Efficient PCS Call Setup Protocols*

Yingwei Cui[1], Derek Lam[2], Jennifer Widom[3], Donald C. Cox[4]
Computer Science & Electrical Engineering Depts.
Stanford University, CA 94305

Abstract

Increasing demand for wireless mobile communications, coupled with limited network resources, has motivated investigation into alternative efficient mobility management solutions. In this paper, we consider wireless PCS *call setup protocols*. We propose a *Lightweight Location Lookup Protocol*, which significantly reduces the network signaling and setup delay during call setup. As an alternative, we extend the *Reverse Virtual Call Setup* technique of [6] and propose a *Reverse Connection Setup* protocol that also performs efficient call setup, while minimizing the cost of failed call attempts. The two new protocols have different advantages, in particular when combined with *location management techniques* including HLR/VLR, HLR/VLR with *partial replication*, and HLR/VLR with *caching*. We study combinations of call setup protocols with location management techniques and compare their performance. Both analytical and simulation results are presented.

[Key Words] Location Management Technique, Call Setup Protocol.

[Technique Areas] Personal Communications Services, Wireless Network Protocols, Mobile Communications Managements.

[1]Corresponding author. Gates 444, Stanford, CA 94305. Tel: 415/723-9429. cyw@cs.stanford.edu.

[2]Tel: 415/725-6563. Fax: 415/723-9251. dlam@wireless.stanford.edu.

[3]Tel: 415/723-7690. Fax: 415/725-2588. widom@cs.stanford.edu.

[4]Tel: 415/723-5443. Fax: 415/725-7126. dcox@wireless.stanford.edu.

*Research partially supported by the National Science Foundation under grant#NCR-9628184 and by Pacific Bell Corporation.

1 Introduction

Mobility management is a challenging problem in wireless Personal Communication Services (PCS) [2]. In a PCS system, mobile users are located in system-defined areas called *cells* that are grouped into *registration zones*. Every user connects with the *base station* in his cell through the wireless medium. Base stations are connected by a fixed wired network and exchange data to perform call setups and deliver calls between different cells. Mobility management allows the PCS system to track users' current locations for delivering incoming calls.

Solutions to the mobility management problem have two dimensions. One is the *Location Management Technique* (LMT) employed, which governs where user location information is stored, and how it is retrieved and updated. The other dimension is the *Call Setup Protocol* (CSP), which defines the steps taken during call setups, such as retrieving the user location information, notifying the user of incoming calls, and setting up the circuit connection through the wired and wireless network paths.

In first-generation wireless systems, such as the North American Advanced Mobile Phone Service (AMPS) [12], the system pages every cell within the network to locate a user for call setup. Because of the radio bandwidth inefficiency of this method, second-generation systems, such as the Global System for Mobile (GSM) communication and the North American Digital Cellular, employ a different approach as defined in the GSM [15] and IS-41 [22] standards. In the IS-41 standard, each user is assigned a permanent *home zone*. At any given time, the user location information is stored in a databases in the user's home zone (the *Home Location Register* or *HLR*), and also in a database in his *current zone* (the *Visitor Location Register* or *VLR*). We call LMTs relying on this database structure *HLR/VLR*. According to the call setup protocol stated in the IS-41 standard, during a call setup, the system first performs database lookups to find the user location information, called *location lookup*. It then sets up the circuit connection from caller to callee, and pages the callee in his cell.

While the IS-41 approach eliminates system-wide paging, which vastly reduces the radio link signaling, it introduces remote database lookups that may incur a large amount of wired network traffic and long call setup delay [13]. Much effort has been dedicated to developing LMTs to improve location lookup performance, as discussed in Section 1.1 below.

In this paper, we consider the entire PCS call setup process, and we propose alternative CSPs to improve the overall call setup performance. We believe that as PCS systems scale to very large numbers of users, network bandwidth will become a significant bottleneck. Thus, our goal is to minimize the extra network signaling and call setup delay introduced by the IS-41 style location lookups. At the same time, we can reduce the cost of failed call attempts (such as when the callee is not reachable). Our contributions are:

1. A complete specification of the call setup process using flow charts. We consider a variety of CSPs including the currently existing and newly proposed ones, combined with different HLR/VLR-based LMTs.

2. A new call setup protocol, called *Lightweight Location Lookup Protocol* (LiLLP), that significantly reduces the network signaling and setup delay associated with the PCS call setup process. The advantage of LiLLP is particularly clear when integrated with an HLR/VLR technique that uses *partial replication* [19].

3. The *Reverse Connection Setup* (ReCS) protocol, an extended version of the *Reverse Virtual Call Setup* (RVC) technique of [6]. ReCS performs efficient call setups and minimizes the cost of failed call attempts, without requiring the global paging techniques that are necessary for RVC.

4. Performance analysis and simulation work that explores the interactions between different CSPs and LMTs, and extracts the best CSP-LMT pair under what we believe to be realistic system conditions.

1.1 Related Work

Much effort has been focused on exploring efficient location management techniques (LMTs). Extensions to standard HLR/VLR techniques, such as HLR/VLR with *partial replication* [1, 14, 19], and HLR/VLR with *caching* [7], have been developed to improve wireless call setup performance. The *local anchoring* scheme [4] has been proposed to limit database update cost in HLR/VLR. The hierarchical database architecture [21] has been considered to achieve *location independent* or *lifelong* numbering. Extensions to the hierarchical scheme, such as HiPER [8] and HOPPER [10], also have been studied to improve location management system performance (in work performed by our group at Stanford).

The idea of *Reverse Virtual Call Setup* was first proposed in [6]. The protocol reduces PCS call setup delay and network cost for failed call attempts using an overlaid global paging network and a reverse connection setup strategy. In this paper, we devise new CSPs and combine existing HLR/VLR-based LMTs with them. Our goal is to explore the CSP-LMT combinations that achieve the best call setup performance, while eliminating the requirement of a global paging network.

1.2 Paper Outline

In Section 2, we describe the PCS environment and establish terminology. In Section 3, three CSPs are specified: the IS-41 standard, our Lightweight Location Lookup Protocol(LiLLP), and our Reverse Connection Setup (ReCS) protocol. We use event flow charts to specify these protocols combined with three LMTs: HLR/VLR, HLR/VLR with partial replication [19], and HLR/VLR with caching [7]. Section 4 gives an analytical comparison of average call setup performance for the nine CSP-LMT combinations. Simulation results are shown in Section 5. We conclude in Section 6.

2 PCS Environment and Terminology

In a wireless PCS system, mobile users are located in wireless cells grouped into registration zones, and they communicate with nearby base stations through the wireless medium. Base stations are connected to a wired network via *Mobile Switching Centers* (MSCs) to exchange data among different cells. When a call is made, the callee's *location information* indicating the user's current zone is retrieved. The callee's *routing information* also is needed, in order for the system to set up the circuit connection. Routing information includes the MSC port the callee is connected to (the *MSC address*) and his temporary user ID for contacting the MSC (the *Temp address*). Finally, *status information*, such as if the callee is busy or if the callee's handset is off, also may be used.

