Evaluation and Comparison of Signaling Protocol Alternatives for the Ultra Flat Architecture

Zoltán Faigl, László Bokor
Mobile Innovation Centre, BME
Budapest, Hungary
zfaigl@mik.bme.hu
goodzi@mik.bme.hu

Pedro Miguel Neves, Ricardo Azevedo Pereira
PT Innovação
Aveiro, Portugal
pedro-m-neves@ptinovacao.pt
ricardo-a-pereira@ptinovacao.pt

Khadija Daoud, Philippe Herbelin
Orange Labs
Issy Les Moulineaux, France
khadija.daoud@orange-ftgroup.com
philippe.herbelin@orange-ftgroup.com

***Abstract*—The Ultra Flat Architecture is a new concept of fixed mobile convergent networks that aims to scale well with the mobile internet traffic explosion prognosticated for the next 5–10 years. This paper investigates the adequacy of three different signaling protocol alternatives for the Ultra Flat Architecture based on operator requirements. The applied evaluation method is the Multiplicative Analytic Hierarchy Process. After the presentation of the evaluation process, we define our evaluation criteria. It is followed by the presentation of the main features of three UFA signaling protocol alternatives. Finally, the terminal scores of the alternatives are analyzed under different circumstances. The SIP-based alternative shows high performance, and low deployment cost. It is adequate for IMS applications. However by the increase of the demand to support the mobility of legacy internet applications, HIP or MIP-based signaling schemes are more suitable to our criteria. The evaluation shows the effect of the criteria weights and the network scenario on the suitability of the alternatives.**

***Keywords*-suitability analysis; decision making; signaling schemes; Ultra Flat Architecture; fixed mobile convergence**

I. Introduction

With the explosive proliferation of mobile communications and wireless computing devices, the scalability property of mobile networks is becoming an increasingly important feature of wireless communication in pervasive networking scenarios. In 3GPP architectures the IP addresses for users are allocated in high-level network components, such as the GGSN in the UMTS networks [1] or the PDN GW in the System Architecture Evolution (SAE) architecture [2]. Furthermore, there are other intermediate network anchor points, such as the SGSN and RNC for UMTS, and SAE GW for SAE. Each of these anchor points maintain a context per Mobile Node (MN) that binds the MN identity, tunnel identifier, required QoS, etc. Network elements are limited in terms of simultaneous active context. With the continuous growth of the traffic, the adaptation of these architectures to the demands, i.e., deployment of an increasing number of equipments, is challenging for operators. The return on investment of the deployed equipments can not be ensured.

The change of the MN's locators (e.g., the IP address) due to vertical handovers between different access technologies, such as the 3GPP, WiFi, and WiMaX networks, must be handled without the degradation of the quality of experience of the users. Mobile IP protocol (MIP) [3] and Proxy Mobile IP (PMIP) [4] represent alternatives to handle terminal mobility in SAE. These protocols introduce the Home Agent (HA), deployed typically in the core network near the PDN GW, to tunnel user traffic to the current location of the MN. These HAs represent anchor points as well for user traffic, leading to scalability issues. For this reason, protocols that realize end-to-end location updates between the communicating peers, such as the Session Initiation Protocol (SIP) [5], the mobile Stream Control Transmission Protocol (m-STCP) [6], or the Host Identity Protocol (HIP) [7] are preferable for handling location updates. For the IP Multimedia Subsystem (IMS) [8], the Proxy Call Service Control Function (P-CSCF) and the Serving CSCF (S-CSCF) entities are considered as anchors for signaling. They are located in the core network and are in charge of maintaining SIP sessions, security and other contexts for registered users.

To solve scalability issues, bottlenecks from packet communication must be removed by eliminating user-plane anchors from the network and bringing IP routing close to the mobile terminals in means 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. Furthermore, advanced mobility and multihoming scenarios should be supported, such as network mobility and per-application mobility. UFA (Ultra Flat Architecture) is a new flat and fully distributed mobile and convergent architecture that has been introduced in our previous work [9] to solve simultaneously mobile networks and IMS [8] scalability issues. UFA is considered as an ultimate step toward flat architectures as it distributes into a single node, called UFA Gateway (UFA_GW), IMS functions as well as mobile data forwarding and control plane functions. The UFA_GW could be implemented in the base stations (BS) or close to the BS.

The UFA requires the redesign of the terminal attachment, service establishment, and mobility procedures. This paper focuses on the evaluation of three possible signaling schemes providing alternatives for these procedures. In [9], [10] UFA uses SIP as the main signaling protocol to provide QoS integrated establishment and mobility procedures. Even if SIP protocols become important in telecommunications, two other

protocol alternatives are envisaged, i.e., the HIP and the MIP based alternative. They fulfill the need for the support of establishment and mobility procedures for non-SIP applications, i.e., the legacy internet applications such as P2P file sharing, video on demand, or real-time multimedia services over the Internet. This paper briefly introduces the SIP, HIP and MIP based signaling schemes for UFA, and evaluates them using a modified Multiplicative Analytic Hierarchy Process (MAHP). The contribution of this paper is that it shows in an objective way the advantages and disadvantages of each signaling schemes based on a criterion set specified by the designers. We introduce simple but essential measures for analyzing the performance, security, deployment features of the alternatives. We have evaluated the alternatives under twenty sub-criteria. They are grouped under four main criteria, i.e., provide high performance, high security, low deployment cost and support for the mobility of non-SIP applications. The evaluation shows that how do the criteria weights and the packet transmission delay of the access network influence the ranking scores. The evaluation process is general, and can be applied in any other domain where a complex matching problem is defined.

