Analysis and Comparison of Location Strategies for Reducing Registration Cost in PCS Networks

Ing-Ray Chen, Tsong-Min Chen and Chiang Lee*

Abstract

In mobile environments, a personal communication service (PCS) network must keep track of the location of each mobile user in order to correctly deliver calls. A basic scheme used in the standard IS-41 and GSM protocols is to always update the location of a mobile user whenever the mobile user moves to a new location. The problem with this approach is that the cost of location update operations is very high especially when the mobile user moves frequently. In recent years, various location management strategies for reducing the location update cost have been proposed. However, the performance issue of these proposed algorithms remains to be investigated. In this paper, we develop two Markov chains to analyze and compare the performance of two promising location update strategies, i.e., the two location algorithm (TLA) and the forwarding and resetting algorithm (FRA). By utilizing the Markov chain, we are able to quickly answer what-if questions regarding the performance of PCS networks under various workload conditions and also identify conditions under which one strategy can perform better than the others.

Using the cost due to location update and search operations between two successive calls to a mobile user as a performance measure, we show that when the mobile user exhibits a high degree of locality and the mobile user's call-to-mobility ratio (CMR) is low, TLA can significantly outperform both FRA and IS-41. On the other hand, when CMR is high, FRA is the winner. Furthermore, unlike TLA which may perform worse than IS-41 at high CMR values, FRA at identified optimal conditions can always perform at least as good as IS-41, suggesting that FRA over TLA should be used as a generic scheme in reducing the location cost in the PCS network to cover all possible per-user CMR values.

Index Terms — Location management algorithms, personal communication services, call to mobility ratio, registration cost evaluation, Markov chains.

*I.R. Chen is with Department of Computer Science, Virginia Tech, Northern Virginia Graduate Center, 7054 Haycock Road, Falls Church, VA 22043, USA. Email address: irchen@cs.vt.edu. T.M. Chen and Chiang Lee are with Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan.

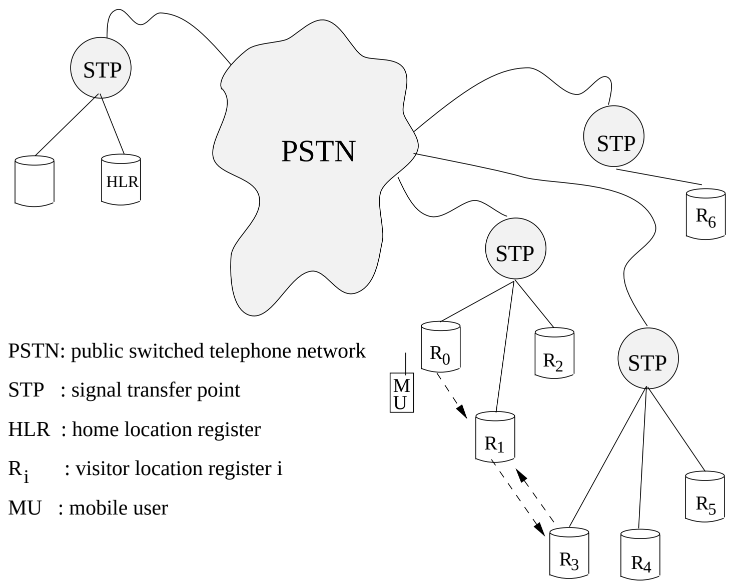

Figure 1: A Hierarchical PCS Network.

1 Introduction

In a Personal Communication Services (PCS) network, a location management scheme must handle two operations efficiently: location registration and call delivery. The former operation occurs when a mobile user moves to a new location and therefore the network must know where it is; the latter operation occurs when there is a call for the mobile user and the network must deliver the call. Since it is possible for a mobile user to move from one place to another while it is being called, the PCS network must track the location of each mobile user in order to correctly deliver calls. A well known basic and simple scheme is to update the location of each mobile user as it moves to a new location. This scheme exists in IS-41 [2] in the United States and GSM [9] in Europe and is commonly known as the basic HLR/VLR location registration scheme.

Under the basic HLR/VLR scheme, a mobile user is permanently registered under a location register called the home location register (HLR). Whenever the mobile user moves to a new location not directly under the control of the HLR, it reports to a visitor location register (VLR) which informs the HLR that the mobile user is now in its area. When there is a call asking for the mobile user, the PCS network always checks with the HLR of the mobile user to know the current VLR of the mobile user and then the call is delivered to the current VLR. Figure 1 shows a hierarchical PCS network as discussed in [3] in which there is one HLR for each mobile user; the mobile user may go to different registration areas under different VLRs (marked R_0, R_1, R_2, etc. in Figure 1). In Figure 1, PSTN refers to the Public Switches Telephone Network and STP refers to Switch Transfer Point. Each STP can connect several VLRs and also possibly connect the HLR with VLRs via the PSTN. Note that only the HLR and VLRs store the mobile user location information. As we can see, when the HLR and VLRs are in different segments of the network, the location update cost can be substantial under the basic HLR/VLR scheme, especially when the mobile user moves frequently since every move involves a connection cost between the HLR and the VLR which the mobile user moves into.

In recent years, various location management strategies for reducing the location update cost have been proposed with the goal of minimizing the PCS network and database loads. When the frequency of the incoming calls is higher than the mobile user's mobility, that is, when call-to-mobility ratio (CMR) is high, the location cache scheme [3] is proposed to reduce the number of locating operations. When CMR is low, on the other hand, it is reported that the forwarding and resetting algorithm (FRA) [4, 10], the alternative location strategy (ALS) [12] and the two location algorithm (TLA) [7, 8] can be used to reduce the location update cost. The analysis of each of these proposed algorithms, however, is done on a per scheme basis alone; it is not clear how these algorithms would fare when they are compared with each other under identical workload conditions. Most work done so far also reported the performance data based on simulation (e.g., [8]) which is laborious and difficult to repeat.

In this paper, we compare the performance of two promising location management schemes, namely, FRA and TLA. The basic idea under FRA is that whenever a mobile user moves to a new VLR area, only a pointer is set-up between the two involved VLRs and there is no need to inform the HLR. Therefore, when a call is delivered, the PCS network must follow a chain of pointers to locate the current VLR. To constrain the search cost, the forwarding chain must be periodically reset. The basic idea under TLA is that the HLR is still updated, but instead of recording only the current VLR in the HLR location database as in the basic scheme, the two most recently visited VLRs are recorded in the HLR database. Therefore, in some cases when a mobile user makes a move such as going back and forth between two VLRs, the location database in HLR does not have to be updated since it keeps the two most recently visited VLRs in the database. We pick these two schemes because reportedly they both perform well under low CMR conditions.

Our approach is based on analytical modeling, that is, we use two separate Markov models to describe the behavior of the PCS network under these two location management schemes separately and then "parameterize" the models (give values to model parameters) for the same workload setting and network condition. Then, we vary workload conditions and assess the situation under which one scheme performs better than the other. Unlike previous performance studies concentrating on only a particular scheme or based on simulation, our approach has the advantage of allowing what-if types of questions to be answered quickly and it can be repeated or modified to analyze other location management location schemes. In this paper, we use the cost due to location update and search operations between two successive calls to a mobile user as a basis to compare these two algorithms. We also use IS-41 as the baseline model for comparison.

The rest of the paper is organized as follows. Section 2 introduces the system model. Section 3 describes TLA in detail and develops a Markov model to describe the behavior of the PCS network operating under TLA. Section 4 does the same as in Section 3 except that the Markov model is developed for the FRA scheme. In addition, since the performance of the FRA scheme strongly depends on the length of the resetting period as a function of the input workload condition, we show how the Markov model developed can be used to determine the optimal resetting period in order to maximize the system performance. Section 5 parameterizes these two separate Markov models so as to compare the performance of these two algorithms under identical conditions. Section 6 summarizes the paper and outlines some possible future research areas.

