在個人通訊服務網路中對於降低註冊成本位址策略之分析比較

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

陳聰敏
Tsong-Min Chen
成功大學資訊工程所
Institute of Information Engineering
National Cheng Kung University
tmchen@DBL.iie.ncku.edu.tw

陳英瑞
Ing-Ray Chen
維吉尼亞州立大學
Department of Computer Science
Virginia Poly. Institute and State University
irchen@plato.vt.cs.edu

摘要

本文使用馬可夫鏈以分析比較兩個位址策略之效能, 即 TLA 和 FRA, 經由此馬可夫鏈我們可以快速明瞭在何種狀況下, 各位址策略之優劣情形。

關鍵字 : 行動計算, 位址管理, 個人通訊網路, 馬可夫鏈。

Abstract

In this paper, we use 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 types of questions about the PCS network performance for various workload conditions and also identify conditions under which one strategy may performance better than the others.

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. A well known basic and simple scheme is to update the location of each mobile user as it moves to a new location. Figure 1 shows a hierarchical PCS network as discussed in [2] in which there is only one HLR for each mobile user, but the mobile user may go to different registration areas under different VLRs.

In recent years, various location management strategies for reducing the location update cost have been proposed with a 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 [2] 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) [3, 8], the alternative location strategy (ALS) [10] and the two location algorithm (TLA) [5, 6] can be used to reduce the location update cost. Most works done so far also reported the performance data based on simulation (e.g., [6]) 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. 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, two most recently visited VLRs are recorded in the HLR 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 the same network structure.

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 switches connecting the HLR to VLRs. For IS-41, all service areas are divided into many registration areas (RA) each corresponding to a VLR.

Figure 1 illustrates a possible hierarchical PCS network as discussed in [2]. The HLR and VLRs (marked R_0, R_1, R_2, etc. in the figure) and even the mobile unit itself each may contain a database for location management.

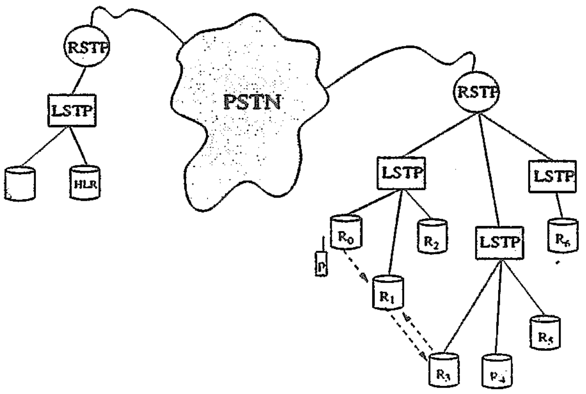

Figure 1: A Hierarchical PCS Network.

The intermediate switches such as RSTPs and LSTPs are only used for connecting the HLR with VLRs.

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 locating-user 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 HLR are updated.

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. Component b is also a binary quantity indicating if the mobile user has just moved to a new registration area. Component c indicates if the location table maintained by the mobile unit is inconsistent with that maintained by the HLR.

Figure 2 shows the Markov model for describing the PCS operating under TLA. 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.

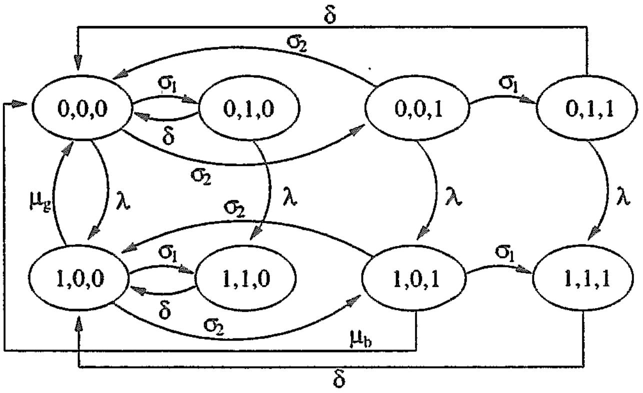

Figure 2: Markov model for PCS network under Two Location Algorithm.

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.

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 $(0,0,0)$.

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 $(i, 1, j)$, $0 \leq i, j \leq 1$, 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 $(i,0,1)$ or from state $(i,0,1)$ to state $(i,0,0)$, $0 \le i \le 1$, with a transition rate σ_2, after which the location table stored in the HLR is inconsistent with that stored in the mobile unit.

