Reducing Signaling Overhead for Femtocell/Macrocell Networks

Huai-Lei Fu, *Student Member*, *IEEE*,
Phone Lin, *Senior Member*, *IEEE*, and Yi-Bing Lin, *Fellow*, *IEEE*

Abstract—Femtocell technology has been proposed to offload user data traffic from the macrocell to the femtocell and extend the limited coverage of the macrocell in mobile communications networks. In existing commercial femtocell/macrocell networks, a widely accepted solution to configure the location areas (LAs) is to partition the femtocells overlapped with a macrocell into small groups and to assign each group with a unique LA ID different from that of the macrocell. Such configuration can reduce the paging cost in the mobility management, but increases registration signaling overhead due to discontinuous coverage of femtocells. To reduce signaling overhead in the femtocell/macrocell network, we propose a delay registration (DR) algorithm that postpones the registration until the delay timer expires when the mobile station (MS) moves into the overlapped femtocell. Analytical models and simulation experiments are proposed to investigate the performance of the DR algorithm. Our analytical models are generally enough to accommodate various MS mobility behaviors. Our study can provide guidelines for the operators to set up a delay timer to reduce signaling overhead while sustaining the traffic offloading capability of the femtocell.

Index Terms—Femtocell, macrocell, mobile communications networks, mobility management

1 Introduction

In a mobile communications network (MCN), the service area is populated with base stations (BSs). The radio coverage of a BS (or a sector of the BS) is called a cell. In an outdoor environment, network operators deploy macrocells with radio coverage area of 0.5 to 2 kilometer-radius. The fast growing population of mobile users leads to an exponential increase in user data traffic demand for the MCN, but the capacity of outdoor macrocells are not sufficient to satisfy such demand.

The femtocell [1], [2], [3] (also called home BS) with low deployment cost was proposed to offload the user data traffic to the macrocell and extend the limited coverage of the macrocell. A femtocell is the radio coverage of a short-range, low-cost, and low-power wireless BS, typically covering an area with the radius of 5 to 20 meters [2]. Femtocells operate in the same licensed spectrum as a macrocell. In femtocell and macrocell deployment environment planned by some mobile network operators, hundreds to thousands of femtocells are overlapped with a macrocell.

In the viewpoint of mobile network operators, traffic offloading is one of the major reasons for femtocell deployment. When a mobile station (MS) resides in a femtocell overlapped with a macrocell, as long as the MS has the authority to access the femtocell, the MS prefers to connect to the femtocell so as to offload traffic to the macrocell.

We name the MCNs with femtocells and macrocells as femtocell/macrocell networks. Fig. 1 depicts an example of the femtocell/macrocell network based on the universal mobile telecommunications system [4]. In this figure, the macrocell $Cell$ (see the solid circle) overlays with the femtocells $cell_1, \ldots, cell_5$ (see the dashed circles). The macrocell connects to the core network (CN; Fig. 1a) through the radio network controller (RNC; Fig. 1b). The femtocells connect to the CN through the femto gateway (Femto GW; Fig. 1c) and a broadband Internet network [5] (Fig. 1d). The Femto GW plays the role of an RNC. Importantly, the service area of femtocells may be discontinued. For example, in Fig. 1, $cell_1, \ldots, cell_5$ can be treated as "islands" that do not overlap with each other.

The cells in the femtocell/macrocell network are grouped into location areas (LAs). Each LA is assigned with a unique LA ID (LAI). The LAs are used for mobility management. The mobility management consists of "registration" and "call termination." Each macrocell or femtocell uses a wireless broadcast channel to broadcast its individually corresponding LAI. The MS listens to the wireless broadcast channel to identify in which LA it resides. When the MS moves from one LA to another (that is, the LAI stored in the MS storage is different from the received LAI), the MS initiates a "registration" to report the LAI of the cell in which it resides to the location database in the CN. This process is also known as "location update." "Call termination" refers to the process of routing an incoming call to an MS. In this process, the CN obtains the LA of the MS by querying the location database, instructs all cells in the LA to "page" the MS, and then sets up the call to the MS through the cell in which the MS responds to the "page." Details of registration and call termination can be found in [6].

- *H.-L. Fu and P. Lin are with the Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106, R.O.C. E-mail: vicfu@pcs.csie.ntu.edu.tw, plin@csie.ntu.edu.tw.*
- *Y.-B. Lin is with Department of Computer Science and Information Engineering, National Chiao Tung University, Engineering Building C, Kuang-Fu Campus, 1001 University Rd., Hsinchu, Taiwan 300, R.O.C. E-mail: liny@csie.nctu.edu.tw.*

Manuscript received 7 Dec. 2011; revised 23 Mar. 2012; accepted 11 May 2012; published online 31 May 2012.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2011-12-0659. Digital Object Identifier no. 10.1109/TMC.2012.132.

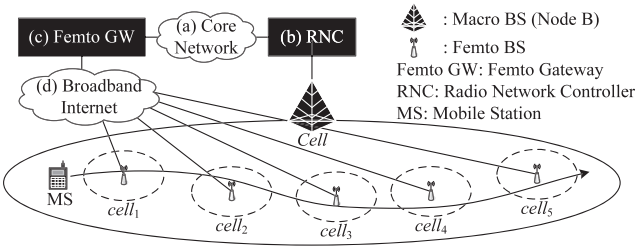


Fig. 1. An example of the femtocell/macrocell network architecture.

There are two alternatives to assign LAs in the femtocell/macrocell network [7], [8]. In the first alternative, all femtocells overlapped with a macrocell are assigned to the same LA as that of the macrocell. If more than hundreds of femtocells are overlapped with the macrocell, all of the femtocells and the macrocell must together page an MS for its incoming call, cumulating high paging cost. In the second alternative, the femtocells overlapped with the macrocell are partitioned into small groups, and each group is assigned with a unique LAI different from that of the macrocell.

Compared with the first alternative, the second alternative significantly reduces the paging cost [8]. Hence, the second alternative is more often exercised in existing commercial MCNs. However, a potential problem of this second alternative is that registration occurs every time the MS moves between the macrocell and the overlapped femtocell, increasing signaling overhead to the network. Take, for example, the network architecture in Fig. 1, where the LA of the macrocell $Cell$ is L_m, and L_f is the LA assigned to the overlapped femtocells $cell_1$, $cell_2, \ldots,$ $cell_5$. Since the service area of these femtocells is discontinuous, when an MS moves from one femtocell to another, it crosses through the macrocell and switches between L_m and L_f, triggering registration. For example, the MS movement $Cell \rightarrow cell_1 \rightarrow Cell \rightarrow cell_2 \rightarrow Cell \rightarrow cell_3 \rightarrow Cell \rightarrow cell_4 \rightarrow Cell \rightarrow cell_5 \rightarrow Cell$ results in as many as 10 registrations.

In this paper, we focus on the location update issue in the femtocell/macrocell networks. To be more specific, we investigate the mobility management for the second alternative of LA layout. The major challenge is that the frequent execution of registration due to small and discontinuous femtocell coverage causes high signaling overhead. Note that some mobile network operators exercise the business models such that private home BSs can be purchased by users, and each home BS can only be accessed by a particular user's mobile devices. In this scenario, not many subsequent registrations will be sent to many femtocells, so the frequent registration issue does not exist.

Following most of location update studies [9], [10], [11], we do not consider channel capacity of femtocells or macrocells, which is an independent issue dealt by admission control. The channel capacity issue is out of the scope of this paper and should be treated separately to investigate original call admission and handoff strategies for femtocell/macrocell networks.

Previous works [12], [13], [14] have addressed the mobility management issue for the macrocell-microcell hierarchical mobile networks. Both [12] and [13] addressed the mobility management issue for the macrocell-microcell hierarchical mobile networks in which microcell radio coverage is continuous, i.e., when an MS moves from one microcell to another, it does not cross any overlapped macrocell. On the other hand, our work addresses the issue caused by coverage discontinuity of femtocells. In [12], the authors proposed a macro-micro paging scheme (based on the predefined delay constraints) to balance the paging load between the macrocell and overlapped microcell. They did not focus on the signaling overhead reduction for registration (i.e., location update), and therefore, the proposed technique does not reduce registration overhead. In [13], the authors proposed a cross-tier registration and paging scheme. In this scheme, a registration is executed only when an MS crosses the boundary between two LAs in the macrocell tier. Based on the paging load, the CN determines whether to page an MS through the macrocell tier or microcell tier. The proposed scheme requires significant modification on the existing mobility management protocols at the CN side, which is considered impractical in the femtocell/macrocell networks. Work [14] proposed the integration of high-tier mobile networks and low-tier mobile networks, and these networks operate in separate mobility management protocols. Work [14] also proposed intelligent algorithms to determine whether the MS should perform the registration operation when the MS switches tiers. The architecture considered in [14] is more complicated than the femtocell/macrocell networks, and the performance is not as good as our approach due to an extra layer of mobility management integration of two networks.

In this paper, we propose the delay registration (DR) algorithm that postpones registration when an MS moves into the overlapped femtocell in the femtocell/macrocell network. Details of the DR algorithm are given in the next section. We propose analytical models and conduct simulation experiments to investigate the performance of the DR algorithm. In the proposed analytical models, we assume general distributions for the residence time periods in the overlapped femtocell and in the nonoverlapped areas of the macrocell. Our analytical model is generally enough to accommodate various MS mobility behaviors. The analytical model is validated against the simulation experiments.