2 System Description

There is no assumption concerning the structure of the PCS network. Conceptually, for a mobile user its HLR is at the higher level while all VLRs that it wanders into from time to time are at the lower level. There may be some STP switches connecting the HLR to VLRs (see Figure 1). For IS-41, all service areas are divided into many registration areas (RA) each corresponding to a VLR. Therefore, when a mobile user moves to a new RA, the mobile user can send the registration information to the new VLR which in turn can perform appropriate update actions, depending on the scheme being used. A VLR or even the mobile unit itself may also keep a small database for location management. The type of database kept by each VLR or mobile user is also different depending on the scheme used. We will discuss this in more detail as we cover FRA and TLA later in Sections 3 and 4.

We assume that a mobile user can cross VLR boundaries freely as it is being called. The time that a particular mobile user stays within a VLR before moving to another one, i.e., dwell time within a VLR, is characterized by an exponential distribution with an average rate of σ. Such a parameter can be estimated using the approach described in [4, 6] on a per user basis. The interarrival time between two consecutive calls to a particular mobile user, regardless of the current location of the user, is also assumed to be exponentially distributed with an average rate of λ. A mobile user thus is characterized by its call-to-mobility ratio (CMR), defined as λ/σ.

Our cost model for evaluating location management schemes follows [7] and includes two components: (a) update cost – the cost of updating the location of the mobile user due to user movements; and (b) query cost – the cost of searching the user in response to a call. The relative magnitudes of these two components vary depending on the location management scheme used. Note that this cost model excludes factors not related to location management such as the service time of a connection after the connection is established or queueing delays due to call blocking events. For a location management scheme X, let X_{reg} be the average cost of the PCS network in servicing a location update operation due to a user movement acrossing registration boundaries and X_{call} be the average cost in finding the mobile user. Furthermore, let X_{cost} be the average cost of the PCS network in servicing the above two types of operations between two consecutive calls. Then,

$$X_{cost} = X_{reg} \times \sigma/\lambda + X_{call} \tag{1}$$

Equation (1) is obtained above because between two consecutive calls, the number of mobility moves across registration boundaries by the mobile user is equal to σ/λ on average. One can imagine that a mobile user moves across registration boundaries for a number of times (σ/λ on average) before receiving a call and then the same pattern repeats again indefinitely. The X_{cost} parameter above gives the total cost incurred to the PCS network in each such repeated period accounting for both the location update and search costs, thus providing a uniform cost measure for fairly comparing all location management schemes. Note that the number of moves corresponds to the number of registration operations, although some of which may not cause any update cost to the PCS network for some location management scheme.

3 Modeling PCS Network under TLA

In this section, we develop a Markov model to describe the behavior of the PCS network operating under TLA as it services location update and search operations of a mobile user.

Under the TLA scheme, a mobile user as well as its HLR each keep a *location table* to store two recently visited VLRs. A time stamp is used to tell which VLR is the most recent VLR visited by the mobile user. When a mobile user moves to a new VLR which is not one of the two in the table, an update operation is initiated by the mobile user so that both the location tables in the mobile unit and in the HLR are updated. Figure 1 shows a mobile user moving from R_0 to R_1, then R_3, and finally back to R_1, i.e., $R_0 \rightarrow R_1 \rightarrow R_3 \rightarrow R_1$. When the mobile user moves from R_0 to R_1, the update operation is performed at the HLR. The same operation is executed when the mobile user moves from R_1 to R_3. However, the last move (i.e. from R_3 to R_1) will not update the location table at the HLR since R_1 is already in the location table. In the last case, the registration cost is considered zero for the PCS network.

Table 1. Notation Used for the TLA Model.

λ	arrival rate of calls to the mobile user, i.e., number of calls per unit time
σ	mobility rate of the mobile user, i.e., the reciprocal of the dwell time of the mobile user staying in a VLR
θ	probability of the mobile user moving back to the previous VLR
CMR	call-to-mobility ratio of the mobile user, i.e., CMR=λ/σ
σ_1	mobility rate of the mobile user moving to a new VLR, i.e., $\sigma_1 = (1-\theta)\sigma$
σ_2	mobility rate of the mobile user moving to the previous VLR, i.e., $\sigma_2 = \theta\sigma$
μ_g	execution rate in locating the mobile user when the location table in the mobile unit is consistent with that in the HLR
μ_b	execution rate in locating the mobile user when the location table in the mobile unit is inconsistent with that in the HLR
δ	execution rate in updating the location table stored in the HLR
TLA_{reg}	average cost of the PCS network under TLA in servicing a registration operation
TLA_{call}	average cost of the PCS network under TLA in locating the mobile user
TLA_{cost}	average cost of the PCS network under TLA in servicing location update and find operations between two consecutive calls
$P_{(i,j,k)}$	the probability that the system stays in state (i,j,k) in equilibrium

The state of a mobile user as it crosses database boundaries while being called can be described by a 3-component state description vector (a, b, c). Component a is a binary quantity indicating whether or not the mobile unit is in the state of being called, with 0 standing idle and 1 standing busy. Component b is also a binary quantity indicating if the mobile user has just moved to a new registration area, with 1 meaning yes and 0 meaning no. The third component, c, indicates if the location table maintained by the mobile unit is inconsistent with that maintained by the HLR, with 0 meaning consistent and 1 meaning inconsistent.

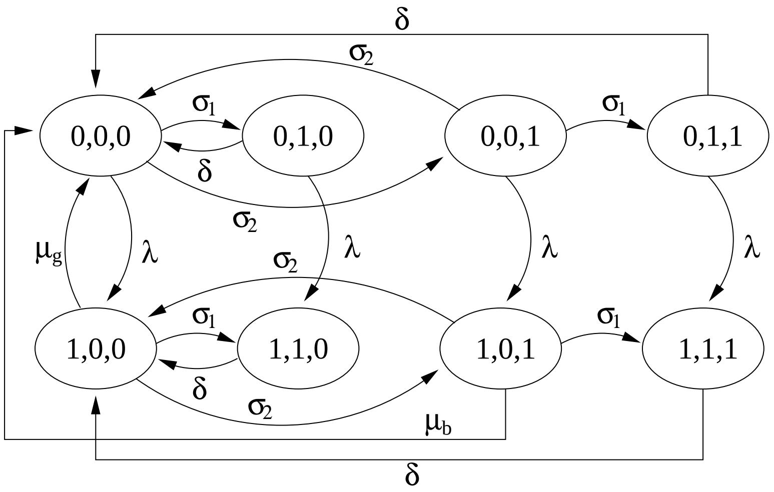

Figure 2: Markov model for PCS network under Two Location Algorithm.

Figure 2 shows the Markov model for describing the PCS operating under TLA. Table 1 shows the notation used for the TLA model. Initially, the mobile user is in the state of $(0, 0, 0)$, meaning that it is not being called and the mobile user has not yet made any move across any registration area boundary. Below, we explain briefly how we construct the Markov model.

First, if the mobile user is in the state of $(0, i, j)$, $0 \leq i, j \leq 1$, and a call arrives, then the new state is $(1, i, j)$, i.e., the mobile user is now in the state of being called. This behavior is modeled by the (downward) transition from state $(0, i, j)$ to state $(1, i, j)$, $0 \leq i, j \leq 1$, with a transition rate of λ.

Second, if the mobile user is in the state of $(1, i, j)$ and another call arrives, then the mobile user will remain at the same state, since the mobile user remains in the state of being called. This behavior is described by a hidden transition from state $(1, i, j)$ back to itself with a transition rate of λ. This type of transition is not shown in Figure 2 since it does not need to be considered when solving a Markov chain [5]. Note that this implies that in state $(1, i, j)$ the number of requests accumulated to locate the mobile user may be greater than one.

Third, if the mobile user is in the state of $(1,0,0)$, it means that the location table stored in the HLR is consistent with that stored in the mobile unit and the mobile unit is in the state of being called. Therefore, the PCS network can service all pending calls simultaneously with a service rate of μ_g. After the service, the new state is $(0,0,0)$.

