Quantitative Analysis of a Hybrid Replication with Forwarding Strategy for Efficient and Uniform Location Management in Mobile Wireless Networks

Ing-Ray Chen, *Member*, *IEEE Computer Society*, and Baoshan Gu

Abstract—A location management scheme in wireless networks must effectively handle both user location update and search operations. Replication and forwarding are two well-known techniques to reduce user search and update costs, respectively, with replication being most effective when the call to mobility ratio (CMR) of the user is high, while forwarding is most effective when the CMR value is low. Thus, based on the user's CMR, the system can adopt a CMR threshold-based scheme such that if the user's CMR is lower than a threshold, then the system applies the forwarding scheme; otherwise, it applies the replication scheme. Applying different location management schemes based on per-user CMR values introduces undesirable high complexity in managing and maintaining location-related information stored in the system as different system support mechanisms must be applied to different users. In this paper, we quantitatively analyze a hybrid replication with forwarding scheme that can be uniformly applied to all users. The most striking feature of the hybrid scheme is that it can determine and apply the optimal number of replicas and forwarding chain length on a per-user basis to minimize the communication cost due to location management operations while still being able to use the same data structure and algorithm to execute location management operations in a uniform way for all users. We develop a stochastic Petri net model to help gather this information and show how the information obtained statically can be used efficiently by the system at runtime to determine the optimal number of replicas and forwarding chain length when given a user's profile. We show that the proposed hybrid scheme outperforms both pure replication and forwarding schemes, as well as the CMR threshold-based scheme under all CMR values.

Index Terms—Mobile computing, location management, personal communication services, replication, forwarding, Petri net, performance evaluation.

◆

1 INTRODUCTION

In a Personal Communication Services (PCS) network, a location management scheme must handle two user location operations efficiently: update and search. The former operation occurs when a mobile host (MH) moves to a new location; the latter operation occurs when there is a call for the MH and the network must find the MH to deliver the call. A well-known basic and simple scheme is to update the location of each MH at its home location register (HLR) whenever it moves to a new visitor location registration (VLR) area. This location management scheme exists in IS-41 [6] in the United States and GSM [11] in Europe, commonly known as the basic HLR/VLR two-tier scheme.

An important research issue for location management is minimizing the network signaling cost associated with location update and search operations under location management strategies as these operations need to be performed frequently. Search operations are related to how often a MH is called, i.e., the call arrival rate, while update operations are related to the MH's mobility rate. Thus, in general, the call to mobility ratio (CMR) parameter, defined as the ratio between a MH's call arrival rate to its mobility rate, captures the MH's call and mobility patterns. Current research on location management focuses on per-user-based algorithms in which location update and search procedures can be adjusted dynamically based on a user's call and mobility patterns [2], [15]. For example, when the frequency of incoming calls is higher than the mobile user's mobility, that is, when the CMR value is high, the caching/replication scheme [8], [13] is effective, while when CMR is low, the forwarding algorithm [4], [9], the paging and location update algorithm [1], and the local anchor algorithm [7] are effective. Thus, under the notion of per-user-based location management, the best algorithm among all can be selected for execution by the system based on the user's CMR value. In addition to lacking a comprehensive comparative study on existing algorithms to identify the best algorithm when given a CMR value in a system environment, this also introduces undesirable high complexity in managing and maintaining location-related information stored in the system as different algorithms may be applied to different users.

Motivated by providing a uniform algorithm that can be generally applied to all users with different CMR values without sacrificing the optimality of individual algorithms, we aim to investigate and analyze hybrid schemes that can

---

- *The authors are with the Computer Science Department, Virginia Tech, Northern Virginia Center, 7054 Haycock Road, Falls Church, VA 22043. Email: irchen@cs.vt.edu and bgu@vt.edu.*

Manuscript received 15 Mar. 2002; revised 14 Aug. 2002; accepted 2 Dec. 2002.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number 10-032002.

combine the benefits of two or more existing schemes. Specifically, in this paper, we develop and analyze a hybrid scheme that combines replication and forwarding techniques. Replication is known to be most effective in reducing user search and update costs when CMR is high because when the call arrival rate to the user is much higher than the mobility rate, the communication cost would be dominated by search operations and the search cost can be reduced by replicating the location of a frequently called MH at selective VLRs from which most calls to the mobile user originate, thus avoiding the cost of querying the HLR (the only copy when replication is not used) for the location of the called MH. Conversely, forwarding is known to be most effective when CMR is low because when the mobility rate is much higher than the call arrival rate, the communication cost would be dominated by location update operations and the update cost can be reduced by simply forming a forwarding chain of VLRs through which the location of the MH can be found from the HLR, thus avoiding updating the HLR upon every update operation. The HLR is updated only when the forwarding chain becomes too long (say after K moves). Our hybrid scheme takes advantages of the benefits of replication and forwarding techniques at high and low CMR values, respectively.

A key concept of the hybrid scheme design is that it will degenerate into the forwarding scheme when the CMR value is sufficiently low and, on the other hand, into the replication scheme when the CMR value is sufficiently high. Under the hybrid scheme, the system applies the optimal number of replicas and the optimal forwarding chain length in order to minimize the total signaling cost, when given the user profile as input characterizing the user's calling and moving patterns. A lookup table is built at static time and applied to all users at runtime to identify the optimal number of replicas and the optimal forwarding chain length to be used for each user. We develop a stochastic Petri net model to help build the table and analyze the time complexity to build this table in the paper. Because of the uniformity property associated with the hybrid scheme, the internal data structure used by the system to manipulate and maintain the location information for all users is the same, thus greatly easing the system maintenance task. Moreover, since typically the per-user profile is kept at the HLR, we can keep the lookup table at the HLR, so the runtime overhead of determining the optimal number of replicas and forwarding chain length by a MH's HLR involves only a table lookup operation which can be executed efficiently.

In addition to having the advantage of uniformity, the hybrid scheme is shown to perform better than either scheme under all CMR values, as well as a binary "CMR threshold-based" scheme that applies the replication technique when a mobile user's CMR is higher than a threshold and applies the forwarding technique otherwise. For this binary CMR threshold-based scheme, we identify the optimal threshold value and show the hybrid scheme outperforms the threshold-based scheme under optimal threshold values.

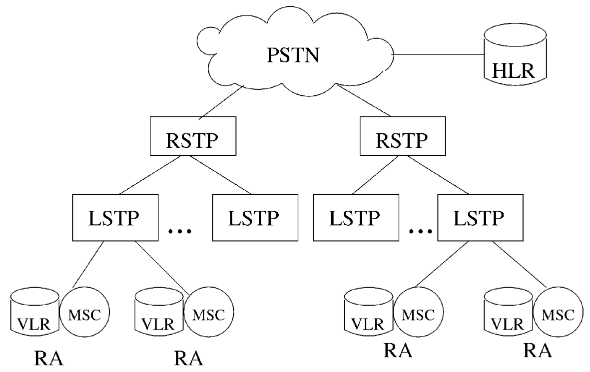

Fig. 1. PCS signaling network architecture.

The rest of the paper is organized as follows: Section 2 describes the system model and assumptions. Section 3 describes our hybrid scheme in detail. Section 4 develops a Petri net model to describe the behavior of a mobile user under the proposed hybrid scheme. Section 5 describes how to use the Petri net model developed to determine the optimal number of replicas and the maximum length of the forwarding chain so as to minimize the cost due to location update and search operations under the hybrid scheme. It also discusses data structures used to store this information at static time as well as look up operations to be performed at runtime to determine the optimal number of replicas and the maximum length of the forwarding chain at runtime, when given a user profile. Section 6 shows numerical data to support our claim that the proposed hybrid scheme performs better than both replication or forwarding schemes as well as the binary CMR threshold-based scheme across all CMR values. Finally, Section 7 summarizes the paper and outlines some future research areas.

