In Proceedings of 12th Annual Meeting of the IEEE / ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Volendam, Netherlands, October 2004

A Performance Study of Session State Re-establishment Schemes in IP-based Micro-mobility Scenarios

T. Chen, G. Schäfer, A. Wolisz
Telecommunication Networks Group
Technische Universität Berlin, Germany
Email: [chen, schaefer, wolisz]@tkn.tu-berlin.de

M. Sortais
Mathematisches Institut, Fak. II MA 7-4
Technische Universität Berlin, Germany
Email: sortais@math.tu-berlin.de

***Abstract*— In order to support seamless handovers in IP-based micro-mobility scenarios, updating a mobile user's location information with an optimized mobility management scheme is necessary. In such a re-registration procedure, an appropriate Quality-of-Service (QoS) provisioning is required to meet end users' expectations. Security measures designed to protect network infrastructure and user data are also inevitable. Some proposals for the seamless handover support have been published and some analytic work on the their performance has been done. However, no proposal has given a complete re-registration procedure considering the security checks, QoS provisioning and mobility management. In this paper, we propose and compare several schemes for a complete re-registration procedure taking these three kinds of operations into account, analyzing their performances in terms of registration latency, packet loss, and CPU processing load at certain nodes.**

I. Introduction

In IP based micro-mobility scenarios, a mobile node (MN) can move freely from one access network to another within a domain while keeping a session alive. It may request the same Quality of Service (QoS) for the active session on the new routing path as he had on the old path. In principle, when the access network receives the request, an authentication check must be taken on it first. Only authenticated re-registration requests can be processed further for the mobility management and the session state re-establishment on the new path.

The authentication check can be performed either by a local security authority (i.e. the local AAA (authentiation, authorization and accounting) server, termed as AAAL) or by an access router (AR) in the access network. In case AAAL performs the authentication, it communicates with an AR by using an AAA protocol; otherwise, AR may get the security information from the old AR by using the Context Transfer (CT) protocol [10], or it may verify a cookie presented in the re-registration request [3].

Mobile IPv4 [13] and Mobile IPv6 [5] were proposed as the main protocols to support IP mobility. To support seamless handovers in IP mobility scenarios, many protocols have been proposed. Fast Handover for Mobile IPv6 (FMIPv6) [8] was designed to minimize handover latency and prevent the degradation of QoS. To minimize the mobility signaling costs outside the access network, Hierarchical Mobile IPv6 (HMIPv6) [14] introduced a local entity (i.e. MAP) for the local mobility management. It was also proposed to integrate FMIPv6 with HMIPv6 in order to minimize the registration latency.

During such a handover, in addition to updating the location information of the MN, QoS information needs to be re-established on the new path, since an application such as VoIP typically requires some Quality of Service support from the new path (i.e. IntServ), which reserves desirable forwarding treatment to certain distinguished packet streams; or it requires enforcement of a policy at the access router to enable the DiffServ Support.

Several analytical works dealing with the performance of mobility management protocols and seamless handover schemes (e.g. Fast Handovers or Context Transfers) have been done. However, no analysis has been carried out for a complete re-registration procedure in a micro-mobility scenario considering the security checks together with mobility management and QoS provisioning.

In this paper, we propose and compare several schemes for a re-registration procedure featuring security checks, QoS provisioning and mobility management, analyzing their performance in terms of registration latency, packet loss, CPU processing load at certain nodes.

II. Related Work

Some analytical studies were carried out for IP-based micro-mobility management. Xie et al. proposed an analytic model for Mobile IP regional registration [16], focusing on determination of the optimal size of regional networks. Woo also proposed an analytic model to investigate the performance of regional registration on registration delay and CPU processing overheaded at mobility nodes [15]. Pack presented a performance study of Hierarchical Mobile IP update and packet delivery costs using random-walk and fluid-flow mobility models [11]. These works focus only on mobility management without taking QoS and security into account.

There have also been analytic works investigating fast handovers and context transfer. Koodli proposed a scheme en-