Four, if the mobile user is in the state of $(1,0,1)$, it means that the location table stored in the HLR is inconsistent with that stored in the mobile unit but there are pending calls waiting to be serviced. Therefore, the PCS network has to spend twice as much time to locate the mobile unit. This behavior is modeled by using a different service rate of μ_b from state $(1,0,1)$ to state $(0,0,0)$. After the service, the new state is $(0,0,0)$ because the location table stored in the HLR is updated after the call delivery service and is therefore consistent with that stored in the mobile unit again.

Lastly, regardless of whether the mobile user is in the state of being called or not, the mobile user can move across a registration area boundary. There are two cases:

1. If the mobile unit moves to a new RA then an update operation has to be performed to the table stored in the HLR. This behavior is modeled by a transition from state $(i,0,j)$ to $(i,1,j)$, $0 \le i,j \le 1$, with a transition rate σ_1, after which the system transits from state $(i,1,j)$ to state $(i,0,0)$ with a transition rate of δ.

2. If the mobile user moves back to the previously visited RA, then there is no update operation required to update the HLR associated with this registration event. This is modeled by a transition from state $(i,0,0)$ to state $(i,0,1)$, $0 \le i \le 1$, with a transition rate σ_2, after which the location table stored in the HLR is inconsistent with that stored in the mobile unit, or from state $(i,0,1)$ to state $(i,0,0)$, $0 \le i \le 1$, also with a transition rate σ_2, after which the location table stored in the HLR on the contrary becomes consistent with that stored in the mobile unit. Note that the update time of the location table stored in the mobile unit with respect to the PCS network is zero (although the mobile unit's location table is updated itself). Therefore, there is no need to model the time needed to perform the registration operation in this case.

The Markov chain shown in Figure 2 is ergodic [5], which means that all states have a non-zero probability. The probability that the system is found in a particular state in equilibrium depends on the relative magnitude of the outgoing and incoming transitions rates. For example, if the call arrival rate λ is much greater than the mobility rate σ, then the probability that the system is found to stay in state $(1, i, j)$ would be much greater than in state $(0, i, j)$. Let TLA_{reg} be the average cost of the PCS network in servicing a registration operation[1] and let TLA_{call} be the average cost in locating the mobile user. Furthermore, let TLA_{cost} be the average cost of the PCS network in servicing the above two types of operations between two consecutive calls. Then,

$$TLA_{reg} = \left(\sum_{i=0}^{1}(P_{(0,0,i)} + P_{(1,0,i)}) \times (1-\theta) \times (1/\delta)\right) + \left(\sum_{i=0}^{1}(P_{(0,1,i)} + P_{(1,1,i)}) \times (1/\delta)\right) \tag{2}$$

$$TLA_{call} = \left(\sum_{i=0}^{1}\sum_{j=0}^{1} P_{(i,j,0)} \times (1/\mu_g)\right) + \left(\sum_{i=0}^{1}\sum_{j=0}^{1} P_{(i,j,1)} \times (1/\mu_b)\right) \tag{3}$$

Therefore, based on Equation (1),

$$TLA_{cost} = TLA_{reg} \times \sigma/\lambda + TLA_{call}. \tag{4}$$

4 Modeling PCS Network Under FRA

In this section, we develop another Markov model to describe the behavior of the PCS network operating under FRA. We consider that the forwarding chain is reset after a call is serviced since the exact location of the mobile unit is known after the call is serviced.

We describe the behavior of the mobile user in this case by two state components: (a) a binary quantity indicating whether or not the mobile unit is in the state of being called, (b) the number of forwarding steps which have accumulated. Figure 3 shows a Markov model

[1] Note that some registration operation takes zero time for the PCS network to serve under TLA. These occur when the mobile unit moves back to a previous VLR. Here, TLA_{reg} is a parameter which stands for the average time for the PCS network under TLA to service registration operations due to moves made by the mobile unit acrossing registration area boundaries.

describing the behavior of a mobile user in the PCS network wherein a state is represented by (a, b) where a is either 0 (standing for IDLE) or 1 (standing for CALLED), while the other component b indicates the number of forwarding steps that has been made since the last reset operation. Initially, the mobile user is in the state of $(0, 0)$, meaning that it is not being called and the number of forwarding steps is zero. Below, we also explain briefly how we construct the Markov model.

Table 2. Notation Used for the FRA Model.

λ	arrival rate of calls to the mobile user
σ	mobility rate of the mobile user
CMR	call-to-mobility ratio of a particular mobile user, i.e., CMR $= \lambda/\sigma$
μ_p	execution rate in setting or following a pointer between two visitor databases
μ_i	execution rate in locating the mobile unit with i pointers on the forwarding chain
k	number of forwarding steps after which a reset operation is performed
m_k	execution rate in performing a reset operation for a forwarding chain with k pointers
FRA_{reg}	average cost of the PCS network under FRA in servicing a registration operation
FRA_{call}	average cost of the PCS network under FRA in locating the mobile user
FRA_{cost}	average cost of the PCS network under FRA in servicing location update and find operations between two consecutive calls
$P_{(i,j)}$	the probability that the system is in a particular state in equilibrium

First, if the mobile user is in the state of $(0, i)$, $0 \leq i < k$, and a call arrives, then the new state is $(1, i)$ in which the number of forwarding steps remains at i but the mobile user is now in the state of being called. This behavior is modeled by the (downward) transition from state $(0, i)$ to state $(1, i)$, $0 \leq i < k$, with a transition rate of λ. A similar transition occurs from state $(0, i + 1^*)$ to state $(1, i + 1^*)$, $0 \leq i < k$, also with a transition rate of λ. (The meaning of $(0, i + 1^*)$ and $(1, i + 1^*)$ will be explained below.)

Second, if the mobile user is in the state of $(1, i)$, the PCS network can service all pending calls simultaneously with a service rate of μ_i. After the service, the new state is $(0, 0)$ since

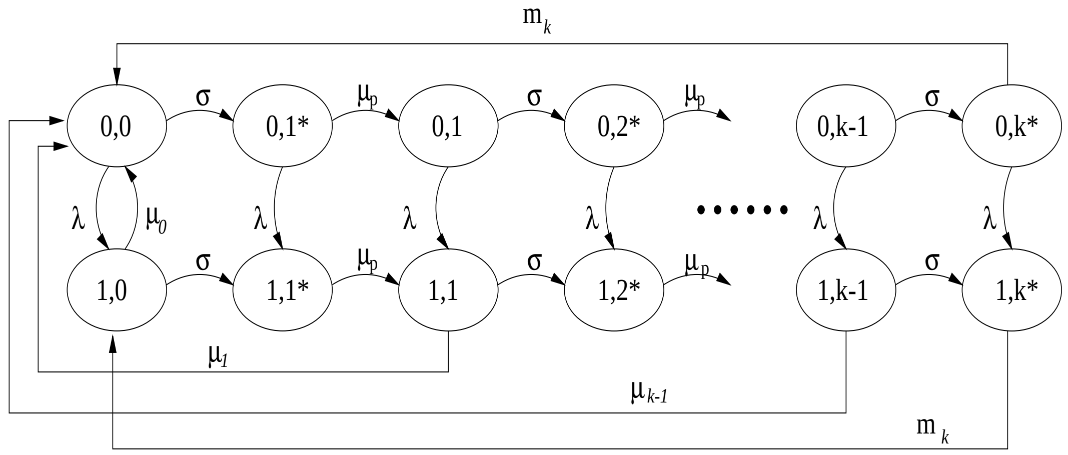

Figure 3: Markov model for PCS Network Under Forwarding and Resetting Algorithm.

all calls have been serviced and the reset operation is performed. This behavior is described by the state transition from state $(1, i)$ to state $(0, 0)$ with a transition rate of μ_i.

Finally, whenever the mobile user is in the state of $(0, i)$ (being idle) or $(1, i)$ (having been called), if the mobile user moves across a registration area boundary, then the new VLR, denoted by v_{i+1}, will determine if a pointer connection or a reset operation has to be performed. This "move" behavior is modeled by a transition from state $(0, i)$ to state $(0, i+1)^*$ (if the mobile user is idle) or from state $(1, i)$ to state $(1, i+1)^*$ (if the mobile user is in the state of being called), with a mobility rate of σ. There are two separate cases to be considered as the mobile user moves across a registration area boundary.

