover events of user $u$, and $h_{[k][l]} \in IH_u$ stands for a handover event with exit and entry cells of $k$ and $l$ respectively. Implicitly the smaller $LP_{mic}^S$ values are the better.

![Figure 5: PA-SABLAF vs. SABAS (left) and Loc. privacy gain vs. cost increment for PA-SABLAF (right)](image)

I have evaluated PA-SABLAF in a further extended version of the mobile environment simulator already introduced in Thesis I.2. Four different scenarios were formed by complex network architectures consisting of several cells, mobile nodes and various compound road grids. Using this environment I have compared my algorithm with its ancestor – the already introduced SABAS which is without any trace of location privacy awareness. Simulation results show (Fig. 5) that PA-SABLAF finds a much better domain structure in terms of the $LP_{mic}^S$ metric for every value of $N_{max}$ compared to the original SABAS. However, we have to pay the price of this benefit: the registration cost is slightly higher in most of the cases with a maximum of 4.8%.
However $LP_{mic}^u$ is able to numerically present the location privacy capabilities of a complete network’s certain micromobility domain structure, it lacks in generality. That is why I have started to evaluate my scheme using more general and widespread location privacy metrics.

**Thesis II.3** [J8], [B4] I have proposed a location privacy metric called $LP_{mic}^u$ in order to adapt the uncertainty-based location privacy metric for localized mobility scenarios and measure the level of obfuscation provided by the built-in location privacy supporting capability of micromobility domain systems. I have developed a privacy aware domain planning algorithm variant called PA$^u$-SABLAF to enhance the domain planning process in terms of the $LP_{mic}^u$ metric. I have shown that PA$^u$-SABLAF is able to improve the domain structure with a significant 30% relative growth in high PoA number domains by raising the possible number of transitions at inter-domain movements.

The uncertainty-based location privacy metric was originally proposed in [33] where authors proposed to measure location privacy of a given user in the system as the attacker’s uncertainty during linking observed events to users. I have adapted their scheme to my model and extended it to be applicable for all the users in the micromobility system.

In a micromobility network the attacker relying on intercepted IP packets can only observe series of crossed domains along the MN’s movements. That is why I split the trajectory of the MN into domain entry and exit points (which are basically the observable events in my threat model) and delimit unobservable path segments between them. As these inside-domain path segments are not traceable based on IP information and assuming that domains contain more than two cells at least, the attacker can only deduce the entry and exit points (so called “flashes”). I assume that transitions are not weighted and the transition probability is the same in every case. Considering $Pr_{A}(d)$ here as the probability of the attacker guessed right when reckoning the actual entry and exit points of crossing domain $d$, a user’s uncertainty-based location privacy metric for a particular domain inside the network can be produced by calculating the entropy of $Pr_{A}(d)$. By calculating this entropy for every domain of every user, and creating the sum of these entropies I get the overall entropy of a micromobility system denoted by $LP_{mic}^u$ (as this metric is an entropy-like measure, the larger values denote the better location privacy support).

$$LP_{mic}^u = - \sum_u \sum_j Pr_{A}(d_j) \times \log_2 \left( Pr_{A}(d_j) \right)$$ (3)

In order to create a more general domain planning scheme based on the criteria of the widespread and universal uncertainty-based location privacy metric I have designed the PA$^u$-SABLAF algorithm, where the greedy phase also considers the crossing rates of all the neighboring transitions besides the crossing rates of the actually examined transition. Since the maximum number of cells in a single micromobility domain is limited by $N_{max}$, the process can always create a structure where cells with big transition rates will create domains and simultaneously their neighbors with reasonably significant number and volume of transitions will form neighboring domains thus increasing the uncertainty of the attacker observing users' domain changes. According to this, PA$^u$-SABLAF will lead the traffic of cells with large transit demands away toward as many edges/edge series as possible. The calculation of the weighted rate based on the above considerations and used in the greedy phase of PA$^u$-SABLAF is as follows.

$$WR_{il}^{k} = CR_{[k][l]} + CR_{[l][k]} + TF_{[l]}$$ (4)

where $CR_{[k][l]}$ stands for the cell border crossing rate from cell $k$ to $l$, and $TF_{[l]}$ is the transition factor of cell $l$ (a cell still waiting to be grouped into a domain). I defined the
transition factor as $TF_l = \sum_{m \in A_l}(CR_{l,m} + CR_{m,l})$ where $A_l$ means the set of all neighbors of cell $l$. Besides this modified weighting and cell selection scheme the PA$^u$-SABLAF algorithm is the same as the method introduced in Thesis II.1.

I have shown in my simulations that PA$^u$-SABLAF achieves serious relative gain in terms of the location privacy metric and the registration cost increment: a more then 30% relative growth can be noticed for location privacy in the $N_{\text{max}} = 6$ case (Fig. 6). Despite this promising result PA$^u$-SABLAF shows the most serious volume of additional registration costs after location privacy aware domain planning: even the smallest cost growth is 27%. However, this is compensated by the remarkable revenues of the $LP_{\text{mic}}^u$ metric.