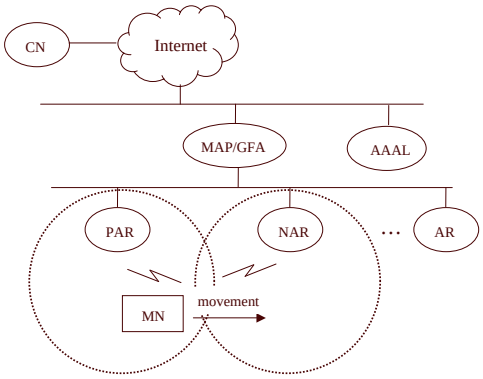

Fig. 1. The Topology for the Analysis

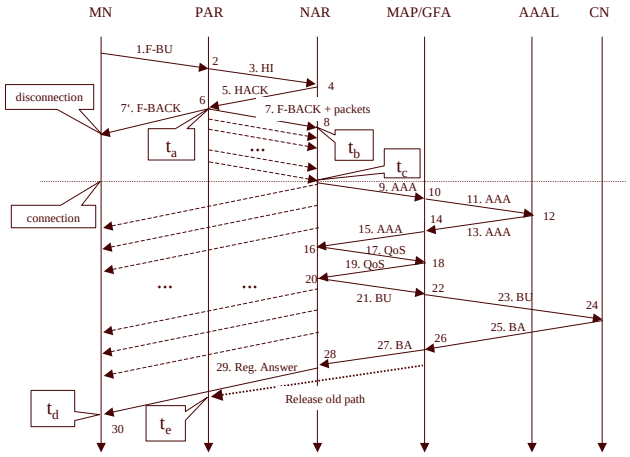

Fig. 2. Signaling Exchanges of Scheme 1

abling an integration of Fast Handovers and Context Transfer in [9]. His performance evaluation showed the benefits of this proposal in achieving seamless handovers. However, for QoS mechanisms that are end-to-end, transferring context at the last hop router may be insufficient to re-initialize completely the mobile host's QoS treatment, since a number of additional routers in the path between the mobile host and corresponding node may also be involved [6]. Therefore, a mechanism to set up a QoS path is still unavoidable, even when CT is used. The performance, including QoS path re-establishment in the registration procedure is unkown.

Costa provided quantitative results for the performance improvements obtained by Hierarchical Mobile IPv6, fast handovers for Mobile IPv6 and their combination in [12]. He considered neither the re-establishment of QoS and security state information, nor the overall performance in a secure, QoS-aware registration procedure.

III. Scheme Description

The topology of the micro-mobility scenario is shown in Figure 1.

An Ethernet bus connects an access network and the core network. In the access network, a Gateway Foreign Agent (GFA) connects ARs via a local bus. GFA acts as mobile anchor point (MAP) in the Hierarchical Mobile IPv6 infrastructure. Each AR dominates a cell receiving registration requests from a mobile user via the wireless interface. Local AAA server (AAAL) is the security authority in the access network, responsible for local authentication and authorization (AA). Any AA check request messages from access routers are sent to AAAL via GFA.

We first work out the possible schemes. In case FMIP is being used as mobility management protocol, either AAA or CT can be used for the authentication check, and they can also serve for the authorization check and DiffServ policy deployment. Therefore, by combining them with a QoS signaling (termed as QoS), we obtain Schemes 1 and 2: $FMIP + AAA + QoS$ and $FMIP + CT + QoS$.

Similarly, when "FMIP+HMIP" is used as mobility management protocol, Schemes 3 and 4 are $FMIP + HMIP + AAA + QoS$ and $FMIP + HMIP + CT + QoS$. When HMIP is used alone for the mobility management, it can be combined with AAA to form Scheme 5. Since it is not optimized to use CT without FMIPv6, HMIPv6 can be combined with a cookie-based approach [2], and this defines Scheme 6.

We take one of the possible protocol runs from the Fast Handovers for Mobile IPv6 mechanism [8] as the FMIP operation in the analysis. That is, MN is assumed to send first a fast Binding Update(F-BU) message when the protocol starts running. MN is also assumed to be able to receive the fast Binding Acknowledgement (F-BACK) before losing its connection to the previous AR (PAR). We hold the same assumptions for all six schemes.

A. Scheme 1: FMIP + AAA + QoS

As shown in Figure 2, when MN receives advertisements from an AR and notices that the signal strength of the advertisement messages is better than the current one, it decides to switch its link connection. Based on the network information in the advertisements, MN obtains a new CoA (NCoA), while still being connected to the previous AR (PAR). It sends a F-BU message to PAR. PAR may first verify the authenticity of the handover request. Then PAR sends a Handover Initiate (HI) message to NAR. Having verified that the NCoA can be used on its link, NAR responds with a Handover Acknowledge (HACK) message to PAR. Thus, PAR validates the MN's NCoA and establishes a bidirectional tunnel between the PAR and NAR. While PAR sends a F-BACK to MN at t_a, it begins forwarding packets intended to MN's previous CoA (PCoA) to the NAR. After receiving the F-BACK message, MN may lose its connectivity with the PAR.

At time t_b, NAR starts to cache all the forwarded packets until MN establishes connectivity with the NAR at time t_c. It is assumed that at time t_c, MN establishes an IP connectivity with NAR, whereas NAR is able to drain the buffer containing the forwarded packets, as well as initiating an AAA process.

Since MN may already transmit its registration request within the F-BU message to PAR, NAR is able to otain the relevant information from PAR and send an AA check message to the local AAA server via GFA. Once the AA