The reminder of this paper is organized as follows. Section II describes related works that show a shift from hierarchical to flat architectures. Section III summarizes the steps of our evaluation process. In Section IV the criterion set is specified under which the alternatives will be evaluated. Section V describes the UFA signaling alternatives and highlights their main aspects regarding the criteria. Section VI presents the network model used for the analysis of the alternatives under different performance criteria. In Section VII the evaluation of the alternatives is presented. Finally, Section VIII concludes the paper and presents our future plans.

II. Related Work

The flat architecture concept for mobile cellular networks is not new. [11] describes that a shift from hierarchical to flat architecture is needed, using one special network entity to provide radio access network functionality, and supplemented by standard IP-based network elements in the core. The paradigm of flat architectures has become an important concept for SAE in the 3GPP architecture evolution. In SAE, the number of functional entities is reduced and restructured compared to the previous 3GPP releases [2]. In Long Term Evolution (LTE) (Release 8) and LTE-Advanced (Release 9), the radio-access network (RAN) is composed only of base stations (BSs, also called eNodeBs) providing the user plane and control plane protocol terminations toward the MN. The BSs are interconnected and directly linked to a gateway managing connections to other IP-based networks. This so-called flat architecture simplifies the user data flow and enables flexible and cost-effective capacity scaling [12], [13]. Previously, the High Speed Packet Access (HSPA) architecture (Release 7) has already made simplifications in the core network, reducing the number of functional elements from four in Release 6 to two in Release 7 [14].

Fixed networks were subject to the same scalability problems. IP connectivity of the users was first provided by centralized equipments. When triple play services took off, network architecture has been modified, pushing the first IP routing function close to the subscribers, e.g., in the Digital Subscriber Line Access Multiplexer (DSLAM) for DSL services [15]. This has solved scalability problems, since central equipments now deal only with the routing or switching of the traffic, furthermore, the number of intelligent edge nodes is mostly influenced by the number of users physically linked to the network. This change reduced the investment costs for the fixed network.

Supporting heterogeneous network layer protocols and different locator families has been addressed in [16], which defines a well scalable architecture for dynamically composable networks. The inseparable bond between the locator and identifier functions of IP address makes it inconvenient or even impossible to design efficient and scalable mobility, multihoming, traffic engineering, routing and security solutions. The general concept of ID/Loc separation aims to eliminate the above problems and limitations by splitting the two roles of IP addresses and such allowing network layer to change locators without interfering with upper layer procedures. The concept gains more and more popularity: several different approaches exist for ID/Loc separation (e.g., LISP [17], SHIM6 [18], FARA [19] or HIP [20]) and it also has recently been introduced in the standardization activities of ITU-T for integration in future network architectures [21], [22]. Some solutions (e.g., [23]) also address scalability issues by eliminating anchor points needed for dynamic mapping and introducing a whole new logical protocol layer as a distributed overlay for translating locators to identifiers.

III. Evaluation Process

Multi-Criteria Decision Analysis is concerned with the evaluation of a finite number of alternatives under a finite number of criteria by a single or a group of decision makers [24]. The Multiplicative Analytic Hierarchy Process (MAHP) [25], used in our study, is a procedure that facilitates the ranking of alternatives by decomposing the complex matching problem to many pairwise comparisons of the alternatives under each criterion.

Fig. 1 illustrates the steps of our evaluation process. Criteria ($C_i, i = 1..M$) mean the requirements under which we evaluate different alternatives. The alternatives ($A_j, j = 1..N$) in this study mean the candidate UFA signaling schemes. The evaluation process results in the terminal scores t_j of the alternatives. Three possible ways are shown in the figure for the specification of the decision maker's input. In case A, the performance metrics of the alternatives are calculated using our network model, as described in Section VI. Some attributes, i.e., security and deployment features of the alternatives are directly assigned, represented by case B. Case C is used for criterion weight calculation.

In decision making procedures, geometric scale is often used in order to approximate the categorization of comparative

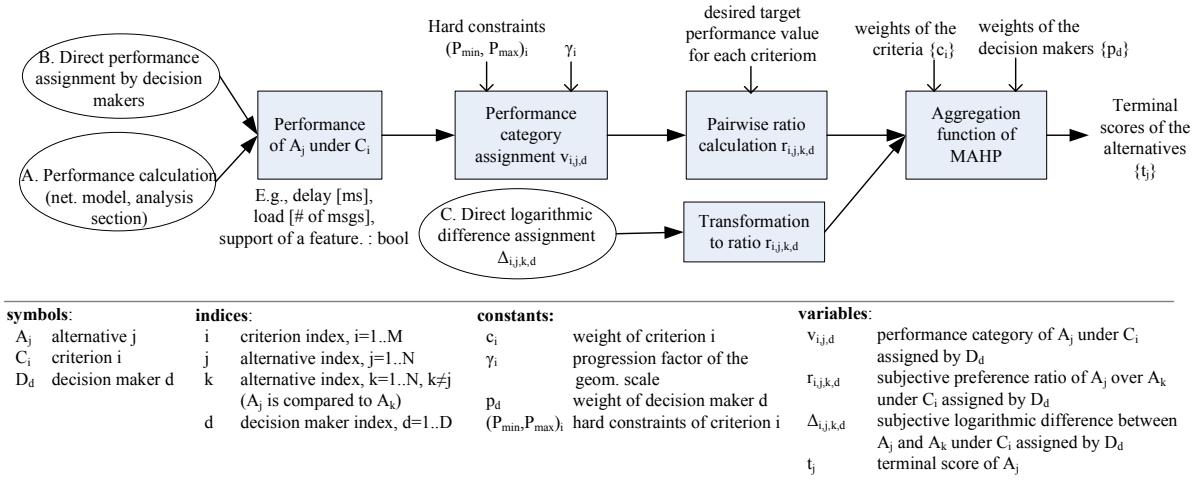


Fig. 1. Evaluation process

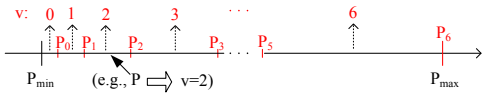


Fig. 2. Comparative human judgment approximated with a geometric scale.

human judgment. The subjective preference is represented with finite integer grades, instead of continuous values. Fig. 2 illustrates such an assignment of categories to real performance values.

