



**HAL**  
open science

## Processing loads analysis of distributed mobility management and SIP-based reachability

Hassan Ali Ahmad, Kashif Munir, Philippe Bertin, Karine Guillouard,  
Meryem Ouzzif, Xavier Lagrange

► **To cite this version:**

Hassan Ali Ahmad, Kashif Munir, Philippe Bertin, Karine Guillouard, Meryem Ouzzif, et al.. Processing loads analysis of distributed mobility management and SIP-based reachability. *Telecommunication Systems*, 2016, 63 (4), pp.681 - 696. 10.1007/s11235-016-0148-4 . hal-01439606

**HAL Id: hal-01439606**

**<https://hal.science/hal-01439606v1>**

Submitted on 2 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Processing loads analysis of distributed mobility management and SIP-based reachability

Hassan Ali-Ahmad<sup>1</sup>, Kashif Munir<sup>2</sup>, Philippe Bertin<sup>3</sup>, Karine Guillouard<sup>3</sup>, Meryem Ouzzif<sup>3</sup>, Xavier Lagrange<sup>4</sup>

**Abstract** Currently, various efforts are being performed on specifying distributed mobility management (DMM) for IPv6 mobile networks. DMM is expected to eliminate the limitations of the current IPv6 mobility protocols and to cope with the rapid increase in mobile data traffic. In current DMM approaches, the mobility anchoring is distributed at the access router level and the mobile node changes dynamically the mobility anchor for new sessions. Although such approaches avoid any network bottleneck and allow optimal routing in most cases, they lack reachability support since the mobile node has no permanent IPv6 address. In this paper, we consider one of the DMM approaches, dynamic mobility anchoring, and study the usage of session initiation protocol (SIP) for providing reachability support. After defining different SIP-based location update modes, we carry out a comparative cost analysis considering new performance metrics related to the processing loads at each network entity. We consider the location update loads on the location server as well as the context, signaling, and tunneling loads on a mobility anchor. We investigate the different modes in several scenarios, allowing the mobile network operator to adopt the preferred mode depending on their architecture.

**Keywords** Distributed mobility management · SIP · IP mobility · Reachability · Performance analysis · Analytical modeling

---

✉ Hassan Ali-Ahmad  
hmaliyahmad@gmail.com

<sup>1</sup> Bretagne Telecom, Chateaubourg, France

<sup>2</sup> King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

<sup>3</sup> Orange Labs, Rennes, France

<sup>4</sup> Telecom Bretagne - IRISA, Rennes, France

## 1 Introduction

Recently, mobile network operators have been experiencing a rapid growth in mobile data traffic. Operators' statistics show that 2012 has seen more mobile data traffic than all the preceding years combined [1]. They also show that the data traffic has doubled on mobile networks between 2012 and 2013. This growth is expected to continue in the coming years [2,3], resulting in an explosion in mobile data traffic. In this context, current network architectures as well as IPv6 mobility protocols are expected to encounter scalability and performance issues since they are deployed in a centralized manner. They rely on a centralized gateway that acts as a mobility anchor and consequently manages all the mobile users' data traffic as well as their contexts. For instance, mobile IPv6 (MIPv6) [4] relies on the home agent (HA) and proxy mobile IPv6 (PMIPv6) [5] relies on the local mobility anchor (LMA). Such centralized mobility anchor is considered as a network bottleneck and single point of failure. It also leads to non-optimal routing as well as large tunneling overhead.

In order to tackle such challenge, a new trend is to flatten networks architectures. Hence, IPv6 mobility management protocols need to be adapted for such evolution. Therefore, there is a need to define novel mobility management mechanisms. Accordingly, the IETF has chartered a new working group called distributed mobility management (DMM) [6] for IPv6 mobility. DMM aims at overcoming such issues by introducing the distributed and dynamic concepts to the mobility management scheme [7]. In DMM, the mobility anchoring is distributed at the access router (AR) level and the mobility anchor is changed dynamically while the user is moving. As a result, most of the data traffic is routed optimally and the tunneling overhead is reduced significantly. However, the mobile user changes its primary IPv6 address

frequently and hence might have several IPv6 addresses simultaneously. This raises the reachability issue in the DMM schemes.

In the literature, several DMM approaches are proposed especially at the IETF DMM working group [8]. These approaches are generally based on Mobile IPv6 (MIPv6) or Proxy Mobile IPv6 (PMIPv6). Several studies focus on the performance evaluation or comparative analysis of different mobility protocols. Part of these studies, e.g. [9–16], consider performance metrics that are related to the traffic loads on the network links (referred to as cost analysis) such as signaling, tunneling, and data packet delivery costs. Other studies, e.g. [15, 17, 18], consider different performance metrics that are related to the delays such as handover latency and end-to-end-delay. Similarly, several studies, consider those cost and delay performance metrics to evaluate DMA or DMM, e.g. [19–25]. Though the results depend of course on the considered parameters, the authors of [20] conclude that DMM is an interesting approach as it improves the data delivery. The authors of [26] consider in addition the costs related to query messages by CN, the costs of refreshing binding updates, registration and data acknowledgement messages but only for their proposed scheme named seamless IP diversity-based generalized mobility architecture (SIGMA) and hierarchical MIPv6 (HMIPv6). There are thus a lot of publications on the mobility signaling and data planes traffic-related costs but there is no study that focuses on the processing loads on the different nodes or servers. Furthermore, the reachability support in the DMM schemes has not been in particular investigated.

Reachability is not required for all services but for services with incoming sessions. Several popular services such as web browsing, video streaming, and file transfer do not require reachability and each mobile node (MN) that has an IP address can use them. On the other hand, voice over IP (VoIP), text messaging, and multimedia conferences are the services where part of the sessions are incoming sessions. These services are all based on the session initiation protocol (SIP) [27], which is now widely adopted as the building block of all IP-based conversational services.

In this paper, we consider dynamic mobility anchoring (DMA) [28–30], which is one of the DMM approaches (the first proposal to the best of our knowledge) and we study how DMA and SIP can be jointly used. The authors of [31] propose a hybrid SIP-MIP mobility management where MIP mobility messages are used to manage SIP mobility in order to avoid duplicated location messages. This requires a tight integration of SIP and MIP. In our work, we consider DMA, which can be based on standard PMIPv6, and avoid any modification at the network layer. The counterpart is the lost of the reachability at the IP layer. However, as explained above reachability is no more necessary for operators at the IP layer but required at the application layer for SIP-based sessions.

Due to the extensive existing literature on traffic costs, we deliberately focus on the analysis of the processing loads on the location server and mobility anchors:

- The location server manages the reachability for all the users. Operators intend to reduce the number of location update requests in order to avoid congestion, failure risks and also to achieve higher reliability.
- One objective of DMA is to include mobility functions in a large number of standard off-the-shelf access routers with limited processing power. Operators thus need to estimate the processing loads generated by mobility functions. These loads are mainly affected by the tunneling mechanisms (creation, release, and modification of a tunnel).

The main contributions of the paper are the following.

- We develop a full analytical model to compute the different processing loads on a mobility anchor as well as on the location server, which are essential with the introduction of DMM. This analytical model allows to compute the lifetime of tunnels and hence the number of contexts, signaling messages, and tunneled packets that a mobility anchor should manage at a time.
- We carry out a comparative performance analysis on the different location update modes in DMM schemes, allowing the mobile network operator to adopt the preferred mode depending on their network architecture.

Note that the analytical model takes into account the fact that there are both periodic events (registration) and random ones (mobility). Furthermore, the model is valid when there are several simultaneous sessions in the same MN and hence it is applicable when the MN is either a mobile terminal or a mobile router that supports multiple sessions of different terminals.

The rest of the paper is organized as follows. After presenting DMA in Sect. 2, we present the different reachability approaches for DMA in Sect. 3. We then detail in Sect. 4 the SIP-based reachability support considering its default mode and a time-based mode. In Sect. 5, we develop the analytical models of the different metrics. Based on these models, we present the performance evaluations in Sect. 6. Finally, we conclude the paper and its results in Sect. 7.

## 2 Dynamic mobility anchoring (DMA)

Dynamic mobility anchoring (DMA) [28–30] is an IPv6 mobility scheme that relies on a distributed architecture. Mobility management is moved to the network edge in order to anchor the traffic closer to the mobile node (MN), rather

than at a centralized entity in the core network. Each access router (AR) is required to have both mobility anchoring and location update functionalities. The AR is then referred to as mobility-capable access router (MAR) [30]. Consequently, it is guaranteed that the MN is always attached to an AR that can act as a mobility anchor. The correspondent node (CN) is not required to have any additional capabilities, i.e. it can be mobility-unaware CN.

Compared to MIPv6/PMIPv6 where all the data traffic is anchored at the same entity ( i.e. the HA/LMA), dynamic anchoring means to change the anchoring point for new sessions. This allows the MN to always initiate new sessions using its current IPv6 address. The data traffic is then routed optimally without tunneling between the MN and CN. If these sessions are terminated before undergoing an IP-handover, then there is no need to keep the MAR where they were initiated as an anchor. On the other hand, if the MN undergoes an IP-handover before terminating these sessions, then the mobility management is needed and hence activated dynamically.