**Figure 6**: PA$^u$-SABLAF vs. SABAS (left) and Loc. privacy gain vs. cost increment for PA$^u$-SABLAF (right)

**Thesis II.4** [8], [B4] I have proposed a location privacy metric called $LP_{\text{mic}}^i$ in order to adapt the traceability-based location privacy metric for localized mobility scenarios and quantify the incapacity of attackers in localizing or tracking mobile nodes in a micromobility domain system. I have developed a privacy aware domain planning algorithm variant called PA$^i$-SABLAF to enhance the domain planning process in terms of the $LP_{\text{mic}}^i$ metric. I have shown that PA$^i$-SABLAF is capable to improve the domain structure with an average gain of 3.9% by transacting and keeping user traffic inside the domains and also decreasing the registration cost in most of the cases.

The traceability-based metric captures the level to which the attacker can track a mobile user with high certainty. In [34] authors define a so called mean distance to confusion metric, which measures the mean distance over which tracking of a user may be possible by the attacker. However, in my model attackers are not able to track mobile users when they are moving inside a particular micromobility domain. It means that domains serve as confusion points, which also implies that mean distance to confusion approaches become vague. This motivated me to create a modified traceability-based metrics called mean distance in confusion ($LP_{\text{u}}^i$) where I measure degree of location privacy as the mean distance over which tracking of a user may not be possible by the attacker. Let $Y_u$ stand for the set of all the untraceable periods for user $u$. Based on this notation the location privacy metric of user $u$ based on mean distance in confusion ($LP_{\text{u}}^i$) can be defined as follows.

$$LP_{\text{u}}^i = \left(\frac{\sum_{(\hat{e}_i, \hat{e}_j) \in Y_u} ||\text{loc}(\hat{e}_i) - \text{loc}(\hat{e}_j)||}{|Y_u|}\right)^{-1}$$  \hspace{1cm} (5)

where $\text{loc}(\hat{e}_i)$ stands for the location at which the event $\hat{e}_i$ occurred. Therefore I calculate the overall traceability-based location privacy metric of a micromobility system ($LP_{\text{mic}}^i$) as follows (the location privacy supporting capability is proportional with the mean distance in
confusion, so here the exponent implies that the smaller values are the better).

\[ L_{mte}^{i} = \sum_{u} \left( \frac{\sum_{e_i \in v_u} \| \text{loc}(e_i) - \text{loc}(\hat{e}_i) \|}{|v_u|} \right)^{-1} \]  \hspace{1cm} (6)

In my PA-SABLAF variant I take the cost constraints into consideration and simultaneously create a domain structure in which mobile users will likely perform inside-domain movements. This can be achieved by increasing the number of “deflector” edges inside the domains. I define an edge or a series of edges as “deflector” if it possesses significant crossing rate and/or it provides input and output for high crossing rates of other edges or series of edges from multiple directions. By inserting cell pairs with deflector edges into the micromobility domains we can enforce that frequent cell sequences of mobile users will likely consist a domain. The calculation of the weighted rate based on the above introduced idea framed for the greedy phase of PA-SABLAF is as follows.

\[
\begin{align*}
\text{if} & \quad E_{[k][l]} \in D_{\psi} \\
\text{then for} & \quad \forall E_{[i][j]} \in E_{[k][l]} \cup A_{[k][l]} \\
\text{do} & \quad WR^{i}_{[i][j]} = CR_{[i][j]} + CR_{[i][j]} + DF
\end{align*}
\]  \hspace{1cm} (7)

where \( E_{[k][l]} \) denotes the edge between cells \( k \) and \( l \), \( D_{\psi} \) means the set of deflector edges containing edges with the upper \( \psi \) percent of all crossing rates in the network, \( A_{[k][l]} \) is the set of neighbors of \( E_{[k][l]} \), \( CR_{[k][l]} \) stands for the cell border crossing rate from cell \( k \) to \( l \), and \( DF \) is a constant called deflector factor used for rewarding certain edges with deflector properties. Besides this special weighting technique of (7) the PA-SABLAF algorithm is basically identical to the method introduced in Thesis II.1.

![Graph](image)

**Figure 7:** PA-SABLAF vs. SABAS (left) and Loc. privacy gain vs. cost increment for PA-SABLAF (right)

The simulation results of PA-SABLAF evaluation are depicted in Fig. 7. This algorithm variant performs a moderate average gain (3.9%) and also shows negative relative gain in the \( N_{\text{max}} = 4 \) case. However, the algorithm enhances the privacy metric together with registration cost in all the other \( N_{\text{max}} \) cases which is a valuable achievement.

As a result of my efforts I can state that the proposed approach proved its power by significantly enhancing the location privacy of users in the network. The total average gain in location privacy for every run of all the three algorithm variants I developed approached 20% at the expense only of a total average 8% growth of the global registration cost (meaning an average 12% relative gain), and there were also distinct cases when the scheme operated with more than 30% relative gain.
4.3. Optimized Solutions for Network Mobility Management

In next generation wireless telecommunication not only single mobile entities have to be taken into account (host or terminal mobility), but also entire mobile networks moving between different subnets need to be maintained as a whole (i.e., network mobility or NEMO) [C5], [C10], [C15]. IPv6 has introduced support for both mobility cases by MIPv6 [2] and NEMO BS [4]. When a host or a moving network has multiple interfaces and/or several IPv6 addresses, it is regarded multihomed, requiring special protocols (e.g., MCoA [35]). With these mobility supporting mechanisms all sessions remain active, even when the mobile node/network changes its subnetwork. However, handovers at network layer usually take several seconds due to the large number of L1/L2/L3 processes. Several proposals exist to overcome the huge delay. Mobile IPv6 Fast Handovers [36] is one example, and there are plenty of other proposals as well (e.g., [17], [37], [B7]. However, according to my best knowledge, none of the existing solutions exploit the benefits of overlapping radio access coverages by managing multiple tunnels and predictive tunnel switching.