The remaining parts of the paper are organized as follows: Details of the DR algorithm are given in Section 2. In Section 3, we describe the analytical models. In Section 4, we evaluate the performance of the DR algorithm and provide guidelines for the setup of the delay timer. Section 5 concludes this paper.

2 The DR Algorithm

This section proposes the DR algorithm. The DR algorithm is exercised at the MS to determine whether the registration should be performed. No modification is required at the network side.

To simplify our description, we consider the following MS moving behavior: A macrocell overlays with several femtocells. Let L_m be the LAI assigned to the macrocell and L_f be the LAI assigned to the overlapped femtocells. Initially, an MS is in the nonoverlapped area in the

TABLE 1
Comparison between the 3GPP Algorithms and the DR Algorithm

	w/o femtocell	3GPP algorithms	DR algorithm
Signaling overhead	N/A	high	low
Offload traffic	N/A	high	medium∼high

macrocell, and L_m is the LAI stored in the location database. At time t, the MS moves into the overlapped femtocell (i.e., the MS can also receive L_f) and stays in the overlapped femtocell for t_f. At $t + t_f$, the MS moves from the overlapped femtocell into the macrocell, and can no longer receive L_f.

In the standard 3GPP algorithms for mobility management, at t one registration is executed to change L_m to L_f in the location database, and at $t + t_f$ another registration is performed to change L_f to L_m. Here, two registrations are performed. Note that if the MS just passes by the femtocell, that is, t_f is as short as only a few seconds, we describe the MS movement as *transient*. If *transient* occurs, it is most likely that no call behavior (call origination/termination) will take place during $[t, t + t_f]$. Thus, these two registrations during the *transient* period may be avoided without significantly affecting traffic offloading capability of the femtocell. To avoid registrations during the *transient* period, we propose the DR algorithm that introduces a delay timer to postpone the registration until the timer expires.

At t, the DR algorithm suspends the registration and starts the delay timer with length t_d. Two cases are considered:

- If $t_d < t_f$, as the delay timer expires at $t + t_d$, the MS is still in the overlapped femtocell. The MS stops the delay timer and initiates the registration to change L_m to L_f.
- Otherwise (i.e., $t_d \geq t_f$), the MS stops the delay timer at $t + t_f$, and no registration occurs during $[t, t + t_f]$.

During the period $[t, t + \min(t_d, t_f)]$, if a call requests arrives, the request is potentially handled through the macrocell instead of through the overlapped femtocell, and the traffic may not be offloaded from the macrocell to the femtocell. Clearly, the longer the delay timer, the more signaling overhead (caused by registration) avoided. But meanwhile, it is more likely that the traffic to the macrocell cannot be offloaded. In the next section, we propose analytical models to study the tradeoff between signaling overhead reduction and traffic offloading capability.

To summarize, in Table 1, we compare the signaling overhead and traffic offloading capability of the MCN without femtocells, femtocell/macrocell networks with the standard 3GPP algorithms, and femtocell/macrocell networks with the DR algorithm. The MCN without femtocells does not have traffic offloading capability nor the signaling overhead. The standard 3GPP has the high signaling overhead and high traffic offloading capability. On the other hand, the DR algorithm has the low signaling overhead, but medium to high traffic offloading capability.

The impacts of the DR algorithm on user experience, device energy consumption, and mobile network operators are discussed as follows: With the DR algorithm, by properly setting the delay timer, the signaling overhead is reduced, and the traffic offloading capability of femtocells can still be achieved. In the femtocell/macrocell networks, more bandwidth is available, and the user will have better experience in call admission. For energy consumption, the DR algorithm reduces the number of registrations executed by an MS, thus saves the MS power consumption. For mobile network operators (or carriers), with the DR algorithm, less registrations (i.e., location database updates) will save more network bandwidth and switch capacity. Note that in most designs of CNs, as indicated in [15], 15 percent of computing power is allocated for location update. Intensive registration due to switching among femtocells and macrocells may exceed 15 percent of computing power and degrade the performance for the CN. Our DR algorithm reduces significant registration overhead in femtocell/macrocell networks.

3 Analytical Models

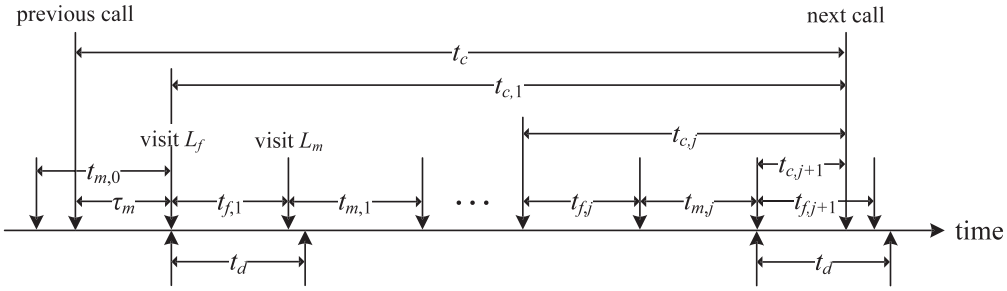
In this section, we propose analytical models to study the impacts of the setup of the delay timer on performance tradeoff between the signaling overhead and traffic offloading capability. Our study provides guidelines for selecting an appropriate t_d such that the signaling overhead decreases, and the traffic offloading capability can still be achieved. The performance metrics derived in our analytical model includes two output measures signaling overhead ratio $r(t_d)$ and potential offload traffic ratio $\rho(t_d)$.

Signaling overhead ratio $r(t_d)$: Define a "crossing" as the event when the MS moves from the macrocell to the overlapped femtocell or vice versa. Consider the intercall arrival time t_c between two consecutive call request arrivals to the MS. Let N_c be the number of crossings the MS has during t_c, and N_r be the number of registrations executed by the MS during t_c. Obviously, $N_c \geq N_r$. If $t_d = 0$ (i.e., the standard 3GPP algorithms are exercised), then $N_r = N_c$. We denote $E[N_r(t_d)|N_c \geq 1]$ for the DR algorithm with delay t_d, and $E[N_r(0)|N_c \geq 1]$ for the 3GPP algorithms. We define $r(t_d)$ as

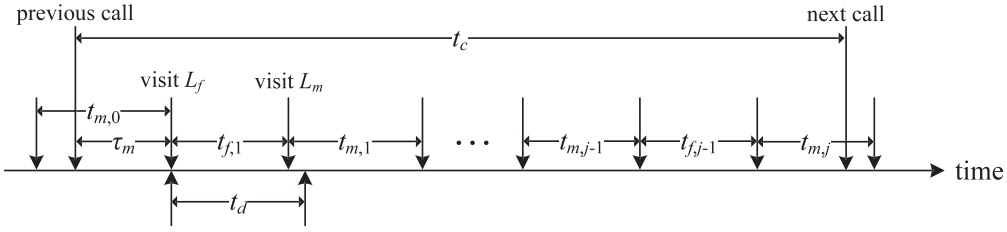
$$r(t_d) = \frac{E[N_r(t_d)|N_c \geq 1]}{E[N_r(0)|N_c \geq 1]} = \frac{E[N_r(t_d)]/\Pr[N_c \geq 1]}{E[N_r(0)]/\Pr[N_c \geq 1]} = \frac{E[N_r(t_d)]}{E[N_r(0)]}, \quad (1)$$

where $0 < r(t_d) \leq 1$. A smaller $r(t_d)$ implies that more signaling overhead for the DR algorithm is reduced.

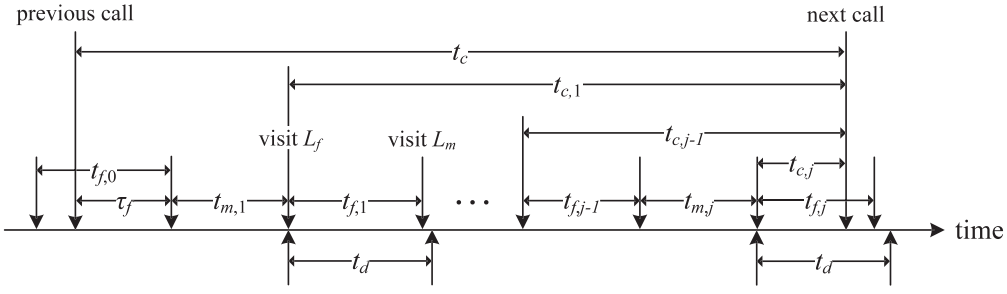
Potential offload traffic ratio $\rho(t_d)$: Let $p(t_d)$ and $p(0)$ denote the probabilities that a call request arrives when the MS registers to the femtocell (i.e., the location information for the MS is the LAI of a femtocell) for the DR algorithm (with



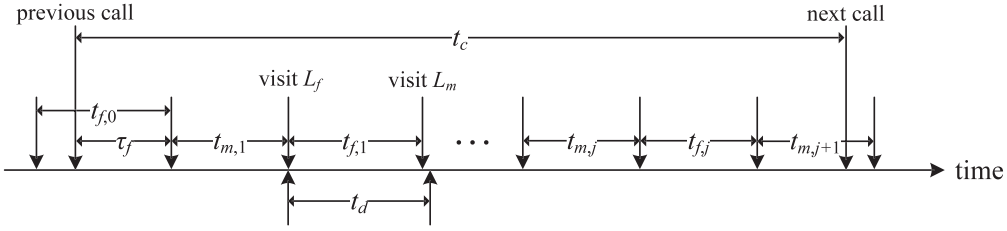
(a) Case Ia: the previous call arrives when the MS is in L_m, and the next call arrives when the MS is in L_f.