Upon an IP-handover, tunneling mechanisms are needed in order to guarantee session continuity for the MN. As the anchoring point is moved to the access network, the tunnel is established in the access network rather than between the access network and the core network. In particular, the tunnel is established between the serving MAR and the anchoring MAR. Then, data packets from the CN destined for the MN are routed naturally to the anchoring MAR by standard IPv6 routing mechanisms, encapsulated by the anchoring MAR to be tunneled to the serving MAR, and then decapsulated by the serving MAR to be delivered to the MN.

on the location server and mobility anchors: If the MN initiates new sessions after a handover, it uses the new IPv6 address. The data traffic of these new sessions is routed optimally. DMA is thus expected to route without tunneling most of the data traffic based on statistics showing that more than 60 % of sessions are non-mobile [32].

In what follows we consider the DMA draft in IETF, described in [30]. This proposal is based on PMIPv6. The new MAR sends on behalf of the MN a proxy binding update (PBU) to the previous MAR, which replies by a proxy binding acknowledgement (PBA). A tunnel is then established between the new MAR and previous MAR of the MN. The data traffic of the sessions that have been initiated at the previous MAR is then routed via this tunnel (see Fig. 1).

In order to be able to send PBU(s), the new MAR needs to know the IPv6 addresses of the MN's previous MAR(s) and the associated active IPv6 addresses of the MN. There are several solutions to provide the information:

- Option 1, relying on the MN. The MN stores for each active session the IPv6 address it uses and the associated anchor. The new MAR sends an Info Query to the MN,

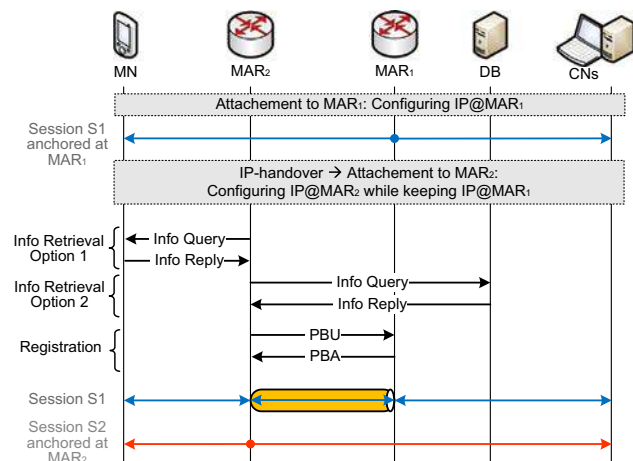


Fig. 1 Mobility management in DMA

which indicates the anchor in an Info Reply. Then, the new MAR sends the PBU(s) on behalf of the MN.

- Option 2, relying on a Database. This database stores ongoing sessions for all the MNs, particularly the active IPv6 addresses of each MN and associated anchors' IPv6 addresses [30]. It should be updated each time an MN configures or releases an IPv6 address. When the MN attaches to a new MAR, this MAR sends the Info Query message to the database, which answers by the Info Reply as explained above.

Option 1 does not introduce any new network entity but the MN is no longer agnostic of the mobility support. Since the MN participates, even partially, in the mobility-related signaling, this option is considered host-based in IETF. Option 2 does not affect the MN and keeps the approach network-based. It introduces a centralized entity in the control plane of the network. The approach is then considered partially distributed in IETF. However, note that as opposed to PMIPv6 where all data packets are routed through the LMA, only a few signaling messages are exchanged here with the data base.

### 3 Reachability support in DMA

In MIPv6 as well as PMIPv6, the MN has always a permanent IPv6 address, the home address (HoA). This allows the MN to be reachable at the IP-level, at any location and at any time. On the other hand, in DMA as well as other DMM schemes, the MN changes its primary IPv6 address frequently due to the dynamic aspect. The MN might have several IPv6 addresses simultaneously but none of these addresses is permanent. This raises the reachability support issue. Hereafter, we discuss the reachability support in DMA at IP-level and applications level.

### 3.1 Reachability support at IP-level

One of the approaches could be to let the MN keep (at least) one permanent IPv6 address, i.e. a home address HoA, and hence the MN is always reachable at the IP-level at that HoA. Consequently, the MN needs (at least) one permanent mobility anchor, which is considered centralized. It is then required to update the binding at that anchor upon each IP-handover. However, when the MN is not attached to that anchor, tunneling mechanisms are required for all the data packets sent by or destined to the MN's HoA.

In fact, this approach is nothing but a mixture of DMA and MIPv6/PMIPv6 as follows:

- The concept of centralized HA/LMA that provides a permanent HoA is used for incoming sessions/calls to support MN's reachability at the IP-level.
- On the other hand, distributed mobility anchors as well as dynamic anchoring concepts are used for all other sessions initiated by the MN itself.

However, similar issues to the ones addressed to current IPv6 mobility protocols (in the DMM problem statement) will be raised, but only for incoming sessions/calls. Incoming sessions will suffer non-optimal routing and tunneling overhead. Moreover, the permanent HA/LMA will be a single point of failure for all incoming sessions/calls, i.e. a single point of failure for reachability support.

In addition, if we assume that the CNs are also mobile and apply the same approach for reachability, then the incoming sessions to the MN are tunneled via its permanent anchor and the outgoing sessions are tunneled via the CNs' permanent anchors. As far as the MN communicates with mobile CNs, MIPv6/PMIPv6 is used more and DMA is used less.

Nevertheless, this approach provides stable IPv6 address for the MN and hence supports reachability at IP-level. It is up to the operator to choose this approach or not.

### 3.2 Reachability support at application level

Here, we consider another approach based on the SIP protocol [27] for reachability support in DMA at the application level. SIP is an application-layer signaling protocol for creating, modifying, and terminating sessions with one or more participants. SIP is well investigated with lot of extensions and already deployed in mobile networks. SIP can provide reachability at the application level for SIP-based sessions.

The main motivation behind this approach is the fact that typical incoming sessions (e.g. voice over IP (VoIP), text messaging, multimedia conferences) to a typical MN are expected to be SIP-based sessions. In other terms, application level reachability is sufficient in typical scenarios. In addition, SIP is an end-to-end protocol, i.e. data packets are

transmitted directly between the MN and its CNs without tunneling. SIP can also support terminal mobility for SIP-based sessions in an end-to-end fashion [33]. Hence, SIP natively avoids issues such as non-optimal routing and tunneling overhead.

*Accordingly, we opt relying on SIP for reachability support in DMA.*

## 4 SIP-based location update

### 4.1 Overview on SIP

In SIP, the communicating nodes are called user agent client (UAC) and user agent server (UAS). For consistency, we keep using MN and CN as identified above.

SIP introduces several logical entities to the network. For each domain (e.g. *example.com*), there is a proxy server and a registrar server (RS). These logical entities are usually co-located when deployed and referred to as SIP server. The SIP servers help to route requests to the MN's current location.

SIP identifies each MN by a unique SIP uniform resource identifier (URI), e.g. *alice@example.com*. In order to associate the SIP-URI with the current IPv6 address, SIP provides a registration function that allows the MN to upload their current IPv6 address to the registrar server. The registrar writes this association/binding into a database and the MN is required to update their IPv6 address. This information in the registrar server is used by the proxy servers to locate the MN when needed, e.g. to forward a request destined for the MN. To conclude, it can be said that SIP can natively provide location management and hence reachability support.

### 4.2 SIP default mode (SDM)

In the SIP default mode (SDM), the MN sends upon initial attachment a REGISTER message to the SIP server. The SIP server writes the binding and replies by an OK message. This binding is valid for a specified lifetime (the default lifetime value is 3600 s [27]). Hence, it should be refreshed periodically. In addition, when changing the IPv6 address, e.g. upon undergoing an IP-handover, the MN is required to perform the REGISTER procedure with the SIP server. Figure 2 illustrates the registrations in the SIP default mode, where  $\delta$  is the REGISTER lifetime.

As the registrar server is always up-to-date, the main advantage of this mode is the optimal routing of incoming sessions' data packets. Tunnels are set up only if the MN undergoes an IP-handover during the session (see Fig. 3). The load on the location sever can be high as because the MN updates its location both at each IP-handover and at each time-period. Such increase of the load could be not consistent

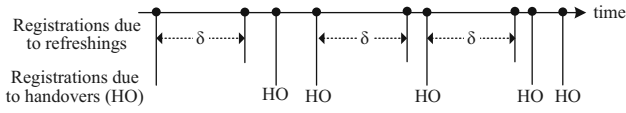


Fig. 2 Registrations in SIP default mode

with the DMM requirements on scalability [34]. Therefore, reducing the location updates rate as much as possible is crucial.

### 4.3 Time-based mode (TBM)

In order to reduce the location update requests sent to the registrar server, a time-based mode (TBM) can be used. In such mode, the MN updates the registrar server only periodically regardless of its movement, as illustrated in Fig. 4. One advantage of this mode is that the application layer (SIP) and the network layer are fully independent. However, the registrar server may not be up-to-date between two consecutive location updates. Consequently, an incoming session might be directed to one of the MN's previous IPv6 addresses that might be no more valid (see Fig. 5). Indeed, the incoming packets are sent to the IPv6 address registered in the registrar server and in practice they pass through the associated mobility anchor (referred to as reachability anchor as shown in Fig. 5). In order to tackle this issue, that anchor should forward those packets to the current MN's IPv6 address.