2 SYSTEM DESCRIPTION

Fig. 1 shows a reference PCS signaling network architecture in the two-tier HLR-VLR structure as in IS-41 [6] and GSM [11]. A Registration Area (RA) can cover a single cell or a group of cells. A mobile switching center (MSC) is used to connect all the cells in one RA. A HLR is responsible for keeping track of a MH's current location as well as its profile. Each MH is permanently associated with a HLR. Conceptually, the HLR of a MH is at a higher level, while all VLRs that the MH wanders into from time to time are at the lower level. There may be some network switches connecting the HLR to VLRs in the mobile network. Each VLR or HLR is connected to the rest of the signaling network through a local signal transfer point (LSTP); one or more LSTPs belonging to one region may be connected to a regional signal transfer Point (RSTP). Separate RSTPs may be connected by a public switched telephone network (PSTN). We assume that each VLR corresponds to one RA. When a mobile user moves to a new RA, the mobile user sends the registration information to the new VLR which, in turn, can perform appropriate update actions, depending on the location management scheme used.

We assume the average communication cost between a VLR and the HLR is equal to the communication cost between any two randomly placed VLRs, represented by

T. The average communication cost between two neighboring VLRs is represented by τ. Clearly, τ is less than T and their values can be calculated by means of a network coverage model (e.g., hexagonal) as in [4]. The time that a particular MH stays in a VLR before moving to another one is characterized by an exponential distribution with an average rate of σ. Such a parameter can be estimated using the approach described in [9] on a per-user basis. The interarrival time between two consecutive calls to a particular MH is also assumed to be exponentially distributed with an average rate of λ. A MH is thus characterized by its CMR value, defined as λ/σ. Note that as the model developed in the paper is based on stochastic Petri nets, the assumption of exponentially distributed times can be relaxed by using Petri net tools that support specifications of general time distributions such as SPNP version 6 [14] and TimeNET version 3 [16].

3 LOCATION MANAGEMENT SCHEMES

In this section, we first give a background on location management schemes based on replication and forwarding. Then, we describe our proposed hybrid scheme that integrates forwarding and replication strategies into one uniform strategy that can be uniformly applied to MHs with vastly different CMR values. For completeness, we also describe the basic HLR-VLR scheme in IS-41 and GSM as a basis for comparison.

3.1 IS-41 Basic HLR/VLR

Under the basic IS-41 HLR/VLR scheme [6], a MH is permanently registered with a HLR. When the MH enters a new VLR area, it reports to the new VLR which in turn informs its HLR by means of a location update operation. The location update operation under IS-41 scheme proceeds as follows:

- When a MH moves into a new RA, it sends a location update message to the current base station which forwards this message to the current serving VLR.
- The current serving VLR forwards the message to MH's HLR.
- The HLR updates the location information of the MH and sends an acknowledgment message together with a copy of the MH's profile to the current serving VLR.
- The HLR sends a location cancellation message to the old serving VLR.
- The old VLR removes all entries belonging to the MH and sends an acknowledgment message to the HLR.

When a call is placed to connect to the MH, the PCS signaling network checks with the HLR of the MH to know the current VLR of the MH and a routing request is sent to the current VLR. A call delivery under IS-41 scheme proceeds as follows:

- The calling MH sends a call initiation message through its base station to its currently serving VLR.
- The VLR determines the associated HLR serving the called MH and sends a location request message to the HLR.
- The HLR determines the callee VLR and sends a route request message to this VLR/MSC.
- The callee VLR sends the route information to the HLR.
- The HLR forwards the route information to the calling VLR. Now, the calling VLR can set up a connection to the callee VLR via the SS7 signaling network using the usual call setup protocol.

3.2 Forwarding

A per-user forwarding strategy was proposed in [9] to reduce the total signaling network cost. It was shown that the forwarding strategy can result in 20-60 percent cost reduction when the user's CMR is low. In the forwarding strategy, when a user moves to a new RA, it updates the HLR only when the current forwarding chain length reaches a predefined constant K. Otherwise, a forwarding pointer is set up from the old VLR to the new VLR. The update procedure is described as follows:

- When a MH moves into a new RA, it sends a location update message to the current base station which then forwards this message to its associated VLR.
- If the current forwarding length is less than the maximum forwarding chain length K, then:
    - The new VLR deregisters the MH at the old VLR, but asks the old VLR to keep a pointer to point to the new VLR.
    - The old VLR sends an acknowledgment to the new VLR.
    - The forwarding length is increased by one.
- Otherwise, the IS-41 basic strategy is followed and the forwarding chain is reset, after which the HLR points to the new VLR directly.

When serving a call to a MH, the HLR is queried first to determine the first VLR at which the MH was registered, and then a chain of forwarding pointers is followed to reach the MH's current VLR. A call delivery under the forwarding strategy proceeds as follows:

- The first VLR is obtained from the callee's HLR as in the IS-41 scheme.
- The callee's current VLR is reached by following the forwarding chain.
- The callee's current MSC/VLR sends user route information to the HLR.
- The HLR forwards route information to the calling MSC. Now, the calling MSC can set up a connection to the callee MSC via the SS7 signaling network using the usual call setup protocol.

3.3 Replication

It has been observed that, even for the case when the number of possible communication areas is very large, the set of communication areas from which calls are made to a MH is relatively static and confined [12], [13]. For example, a study on e-mail traffic patterns [12] indicated that about 80 percent of mails are from three most frequently calling communication sites. If we place location replicas in these three sites, we can reduce the search cost by 80 percent. Another study on tracing actual calls over a six-month period in Stanford [13] showed that more than 70 percent of the calls made by callers in a week are to their top five callees. Also, 80 percent of the calls made by callers in one day are to their top three callees. If we store the top three callees' locations in the VLR of the caller, then 80 percent of the remote search cost can be eliminated. These results suggested that by using replication, a significant call latency and network cost can be eliminated when the user's CMR is high. Under the replication strategy, the location of a specific MH is replicated at selected sites. A location update operation due to mobility proceeds as follows:

- The MH sends a location update message to the serving MSC/VLR which forwards it to the MH's HLR as in IS-41.
- The HLR updates the location of the MH in its table. In addition, it also sends an update message to all VLRs where a replica of the MH's profile is stored. All other steps are the same as in IS-41.

A call delivery operation under the replication strategy proceeds as follows:

- The calling MH sends a call initiation message to its currently serving MSC/VLR through its base station.
- If the MSC/VLR stores a location replica, it contacts the callee's current MSC directly and gets the routing information. Otherwise the IS-41 procedure is followed.

3.4 Hybrid Replication with Forwarding Strategy

We first observe that the forwarding strategy is most beneficial when the user's CMR is below a threshold. The effect is especially pronounced when τ/T is small. On the other hand, the replication strategy attempts to exploit locality in calling patterns to reduce the call delivery cost, and is most beneficial when the user's CMR value is above a threshold.

We analyze a hybrid strategy that combines per-user replication and forwarding. The basic idea is that under a low CMR, the hybrid strategy attempts to reduce the total cost by replacing expensive HLR update operations with adjacent VLR forwarding pointer operations, thus behaving like the forwarding strategy. Under a modest CMR, the hybrid strategy combines the benefits from both forwarding and replication strategies. Finally, under a high CMR, the hybrid strategy behaves like the replication strategy by exploiting call locality to reduce the total signaling cost.

Under the hybrid scheme, the system maintains N replicas and a maximum forwarding chain length of K for each user. How many replicas to be used depends on a user's CMR value and its call arrival profile, e.g., if 50 percent of the total calls to the user originate from a single VLR (so $N = 1$), 70 percent of the total calls originate from 2 VLRs ($N = 2$), and 80 percent may come from three VLRs ($N = 3$), then the call arrival profile can be represented by a series of (N, P) values, e.g., (1, 50 percent), (2, 70 percent), and (3, 80 percent). This per-user call arrival profile information is kept at the MH's HLR as part of the user profile.