1. If $0 \le i < k-1$, then the new visitor database v_{i+1} simply sets up a pointer connection between v_i and v_{i+1}. This behavior is modeled by a transition from state $(1, i+1)^*$ to state $(1, i+1)$ (if busy) or from state $(0, i+1)^*$ to state $(0, i+1)$ (if idle) with a transition rate of μ_p. Note that we assume that when a mobile user moves across a visitor database boundary, it will make sure that the pointer connection is established properly between the two involved VLRs before it will cross another registration area boundary.

2. If $i = k-1$, then the new visitor database v_{i+1} knows that the mobile unit has made

k boundary moves. Therefore, a reset operation will be invoked. This is modeled by a transition from state $(1,k)^*$ to state $(1,0)$ (if busy) or from state $(0,k)^*$ to state $(0,0)$ (if idle) with a transition rate of m_k.

Now let FRA_{reg} be the average cost of the PCS network under FRA in servicing a registration operation[2]; FRA_{call} be the average cost in locating the mobile user; and let FRA_{cost} be the average cost of the PCS network in servicing the above two types of operations between two consecutive calls. Then,

$$\begin{aligned} FRA_{reg} = & \left(P_{(0,k-1)} + P_{(1,k-1)} + P_{(0,k)^*} + P_{(1,k)^*}\right) \times (1/m_k) \\ & + \left(\sum_{i=0}^{k-2}(P_{(0,i)} + P_{(1,i)}) + \sum_{i=1}^{k-1}(P_{(0,i)^*} + P_{(1,i)^*})\right) \times (1/\mu_p) \end{aligned} \tag{5}$$

$$\begin{aligned} FRA_{call} = & \left(\sum_{i=0}^{k-1}(P_{(0,i)} + P_{(1,i)}) \times (1/\mu_i)\right) + \left(\sum_{i=0}^{k-2}(P_{(0,i+1)^*} + P_{(1,i+1)^*}) \times (1/\mu_{i+1})\right) \\ & + (P_{(0,k)^*} + P_{(1,k)^*}) \times (1/\mu_0) \end{aligned} \tag{6}$$

$$FRA_{cost} = FRA_{reg} \times \sigma/\lambda + FRA_{call} \tag{7}$$

Equation (7) above yields FRA_{cost} as a function of k. For a given set of parameter values, we can first compute the values of $P_{(i,j)}$ for all states and then use Equation (7) to determine the best value of k that can minimize the cost. Of course, different PCS network structures may give different parameter values and thus may yield different optimal k values.

5 Analysis and Comparison

In this section, we compare TLA, FRA and IS-41 under identical set of conditions using the Markov models developed in the last two sections. We list the network cost parameters in Table 3. The first two parameters apply to all schemes, while the next two parameters apply to FRA specifically.

[2]Note that a registration operation may execute at rate μ_p or m_k depending on whether a reset operation is performed or not in response to the location registration (or movement) event.

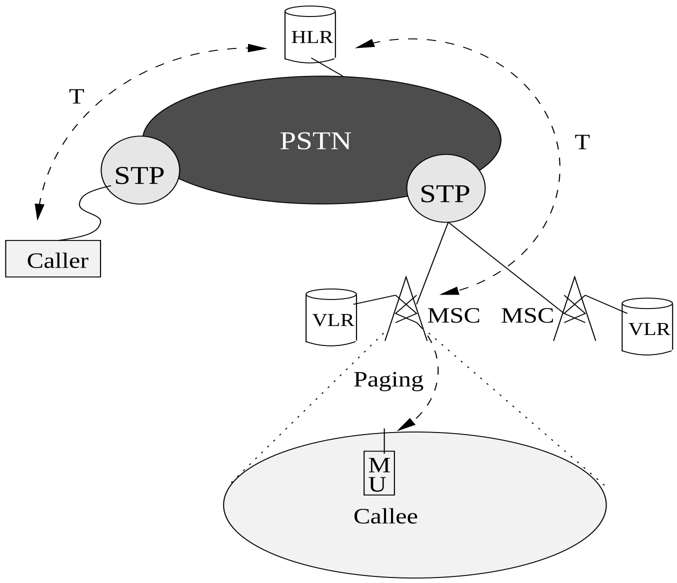

Figure 4: Caller to Callee Path to Illustrate Relation between Parameters U and T.

Table 3. Network Cost Parameters.

U	the average (round trip) search cost for finding the mobile user under the basic scheme; it is also the search cost under the TLA scheme when the HLR's location table is consistent with that stored in the mobile unit, as well as the search cost under the FRA scheme to find the first VLR on the forwarding chain and then to locate the user under it
T	the average (round trip) communication cost between HLR and VLR; it is also the registration cost under the basic scheme, the registration cost to update the location table in the HLR under the TLA scheme, as well as the average cost of a reset operation under the FRA scheme.
τ	the average (round trip) communication cost between VLR and VLR; it is also the cost of a registration operation (i.e., a location update operation) under the FRA scheme before a reset operation is performed
α	the ratio of the VLR-VLR communication cost to the VLR-HLR communication cost, i.e., $$\alpha = \frac{\tau}{T}$$ The physical meaning of α is that it corresponds to the amount of saving in registration cost as a result of setting-up a VLR-VLR pointer instead of performing a VLR-HLR reset operation in the FRA scheme

As illustrated in Figure 4, the relationship between U and T is as follows:

$$U = 2T \ + \ P_c$$

where P_c is the average paging cost needed to page a mobile user under a VLR. Specific values of U, T and τ above can be obtained by considering specific network coverage models (see for example [1]). In this paper, we fix the value of U at 1 as used in [7], but use different values of T and τ to study their effects. Note that the range of T is between 0 to 0.5 since we fix U at 1. Recall that U corresponds to the basic search cost under IS-41.

For all the performance data reported below, we use the cost measure defined in Equation 1. All costs are normalized with respect to $U = 1$; therefore, a cost value seen hereafter is expressed in terms of U, e.g., 3 means 3 U. The average network cost of the PCS network under IS-41, based on Equation 1, is given by

$$IS-41_{cost} = IS-41_{reg} \times \sigma/\lambda + IS-41_{call} \tag{8}$$

where $IS-41_{reg} = T$ and $IS-41_{call} = U$, as defined in Table 3; $IS-41_{cost}$ will be used as a baseline cost measure with which TLA_{cost} and FRA_{cost} (defined by Equations 4 and 7 respectively) are compared. A final note is that the cost measure defined in Equation 1 accounts for the network cost incurred due to location management in a *repeated* period between two consecutive calls, so even a cost difference of 0.1 U between two location management schemes should be considered significant since the effect is cumulative.

5.1 Parameterization of the TLA Markov Model

There are six parameters in the TLA Markov model (see Figure 2), i.e., μ_g, μ_b, δ, σ, λ, and θ (see Table 1 for their meanings). Of these six parameters, σ, λ and θ are mobile-user dependent parameters and will be studied in the paper by changing their values; on the other hand, μ_g, μ_b, and δ are network structure dependent and can be parameterized as follows:

$$\mu_g = \frac{1}{U}$$

$$\mu_b = \frac{1}{U+T+P_c} = \frac{1}{2U-T}$$
$$\delta = \frac{1}{T}$$

where U and T are as defined in Table 3.

5.2 Parameterization of the FRA Markov Model

There are five parameters in the FRA Markov model (see Figure 3), i.e., μ_p, μ_i (i from 0 to k-1), m_k, σ and λ (see Table 2 for their meanings). Of these five parameters, σ and λ are again mobile-user dependent and will be studied in the paper by changing their values. All other parameters are network structure dependent and can be parameterized as follows:

$$m_k = \frac{1}{T}$$
$$\mu_p = \frac{1}{\tau}$$
$$\mu_i = \frac{1}{U + (i \times \tau)}$$

Note that in parameterizing m_k, we assume that the obsolete VLR database records stored in the old VLRs along the forwarding chain will be deleted automatically when they are replaced by future new database records, so there will be no explicit de-registration messages sent from the HLR or from the new VLR to delete the obsolete database records. Note that in this case

$$\alpha = \frac{\tau}{T} = \frac{m_k}{\mu_p}$$

5.3 Performance of TLA

Figure 5 shows the average costs (with respect to $U = 1$) of the PCS network under TLA and IS-41 as a function of θ and CMR values with T fixed at 0.2. (Recall that CMR = λ/μ.) The data for TLA on the figure were obtained by first solving the Markov model shown in Figure 2 using the SHARPE software package [11] to obtain the $P_{(i,j,k)}$ for each state (i, j, k) and then by computing TLA_{cost} based on Equation (4). The results correlate well with those reported in [7].