In other words, the reachability anchor should be maintained as one of the MN's anchors until the MN's IPv6 address is updated again. During this period, it is required

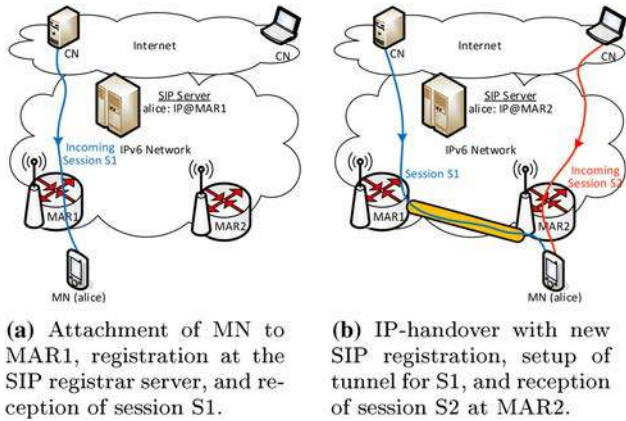


Fig. 3 Tunnels in DMA with SIP default mode

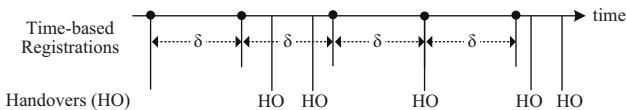
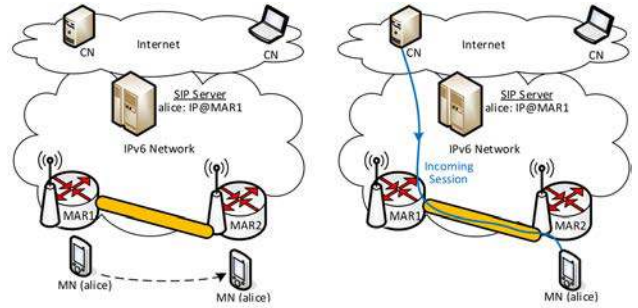


Fig. 4 Registrations in SIP time-based mode



(a) Attachment of MN to MAR1, SIP registration, then IP-handover & setup of reachability tunnel. (b) Reception of incoming sessions via MAR1 before updating the SIP registrar.

Fig. 5 Tunnel with the reachability anchor in DMA with time-based mode

to update the MN's binding at the reachability anchor upon each IP-handover. This can be achieved in a similar manner as updating the binding at the other MN's anchors through one of the two DMA options defined above.

### 4.4 Failure of the SIP registrar

As the registrar is used to locate the MN, it can appear as a single point of failure. However, the registrar has only to store a simple association between a URI and an IP address. Hence, it is possible to duplicate the registrar in order to have a resilient reachability management. Each registration is stored by two registrars using well-known synchronisation mechanisms such as virtual IP address. Such a solution can be used for both SDM and TBM even that it suits more TBM as the number of messages is limited and the location update frequency for a given MN is low. Hence, forwarding all the location requests to a secondary registrar in case of failure of the primary one can provide a continuous service to customers.

## 5 Performance analysis

### 5.1 Preliminaries

We consider a homogeneous regular network that is divided into hexagonal areas. Each area includes one MAR acting as the mobility anchor for that area. One or more radio access networks (e.g. cellular base stations or WiFi access points) can be connected to the same MAR. We then consider the well-known fluid-flow mobility model to represent the MNs' mobility between the areas. Under this model, it is assumed that the direction of the movement is uniformly distributed over the range  $[0, 2\pi)$ , and that the area residence time is exponentially distributed. The area crossing rate for an MN

is expressed as [35]:  $\alpha = \frac{E[v]L}{\pi S}$ , where  $E[v]$  is the MN's average speed,  $L$  is the area perimeter, and  $S$  is the area size. For hexagonal areas,  $L = 6r$  and  $S = 3\sqrt{3}r^2/2$  where  $r$  is the radius of the hexagon. Hence,  $L = 6\sqrt{2/(3\sqrt{3})}\sqrt{S}$ . The crossing rate can be simplified as:  $\alpha = \frac{\sqrt{8\sqrt{3}} E[v]}{\pi \sqrt{S}}$ .

As a traffic model, we assume that the session arrival to an MN follows a Poisson process with mean rate  $\lambda_s$ , i.e. the inter-arrival time between sessions is exponentially distributed with this rate. We assume also that the duration of a typical session is exponentially distributed with mean rate  $\mu_s$ . Then, we denote by  $\rho$  the proportion of incoming sessions (i.e. sessions initiated by the CN and not the MN). Hence, the arrival of incoming sessions is a Poisson process with rate  $\rho\lambda_s$  and the arrival of outgoing sessions is also a Poisson process with rate  $(1 - \rho)\lambda_s$ .

For any random variable  $X$ , we denote by  $F_X(t)$  the cumulative distribution function (CDF) of  $X$ , by  $\bar{F}_X(t)$  its complementary CDF (CCDF), by  $f_X(t)$  its probability density function (PDF) and by  $E[X]$  its expectation. Note that  $\bar{F}_X(t) = 1 - F_X(t)$  and  $f_X = dF_X/dt$ . In the problem we consider, all random variables are defined over  $\mathfrak{R}^+$ . Furthermore, we define  $X^+$  as the non zero values of  $X$ . Hence  $\bar{F}_{X^+}(t) = \bar{F}_X(t)/P(X \neq 0)$  if  $t > 0$  and  $\bar{F}_{X^+}(0) = 1$ . From elementary computations, we can deduce

$$F_{X^+}(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \frac{F_X(t) - P(X=0)}{1 - P(X=0)} & \text{if } t > 0 \end{cases} \quad (1)$$

We have  $E[X^+] = \int_0^\infty \bar{F}_{X^+}(t)dt$  if the latter integral is defined.

The main symbols used in the analytical models are listed in Table 1.

## 5.2 Signaling loads on the location server

In this subsection, we develop the analytical models of the signaling loads on the location server in both TBM and SDM cases. Through this metric, we can evaluate the scalability of the location server and how much is the gain in TBM compared to SDM. We consider that the location server is a SIP server or a SIP-like server.

### 5.2.1 TBM case

In TBM, each MN sends a REGISTER to the server every  $\delta$  seconds. Let  $N_{MN}$  be the total number of MNs in the network. Let  $\theta^{\text{TBM}}$  be the overall arrival rate of REGISTER requests on the SIP server. We have

$$\theta^{\text{TBM}} = N_{MN} \frac{1}{\delta}. \quad (2)$$

**Table 1** List of main symbols

Symbol	Description
Symbols used for an MN	
$\alpha$	Area crossing rate
$\delta$	REGISTER message lifetime
$\lambda_s$	Average arrival rate of sessions
$\mu_s$	Session service mean rate
$\rho$	Probability that a new session is an incoming session
$\pi_i^{\text{IS}}(t)$	Probability of having $i$ active incoming sessions at time $t$
$\pi_i$	Probability of having $i$ sessions in the set of the current area at any time
$\pi_i^{\text{OS}}$	Probability of having $i$ outgoing sessions in the set of the current area at any time
Symbols used for a set of MNs	
$N$	Number of MNs in an area at any time
$E[N]$	Expected number of MNs in an area at any time
$N_{MN}$	Total number of MNs in the network
$\lambda_{MN}$	Arrival rate of MNs to an area
Symbols used for a tunnel	
$\tau^x$	Mean lifetime of a tunnel in $x$ mode ( $x = \text{TBM or SDM}$ )
$\lambda_t^x$	Arrival rate of tunnels in $x$ mode
$M^x$	Average number of tunnels on a mobility anchor in $x$ mode

### 5.2.2 SDM case

In SDM, the MN sends a REGISTER to the SIP server upon each IP-handover and also for refreshing. The IP-handover rate is equal to the area crossing rate, which is  $\alpha$ . The refreshing rate is derived as follows.

Let  $N_{\text{REG}}$  be the number of REGISTER messages sent by an MN in a given area and  $T_d$  be the area dwell time. Random variable  $T_d$  is exponentially distributed with parameter  $\alpha$ . Then,  $N_{\text{REG}}$  is expressed as  $\lfloor T_d/\delta \rfloor$ .

Let  $p_k$  be the probability that exactly  $k$  REGISTER messages are sent by an MN during its area residence time. It can be calculated as follows:

$$\begin{aligned} p_k &= P\left(\left\lfloor \frac{T_d}{\delta} \right\rfloor = k\right) \\ &= P(k\delta \leq T_d < (k+1)\delta) \\ &= (e^{-\alpha\delta})^k (1 - e^{-\alpha\delta}). \end{aligned}$$

which is a geometric-related distribution. Thus,

$$E[N_{\text{REG}}] = \frac{e^{-\alpha\delta}}{1 - e^{-\alpha\delta}}.$$

The refreshing rate is thus given by  $\alpha E[N_{\text{REG}}]$ .