When given this per-user profile information and by applying the model and methodology described in Sections 4 and 5, the HLR determines the user's best (N, K) combination by performing a table lookup operation at runtime with the goal to minimize the overall cost associated with location search and update operations for the user. Each of the N replicas stores the identical content, i.e., a pointer to the first VLR of the forwarding chain (say, v_0). Note that a location update operation would not change this content stored in replicas unless the current forwarding chain length reaches K. Also, once the best (N, K) is determined for each MH, the HLR knows exactly which calling VLRs keep a replica by consulting the MH's call arrival profile. For example, if the best N value is 2, then only the two VLRs with the highest calling probability to the MH will each keep a replica; All other VLRs will not keep a replica. Recall that the call arrival profile regarding that VLRs have the highest calling probability to the MH is part of the user profile kept by the HLR. The HLR then is responsible for updating v_0 to all N replicas when a reset operation on the forwarding chain is performed at the HLR. At all time, all replicas only know v_0 and there is no need to do any kind of table lookup operations by replicas.

A location update under our hybrid strategy proceeds as follows:

- When a MH moves into a new visitor location area, it sends a location update message to the new VLR/MSC via its current base station.
- The new VLR/MSC examines the current forwarding chain length. If it reaches K, the HLR is updated to point to the new VLR as in IS-41, the forward chain is reset as in the forwarding scheme, and all N replicas are updated to store the location of the new VLR as in the replication scheme. This is the only condition under which replicas are being updated upon a location update operation since the first VLR of the forwarding chain has been changed as a result of the reset operation being executed when the maximum forwarding chain length K is reached.
- Otherwise:
  - The MH registers at the new VLR and the new VLR deregisters MH at the old VLR;
  - The old VLR sets a pointer to point to the new VLR, and then sends an ACK and MH's location profile to the new VLR;
  - The current forwarding chain length is increased by one.
  - The replicas are not changed.

When a call is placed to a MH, the caller's local VLR is searched first to see if it stores a location replica. If a replica is found (i.e., a replica hit), the caller will use the local information to get the first VLR and follow the forwarding chain to contact the callee's current VLR instead of contacting the callee's HLR. If a location replica is not found (i.e., a replica miss), the HLR is queried to determine the first VLR at which the callee was registered, and then the chain of forwarding pointers is followed to reach the current VLR.

A call delivery in the hybrid strategy proceeds as follows:

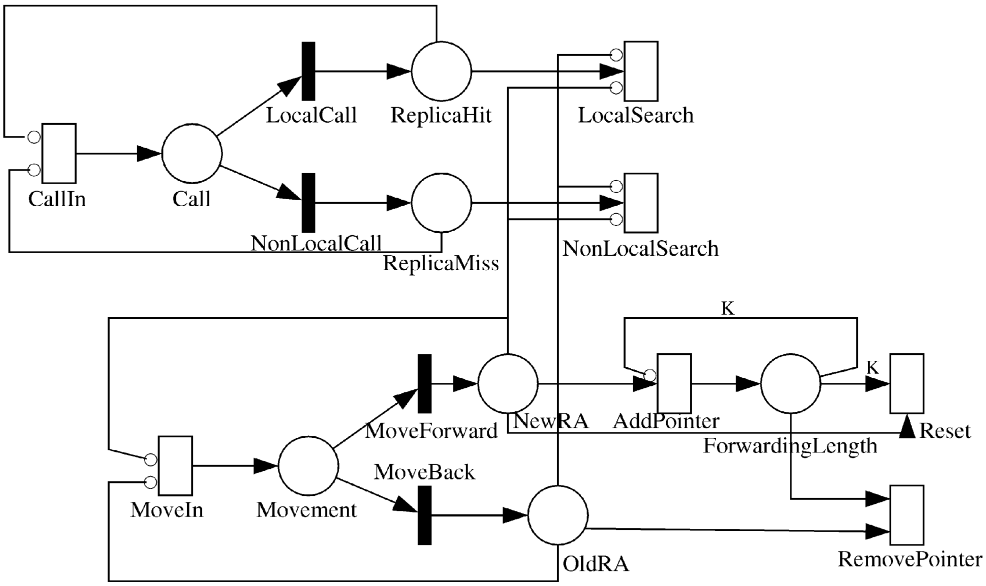

Fig. 2. SPN model for hybrid strategy.

- The calling MH sends a call initiation message to its current serving VLR/MSC through its base station.
- The VLR/MSC checks if it has a location replica of the callee. If a replica is found, the caller's VLR sends a routing request message directly to reach the first VLR in the forwarding chain. Otherwise, the callee's HLR is contacted which sends a routing request message to the first VLR.
- The query message reaches the callee's current serving VLR/MSC by following the forwarding chain.
- The callee's current serving VLR/MSC sends routing information to the calling VLR/MSC (if it is a replica hit) or its HLR (if it is a replica miss), which forwards it to the calling VLR/MSC. Now, the calling MSC can set up a connection to the called MSC via the SS7 signaling network using the usual call setup protocol.

4 MODEL

In this section, we develop a stochastic Petri net (SPN) model to study the performance of the hybrid strategy compared with both replication forwarding schemes, as well as the binary CMR threshold-based scheme, for users with different call and mobility patterns. We use the average cost of the PCS signaling network between two consecutive calls as the performance metric. That is, let C_{update} be the average cost of the signaling network in serving a location update operation and C_{call} be the average cost of the signaling network in locating a MH. Then, $C_{total} = C_{call} + C_{update} \times (\frac{\sigma}{\lambda})$, where σ/λ is the average number of update operations issued by the user between two consecutive location search operations. Note that C_{total} obtained above is a repeated measure on a cycle by cycle basis (with each cycle spanning the average time period between two consecutive calls to the MH), so the cumulative effect can be significant even for a small difference in C_{total} as we compare two location management algorithms.

The SPN model of the hybrid strategy is shown in Fig. 2 with the MH's incoming call pattern being modeled in the upper part and its mobility pattern being modeled in the lower part. These two parts interconnect with each other by several inhibitor arcs to model the fact that the system will perform location updates before location queries if they come simultaneously.

Table 1 gives the notation for model parameters used in the paper. Table 2 gives the meaning of places defined in the SPN model. Table 3 shows transition rates (for timed transitions) or probabilities (for immediate transitions) assigned to transitions defined in the SPN model.

The SPN model is constructed as follows:

- When a call arrives, a token is placed in place `Call`. The immediate transition `LocalCall` and `NonLocalCall` are enabled with the probabilities of P_h and P_m, respectively.
    - If a location replica exists (a replica hit with probability of P_h) in the caller's VLR, the token will flow through transition `LocalCall` to `ReplicaHit` which in turn will be serviced by a `LocalSearch` procedure in the hybrid strategy. The `LocalSearch` procedure will directly find the first VLR via the local replica and subsequently find the current VLR (where the MH currently resides) by following the forwarding chain whose length is indicated by the number of tokens stored in place `ForwardingLength`, i.e., $mark$(`ForwardingLength`), where $mark$(`P`) stands for the number of tokens in place `P`. This is modeled by making the execution rate of `LocalSearch` marking-dependent based on

$$mark(\texttt{ForwardingLength}),$$

TABLE 1
Notation Used in the Paper

Symbol	Meaning
λ	arrival rate of calls to a particular MH.
σ	mobility rate of a particular MH.
CMR	call-to-mobility ratio of a particular MH, i.e., λ/σ.
P_f	probability that a particular MH moves to a new VLR.
P_b	probability that a particular MH returns to the last-visited VLR; $P_b = 1 - P_f$.
P_h	probability that a call is serviced by the local replica, i.e., the hit probability.
P_m	probability that a local replica is not found to service a call, i.e., the miss probability; $P_m = 1 - P_h$.
μ_p	execution rate to set up, delete, or travel a pointer between two adjacent VLRs.
$\mu_{h,i}$	execution rate to find the current VLR where the MH resides for a forwarding chain of length i when there is a replica hit.
$\mu_{m,i}$	execution rate to find the current VLR where the MH resides for a forwarding chain of length i when there is a replica miss.
K	maximum number of forwarding steps after which a reset operation is performed
N	number of replicas
$\mu_{r,N}$	execution rate to reset a forwarding chain of length K so that after the reset operation is performed, all N replicas and the HLR are updated to point to the new VLR, and the old forwarding chain is released.
C_{update}	average cost of the signaling network in serving a location update operation.
C_{call}	average cost of the signaling network in locating a MH.
C_{total}	the average cost of the signaling network in serving location update and MH locating between two consecutive calls, i.e. $C_{call} + C_{update}/CMR$.