In order to enhance NEMO solutions, I have followed two approaches. On the one hand I have extended standard IPv6-based network mobility by forming a framework based on a special, multi-tunnel based, predictive, seamless handover solution (Thesis III.1 and III.2). On the other hand I have further extended Host Identity Protocol (HIP) and my introduced µHIP scheme by developing and evaluating a HIP-based NEMO protocol (Thesis III.3).

Thesis III.1 [C19], [C25], [B7] I have developed a location information aided predictive mobility management framework with an efficient handover execution scheme for multihomed NEMO BS configurations, which combines the benefits of MCoA with a new prediction-driven cross-layer management entity allowing NEMO BS mobile routers to operate using always the best available access networks and to perform seamless handovers when multiple overlapping radio coverages are available.

In the proposed scheme I use Flow Bindings [38] to direct the whole traffic of the MR through one active egress interface. In this way the solution loses the benefits of redundant interfaces, but gains the possibility to use inactive interfaces for handover preparation, i.e., selecting appropriate access network, performing lower layer connections and acquiring new IPv6 addresses. Therefore the scheme requires several interfaces for operation. Some of the interfaces are used for normal communication (“active”), the others are used for handover preparation (“inactive”).

![Figure 8: The proposed framework (left) and the handover execution protocol (right)](image-url)
The activation of a new interface must be accurately synchronized with the deactivation of the old one. The activation/deactivation procedure means simultaneous reallocation of NEMO BS tunnels. It is performed by properly scheduled flow binding policy control messages on the HA and the MR.

The proposed framework (Fig. 8, left) has three main components: Access Network Predictor (ANP), Handover Manager on the MR (HM-MR) and on the Home Agent (HM-HA). I do not claim all the functional entities are my results; however the overall framework and the design of the predictive handover execution scheme are. The ANP is responsible for maintaining a database containing information of access networks, and sending periodic prediction messages to the HM-MR module based on the current velocity vector and the contents of the database associated with the predicted geographical location. In order to avoid the explosion of the size of the access network database, the received GNSS coordinates are rounded (the longitude and latitude values are multiplied by 10,000 and rounded to the closest integer), therefore instead of a continuous space they form a limited set with members called raster points inside a raster net. The database is kept up-to-date by the Measurement Unit residing in the Handover Manager, which passively monitors the available access networks via one of its passive interfaces, periodically sending network availability and performance indicators such as SNR and IPv6 prefix to the Access Network Predictor. Based on the predictions received from the ANP, the Connection Manager may decide that the currently active access network will no longer be the best available network in the predicted timeframe. When the HM decides to perform a handover, in order to use the benefits of MCoA, the following steps are executed. Using one of the inactive interfaces the HM connects to the new access network and establishes a new Mobile IPv6 binding. At this stage, the current and new access networks are both connected and Mobility Tunnels are established between the MR and the HA. Handing over to the new access network is entirely based on Flow Bindings, which in this case means that all flows are moved from one interface to another. To avoid asymmetric routing, the MA and HA has to modify their bindings simultaneously, in a timely manner. The schedule is communicated by the Flow Binding modules in predictive Flow Binding Update/Acknowledgement messages (Fig. 8, right). When the changes of flow bindings are executed, the new interface is marked as active, while the rest of the communication interfaces are set to inactive mode. The mobile network nodes (MNNs) inside the NEMO will always and transparently use the communication path spanned by the active interface (Fig. 8, left). Different Handover Policies may have different effects on handover strategies.

The proposed framework and handover execution protocol strongly relies on the prediction accuracy which depends on the rasterization scheme working inside the ANP module. That is why I have started to analyze the limitations of the overall architecture inherited by possible wrong positioning on the raster net inside the ANP.

**Thesis III.2** [C25], [J16] I have developed a probabilistic system model for the ANP module and proposed an appropriate rasterization scheme where the probability of wrong positioning on the raster remains below 1%.

Assume that we have a set of raster points given as \( S = \{x_1, x_2, ..., x_{\infty}\} \). \( x_i \) represents the \( i \)th point which is a geographical position with two coordinates: one on the west-east axis and one on the north-south axis. \( S \) is an infinite but countable set. The members of the set are constant: they are given by the actual raster size. Assume that we are at a geographical position \( x_0 \) (\( x_0 \) can be given by god – no possibility to measure it exactly). We have a GNSS measurement equipment and want to figure out, what \( x_0 \) is. We make measurements and we
get $\eta$ as an estimate, which is not exact of course. $\eta$ is a random number (Gaussian, due to the large number of independent effects), with expectation of $x_0$ and covariance matrix $C$:

$$\mathbb{E}[\eta] = x_0$$

$$Pr(\eta \leq y|x_0, C) = \Phi(y, x_0, C) = \int_{\infty}^{y_1} \int_{\infty}^{y_2} \frac{1}{\sqrt{2\pi} \sqrt{(detC)}/2} e^{-(x-x_0)^T C^{-1}(x-x_0)} dx_1 dx_2$$

Note that in the last equation we introduced a new notation, $\Phi$. Also note that $\eta \leq y$ means all $\eta$ points where both coordinates are less than or equal to the ones of $y$.