Let  $\theta^{\text{SDM}}$  be the arrival rate of REGISTER requests on the SIP server in SDM. It is the sum of the handover rate and refreshing rate (cf. Fig. 2), multiplied by the number of MNs. Hence, it is expressed as follows:

$$\theta^{\text{SDM}} = N_{\text{MN}} \alpha (1 + E[N_{\text{REG}}]) = \frac{N_{\text{MN}} \alpha}{1 - e^{-\alpha\delta}}. \quad (3)$$

### 5.3 Tunnel lifetime

After modeling the signaling loads on the location server, we need to model the tunneling-related aspects such as the number of tunnels and generated loads on a mobility anchor. First of all, we develop the analytical model of tunnel lifetime in different cases in order to be able to proceed after developing the other models. We first only consider the time-based mode (TBM). The SIP default mode (SDM) is deduced after from the analysis made for TBM.

#### 5.3.1 Overview

In TBM, the process is strictly periodic. We can then analyze only what happens in period  $[t_0, t_0 + \delta)$ . Let's consider an MN that sends a REGISTER at time  $t_0$  in MAR1 area. Let  $t_1$  be the time when the MN leaves MAR1 area and simultaneously enters another area (see Fig. 6). Thus,  $T_r = t_1 - t_0$  is a random variable that represents the residence time after REGISTER is sent. If  $T_r \geq \delta$ , then no new tunnel is created for reachability within period  $[t_0, t_0 + \delta)$ . If  $T_r < \delta$ , a new tunnel is set up when the mobile leaves the area to provide reachability. Let  $X$  be the duration between  $t_1$  and the instant of time of the next REGISTER, i.e.,  $X = \delta - T_r$ . The tunnel is maintained at least for duration  $X$ . If there are active incoming sessions, the tunnel is further maintained until all incoming sessions are terminated. Let  $Y$  be the duration necessary to finish all incoming sessions that are active at  $t_0 + \delta$ . The tunnel lifetime is greater than or equal to  $X + Y$ . Note that if  $T_r \geq \delta$  then  $X = 0$  and thus  $Y = 0$ .

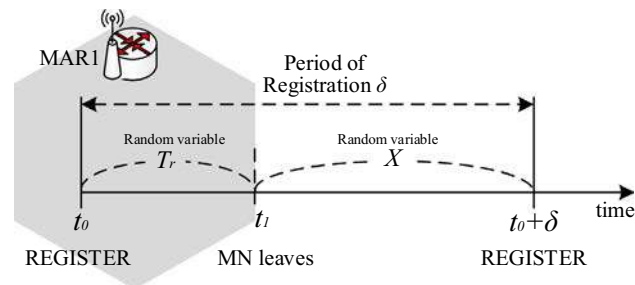


Fig. 6 Illustration of  $t_0$ ,  $t_1$  and  $\delta$  on the time axis

When the MN leaves the area at time  $t_1$ , there might be some active outgoing sessions. Let  $Z$  be the duration necessary to finish all outgoing sessions that are active at  $t_1$ . The tunnel should be maintained as long as there are active outgoing sessions. The tunnel lifetime is thus greater than or equal to  $Z$ . If the MN makes several successive handovers within  $[t_0, t_0 + \delta)$ , a new tunnel for outgoing sessions can be created at each handover ( $Z \neq 0$ ) as opposed to reachability for which only one new tunnel is created at time  $t_1$ .

The same tunnel is maintained for all functions, i.e. reachability, incoming sessions and also outgoing sessions. Hence, the tunnel lifetime, denoted by  $R$ , is expressed as follows:

$$R = \max(X + Y, Z). \quad (4)$$

A scenario is illustrated in Fig. 7.

An MN sends a REGISTER message  $R1$  in MAR1 area and its next REGISTER message  $R2$  in MAR3 area. Two outgoing sessions are set up at MAR1 and are still active when the MN leaves MAR1. The tunnel is maintained at least until the second outgoing session is terminated. The MN sends a register when it is in MAR3 area. At that time, incoming session 5 is still active. The tunnel is maintained until session 5 is terminated. Note that when the MN leaves MAR2, no new reachability tunnel is created ( $X = 0$  and then  $Y = 0$ ). A tunnel is created but only because outgoing session 4 was set up in MAR2 and is still active.

Variables  $Y$  and  $Z$  are clearly random with positive or zero values. As we consider the time when an MN leaves an area, the instant of time of the last REGISTER can be defined as a random variable. Hence,  $X$  is also a non negative random variable. Note that (4) is valid for all cases, including when  $X = Y = 0$  or  $Z = 0$ . In order to compute the CDF of the tunnel lifetime, we first compute the CDF of  $X$ ,  $Y$  and  $Z$  in the following.

#### 5.3.2 Lifetime of reachability tunnel

In this section we compute the CDF of  $X$  and  $X + Y$ , which are used later in the paper. Let us consider an MN that sends a

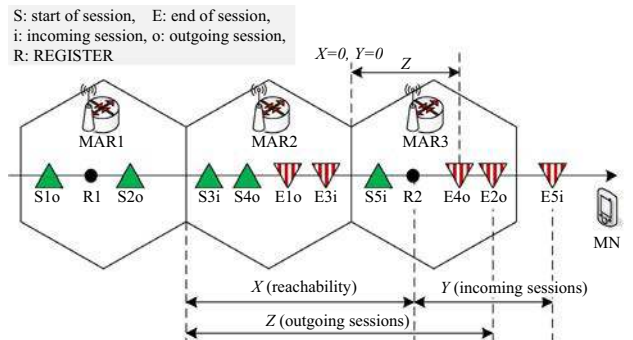


Fig. 7 Illustration of  $X$ ,  $Y$ , and  $Z$  with respect to different events (sessions, handovers, registers)



REGISTER at  $t_0$  in MAR1. No tunnel is created if the MN stays in MAR1 till the next REGISTER, i.e. if  $T_r \geq \delta$ . As the area residence time is a memory-less process, its CDF is  $1 - \exp(-\alpha t)$ . Hence, the probability of no tunnel created for reachability during period  $\delta$  is  $\exp(-\alpha\delta)$ . The probability to create at least one tunnel is  $1 - \exp(-\alpha\delta)$ . Note that in the case of subsequent handovers, no new tunnel is created for reachability as the existing one is used. Then,  $1 - \exp(-\alpha\delta)$  is the probability to create exactly one tunnel during  $[t_0, t_0 + \delta)$ . If a tunnel is created, then:

$$P(T_r < t | \text{tunnel created}) = \begin{cases} \frac{1 - e^{-\alpha t}}{1 - e^{-\alpha\delta}} & \text{if } 0 \leq t \leq \delta, \\ 1 & \text{if } t > \delta. \end{cases}$$

The tunnel is created at  $t_1$  and maintained until  $t_0 + \delta$ . Hence,  $X^+ = t_0 + \delta - t_1 = \delta - T_r$  and:

$$F_{X^+}(t) = \begin{cases} 1 - \frac{1 - e^{-\alpha(\delta-t)}}{1 - e^{-\alpha\delta}} & \text{if } 0 \leq t \leq \delta, \\ 1 & \text{if } t > \delta. \end{cases} \quad (5)$$

The PDF of  $X^+$  can be easily deduced from the CDF by derivation. We thus have

$$E[X^+] = \int_0^\delta t f_{X^+}(t) dt = \frac{\delta}{1 - e^{-\alpha\delta}} - \frac{1}{\alpha}. \quad (6)$$

During interval  $[t_0, t_0 + \delta)$ , the MN experiences  $\alpha\delta$  area crossings on average and creates  $1 - \exp(-\alpha\delta)$  tunnels for reachability on average. As a probability can be seen as the number of events divided by the number of experiences, the probability to create a tunnel for reachability given that the MN leaves an area is:

$$P(X \neq 0) = \frac{1 - e^{-\alpha\delta}}{\alpha\delta}. \quad (7)$$

### 5.3.3 Lifetime of incoming-sessions tunnel

We define the current set of sessions on an MN as the set of all those sessions which started in the current MAR area being visited by the MN and which are still active. Let us assume there is a set of  $k$  active incoming sessions that were set up between  $t_0$  and  $t_0 + \delta$  and that are still alive at time  $t_0 + \delta$ . Let  $Y_k$  be the service-time of such a set. The set is released when all  $k$  incoming sessions of the set are terminated. We have:

$$F_{Y_k}(t) = P(Y_k \leq t) = (1 - e^{-\mu_s t})^k, \quad (8)$$

where  $k \geq 0$ . We then have:

$$F_Y(t) = \sum_{k=0}^{\infty} \pi_k^{\text{IS}}(t_0 + \delta) (1 - e^{-\mu_s t})^k, \quad (9)$$

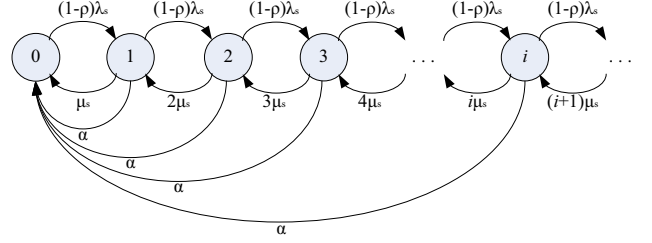


Fig. 8 CTMC of the system (a state is defined as the number of active outgoing sessions in the current set of an MN)

where  $\pi_k^{\text{IS}}(t_0 + \delta)$  is the probability that an MN has  $k$  active incoming sessions at time  $t_0 + \delta$ .