TABLE 2
Meaning of Places

Place	Meaning
`Call`	$mark(\texttt{Call})$=1 indicates that a call has just arrived.
`ReplicaHit`	$mark(\texttt{ReplicaHit})$=1 indicates that the call can be serviced by the local replica.
`ReplicaMiss`	$mark(\texttt{ReplicaMiss})$=1 indicates that the call cannot be serviced locally because a replica does not exist.
`Movement`	when $mark(\texttt{Movement})$=1, it indicates that the MH just makes a move across the VLR boundary.
`NewRA`	when $mark(\texttt{NewRA})$=1, it indicates a forward movement has just been made.
`OldRA`	when $mark(\texttt{OldRA})$=1, it indicates a backward (return to the previously visited RA) movement has just been made.
`ForwardingLength`	$mark(\texttt{ForwardingLength})$ indicates the current length of the forwarding chain.

that is, $rate(\texttt{LocalSearch}) = \mu_{h,i}$, where

$$i = mark(\texttt{ForwardingLength}).$$

- If a location replica does not exist (a replica miss with probability P_m) in the caller's VLR, the token will flow through transition `NonLocalCall` to `ReplicaMiss` which, in turn will be serviced by a `NonLocalSearch` procedure in the hybrid strategy. The `NonLocalSearch` procedure will go to the HLR to locate the first VLR and, subsequently, find the current VLR (where the MH currently resides) by following the forwarding chain whose length is indicated by the number of tokens stored in place `ForwardingLength`, i.e.,

  $$mark(\texttt{ForwardingLength}).$$

  This is also modeled by making the execution rate of `NonLocalSearch` marking-dependent based on $mark(\texttt{ForwardingLength})$, that is, $rate(\texttt{NonLocalSearch}) = \mu_{m,i}$, where

  $$i = mark(\texttt{ForwardingLength}).$$

- When the MH moves across a RA boundary, a token is placed in place `Movement`.
  - If the MH moves to a new RA, the immediate transition `MoveForward` will consume the token, after which a token will be placed in `NewRA`. In this case, transitions `LocalSearch` and `NonLocalSearch` will subsequently be disabled. Either `AddPointer` or `Reset` is enabled based on the current length of the forwarding chain indicated by

TABLE 3
Transition Rates or Probabilities

Transitions	Rate or probability
`CallIn`	λ
`LocalCall`	P_h
`NonLocalCall`	P_m
`LocalSearch`	$\mu_{h,i}$
`NonLocalSearch`	$\mu_{m,i}$
`MoveIn`	σ
`MoveForward`	P_f
`MoveBack`	P_b
`AddPointer`	μ_p
`RemovePointer`	μ_p
`Reset`	$\mu_{r,N}$

$$mark(\texttt{ForwardingLength})$$

as follows:

a. If the current forwarding length is less than the predefined maximum forwarding length K, transition `AddPointer` is enabled. A forwarding pointer between the two adjacent VLRs is set up and a token is added to `ForwardingLength`.
b. If the current forwarding length is equal to K, transition `AddPointer` is disabled and transition `Reset` fires. A reset operation is performed which resets the forwarding chain and updates all replicas. Transition `Reset` consumes the token stored in `NewRA`, as well as the K tokens stored in `ForwardingLength`. After the reset operation, the number of token in `ForwardingLength` is zero, i.e.,

$$mark(\texttt{ForwardingLength}) = 0.$$

- If the MH moves back to the previous RA, the immediate transition `MoveBack` will consume the token, after which one token will be placed in `OldRA`. At this time, there should exist at least one token in `ForwardingLength`. Therefore, `RemovePointer` fires and consumes one token from `ForwardingLength` and one token from `OldRA`, thus reducing the current forwarding length by one. This models the fact that the forwarding chain length can be reduced by one, if the MH moves back to the last visited VLR.

The Petri net model shown in Fig. 2 will generate a semi-Markov model that contains a number of states with each state being represented by a 5-component tuple (`ReplicaHit`, `ReplicaMiss`, `NewRA`, `OldRA`, `ForwardingLength`). The `ReplicaHit` component will take on the value of either 1 or 0, with 1 meaning a call just arrives and there is a replica hit because of the existence of a local replica, and 0 meaning otherwise. The `ReplicaMiss` component will also take on the value of either 1 or 0, with 1 meaning a call just arrives and there is a replica miss because of the nonexistence of a local replica, and 0 meaning otherwise. Note that a value of 1 can appear in either of these two components, but not in both at the same time. When both components contain the value of 0, it means that there is no call arrival in the state. The `NewRA` component will take on the value of 1 or 0, with 1 meaning a movement crossing a VLR boundary just occurs and it is a forward movement, and 0 meaning otherwise. The `OldRA` component will take on the value of 1 or 0, with 1 meaning a movement crossing a VLR boundary just occurs and it is a backward movement, and 0 meaning otherwise. When both `NewRA` and `OldRA` contain the value of 0, it means that there is no movement. The `ForwardingLength` component will take the value in the range of 0 to K indicating the length of the forwarding length in the state. For example, (0, 1, 0, 0, 3) represents a state in which a call just arrives with a replica miss and the forwarding length is 3. Thus, the system needs to go to the HLR to find the first VLR of the forwarding chain and then follows the forwarding chain of length 3 to find the current serving VLR of the MH to deliver the call in that state. It should be noted that the number of replicas N, replica hit/miss probability and length of the forwarding chain K will affect the transition rates of the Petri net model, thus affecting the probability that the system is found in a particular state in the steady state. Moreover, the value of K will affect the total number of states that exist in the underlying state model. For the Petri net model shown in Fig. 2, the total number of states is less than 200 for K in the range of 0 to 10 (maximum forwarding length). Using an evaluation tool such as SPNP [14] designed to solve thousands of states, we can solve the Petri net model very efficiently.

5 METHODOLOGY

In this section, we describe how to use the SPN model developed to evaluate the hybrid scheme proposed. Our goal is to determine the best K value (the maximum length of the forwarding chain) and the best N value (the replica number) to minimize the average cost of the PCS network between two consecutive calls, when given the profile of a MH.

5.1 C_{total} Calculation

Suppose that there are altogether N_s states in the underlying semi-Markov model of the Petri net. Let P_i be the steady state probability that the system is found in state i, as solved by SPNP. The average cost of the PCS signaling network in serving location update and call delivery operations between two consecutive calls can be obtained by assigning "cost" values to states of the system.