(b) Case Ib: the previous call arrives when the MS is in L_m, and the next call arrives when the MS is in L_m.



(c) Case IIa: the previous call arrives when the MS is in L_f, and the next call arrives when the MS is in L_f.



(d) Case IIb: the previous call arrives when the MS is in L_f, and the next call arrives when the MS is in L_m.

Fig. 2. The timing diagrams for MS movement and call activities.

delay period t_d) and for the 3GPP algorithms, respectively. With the DR algorithm, the traffic offloading capability degrades due to the delay timer, i.e., less call request arrivals when the MS registers to the femtocell than that with the 3GPP algorithms. Hence, we have $p(t_d) \leq p(0)$. Define $\rho(t_d)$ as

$$\rho(t_d) = \frac{p(t_d)}{p(0)}, \tag{2}$$

where $0 < \rho(t_d) \leq 1$. A larger $\rho(t_d)$ implies that the DR algorithm causes less degradation for the traffic offloading capability of the femtocell.

We model the behavior of an MS in the femtocell/macrocell network as follows: Consider the timing diagram in Fig. 2. Without loss of generality, the macrocell belongs to LA L_m, and the femtocells belongs to L_f. The radio coverage areas of femtocells in L_f are discontinuous. When the MS passes through these discontinuous femtocells, it switches between L_m and L_f. More specifically, the MS stays in L_f for a period t_f and then moves to L_m for another period t_m.

During the timer period t_c between two consecutive call request arrivals, for $i \geq 1$, let $t_{m,i}$ be the ith time that the MS visits L_m after the previous call request arrival, and $t_{f,i}$ be the ith time that the MS visits L_f after the previous call

request arrival. By convention, when $i = 0$, $t_{m,i}$ ($t_{f,i}$) refers to the time period when the previous call arrives in a macrocell (in a femtocell). In the DR algorithm, each $t_{f,i}$ is associated with the delay timer with length t_d. In the analytical model, we make the following three assumptions:

- *A1*. The call request arrivals to an MS form a Poisson process with rate λ, i.e., the intercall arrival time t_c is exponentially distributed with the density function $f(t) = \lambda e^{-\lambda t}$. The Poisson process assumption is widely used in teletraffic analysis to model the call request arrival behavior [16], [17].
- *A2*. The $t_{m,i}$ and $t_{f,i}$ random variables are i.i.d. with the general density function $f_m(t)$ and $f_f(t)$, the mean $1/\eta_m$ and $1/\eta_f$, and the Laplace transform $f_m^*(s)$ and $f_f^*(s)$, respectively.
- *A3*. The t_d random variable is exponentially distributed with the mean $1/\theta$ and the density function $f_d(t) = \theta e^{-\theta t}$. From implementation perspectives, fixed t_d seems to be a convenient choice. The complexities of generating fixed t_d and exponential t_d are basically the same in terms of implementation cost [18]. We will extend exponential t_d to fixed t_d in simulation experiments.

During t_c, let $t_{c,i}$ be the time period between when the MS visits L_f for the ith time and when the next call request arrives. The density function of $t_{c,i}$ is denoted by $f_{c,i}(t)$. Since t_c is exponentially distributed, from the PASTA property of Poisson processes [19], $t_{c,i}$ has the same distribution as t_c, i.e., the density function $f_{c,i}(t) = \lambda e^{-\lambda t}$. Fig. 2 illustrates four scenarios that may occur during period t_c. In Fig. 2a, the previous call request arrives when the MS is in L_m, and the next call request arrives when the MS is in L_f. In Fig. 2b, both call requests arrive at L_m. In Fig. 2c, both call requests arrive at L_f. In Fig. 2d, the previous call request arrives at L_f, and the next call request arrives at L_m.

The notations used in the analytical models are summarized as follows:

- t_c: The time period between two consecutive call request arrivals to the MS. The density function of t_c is $f(t) = \lambda e^{-\lambda t}$.
- $t_{c,i}$: The time period between when the MS visits L_f for the ith time and when the next call request arrives. The density function of $t_{c,i}$ is $f_{c,i}(t) = \lambda e^{-\lambda t}$.
- t_d: The delay period. The density function of t_d is $f_d(t) = \theta e^{-\theta t}$.
- $t_{f,i}$: The ith time that the MS visits L_f after the previous call request arrival (i.e., the MS is in the overlapped femtocell). The density function, mean, and Laplace transform of $t_{f,i}$ are $f_f(t)$, $1/\eta_f$, and $f_f^*(s)$, respectively.
- $t_{m,i}$: The ith time that the MS visits L_m after the previous call request arrival (i.e., the MS is in the non-overlapped area of the macrocell). The density function, mean, and Laplace transform of $t_{m,i}$ are $f_m(t)$, $1/\eta_m$, and $f_m^*(s)$, respectively.
- N_c: The number of crossings for the MS during t_c.
- N_r: The number of registrations executed by the MS during t_c.
- $r(t_d)$: The signaling overhead ratio for the DR algorithm.
- $\rho(t_d)$: The potential offload traffic ratio for the DR algorithm.
- τ_m: The time period between when the previous call request arrives at L_m and when the MS visits L_f for the first time.

3.1 Derivation of Signaling Overhead Ratio $r(t_d)$

To derive the $r(t_d)$ performance, we first derive $E[N_r(0)]$ and then $E[N_r(t_d)]$, respectively, described in Sections 3.1.1 and 3.1.2. Applying the derivation results, (17) and (30), into (1), we have

$$r(t_d) = \frac{\theta[1 - f_f^*(\lambda + \theta)]}{\lambda + \theta}. \tag{3}$$

3.1.1 Derivation of $E[N_r(0)]$

If $t_d = 0$ (i.e., the 3GPP algorithms are exercised), the MS immediately executes a registration when it moves from L_m to L_f or moves from L_f to L_m. Then, we have

$$E[N_r(0)] = E[N_c].$$

We consider two cases to derive $\Pr[N_c = k]$ for $k \geq 1$:

Case I: The previous call arrives when the MS is in L_m. See Figs. 2a and 2b.

Case II: The previous call arrives when the MS is in L_f. See Figs. 2c and 2d.

Let $N_{I,c}$ and $N_{II,c}$ be the number N_c conditioning on Cases I and II, respectively. We express $\Pr[N_c = k]$ by

$$\begin{aligned}\Pr[N_c = k] = {} & \Pr[N_{I,c} = k] \Pr[\text{Case I}] \\ & + \Pr[N_{II,c} = k] \Pr[\text{Case II}].\end{aligned} \tag{4}$$

As shown in Fig. 2, the timing diagram for the MS alters between $t_{m,i}$ and $t_{f,i}$. According to the alternating renewal process [20], we have

$$\Pr[\text{Case I}] = \frac{E[t_{m,i}]}{E[t_{m,i}] + E[t_{f,i}]} = \frac{\eta_f}{\eta_m + \eta_f} \tag{5}$$

and

$$\Pr[\text{Case II}] = \frac{\eta_m}{\eta_m + \eta_f}. \tag{6}$$

We apply (5) and (6) into (4) to yield

$$\Pr[N_c = k] = \frac{\eta_f \Pr[N_{I,c} = k] + \eta_m \Pr[N_{II,c} = k]}{\eta_m + \eta_f}, \tag{7}$$

where $\Pr[N_{I,c} = k]$ and $\Pr[N_{II,c} = k]$ are derived as follows:

The derivation of $\Pr[N_{I,c} = k]$: Consider Case I in Figs. 2a and 2b. Let τ_m be the time period between when the previous call request arrives at L_m and when the MS visits L_f for the first time, i.e., τ_m is the residual life of $t_{m,0}$. Let $r_m(t)$ be the density function of τ_m. From the residual life theorem [20] and the previous work [10], the Laplace transform of $r_m(t)$ is obtained by

$$r_m^*(s) = \left(\frac{\eta_m}{s}\right)[1 - f_m^*(s)]. \tag{8}$$

Let $t_{M,k} = \tau_m + \sum_{i=1}^{k}(t_{f,i} + t_{m,i})$ and

$$t_{F,k} = \tau_m + t_{f,1} + \sum_{i=1}^{k}(t_{m,i} + t_{f,i+1}).$$

By convention, $t_{M,0} = \tau_m$ and $t_{F,0} = \tau_m + t_{f,1}$. Assume that $t_{M,k}$ and $t_{F,k}$ have the density functions $f_{M,j}(t)$ and $f_{F,j}(t)$ with the Laplace transforms $f^*_{M,k}(s)$ and $f^*_{F,k}(s)$, respectively. From the convolution rule of Laplace transform, we have

$$f^*_{M,k}(s) = r^*_m(s)[f^*_m(s)f^*_f(s)]^k \tag{9}$$

and

$$f^*_{F,k}(s) = r^*_m(s)f^*_f(s)[f^*_m(s)f^*_f(s)]^k. \tag{10}$$

In the following, we consider two cases, Case Ia (i.e., the next call request arrives when the MS is in L_f) and Case Ib (i.e., the next call request arrives when the MS is in L_m) to derive $\Pr[N_{I,c} = k]$.