As the chosen value of  $\delta$  is larger than the session duration and the inter-arrival, the system reaches the steady-state and hence the distribution of  $\pi_i^{\text{IS}}$  is a Poisson distribution. We thus have:

$$\pi_i^{\text{IS}}(t_0 + \delta) \approx e^{-\frac{\rho\lambda_s}{\mu_s}} \frac{\left(\frac{\rho\lambda_s}{\mu_s}\right)^i}{i!}. \quad (10)$$

By combining (9) and (10), we obtain after several elementary steps:

$$F_Y(t) \approx e^{-\frac{\rho\lambda_s}{\mu_s}} e^{-\mu_s t}. \quad (11)$$

### 5.3.4 Lifetime of outgoing-sessions tunnel

The CDF of a set of  $k$  outgoing sessions is computed similarly to the one of incoming sessions (see (8)):

$$F_{Z_k}(t) = P(Z_k \leq t) = (1 - e^{-\mu_s t})^k. \quad (12)$$

Equation (9) is also valid for outgoing sessions but with some slight modifications as follows:

$$F_Z(t) = \sum_{i=0}^{\infty} \pi_i^{\text{OS}} (1 - e^{-\mu_s t})^i, \quad (13)$$

where  $\pi_i^{\text{OS}}$  is the probability that an MN that leaves an area has  $i$  active outgoing sessions that were set up in that area.

Note that  $P(Z = 0)$  is the same as the probability that a departing MN has no active outgoing session in its current set. Therefore:

$$P(Z = 0) = \pi_0^{\text{OS}}. \quad (14)$$

The tunnel is used only by outgoing sessions that were set up in MAR1 and that are still alive when the MN leaves MAR1. The number of sessions set up in the same area is a continuous-time Markov chain (CTMC), as shown in Fig. 8.

The following equilibrium equations can be written to find out the steady-state probability vector  $\pi^{\text{OS}}$ :

$$\begin{cases} ((1-\rho)\lambda_s + i\mu_s + \alpha)\pi_i^{\text{OS}} = \\ (1-\rho)\lambda_s\pi_{i-1}^{\text{OS}} + (i+1)\mu_s\pi_{i+1}^{\text{OS}}, \quad i \geq 1, \\ (1-\rho)\lambda_s\pi_0^{\text{OS}} = \alpha \left( \sum_{j=1}^{\infty} \pi_j^{\text{OS}} \right) + \mu_s\pi_1^{\text{OS}}, \\ \sum_{j=0}^{\infty} \pi_j^{\text{OS}} = 1. \end{cases} \quad (15)$$

System of equations (15) can easily be solved by numerical methods.

### 5.3.5 Global expression of tunnel lifetime

We now compute the CDF of  $R$  and  $R^+$ . The probability to create a tunnel is given by  $P(R \neq 0)$ . Using (4), we deduce:

$$P(R = 0) = P(X = 0) \times P(Z = 0). \quad (16)$$

If  $X = 0$  then  $Y = 0$  and  $\max(X + Y, Z) = Z$ . If  $X \neq 0$  then  $X = X^+$  and  $F_{\max(X+Y, Z)}(t) = F_{X+Y}(t) \times F_Z(t)$ . Hence, using (1), we can write (with short notation  $P_{X0} = P(X = 0)$  and  $P_{\bar{X}0} = 1 - P_{X0}$ ):

$$\begin{aligned} F_{R^+}(t) \\ = \frac{-P_{X0}P_{Z0} + P_{X0}F_Z(t) + P_{\bar{X}0}F_{X+Y}(t)F_Z(t)}{1 - P_{X0}P_{Z0}} \end{aligned} \quad (17)$$

if  $t > 0$  and  $\bar{F}_{R^+}(0) = 0$ .

As  $X^+$  and  $Y$  are non negative random variables, we have:

$$F_{X+Y}(t) = \int_0^t F_Y(t-u)f_{X^+}(u)du. \quad (18)$$

By combining (17), (5), (9), (13) and (18), we can compute the CDF of the tunnel lifetime. With similar considerations, we can compute the average value of  $R^+$ , which is as follows:

$$E[R^+] = \frac{P_{X0}E[Z] + (1-P_{X0})E[\max(X^+ + Y, Z)]}{1 - P_{X0}P_{Z0}}, \quad (19)$$

which can finally be expressed as follows:

$$\begin{aligned} \tau^{\text{TBM}} &= \frac{P_{X0}}{1 - \pi_0^{\text{OS}}P_{X0}} \\ &\times \left( \sum_{i=1}^{\infty} \pi_i^{\text{OS}} \sum_{j=1}^i \frac{(-1)^{j-1} \binom{i}{j}}{j\mu_s} + \frac{P_{\bar{X}0}}{P_{X0}} \int_0^{\infty} (1 - F_R(t))dt \right) \end{aligned} \quad (20)$$

where  $\binom{i}{j} = \frac{i!}{j!(i-j)!}$ ,  $\pi_i^{\text{OS}}$  is calculated using (15), and  $F_R(t)$  is expressed as

$$F_R(t) = \sum_{i=0}^{\infty} \pi_i^{\text{OS}} (1 - e^{-\mu_s t})^i \int_0^{\min(t, \delta)} g(u)du \quad (21)$$

$$\text{with } g(u) = \frac{\alpha e^{\alpha(\delta-u)}}{e^{\alpha\delta} - 1} e^{-\frac{\rho\lambda_s}{\mu_s} e^{-\mu_s(t-u)}}.$$

In SDM, a tunnel is maintained only for providing session-continuity as the SIP server is updated whenever an MN joins a new MAR. SDM is then analyzed like TBM with  $\rho = 0$ . Hence, several equations are simplified and we have:

$$\tau^{\text{SDM}} = \frac{1}{1 - \pi_0^{\text{OS}}} \sum_{k=1}^{\infty} \pi_k^{\text{OS}} \sum_{j=1}^k (-1)^{j-1} \binom{k}{j} \frac{1}{j\mu_s}. \quad (22)$$

## 5.4 Context loads on a mobility anchor

After developing the tunnel lifetime expression, we are able now to calculate the average number of simultaneous tunnels on a mobility anchor. This metric represents the size of binding cache table because each tunnel has an entry for its context. As the size of binding cache table increases, the scalability issues and failure risks raise more. Moreover, the latency of a binding lookup increases, where such lookup is performed for each tunneled data packet. Consequently, this metric reflects the context loads on the mobility anchor.

### 5.4.1 TBM case

Let  $\lambda_{\text{MN}}$  be the arrival rate of MNs in an area. As it is proportional to the average number of MNs in an area, we simply have  $\lambda_{\text{MN}} = \alpha E[N]$ . In the model, an MN that leaves an area immediately enters another area. The departure rate of MNs is then  $\lambda_{\text{MN}}$ . Each MN that leaves an area sets up a tunnel with probability  $P(R \neq 0)$ . Let  $\lambda_t^{\text{TBM}}$  be the arrival rate of the tunnels on an MAR in TBM. By combining (16), (7), (14), we have:

$$\lambda_t^{\text{TBM}} = \alpha E[N] \left[ 1 - \pi_0^{\text{OS}} \left( 1 - \frac{1 - e^{-\alpha\delta}}{\alpha\delta} \right) \right]. \quad (23)$$

A tunnel has two end-points, one in the anchor MAR and the other in the visited MAR. As we assume a regular homogeneous network, the average number of anchor endpoints in an MAR is equal to the average number of visited endpoints. The average number  $M^{\text{TBM}}$  of tunnels on an MAR in TBM is

$$M^{\text{TBM}} = 2 \lambda_t^{\text{TBM}} \tau^{\text{TBM}} \quad (24)$$

where  $\tau^{\text{TBM}}$  is given by (20).

#### 5.4.2 SDM Case

In SDM, a tunnel is created if the MN leaves an area, where at least one session started and if that session is still active. The arrival rate  $\lambda_t^{\text{SDM}}$  of tunnels in SDM is then:

$$\lambda_t^{\text{SDM}} = \alpha E[N](1 - \pi_0), \quad (25)$$

where  $\pi_0$  given by (15) is the probability of having no session in the MN for  $\rho = 0$ . The average number  $M^{\text{SDM}}$  of tunnels on an MAR in SDM is

$$M^{\text{SDM}} = 2 \lambda_t^{\text{SDM}} \tau^{\text{SDM}} \quad (26)$$

where  $\tau^{\text{SDM}}$  is given by (22).

#### 5.5 Signaling loads on a mobility anchor

In this section, we study the signaling loads on an MAR, i.e. the arrival rate of different requests on an MAR. We consider three types of requests on an MAR that are related to IPv6 address allocation as well as to tunneling mechanisms, as follows.

- A request for a new IPv6 address by a newly joined MN.
- Requests for the creation/release of an anchor-endpoint of a tunnel and the corresponding creation/release of a visited-endpoint of the tunnel (total of four requests for each tunnel as there are two endpoints of a tunnel that are created and released).
- For a tunnel modification event, there is a request for the creation of a new visited-endpoint of a tunnel on the new visited MAR, a request from the new visited MAR to the anchoring MAR (of the tunnel) for the modification of the tunnel, and a request from the anchoring MAR (of the tunnel) to the old visited MAR for the release of the old visited-endpoint of the tunnel (total of three requests for a modification of the tunnel).