Let $C_{i,call}$ be the search cost assigned to state i given that a search operation is being serviced in state i. Then, the average search cost is given by

$$C_{call} = \sum_{i=1}^{N_s} P_i C_{i,call}. \qquad (1)$$

TABLE 4
Minimum Communication Cost Table

	0%	P=5%	...	P=100%
N=0	$C^{min}_{total}(0{,}0\%{,}K_{opt})$	-	-	-
N=1	-	$C^{min}_{total}(1{,}5\%{,}K_{opt})$	...	$C^{min}_{total}(1{,}100\%{,}K_{opt})$
...	-	...	...	...
N=10	-	$C^{min}_{total}(10{,}5\%{,}K_{opt})$	...	$C^{min}_{total}(10{,}100\%{,}K_{opt})$

Similarly, let $C_{i,update}$ be the update cost assigned to state i given a location update operation is being serviced in state i. Then, the average location update cost is given by

$$C_{update} = \sum_{i=1}^{N_s} P_i C_{i,update}. \tag{2}$$

Below, we use the syntax `if-then-else` and `condition?yes-value:no-value` as in programming language C to specify the cost assignments to states. For the search operation, if the calling VLR contains a location replica, then the local search cost applies; otherwise, the nonlocal search cost applies. Thus,

$$\begin{aligned} C_{i,call} = {} & \texttt{if } \mathit{enabled}(''\texttt{LocalSearch}'') \\ & \texttt{then } \frac{1}{\mathit{rate}(''\texttt{LocalSearch}'')} \\ & \texttt{else if } \mathit{enabled}(''\texttt{NonLocalSearch}'') \\ & \texttt{then } \frac{1}{\mathit{rate}(''\texttt{NonLocalSearch}'')} \\ & \texttt{else } \frac{P_h}{\mathit{rate}(''\texttt{LocalSearch}'')} + \\ & \frac{P_m}{\mathit{rate}(''\texttt{NonLocalSearch}'')}, \end{aligned}$$

where $\mathit{enabled}(''\texttt{T}'')$ means that transition `T` is enabled; $\mathit{rate}(''\texttt{T}'')$ stands for the rate at which transition `T` fires in the Petri net (see Table 3 for rates associated with transitions). When $\mathit{enabled}(''\texttt{T}'')$ is true, it means that the system is in a state in which the event associated with transition `T` is occurring. Thus, when $\mathit{enabled}(''\texttt{LocalSearch}'')$ is true, it means that the system is in a state in which a local replica exists and, consequently, the rate at which the system serves the call is $\mathit{rate}(''\texttt{LocalSearch}'')$. Conversely, when

$$\mathit{enabled}(''\texttt{LocalSearch}'')$$

is false, it means that the system is in a state in which a local replica does not exist and, consequently, the rate at which the system serves the call is $\mathit{rate}(''\texttt{NonLocalSearch}'')$. That is, the system needs to go to the HLR to find the location of the called user.

For the update operation, if the maximum length K is reached, then the cost of `Reset` applies; else, if the movement is a forward movement, then the cost of `AddPointer` applies; else, if the movement is a backward movement, then the cost of `RemovePointer` applies. Thus,

$$\begin{aligned} C_{i,update} = {} & \texttt{if } \mathit{enabled}(''\texttt{AddPointer}'') \\ & \texttt{then } \frac{1}{\mathit{rate}(''\texttt{AddPointer}'')} \\ & \texttt{else if } \mathit{enabled}(''\texttt{RemovePointer}'') \\ & \texttt{then } \frac{1}{\mathit{rate}(''\texttt{RemovePointer}'')} \\ & \texttt{else if } \mathit{enabled}(''\texttt{Reset}'') \\ & \texttt{then } \frac{1}{\mathit{rate}(''\texttt{Reset}'')} \\ & \texttt{else } \frac{P_f}{\mathit{mark}(''\texttt{ForwardingLength}'')=K?\mathit{rate}(''\texttt{Reset}''):\mathit{rate}(''\texttt{AddPointer}'')} \\ & + \frac{P_b}{\mathit{rate}(''\texttt{RemovePointer}'')}. \end{aligned}$$

Finally, we obtain C_{total} as follows:

$$C_{total} = C_{call} + C_{update}/CMR. \tag{3}$$

5.2 $C^{min}_{total}(N, P, K_{opt})$ Calculation

A MH's user profile is characterized by its CMR value and (N, P) value sets describing the call arrival profile to the MH. For example, if out of the total number of calls received by a MH, 50 percent comes from VLR 1, 70 percent comes from VLRs 1 and 2, and 80 percent comes from VLRs 1, 2, and 3, then it means that one replica (resided in VLR 1) can provide 50 percent local replica hit, two replicas (resided in VLRs 1 and 2) can provide 70 percent local replica hit, and three replicas (resided in VLRs 1, 2, and 3) can provide 80 percent local replica hit. Obviously, the local replica hit ratio will be 0 percent when the number of replicas N is zero under which the hybrid strategy becomes a pure forwarding strategy. In the scenario described above, the (N, P) value sets characterizing the call arrival profile would be (1, 50 percent), (2, 70 percent), and (3, 80 percent).

Here, we note that a different combination of (N, P, K) will result in a different C_{total} being calculated based on (3) because C_{call} and C_{update} depend on the state probabilities P_is calculated and P_is themselves in turn depend on the (N, P, K) combination considered. To get the optimal N and K, we first consider all possible combinations of replica hit ratio P and replica number N, e.g., with P in the range of [0 percent, 100 percent] in 5 percent increment, and N in the range of [0, 10] in an increment of 1. Of course, $P = 0\%$ when $N = 0$. Using the SPN model developed, we can statically obtain the minimum communication costs for possible combinations of (N, P) as shown in Table 4.

In Table 4, $C^{min}_{total}(N, P, K_{opt})$ represents the minimum communication cost under a specified CMR value. The "not applicable" cases are marked with "-". As mentioned,

TABLE 5
Finding $C_{total}^{min}(N, P, K_{opt})$

K	$C_{total}(N, P, K)$
0	$C_{total}(N, P, 0)$
1	$C_{total}(N, P, 1)$
...	...
n_K	$C_{total}(N, P, n_K)$

when $N = 0$, there is no replica and the hybrid strategy proposed degenerates into a pure forwarding strategy.

To obtain $C_{total}^{min}(N, P, K_{opt})$, we utilize the SPN model developed to find the optimal forwarding length K when given N and P. This is achieved by calculating C_{total} based on (3) under each possible K value in the range of $[0, n_K]$ as shown in Table 5 with n_K representing the maximum allowable forwarding chain length by the system for all MHs.

We then obtain $C_{total}^{min}(N, P, K_{opt})$ by taking the minimum C_{total} among all, i.e.,

$$C_{total}^{min}(N, P, K_{opt}) = \min_{K \in (0, n_K)} C_{total}(N, P, K).$$

When $K = 0$, the maximum forwarding length is zero and the hybrid strategy degenerates to the pure replication strategy. If both N and K are equal to zero, the hybrid strategy becomes the IS-41 HLR/VLR basic scheme.

As an example, consider $(N, P) = (1, 50\%)$ under $CMR = 1.0$. Table 6 shows $C_{total}(1,\ 50\%, K)$ values at different forwarding chain length Ks. We see in this case, the optimal K value is equal to 1 since $C_{total}^{min}(1, 50\%, K_{opt}) = C_{total}(1, 50\%, 1) = 1.32$ is the lowest among all.

Let the number of possible CMR values be $n_{CMR} = 160$ in a typical range 0.1 to 16 in an increment of 0.1. Let the number of possible P values be $n_P = 21$ in the range of 0 to 1 in an increment of 0.05 (as in Table 4). Also, let the number of possible N values be $n_N = 11$ in the range of 0 to 10 in an increment of 1 (as in Table 4) and let the maximum number of possible K values to be examined be $n_K = 11$ in the range of 0 to 10 in order to determine the best $C_{total}^{min}(N, P, K_{opt})$ for each possible pair of (N, P). Then, as we need to build a table similar to Table 4 for each possible CMR value and each table entry requires the SPN model developed in the paper to be executed a maximum of n_K times, the total number of times the SPN model developed must run to build all tables is upper bounded by

$$n_{CMR} \times n_P \times n_N \times n_K = 406,560.$$

Using an Ultra-10 Sun machine, it takes about 10 hours of real time to build all tables. Note that again these tables are built only once at static time and can be applied to all users by performing table lookup operations by the HLR at runtime to determine the optimal K and N values for minimizing the cost associated with location update and search operations for individual users.