Figure 5 indicates that in this case when the CMR value is small, TLA can significantly outperform IS-41. For example, when CMR is 0.1 and θ is 0.3, TLA reduces 22% of the cost of IS-41; when CMR is 0.1 and θ is 0.7, TLA reduces the cost even further by 62%. It demonstrates that if CMR is small and the mobile user's move behavior exhibits a high degree of locality (i.e., when θ is closer to 1), then TLA will significantly outperform IS-41, largely due to the fact that (a) the network spends less time updating the user's location since most moves are local and do not require the table stored in the HLR to be updated; (b) call arrivals are rare, so even the HLR's table is not up-to-date most of the time, the location cost can still be properly amortized.

When CMR is relatively large and θ is also large, however, TLA may perform worse than IS-41. For example, when CMR =1.5 and θ=0.5 or 0.7, the average cost of TLA is higher than the cost of IS-41. The increased cost is due to the fact that there is a higher probability that the system will stay in state $(1, i, 1)$ in which there are calls waiting to be serviced but the HLR's database is out-of-date. In this case, the saving in updating the location of the mobile user cannot catch up with the extra search cost needed to locate the mobile user because the HLR will use the inconsistent database to locate the user and will search the user using twice as much time. Figure 5 shows that after CMR is larger than a threshold value, IS-41 performs better than TLA, regardless of the magnitude of the θ value.

The above result, however, is for the case in which $T = 0.2$, that is, when the VLR-HLR communication cost is relatively small compared to $U = 1$. As the VLR-HLR communication cost increases, the advantage of IS-41 over TLA disappears quickly. Figure 6 shows that when $T = 0.4$, TLA can perform better than IS-41 even when CMR goes as high as 1.0. In conclusion, Figures 5 and 6 exhibit the following performance characteristics of TLA: TLA is better than IS-41 in reducing the network cost due to location management when T is small and θ is large; the effect is especially pronounced when the mobile user has a small CMR value.

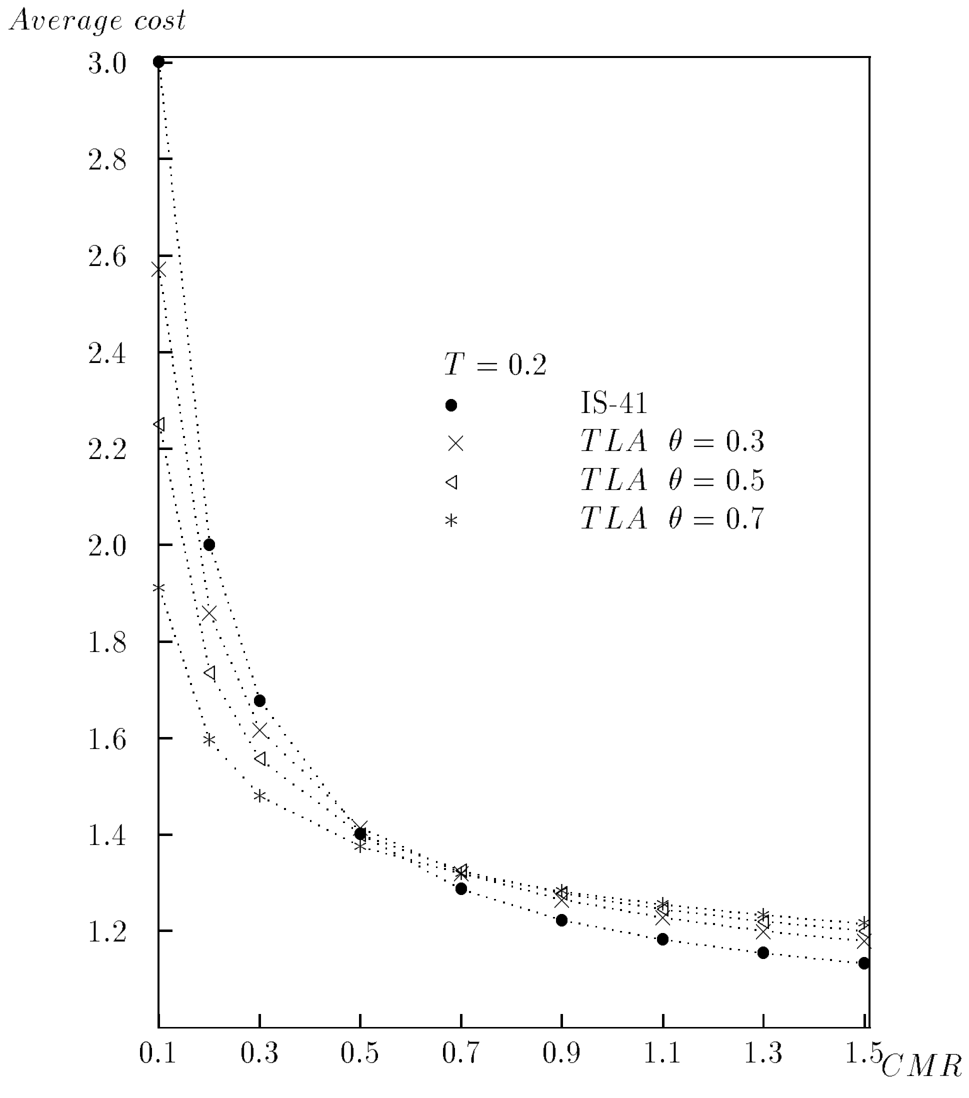

Figure 5: Comparing TLA with IS-41 under T=0.2 with Different θ and CMR Values.