Let  $\lambda_{\text{Req}}^{\text{TBM}}$  be the arrival rate of requests on an MAR in TBM. In steady-state, corresponding to the above three requests types, it can be computed as follows:

$$\lambda_{\text{Req}}^{\text{TBM}} = \lambda_{\text{MN}} + 4\lambda_t^{\text{TBM}} + 3\lambda_t^{\text{TBM}}\alpha\tau^{\text{TBM}}, \quad (27)$$

where  $\lambda_{\text{MN}} = \alpha E[N]$  and  $\alpha\tau^{\text{TBM}}$  is the average number of modifications of an anchor-endpoint of a tunnel on an MAR in TBM.

Let  $\lambda_{\text{Req}}^{\text{SDM}}$  be the arrival rate of requests on an MAR in SDM. Similar to Eq. (27), it can be computed as follows:

$$\lambda_{\text{Req}}^{\text{SDM}} = \lambda_{\text{MN}} + 4\lambda_t^{\text{SDM}} + 3\lambda_t^{\text{SDM}}\alpha\tau^{\text{SDM}}, \quad (28)$$

where  $\alpha\tau^{\text{SDM}}$  is the average number of modifications of an anchor-endpoint of a tunnel on an MAR in SDM.

#### 5.6 Tunneling loads on a mobility anchor

First, we calculate the total number of data packets that are handled by an MAR. The average number of MNs per MAR's area is  $E[N]$ . For each MN, the mean arrival rate of sessions is  $\lambda_s$ . Each session generates  $E[P]$  data packets on average. Then, the total number of data packets handled by an MAR per second is  $E[N]\lambda_s E[P]$  on average.

Some of the data packets are tunneled and hence need a specific treatment. In particular, the MAR needs to add or remove the tunneling overhead before routing the data packet. In order to accomplish this, the MAR needs to look up the binding cache table for each data packet. This generates additional processing loads. Note that these loads are different than the contexts loads (i.e. number of tunnels on a mobility anchor) derived above.

Hereafter, we calculate the probability that a data packet is tunneled and then the average number of tunneled data packets handled by a mobility anchor per second.

##### 5.6.1 SDM case

In SDM, a new (outgoing or incoming) session is anchored at the current MAR where the session is initiated. Before undergoing any IP-handover, the data packets of the session are routed optimally without tunneling. After undergoing one or more IP-handovers, the data packets are tunneled between the new MAR and anchoring MAR of that sessions. The probability that a traffic is a tunneled traffic is expressed as follows [21]:

$$P_t = \frac{\alpha}{\mu_s + \alpha}. \quad (29)$$

Consequently, the average number of tunneled data packets handled by an MAR per second is:

$$D_{\text{tunneled}}^{\text{SDM}} = P_t E[N]\lambda_s E[P]. \quad (30)$$

##### 5.6.2 TBM case

In TBM, the outgoing sessions are treated similarly to SDM. On the other hand, the incoming sessions are treated depending on the MN's registration status. First of all, the probability of using tunneling for an incoming session is  $1 - e^{-\alpha\delta}$ . Then, there are two cases:

- If the registered IP address in the location server is the MN's current one, then it is treated similarly to SDM.

This occurs when the incoming session arrives between  $t_0$  and  $t_1$  and hence with probability  $\frac{E[T_r]}{\delta} = 1 - \frac{E[X^+]}{\delta}$ .

- If the registered IP address in the location server is a previous one, then all the data packets of that incoming session are tunneled. This occurs when the incoming session arrives between  $t_1$  and  $\delta$  and hence with probability  $\frac{E[X^+]}{\delta}$ .

Accordingly, the probability that a traffic is a tunneled traffic in TBM is:

$$P_{t,TBM} = (1 - \rho)P_t + \rho(1 - e^{-\alpha\delta}) \left( \frac{\delta - E[X^+]}{\delta} P_t + \frac{E[X^+]}{\delta} \right). \quad (31)$$

Consequently, the average number of tunneled data packets handled by an MAR per second is:

$$D_{\text{tunneled}}^{\text{TBM}} = P_t^{\text{TBM}} \lambda_s E[P]. \quad (32)$$

## 6 Numerical results

After defining and deriving the different metrics, we show in this section the numerical results. First, we consider the network total area to be 300,000 km<sup>2</sup>, which is close to the densely populated area of France (total area of France is 500,000 km<sup>2</sup>). We also consider the total number of MNs to be 30 million, which is almost the number of mobile subscribers of the French telecommunications operator Orange in France.

Then, we investigate several scenarios as shown in Table 2, considering anchor distribution levels, from a rather centralized approach with 10 anchors to a decentralized approach with 10<sup>6</sup> anchors. Note that with 3 × 10<sup>5</sup> anchors, the area per

anchor is 1 km<sup>2</sup>, which corresponds to 0.33 km<sup>2</sup> radio cells as tri-sectorization is generally used. In other words, each base station is a MAR in that case. We get  $\alpha = 2.06 \times 10^{-3} \text{ s}^{-1}$  at 1 m/s speed for a radio cell and the handover probability is thus  $\alpha/(\alpha + \mu) = 0.4$ . In other words, about 60% sessions are non-mobile, which is consistent with the statistics of [32].

When it is relevant, we compare our results with SIP over PMIP. In present mobile network architectures, there is generally one national Packet Data Gateway (PGW), which acts as a LMA, and a set of Serving Gateway (SGW) that act as MAG.

In order to validate the numerical results, we have developed a discrete event-based simulator in C++. As it is not possible to simulate the area of France with 30 million MNs, we consider a wrap-around hexagon network structure of 19 MAR areas. The MAR's coverage area as well as the number of MNs in each MAR area vary depending on the scenario in such a way that these 19 MAR areas represent a part of the full network described above. For example, in the scenario of 1000 MAR areas in France, the MAR's coverage area is  $\frac{300,000}{1000} = 300 \text{ km}^2$  and the number of MNs in each MAR area is  $\frac{30 \times 10^6}{1000} = 30,000$ . We made 20 simulations runs and computed the 95% confidence intervals. As it can be seen in Figs. 9, 10, and 11, the confidence intervals are very small in most cases.

### 6.1 Signaling loads on the location server

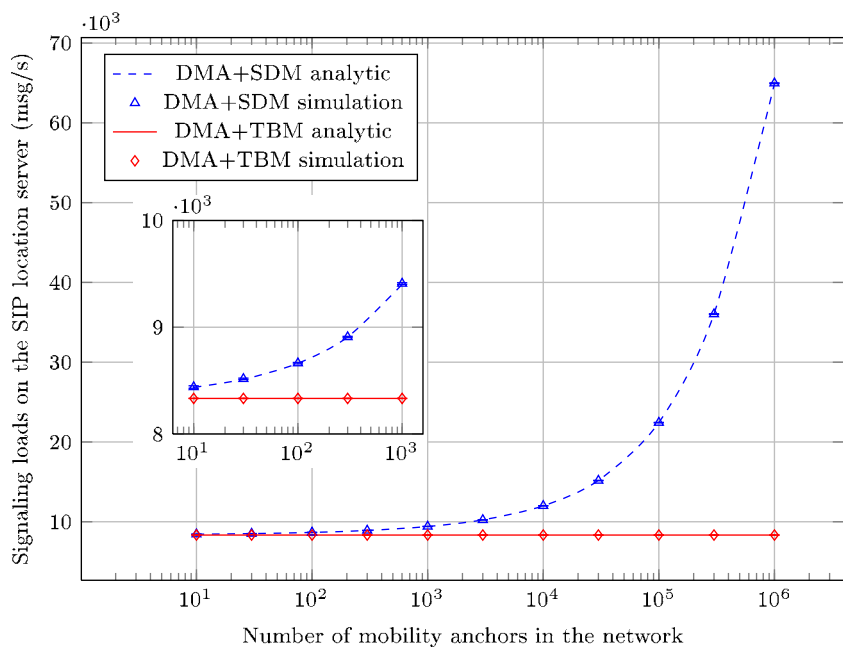
Based on (3) and (2) Fig. 9 shows the arrival rate of location updates on the location server (in message per second) with respect to the number of mobility anchors in the network.