Acceptable performance values of alternatives under a criterion C_i form a closed continuous or discrete interval, with borders $P_{min,i}$ and $P_{max,i}$. We call these borders as hard constraints, i.e., if an alternative performs outside the interval, then it will get zero for terminal score at the end of the process. If A_j performs inside the interval of C_i, then we assign a performance category $v_{i,j,d}$ to the alternative using a seven-grade geometric scale that covers this interval. The progression factor of the scale is 2^γ, which gives the ratio between two consecutive steps of the scale. For each criterion, Eq. 1 or 2 were applied for performance category assignment, resulting in a seven-grade scale starting at P_{min} or P_{max}, respectively.

$$v_{i,j,d} = \left\lceil 6 + \frac{1}{\gamma}\log_2 \frac{P_{i,j,d} - P_{min,i}}{P_{max,i} - P_{min,i}} \right\rceil \tag{1}$$

$$v_{i,j,d} = \left\lceil 6 + \frac{1}{\gamma}\log_2 \frac{P_{max,i} - P_{i,j,d}}{P_{max,i} - P_{min,i}} \right\rceil \tag{2}$$

Note, that Eq. 1 results a scale represented in Fig. 2 with starting point P_{min}.

MAHP aggregation function needs as input the subjective preference ratios of alternative A_j over A_k under criterion i defined by decision maker D_d. These ratios represent that how much better is A_j than A_k. For each criterion we can define high or low performance values as the desired target values. E.g., for handover delay or network overhead the desired performance value is $P_{min,i} = 0$ second, however, for the existence of security features represented by booleans, the desired value is the high value, i.e., support of a given security service ($P_{max,i} = 1$). Hence, the performance category values on the scale might increase or decrease toward the desired target. Eq. 3 and Eq. 4 present the calculation of the subjective preference ratios for the two cases, respectively. $\Delta_{i,j,k,d}$ is the subjective logarithmic difference between A_j and A_k under C_i defined by D_d. Hence, four types of preference ratio calculations are possible with geometric scales, given by the combinations of Eq. 1 or Eq. 2 and Eq. 3 or Eq. 4.

$$r_{i,j,k,d} = 2^{\gamma(v_{i,j,d} - v_{i,k,d})} = 2^{\gamma\Delta_{i,j,k,d}} \tag{3}$$

$$r_{i,j,k,d} = 2^{\gamma(v_{k,i,d} - v_{j,i,d})} = 2^{-\gamma\Delta_{i,j,k,d}} \tag{4}$$

Given the subjective preference ratios of the decision makers for all alternative pairs under a subset of criteria, the terminal score of an alternative is calculated with the aggregation function depicted in Eq. 5. The weights of the criteria (c_i), and the weights of the decision makers (p_d) are used as inputs.

$$t_j = \prod_{k=1}^{N}\prod_{i=1}^{M}\prod_{d=1}^{D} r_{i,j,k,d}^{p_d c_i / N} \tag{5}$$

IV. Decision Criteria

Our objective was to find the most relevant criteria for the evaluation of the alternatives. Good criteria are the ones that show well the costs and benefits of the alternatives regarding the objectives. We defined four main criteria, i.e., provide (1) high performance, (2) high security, (3) low deployment cost, and (4) support for the mobility of non-SIP applications. There are fourteen sub-criteria under them, organized in a decision tree, illustrated in Fig. 3.

The criteria weights were defined in an iterative process, based on the consensus of five decision makers using direct logarithmic difference assignment (as depicted by case C in Fig. 1). In this process, decision makers assign subjective difference values ($\Delta_{j,k,d} = \{-6, .., -2, -1, 0, 1, 2, .., 6\}$) between criteria j and k, $j = 1..M-1, k = j..M$ for each

	Criterium weights	Measurement units	Hard constr. Pmin, Pmax	γ	Scale type [# of Eq]
Objective: rank UFA signaling protocols	1	-	-	-	-
Low performance cost	0.2894	-	-	-	-
Minimize real-time service session interruption delay	0.5665	ms	[0,200]	0.5	1 & 4
Low handover preparation and execution overhead	0.2832	# of msgs.	[0,25]	0.5	2 & 3
Low session establishment procedure overhead	0.0795	# of msgs.	[0,25]	0.5	2 & 3
Low network attachment procedure overhead	0.0708	# of msgs.	[0,25]	0.5	2 & 3
High security	0.2107	-	-	-	-
Must support mutual authentication	0.3224	boolean	[1,1]	0.5	2 & 4
Must protect signaling traffic bw. MN and UFA_GW	0.2395	boolean	[1,1]	0.5	2 & 4
Must protect data traffic bw. MN and UFA_GW	0.2169	boolean	[1,1]	0.5	2 & 4
May resist DoS from MNs to the UFA_GWs	0.1322	boolean	[0,1]	0.5	2 & 4
May resist MiTM attacks	0.0890	boolean	[0,1]	0.5	2 & 4
Easy deployment	0.2418	-	-	-	-
Low complexity of the MN on signalling plane	0.5802	# of modules	[0,6]	0.5	2 & 3
Low complexity of the MN on user plane	0.3256	# of modules	[0,6]	0.5	2 & 3
Low complexity of the UFA_GW on signalling plane	0.0228	# of modules	[0,6]	0.5	2 & 3
Low complexity of the UFA_GW on user plane	0.0256	# of modules	[0,6]	0.5	2 & 3
Low amount of changes in the IMS core components	0.0457	# of modules	[0,6]	0.5	2 & 3
Support non-SIP based applications	0.2580	boolean	[0,1]	0.5	2 & 4