The database uses the raster points only. Thus, based on the measured value $\eta$ we can choose the closest raster point as

$$\xi(t) = \arg\min_{\eta \in \mathbb{R}^2} \parallel \eta(t) - x \parallel$$

Here, the time dependence have been also added as $(t)$, and $\parallel \cdot \parallel$ measures the absolute distance. With the help of God (knowing $x_0(t)$), we would get the perfect estimate $x(t)$ as

$$x(t) = \arg\min_{\eta \in \mathbb{R}^2} \parallel x_0(t) - x \parallel$$

Then the formalized question: what is the probability of making a wrong estimate?

$$Pr(\xi(t) \neq x(t)) = ?$$

Note that both (10) and (11) are non-linear operations, making it difficult to analyse the problem. The following subsection is about evaluating this probability.

Fig. 9 shows the general geographical setup. As the raster net is self similar, we can put it into the centre of the coordinate system. The area of $T$ is defined as

$$T = \{(i,j), \text{ where } -a \leq i \leq +a \text{ and } -b \leq j \leq +b\}$$

![Figure 9: Raster net setup of the probability model](image)

Assuming that the real geographical position $(x_0)$ is equally probable at any position, the probability of making a wrong estimate $Pr(\xi(t) \neq x(t))$, equals the probability that the real position is inside the grey area $T$ ($x_0(t) \in T$), and the measured point is outside of $T$ ($\eta(t) \notin T$):

$$Pr(\xi(t) \neq x(t)) = Pr(x_0(t) \in T \cap \eta(t) \notin T) = 1 - Pr(x_0(t) \in T \cap \eta(t) \in T)$$

The probability of $\eta$ falling into $T$ can be computed as
Following equation (14), the Bayes' rule and our positioning error constraint, we get

\[
0.01 \geq Pr(\xi(t) \neq x(t)) = 1 - \frac{1}{4ab} \int_{-a}^{a} \int_{-b}^{b} Pr(\eta \in \mathcal{T}|x_0, C) \, dj \, di
\]  

(16)

Considering a GPS system for our GNSS measurements with a horizontal positioning error of \( \sigma = 5m \) (standard deviation), and taking into consideration that the length in one minute of longitude depends on the latitude (which is \( \sim 47.5^\circ \) N for Budapest), we get that the appropriate raster net is around 18.2m \( \times \) 27m.

In order to provide network mobility support in the Host Identity Layer, I have further extended my \( \mu \)HIP framework (already introduced in Thesis I.3) and built a novel, HIP-based NEMO protocol (called HIP-NEMO) upon it. Although there were some proposals like [39] before HIP-NEMO, my work can be considered as the first complete and pure HIP-based NEMO solution.

**Thesis III.3** [C6], [C17], [B5], [J2], [J14], [J17] *I have proposed a Host Identity Protocol based network mobility solution (HIP-NEMO) that introduces a new network entity called Mobile Rendezvous Point (mRVS) in charge of providing NEMO services for HIP-aware MNNs. The scheme eliminates suboptimal user-plane tunneling known in NEMO-BS and ensures efficient communication of the mobile network. I have shown by extensive simulations built on complex protocol models that my proposed HIP-NEMO scheme outperforms the standard NEMO-BS network mobility management solution with a cumulated average throughput gain of 211% and provides a significant functional extension to the basic HIP protocol without considerable cost impacts.*

In HIP-NEMO the introduced mRVS provides continuous connectivity for the served MNNs and also for other moving networks connected to it. Moreover, the mRVS is in charge to be the signalling proxy for MNNs: the MNNs delegate their signalling rights to the mRVS thus it can efficiently control any mobility scenario. As the initiations, the MNNs register themselves at the mRVS and delegate their signalling rights to it (Fig. 10, left).

The mRVS maintains a mapping between the HIT of the MNNs and their IP addresses valid inside the moving network (\( IP_{\text{MNN}} \)). This mapping entry is then associated with an IP address assigned by the mRVS (\( IP_0_{\text{MNN}} \)). The purpose of this address is exactly the same as
its counterpart in case of LRVSs in μHIP: it is valid outside the moving network and is exchanged with the MNN's actual IP address by the mRVS inside the NEMO. After the initial registration the mRVS does the necessary signalling for the MNN by informing the RVS of the MNN.

As the initial steps are completed it is possible to trigger a communication session with one of the MNNs in the mobile network (Fig. 10, right). The first packet of CN-MNN Base Exchange is sent to the MNN’s RVS which forwards the packet towards the mRVS according to its mapping. The I1 packet [7] is intercepted by the mRVS and the destination address is changed to MNN’s inside moving network IP address. Finally the MNN receives the message and answers with the R1 according to the HIP BEX procedure [7]. The source address is changed at the mRVS to the globally reachable one and sent directly to the peer. The connection setup finishes in the regular way, but with the mediator role of mRVS.