Location Management Techniques (LMTs) maintain user information in databases in order to perform efficient call setup. The current LMT standards are HLR/VLR schemes, as introduces in Section 1. In these schemes, each user's routing information is stored in VLR database at his current zone. In addition, an identifier for the user's VLR (the *VLR address*), which also indicates the user's current location, is stored in the HLR database at the user's home zone. Each user's HLR address is encoded in his phone number. Therefore, the system can always locate a user's HLR from the phone number, then locate his VLR, and finally retrieve the routing information. Entries in the HLR and VLR can be combined for users currently located in their home zone.

HLR/VLR with *partial replication* [1, 19] (hereafter called the Replication LMT) is one extension of the HLR/VLR standards. It replicates some users' VLR addresses in additional registration zones, so that callers from those zones can obtain the user's VLR address with one local lookup, thereby reducing the overall database lookups and network messages. The replicas are kept up-to-date by the system. HLR/VLR with *caching* [7] (hereafter called the Caching LMT) is another extension of HLR/VLR schemes. If it is discovered that a zone has many calls directed to a given user, the user's VLR address is cached in the local database. This also increases local lookup ratio and saves network messages before the cached entry becomes out-of-date.

Figure 1 shows the infrastructure of a PCS system that employs HLR/VLR-based techniques. The tables in Figure 1 indicate the user information databases, which combine the HLR, VLR, and cache or replication depository of a zone into one database. In the figure, we can see that user Y is visiting zone 3, and therefore his routing and status information is stored in the database there (the first row in HLRVLR-3). Zone 2 is Y's home zone, and stores Y's VLR address in its local database (the first row in HLRVLR-2). In addition, a copy of Y's VLR address is stored in zone 1's cache (the second row in HLRVLR-1). User X is currently located in his home zone, zone 1. His HLR entry and VLR entry are combined, and stored in zone 1 (the first row in HLRVLR-1).

Call Setup Protocols (CSPs) coordinate information exchange during the *call setup process*. The call setup process has two stages: the *location lookup* stage, in which the system finds the callee's current location (possibly with routing and status) information, and the *connection setup* stage that sets up the physical circuit connections between caller and callee. To optimize the call

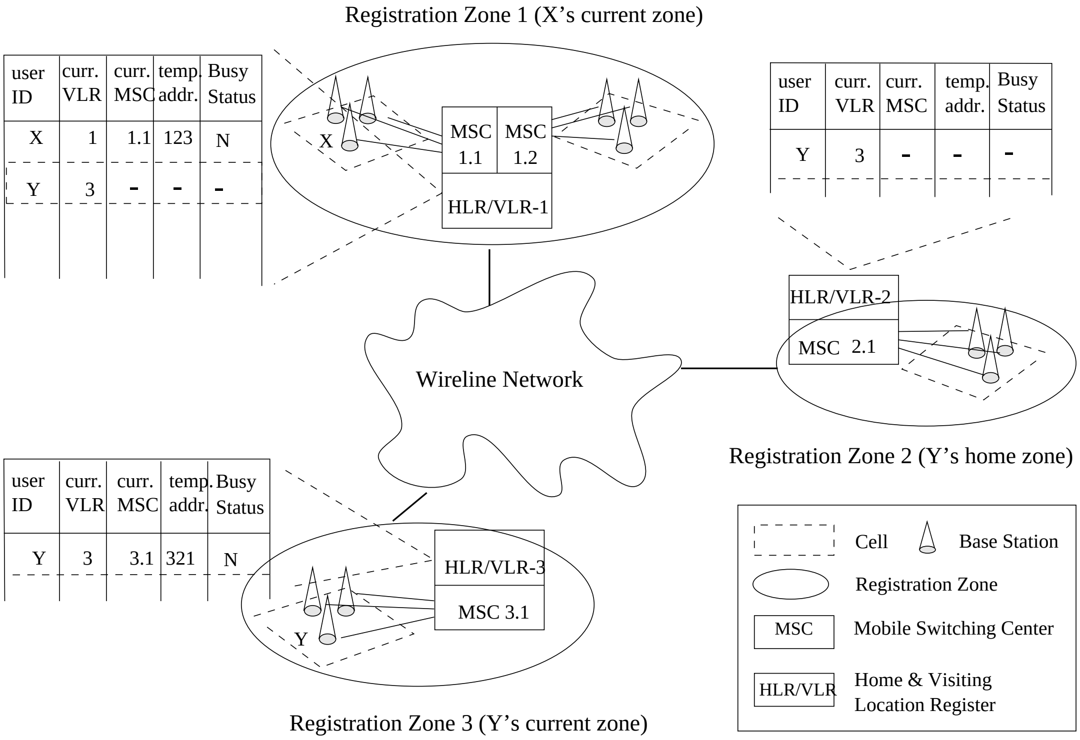

Figure 1: PCS Network Architecture

setup process, it is important to consider the two stages together, in order to utilize the location information most effectively. In current CSP standards, including that in IS-41, location lookup queries are sent from the caller's MSC to his VLR, then to the callee's HLR and then VLR, until the callee's routing information is retrieved. The connection is then set up from the caller to the callee accordingly.

In this paper, we develop new CSPs and explore how they interact with various HLR/VLR-based LMTs in the entire call setup process.

3 Call Setup Protocols

In this section, we describe two new CSPs: the *Lightweight Location Lookup Protocol* (LiLLP) and the *Reverse Connection Setup* (ReCS) protocol. We also study the IS-41 standard CSP for comparison. We combine each CSP with the three LMTs introduced in Section 2: HLR/VLR, Replication, and Caching. We specify each of the nine CSP-LMT combination schemes in a flow chart. The database and network performance of each combination is studied in later sections of the paper.

We specify our protocols based on the following general scenario, in which PCS user X is calling user Y in the system shown in Figure 1. We assume:

- X is currently located in zone 1.
- Y has home zone 2 and is currently located in zone 3.
- HLR and VLR are combined as database HLR/VLR-i in zone i, $i = 1..3$.
  In the rest of the paper, we refer the database at the user's home zone as HLRi, and the user's current zone's database as VLRi.
- The distance between zone 1 and zone 2 is l network hops.
  ($l = 0$ means X is in Y's home zone.)
- The distance between zone 1 and zone 3 is m network hops.
  ($m = 0$ means X and Y are in the same zone.)
- The distance between zone 2 and zone 3 is n network hops.
  ($n = 0$ means Y is at home.)

3.1 Call Setup Procedures

Several steps are taken during a call setup, and we define each step as a separate procedure. For each procedure, the number of database lookups, wired network messages, and network hops are counted to measure call setup performance. When a database is queried for user location information, it counts as one database lookup. Every remote database lookup, connection setup request, or acknowledgment contributes one network message. An inter-zone message may take multiple network hops to reach its destination. We do not include wireless signaling (e.g., paging) or local messages (e.g., messages between the local VLR and MSC) in our analysis.

Figure 2 lists the call setup procedures. Event counters formatted as [**database lookups, network messages, network hops**] are presented in the right column, to capture the cost of each step. Total network and database costs are calculated for each procedure. Different sets of procedures are used in different call setup protocols: we specify the IS-41's CSP in Section 3.2, LiLLP in Section 3.3, and ReCS in Section 3.4.