Fig. 3. Criterion tree.

branch of the tree. The transformation of the logarithmic differences to subjective preference ratios can be calculated with Eq. 3. $\gamma = 1$ was used for each criterion. For this γ, if $\Delta_{j,k,d} = 6$ then C_j is 64 times more important than C_k in D_d's opinion. Given the subjective preference ratios, we can calculate the terminal scores of the criteria with the aggregation function in Eq. 5. Finally, the normalized weights (c_j) are calculated with Eq. 6.

$$c_j = \frac{t_j}{\sum_{j=1}^{M} t_j} \tag{6}$$

The obtained criteria weights are presented in the first column of Fig. 3. As it can be seen, the dominating sub-criteria under the four main criteria are the low real-time service interruption delay due to vertical handovers; the mutual authentication, signaling and user data protection between the MN and the UFA_GW; the low number of additional modules to deploy in the MN in the control plane and user plane; and the support for non-SIP applications, respectively. Each message overhead sub-criterion under the performance criterion contains five sub-sub-criteria that is not illustrated in the figure. They evaluate the message overheads of the alternatives in different parts of the networks as given later in the first column of Tab. III. The message overhead criterion for the network part between the MN and the UFA_GW (i.e., part I and VII) dominates the five sub-sub-criteria. The criterion tree presents the hard constraints for every criterion, which was defined based on specifications, such as the maximum real-time service interruption delay, or by the authors. The type and progression factor of the scales were carefully defined in order to reflect quality of experience differences with different categories.

Some of the main requirements of the UFA, such as high scalability, self-configuring and self-optimizing network were not included due to the following reasons. Good scalability is

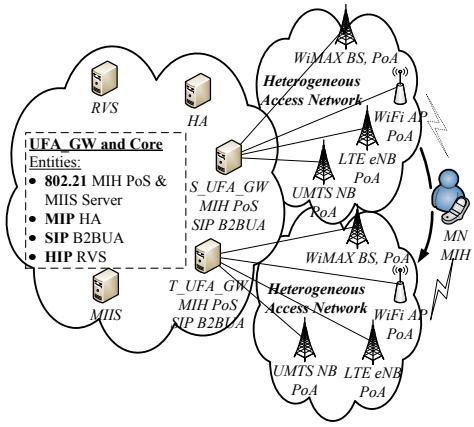


Fig. 4. Ultra Flat Architecture.

mainly ensured by the structure of the architecture, and influenced by the user traffic. Hence, the signaling schemes would not make much difference under the scalability criterion. The other previously mentioned criteria are difficult to evaluate in this phase of the project.

V. UFA Signaling Alternatives

Fig. 4 illustrates the UFA reference scenario. Each signaling scheme makes use of the 802.21 Media Independent Handover (MIH) framework [26] that provides procedures for carrying out network controlled pro-active handovers over multiple radio access technologies, such as WiFi, WiMax, or 3GPP access. S_UFA_GW and T_UFA_GW are the source and target UFA_GWs, respectively, between which the MN performs the handover. When the wireless signal decreases below a threshold level, the MN signals the S_UFA_GW, i.e., the MIH Point of Service (PoS), to prepare the handover to a new access network, i.e., the candidate MIH Point of Access (PoA). The S_UFA_GW discovers accessible PoAs and queries about their available resources for the MN's sessions. The decision on the target PoA and the T_UFA_GW is made by the UFA_Cross_Layer (UFA-cl) module located at the S_UFA_GW. The S_UFA_GW allocates the required Layer-2 (L2) resources at the selected PoA. After these steps, the network prepares itself for the L3 handover of the MN using the handover execution procedures detailed in the following subsections. After the contexts are established at the T_UFA_GW, the MN and its Correspondent Nodes (CNs), the S_UFA_GW sends a MIH Net Handover Commit request to the MN, which triggers the physical L2/L3 handover. After successful handover the S_UFA_GW is notified by the MIH completion messages to release the network resources maintained for the MN.

A. SIP Based Alternative

For SIP based alternative, referred to as SIP, UFA proposes to distribute the first IP router, PCRF and P-CSCF functions of IMS in the UFA_GW. Nevertheless, the SIP layer acts as a Back-To-Back User Agent (B2BUA), named U-CSCF, in order to control SIP messages and perform their modification

according to the operator policies. The authentication and key agreement of the MN to the network is based on IMS AKA [27]. Thanks to this control and because the UFA_GWs simultaneously manage the SIP layer and all the network resource information, adapting applications to resource availability become possible for session establishment procedures, considering QoS constraints [10]. Handover execution of a MN from a S_UFA_GW to a T_UFA_GW is based on the SIP re-INVITE procedures. They transfer IPsec contexts as well from the S_UFA_GW to the T_UFA_GW.

B. HIP Based Alternative

HIP based alternative, referred to as HIP, requires the deployment of HIP naming service in the UFA core network, e.g. a HIP-capable DNS (hDNS) and HIP Rendezvous Service (RVS). It brings cross-layer HIP modules in the UFA_GWs, MNs and Correspondent Nodes (CNs). HIP Base Exchange (BEX) and Update procedures deal with dynamic negotiation of IPsec security associations between the MN and the UFA_GW to protect user data and mutually authenticate the MN and the network. The handover execution procedure is started by the S_UFA_GW. HIP and IPsec contexts are established between the T_UFA_GW and the MN's CNs, furthermore, between the T_UFA_GW and the MN, using the mediation of the S_UFA_GW. This is possible due to the delegation of HIP signaling rights from the MN and from the T_UFA_GW to the S_UFA_GW [28]. Context Transfer Protocol [29] is used to transfer the HIP and IPsec contexts from the S_UFA_GW to the T_UFA_GW and the MN. As the contexts are in their place the MN is notified by MIH Net commit request to attach to the new PoA. A detailed description of the HIP based UFA signaling scheme is given in [30].