5.3 Using Minimum Communication Cost Table

For a given MH, we can obtain its CMR value and call patterns from its profile. Here, we present an example of how to use the minimum communication cost table obtained statically to determine the optimal K and N. Suppose that one MH has $CMR = 1$ and its incoming call pattern is as follows:

TABLE 6
Finding $C_{total}^{min}(1,\ 50\%, K_{opt})$

K	C_{total}(1,50%,K)
0	1.75
1	1.32
2	1.33
3	1.45
4	1.64
5	1.86
6	2.12
7	2.40

- $P = 0$ percent when $N = 0$ (trivial condition) and
- $P = 50$ percent when $N = 1$.

The MH in this case has only two possible (N, P) combinations, so we only need to compare $C_{total}^{min}(0, 0\%, K_{opt})$ with $C_{total}^{min}(1, 50\%, K_{opt})$ to determine this MH's optimal K and N. Follow our last scenario

$$C_{total}^{min}(1, 50\%, K_{opt}) = C_{total}(1, 50\%, 1) = 1.32.$$

Suppose that $C_{total}^{min}(0, 0\%, K_{opt}) = C(0, 0\%, 1) = 1.20$ and this information is also stored in the table. Now, we simply compare the values of $C_{total}^{min}(0, 0\%, K_{opt})$ and $C_{total}^{min}(1, 50\%, K_{opt})$. Since

$$C_{total}^{min}(0, 0\%, K_{opt}) < C_{total}^{min}(1, 50\%, K_{opt}),$$

the optimal K is equal to the K_{opt} in $C_{total}^{min}(0, 0\%, K_{opt})$ which is 1 and the optimal N is equal to 0 which means no replica at all. In other words, for this particular MH, it is better that we don't use any replica and instead just simply use forwarding with $K = 1$ to minimize the cost associated with location update and search operations.

6 ANALYSIS

In this section, we first discuss the parameterization process, i.e., how to estimate values for the parameters of the SPN model in Fig. 2. Then, we present the analysis results with physical interpretations given.

6.1 Parameterization

Let the average communication time (single trip) between the HLR and a VLR or between two random VLRs be T and the average communication time (single trip) between two neighboring VLRs be τ. These two parameters can be estimated by considering a network coverage (e.g., hexagonal) model characterizing the underlying wireless network [4]. As a case study, we consider $\tau/T = 0.3$. Also, consider the case that the call arrival rate λ is 1.4/hr/MH as in [10]. Therefore, the mobility rate $\sigma = 1.4/CMR$, where CMR is given from the MH's profile. Also, assume that we can obtain the local replica hit probability P_h from the MH's profile. The local replica miss probability $P_m = 1 - P_h$. For example, if we consider the combination $(N, P) = (3, 80\%)$, then $P_h = 0.8$.

The move forward probability P_f and move backward probability P_b are also network structure dependent and their values will be different depending on the network coverage model considered. Suppose that the network structure is again modeled by the hexagonal coverage model and that the MH moves randomly to one of its neighbors with equal probability, i.e., 1/6 for the hexagonal network coverage model. Then, we can calculate these two probabilities as:

$$P_f = \frac{5}{6}, \tag{4}$$

$$P_b = \frac{1}{6}. \tag{5}$$

The communication cost to set up, delete, or travel a pointer connection between two neighboring VLRs is:

$$\begin{aligned} C_p &= cost(src_{VLR} \rightarrow dest_{VLR}) + cost(dest_{VLR} \rightarrow src_{VLR}) \\ &= \tau + \tau \\ &= 2\tau. \end{aligned}$$

Therefore, the execution rate μ_p to set up or delete a pointer connection between two neighboring VLRs is:

$$\mu_p = \frac{1}{2\tau}. \tag{6}$$

The communication cost for a call delivery under a replica hit condition is:

$$\begin{aligned} C_h &= cost(caller_{VLR} \rightarrow first_{VLR}) \\ &+ cost(travel\_a\_pointer\_connection) \\ &\times current\_forwarding\_length \\ &+ cost(current_{VLR} \rightarrow caller_{VLR}) \\ &= T + 2i\tau + T \\ &= 2T + 2i\tau. \end{aligned}$$

Here, i is the current forwarding chain length corresponding to the number of tokens contained in place `ForwardingLength`, that is, $mark$(`ForwardingLength`). Therefore, the execution rate for a call delivery operation under a replica hit condition is:

$$\mu_{h,i} = \frac{1}{2T + 2i\tau}. \tag{7}$$

The communication cost for a call delivery operation under a replica miss condition is:

$$\begin{aligned} C_m &= cost(caller_{VLR} \rightarrow callee_{HLR}) \\ &+ cost(callee_{HLR} \rightarrow first_{VLR}) \\ &+ cost(travel\_a\_pointer\_connection) \\ &\times current\_forwarding\_length \\ &+ cost(current_{VLR} \rightarrow callee_{HLR}) \\ &+ cost(callee_{HLR} \rightarrow caller_{VLR}) \\ &= T + T + 2i\tau + T + T \\ &= 4T + 2i\tau. \end{aligned}$$

Here, i again is the current forwarding chain length corresponding to $mark$(`ForwardingLength`). Therefore, the execution rate for a call delivery operation under a replica miss condition is:

$$\mu_{m,i} = \frac{1}{4T + 2i\tau}. \tag{8}$$

The communication cost for a reset operation is:

$$\begin{aligned} C_r &= cost(current_{VLR} \rightarrow HLR) \\ &+ cost(HLR \rightarrow first_{VLR}) \\ &+ cost(travel\_a\_pointer\_connection) \\ &\times maximum\_forwarding\_length \\ &+ cost(last_{VLR} \rightarrow HLR) \\ &+ cost(HLR \rightarrow current_{VLR}) \\ &+ (cost(HLR \rightarrow all\_replicas) \\ &+ cost(all\_replicas \rightarrow HLR)) \\ &= T + T + 2K\tau + T + T + 2NT \\ &= 4T + 2K\tau + 2NT. \end{aligned}$$

Here, K is the maximum forwarding chain length and N is the number of replicas. Consequently, the execution rate for a reset operation is:

$$\mu_{r,N} = \frac{1}{4T + 2K\tau + 2NT}. \tag{9}$$

6.2 Example

In this section, we present a detailed analysis and numerical data obtained for a case study to demonstrate the effectiveness of our approach. The input is the instant value of CMR and a series of (N, P) value sets characterizing the call arrival profile of a MH. The output is the optimal values of N and K identified by our hybrid scheme. We show that not only can we easily determine the best number of replicas N and forwarding chain length K for minimizing C_{total}, but also the C_{total} value obtained is better than that obtained under either replication or forwarding, as well as that obtained under the CMR threshold-based scheme.

Consider a MH's call pattern given as follows:

- $P = 0\%$ when $N = 0$ (trivial condition),
- $P = 50\%$ when $N = 1$,
- $P = 70\%$ when $N = 2$, and
- $P = 80\%$ when $N = 3$.

We first use the SPN model developed to obtain the optimal forwarding length K under these four given (N, P) combinations. Specifically, we use SPNP [14] to solve the SPN model in Fig. 2 based on (3). Fig. 3 shows the case in which $(N, P) = (2, 70\%)$ with the K value ranging from 0 to 10 under different CMR ratios. We normalize the cost in the Y-coordinate with respect to T. Thus, a ratio of $\tau/T = 0.3$ means that τ is set to $0.3T$. The result shows that the optimal K is 8 when $CMR = 0.25$, 4 when $CMR = 1$ and 2 when $CMR = 4$.