3.2 IS-41

In the IS-41 standard, when X calls Y, X initiates a local lookup at his own VLR (VLR1) for Y's information. If Y's information is not found, VLR1 sends query to Y's home HLR (HLR2) where Y's VLR address is always stored. HLR2 then retrieves Y's routing and status information from Y's current VLR (VLR3) and sends it back to VLR1. If Y is not busy, the connection request is sent from X's MSC to Y's MSC, and the network path is reserved. Y is then paged in his cell. The call setup is completed once X receives response from Y; otherwise the call setup fails. Figure 3 illustrates the call setup process in IS-41.

Location Lookup Procedures:

Procedure P1: X does local lookup in VLR1	total $\rightarrow [1,0,0]$
(a) X sends location lookup message to local VLR1.	$\rightarrow [0,0,0]$
(b) VLR1 looks up Y's location information.	$\rightarrow [0,0,0]$
(c) VLR1 replies to X.	$\rightarrow [0,0,0]$
Procedure P2: VLR1 queries VLR3	total $\rightarrow [1,2,2l]$
(a) VLR1 sends location lookup message to VLR3.	$\rightarrow [0,1,l]$
(b) VLR3 looks up Y's MSC and Temp address.	$\rightarrow [1,0,0]$
(c) VLR3 replies to VLR1.	$\rightarrow [0,1,l]$
Procedure P3: VLR1 queries HLR2	total $\rightarrow [1,2,2m]$
(a) VLR1 sends location lookup message to HLR2.	$\rightarrow [0,1,m]$
(b) HLR2 looks up Y's VLR address.	$\rightarrow [1,0,0]$
(c) HLR2 replies to VLR1.	$\rightarrow [0,1,m]$
Procedure P4: HLR2 queries VLR3	total $\rightarrow [1,2,2n]$
(a) HLR2 sends location lookup message to VLR3.	$\rightarrow [0,1,n]$
(b) VLR3 looks up Y's MSC and Temp address.	$\rightarrow [1,0,0]$
(c) VLR3 replies to HLR2.	$\rightarrow [0,1,n]$

Connection Setup Procedures:

Procedure P5 (P8): X connects to Y (X, Y in same zone).	total $\rightarrow [0,2,2l]([0,0,0])$
(a) X sends connection setup request to Y's MSC.	$\rightarrow [0,1,l]([0,0,0])$
(b) Y's MSC pages Y.	$\rightarrow [0,0,0]([0,0,0])$
(c) Y's MSC replies to X if Y is reachable.	$\rightarrow [0,1,l]([0,0,0])$
Procedure P6: X connects to VLR3	total $\rightarrow [1,2,2l]$
(a) X sends connection setup request to VLR3.	$\rightarrow [0,1,l]$
(b) VLR3 looks up Y's MSC and Temp address.	$\rightarrow [1,0,0]$
(c) VLR3 forwards connection setup message to Y's MSC.	$\rightarrow [0,0,0]$
(d) Y's MSC pages Y.	$\rightarrow [0,0,0]$
(e) Y's MSC replies to X if Y is reachable.	$\rightarrow [0,1,l]$
Procedure P7 (P9): Y connects to X (X, Y in same zone).	total $\rightarrow [0,2,2l]([0,0,0])$
(a) Y sends connection setup request to X's MSC.	$\rightarrow [0,1,l]([0,0,0])$
(b) X's MSC replies to Y.	$\rightarrow [0,1,l]([0,0,0])$

Figure 2:Call Setup Procedures

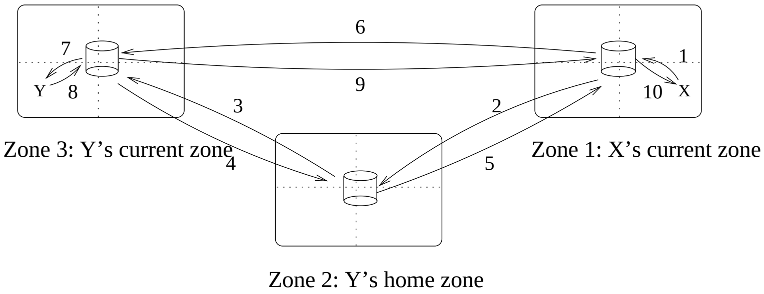

Figure 3: IS-41 Call Setup Process

We are interested in how the CSPs interact with different LMTs during call setup. Thus, we provide flow charts in Figures 9, 10, and 11 (in the Appendix) that specify call setup processes using IS-41's CSP combined with HLR/VLR, Replication, and Caching, respectively. The flowcharts present the entire call setup process, from the "start" of the location lookup stage to the "end" of the connection setup stage. Each oval in the charts indicates the invocation of a procedure from Figure 1. Both the procedure name and event counters for the procedure are noted to indicate the activities and cost of the step. The diamonds indicate conditional points that lead to different execution branches under different situations, with "[Y]" indicating the positive (yes) branch and "[N]" indicating the negative (no) branch. The terms on the edges, such as "q" and "$1 - q$", indicate the probability assumptions for our analysis and will be discussed further in Section 4.

HLR/VLR-based LMTs store only the user's current VLR address in his HLR, so that the HLR need not be updated when the user moves between different cells within one registration zone. However, with such a scheme, if the call setup protocol insists on obtaining full routing information before setting up the connection, then it must query both the callee's HLR and VLR, which may be in different zones, during the location lookup stage (as shown in Figure 3). This may lead to long setup delay and heavy network traffic. In our proposed protocols, we show that it is possible to reduce the location lookup stage, without significantly increasing the connection setup stage, thereby decreasing overall setup delay and network load.

3.3 LiLLP

In our *Lightweight Location Lookup Protocol* (LiLLP), the location lookup stage in the call setup from X to Y retrieves only Y's VLR address, instead of his entire routing information, thereby reducing the tasks of the location lookup stage. After obtaining Y's VLR address, the system sends a connection setup request to Y's VLR and reserves network paths. Upon receiving the connection setup request, Y's VLR forwards the request to Y's MSC, according to the routing information stored in the VLR. Y's MSC, which is co-located with the VLR in most current systems, pages Y in his cell. The connection setup is completed once Y responds. Figure 4 shows the LiLLP call setup process. As in IS-41, LiLLP can be combined with different LMTs.

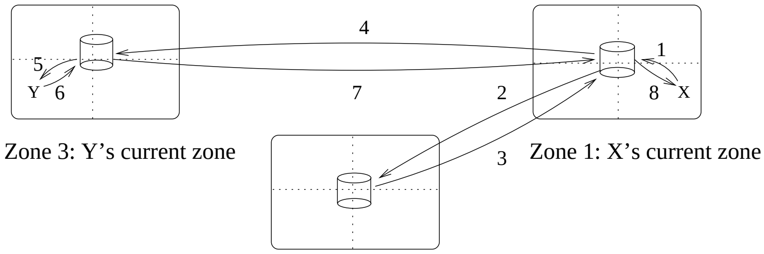

Figure 4: LiLLP Call Setup Process

Figures 12, 13, and 14 (in the Appendix) specify the entire call setup processes of LiLLP integrated with the three LMTs.