C. MIP Based Alternative

For the MIP based alternative, referred to as MIP, the MIP HA is located in the UFA core network. Before the physical handover, the MN obtains a Care-of-Address (CoA) from the T_UFA_GW via the S_UFA_GW with a MIH Link action request message, and registers it on the MIP HA with the Binding Update process. As a result, an IP-in-IP tunnel is configured between the HA and the MN target access interface to transfer the downlink data packets. Therefore, the packets sent by the CN to the MN home address are intercepted by the HA and thereafter sent toward the MN CoA through the MIP tunnel. For route optimization, immediately after receiving a downlink packet of the CN, the MN updates the CN binding cache with the new CoA. Thereafter packets are sent directly between the MN CoA and the CN, without being intercepted by the HA. Mutual authentication and security association negotiation between the MN and the UFA_GW is performed by IKEv2 [31] protocol using the EAP-AKA method. The update of IPsec security associations between the MN and the UFA_GW is performed with MOBIKE [32].

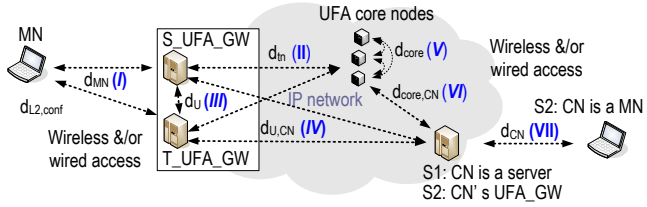


Fig. 5. Network model.

D. Properties of the Alternatives

In this part we summarize the properties of the alternatives regarding the criteria. Easy deployment is one of the main criteria. We introduced the following measure for deployment complexity. Assuming a default TCP/IP stack in each UFA node, which supports MIP and IPsec, we calculated the surplus protocol modules in user and control plane in the MNs, UFA_GWs, and UFA core network. The subjective preference ratios of the alternatives under the deployment criteria were calculated using Eq. 2 and 3. Hence, the linear increase of additional modules causes an exponential decrease of performance categories. Tab. I summarizes the modules needed in the UFA nodes for the different signaling schemes.

The security features of the alternatives are summarized in Tab. II. Security criteria are measured with performance values 1 for support and 0 for no support of the required security service. SIP, HIP and MIP alternatives are based on different authentication and key agreement protocols given in the table. All alternatives use IPsec tunnels for the protection of signaling and user traffic between the MN and the UFA_GW.

The performances of the alternatives under the low message overhead criteria are measured with the number of required messages for one successful network attachment, session establishment and handover, respectively, at different parts of the network. The performance of the signaling schemes under these sub-criteria is detailed in Tab. III. The network parts are illustrated in Fig 5.

The performance of the alternatives under the low service interruption delay criterion is calculated by summing the transfer delays of the messages causing real-time session interruption, and the L2 configuration delay of the MN. The performance of the alternatives under the criterion for the support of non-SIP applications is measured using 0 for no support and 1 for support. The SIP, HIP, and MIP based signaling schemes get 0, 1, 1 performance values under this criterion, respectively.

VI. Network Model

The analysis of the alternatives under performance criteria is based on the network model presented in Fig. 5. It is composed of the MN, the UFA_GWs in charge of MN or CN attachment and the UFA core nodes. We suppose in our model that all UFA equipments are linked to a full mesh IP network. We analyzed two different scenarios, i.e., in the first scenario (S1) the CN is a server, e.g., an IMS IPTV server or a HTTP server, in the second scenario (S2) the CN is a MN attached to an

TABLE I
SURPLUS PROTOCOL MODULES FOR THE ALTERNATIVES BEYOND THE TCP/IP STACK.

	SIP	HIP	MIP
MN cp.	SIP, 802.21, UFA-cl	HIP, SIP, 802.21, UFA-cl, uHIP	IKEv2, SIP, 802.21, UFA-cl
MN up.	SIP-based applications	–	–
UFA cp.	U-CSCF, 802.21, UFA-cl	HIP, 802.21, U-CSCF, UFA-cl, uHIP	IKEv2, 802.21, U-CSCF, UFA-cl
UFA up.	–	–	–
IMS nodes	S-CSCF, I-CSCF, HSS MIIS	S-CSCF, I-CSCF, HSS hDNS , RVS, MIIS	S-CSCF, I-CSCF, HSS HA, MIIS

cp. – control plane, up. – user plane; U-CSCF – SIP B2BUA, UFA-cl: UFA cross-layer communication module
uHIP – UFA_GWs perform HIP Updates from MNs to CNs; hDNS – DNS extended with name to HIT resolution

TABLE II
SECURITY FEATURES OF THE ALTERNATIVES.

	SIP	HIP	MIP
mutual authentication	IMS AKA	HIP BEX	IKEv2
signaling protection on wl.	IPsec	IPsec	IPsec
user data protection on wl.	IPsec	IPsec	IPsec
DoS resistance	false	true	true
MiTM resistance	true	true	true

wl.: network part bw. the MN and the UFA_GW, it may contain wireless link

TABLE III
NUMBER OF MESSAGES FOR CONTROL PLANE PROCEDURES.

Network part	SIP	HIP	MIP
Message numbers during service interruption			
d_{conf}	1	1	1
I,VII	1	2	2
Handover preparation and execution			
I,VII	10	11	15
II,VI	10	7	10
III	6	10	4
IV	0	3	2
Session establishment			
I,VII	5	7	11
II,VI	10	12	13
IV	0	3	4
V	8	8	9
Network attachment			
I,VII	6	11	18
II,VI	4	8	10
V	12	6	6

UFA_GW. The parameters of the model are summarized in Tab. IV.