In TBM, the location update loads on the location server are constant in all scenarios since it does not depend on the area size but only on the registration lifetime. Based on the SIP standard, the default value of the registration lifetime is 3600 s and hence the arrival rate of registrations for one MN is 1/3600 per second and for all MNs

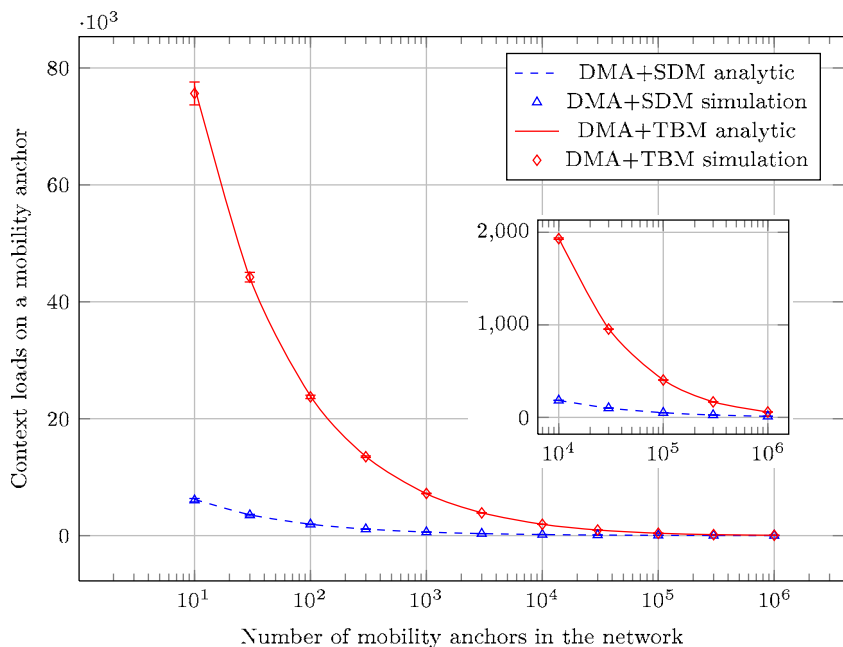
**Table 2** System parameters settings

Parameter		Value or expression
$S_T$	Area of the network	$3 \times 10^5 \text{ km}^2$
$N_{MN}$	Total number of MNs	$30 \times 10^6$
$\delta$	REGISTER message lifetime	3600 s [27]
$\lambda_s$	Session arrival rate	$5/3600 \text{ s}^{-1}$
$1/\mu_s$	Average session duration	350 s
$\rho$	Probability of incoming session	0.5
$E[v]$	MN's average speed	1 m/s
$E[P]$	Average number of packets in a session	1000
$N_T$	Number of MARs	$[10^1, 10^6]$
$S$	MAR's coverage area	$S_T/N_T$
$E[N]$	Expected number of MNs per area	$N_{MN}/N_T$
$\alpha$	Area crossing rate per MN	$\frac{\sqrt{8\sqrt{3}}}{\pi} \frac{E[v]}{\sqrt{S}}$

**Fig. 9** Signaling loads on the location server



**Fig. 10** Context loads on a mobility anchor



is  $30 \times 10^6 / 3600 \simeq 8333$  per second. In SDM, the arrival rate of requests highly depends on the area size since the MN sends a registration upon each IP-handover and also for refreshing. Consequently, it appears that the location update loads increase significantly for relatively small areas since the IP-handovers are more frequent.

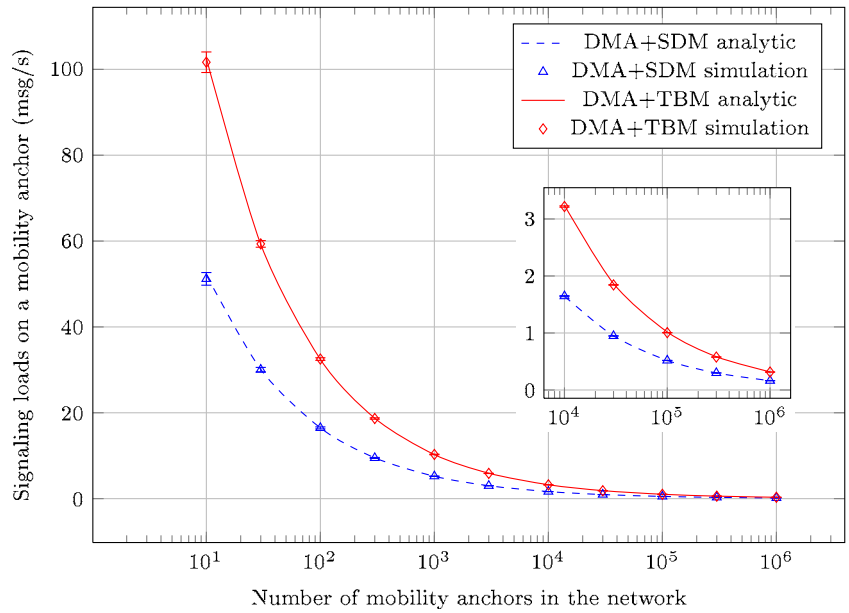
## 6.2 Context loads on a mobility anchor

Figure 10 shows the context loads on a mobility anchor with respect to the number of mobility anchors in the network. It was obtained from (24) and (26).

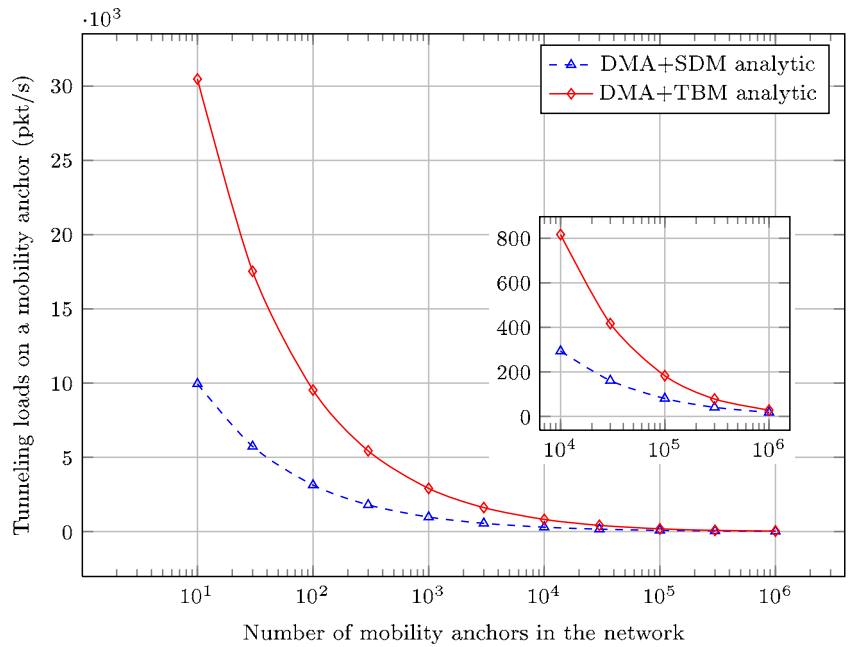
As described before, the context loads depend on the number of tunnels at an anchor. In SDM, there is no need for reachability tunnels since the location server is always up-to-date. On the other hand, there is a need in TBM. Consequently, the average number of tunnels is always smaller for SDM than for TBM. However, as the number of mobility anchors increases, the difference between SDM and TBM diminishes. Hereafter, we clarify the reason behind this.

The period of registration is much larger than the mean session duration. For large areas, the probability that an MN sends a registration before its departure from the area is high. Hence, the probability that a tunnel is generated to provide at

**Fig. 11** Signaling loads on a mobility anchor



**Fig. 12** Tunneling loads on a mobility anchor



least the reachability service is high for each IP-handover of the MN with TBM and the mean lifetime of a tunnel is mainly influenced by the percentage of the tunnels which at least provide the reachability service. For small areas, there is a high probability that an MN has several IP-handovers between two successive registrations. Hence, lot of IP-handovers do not generate a new tunnel (and only modify an existing tunnel). The benefit of SDM is then reduced.

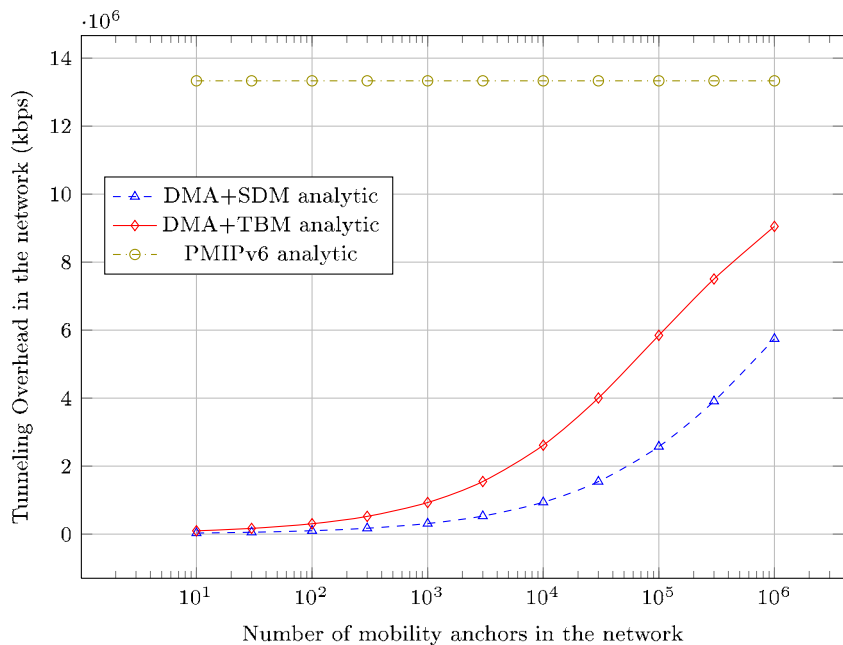
Compared to centralized schemes, both DMA modes reduce the context loads significantly. In a centralized scheme such as PMIPv6, the LMA manages the context of all the MNs, i.e. 30 millions in our case. With  $3 \times 10^5$  anchors, each

anchor has to manage only 26 contexts with SDM and 84 contexts with TBM. The total number of tunnels is  $3.9 \times 10^6$  with SDM and  $25.2 \times 10^6$  with TBM in the whole network. Though TBM gives the same order of magnitude as PMIP ( $30 \times 10^6$  tunnels), DMA gives a clear benefit in both modes as the tunnel management is shared between a large number of nodes.