Handover situations are also straightforward to handle in HIP-NEMO: as mRVS has right to signal on behalf of its MNNs, it can send HIP UPDATE packets to the RVS and to all of the communication partners of every MNN. The UPDATE packet sent to the RVS contains the LOCATOR parameter, which holds only the new address that mRVS uses on its new location. On reception of the UPDATE packet, the RVS will update the binding of the mRVS. Furthermore, all other bindings are updated that are indexed by the HIT of the mRVS.

In order to evaluate my proposed HIP-NEMO scheme and compare its performance against the standard and widespread NEMO-BS solution as reference, I have extended the HIPSim++ framework [C17] with the protocol models and the required NEMO scenarios.

I have defined one mobile network with one MNN and one mRVS/MR for both the HIP-NEMO/NEMO-BS scenarios, respectively. This mobile network was able to perform handovers between different access points in the home and foreign networks. In the case of NEMO BS there was a Home Agent (HA) in the topology for the MR, and in the case of HIP-NEMO there was an RVS defined. After the simulation start, a CN initiated UDP or TCP connections with the MNN, and then started to send data packets. By using a special node in the topology between the home network (containing the HA and the RVS) and the foreign networks, additional delay from 0 to 300 ms was introduced into the test setup. The gathered and depicted results in Fig. 11 and 12 were rendered using the averages of a total of 10000 independent measurements each.

TCP results (Fig. 11) show that the frequency of handovers has tremendous impact on both solutions. The TCP throughput during the movement decreases significantly with higher...
handover frequency. However, as in case of NEMO BS all signalling and data traffic flows through the HA, the rate of that decrease is higher than it is when HIP-NEMO is used. My solution builds near-optimal routes between communicating entities and eliminates user-plane tunnels from the communication path which also results in the fact that user-plane performance will not be tainted by the increasing topological distance between the home and foreign networks (i.e., between the RVS and the HIP supported mobile network). The cumulated average throughput gain of HIP-NEMO is around 211%.

UDP results (Fig. 12) further strengthen the above statements. The effects of one handover are independent from the added delay in case of HIP-NEMO, while NEMO-BS shows notable degradation when the mobile network is more far away from home. The average performance gain of HIP-NEMO in terms of lost UDP packets for the high data rate and high added delay cases is more than 19%.

![Figure 12: Simulation results for the UDP packet loss measurement](image)

4.4. Schemes for Distributed and Flat Mobility Management

According to industry prognostications like [1] it is highly expected that due to their centralized (anchor-based) design, mobile Internet architectures currently being under deployment or standardization will not scale particularly well to efficiently handle the challenges. Motivated by the above reasoning, novel mobile architectures and accompanied protocols started to emerge where bottlenecks from packet communication is removed by eliminating user-plane anchors from the network and bringing pure IP routing close to the mobile terminals in terms of physical location in the architecture. Decentralized, robust, self-configuring and self-optimizing network structures are envisioned with reduced operation expenditure (OPEX), improved system capacity and energy efficiency. To enhance scalability of the core network in mobile Internet architectures, the Ultra Flat architecture (UFA) has been introduced by Khadija Daoud et al. [5]. Their proposal reduces the number of network nodes to only one serving node called the UFA Gateway (UFA GW) and traditional user and control plane functions are distributed in such UFA GWs deployed at the edge of the architecture, close to the subscribers. The main characteristic of this proposal is that the execution of handovers is managed by the network via the Session Initiation Protocol (SIP) operating within the frame of the IP Multimedia Subsystem (IMS) [C20], [J5]. Even though SIP is a very powerful signalling solution for UFA, it is not applicable for non-SIP (i.e.,
legacy Internet) applications and the published SIP-based UFA scheme also does not comply with ITU-T's recommendation of requirements for ID/Loc separation in future networks [6].

This motivated my co-authors and me to work on an alternative signalling scheme for the Ultra Flat Architecture based on the promising ID/Loc separation method of the Host Identity Protocol. My contribution to the developed HIP-based Ultra Flat Architecture (UFA-HIP) lies in the design of its novel, HIP mobility management, signalling delegation, context transfer and cross-layer interworking based system framework by the definition of the main functional elements, their structure and roles inside the control plane, in the development of a proactive, distributed, 802.21 MIH / HIP-based handover initiation, preparation, execution and completion protocol, and in the performance evaluation of the UFA-HIP mobility management scheme by simulations. The definition of general purpose HIP delegation services, L2, L3 terminal attachment and session establishment procedures and a SIP-extended handover completion phase of UFA-HIP are the results of my co-author Zoltán Faigl.

Thesis IV.1 [C21], [C22], [J6], [J9], [J12], [J13] I have proposed a Host Identity Protocol based system framework for the Ultra Flat Architecture (called UFA-HIP), which completely eliminates centralized IP anchors between Point of Access (PoA) nodes and correspondent nodes, places network functions at the edge of the transit and access networks (close to PoAs) in the UFA-HIP GWs, integrates 802.21 MIH and HIP features to provide efficient inter-GW mobility management, and incorporates signalling delegation and context transfer to reduce the number of HIP Base Exchanges (BEX) between the MN and the network and within the network and also to remove overhead from wireless links by shifting significant part of signalling overhead of MNs from the air interface to the wired UFA-HIP segments.