Effectively, LiLLP omits the remote database lookups for the detailed routing information during the location lookup stage (messages 3 and 4 in Figure 3) and instead obtains the information "along the way" during connection setup. As we will see, this approach can reduce overall network signaling and setup delay. The advantage of our lightweight scheme is particularly clear when integrated with the Replication LMT. We can see from Figure 4 that if X's VLR has Y's current VLR address replicated in its database, then X can obtain the address with one local lookup and directly set up the connection with Y's VLR. Thus, only two network messages are needed to complete the entire setup process.

Since LiLLP reserves network paths before retrieving Y's status information, LiLLP may waste network resources by reserving connections to unavailable users. The probability that a user is busy is likely to be small, but the probability that a user's handset is off may be significant in today's systems. In fact, the same problem occurs in IS-41, and is addressed by the *Reverse Connection Setup Protocol* proposed in the next section.

3.4 ReCS

The technique of *Reverse Virtual Call Setup* (RVC) was introduced in [6]. The authors proposed that during call setup, the caller's request should be sent to the callee's HLR, which invokes global paging of the callee from there. If the callee answers the page, connection setup is initiated. Otherwise, a message indicating the failure is sent back to the caller. This scheme reduces the cost of failed call attempts and call setup delay. However, it requires another overlaid global paging network and introduces extra wired or wireless signaling. We propose an alternative to RVC, in which the system pages the callee in his current visiting zone during the location lookup stage. We call this new protocol the *Reverse Connection Setup* (ReCS) protocol.

In the ReCS protocol, when a call is made from X to Y, the system first finds Y's current VLR address (VLR3) in X's local database or Y's HLR, and sends the location lookup request to VLR3. VLR3 then looks at its database for Y's routing information, and directly pages Y instead of sending the routing information back to X as in IS-41. The circuit connection setup is initiated from Y to X automatically (which is the reverse of the call request direction) after Y answers. Figure 5 shows the call setup process using ReCS. Figures 15, 16, and 17 (in the Appendix) specify ReCS combined with the three LMTs.

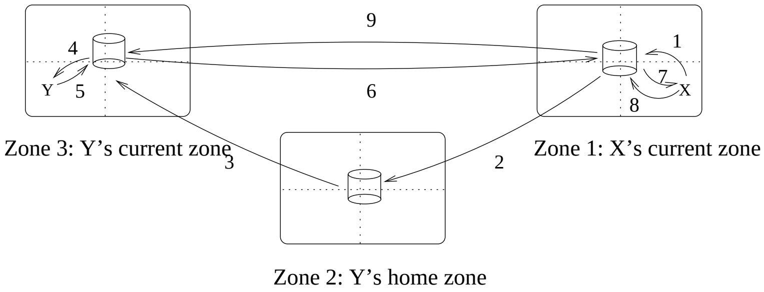

Figure 5: ReCS Call Setup Process

By doing reverse connection setup, messages that notify the caller of the callee's routing information (messages 4 and 5 in Figure 3) are eliminated. This leads to lower setup delay and signaling than the protocol employed in IS-41. Also, since the ReCS protocol pages the callee before setting up the connection, no network path is reserved if the callee 's handset is off[1]. Thus, the cost of failed call attempts is significantly reduced.

4 Analytical Comparison

In this section, we perform analytical comparisons between the three CSPs (our proposed LiLLP and ReCS, and the standard IS-41), when combined with the three different LMTs. We only consider average call setup (i.e. location lookup + connection setup) performance here. Empirical comparisons of overall system performance of database and network activities for the CSP-LMT schemes, including location lookups and updates, are shown by simulation results in Section 5.

Figure 6 summarizes the network message exchanges between the caller and callee HLR and VLRs during the call setup process for the three CSPs. We notice immediately that at least two messages are saved by using LiLLP or ReCS. Assuming each message takes the same time to reach its destination, LiLLP and ReCS can achieve much shorter setup delay. However, the scenario becomes more complicated with different traffic conditions, and when the CSPs are combined with different LMTs. Thus, we consider the probabilities of different situations as following:

- The probability that Y's handset is off be o.

[1]We are assuming here and in the rest of the paper that the user is unreachable only when his handset is off. However, there may be other reasons, such as the user is out of range, that can cause the unreachability of a user.

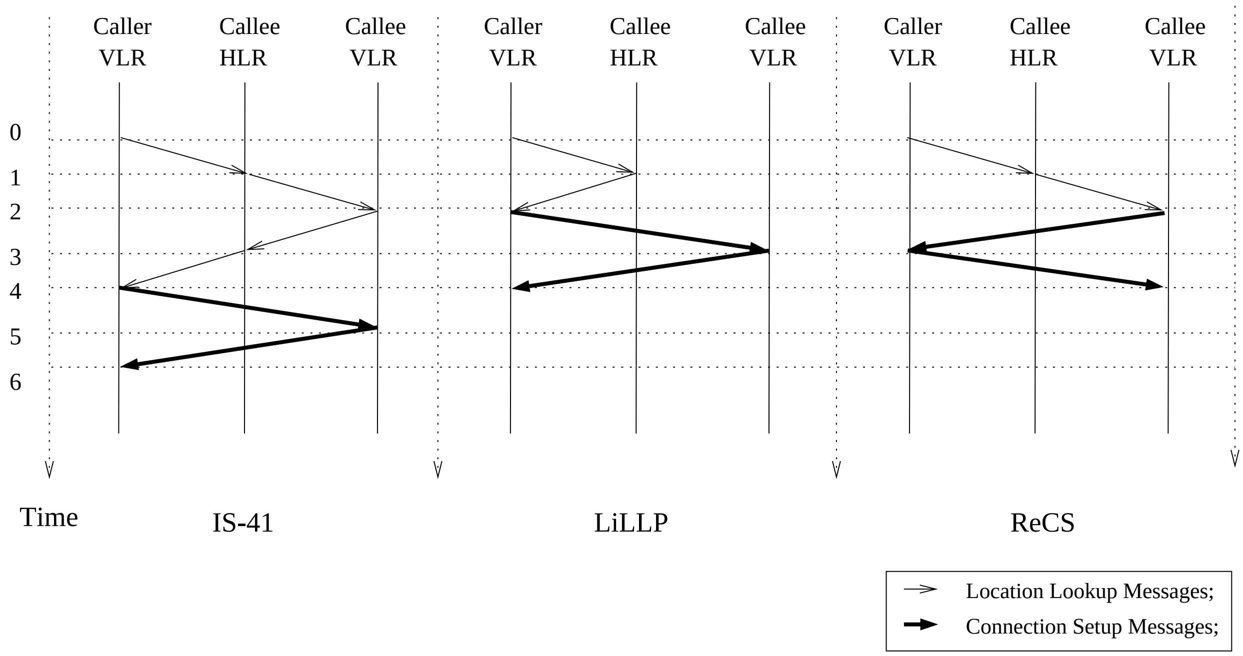

Figure 6: Call Setup Network Messages

- The probability that Y is busy be p.
- The probability that Y is in the same zone as X be q.
- The probability that Y is in his home zone be s.
- The probability that X is in Y's home zone be t.