VII. EVALUATION AND COMPARISON OF THE ALTERNATIVES

During the evaluation of the session establishment and handover procedures of the alternatives, we fixed the number of concurrent application level sessions to one, between the MN and the CN. We supposed to have a HTTP page request and a SIP-based VoIP session establishment for scenarios S1 and S2, respectively.

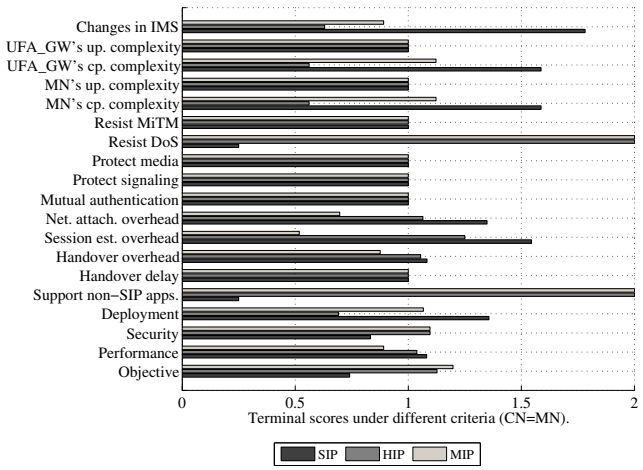
Fig. 6 shows the terminal scores of the alternatives under the main and sub-criteria for both network scenarios. The results differ only under the criterion for low message overhead during session establishment. The weights of the criteria and the network parameters were fixed as given in Fig. 3 and Tab. IV. The overall terminal scores of the signaling schemes can be seen under the objective criterion: MIP is followed by HIP, and SIP in decreasing order. The terminal score of SIP

TABLE IV
NETWORK PARAMETERS.

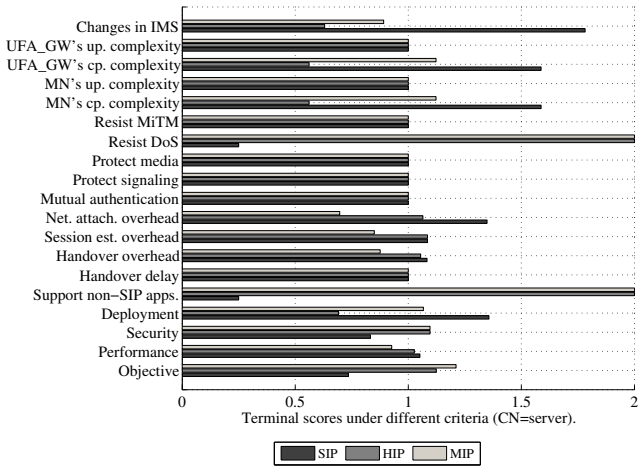
Symbol	Delay [ms]	Description
d_{MN}	[1..100] 10	access network delay bw. MN and Source or Target UFA_GW
d_{CN}	[5..100] 10	access network delay bw. CN and UFA_GW ($d_{MN} = d_{CN}$)
d_{tn}	[5..50] 30	IP network delay bw. UFA_GW of the MN and the UFA core nodes
d_{core}	10	delay bw. the UFA core nodes
d_U	20	delay bw. Source and Target UFA_GW
$d_{core,CN}$	S1:d_{core} S2:d_{tn}	delay bw. UFA core elements and the UFA_GW of the CN
$d_{U,CN}$	S1,S2:d_{tn}	delay bw. the Source or the Target UFA_GW and the CN (or CN's UFA_GW)
$d_{L2,conf}$	50	L2 configuration delay of the MN during access to a new UFA_GW

is low because of the relatively high weight of the support non-SIP applications criterion. The terminal scores of HIP and MIP are influenced heavily by the deployment criteria. We can see that the HIP alternative performs bad under the sub-criteria for low deployment cost. This is due to the high number of additional modules required by HIP. SIP performs very good, and MIP is between the two alternatives under the deployment criteria. Under the security sub-criteria, HIP and MIP perform better than SIP, because IKEv2 and HIP have built-in DoS resistance features, based on cookies and puzzle-challenges, respectively. The IMS AKA does not mitigate DoS attacks in case of the SIP alternative. The message overload of SIP is the smallest for all procedures. It is followed by HIP and MIP, respectively. The explanation is that security control, IMS registration, VoIP session establishment, location updates, and context transfers are highly integrated within the SIP messages. MIP and HIP must execute their own establishment and location update procedures, besides the application level registration and session establishment procedures. Note, that HIP and MIP become more effective than SIP as soon as multiple applications must be updated at the same CN. While SIP provides location updates in an end-to-end fashion in the application level, the other two alternatives need only to execute the location updates once per CN. Every alternative performs equally under the low service interruption delay criterion.

Fig. 7 illustrates the sensitivity of terminal scores of the alternatives to the transmission delay (d_1) of the network part between the MN and its UFA_GWs (parts I and VII in Fig. 5). Fig. 7a presents the terminals scores of the alternatives

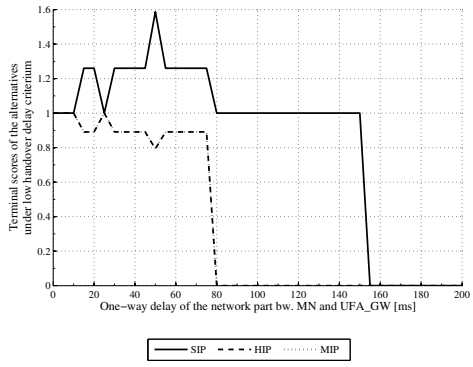


(a) CN is a MN.

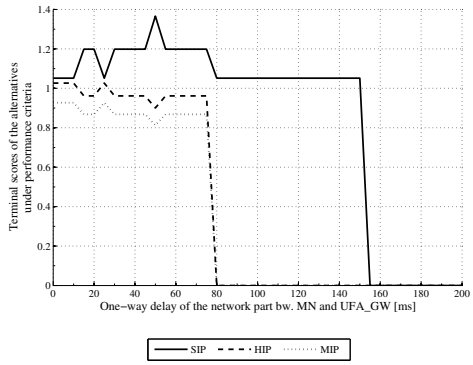


(b) CN is a server in the core network.