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. Let TLA_{reg} be the average cost of the PCS network in servicing a registration operation 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,

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

$$TLA_{call} = \left(\sum_{i=0}^{1}\sum_{j=0}^{1} P_{(i,j,0)} \times (1/\mu_g)\right) + \left(\sum_{i=0}^{1}\sum_{j=0}^{1} P_{(i,j,1)} \times (1/\mu_b)\right) \quad (2)$$

$$TLA_{cost} = TLA_{reg} \times \sigma/\lambda + TLA_{call} \quad (3)$$

Equation (3) is obtained above because between two consecutive calls, the number of mobility moves across VLR boundaries by the mobile user is equal to σ/λ on average. 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 depending on whether the location table in the HLR needs to be updated or not.

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.

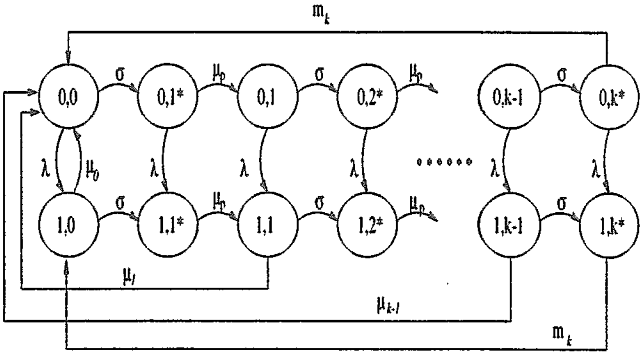

Figure 3: Markov model for PCS Network Under Forwarding and Resetting Algorithm.

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 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 explain briefly how we construct the Markov model.

First, if the mobile user is in the state of $(0,i)$, $0 \le i < k$, 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 $(1,i)$, $0 \le i < k$, with a transition rate of λ.

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 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.

Now let FRA_{reg} be the average cost of the PCS network under FRA in servicing a registration operation , 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,

$$\begin{aligned} FRA_{reg} = & \\ & \left(P_{(0,k-1)} + P_{(1,k-1)} + P_{(0,k)^*} + P_{(1,k)^*}\right) \times (1/m_k) \\ & + \\ & \left(\sum_{i=0}^{k-2}(P_{(0,i)} + P_{(1,i)}) + \sum_{i=1}^{k-1}(P_{(0,i)^*} + P_{(1,i)^*})\right) \times (1/\mu_p) \end{aligned} \qquad (4)$$

$$\begin{aligned} FRA_{call} = & \left(\sum_{i=0}^{k-1}(P_{(0,i)} + P_{(1,i)}) \times (1/\mu_i)\right) \\ & + \left(\sum_{i=0}^{k-2}(P_{(0,i+1)^*} + P_{(1,i+1)^*}) \times (1/\mu_{i+1})\right) \\ & + (P_{(0,k)^*} + P_{(1,k)^*}) \times (1/\mu_0) \end{aligned} \qquad (5)$$

$$FRA_{cost} = FRA_{reg} \times \sigma/\lambda + FRA_{call} \qquad (6)$$

Equation (6) 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 (6) 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.
U is the average cost for locating the mobile user under the basic scheme.
T is the average communication cost between HLR and VLR.
τ is the average communication cost between VLR and VLR.
Specific values of these network communication cost parameters can be obtained by considering specific network coverage models (see for example [1]).

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

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

$$\mu_b = \frac{1}{U+T}$$

and

$$\delta = \frac{1}{T}$$

5.2 Parameterization of the FRA Markov Model

There are also 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. 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 obsolete VLA pointer records in old VLRs along the forwarding chain will be deleted automatically when they are replaced by future new pointer records, so there will be no explicit deregistration messages sent from the HLR or from the new VLR to delete obsolete pointer records. To facilitate discussions, we introduce another parameter α which is simply the ratio of the VLR-VLR communication cost to the VLR-HLR communication cost.

5.3 Performance of TLA

Figure 4 shows the average costs of the PCS network under TLA and IS-41 as a function of θ parameter values and CMR values. (Recall that CMR $= \lambda/\mu$.) The data for TLA on the figure were obtained by first solving the Markov model shown in Figure 2 using the SHARPE software package [9] to obtain the $P_{(i,j,k)}$ for each state

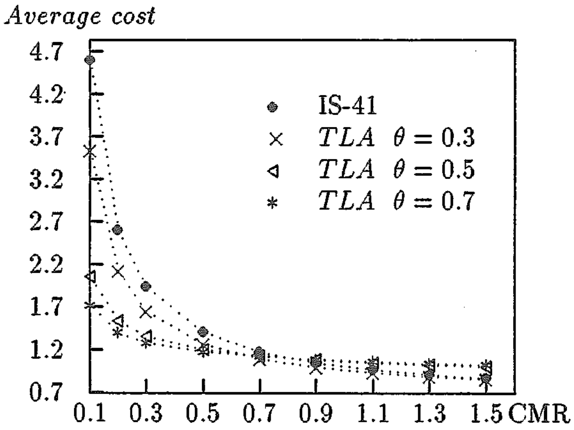

Figure 4: Comparing TLA with IS-41 under different θ and CMR values.