In addition, for the Replication and Caching LMTs, we assume:

- The probability that X has a replica or cache of Y's VLR address is r.
- The probability of a cache hit is h.

These probabilities are indicated in the flowcharts of Figures 10–18. We calculate for each CSP-LMT combined scheme the average number of network messages and database lookups per call setup based on the flow charts, and present the results in Tables 1 and 2. The numbers are represented as formulas based on the probabilities specified above. They are indicative of the call setup delay, and partly decide the network and database loads associated with call setup. We use the following shorthands in the formulas:

$$\bar{h} = 1 - h,\ \bar{o} = 1 - o,\ \bar{p} = 1 - p,\ \bar{q} = 1 - q,\ \bar{r} = 1 - r,\ \bar{s} = 1 - s,\ \bar{t} = 1 - t,\ \overline{rh} = 1 - rh.$$

We see from Table 2 that LiLLP and ReCS incur the same number of database lookups as IS-41. We compare the network performance of LiLLP and ReCS with the IS-41 standard by presenting

	IS-41	LiLLP	ReCS
HLR/VLR	$2\bar{q}(1+\bar{s}\bar{t}+\bar{p})$	$2\bar{q}(1+\bar{s}\bar{t}+\bar{p}s\bar{t})$	$\bar{q}(1+\bar{s}\bar{t}+2\bar{o}\bar{p})$
Replication	$2\bar{q}(1+\bar{r}\bar{s}\bar{t}+\bar{p})$	$2\bar{q}(1+\bar{r}\bar{s}\bar{t}+\bar{p}\bar{r}s\bar{t})$	$\bar{q}(1+\bar{r}\bar{s}\bar{t}+2\bar{o}\bar{p})$
Caching	$2\bar{q}(1+\overline{rh}\bar{s}\bar{t}+\bar{h}r\bar{t}+\bar{p})$	$2\bar{q}(1+\overline{rh}\bar{s}\bar{t}+\bar{h}r\bar{t}+\bar{p}\overline{rh}s\bar{t}+h\bar{p}r\bar{t})$	$\bar{q}(1+\overline{rh}\bar{s}\bar{t}+\bar{h}r\bar{t}+2\bar{o}\bar{p})$

Table 1: Average Network Messages per Call Setup

	IS-41, LiLLP, or ReCS
HLR/VLR	$1+\bar{q}(1+\bar{s}\bar{t})$
Replication	$1+\bar{q}(1+\bar{r}\bar{s}\bar{t})$
Caching	$1+\bar{q}(1+\overline{rh}\bar{s}\bar{t}+\bar{h}r\bar{t})$

Table 2: Average Database Reads per Call Setup

the differences of the numbers of call setup network messages in Table 3. We can see that each formula in the table has a positive value (based on the fact that all the probabilities are ≤ 1), which means that LiLLP and ReCS incur a strictly smaller average number of network messages per call setup than IS-41, regardless which LMT is used.

	IS-41 − LiLLP	IS-41 − ReCS
HLR/VLR	$2\bar{p}\bar{q}(1-s\bar{t})$	$\bar{q}(1+\bar{s}\bar{t}+2o\bar{p})$
Replication	$2\bar{p}\bar{q}(1-\bar{r}s\bar{t})$	$\bar{q}(1+\bar{r}\bar{s}\bar{t}+2o\bar{p})$
Caching	$2\bar{p}\bar{q}(t+\overline{rh}\bar{s}\bar{t})$	$\bar{q}(1+\overline{rh}\bar{s}\bar{t}+r\bar{h}\bar{t}+2o\bar{p})$

Table 3: Performance Comparison Between CSPs

The relative performance of LiLLP and ReCS relies on specific probability assumptions. To further investigate the analytical formulas, let us make the following assumptions about the system conditions:

$$p=\tfrac{1}{8},\ q=\tfrac{1}{8},\ s=\tfrac{1}{2},\ t=\tfrac{3}{8},\ r=\tfrac{1}{2},\ h=\tfrac{7}{8}.$$

We learned from simulations that the caching (or replication) ratio $r=\frac{1}{2}$ can be achieved with fewer than five replicas for each user. For the probability that a user's handset is off, we consider two possible scenarios. One is the ideal $o=0$ in which handsets are always on. The other is $o=\frac{1}{2}$ proposed in [18], which is more realistic in today's systems due to power limitations. Tables 4 and 5 show the average numbers of network messages and database reads per call setup under the above default assumptions.

We have seen in Table 3 that the larger the probability o is, the more network messages ReCS saves. Table 4 further shows that when $o=\frac{1}{2}$, ReCS schemes achieve the best network

	IS-41	LiLLP	ReCS ($o = 0$)	ReCS ($o = \frac{1}{2}$)
HLR/VLR	3.82	2.78	2.68	1.91
Replication	3.55	2.26	2.54	1.78
Caching	3.66	2.81	2.44	1.83

Table 4: Average Network Messages Per Call Setup (under default assumptions)

	IS-41, LiLLP or ReCS
HLR/VLR	2.15
Replication	2.01
Caching	2.06

Table 5: Average Database Reads per Call Setup (under default assumptions)

performance, reducing up to 1/2 of the network messages in IS-41. However, when $o = 0$, the scheme combining LiLLP with the Replication LMT achieves better performance in the average number of network messages and database lookups. Since the Caching LMT always connects to the cached VLR address before connection setup to ensure that the cached address is correct, it incurs one extra remote VLR lookup for each call setup in LiLLP and loses its advantage. Therefore, using caching in LiLLP is not worthwhile.

The numbers in Tables 4 and 5 will certainly change as the probability assumptions vary. For example, the higher the replication ratio, the better the Replication scheme performs. However, update cost increases when more replicas are added. Neither LiLLP nor ReCS itself incurs extra update cost, but the Replication LMT may introduce extra update cost in order to keep the replicated VLR addresses up-to-date. The overall database and network performance is studied via simulations in the next section.

5 Simulations

We have studied the performance of the call setup protocols, including our proposed LiLLP and ReCS protocols, by simulations. As in the previous sections, we combine the call setup protocols with different location management techniques, and compare their performance with respect to database accesses and network activity. Our simulations take into account the update cost when users move across zones, which was not considered when analyzing the average call setup performance in Section 4.

In this section, we first introduce our *Pleiades* simulator [8, 9] used to perform our simulations. The topology models and network traffic models that we used to simulate the PCS environment are also briefly presented. Finally, simulation results comparing the schemes are shown.

5.1 Pleiades Simulator

We perform simulations using our PCS system simulator, *Pleiades* [17]. Pleiades is a highly configurable discrete event simulator developed at Stanford. The simulator models PCS location management and call setup processes. An input script file is used to describe the geographical and network topology models as well as different user call and mobility models. The statistics of the network and database activities are written to an output file for analysis.