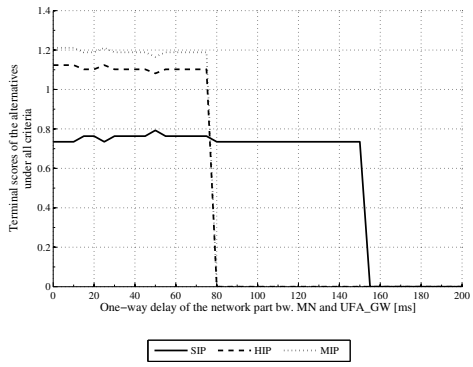
Fig. 6. Terminal scores of the alternatives under fixed network parameters specified in Tab. IV.



(a) Terminal scores under the criterion for low service interruption time during handovers.



(b) Terminal scores under the performance criteria.



(c) Overall terminals scores.

Fig. 7. Sensitivity of the ranking of alternatives to d_1 delays.

under the criterion for low service interruption delay. It heavily influences the terminal scores under the performance criteria in Fig. 7b, and the overall terminal scores in Fig 7c. If the service interruption delay of an alternative gets higher than the maximum hard constraint ($P_{max} = 200ms$), the alternative fails, and is assigned with zero terminal scores. HIP and MIP perform the same under the low service interruption delay criterion. They exceed the maximum hard constraint at $d_1 > 65$ms. SIP performs in the acceptable interval if $d_1 < 150$ ms.

Fig. 8 presents the sensitivity of the ranking of the alternatives to the variation of the weights of the main criteria under fixed network parameters that are given in Tab. IV. For all evaluations, while the weight c_i of the perturbed criterion is increased from zero to one, the weights of the other three main criteria are decreased in a proportional way, so that the sum of the main criteria weights is one. The differences between the gradients of the curves show that how the importance of a criterion influence the judgment of an alternative. SIP is prominent under the performance and the deployment criteria as depicted in Fig. 8a and Fig. 8c. The security criteria do not influence the judgment of the alternatives, in Fig. 8b. SIP's preference decreases as the importance of non-SIP application support increases in Fig. 8d. HIP is better in performance than MIP, in Fig. 8a, but much worse under the low deployment cost criterion in Fig. 8c. They become suitable as the importance to support non-SIP applications increases in Fig. 8d.

VIII. Concluding Remarks and Future Work

In this paper the suitability of three different signaling schemes was analyzed for the UFA. The alternatives are mainly based on the SIP, HIP and MIP protocols, and use

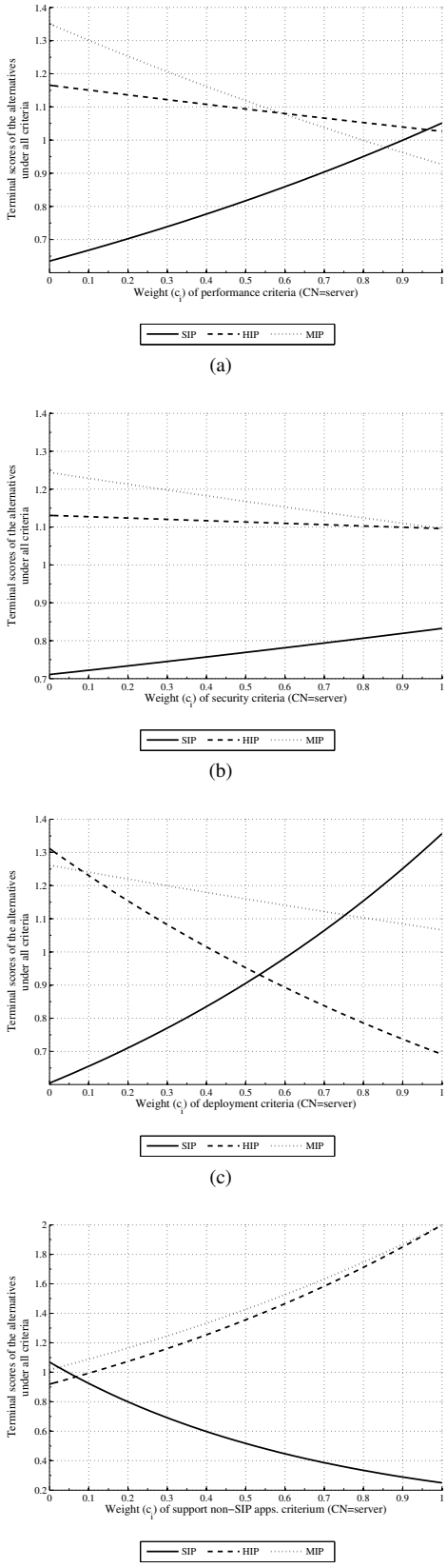


Fig. 8. Terminal scores of the alternatives under perturbed criteria weights.

the 802.21 MIH protocol for handover preparation, in order to provide proactive, network controlled handovers. In this phase of the project, criteria for high performance, high security, low deployment cost and the support of the mobility of non-SIP applications were interesting. Even if the measurement of some criteria was simplified, the evaluation had the benefit of seeing clearer the objectives, consider systematically the features of the alternatives, and discover their deficiencies. The evaluation also shows that criteria weights and network parameter values highly influence the decision. The results show that the SIP-based alternative performs very good under performance and deployment criteria. HIP and MIP perform slightly better than SIP under the security criteria. HIP performs better in performance than MIP. It may be a good direction to use the combination of these signaling protocols, e.g., SIP and HIP or SIP and MIP, but the combined scenarios should be evaluated. We expect that they will perform better than the analyzed alternatives, since non-SIP applications will be supported, and the advantages of SIP still can be used for IMS applications. However, their deployment complexity is higher. We plan to investigate the adequacy of these schemes to support per application mobility scenarios, and requiring simultaneous maintenance of SIP and non-SIP sessions.