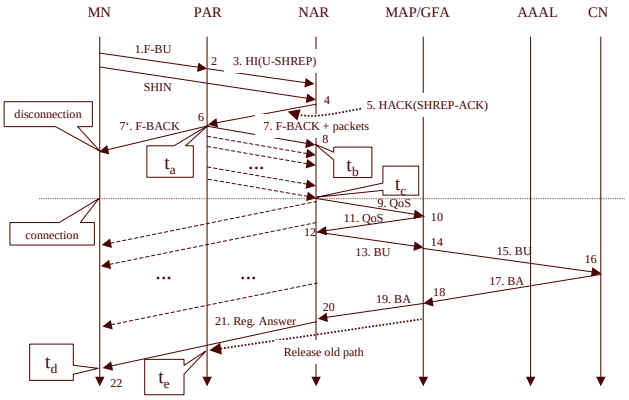

Fig. 3. Signaling Exchanges of Scheme 2

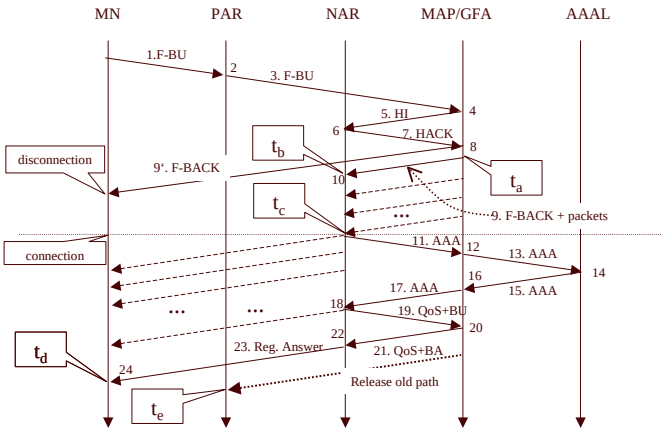

Fig. 4. Signaling Exchanges of Scheme 3

checks are successful, the DiffServ policy may be transmitted from AAAL to NAR. Meanwhile, the NAR initiates a QoS resource reservation (e.g. bandwidth) on the path from NAR to the cross-over router, which is the GFA in our case. When the resource is reserved successfully for the session, it performs a binding update process to CN on behalf of MN. When NAR gets a successful BA message, it sends a registration answer to MN. At time t_d, MN receives packets routed directly to its NCoA address. At time t_e, PAR receives a message from GFA to tear down the old path. The corresponding details are given in Table I.

To simplify the calculation, we assume that the packet size of each signaling message is 200 bytes except for the first F-BU, which is assumed to be 300 bytes; the processing time at each node is 200 μs except for the forwarding action, which takes only 20 μs; CN is assumed to be 10 ms away from the access network. The time interval between "disconnection" with PAR and "connection" with NAR (namely link switching delay and IP connectivity latency) is assumed to be 80 ms based on [9]. We hold the same assumptions for all the evaluated schemes.

B. Scheme 2: FMIP + CT + QoS

The combined procedure of FMIP and CT for a seamless handover operation is shown in Figure 3. When MN generates its NCoA, it sends a F-BU message to PAR. In this message, MN indicates its desire for context transfer. After PAR has verified the authenticity of the request, it sends a HI message including all the relevant feature contexts, as well as the authentication option and an "unsolicited Seamless Handover Reply" (U-SHREP) option. When MN sends a "Seamless Handover Initiate" (SHIN) message to NAR requesting context transfer, NAR must verify that the authentication data present in the SHIN message matches what was supplied by PAR in U-SHREP. When the check passes, NAR may send a SHREP-ACK option back to PAR in the HACK message. Thus, a bidirectional tunnel between the PAR and NAR is established.

Since it is assumed that there is a security association between PAR and NAR, the authentication and authorization information of the MN has been transferred securely. Hence NAR can perform the security checks without involving the local AAA server. Therefore, NAR can initiate the resource reservation process directly on the path from NAR to GFA, and then perform a binding update process to CN on behalf of MN.

t_a, t_b, t_c, t_d and t_e bear the same significance as in Scheme 1.

C. Scheme 3: FMIP + HMIP + AAA + QoS

In HMIPv6, a local anchor point (MAP) is placed in the local network, allowing Mobile IPv6 to benefit from reduced mobility signaling with external networks. The MAP server is then essentially a local home agent, whereas in FMIPv6, the ARs act as local Home Agents which hold binding caches for MNs and receive Binding Updates, so that these ARs function like the MAP specified in HMIPv6. It is also quite possible to have ARs communicating through an aggregation router instead of being directly connected. An aggregation router is then the ideal situation for the MAP functionality [14], so that using MAP in the aggregation router would improve the efficiency of Fast Handovers.

In Scheme 3, where we consider the combination of FMIP and HMIP, it is possible to embed QoS signaling in the binding update process to minimize the registration latency [4]. Therefore, in the HMIP schemes, we integrate the QoS signaling with the binding update process.

As happens for any FMIP scheme, NAR has to consume storage space during the period of IP connectivity, and the transferred packets from NAR to MN have no QoS guarantee or SA protection.

The time at which MN disconnets from PAR is assumed to be t_a, when MAP fowards packets to NAR instead of PAR. t_b is the time at which NAR starts caching forwarded packets from MAP. NAR forwards the cached packets to MN at t_c, and it initates an AAA process meanwhile. t_d is the endpoint of the handover operation, and the old QoS path is completely torn down at t_e.