Case Ia: See Fig. 2a. The previous call request arrives during $t_{m,0}$, and the next call request arrives during $t_{f,j+1}$ where $j \geq 0$. In this case, $t_{M,j} < t_c < t_{F,j}$, and k must be an odd number, i.e., $k = 2j+1$. Then, we have

$$\begin{aligned}\Pr[N_{I,c} = 2j+1] &= \Pr[t_{M,j} < t_c < t_{F,j}] \\ &= \Pr[t_c > t_{M,j}] - \Pr[t_c > t_{F,j}] \\ &= \int_{t_c=0}^{\infty}\int_{t_{M,j}=0}^{t_c} f_{M,j}(t_{M,j})\lambda e^{-\lambda t_c} dt_{M,j}dt_c \\ &\quad - \int_{t_c=0}^{\infty}\int_{t_{F,j}=0}^{t_c} f_{F,j}(t_{F,j})\lambda e^{-\lambda t_c} dt_{F,j}dt_c \\ &= f^*_{M,j}(\lambda) - f^*_{F,j}(\lambda).\end{aligned} \tag{11}$$

Apply (8), (9), and (10) into (11) to have

$$\begin{aligned}\Pr[N_{I,c} = 2j+1] = &\left(\frac{\eta_m}{\lambda}\right)[1 - f^*_m(\lambda)][1 - f^*_f(\lambda)] \\ &[f^*_m(\lambda)f^*_f(\lambda)]^j.\end{aligned} \tag{12}$$

Case Ib: See Fig. 2b. The previous call request arrives during $t_{m,0}$, and the next call request arrives during $t_{m,j}$ where $j \geq 1$. In this case, $t_{F,j-1} < t_c < t_{M,j}$, and k must be an even number, i.e., $k = 2j$. Similar to the derivation of (12), we have

$$\begin{aligned}\Pr[N_{I,c} = 2j] &= \Pr[t_{F,j-1} < t_c < t_{M,j}] \\ &= \left(\frac{\eta_m}{\lambda}\right) f^*_f(\lambda)\left[1 - f^*_m(\lambda)\right]^2 \\ &\quad \left[f^*_m(\lambda)f^*_f(\lambda)\right]^{j-1}.\end{aligned} \tag{13}$$

The derivation of $\Pr[N_{II,c} = k]$: Consider Case II in Figs. 2c and 2d. The derivation for $\Pr[N_{II,c} = k]$ is similar to that for $\Pr[N_{I,c} = k]$ except that in Case II, the previous call request arrives during $t_{f,0}$ and then the MS alters between the time period pairs (t_f, t_m). Thus, we have

$$\begin{aligned}\Pr[N_{II,c} = 2j] = &\left(\frac{\eta_f}{\lambda}\right) f^*_m(\lambda)[1 - f^*_f(\lambda)]^2 \\ &[f^*_m(\lambda)f^*_f(\lambda)]^{j-1},\end{aligned} \tag{14}$$

and

$$\begin{aligned}\Pr[N_{II,c} = 2j+1] = &\left(\frac{\eta_f}{\lambda}\right)[1 - f^*_f(\lambda)][1 - f^*_m(\lambda)] \\ &[f^*_m(\lambda)f^*_f(\lambda)]^j.\end{aligned} \tag{15}$$

Apply (12), (13), (14), and (15) into (7), and for $k \geq 1$, we have

$$\Pr[N_c = k] = \begin{cases} \left[\dfrac{2\eta_m\eta_f}{\lambda(\eta_m+\eta_f)}\right]\left[1 - f^*_m(\lambda)\right]\left[1 - f^*_f(\lambda)\right] \\ \left[f^*_m(\lambda)f^*_f(\lambda)\right]^{\frac{k-1}{2}}, \\ \text{odd } k; \\ \left[\dfrac{\eta_m\eta_f}{\lambda(\eta_m+\eta_f)}\right]\Big\{f^*_f(\lambda)\left[1 - f^*_m(\lambda)\right]^2 \\ +f^*_m(\lambda)\left[1 - f^*_f(\lambda)\right]^2\Big\}\left[f^*_m(\lambda)f^*_f(\lambda)\right]^{\frac{k}{2}-1}, \\ \text{even } k. \end{cases} \tag{16}$$

Then, from (16), we have

$$\begin{aligned}E[N_r(0)] = E[N_c] &= \sum_{k=1}^{\infty} k\Pr[N_c = k] \\ &= \frac{2\eta_m\eta_f}{\lambda(\eta_m+\eta_f)}.\end{aligned} \tag{17}$$

3.1.2 Derivation of $E[N_r(t_d)]$

Consider the time period $t_{f,i}$ for $i \geq 1$. Let α be the probability that the MS executes the registration during $t_{f,i}$ (i.e., $t_d < t_{f,i}$) under the condition that the next call request does not arrive during $t_{f,i}$, that is,

$$\begin{aligned}\alpha = \Pr[t_d < t_{f,i} | t_{c,i} > t_{f,i}] &= \frac{\Pr[t_d < t_{f,i} < t_{c,i}]}{\Pr[t_{c,i} > t_{f,i}]} \\ &= \frac{\int_{t_d=0}^{\infty}\int_{t_{f,i}=t_d}^{\infty}\int_{t_{c,i}=t_{f,i}}^{\infty} \lambda e^{-\lambda t_{c,i}} f_f(t_{f,i})\theta e^{-\theta t_d} dt_{c,i}dt_{f,i}dt_d}{\int_{t_{f,i}=0}^{\infty}\int_{t_{c,i}=t_{f,i}}^{\infty} \lambda e^{-\lambda t_{c,i}} f_f(t_{f,i})dt_{c,i}dt_{f,i}} \\ &= \frac{f^*_f(\lambda) - f^*_f(\lambda+\theta)}{f^*_f(\lambda)}.\end{aligned} \tag{18}$$

Let β be the probability that the MS executes the registration before the next call request arrives under the condition that the next call request arrives during $t_{f,i}$, and we have

$$\begin{aligned}\beta = \Pr[t_d < t_{c,i} | t_{c,i} < t_{f,i}] &= \frac{\Pr[t_d < t_{c,i} < t_{f,i}]}{\Pr[t_{c,i} < t_{f,i}]} \\ &= \frac{\int_{t_d=0}^{\infty}\int_{t_{c,i}=t_d}^{\infty}\int_{t_{f,i}=t_{c,i}}^{\infty} f_f(t_{f,i})\lambda e^{-\lambda t_{c,i}}\theta e^{-\theta t_d} dt_{f,i}dt_{c,i}dt_d}{1 - f^*_f(\lambda)} \\ &= \frac{\left(\frac{\theta}{\lambda+\theta}\right) + \left(\frac{\lambda}{\lambda+\theta}\right)f^*_f(\lambda+\theta) - f^*_f(\lambda)}{1 - f^*_f(\lambda)}.\end{aligned} \tag{19}$$

Let $N_{I,r}(t_d)$ and $N_{II,r}(t_d)$ denote $N_r(t_d)$ conditioning on Cases I and II, respectively. Then, $\Pr[N_r(t_d) = k]$ is expressed as

$$\begin{aligned}\Pr[N_r(t_d) = k] &= \Pr[N_{I,r}(t_d) = k]\Pr[\text{CaseI}] \\ &\quad + \Pr[N_{II,r}(t_d) = k]\Pr[\text{CaseII}] \\ &= \frac{\eta_f \Pr[N_{I,r}(t_d) = k] + \eta_m \Pr[N_{II,r}(t_d) = k]}{\eta_m + \eta_f}.\end{aligned} \tag{20}$$

$\Pr[N_{I,r}(t_d) = k]$ and $\Pr[N_{II,r}(t_d) = k]$ are derived as follows:

The derivation of $\Pr[N_{I,r}(t_d)=k]$: Similar to the derivation for $\Pr[N_{I,c}=k]$, we consider Case Ia (i.e., the next call request arrives when the MS is in L_f) and Case Ib (i.e., the next call request arrives when the MS is in L_m) to derive $\Pr[N_{I,r}(t_d)=k]$.

Case Ia: The previous call request arrives during $t_{m,0}$, and the next call request arrives during $t_{f,j+1}$. See Fig. 2a. In this case, the MS has $N_{I,c}=2j+1$ crossings. The MS visits the macrocell for j times and the overlapped femtocell for $j+1$ times (i.e., $t_{f,1}$, $t_{m,1}$, $t_{f,2}$, $t_{m,2},\ldots,t_{f,j}$, $t_{m,j}$, $t_{f,j+1}$). For $1\le i\le j$, if the MS executes a registration during $t_{f,i}$ (i.e., $t_d<t_{f,i}$), then after $t_{f,i}$ (i.e., at the beginning of $t_{m,i}$), a registration will be executed. Otherwise (i.e., the MS does not execute a registration during $t_{f,i}$; i.e., $t_d>t_{f,i}$), after $t_{f,i}$ (i.e., at the beginning of $t_{m,i}$), no registration will be executed.