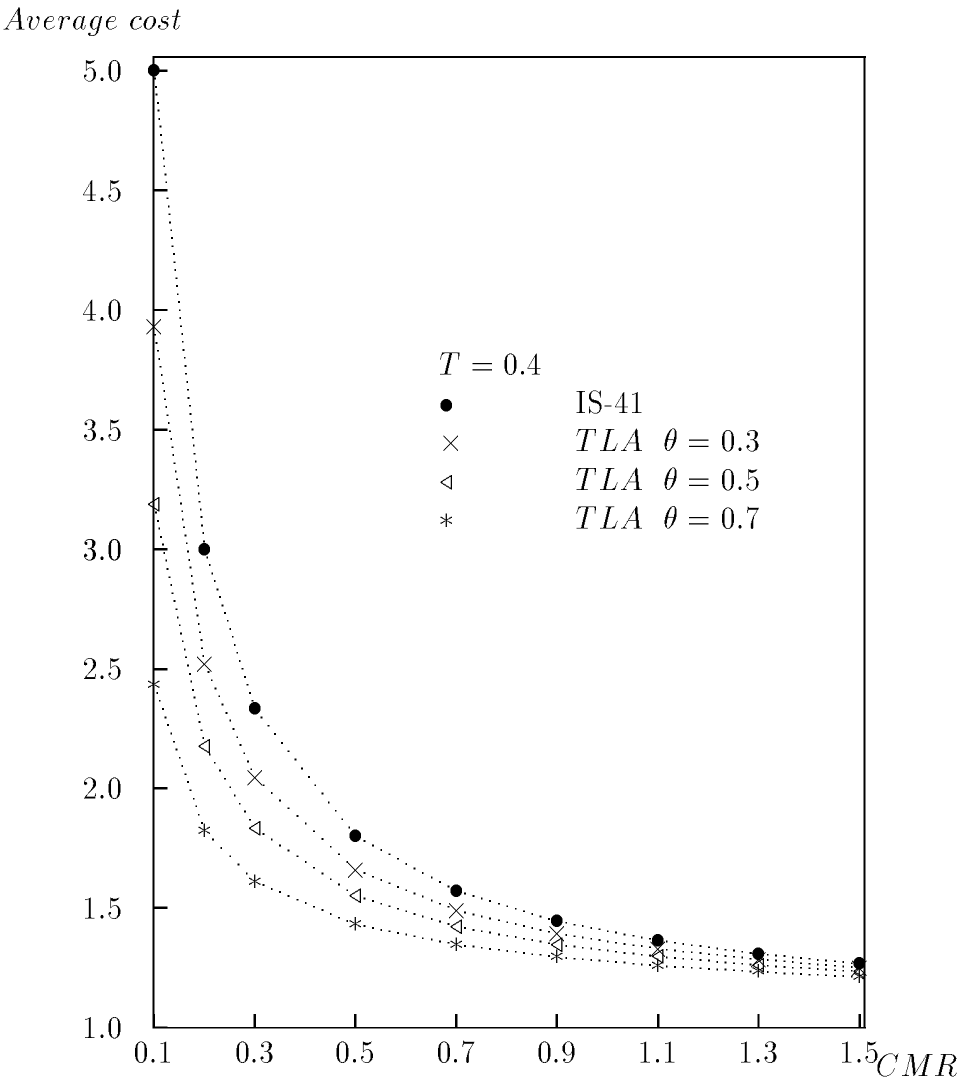

Figure 6: Comparing TLA with IS-41 under T=0.4 with Different θ and CMR Values.

5.4 Performance of FRA: Searching for Optimal Conditions

Figures 7 and 8 show FRA_{cost} vs k (number of forwarding steps after which a reset operation is performed) as a function of CMR, under $\alpha = 0.3$ and $\alpha = 0.7$, respectively. These two diagrams are based $T = 0.2$; The results under $T = 0.4$ are similar and thus are not shown to avoid repetition. The data were obtained by first solving the Markov model shown in Figure 3 to obtain $P_{(i,j)}$ for each state (i, j) and then by computing FRA_{cost} based on Equation (7).

Figure 7 is for the case when the VLR-VLR communication cost is relatively small ($\alpha = 0.3$), while Figure 8 is for the case when the VLR-VLR communication cost is relatively large compared with the VLR-HLR communication cost ($\alpha = 0.7$). This difference is reflected by the fact that the optimal values of k are higher in Figure 7 than those in Figure 8 under the same set of CMR values. The most important conclusion here is that there exists an optimal k value under which the cost is minimized for each given CMR value. From observing data shown in Figures 7 and 8, we also conclude that (a) for any two network structures under the same CMR value, the structure with a smaller α value (Figure 7) is likely to have a higher optimal k value than the structure with a larger α ratio (Figure 8), and (b) the optimal k value decreases as CMR increases. Another note that is worth mentioning is that the cost difference between the optimal and nonoptimal cases is significant, e.g., the difference can be as high as 1 when CMR = 0.1. Also note that the effect is cumulative since the cost measure is based on Equation 1 which accounts for the cost incurred for location management between two consecutive calls only.

5.5 Comparing TLA, FRA and IS-41

In this subsection, we compare the performance of TLA, FRA and IS-41 under identical conditions.

We make the following observations as we compare TLA, FRA and IS-41.

1. The CMR ratio is essentially a per-user parameter, so we must consider a wide CMR value range to cover all possible cases.

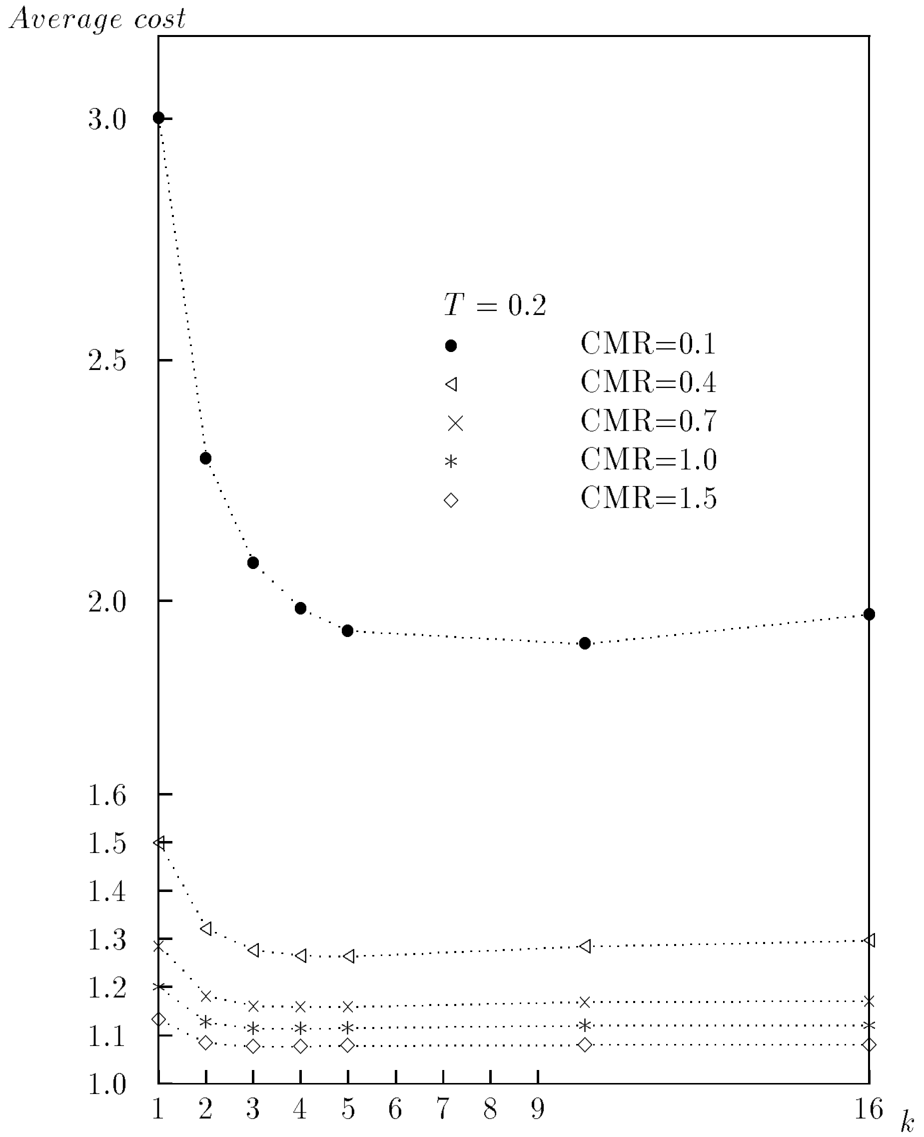

Figure 7: Optimal Number of Forwarding Steps for FRA under α=0.3 and T=0.2.