TABLE I

DETAILS OF SCHEME 1

Step	Name	Time (μs)	Packet Size bytes	Link (Mbps)	Remarks
1	F-BU	-	300	54	F-BU is transferred over a 54 Mbps wireless link
2	PT_{PAR}	200	-	-	PAR takes 200 μs to process the message, including verifying authenticity of the request
3	HI	-	200	100	HI is transferred over a 100 Mbps wired link
4	PT_{NAR}	200	-	-	NAR takes 200 μs to processes the message
5	HACK	-	200	100	HACK is transferred over a 100 Mbps wired link
6 (t_a)	PT_{PAR}	200	-	-	PAR takes 200 μs to process the message
7	F-BACK + packets	-	200+packet size	100	PAR sends F-BACK and forwards packets to NAR via GFA; meanwhile, it sends a F-BACK to MN
8 (t_b)	PT_{NAR}	200	-	-	NAR takes 200 μs to process the message. NAR begins to cache the forwarded packets
9 (t_c)	AAA	-	200	100	NAR sends an AA check message to AAAL via GFA
10	PT_{FGA}	20	-	-	GFA takes 20 μs to forward the message
11	AAA	-	200	100	NAR sends an AA check message to AAAL via GFA
12	PT_{AAAL}	200	-	-	AAAL takes 200 μs to perform the AA check
13	AAA	-	200	100	AAAL sends an AA check response message to NAR via GFA
14	PT_{FGA}	20	-	-	GFA takes 20 μs to forward the message
15	AAA	-	200	100	AAAL sends an AA check response message to NAR via GFA
16	PT_{NAR}	200	-	-	NAR takes 200 μs to process the message. NAR initiates a resource reservation message
17	QoS	-	200	100	NAR sends an resource reservation message to GFA
18	PT_{FGA}	200	-	-	GFA takes 200 μs to process the message
19	QoS	-	200	100	GFA sends an resource reservation response message to NAR
20	PT_{NAR}	200	-	-	NAR takes 200 μs to process the message. NAR initiates a binding update message
21	BU	-	200	100	NAR sends an binding update message to CN
22	PT_{FGA}	20	-	-	GFA takes 20 μs to forward the message
23	BU	10,000	200	100	it is assumed to take 10 ms to transmit the BU message to CN
24	PT_{CN}	200	-	-	CN takes 200 μs to perform the BU operation
25	BA	10,000	200	100	it is assumed to take 10 ms to transmit the BA message to GFA
26	PT_{FGA}	20	-	-	GFA takes 20 μs to forward the message. GFA releases the old path
27	BA	-	200	100	GFA forwards the BA message to NAR
28	PT_{NAR}	200	-	-	NAR takes 200 μs to process the message. NAR sends a registration answer message to MN
29	Reg. Answer	-	200	54	Registration message is transferred over 54 Mbps wireless channel
30 (t_d)	MN receives the Registration message				

D. Scheme 4: FMIP + HMIP + CT + QoS

The signaling exchange flow of the combined procedure of FMIP and CT in a Hierarchical Mobile IPv6 architecture is shown in Figure 5.

MN sends a F-BU message to MAP via PAR. Since CT is enabled, after verifying the authenticity of the request, PAR transfers the related context to MAP with the message. MAP then sends a HI message containing SHREP to NAR with the transferred context. Thus, when receiving a request from MN, NAR is able to perform authentication and authorization checks and also deploy the DiffServ Policy. Then if the checks pass, NAR sends a HACK(SHREP-ACK) message to MAP. At time t_a, MAP starts forwarding data packets that were destined to MN's old address to NAR. NAR needs to cache the forwarded packets for the period between t_b and t_c, until MN establishes connection with NAR. Since security checks have been done, NAR then initiates directly a joint QoS and mobility process. At time t_d, when MN receives the registration answer message, the procedure is complete.

E. Scheme 5: HMIP + AAA + QoS

Here FMIP is not being deployed, so that when MN loses its connection to PAR it has to re-establish a path with NAR. When MN receives advertisements from NAR and decides to move, it generates its NCoA based on the network prefix information contained in the advertisements. MN needs the same link switching time as in the FMIP cases to gain IP connectivity, which is assumed to be 80 ms according to [9].

As shown in Figure 6, NAR first caches some state information such as requested QoS and BU, and initiates an AAA process with AAAL. After a successful security check, while deploying DiffServ policy, NAR may perform resource reservation and binding update in a combined manner. When the process succeeds, NAR sends a registration answer message

to MN. The old path is released by MAP after the successful resource reservation and binding update operations.

F. Scheme 6: HMIP + Cookie + QoS

NAR first performs a cookie verification. A cookie is granted by an access router after MN's first successful inter-domain handover. The cookie is used to gain access during an intra-domain handover. If the cookie check passes, NAR initiates a combined QoS and BU process and a re-authorization process at the same time. When receiving a positive binding ackowledgement message from MAP, NAR sends a registration answer message to MN, along with a set of SA parameters; this enables setting up a temporary IPSec tunnel without having to wait for the result of the re-authorization process, whenever this result has not yet reached AR.

The handover procedure is regarded to be complete when MN receives the registration message at time t_b.

When the result of the re-authorization process arrives at NAR, NAR performs an authentication check by using the session key. If the security checks pass, NAR generates a new cookie, encrypts it with the session key, and transmits the encrypted cookie to MN, along with a parameter for a new IPSec tunnel, in a registration refresh message. Upon receiving this message at time t_c, MN starts using the new SA to protect the user data and sends a registration refresh reply message to NAR.

The procedure details are illustrated in Figure 7.

IV. Performance Evaluation

In order to analyze the six Schemes presented earlier, we shall use the metrics of Total Response Time (TR), Packet Loss and CPU Processing Load.

A. Total Response Time

Here we are interested in computing the *mean value* of the Total Response Time TR, TR denoting the amount of time elapsed between sending of the first bit of a registration request and reception of the last bit of the corresponding registration response.

We consider the moment where this last bit is received as a period endpoint, since MN can receive user data with QoS guarantee and security protection after this moment. Therefore, the TR metric reflects the time needed for complete mobility management and session state re-establishment.

Although MN may receive user data from NAR much before the whole procedure is complete, all the forwarded data has no QoS guarantee and security protection. Forwarding packets from PAR to NAR enables one to avoid losing packets, which may be crucial to certain packet-loss sensitive applications.