Among $N_{I,c}=2j+1$ crossings, for $N_{I,r}(t_d)=k$, if k is an odd number, a registration is executed during $t_{f,j+1}$. Otherwise (i.e., k is an even number), a registration is not executed during $t_{f,j+1}$. Then, we have

$$\Pr[N_{I,r}(t_d)=k|N_{I,c}=2j+1]=\begin{cases}\binom{j}{\frac{k-1}{2}}\alpha^{\frac{k-1}{2}}(1-\alpha)^{j-\frac{k-1}{2}}\beta, & \text{odd } k;\\ \binom{j}{\frac{k}{2}}\alpha^{\frac{k}{2}}(1-\alpha)^{j-\frac{k}{2}}(1-\beta), & \text{even } k.\end{cases}\tag{21}$$

Case Ib: The previous call request arrives during $t_{m,0}$, and the next call request arrives during $t_{m,j}$ where $j\ge 1$. See Fig. 2b. In this case, the MS has $N_{I,c}=2j$ crossings. The MS visits the overlapped femtocell for j times and the macrocell for j times (i.e., $t_{f,1}$, $t_{m,1}$, $t_{f,2}$, $t_{m,2},\ldots,t_{f,j}$, $t_{m,j}$). Among $N_{I,c}=2j$ crossings, for $N_{I,r}(t_d)=k$, k must be an even number. Then, similar to the derivation of (21), we have

$$\Pr[N_{I,r}(t_d)=k|N_{I,c}=2j]=\binom{j}{\frac{k}{2}}\alpha^{\frac{k}{2}}(1-\alpha)^{j-\frac{k}{2}}.\tag{22}$$

From (21) and (22), we have

$$\Pr[N_{I,r}(t_d)=k]=\begin{cases}\sum_{j=\frac{k-1}{2}}^{\infty}\Pr[N_{I,r}(t_d)=k|N_{I,c}=2j+1]\Pr[N_{I,c}=2j+1]; & \text{odd } k,\\ \sum_{j=\frac{k}{2}}^{\infty}\Pr[N_{I,r}(t_d)=k|N_{I,c}=2j+1]\Pr[N_{I,c}=2j+1] & \\ \quad+\sum_{j=\frac{k}{2}}^{\infty}\Pr[N_{I,r}(t_d)=k|N_{I,c}=2j]\Pr[N_{I,c}=2j], & \text{even } k.\end{cases}\tag{23}$$

Applying (12), (13), (21), and (22) into (23), we obtain

$$\Pr[N_{I,r}(t_d)=k]=\begin{cases}\dfrac{\eta_m\beta\left[1-f_m^*(\lambda)\right]\left[1-f_f^*(\lambda)\right]\left[\alpha f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k-1}{2}}}{\lambda\left[1-(1-\alpha)f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k+1}{2}}}, & \text{odd } k;\\ \dfrac{\eta_m\left[1-f_m^*(\lambda)\right]\left[1-(1-\beta)f_m^*(\lambda)f_f^*(\lambda)-\beta f_m^*(\lambda)\right]}{\lambda f_m^*(\lambda)\left[1-(1-\alpha)f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k}{2}+1}}\left[\alpha f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k}{2}}, & \text{even } k.\end{cases}\tag{24}$$

The derivation of $\Pr[N_{II,r}(t_d)=k]$: We consider Case IIa (i.e., the next call request arrives when the MS is in L_f) and Case IIb (i.e., the next call request arrives when the MS is in L_m) to derive $\Pr[N_{II,r}(t_d)=k]$.

Case IIa: The previous call request arrives during $t_{f,0}$, and the next call request arrives during $t_{f,j}$ where $j\ge 1$. See Fig. 2c. In this case, the MS has $N_{II,c}=2j$ crossings. The MS visits the overlapped femtocell for j times and the macrocell for j times (i.e., $t_{m,1}$, $t_{f,1}$, $t_{m,2}$, $t_{f,2},\ldots,t_{m,j}$, $t_{f,j}$). At the beginning of $t_{m,1}$ (after $t_{f,0}$), the MS executes a registration if the MS has executed a registration during $t_{f,0}$. For $1\le i<j$, if the MS executes a registration during $t_{f,i}$ (i.e., $t_d<t_{f,i}$), then after $t_{f,i}$ (i.e., at the beginning of $t_{m,i+1}$), a registration is executed. Otherwise (i.e., the MS does not execute a registration during $t_{f,i}$; i.e., $t_d>t_{f,i}$), then after $t_{f,i}$, at the beginning of $t_{m,i+1}$, no registration is executed.

Among $N_{II,c}=2j$ crossings, for $N_{II,r}(t_d)=k$, if k is an odd number, a registration is executed during $t_{f,0}$ and no registration is executed during $t_{f,j}$ or no registration is executed during $t_{f,0}$ and a registration is executed during $t_{f,j}$. Otherwise (i.e., k is an even number), the registrations are executed during $t_{f,0}$ and $t_{f,j}$ or no registration is executed neither during $t_{f,0}$ nor during $t_{f,j}$. Then, we have

$$\Pr[N_{II,r}(t_d)=k|N_{II,c}=2j]=\begin{cases}\binom{j-1}{\frac{k-1}{2}}2\alpha^{\frac{k-1}{2}}(1-\alpha)^{j-\frac{k-1}{2}-1}\beta(1-\beta), & \text{odd } k;\\ \binom{j-1}{\frac{k}{2}-1}\alpha^{\frac{k}{2}-1}(1-\alpha)^{j-\frac{k}{2}}\beta^2 & \\ \quad+\binom{j-1}{\frac{k}{2}}\alpha^{\frac{k}{2}}(1-\alpha)^{j-\frac{k}{2}-1}(1-\beta)^2, & \text{even } k.\end{cases}\tag{25}$$

Case IIb: The previous call request arrives during $t_{f,0}$, and the next call request arrives during $t_{m,j+1}$ where $j\ge 0$. See Fig. 2d. In this case, the MS has $N_{II,c}=2j+1$ crossings where the MS visits the overlapped femtocell for j times and the macrocell for $j+1$ times (i.e., $t_{m,1}$, $t_{f,1}$, $t_{m,2}$, $t_{f,2},\ldots,t_{m,j}$, $t_{f,j}$, $t_{m,j+1}$). Among $N_{II,c}=2j+1$ crossings, for $N_{II,r}(t_d)=k$, if k is an odd number, a registration is executed during $t_{f,0}$. Otherwise (i.e., k is an even number), no registration is executed during $t_{f,0}$. Then, we have

$$\Pr[N_{II,r}(t_d)=k|N_{II,c}=2j+1]=\begin{cases}\binom{j}{\frac{k-1}{2}}\alpha^{\frac{k-1}{2}}(1-\alpha)^{j-\frac{k-1}{2}}\beta, & \text{odd } k;\\ \binom{j}{\frac{k}{2}}\alpha^{\frac{k}{2}}(1-\alpha)^{j-\frac{k}{2}}(1-\beta), & \text{even } k.\end{cases}\tag{26}$$

From (25) and (26), we rewrite

$$\Pr[N_{II,r}(t_d)=k] = \begin{cases} \sum_{j=\frac{k+1}{2}}^{\infty} \Pr[N_{II,r}(t_d)=k|N_{II,c}=2j]\Pr[N_{II,c}=2j] \\ +\sum_{j=\frac{k-1}{2}}^{\infty} \Pr[N_{II,r}(t_d)=k|N_{II,c}=2j+1] \\ \Pr[N_{II,c}=2j+1], \\ \text{odd } k; \\ \sum_{j=\frac{k}{2}}^{\infty} \Pr[N_{II,r}(t_d)=k|N_{II,c}=2j]\Pr[N_{II,c}=2j] \\ +\sum_{j=\frac{k}{2}}^{\infty} \Pr[N_{II,r}(t_d)=k|N_{II,c}=2j+1] \\ \Pr[N_{II,c}=2j+1], \\ \text{even } k. \end{cases} \quad (27)$$

Apply (14), (15), (25), and (26) into (27) to yield

$$\Pr[N_{II,r}(t_d)=k] = \begin{cases} \left(\frac{\eta_f\beta\left[1-f_f^*(\lambda)\right]\left[\alpha f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k-1}{2}}}{\lambda\left[1-(1-\alpha)f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k+1}{2}}}\right) \\ \left[1+(1-2\beta)f_m^*(\lambda)-2(1-\beta)f_m^*(\lambda)f_f^*(\lambda)\right], \\ \text{odd } k; \\ \left(\frac{\eta_f\left[1-f_f^*(\lambda)\right]\left[\alpha f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k}{2}}}{\lambda\alpha f_f^*(\lambda)\left[1-(1-\alpha)f_m^*(\lambda)f_f^*(\lambda)\right]^{\frac{k}{2}+1}}\right) \\ \left\{\beta\left[1-f_f^*(\lambda)\right]\left[(\alpha+\beta-2\alpha\beta)f_m^*(\lambda)f_f^*(\lambda)-\beta\right]\right. \\ \left.-\alpha(1-\beta)f_f^*(\lambda)\left[1-f_m^*(\lambda)f_f^*(\lambda)\right]\right\}, \\ \text{even } k. \end{cases} \quad (28)$$