### 6.3 Signaling loads on a mobility anchor

Figure 11 shows the signaling loads on a mobility anchor (in message per second) with respect to the number of mobility

**Fig. 13** Tunneling overhead in kbps in the network



anchors in the network. It was obtained based on (27) and (28). It appears that these loads are more in TBM than in SDM especially for large areas. This difference diminishes for small areas, i.e for high number of anchors.

The requests sent to an anchor include several types such as IPv6 address configuration, creation/release of a tunnel, and modification of a tunnel (cf. Sect. 5.5). The former does not depend on the mode (similar in both SDM and TBM) and hence the tunneling mechanisms are the main reason behind the variation in signaling loads between SDM and TBM.

When regarding the contribution of each type in the overall loads, we can notice the following. For large areas, the tunnel creation/release requests are the main contributor. For small areas, the tunnel modification requests are the main contributor.

#### 6.4 Tunneling loads and tunneling overhead

For each tunneled data packet, the MAR needs to perform a binding cache lookup and then to treat (add or remove) the tunneling overhead of that packet. This generates additional processing loads on the MAR and hence the operator needs to estimate these loads in the different scenarios. Figure 12 shows the tunneling loads given by (30) and (32) on a mobility anchor (in packet per second) with respect to the number of mobility anchors in the network. In particular, it presents the average number of tunneled data packets handled by an anchor per second. For relatively large areas, these loads are much higher in TBM than in SDM. For smaller areas, TBM and SDM tend to perform likewise.

Each tunneled data packet has an additional overhead, i.e. IPv6 header, of 40 bytes. With PMIPv6, all packets in the

network have the 40-byte overhead. In SDM and TBM, we multiply the tunneling load given by (30) and (32) by the number of anchors and by  $40 \times 8$  to get the tunneling overhead in bps in the whole network. The result is shown in Fig. 13 for both TBM and SDM modes and also for PMIPv6. This represents the traffic loads on the network links due to tunneling. As the MAR's coverage area decreases, a session experiences more handovers on average and hence the tunneled traffic increases. However, the routing is more performant since the MARs are closer to each other. The tunneling overhead traffic in TBM is higher than that in SDM since in TBM the location server is not updated at each handover and hence some of the incoming sessions are tunneled entirely. On the other hand, all the data packets are tunneled in PMIPv6 and hence it has higher overhead.

## 7 Conclusions

In this paper, we have carried out a novel performance analysis on DMA, which is one of the main DMM schemes for IPv6 mobile networks. After presenting DMA, we have discussed the reachability support at IP and upper layers and also investigated different modes of location update, namely SIP default mode (SDM) and time-based mode (TBM).

The results show that TBM can assure stable loads on the location server compared to SDM. This gain of TBM on SDM is more crucial with dense deployment of mobility anchors. On the other hand, SDM generally introduces less loads on a mobility anchor compared to TBM especially with non-dense deployment of mobility anchors. To sum

up, TBM is more advantageous with highly-dense deployment of mobility anchors since it reduces the loads on the location server significantly compared to SDM while not much increasing the loads on the mobility anchors. Besides, SDM is more advantageous with non-dense deployment of mobility anchors since it reduces the loads on the mobility anchors while not much increasing the loads on the location server. The obtained simulation results guide the mobile network operator firstly to estimate the loads on different network entities in order to deploy the appropriate equipment, and secondly to choose the most suitable protocols and modes for their network architecture and deployment scenario.

## References

1. GSM Association (GSMA). (2013). The Mobile Economy 2013.
2. Ericsson. (2013). Ericsson Mobility Report.
3. Cisco. (2013). Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017.
4. IETF RFC 3775. (2011). Mobility support in IPv6, Jun. 2004. Updated by: IETF RFC 6275.
5. IETF RFC 5213. (2008). Proxy Mobile IPv6
6. IETF, DMM Charter, Available online: <http://datatracker.ietf.org/wg/dmm/charter/>
7. Chan, Chan, Yokota, H., Xie, J., Seite, P., & Liu, D. (2011). Distributed and dynamic mobility management in mobile internet: Current approaches and issues. *Journal of Communications*, 6(1), 4–15.
8. IETF, DMM Documents, Available online: <http://datatracker.ietf.org/wg/dmm/>
9. Lee, J. H., et al. (2010). Cost analysis of IP mobility management protocols for consumer mobile devices. *IEEE Transactions on Consumer Electronics*, 56(2), 1010–1017.
10. Pack, S., et al. (2007). Adaptive Route Optimization in Hierarchical Mobile IPv6 Networks. *IEEE Transactions on Mobile Computing*, 6(8), 903–914.
11. Skorepa, M., & Klugl, R. (2013). Enhanced analytical method for IP mobility handover schemes cost evaluation. *Telecommunication Systems*, 52(3), 1573–1582.
12. Ali-Ahmad, H., Lagrange, X., Ouzzif, M., & Bertin, P. (2013). Comparative cost analysis on global mobility management schemes in IPv6 mobile networks. In *VTC*. IEEE.
13. Lee, J. H., Han, Y. H., Gundavelli, S., & Chung, T. M. (2009). A comparative performance analysis on Hierarchical Mobile IPv6 and Proxy Mobile IPv6. *Telecommunication Systems*, 41(4), 279–292.
14. Wang, Q., & Abu-Rgheff, M. A. (2006). Signalling analysis of cost-efficient mobility support by integrating Mobile IP and SIP in all IP wireless networks. *International Journal of Communication Systems*, 19(2), 225–247.
15. Makaya, C., & Pierre, S. (2008). An analytical framework for performance evaluation of IPv6-based mobility management protocols. *IEEE Transactions on Wireless Communications*, 7(3), 972–983.
16. Munir, K., Lagrange, X., Bertin, P., Ouzzif, M., & Guillouard, K. (2015). Performance analysis of mobility management architectures in cellular networks. *Telecommunication Systems*, 59(2), 211–227.
17. Lee, J. H., Bonnin, J. M., You, I., & Chung, T. M. (2013). Comparative handover performance analysis of IPv6 mobility management protocols. *IEEE Transactions on Industrial Electronics*, 60(3), 1077–1088.
18. Lee, K. H., Lee, H. W., Ryu, W., & Han, Y. H. (2013). A scalable network-based mobility management framework in heterogeneous IP-based networks. *Telecommunication Systems*, 52(4), 1989–2002.
19. Giust, F., Bernardos, C. J., & de la Oliva, A. (2014). Analytic evaluation and experimental validation of a network-based IPv6 distributed mobility management solution. *IEEE Transactions on Mobile Computing*, 13(11), 2484–2497.
20. Condeix, T., & Sargento, S. (2014). Studying the integration of distributed and dynamic schemes in the mobility management. *Computer Networks*, 60, 46–59.
21. Ali-Ahmad, H., Ouzzif, M., Bertin, P., & Lagrange, X. (2014). Performance analysis on network-based distributed mobility management. *Wireless Personal Communications (WPC)*, 74(4), 1245–1263.
22. Bertin, P., Bonjour, S., & Bonnin, J.M. (2009). An evaluation of dynamic mobility anchoring. In *VTC*. IEEE.
23. Louin, P., & Bertin, P. (2011). Network and host based distributed mobility. In *WPMC* IEEE:.
24. Ali-Ahmad, H., Ouzzif, M., Bertin, P., & Lagrange, X. (2013). Distributed mobility management: approaches and analysis. In *Proceedings of the IEEE ICC—telecommunication standards*.
25. Gohar, M., & Koh, S. J. (2012). Network-based distributed mobility control in localized mobile LISP networks. *IEEE Communications Letters*, 16(1), 104–107.
26. Hossain, M. S., & Atiquzzaman, M. (2013). Cost analysis of mobility protocols. *Telecommunication Systems*, 52(4), 2271–2285.
27. IETF RFC 3261. (2002). SIP: Session initiation protocol.
28. Bertin, P., Bonjour, S., & Bonnin, J. M. (2008). A distributed dynamic mobility management scheme designed for flat IP architectures. In *NTMS*. IEEE.
29. Bertin, P., Bonjour, S., & Bonnin, J.M. (2009). Distributed or centralized mobility? In *GLOBECOM*. IEEE.
30. Seite, P., & Bertin, P. (2014). Distributed mobility anchoring, draft-seite-dmm-dma-07, IETF.
31. Sadhukhan, P., Das, P. K., & Saha, S. (2013). Hybrid mobility management schemes integrating mobile IP and SIP for seamless invocation of services in All-IP network. *Telecommunication Systems*, 52(4), 2027–2046.
32. Cisco. (2010). Cisco visual networking index: Global mobile data traffic forecast update, 2009–2014.
33. Schulzrinne, H., & Wedlund, E. (2000). Application-layer mobility using SIP. *ACM Mobile Computing and Communications Review*, 4(3), 47–57.
34. RFC 7333. (2014). Requirements for distributed mobility management.
35. Akyildiz, I. F., & Wang, W. (2002). A dynamic location management scheme for next-generation multitier PCS systems. *IEEE Transactions Wireless Communications*, 1(1), 178–189.