We shall be using a *homogeneous* fluid flow mobility model, so that the number of mobile users entering a given cell during one second is a Poisson distributed random variable with mean value x. Our homogeneity assumption is valid both in space and time, so that the distribution of this random variable does not depend on the particular cell under consideration, nor on the particular (one second long) time interval.

Looking back at Figure 2 (Signaling exchanges in Scheme 1), one may first notice that the variable TR_1 may be expressed as a sum of random variables (queueing times) and deterministic variables (transmission and processing times). In the technical framework we are considering, one may then compute e.g. t_1 (time needed for the transmission of a 300 bytes long message over the wireless link) as

$$t_1 = (8 \cdot 300) \cdot (54 \cdot 10^6)^{-1} s,$$

and all the other transmission times may be computed in a similar way, assuming the wireless channels have a theoretical capacity of $54Mbps$, and the wired channels a 100 Mbps capacity.

Referring to [9], we are also assuming that the time during which NAR holds the packets forwarded by PAR before draining the corresponding buffer has mean value

$$\Delta_1 = 80ms,$$

and that the mean value of the time needed for the transmission of a Binding Update message from MAP to CN and reception of the corresponding answer adds up to $\Delta_2 = t_{23} + t_{24} + t_{25} = 20,2ms$.

It then remains to compute the mean values $E[W_{AR}]$, $E[W_{MAP}]$ and $E[W_{AAAL}]$ of the time that a job needs to spend queueing at an AR server (resp. at MAP, at AAAL) before undergoing service. (According to our homogeneity assumptions, the queueing times at each of the AR servers

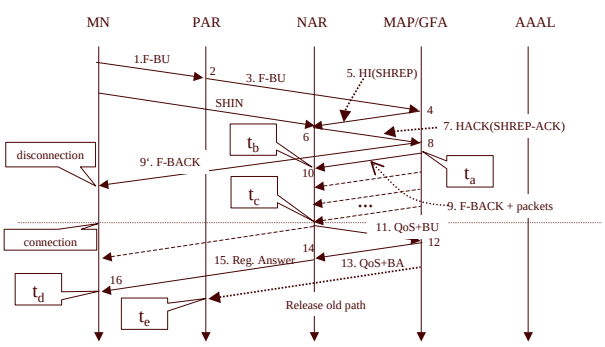

Fig. 5. Signaling Exchanges of Scheme 4

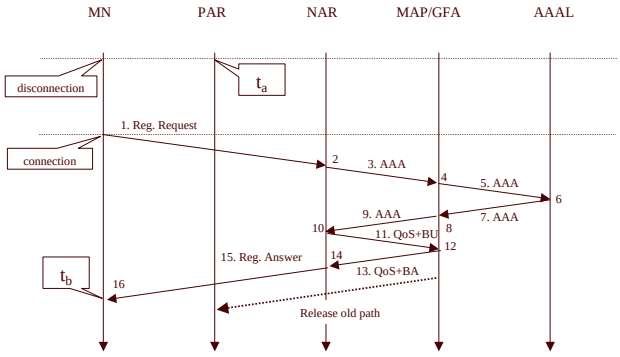

Fig. 6. Signaling Exchanges of Scheme 5

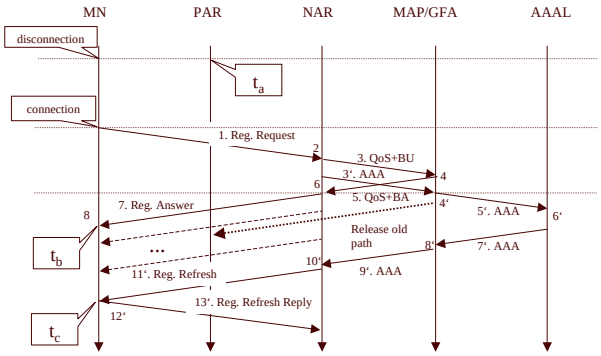

Fig. 7. Signaling Exchanges of Scheme 6

are identically distributed).

Let us first evaluate the mean waiting time at the AAAL server, since this is the simplest of all queueing times to be analyzed. Indeed, according to the Pollaczek-Khinchine formula (see e.g. Chapter 3 in [1] or Chapter 3 in [7]), the mean residual time at AAAL may be expressed as

$$E\left[R_{AAAL}\right] = \frac{\lambda^*_{AAAL}}{2} \cdot t_{12}^2$$

Here t_{12} still denotes the time required for the execution of task no.12 by AAAL, and

$$\lambda^*_{AAAL} = \lim_{T\to\infty} \frac{1}{T} E\left[\sharp \text{ of job arrivals at AAAL}\right]$$

is the asymptotic mean frequency of arrivals at AAAL. Assuming that the MAP and AAAL servers have been assigned to a region consisting of N cells, one has

$$\lambda^*_{AAAL} = N \cdot x,$$

whereas $\lambda^*_{AR} = x$.