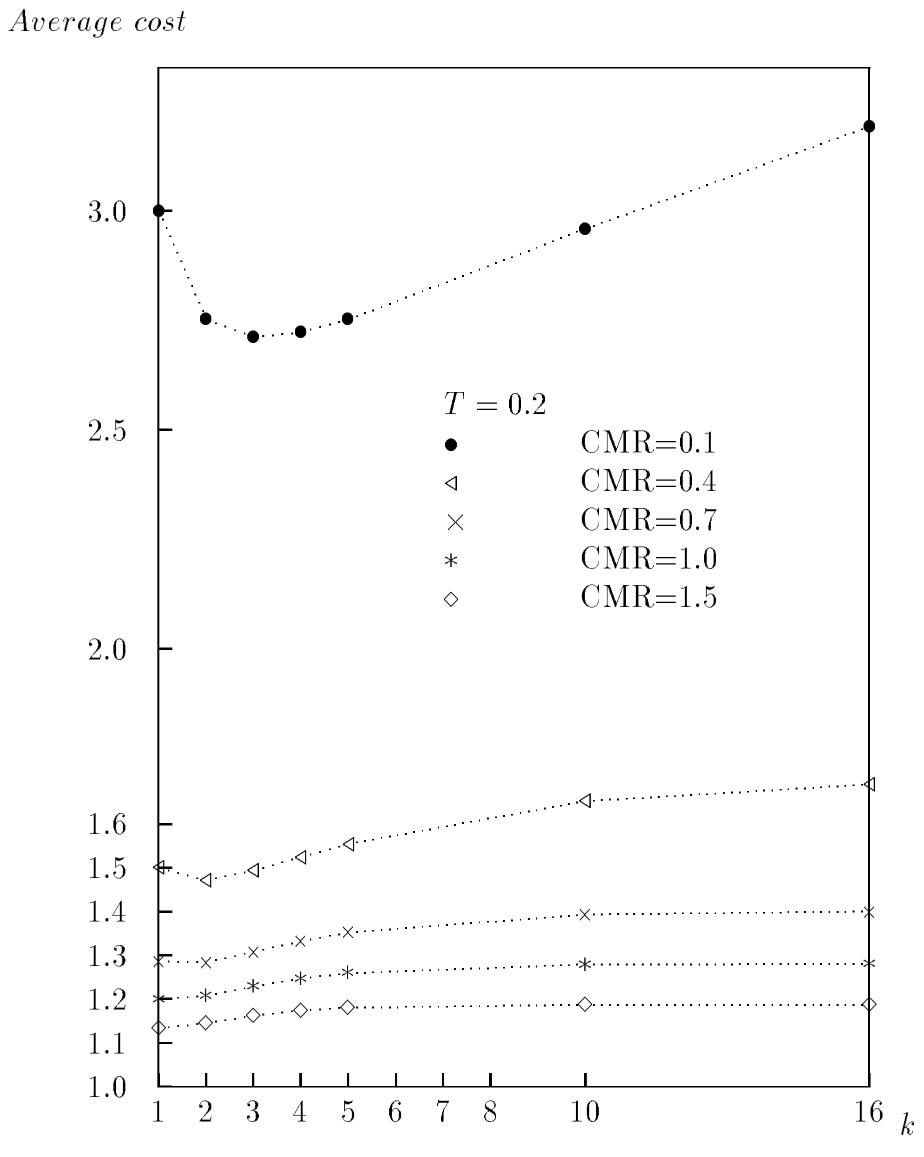

Figure 8: Optimal Number of Forwarding Steps for FRA under α=0.7 and T=0.2.

2. We like to compare TLA, FRA and IS-41 under identical conditions for the *same* user under the *same* CMR ratio. We note that all three schemes depend on U, T (see Table 3) and CMR. The cost under IS-41 can be calculated when the values of these three parameters are given based on Equation 8. TLA has one more parameter, namely, θ, and FRA also has one additional parameter, namely, α.

3. Although θ is a parameter inherently associated with TLA, we observe its relationship with α associated with FRA as follows: If θ is high (i.e., more local movements by the mobile user), then it is likely that α is low because the VLR-VLR communication cost is low compared with the VLR-HLR communication cost. This is so because the set of VLRs which the mobile user wanders into frequently are most likely to be under the same network switch (i.e., a STP in Figure 1). In other words, when θ is high the probability of the mobile user moving from one VLR covered by one STP to another VLR covered by another STP is low since its movements are mostly local. This effectively lowers the VLR-VLR communication cost relative to the VLR-HLR communication cost, thus resulting in a low α value under the FRA scheme. Therefore, when comparing TLA with FRA, we should compare TLA with high θ to FRA with low α, and vice versa, in order to get a fair, meaningful comparison.

Figure 9 draws the costs of FRA, TLA and IS-41 as a function of CMR under selected network conditions. T is fixed at 0.2 relative to $U = 1$. The cost on the y-axis is again defined in terms of the cost of U, e.g., 4 means that it is 4 times as much of the cost of U as defined in Table 3. The data points shown for FRA are at optimal k values. Figure 10 shows a similar comparison for the case when T is 0.4. Figures 9 and 10 suggest the following results:

1. When the CMR value is small, TLA with high-locality movements performs better than FRA and IS-41; this is especially so when T is large relative to U. For example, when CMR=0.1 and $T = 0.4$ (Figure 10), TLA with θ=0.7 has a lower cost than FRA with α=0.3 or 0.7.

2. If the CMR is greater than 0.7, then FRA with a lower α starts to outperform TLA and IS-41, especially when T is small relative to U, e.g., when CMR=1.0 and $T = 0.2$ (Figure 9) FRA with α=0.3 outperforms both IS-41 and TLA with θ=0.7.

3. Unlike TLA, FRA appears to perform better than or at least as good as IS-41 for all CMR values, even when α is a relatively large number (i.e., closer to 1). This result seems to favor FRA over TLA as a generic scheme in reducing the network cost due to location management for all possible CMR values. A reason behind this is that FRA at optimal k values can balance the location update cost with the location search cost better than the TLA scheme, by effectively trading off one cost factor for the other. At low CMR values, since user movements occur more frequently than call arrivals, FRA can effectively reduce the location update cost by maintaining a long forwarding chain at the expense of a long search cost. Conversely, at high CMR values, since calls arrive more frequently than user movements, it will effectively reduce the search cost by using a short forwarding chain. Thus, for all CMR values FRA is likely to maintain a performance level as good as IS-41.

Neither FRA nor TLA dominates under all conditions, although FRA appears to be a feasible scheme that can guarantee a performance level at least as good as IS-41 for all CMR values. It may also be possible to build a table based on the analysis result presented here and dynamically switch to the best location update algorithm as the per-user workload condition at the run-time changes so as to maximize the overall PCS network performance.

6 Conclusion and Future Work

In this paper, we developed two separate Markov models to analyze the performance characteristics of the PCS network operating under the Two Location Algorithm (TLA) and Forwarding and Resetting Algorithm (FRA) for location management, respectively. We used the total amount of time needed for the PCS network to service "location update" and "find" operations between two successive calls for the mobile user as a cost measure.