The proposed HIP-based Ultra Flat Architecture (UFA-HIP) system framework defines seven main building blocks (Fig. 13): 1) several access networks (both wired and wireless), 2) an IP/MPLS transit network, 3) HIP capable UFA GWs (UFA-HIP GWs) controlling main network functions, 4) an optimized terminal attachment scheme with cross-layer access authorization, 5) a session establishment protocol, 6) a handover initiation, preparation, decision and execution subsystem based on the IEEE 802.21 Media Independent Handover (MIH) standard [40] and extended HIP functionalities, and 7) a HIP-based control network.

Heterogeneous access networks provide the air interface for MNs making them able to connect to the core infrastructure (and to the Internet) anytime, anywhere. The IP/MPLS transit network is the operator’s backbone including routers and core network elements (for service and configuration provision, 802.21 services etc.), and natively connecting UFA to the global backbone (i.e., the Internet).

In this system centralized IP anchors between PoAs and correspondent nodes are totally removed, and network functions are placed at the edge of the transit and access networks (close to the PoAs) in the UFA-HIP GWs. The solution applies HIP for IPsec security association (SA) establishment between the MNs and UFA-HIP GWs, and between the UFA-HIP GWs. IP-level handovers are prepared and executed using HIP delegation services [C21] and CXTP-based context transfer [41]. The main tasks of UFA-HIP GWs in my proposal:

1. Performing fast cross layer (L2 and HIP-level) access authorization.
2. Actively interacting with hosts through delegation-based HIP and IPsec association management and context transfer for optimized message exchange in HIP-based UFA mobility and multihoming. (Note that the proposed system framework transports end-to-end flows between MNs and CNs in a hop-by-hop manner. The middle-hops are the UFA-HIP GWs, i.e., the delegates of the end peers).
3. Performing the actual mapping/routing between outer header IPsec tunnels based on inner header identifiers.
4. Coordination of resource allocation, load balancing, and handover decisions with the help of the UFA-HIP cross-layer module (MIHU in the MIH taxonomy).

I proposed the use of HITs in inner IP headers for the identification of flows in UFA-HIP GWs, with the same purpose as the Control Plane Header (CPH) in [42]. Without delegation, maintaining end-to-end security associations (SAs) between every communicating peers would be required, as in the SPINAT-based frameworks [42].

There are control functions which are not part of the UFA-HIP GWs and remain in the core network. The optimal location of these functions is subject to further research. Such functions are IP Multimedia Subsystem (IMS), the Home Subscriber Server (HSS), the authentication, authorization and accounting (AAA) servers, service and configuration provision (DHCP), Media Independent Information Service (MIIS). Existing service platforms and application servers remain centralized as well.

The IEEE 802.21 MIH management subsystem inside my proposed framework handles handover preparation issues and relating signaling tasks in order to initiate proactive HIP handover procedures in the UFA and to support network and mobile controlled handover decision. In my system framework, network initiated 802.21 handover preparation procedures are triggered by the serving UFA-HIP GWs (refer to Appendix C.2 in [40]).

The control network (HIP-based addressing and mobility support) contains a HIP-compatible Domain Name System [43] and a HIP Control Plane which stores and distributes dynamic and frequently changing binding information between host identities and locators of all actively communicating (mobile) hosts in UFA-HIP. This control plane might be a conventional RVS park or a complete distributed HIP signalling architecture like Hi³ [44]. The records managed here are provided by the UFA-HIP GWs using their own global locators as location information to be bounded with identities of their actively interacting partners.

HIP BEX and Update procedures deal with dynamic negotiation of IPSec security associations between the MN and the UFA-HIP GW to protect user data and mutually authenticate the MN and the network. The handover preparation and initiation subsystem handles handover preparation issues and relating signalling tasks in order to initiate proactive HIP handover procedures in the UFA and to support both network and mobile controlled handover decisions. The handover execution procedure is started by the source UFA GW (S UFA GW). HIP and IPsec contexts are established between the target UFA GW (T UFA GW) and the MN’s CNs, furthermore, between the target UFA GW and the MN, using the
mediation of the S UFA GW. This is possible due to the delegation of HIP signalling rights from the MN and from the target UFA GW to the source UFA GW. Context Transfer Protocol [41] is used to transfer the HIP and IPsec contexts from the source UFA GW to the target UFA GW and the MN. As the contexts are in their place the MN is notified by the handover preparation and initiation subsystem to attach to the new PoA.

The proposed distributed mobility management scheme anticipates live registration of MN to the network, live registration between target and source UFA-GWs, an already completed setup phase where the UFA GW configures QoS thresholds on the MN using MIH procedures [40], and a special algorithm deciding and triggering to start the handover process to one of the candidate UFA GWs (C UFA GW) before the connectivity on the S UFA GW’s PoA is lost. After receiving this trigger message the 802.21 MIH handover initiation starts (Fig. 14 left part) with the following sub-phases. (1) Discovery: during this phase, the list of candidate UFA GWs is obtained through the 802.21 Media Independent Information Service (MIIS) [40]. (2) Query: in this phase, the mobility decision algorithm acquires all QoS metrics for all available candidate UFA GWs. (3) Selection: the mobility decision algorithm running either on the network or in the terminal side, decides for the target network (i.e., T UFA GW). The S UFA GW queries the MIIS about the available neighbouring networks, then asks the MN to narrow the list of candidate access networks, and finally checks the available resources at each C UFA GW. Thereafter, it decides the selected T UFA GW for the handover procedure.