According to Little's Theorem, the mean number of jobs to be found queueing at AAAL is then

$$n_{AAAL} = \lambda^*_{AAAL} \cdot E\left[W_{AAAL}\right],$$

so that the mean value of the time a job needs to queue at AAAL before being executed is given by

$$\begin{aligned} E\left[W_{AAAL}\right] &= n_{AAAL} \cdot t_{12} + E\left[R_{AAAL}\right] \\ &= \frac{(\lambda^*_{AAAL}/2)\cdot t_{12}^2}{1-\lambda^*_{AAAL} t_{12}} \end{aligned}$$

In the very same way, one may use the P-K formula to show that

$$E\left[R_{MAP}\right] = \frac{\lambda^*_{MAP}}{2} \sum_{k=0}^{4} t_{10+4k}^2$$

and then combine it with Little's Theorem in order to obtain:

$$E\left[W_{MAP}\right] = \frac{(\lambda^*_{MAP}/2) \cdot \sum_{k=0}^{4} t_{10+4k}^2}{1 - \lambda^*_{MAP} \sum_{k=0}^{4} t_{10+4k}}$$

Considering anyone of the AR servers in the system, it then remains to compute the mean value of the time a job needs to queue at this server before being treated. Clearly, this AR server is busy treating both jobs corresponding to users entering its cell as well as jobs corresponding to users exiting its cell, and the mean frequency of user exits is just the same as the mean frequency of new entrances, namely x users per second. Proceeding as before, one obtains

$$E\left[W_{AR}\right] = \frac{(x/2)\left\{t_2^2 + t_4^2 + t_6^2 + t_8^2 + t_{16}^2 + t_{20}^2 + t_{28}^2\right\}}{1 - x\left\{t_2 + t_4 + t_6 + t_8 + t_{16} + t_{20} + t_{28}\right\}}$$

The mean Total Response Time corresponding to Scheme 1 may finally be expressed as

$$\begin{aligned} E\left[TR_1\right] = 7 \cdot E\left[W_{AR}\right] + 5 \cdot E\left[W_{MAP}\right] + E\left[W_{AAAL}\right] \\ + \Delta_1 + \textstyle\sum_{i=1}^{29} t_i \end{aligned}$$

The very same methods may then also be used in order to compute the mean values of the Total Response Times $E\left[TR_2\right]$, ..., $E\left[TR_6\right]$ corresponding to all five other Schemes.

B. CPU Processing Load

Considering a server S in the context of Scheme k, one may define the CPU Processing Load $CPU_S^{(k)}$ as the proportion of time during which server S is kept busy processing different jobs in Scheme k. Mathematically speaking, this metric is then simply a linear function of the mean frequency of new arrivals per cell.

One may for example compute and compare the metrics $CPU_{MAP}^{(k)}$ for $k = 1, 2, \ldots, 6$. In the case of Scheme 1, we obtain (using Fig.2 and Table I):

$$CPU_{MAP}^{(1)} = (Nx) \cdot \sum_{k=0}^{4} t_{10+4k},$$

x denoting the mean number of new arrivals per second in a given cell, and the service times t_i above being expressed in seconds. The metrics $CPU_{MAP}^{(2)}$, ..., $CPU_{MAP}^{(6)}$ may then be expressed in the same fashion.

C. Packet Losses and Storage Space

As shown in Figure 2, it is assumed that MN loses its L2 connection with PAR at t_a. At the same time, PAR starts forwarding User Data to NAR. Once a connection is re-established between MN and NAR at t_c, NAR transmits the forwarded packets to MN. Although the forwarded packets have no QoS guarantee and security protection, they will eventually be received by MN. Therefore, there are no packet losses during the registration procedure in the FMIP schemes, and NAR only needs to store the forwarded packets for the period separating t_a and t_c.

In contrast, when we assume that the handover is performed in a "break before make" manner, MN loses its L2 connection with PAR at t_a, as shown in Figure 5 and Figure 6. It regains the connection with NAR when the registration procedure is complete, i.e. at t_b. In between, i.e. during the Interruption Time, the User Data is simply lost.

Schemes 5 and 6 thus have a specific drawback, namely that of inducing a packet loss during the Re-Registration Procedure. In order to quantify these losses, we have considered the situation where a Mobile User entering a Cell is currently

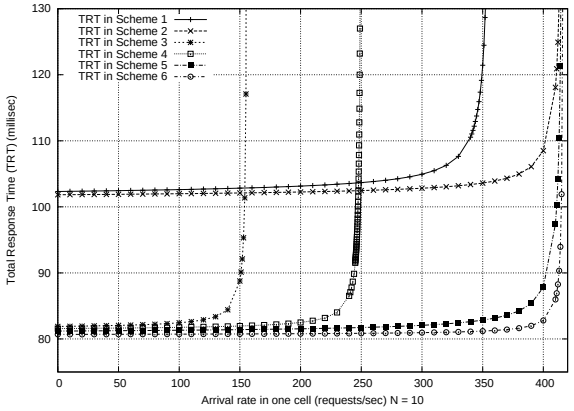

Fig. 8. Total Response Time N = 10

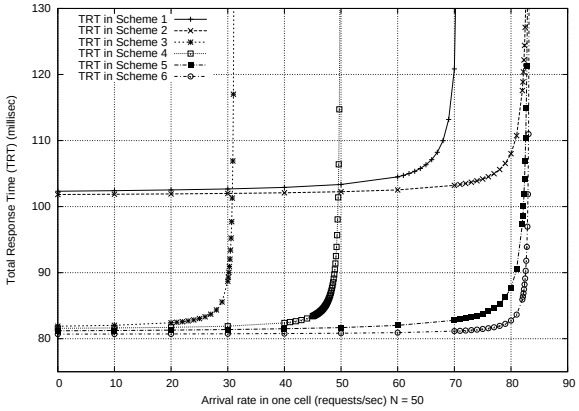

Fig. 9. Total Response Time N = 50

receiving data using File Transfer Protocol (FTP) with the rate of $88Kbps$.