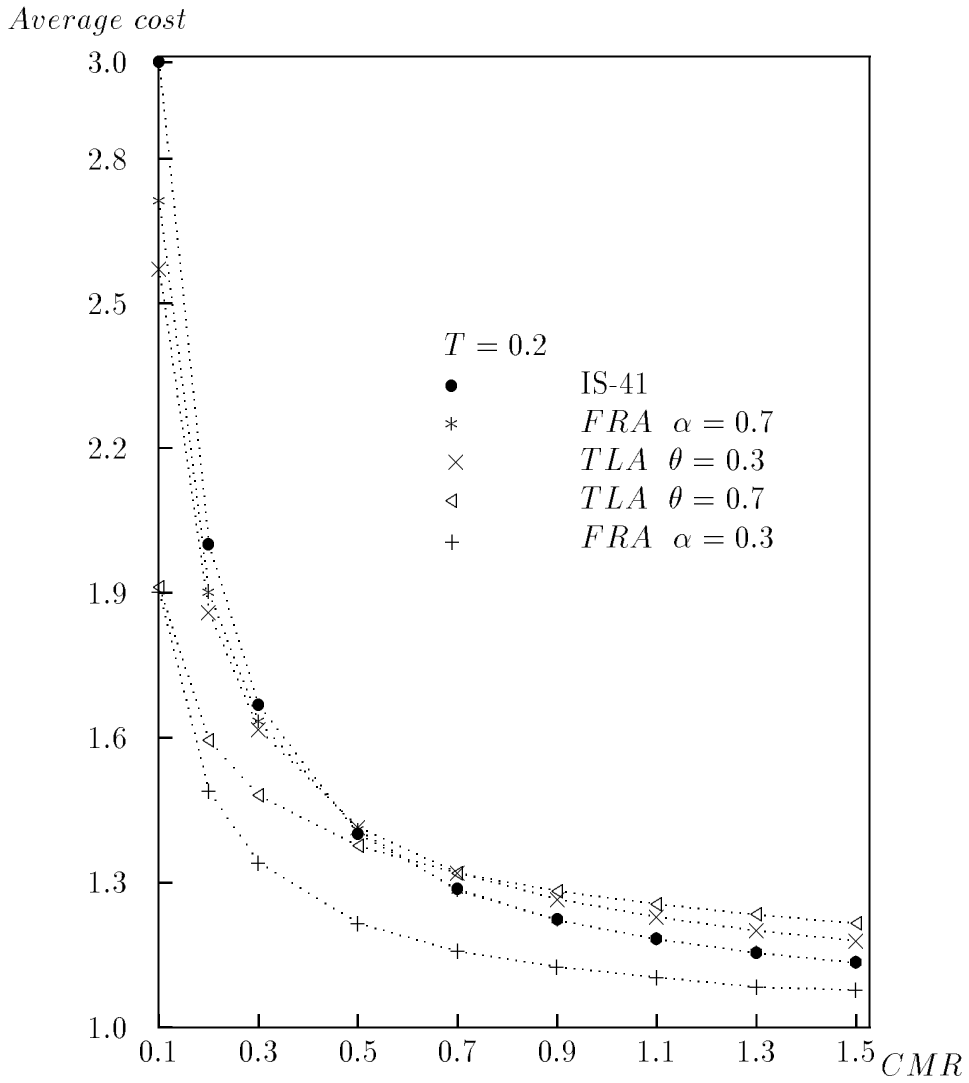

Figure 9: Comparison of IS-41, FRA (Under Optimal Settings) and TLA under $T = 0.2$.

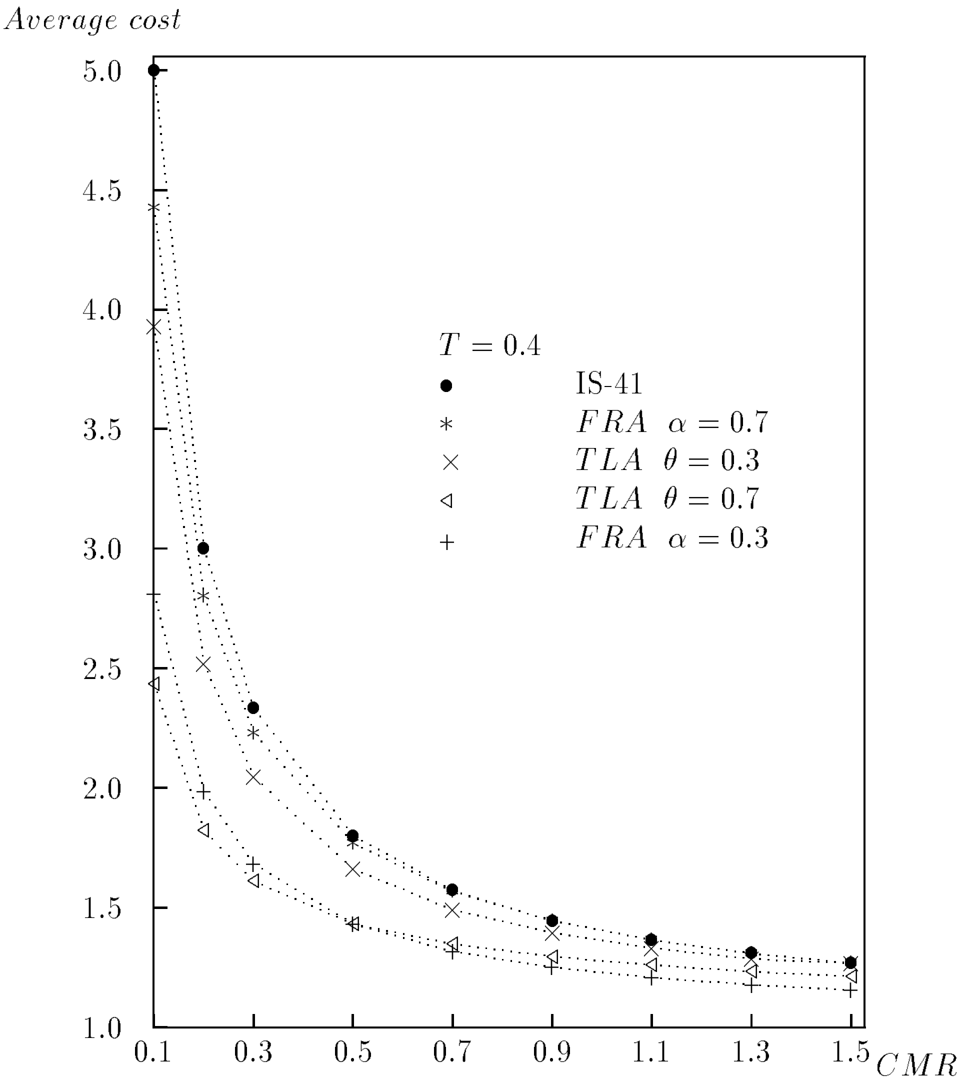

Figure 10: Comparison of IS-41, FRA (Under Optimal Settings) and TLA under $T = 0.4$.

We observed that if the mobile user exhibits a high degree of locality as it moves across registration area boundaries and also CMR is small, TLA can significantly outperform both FRA and IS-41, especially when the VLR-HLR communication cost is low. On the other hand, when CMR is large, it appears that FRA is the winner. Furthermore, our analysis result suggests that unlike TLA which may perform much worse than IS-41 at high CMR values, FRA at optimal k values can always perform as good as IS-41, even at high CMR values and high α values. The exact conditions under which one scheme is superior than the others and by how much can be assessed by using the Markov models developed in the paper.

Some future research areas related to this paper include (a) parameterizing network structure dependent parameters by considering specific network coverage models such as hexagonal and mesh models; this would allow us to estimate network structure dependent parameters more accurately and also to better relate the θ parameter in TLA and the α parameter in FRA; (b) generalizing the analysis method so that it can be applicable to all location management algorithms.

References

[1] I. R. Chen, T. M. Chen and C. Lee, "Performance Characterization of Forwarding Strategies in Personal Communication Networks," *21th IEEE International Conference of Computer Software and Application (COMPSAC '97),* Washington D.C., Aug. 1997, pp. 137-142.

[2] EIA/TIA, *Cellular Radio Telecommunication Inter system Operations,* Technical Report IS-41 (Revision B), EIA/TIA, July 1991.

[3] R. Jain, Y.B. Lin, C. Lo and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE Journal on Selected Areas in Communications,* Vol. 12, No. 8, Oct. 1994, pp. 1434-1444.

[4] R. Jain, Y.B. Lin, C. Lo and S. Mohan, "A forwarding strategy to reduce network impacts of PCS," *14th Annual Joint Conference of the IEEE Computer and Communications Societies, (IEEE INFOCOM '95)*, 1995, Vol. 2, pp. 481-489.

[5] L. Kleinrock, *Queueing Systems, Vol. 1: Theory*, John Wiley and Sons, 1975.

[6] W.R. Lai and Y.B. Lin, "Mobility database planning for PCS," *1996 Workshop Distributed System Technologies and Applications,* Tainan, Taiwan, 1996, pp. 263-269.

[7] Y.B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Transactions on Networking,* Vol. 5, No. 1, February, 1997, pp. 25-33.

[8] P. Lin and W.N. Tsai, "A Simulation Study of Multiple-Location Scheme for PCS Mobility Management," *3rd Workshop on Mobile Computing*, Taiwan, 1997, pp. 37-45.

[9] M. Mouly and M.B. Pautet, *The GSM System for Mobile Communications*, 49 rue Louise Bruneau, Palaiseau, France, 1992.

[10] S. Rao, B. Gopinath and D. Kurshan, "Optimizing call management of mobile units," *3rd IEEE Inter. Symp. Personal, Indoor and Mobile Communications*, 1992, pp. 225-229.

[11] R. Sahner, K. S. Trivedi and A. Puliafito, *Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package,* Kluwer Academic, 1996.

[12] S. Tabbane, "An Alternative Strategy for Location Tracking," *IEEE Journal on Selected Area in Communication,* Vol. 13, No. 5, June 1995.