Figure 14: 802.21 MIH handover initiation (left) and HIP handover preparation phase 1/2 (right)

After the selection of the target PoA and the T UFA GW, the necessary HIP and IPsec contexts are proactively established in the network by the S UFA GW using Type 1 and Type 2 HIP delegation services [C21] in the HIP handover preparation phase. As depicted in Fig. 14 right part, the S UFA GW initiates a Type 2 Mandated Action Request on behalf of the MN for handing off MN’s sessions. It triggers a bulk Type 1 Delegation Action Request sent...
back to the S UFA GW. With this request the T UFA GW authorizes the S UFA GW for the establishment of HIP and IPsec connections with the MN’s peers in the name of the T UFA GW. Then the S UFA GW sends the security contexts to the T UFA GW using CXT protocol protected with IPsec. Hence, the number of HIP BEX procedures can be reduced and replaced by HIP Updates. After the successful context transfer, the T UFA GW updates the traffic forwarding policies for the MN at the CNs, RVS, and the MN, as illustrated in Fig. 15, left. Firstly, Type 2 Mandated Action Requests are sent by the T UFA GW to the CNs and the RVS in the MN’s name. After updating the MN’s peers, the T UFA GW informs the S UFA GW with a Type 2 Mandated Action Response to prepare for the redirection of the sessions. The T UFA GW updates its HIT-based traffic forwarding table [J6] to receive traffic from MN’s peers and send packets towards the S UFA GW. The S UFA GW also updates the MN’s and its own local HIT-based traffic mapping table: the traffic coming from the MN, related to the sessions that will be handed off soon, must be mapped to the IPsec tunnel that has the T UFA GW on the other side. The MN delays the activation of forwarding its traffic to the T UFA GW. Therefore, the traffic of the MN passes through the source and target UFA GW until the physical handover completes.

![Figure 15: HIP handover preparation phase 2/2 (left) and Handover execution/completion phase (right)](image)

Fig. 15 right illustrates the handover execution phase for my proposed UFA-HIP scheme. After HIP handover preparation phase, L2 handover execution procedure is initiated by the MIH_N2N_HO_Commit and MIH_Net_HO_Commit request messages of 802.21 MIH [40], towards the T UFA GW and the MN, respectively. Then, the MN attaches on L2 to the target L2 PoA. The last phase is initiated by the MN when it physically attaches to the target L2 PoA and UFA GW. The MN signals to the T UFA GW that the handover was successfully executed and S UFA GW and L2 PoA can release the resources maintained for the MN’s handed off sessions. This last phase is executed by the 802.21 MIH protocol’s MIH_N2N_HO_Complete procedure. Finally, the traffic forwarding policy must be updated for the MN in the source and target UFA GWs to exclude the S UFA GW from the path.

The evaluation of the developed scheme was performed in the extended version of the INET/OMNeT++ based open source HIPSim++ [C17] simulation environment.

**Thesis IV.3** [C17], [J14], [J17] I have shown that the proposed proactive UFA-HIP handover preparation and execution protocol reduces the handover latency with an average 67% and thus almost totally eliminates the effects of frequent inter-GW mobility events in distributed or flat mobile networking architectures. I have also revealed the scale of the benefits exploited from the scheme by UDP and TCP applications. The number of lost UDP packets is 55% less
in average, while the average TCP throughput gain of the distributed scheme is above 60% compared to the legacy solutions.

I have used standard HIP and MIPv6 mobility management solutions as reference: the mobile host (MN) changed its PoA by connecting to another Wi-Fi access point (AP) due to its movement. As the APs were connected to different access routers advertising different IPv6 prefixes, the IPv6 address of the MN was changed after reattachment. In the HIP case, standard RFC 5206 mechanisms were applied to handle this mobility situation by running the HIP UPDATE process [9], while in the MIPv6 case standard routing optimization mechanisms (RO) [2] were switched ON and OFF in two sub-cases. In the UFA-HIP scenario two HIP-capable UFA GW nodes replace the legacy access routers and control their PoAs (Wi-Fi AP 1 and 2). In this scenario the MN uses active Signal-to-Noise Ratio measurements and a threshold value to trigger handover preparation. HIP functions (signaling delegation and context transfer) were implemented as extensions to HIPSim++, while the model of 802.21 MIH framework was built on the Notification Board toolset of INET/OMNeT++ [45].

![Graph](image.png)

**Figure 16:** Handover latency of the UFA-HIP scheme

In all the above scenarios the MN is able to migrate between the different APs with a constant speed such provoking handovers situations. By inducing 100 independent handovers during simulation runs I have measured three key performance indicators in three different sub-scenarios. Fig. 16/a presents the Handover Latency as the average of the 100 handover series for every RA interval within its 99% confidence interval. Measurements show that UFA-HIP handover performance is independent of the subsidiary IP layer mechanisms (i.e., delays of acquiring IP address, duplicate address detection, etc.) and keeps service interruption delay slightly above 1 sec. Measurements show that the service interruption delay of UFA-HIP is independent from the configuration delay in the target network (i.e., RA interval), and about 67% smaller than the reference protocols case in average, thanks to the advanced proactive operation which basically reduces the handover disruption to the Layer 2 (re-)attachment delay. Fig. 16/b presents the Handover Latency in function of the RA interval and the delay between source and target UFA-HIP GWs. The graph depicts that the performance of my proposed proactive handover solution is independent from the topological distance of the S and T UFA GWs.