Our simulation models are described in [9, 11] and summarized here. Our system topology is modeled after the ten largest cities of the United States. Each city is divided into a number of registration zones and wireless cells according to the city size. Each registration zone is approximately 400 square miles with about 20,000 subscribers. We used a hierarchical network topology, in which cells in each city are first connected with each other, then connected to other cities through higher-level network nodes. We have captured realistic traffic conditions by using time-varying traffic models. The call model includes various call traffic types, call volume changes throughout the day, and also incorporates callee distribution. We have instantiated our call models with parameters that were derived from call traffic data obtained from our local university telephone exchange [20]. The Metropolitan Mobility Model (METMOD) and the National Mobility Model (NATMOD) [9] are used for modeling movements within and between the cities. They are derived from transportation surveys [5], vehicle traffic data [16] and commercial airline data [3].

5.2 Results

In our simulations, we consider nine CSP-LMT combinations as discussed in Sections 3 and 4. For Replication and Caching, user i's VLR address is eligible to be replicated or cached in zone j only if the *local call to mobility ratio* ($LCMR_{ij}$) is higher than the optimal thresholds [7, 19]. A maximum of five VLR address replicas for each user are allowed in the Replication LMT. We also assume the ideal situation where users are always reachable by the system ($o = 0$) in our simulations.

We performed multi-day simulations and show results for a representative 24-hour period after start-up transients. Figures 7 and 8 depict the total network traffic and total database accesses per second versus simulation time. The peaks in the figures indicates the peak calling and movement activities during the simulated day.

Our graphs capture all database and network events, including location lookups and updates as well as connection setups, for all subscribers' calling and mobility activities. From Figure 7, we see that LiLLP integrated with Replication incurs the least network signaling, and is about 2/3 of the signaling in the IS-41 schemes. ReCS also achieves good performance. In Figure 8, we can see that LiLLP and ReCS don't incur extra database accesses, and the database performance achieved by using the three different LMTs are also similar. These results are consistent with our analysis in Section 4.

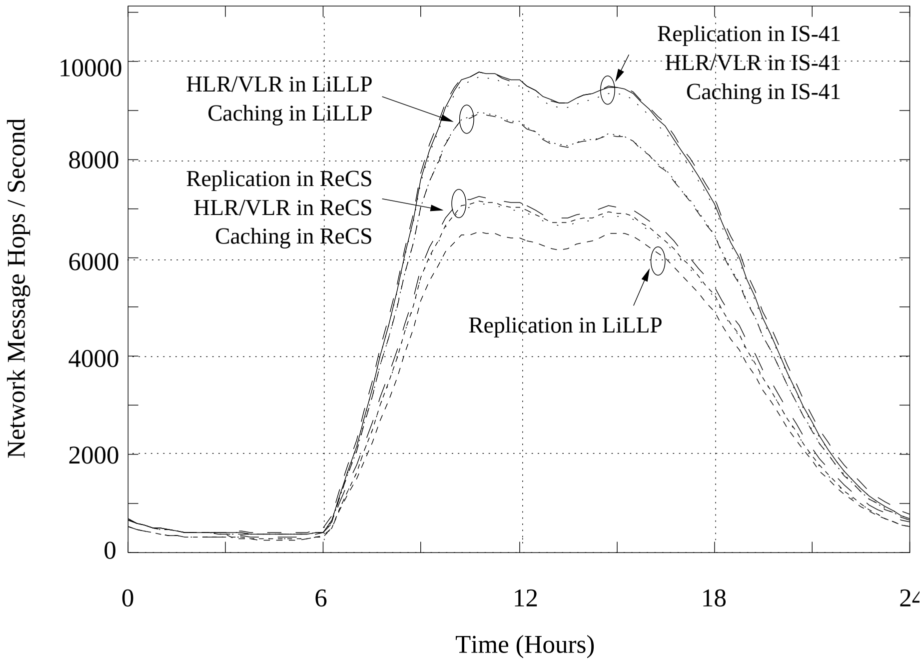

Figure 7: Total Network Signaling Hops

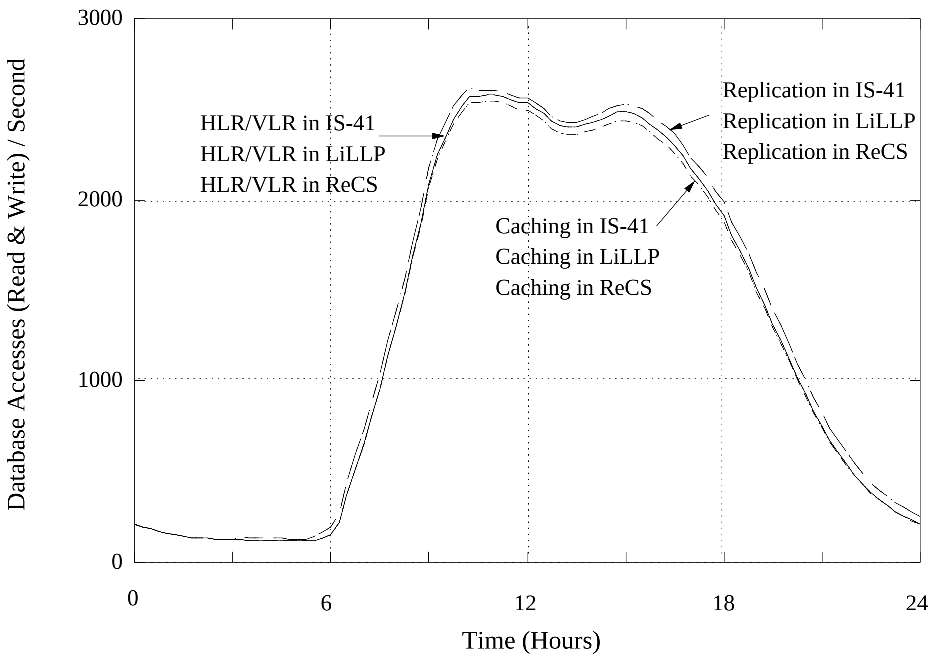

Figure 8: Total Database Accesses

6 Conclusions

We have proposed two new protocols for PCS call setup: the Lightweight Location Lookup Protocol (LiLLP) and Reverse Connection Setup (ReCS) protocol. We compared our protocols with the existing call setup protocol in the IS-41 standard. We carefully illustrated the interactions between the call setup protocols and different location management techniques using flow charts. Through analysis and simulations, we have shown that the newly proposed protocols can reduce the setup delay and decrease network signaling in the call setup process by up to 1/2, while no extra database accesses or update costs are required. We also discussed the best CSP-LMT schemes under what we believe to be realistic system conditions.

Acknowledgments

We are grateful to Bora Akyol, Jan Jannink, and Narayanan Shivakumar for numerous fruitful discussions.

References

[1] B. R. Badrinath and T. Imielinski. Replication and mobility. In *Second Workshop on the Management of Replicated Data*, pages 9–12. IEEE, November 1992.

[2] D. C. Cox. Wireless personal communications: What is it? *IEEE Personal Communications*, pages 20–35, April 1995.

[3] Origin and destination survey, t-100 domestic market data. Department of Transportation. Dataset for periods 10/91-12/91 and 10/93-12/93.

[4] J.S.M. Ho and I.F. Akyildiz. Local anchor scheme for reducing location tracking costs in PCN. In *1st ACM International Conference on Mobile Computing and Networking (MOBICOM'95)*, pages 170–180, Berkeley, California, 1995.