The probability $\Pr[N_r(t_d)=k]$ can be obtained by applying (24) and (28) into (20). Then, we have

$$\begin{aligned} E[N_r(t_d)] &= \sum_{k=1}^{\infty} k\Pr[N_r(t_d)=k] \\ &= \frac{2\eta_m\eta_f\left[\beta+(\alpha-\beta)f_f^*(\lambda)\right]}{\lambda(\eta_m+\eta_f)(\lambda+\theta)}. \end{aligned} \quad (29)$$

Applying (18) and (19) into (29), we have

$$E[N_r(t_d)] = \frac{2\eta_m\eta_f\theta[1-f_f^*(\lambda+\theta)]}{\lambda(\eta_m+\eta_f)(\lambda+\theta)}. \quad (30)$$

3.2 Derivation of Potential Offload Traffic Ratio $\rho(t_d)$

In the 3GPP algorithms, the LAI stored in the location database is L_f as long as the MS is in the overlapped femtocell, and the same applies for L_m as long as the MS is in the macrocell. When a call request arrives during t_f and t_m, by querying the LAI information in the location database, the call request is processed by the overlapped femtocell and macrocell, respectively. Since the MS alters between t_m and t_f, from the alternating renewal process [20], the probability $p(0)$ that a call arrives during t_f is

$$p(0) = \frac{E[t_f]}{E[t_m]+E[t_f]} = \frac{\eta_m}{\eta_m+\eta_f}. \quad (31)$$

As shown in Fig. 2, the MS shifts between staying in the macrocell and the overlapped femtocell for t_m and t_f. Suppose that the MS moves from the macrocell to the overlapped femtocell at t, and leaves the femtocell at $t+t_f$. In the DR algorithm, at t, the registration is suspended until the delay timer with length t_d expires. During the period $[t, t+\min(t_d,t_f)]$, the LAI stored in the database is L_m, and if a call request arrives, it is processed by the macrocell. During the period $[t+\min(t_d,t_f), t+t_f]$, the LAI stored in the database is L_f, and if a call request arrives, it is processed by the femtocell. Let $t'_m = t_m + \min(t_d,t_f)$, and $t'_f = t_f - \min(t_d,t_f)$, and we have $E[t'_m] = E[t_m] + E[\min(t_d,t_f)]$ and $E[t'_f] = E[t_f] - E[\min(t_d,t_f)]$. In other words, the LAI is alternatively changed between t'_m and t'_f. During t'_f, all call request arrivals are processed by the femtocell. Therefore, $p(t_d)$ is the probability that a call arrives during t'_f. According to the alternating renewal process, we have

$$p(t_d) = \frac{E[t'_f]}{E[t'_m]+E[t'_f]} = \frac{E[t_f]-E[\min(t_d,t_f)]}{E[t_m]+E[t_f]}, \quad (32)$$

where

$$\begin{aligned} E[\min(t_d,t_f)] &= \int_{t_f=0}^{\infty}\int_{t_d=0}^{t_f} t_d f_f(t_f)\theta e^{-\theta t_d} dt_d dt_f \\ &\quad + \int_{t_f=0}^{\infty}\int_{t_d=t_f}^{\infty} t_f f_f(t_f)\theta e^{-\theta t_d} dt_d dt_f \\ &= \frac{1-f_f^*(\theta)}{\theta}. \end{aligned}$$

Then, (32) is rewritten as

$$p(t_d) = \frac{\eta_m[\theta-\eta_f+\eta_f f_f^*(\theta)]}{\theta(\eta_m+\eta_f)}. \quad (33)$$

Then, we apply (31) and (33) into (2) to yield

$$\rho(t_d) = \frac{\theta-\eta_f+\eta_f f_f^*(\theta)}{\theta}. \quad (34)$$

4 Performance Evaluation

In this section, we first analyze the accuracy of the analytical models proposed in this paper through the simulation experiments, and then study the $r(t_d)$ and $\rho(t_d)$ performances of the DR algorithm.

As shown in (3) and (34), $r(t_d)$ and $\rho(t_d)$ are obtained if the closed-form expressions of $f_m^*(s)$ and $f_f^*(s)$ exist. In this study, we apply the Gamma distribution for the residence times t_m and t_f. The Gamma distribution is selected because it can approximate many types of distributions, it has a closed-form expression for its Laplace transform, and it has been widely used in many previous works (e.g., [9], [10], [11]) to reflect the MS mobility.

Suppose that the Gamma density functions $f_m(t)$ and $f_f(t)$ are with the means $1/\eta_m$ and $1/\eta_f$, the shape parameters γ_m and γ_f, the variances $v_m = 1/(\gamma_m\eta_m^2)$ and $v_f = 1/(\gamma_f\eta_f^2)$, and the Laplace transforms

$$f_m^*(s) = \left(\frac{\gamma_m\eta_m}{\gamma_m\eta_m+s}\right)^{\gamma_m} \text{and } f_f^*(s) = \left(\frac{\gamma_f\eta_f}{\gamma_f\eta_f+s}\right)^{\gamma_f}.$$

TABLE 2
Validation of the Simulation and Analysis Results

$E[t_c] = 1/\lambda = 50$ minutes, $\eta_f = 200\lambda$, $v_f = 1000/\eta_f^2$, $\eta_m = 50\lambda$, $v_m = 10/\eta_m^2$						
$E[t_d]$ (unit: $1/\lambda$)		1/10	1/20	1/30	1/40	1/50
α	Analytic	0.0022311	0.00286756	0.00325279	0.00352987	0.00374638
	Simulation	0.00223434	0.00287484	0.0032586	0.00353166	0.00375318
	Error (%)	0.145318	0.253703	0.178593	0.0506395	0.18136
β	Analytic	0.795992	0.876239	0.909232	0.927603	0.939431
	Simulation	0.796412	0.875418	0.909573	0.927354	0.939321
	Error (%)	0.0527235	0.0936923	0.0374611	0.0268176	0.0117352
$r(t_d)$	Analytic	0.00365206	0.00443103	0.00487464	0.00518411	0.0054214
	Simulation	0.0036542	0.00441135	0.00488304	0.00520538	0.00543302
	Error (%)	0.0587348	0.444215	0.172404	0.410383	0.214335
$\rho(t_d)$	Analytic	0.921518	0.953955	0.966635	0.973554	0.977959
	Simulation	0.921672	0.95301	0.966418	0.97411	0.977663
	Error (%)	0.0167304	0.0990423	0.0225196	0.0571416	0.0303185

Then, we have

$$r(t_d) = \frac{\theta - \theta\left(\frac{\gamma_f \eta_f}{\gamma_f \eta_f + \lambda + \theta}\right)^{\gamma_f}}{\lambda + \theta} \tag{35}$$

and

$$\rho(t_d) = \frac{\theta - \eta_f + \eta_f\left(\frac{\gamma_f \eta_f}{\gamma_f \eta_f + \theta}\right)^{\gamma_f}}{\theta}. \tag{36}$$

We develop the simulation model for the DR algorithm based on the discrete event-driven approach, one widely used in MCN studies (e.g., [17], [21]). The simulation model is similar to that in [17], so the details are not presented in this paper. In our study, the input parameters η_m, η_f, and θ are normalized by λ. For example, if we set the expected intercall arrival time $E[t_c] = 1/\lambda = 50$ minutes [14], $\theta = 10\lambda$ means that the expected delay period $E[t_d] = 1/\theta = 5$ minutes. In other words, our study results are dependent on the call arrival rate λ. For the DR algorithm to be effective, λ must be estimated very accurately. Accurate λ measurements are available in existing CN nodes (i.e., mobile switching centers or GPRS support nodes).

The analytical and simulation results are validated against each other. As shown in Table 2, the errors between the analytical and simulation results fall within 1 percent, demonstrating consistent findings from both our analytical models and simulation experiments.

In the following, we investigate the effects of input parameters on $r(t_d)$ and $\rho(t_d)$ for the DR algorithm. Section 4.1 studies the effects of MS mobility. Section 4.2 studies the effects of t_d with fixed and exponential setups.

4.1 Effects of MS Mobility

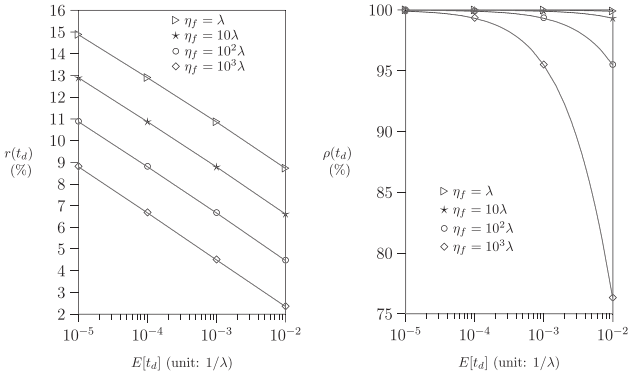
This section studies the effects of MS mobility behaviors on the $r(t_d)$ and $\rho(t_d)$ performances. From (3) and (34), it is clear that $r(t_d)$ and $\rho(t_d)$ are independent from the distribution for t_m, indicating that the MS mobility behavior in nonoverlapped areas of the macrocell does not affect the $r(t_d)$ and $\rho(t_d)$ performances. Therefore, in the following, we only study the effects of MS mobility behavior in the overlapped femtocell.