Fig. 17/a shows how much UDP packet was lost during a handover in a HIP, MIPv6 RO ON, MIPv6 RO OFF, and UFA-HIP based system. The points on the graph represent the average UDP packet loss of 100 handovers for every offered datarate value and depicted within a 99% confidence interval. The differences and similarities of the examined protocols’
handover performance are clearly observable in the UDP transport layer. The power of my proactive, context-transfer based distributed solution designed for UFA is highlighted by the fact that the number of lost UDP packets is 55% less in average for the UFA-HIP case compared to the legacy HIP and MIPv6 aggregate performances.

Fig. 17/b depicts the TCP throughput proportion of the four protocols under analysis in a one minute communication session between the MN and the CN experienced at different handover frequencies from 0 to 9. The gain of my UFA-HIP solution is eye-catching especially when the circumstances are deteriorating (i.e., the number of handovers is increasing): in case of the highest handover frequency UFA-HIP shows more than 175% gain in TCP throughput, but also the average gain of the distributed scheme is above 60%.

Figure 17: Performance of UDP and TCP applications in the UFA-HIP handover scheme

With the help of extensive simulations I proved that the handover latency and the number of lost packets during handoffs can be significantly reduced while the TCP throughput can be considerably increased in case of my UFA-HIP proposal, meaning that relevant improvements in handover performance can be achieved besides the enhanced scalability when applying intelligent distributed HIP gateways in the mobile network.

5. Application of New Results

Supporting localized or micromobility management, location privacy enhancement, network mobility management and distributed mobility management are very important scenarios for emerging application areas and use-cases in the all-IP world of modern telecommunications. In my dissertation I have presented new schemes, protocols and algorithms to support these scenarios, improve the performance of legacy solutions in these use-cases and such increase the quality of mobile applications in general.

The proposed anycast based micromobility management method and the introduced anycast domain planning scheme would easily represent a convenient hop-by-hop routing based but still scalable technique for mobile operators to deploy transparent and built-in micromobility management. However, my proposal requires the standardization and widespread implementation of IPv6 anycast routing and group management protocols.

The introduced location privacy aware domain planning algorithms would provide an opportunity for mobile operators to configure micromobility domains and define gateway placement policies in distributed architectures in a way which guarantees a near optimum tradeoff between the registration signaling load and the paging cost, while also maximizing the domain’s location privacy capabilities based on different considerations. Domain planning
is an important issue in the design of future’s highly distributed or even flat mobile networks, since IP address changes will occur much frequently in such architectures, therefore structure of domains will show even more serious impacts on IP-dependent location privacy and also on the total mobility management cost of mobile nodes. Of course operators will only apply such techniques if the user awareness for location privacy protection will increase above a certain level, such demanding the application of strong privacy enhancing technologies even during the network planning phase.

My proposed, location information aided predictive mobility management framework and handover scheme extends the standardized NEMO BS solution for network mobility, and combines the benefits of MCoA with a new prediction-driven cross-layer management entity. I have shown that with an appropriate setup the prediction engine will not suffer from the errors of wrong positioning measurements, which makes my proposed system a solid, trustworthy and practical extension of NEMO BS in multihomed configurations. The scheme was successfully applied in the BOSS project [46] and served as the main mobility management solution for the on board wireless secured video surveillance system designed by the BOSS consortium for railway carriages.

The separation of locator and identifier information is probably one of the main evolution trends of the future Internet. The Host Identity Protocol is a security control protocol providing true, cryptographic ID/Loc separation, IP-mobility and multihoming. In HIP-enabled nodes, applications use persistent host identities instead of IP addresses for addressing. Any type of mobility is hidden from the application and transport layer. In current 3GPP networks, non-3GPP access is protected by IKEv2 and IPsec protocols. HIP could replace IKEv2 currently defined for non-3GPP access as network access security protocol, if it performs better in L3 re-authentication and IPsec security association establishment procedures. Seamless inter-system handover between non-3GPP accesses is not covered by current 3GPP standards, however HIP could also support seamless inter-system handover between non-3GPP access networks. In distributed or flat architectures, containing multiple distributed P-GWs, intra-3GPP mobility will lead to frequent inter-PGW handovers. That is the reason why my proposed HIP based micromobility and UFA-HIP distributed mobility could play an essential role in future mobile Internet architectures, and also this is the motivation to provide NEMO support also in the host identity layer by my introduced HIP-NEMO scheme. However, introduction of HIP technologies in current or evolving mobile architectures is not an easy job: the structural modifications inside the common TCP/IP protocol stack raises serious deployment concerns which should be tackled for widespread application of HIP based networking solutions. In case of HIP delegation-based services, support of non-HIP enabled peers can be solved by example using HIP proxies [47]. My HIP-based schemes were successfully applied in project MEVICO [48] as the building blocks of a possible green-field alternative to support mobile networks evolution towards a more distributed architecture and enhanced individual communications experience.

References


Publications

Journal papers


Hungarian journal papers


Book chapters


Conference papers