[5] P. S. Hu and J. Young. *1990 NPTS Databook: Nationwide Personal Transportation Survey.* Federal Highway Administration, November 1993.

[6] C.-L. I, G. P. Pollini, and R. D. Gitlin. PCS mobility management using the reverse virtual call setup algorithm. *IEEE/ACM Transactions on Networking*, 5(1):13–23, February 1997.

[7] 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*, 12(8), October 1994.

[8] J. Jannink, D. Lam, N. Shivakumar, J. Widom, and D. C. Cox. Efficient and flexible location management techniques for wireless communication systems. In *2nd ACM International*

Conference on Mobile Computing and Networking (MOBICOM'96), pages 38–49, Rye, New York, November 1996.

[9] D. Lam, D. C. Cox, and J. Widom. Teletraffic modeling for Personal Communications Services. *IEEE Communications Magazine Special Issue on Teletraffic Modeling, Engineering and Management in Wireless and Broadband Networks*, 35(2):79–87, February 1997.

[10] D. Lam, Y. Cui, D. Cox, and J. Widom. A location management technique to support lifelong numbering in Personal Communication Services. *IEEE Global Telecommunications Conference*, November 1997. To appear.

[11] D. Lam, J. Jannink, D. C. Cox, and J. Widom. Modeling location management for Personal Communication Services. In *Proceedings of 1996 IEEE International Conference on Universal Personal Communications (ICUPC96)*, pages 596–601, September 1996.

[12] W. C. Y. Lee. *Mobile Cellular Telecommunications Systems.* McGraw-Hill, 1989.

[13] K. K. Leung and Y. Levy. Data replication schemes for global network. In *15th Internation Teletraffic Congress (ITC15)*, pages 211–222, June 1997.

[14] K. K. Leung and Y. Levy. Global mobility management by replicated databases in Personal Communication Networks. *IEEE Journal On Selected Areas In Communications*, 1997.

[15] M. Mouly and M.-B. Pautet. *The GSM System for Mobile Communications.* Palaiseau, France, 1992.

[16] 1990 commute summary. Metropolitan Transportation Commission, Lotus 123 format spreadsheet, August 1990. S.F. Bay area transportation measurements.

[17] Peiades homepage: http://www-db.stanford.edu/pleiades/. Stanford University.

[18] G. P. Pollini, S. Meier-Hellstern, and D. J. Goodman. Signaling traffic volume generated by mobile and personal communications. *IEEE Communication Magazine*, pages 60–65, June 1995.

[19] N. Shivakumar and J. Widom. User profile replication for faster lookup in mobile environments. In *1st ACM International Conference on Mobile Computing and Networking (MOBICOM'95)*, pages 161–169, Berkeley, California, 1995.

[20] Telephone call traffic data set, 3/95–9/95. S. Phillips, personal communication, October 1995. Encrypted ID numbers.

[21] J. Z. Wang. A fully distributed location registration strategy for universal personal communication systems. *IEEE Journal on Selected Areas in Communications*, 11(6):850–860, August 1993.

[22] J. I. Yu. IS–41 for mobility management. In *First International Conference on Universal Personal Communications, (ICUPC92)*. IEEE, 1992.