Effects of mean of femtocell residence time $1/\eta_f$: In Fig. 3, we study the effects of the femtocell residence time, where $E[t_d]$ is set from $10^{-5}/\lambda$ (i.e., 0.03 seconds) to $10^{-2}/\lambda$ (i.e., 30 seconds), $\eta_m = 25\lambda$ (i.e., $E[t_m] = 2$ minutes), and $v_m = 1/\eta_m^2$. We set $v_f = 100/\eta_f^2$ to simulate the real MS mobility behavior that the MS either stays in the femtocell for a long period or just passes by the femtocell.

As shown in Fig. 3a, $r(t_d)$ decreases as η_f increases. A larger η_f implies that the MS stays in the overlapped femtocell for a shorter period. This is when a transient phenomenon might likely occur, during which more registration traffic can be avoided in the DR algorithm. Fig. 3a also indicates that the longer we set the delay timer, the more registration avoided, i.e., $r(t_d)$ decreases as $E[t_d]$ increases. In Fig. 3a, we observe that the DR algorithm reduces at least 85 percent of registration signaling overhead.

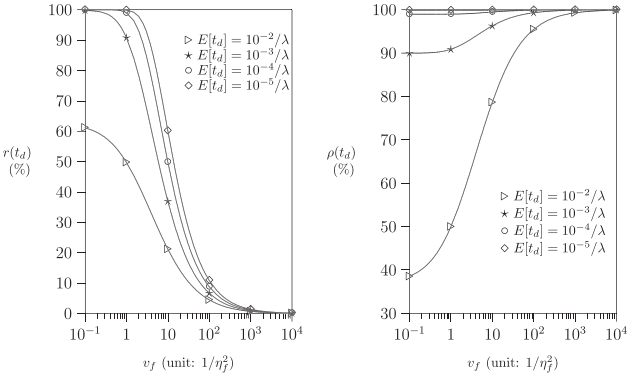
On the other hand, in Fig. 3b, $\rho(t_d)$ decreases as η_f increases, i.e., with the delay timer, the transient phenomenon reduces traffic offloading capability of the femtocell. We observe that the DR algorithm causes at most 24 percent of the degradation of the traffic offloading capability (i.e., $\rho(t_d) \approx 76\%$ when $\eta_f = 10^3\lambda$ and $E[t_d] = 10^{-2}/\lambda$). To summarize, the DR algorithm significantly reduces the registration signaling overhead. Meanwhile, the DR algorithm sustains good performance for the traffic offloading capability of the femtocell.

In the following, we discuss how to set up the delay timer adapting to different MS mobility behaviors to achieve better $r(t_d)$ performance while minimizing the loss in traffic offload capability. Observe the "$\diamond$" curves in Figs. 3a and 3b, where $\eta_f = 10^3\lambda$ (i.e., $E[t_f] = 3$ seconds). In this mobility scenario, the MS stays in the overlapped femtocell for transient periods. Fig. 3a indicates how the $r(t_d)$ values drop linearly (i.e., better $r(t_d)$ performance is obtained; from 9 to



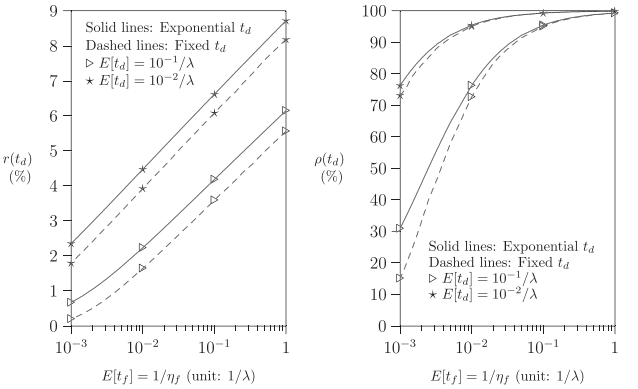
(a) Performance of Signaling Overhead Ratio (b) Performance of Potential Offload Traffic Ratio

Fig. 3. The effects of η_f on $r(t_d)$ and $\rho(t_d)$ ($\eta_m = 25\lambda$, $v_m = 1/\eta_m^2$, $v_f = 100/\eta_f^2$).



(a) Performance of Signaling Overhead Ratio (b) Performance of Potential Offload Traffic Ratio

Fig. 4. The effects of v_f on $r(t_d)$ and $\rho(t_d)$ ($\eta_m = 25\lambda$, $\eta_f = 100\lambda$, $v_m = 1/\eta_m^2$).



(a) Performance of Signaling Overhead Ratio (b) Performance of Potential Offload Traffic Ratio

Fig. 5. The effects of fixed and exponential t_d on $r(t_d)$ and $\rho(t_d)$ ($\eta_m = 25\lambda$, $v_m = 1/\eta_m^2$, $v_f = 100/\eta_f^2$).

2.3 percent) as $E[t_d]$ increases from $10^{-5}/\lambda$ to $10^{-2}/\lambda$. However, Fig. 3b indicates that when $E[t_d] \leq 10^{-3}/\lambda$, the $\rho(t_d)$ performance decreases slightly (from 100 to 95 percent) as $E[t_d]$ increases, but when $E[t_d] > 10^{-3}/\lambda$, the $\rho(t_d)$ performance drops very quickly (from 95 to 76 percent). To summarize, we prefer to set $E[t_d] \leq 10^{-3}/\lambda$ (i.e., 3 seconds).

For other mobility scenarios, $\eta_f = \lambda$, $\eta_f = 10\lambda$, and $\eta_f = 100\lambda$ (see "$\triangleright$", "$\star$", and "$\circ$" curves), we prefer to set $E[t_d] = 10^{-2}/\lambda$ (i.e., 30 seconds) because when $E[t_d] = 10^{-2}/\lambda$, we achieve the best $r(t_d)$ performance with loss of traffic offloading capability no larger than 5 percent (i.e., $\rho(t_d) = 95\%$).

Effects of variance of femtocell residence time v_f: In Fig. 4, we study the effects of the variance v_f of the femtocell residence time, where $E[t_d]$ is set from $10^{-5}/\lambda$ (i.e., 0.03 seconds) to $10^{-2}/\lambda$ (i.e., 30 seconds), $\eta_m = 25\lambda$, $\eta_f = 100\lambda$, and $v_m = 1/\eta_m^2$. As v_f increases, it is more likely to observe an MS with short and long residence time in an overlapped femtocell, so the MS mobility behavior in the femtocell is more "dynamic."

For short t_f periods, it is more likely that an MS moves out of an overlapped femtocell before the delay timer t_d expires. The MS has less chance to execute the registration in the overlapped femtocell, reducing more signaling overhead caused by registration. Therefore, we observe $r(t_d)$ decreases as v_f increases in Fig. 4a.

On the other hand, for longer t_f periods, the MS is more likely to have call requests through the long-residence femtocell. More requests are potentially processed by the femtocell. Therefore, larger $\rho(t_d)$ is observed as v_f increases in Fig. 4b.

To summarize, when the MS mobility is more dynamic, the DR algorithm can work more effectively (i.e., both $r(t_d)$ and $\rho(t_d)$ have better performance when v_f is larger).

In addition to the Gamma distribution, in this study, we also considered the Weibull distribution for MS mobility behaviors, which has also been widely used to approximate real MS mobility patterns in many MCN studies (e.g., [22], [23]). Note that the Weibull distribution does not have a closed-form expression for its Laplace transform [24]. Therefore, the closed-form expressions of $r(t_d)$ and $\rho(t_d)$ for the Weibull distribution do not exist in our analytical models. Instead, we run simulation experiments to study effects of MS mobility on $r(t_d)$ and $\rho(t_d)$ for the Weibull distributed residence times. We observe similar performance trends for both the Weibull and Gamma distributions, and thus, we do not include the performance evaluation for the Weibull distributed residence times.

4.2 Effects of Fixed and Exponential t_d

In Fig. 5, based on the simulation experiments, we study $r(t_d)$ and $\rho(t_d)$ against $E[t_f]$ for fixed and exponential t_d, where $\eta_m = 25\lambda$, $v_m = 1/\eta_m^2$, $v_f = 100/\eta_f^2$, and t_f is Gamma distributed. We observe that the performance trends of $r(t_d)$ and $\rho(t_d)$ for fixed t_d are similar to those for exponential t_d.

In Fig. 5a, when $E[t_d] = 10^{-1}/\lambda$ and $E[t_d] = 10^{-2}/\lambda$, $r(t_d)$ for fixed t_d is about 0.5 percent lower than that for exponential t_d. As $E[t_f]$ increases from $10^{-3}/\lambda$ to $1/\lambda$, this difference remains the same.

On the other hand, in Fig. 5b, as $E[t_f]$ increases from $10^{-3}/\lambda$ to $1/\lambda$, the difference between $\rho(t_d)$ of fixed t_d and that of exponential t_d diminishes from 2.5 to 0 percent for $E[t_d] = 10^{-2}/\lambda$, and from 15 to 0 percent for $E[t_d] = 10^{-1}/\lambda$. This difference is larger when $E[t_d]$ is longer.

To summarize, the performance trends for $E[t_d] = 10^{-1}/\lambda$ and $E[t_d] = 10^{-2}/\lambda$ are very similar. To achieve better $r(t_d)$ and $\rho(t_d)$, we suggest to use exponential t_d setup when $E[t_f] < 10^{-1}/\lambda$ and use fixed t_d setup when $E[t_f] \geq 10^{-1}/\lambda$.