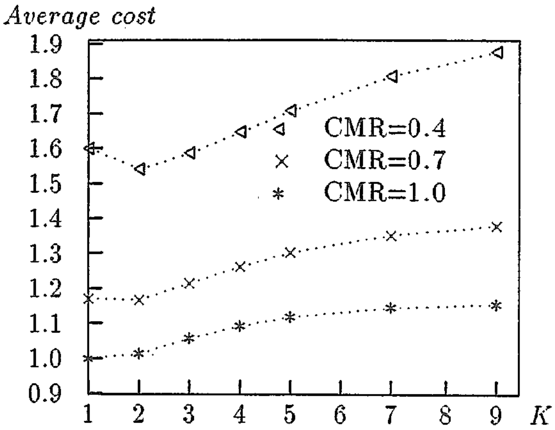

Figure 5: Optimal Number of Forwarding Steps for FRA under α=0.7.

(i, j, k) and then by computing TLA_{cost} based on Equation (3). The results correlate well with those reported in [5].

Figure 4 indicates that when the CMR value is small, TLA indeed can significantly outperform IS-41. 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. When CMR is relatively large and θ is also large, however, TLA performs 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. Figure 4 shows that after CMR is larger than a threshold value, IS-41 performs better than TLA, regardless of the magnitude of the θ value.

5.4 Performance of FRA: Searching for Optimal Conditions

Figures 5 show FRA_{cost} vs k (number of forwarding steps after which a reset operation is performed) as a function of CMR. 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 (6).

Figure 5 is for the case when the VLR-VLR communication cost is relatively small (large correspondingly) compared with the VLR-HLR communication cost. The most important conclusion here is that there exists an optimizing k value under which the cost is minimized for each given CMR value.

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 first note that although θ is a parameter inherently associated with the TLA scheme, it also bears some relationship with α. That is, 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 going to be low compared with the VLR-HLR communication cost since all VLRs are nearby. Therefore, when comparing TLA with FRA, we should compare TLA with high θ to FRA with low α, and vice versa. Figure 6 draws the costs of FRA, TLA and IS-41 as a function of CMR under identical network conditions. The data points shown for FRA are at optimizing k values. Figure 6 suggests the following results:

1. When CMR value is small, TLA with high-locality movements performs better than FRA and IS-41. For example, when CMR=0.1, 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 0.3, then FRA with a lower α outperforms TLA and IS-41, i.e., when

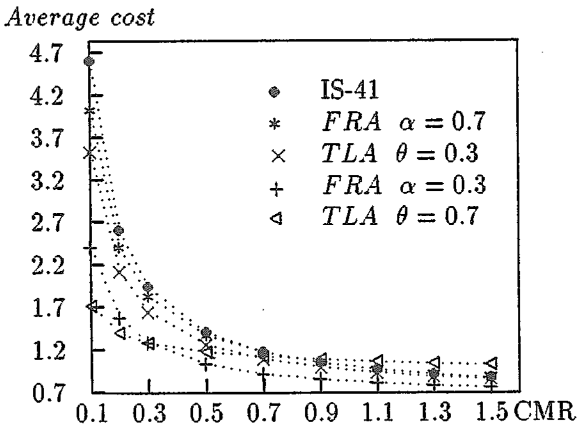

Figure 6: The comparison of IS-41, FRA, and TLA.

CMR=1.5, FRA with α=0.3 is better than both TLA and IS-41.

3. Unlike TLA, FRA appears to perform better than or at least as good as IS-41 under all conditions, even when α is a relatively large number (i.e., closer to 1). This result seems to favor FRA over TLA as a more general scheme under all conditions.

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. 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 workload condition at the run-time is detected on a per-user basis 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 using Two Location Algorithm (TLA) and Forwarding and Resetting Algorithm (FRA) for location management, respectively. 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. 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 performance worse than IS-41 at high CMR values, FRA at optimizing 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.

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.

[2] 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.

[3] 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.

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

[5] Y.B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Transactions on Networking*, 5(1):25-33, February, 1997.

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

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

[8] 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.

[9] 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.

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