ACKNOWLEDGMENT

Ultra Flat Architecture is being studied in P1857 Eurescom project and interests France Telecom, Portugal Telecom, Mobile Innovation Centre Hungary.

REFERENCES

[1] 3GPP, "General Packet Radio Service (GPRS); Service description," TS 23.060, Sep. 2009.

[2] ——, "3GPP System Architecture Evolution: GPRS enhancements for LTE access," TS 23.401, Sep. 2009.

[3] D. B. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," RFC 3775, June 2004.

[4] S. Gundavelli *et al.*, "Proxy Mobile IPv6," RFC 5213, Aug. 2009.

[5] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," RFC 3261, Jun. 2002.

[6] S. J. Koh and S. W. Kim, "mSCTP for Vertical Handover Between Heterogeneous Networks," in *Web and Communication Technologies and Internet-Related Social Issues - HSI 2005*. Springer, 2005, vol. 3597, pp. 28–36.

[7] P. Nikander, T. Henderson, C. Vogt, and J. Arkko, "End-Host Mobility and Multihoming with the Host Identity Protocol," RFC 5206, Apr. 2008.

[8] 3GPP, "IP Multimedia Subsystem (IMS)," TS 23.228, Sep. 2009.

[9] K. Daoud, P. Herbelin, K. Guillouard, and N. Crespi, "Performance and Implementation of UFA: a SIP-based Ultra Flat Mobile Network Architecture," in *Proceedings of PIMRC 2009*, Tokyo, Japan, September 13–16, 2009, pp. 1–6.

[10] K. Daoud, P. Herbelin, and N. Crespi, "One-Node Based Mobile architecture for better QoS control," in *Proceedings of the 1st IFIP Wireless Day Conference 2008*, Dubai, UAE, November 24–27, 2008, pp. 1–5.

[11] P. Bosch, L. Samuel, S. Mullender, P. Polakos, and G. Rittenhouse, "Flat cellular (umts) networks," in *Proceedings of WCNC 2007*, Hong Kong, March 11–15, 2007, pp. 3861 –3866.

[12] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description," TS 36.300, Sep. 2009.

[13] A. Osseiran *et al.*, "The road to imt-advanced communication systems: state-of-the-art and innovation areas addressed by the winner+ project," *Comm. Mag.*, vol. 47, no. 6, pp. 38–47, 2009.

[14] H. Holma *et al.*, "High-speed packet access evolution in 3gpp release 7 [topics in radio communications]," *Comm. Mag.*, vol. 45, no. 12, pp. 29 –35, 2007.
[15] D. Forum, "Migration to Ethernet-Based DSL Aggregation," TR-101, Apr. 2006.
[16] S. Schmid *et al.*, "TurfNet: An Architecture for Dynamically Composable Networks," in *Autonomic Communication*. Springer, 2005, vol. 3457, pp. 94–114.
[17] D. Farinacci *et al.*, "Locator/ID Separation Protocol (LISP)," IETF Draft, Jan. 2010.
[18] E. Nordmark and M. Bagnulo, "Shim6: Level 3 Multihoming Shim Protocol for IPv6," RFC 5533, Jun. 2009.
[19] D. Clark *et al.*, "FARA: reorganizing the addressing architecture," *ACM SIGCOMM Computer Communication Review*, vol. 33, no. 4, pp. 313 – 321, 2003.
[20] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson, "Host Identity Protocol," RFC 5201, Apr. 2008.
[21] ITU-T, "General requirements for ID/locator separation in NGN," ITU-T Draft Recommendation, ITU-T Y.2015 (Y.ipsplit), Feb. 6, 2009.
[22] V. P. Kafle, H. Otsuki, and M. Inoue, "An ID/Locator Split Architecture for Future Networks," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 138–144, 2010.
[23] A. Gurtov *et al.*, "Hi3: An efficient and secure networking architecture for mobile hosts," *Journal of Computer Communications*, vol. 31, no. 10, pp. 2457–2467, 2008.
[24] F. A. Lootsma, *Multi-criteria decision analysis via ratio and difference judgement*, ser. Applied Optimization. Dordrecht: Kluwer Academic, 1999, vol. 29.
[25] ——, "Scale sensitivity in the multiplicative AHP and SMART," *Journal of Multi-Criteria Decision Analysis*, vol. 2, no. 2, pp. 87–110, 1993.
[26] IEEE, "IEEE Standard for Local and metropolitan area networks- Part 21: Media Independent Handover," IEEE Std 802.21-2008, January 21 2009.
[27] 3GPP, "3G security; Access security for IP-based services," TS 33.203, Sep. 2009.
[28] P. Nikander and J. Arkko, "Delegation of Signalling Rights," in *Security Protocols*, ser. Lecture Notes in Computer Science, B. Christianson, B. Crispo, J. A. Malcolm, and M. Roe, Eds. Springer, 2004, vol. 2845, pp. 575–586.
[29] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli, "Context Transfer Protocol (CXTP)," RFC 4067, Jul. 2005.
[30] L. Bokor, Z. Faigl, and S. Imre, "A Delegation-based HIP Signaling Scheme for the Ultra Flat Architecture," in *Proc. of IWSCN'10*, Karlstad, Sweden, May 26–28, 2010, pp. 1–8.
[31] C. Kaufman, "Internet Key Exchange (IKEv2) Protocol," RFC 4306, Dec. 2005.
[32] P. Eronen, "IKEv2 Mobility and Multihoming Protocol (MOBIKE)," RFC 4555, Jun. 2006.