5 Conclusions

In this paper, we proposed a DR algorithm to reduce signaling overhead caused by frequent registrations, while noticing the slight decrease in traffic offloading capability of femtocells. To avoid registrations during the transient period, in the DR algorithm, we introduce a delay timer to postpone the registration until the timer expires. We conducted analytical models and simulation experiments to study the performance of the DR algorithm in terms of the signaling overhead ratio $r(t_d)$ and potential offload traffic ratio $\rho(t_d)$. The analytical model is generally enough to accommodate various MS mobility behaviors. Our performance study can provide network operators with guidelines to configure the delay timer. Our study indicates that the DR algorithm can significantly reduce the signaling overhead with slight loss of traffic offloading capability of the femtocells. Moreover, when the MS mobility is more dynamic, the DR algorithm can work more effectively, i.e., lower signaling overhead ratio and higher potential offload traffic ratio.

ACKNOWLEDGMENTS

The authors thank the four anonymous reviewers for their valuable comments, which significantly improved the quality of this paper. Phone Lin's work was supported in part by NSC 100-2221-E-002-176, NSC 100-2221-E-002-184, NSC 100-2219-E-002-015, NSC 101-2219-E-002-002, NSC 100-2219-E-007-010, the National Science Council, National Taiwan University, Intel Corporation under Grants NSC 100-2911-I-002-001 and 101R7501, Chunghwa Telecom, Arcadyan Technology Corporation, and ICL/ITRI. Yi-Bing Lin's work was supported in part by NSC 100-2221-E-009-070, Chunghwa Telecom, IBM, Arcadyan Technology Corporation, ICL/ITRI, Nokia Siemens Networks, and the MoE ATU plan.

REFERENCES

[1] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Architecture for 3G Home Node B (HNB); Stage 2 (Release 9)," 3G TS 25.467, Sept. 2009.
[2] Femto Forum, http://www.femtoforum.org, 2013.
[3] V. Chandrasekhar, J.G. Andrews, and A. Gatherer, "Femtocell Networks: A Survey," *IEEE Comm. Magazine,* vol. 46, no. 9, pp. 59-67, Sept. 2008.
[4] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Network Architecture (Release 9)," 3G TS 23.002, Nov. 2009.
[5] Broadband Forum, http://www.broadband-forum.org, 2013.
[6] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Location Management Procedures (Release 8)," 3G TS 23.012, Sept. 2009.
[7] H. Claussen, I. Ashraf, and L.T.W. Ho, "Dynamic Idle Mode Procedures for Femtocells," *Bell Labs Technical J.,* vol. 15, pp. 95-116, Sept. 2010.
[8] Y. Lei and Y. Zhang, "Efficient Location Management Mechanism for Overlay LTE Macro and Femto Cells," *Proc. IEEE Int'l Conf. Comm. Technology and Applications,* pp. 420-424, Dec. 2009.
[9] I.F. Akyildiz, J.S.M. Ho, and Y.-B. Lin, "Movement-Based Location Update and Selective Paging for PCS Networks," *IEEE/ACM Trans. Networking,* vol. 4, no. 4, pp. 629-638, Aug. 1996.
[10] Y.-B. Lin, "Reducing Location Update Cost in a PCS Network," *IEEE/ACM Trans. Networking,* vol. 5, no. 1, pp. 25-33, Feb. 1997.
[11] P. Lin and Y.-B. Lin, "Implementation and Performance Evaluation for Mobility Management of a Wireless PBX Network," *IEEE J. Selected Areas in Comm.,* vol. 19, no. 6, pp. 1138-1146, June 2001.
[12] Y. Xiao and M. Guizani, "Optimal Paging Load Balance with Total Delay Constraint in Macrocell-Microcell Hierarchical Cellular Networks," *IEEE Trans. Wireless Comm.,* vol. 5, no. 8, pp. 2202-2209, Aug. 2006.
[13] X. Wu, B. Mukherjee, and B. Bhargava, "A Crossing-Tier Location Update/Paging Scheme in Hierarchical Cellular Networks," *IEEE Trans. Wireless Comm.,* vol. 5, no. 4, pp. 839-848, Apr. 2006.
[14] K.I. Park and Y.-B. Lin, "Reducing Registration Traffic for Multi-Tier Personal Communications Services," *IEEE Trans. Vehicular Technology,* vol. 46, no. 3, pp. 597-602, Aug. 1997.
[15] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures.* John Wiley and Sons, 2001.
[16] Y. Fang, "Movement-Based Mobility Management and Trade Off Analysis for Wireless Mobile Networks," *IEEE Trans. Computers,* vol. 52, no. 6, pp. 791-803, June 2003.
[17] P. Lin, Y.-B. Lin, and J.-Y. Jeng, "Improving GSM Call Completion by Call Re-Establishment," *IEEE J. Selected Areas in Comm.,* vol. 17, no. 7, pp. 1305-1317, July 1999.
[18] Y.-B. Lin and P. Lin, "Performance Modeling of Location Tracking Systems," *ACM Mobile Computing and Comm. Rev.,* vol. 2, pp. 24-27, July/Aug. 1998.
[19] F. Baccelli, S. Machiraju, D. Veitch, and J.C. Bolot, "The Role of PASTA in Network Management," *Proc. ACM SIGCOMM,* pp. 231-242, Sept. 2006.
[20] R. Nelson, *Probability, Stochastic Processes, and Queueing Theory.* Springer-Verlag, 1995.
[21] Y.-B. Lin and S.-R. Yang, "A Mobility Management Strategy for GPRS," *IEEE Trans. Wireless Comm.,* vol. 2, pp. 1178-1188, Nov. 2003.
[22] H.-N. Hung, P.-C. Lee, and Y.-B. Lin, "Random Number Generation for Excess Life of Mobile User Residence Time," *IEEE Trans. Vehicular Technology,* vol. 55, no. 3, pp. 1045-1050, May 2006.
[23] F. Khan and D. Zeghlache, "Effect of Cell Residence Time Distribution on the Performance of Cellular Mobile Networks," *Proc. IEEE Vehicular Technology Conf.,* pp. 949-953, May 1997.
[24] M. Fischer, D. Gross, D. Masi, and J. Shortle, "Analyzing the Waiting Time Process in Internet Queueing Systems with the Transform Approximation Method," *Telecomm. Rev.,* vol. 12, pp. 21-32, 2001.

Huai-Lei Fu received the bachelor and master degrees in computer science and information engineering (CSIE) from Tatung University and Yuan Ze University (YZU) in 2005 and 2007, respectively. He is currently working toward the PhD degree in the Department of CSIE, National Taiwan University (NTU). His research interests include machine-to-machine communications network, multicast and broadcast service, mobility management, performance modeling, and wireless sensor network. He has received the YZU Academic Silver Medal Award in 2006, the YZU School Work Silver Medal Award in 2006, the Y. Z. Hsu Scholarship Award in 2006 and 2007, and the NTU Outstanding Student Award (Academic Category) in 2009. He is a student member of the IEEE.

Phone Lin received the BSCSIE and PhD degrees from National Chiao Tung University in 1996 and 2001, respectively. He is a professor at National Taiwan University (NTU) in the Department of Computer Science and Information Engineering, Graduate Institute of Networking and Multimedia, Telecommunications Research Center, Optoelectronic Biomedicine Center, and Intel-NTU Connected Context Computing Center. He has served on the editorial boards of many journals, including the *IEEE Transactions on Vehicular Technology*, *IEEE Wireless Communications*, and *ACM/Springer Wireless Networks*, has been a guest editor for *IEEE Wireless Communications* and *ACM/Springer Mobile Networks and Applications*, and has been involved with several prestigious conferences, such as serving as the technical program chair for WPMC 2012. He has received numerous awards, including the 2010 Junior Researcher Award from Academia Sinica, the 2009 Ten Outstanding Young Persons Award of Taiwan, the 2007 IEEE ComSoc Asia-Pacific Young Researcher Award, the 2006 Youth Engineer Award of the Chinese Institute of Electrical Engineering, the 2005 Wu Ta You Memorial Award of NSC, and the 2005 Fu Suu-Nien Award of NTU. He is a senior member of the IEEE and a member of the ACM.

Yi-Bing Lin is the vice president and a lifetime chair professor at National Chiao Tung University. He serves on the editorial board of the *IEEE Transactions on Vehicular Technology* and has been a chair for prestigious conferences including ACM MobiCom 2002 and a guest editor for several journals including *IEEE Transactions on Computers*. He is the author of the books *Wireless and Mobile Network Architecture* (Wiley, 2001), *Wireless and Mobile All-IP Networks* (John Wiley, 2005), and *Charging for Mobile All-IP Telecommunications* (Wiley, 2008). He has received numerous research awards, including the 2005 NSC Distinguished Researcher Award, the 2006 Academic Award from the Ministry of Education, the 2008 Award for Outstanding Contributions in Science and Technology from the Executive Yuan, the 2011 National Chair Award, and the TWAS Prize in Engineering Sciences in 2011 from the Academy of Sciences for the Developing World. He is on advisory or review boards for the Ministry of Economic Affairs, Ministry of Education, Ministry of Transportation and Communications, and National Science Council. He is a member of the board of directors for Chunghwa Telecom. He is a fellow of the IEEE, AAAS, ACM, and IET.