Fig. 4 summarizes the effects of N and CMR on K. We first discuss the effect of CMR on K. As we can see from Fig. 4, when the CMR value becomes larger or, equivalently stated, when the mobile user is called more often than it crosses VLR boundaries, the system would incur a lower communication cost with a smaller K value. On the other

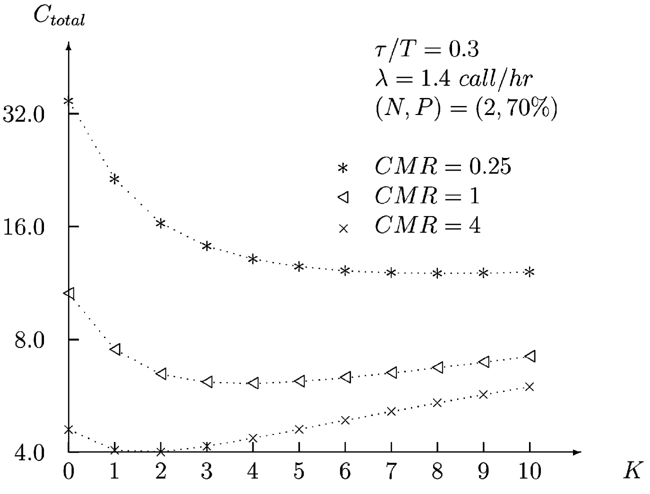

Fig. 3. Finding the optimal K under a constant number of replicas N.

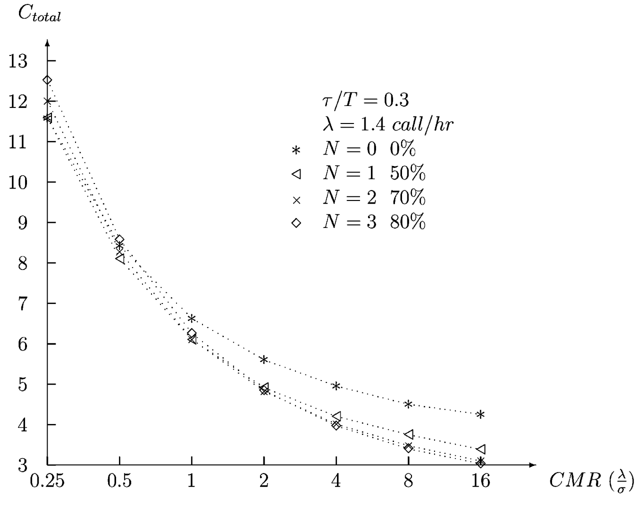

Fig. 5. Cost versus number of replica N at optimal K.

hand, with a smaller CMR value, the system would perform better with a larger K value. For example, with N fixed at 2, the optimal K is 0 (no forwarding at all) when $CMR = 16$ and then becomes 8 when $CMR = 0.25$. The interpretation of this result is clear: When CMR is low, the MH is not called very often relative to its mobility, so it is not judicious to minimize the call delivery cost by performing very costly resetting operations frequently.

Fig. 4 also shows the effect of N. Specifically, when N is large, it is better that K is a large value; otherwise, it is better that K is a small value. For example, with CMR fixed at 1 in Fig. 4, the optimal K is 4 when N is 3, and then drops to 2 when N is 0 (e.g., no replica at all). This result is attributed to the fact that the resetting operation cost depends on N. The larger the replica number N, the larger the resetting cost. In this case, the forwarding length tends to be large to avoid the high resetting cost.

Fig. 5 summarizes the total communication cost incurred due to location update and search operations under different replica numbers N at their respective optimal K values as determined from Fig. 4. From Fig. 5, we can easily obtain the optimal N and K values for the example MH profile. For example, when $CMR = 0.25$, the lowest communication cost can be obtained at $N = 0$ and $K = 5$; when $CMR = 16$, the optimal N increases to 3 while the optimal K decreases to 0. Fig. 5 also demonstrates the superiority of the hybrid strategy because it encomposes the advantages of both the forwarding and replication strategies. When the call arrival rate is low compared with the mobility rate (i.e., when CMR is low), it behaves like a pure forwarding strategy with $N = 0$. For example, when $CMR = 0.25$, $(N, P, K_{opt}) = (0, 0\%, 5)$ for this MH. When the call arrival rate is high compared with the mobility rate (i.e., when CMR is high), it behaves like a pure replication strategy with $K = 0$. For example, when $CMR = 16$, $(N, P, K_{opt}) = (3, 80\%, 0)$ for this MH.

To further demonstrate the performance gain compared with both pure forwarding and replication schemes. We consider a "threshold-based" scheme such that if the user's CMR is less than a threshold CMR value, then the pure forwarding scheme is applied to the user since it is known that the pure forwarding scheme performs excellent under low CMR values. On the other hand, if the user's CMR value is higher than or equal to the threshold CMR value, then the pure replication scheme is used since it is

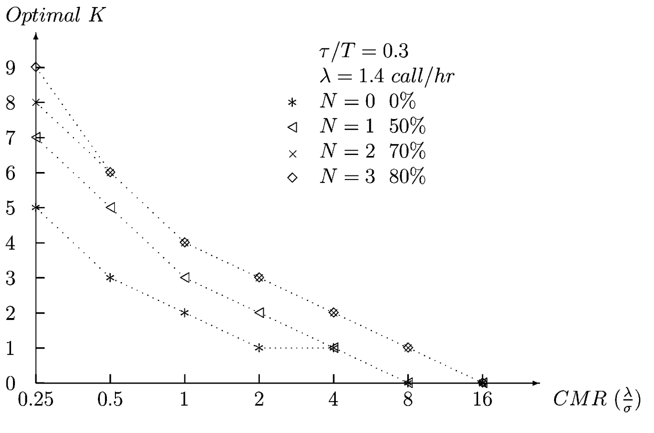

Fig. 4. Optimal K under different N and CMR values.

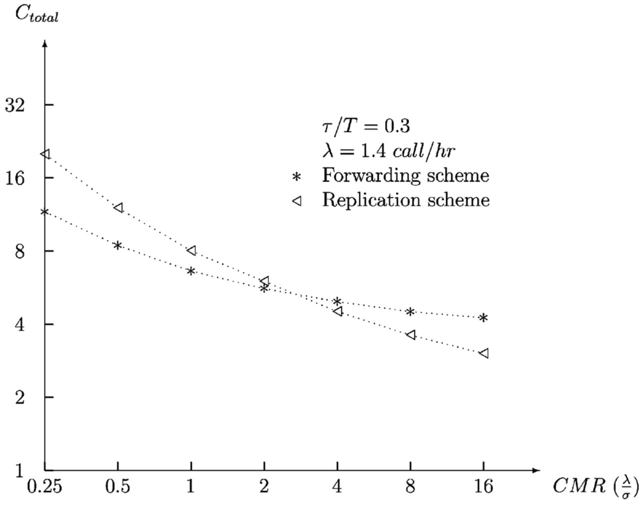

Fig. 6. Pure forwarding scheme versus pure replication scheme.

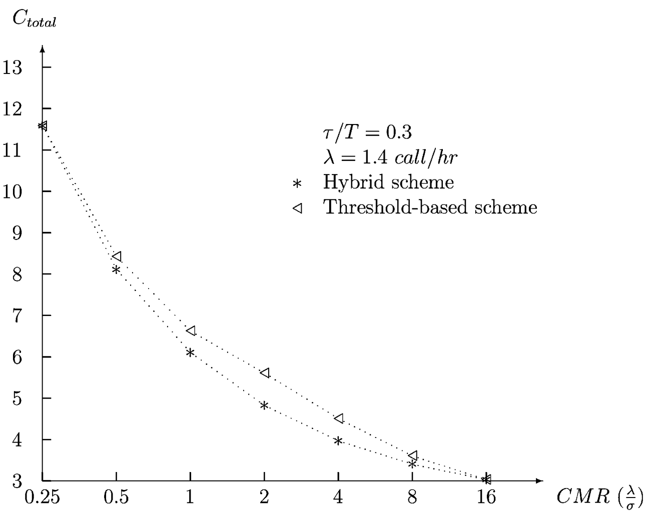

Fig. 7. Comparing hybrid versus threshold-based schemes.

known that the pure replication scheme performs excellent under high CMR values. Such a threshold-based scheme has the advantage of simplicity compared with the hybrid scheme at the expense of uniformity.