The mean packet losses in Schemes 5 and 6 may then be computed simply by multiplying the mean value of the Total Response Time by the appropriate rate. In contrast to these losses occuring in Schemes 5 and 6, one could believe that the requirement of storing temporarily some User Data at MAP in Schemes 1, 2, 3 and 4 might rapidly cause some buffer overflows, but this is actually not the case. Indeed, considering the buffer space required for the temporary storage of User Data at MAP, one may easily come to the conclusion that in the worst case of a Region consisting of 50 Cells, the buffer space required at MAP for such storage purposes does not exceed a few Mbits.

V. Numerical Results and Discussion

Considering first the metric of Total Response Time, we give three comparative plots as shown in Figures 8 and 9, where the MAP and AAAL servers are being allocated to regions consisting successively of $N = 10$ and $N = 50$ cells.

In both cases, it turns out that the HMIP Schemes (i.e. Schemes 3, 4, 5, 6) have a better TR performance than Schemes 1 and 2 in case of moderate traffic intensity, due to the absence of a Binding Update operation involving CN. This shows that Hierachical Mobile IP can minimize the response time in micro-mobility senarios.

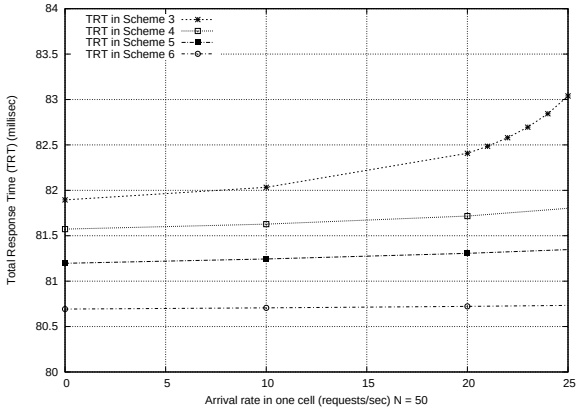

Fig. 10. Total Response Time in case of Low Traffic Intensity N = 50

On the other hand, combining HMIP with FMIP in order to avoid this BU procedure also has the effect of inducing a heavier workload on the MAP server, so that in Schemes 3 and 4 saturation is reached earlier than in the other Schemes (i.e. Schemes 1, 2, 5 and 6).

Moverover, as shown in Figure 10, in case of low traffic intensity (i.e. before the saturation points), the TR values in Schemes 3 and 4 are slightly higher than those appearing in Schemes 5 and 6, because MAP has more operations to handle in a fast handover context.

When comparing Schemes 1 and 2 or Schemes 3 and 4, we observe that Context Transfer plays a positive role in reducing the Total Response Time and yielding an improved tolerance to higher traffic intensities.

On the other hand, the TR metric in Scheme 6 (which has a single round trip) is only slightly better than that of Scheme 5 (which has two round trips), since the time for the first round trip in Scheme 5 is very short compared to the link swiching delay.

This difference in the workloads at MAP may also be appreciated through an examination of the CPU Processing Load at MAP in all six Schemes. Figure 11 shows that the GFA/MAP server is more solicited in Schemes 3 and 4 than in the other schemes, because of a heavier load induced by FMIP+HMIP operations.

When comparing Schemes 3 and 4, one may observe that CT is slightly preferable to AAA since CT requires a milder involvement of the MAP server.

Our final plot (Figure 12) features one of the main drawbacks of the HMIP Schemes compared with the FMIP Re-Registration procedures: Schemes 5 and 6 introduce a data Packet Loss which could be intolerable in certain situations such as downloading a file by using FTP, whereas there may be no packet getting lost when FMIP is used. Taking fluctuations into account, a *mean value* of $1Kb$ for the packet losses signifies that several Kbits may be lost during a single Re-Registration. The packet loss in Scheme 6 is slightly lower than in Scheme 5 due to the shorter TR in Scheme 6.