Fig. 6 identifies the cross-over threshold CMR value (at 2.8) below which the pure forwarding scheme (at $N = 0$ and the optimal K value) performs better than the pure replication scheme (at $K = 0$ and the optimal N value) and vice versa for the same example case considered. Such a threshold CMR value for the threshold-based scheme again can be determined statically using the methodology described earlier since both pure forwarding and replication schemes are encompassed by the more general hybrid scheme. Fig. 7 displays the cost difference between the cost obtained under the threshold-based scheme just described and that obtained under the hybrid replication with forwarding scheme. The curve for the threshold-based scheme in Fig. 7 is obtained by combining the lower-cost portions of the two curves in Fig. 6. We see that the hybrid scheme outperforms the threshold-based scheme over a wide range of CMR values because the hybrid scheme allows the best N and K values to be identified to minimize the total communication cost due to location management operations. The cost difference is close to zero when the user's CMR value is very small or very large because, in these cases, the hybrid scheme degenerates to a pure forwarding scheme and a pure replication scheme, respectively.

7 CONCLUSION

We have quantitatively analyzed a hybrid replication with forwarding location management scheme which can be uniformly applied to all MHs with different CMR values in wireless networks. The most striking feature of the hybrid scheme is that it can determine and apply the optimal N and K values on a per-user basis to minimize the communication cost due to location management operations while still being able to use the same data structure and algorithm to execute location management operations in a uniform way for all users. We developed a Petri net model to allow the calculation of $C^{min}_{total}(N, P, K_{opt})$ at different (N, P) combinations and showed how the results obtained at static time can be summarized into several minimum communication cost tables (with one table for each CMR value) to allow the best N and K to be determined at runtime, when given a MH's CMR and calling patterns. Such tables are generated statically without any knowledge of individual users' CMR values and are not changed until the network configuration changes. Furthermore, they can also be used to determine the best N and K values in a reactive manner for a MH having different CMR values in different time periods. The runtime overhead for determining the best N and K values by the HLR involves only a table lookup operation which can be executed efficiently. We demonstrated that the hybrid scheme performs not only better than either the replication or forwarding scheme, but also better than a binary CMR threshold-based scheme for all CMR values.

Some future research areas related to this paper include 1) investigating the possibility of combining other location management techniques (in addition to replication and forwarding) into another hybrid scheme that can encompass advantages of several schemes without compromising system complexity and 2) applying the methodology developed in the paper to hierarchical location management structures.

ACKNOWLEDGMENTS

This research is partially supported by a US National Science Foundation grant #9987586, a Microsoft Research grant, and an Intel grant.

REFERENCES

[1] I.F. Akyildiz, J.S.M. Ho, and Y.B. Lin, "Movement-Based Location Update and Selective Paging for PCS Network," *IEEE/ACM Trans. Networking,* vol. 4, pp. 629-638, 1996.

[2] I.F. Akyildiz, J. McNair, J.S.M. Ho, I. Uzunalioglu, and W. Wang, "Mobility Management in Next Generation Wireless Systems," *Proc. IEEE,* vol. 87, pp. 1347-1384, 1999.

[3] I.R. Chen and B. Gu, "A Comparative Cost Analysis of Degradable Location Management Algorithms in Wireless Networks," *The Computer J.,* vol. 45, no. 3, pp. 304-319, 2002.

[4] I.R. Chen, T.M. Chen, and C. Lee, "Performance Evaluation of Forwarding Strategies for Location Management in Mobile Networks," *The Computer J.,* vol. 41, no. 4, pp. 243-253, 1998.

[5] G. Cho and L.F. Marchall, "An Efficient Location and Routing Scheme for Mobile Computing Environments," *IEEE J. Selected Areas in Comm.,* vol. 13, no. 5, pp. 868-879, 1995.

[6] EIA/TIA, "Cellular Radio Telecommunication Inter System Operations,"Technical Report IS-41 (Revision B), 1991.

[7] J.S.M. Ho and I.F. Akyidliz, "Local Anchor Scheme for Reducing Signaling Costs in Personal Communications Networks," *IEEE/ACM Trans. Networking,* vol. 4, no. 5, pp. 709-725, 1996.

[8] R. Jain, Y.B. Lin, C. Lo, and S. Mohan, "A Caching Strategy to Reduce Network Impacts of PCS," *IEEE J. Selected Areas in Comm.,* vol. 12, no. 8, pp. 1434-1444, 1994.

[9] R. Jain, Y.B. Lin, C. Lo, and S. Mohan, "A Forwarding Strategy to Reduce Network Impacts of PCS," *Proc. 14th Ann. Joint Conf. IEEE Computer and Comm. Soc. (IEEE INFOCOM '95),* pp. 481-489, 1995.

[10] S. Mohan and R. Jain, "Two User Location Strategies for Personal Communications Services," *IEEE Personal Comm.,* vol. 1, no. 1, pp. 42-50, 1994.

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

[12] S. Rajagopalan and B.R. Badrinath, "An Adaptive Location Management Strategy for Mobile IP," *Proc. First Ann. Int'l Conf. Mobile Computing and Networking,* pp. 170-180, 1995.

[13] N. Shivakumar, J. Jannink, and J. Widom, "Per-User Profile Replication in Mobile Environments: Algorithms, Analysis and Simulation Results," *ACM-Baltzer J. Mobile Networks and Nomadic Applications (MONET),* vol. 2, no. 2, pp. 129-140, 1997.

[14] K.S. Trivedi, G. Ciardo, and J. Muppala, "User Manual," *SPNP Version 6,* Dept. of Electrical Eng., Duke Univ., Durham, N.C., 1999.

[15] V.W.S. Wong and V.C.M. Leung, "Location Management for Next-Generation Personal Communications Networks," *IEEE Network,* vol. 14, no. 5, pp. 18-24, 2000.

[16] A. Zimmermann, "User Manual 3. 0," *TimeNET: A Software Tool for the Performability Evaluation with Stochastic Petri Nets,* TU Berlin, 2001.

Ing-Ray Chen received the BS degree from the National Taiwan University, Taipei, Taiwan, and the MS and PhD degrees in computer science from the University of Houston, University Park, Houston, Texas. He is currently an associate professor in the Department of Computer Science at Virginia Tech. His research interests include mobile computing, pervasive computing, multimedia, distributed systems, real-time intelligent systems, and reliability and performance analysis. Dr. Chen has served on the program committees of numerous conferences, including being program chair of the 14th IEEE International Conference on Tools with Artificial Intelligence in 2002 and the Third IEEE Symposium on Application-Specific Systems and Software Engineering Technology in 2000. Dr. Chen currently serves as an associate editor for *IEEE Transactions on Knowledge and Data Engineering*, *The Computer Journal*, and *International Journal on Artificial Intelligence Tools.* He is a member of the IEEE Computer Society and the ACM.

Baoshan Gu received the BS degree from the University of Science and Technology of China, Hefei, China, in 1992 and the MS degree in computer science from The Institute of Computing Technology, Chinese Academia of Science, Beijing, China, in 1995. From 1995 to 2000, he was a research and development engineer in the Institute of Computing Technology, Chinese Academia of Science. He is currently pursuing his PhD degree in the Department of Computer Science, Virginia Tech, where he is a research assistant in the Systems and Software Engineering Laboratory. His research interests include next-generation wireless system architectures, design and evaluation of location and service management schemes in mobile computing environments, and mobile multimedia systems.