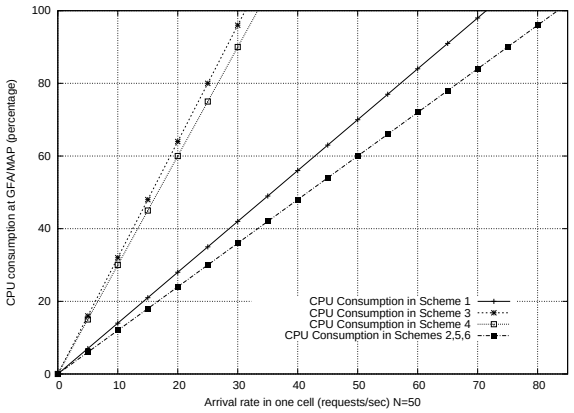

Fig. 11. CPU Consumption N = 50

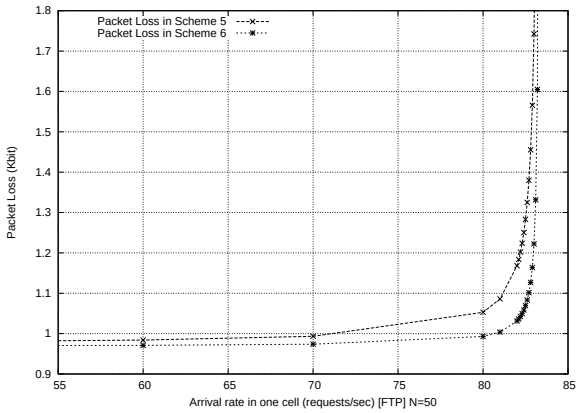

Fig. 12. Packet Loss of FTP Traffic N = 50

VI. Conclusion

In the present paper, we have been comparing six Re-Registration Schemes featuring FMIP, Context Transfer, HMIP, using the metrics of Total Response Time (TR), CPU Consumption at the MAP/GFA Server and Packet Loss (PL).

It turns out that Schemes 5 and 6 (based on HMIP) provide lower Total Response Times, with a better tolerance to higher values of the Traffic Intensity. These two Schemes may also be seen to offer more robustness when comparing the CPU Processing Loads induced by each of the six Schemes at the MAP/GFA Server. Scheme 6 should be preferred when taking also robustness against DoS attacks into consideration, because of the cookie scheme. Moreover, in Scheme 6, PL and TR are further reduced by parallelizing the security check process and the QoS+BU process [3].

On the other hand, when considering lower values of the Traffic Intensity and restricting our attention to the TR metric, the comparison becomes slightly favorable to Schemes 3 and 4, in which FMIP and HMIP are being integrated.

Furthermore, one should also realize that in Schemes 5 and 6 several Kbits of User Data may be lost during a single Re-Registration, even when considering intermediate values of the Traffic Intensity and FTP Data only, which might lead one to choose one of the first three FMIP Schemes (which are able to transmit all User Data to MN) for certain Packet Loss sensitive applications.

From the plots of all metrics, we conclude that the bundling feature of context transfer on FMIP can improve the performance. Assigning fewer Cells to each MAP server results in an improved tolerance to higher traffic intensities and less Packet Losses. However, in the FMIP schemes, if the data packets of real-time applications are buffered for a longer time period than the admissible end-to-end delay, they may become useless. Even when the buffered packets are forwarded to MN within the admissible delay, they may be delivered on the new QoS path from GFA/MAP to NAR without appropriate treatments since the new QoS path may have not been established.

References

[1] D. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, 2nd edition, 1991.

[2] T. Chen, S. Hermann, and G. Schaefer. Secure, QoS-enabled Mobility Support in All-IP Networks
- Final Report of the SeQoMo/SAM Project. Technical report, TKN TU-Berlin, 1993.

[3] T. Chen, G. Schaefer, C. Fan, S. Adams, M. Sortais, and A. Wolisz. Denial of Service Protection for Optimized and QoS Handover Based on Localized Cookies. In *proceedings of The 5th European Wireless Conference Mobile and Wireless Systems beyond 3G EW2004*, 2004.

[4] X. Fu, H. Karl, and C. Kappler. Qos-conditionalized handoff for mobile ipv6. In *Proc. of the Second IFIP-TC6 Networking Conf. - Networking2002*, pages 721–730, Pisa, Italy, May 2002. Springer-Verlag.

[5] D. Johnson, C. Perkins, and J. Arkko. Mobility Support in IPv6, June 2003. Internet Draft, draft-ietf-mobileip-ipv6-24.txt.

[6] J. Kempf. Problem Description: Reasons For Performing Context Transfers Between Nodes in an IP Access Network, September 2002. RFC3374.

[7] L. Kleinrock. *Queueing Systems, Vol. 1: Theory*. Wiley, 1975.

[8] R. Koodli. Fast Handovers for Mobile IPv6, March 2003. Internet Draft, draft-ietf-mobileip-fast-mipv6-06.txt.

[9] R. Koodli and C. Perkins. Fast Handovers and Context Transfers in Mobile Networks. *ACM Computer Communication Review*, Vol. 31(No. 5), oct 2001.

[10] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli. Context Transfer Protocol, June 2003. Internet Draft, draft-ietf-seamoby-ctp-03.txt.

[11] S. Pack and Y. Choi. A Study on Performance of Hierarchical Mobile IPv6 in IP-based Cellular Networks. *IEICE Transactions on Communications*, Vol. E87-B(No. 3), mar 2004.

[12] X. Perez-Costa, M. Torrent-Morenoa, and H. Hartenstein. A performance comparison of Mobile IPv6, Hierarchical Mobile IPv6, Fast Handovers for Mobile IPv6 and their combination. *ACM MC2R*, nov 2003.

[13] C. Perkins. IP Mobility Support for IPv4, August 2002. RFC 3344.

[14] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier. Hierarchical Mobile IPv6 mobility management (HMIPv6). Internet-Draft: draft-ietf-mipshop-hmipv6-01.txt, February 2004.

[15] M. Woo. Performance Analysis of Mobile IP Regional Registration. *IEICE Transactions on Communications*, Vol. E86-B(No. 2):pp. 472–478, feb 2003.

[16] J. Xie and I. Akyildiz. A Distributed Dynamic Regional Location Management Scheme for Mobile IP . *IEEE Transactions on Mobile Computing*, Vol. 1(No. 3):pp. 163–175, July 2002.