A Flow charts

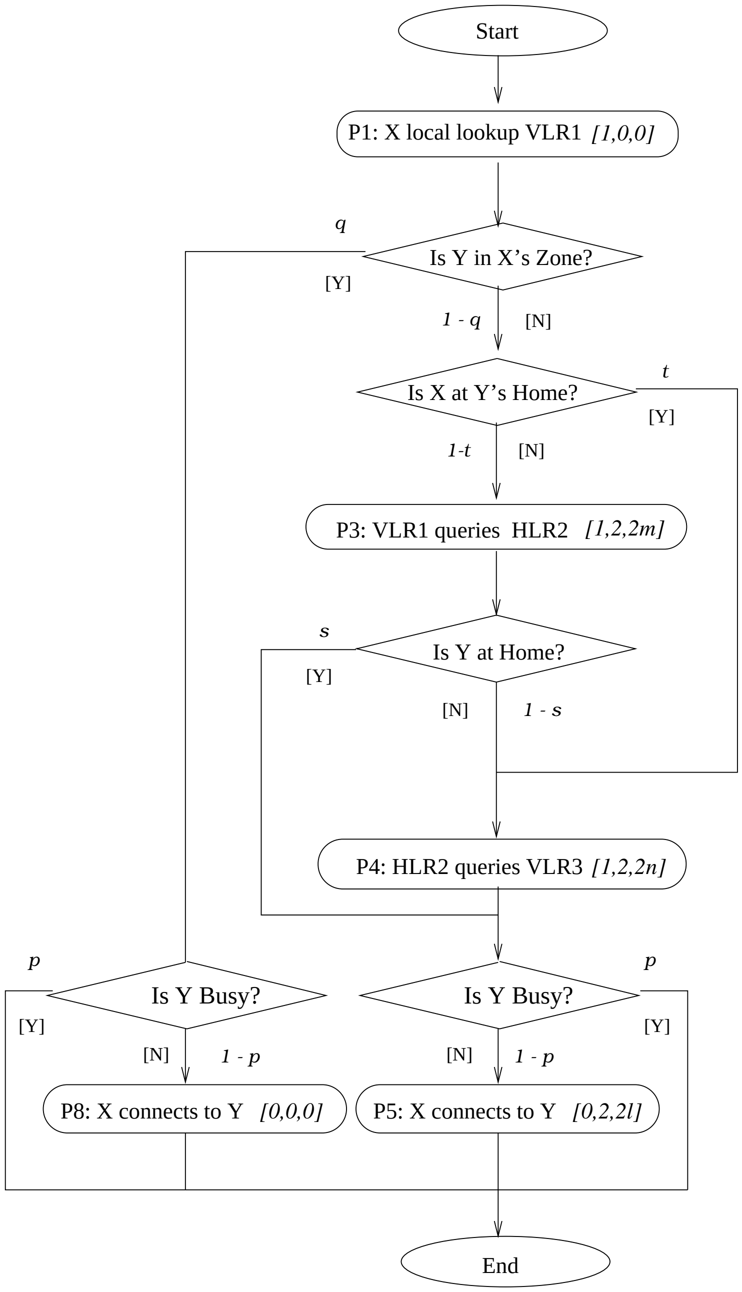

Figure 9: HLR/VLR in IS-41

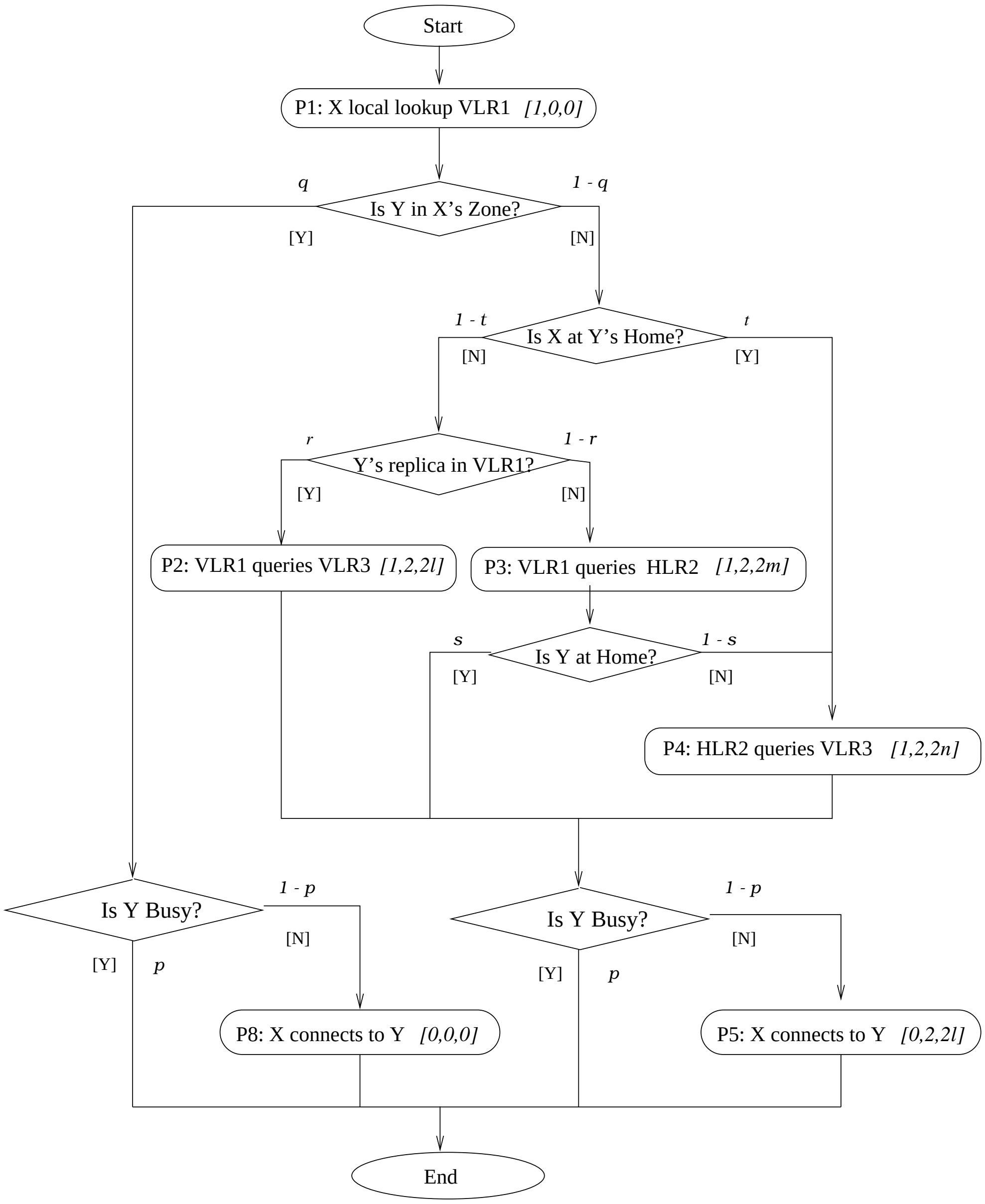

Figure 10: HLR/VLR with Replication in IS-41

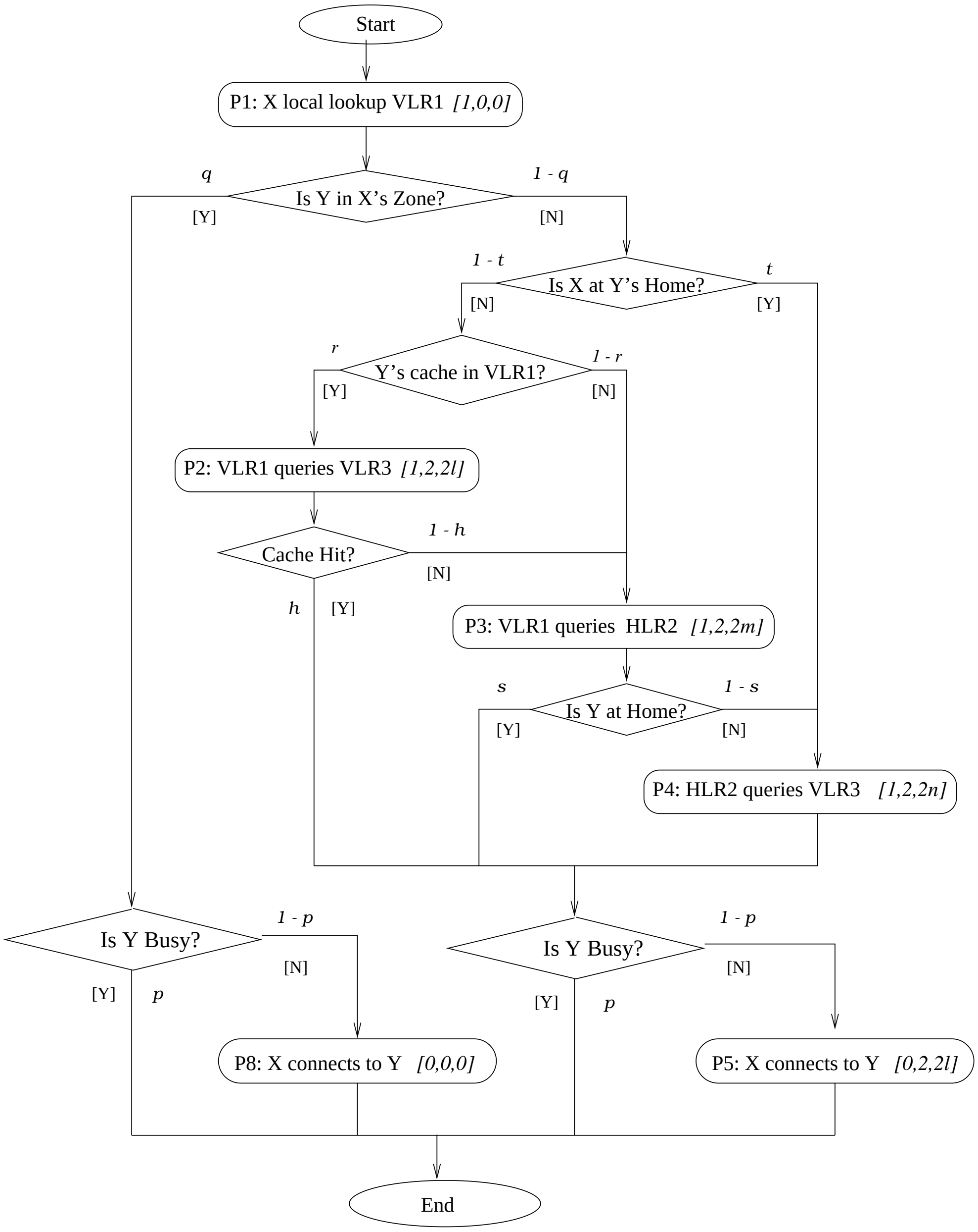

Figure 11: HLR/VLR with Caching in IS-41

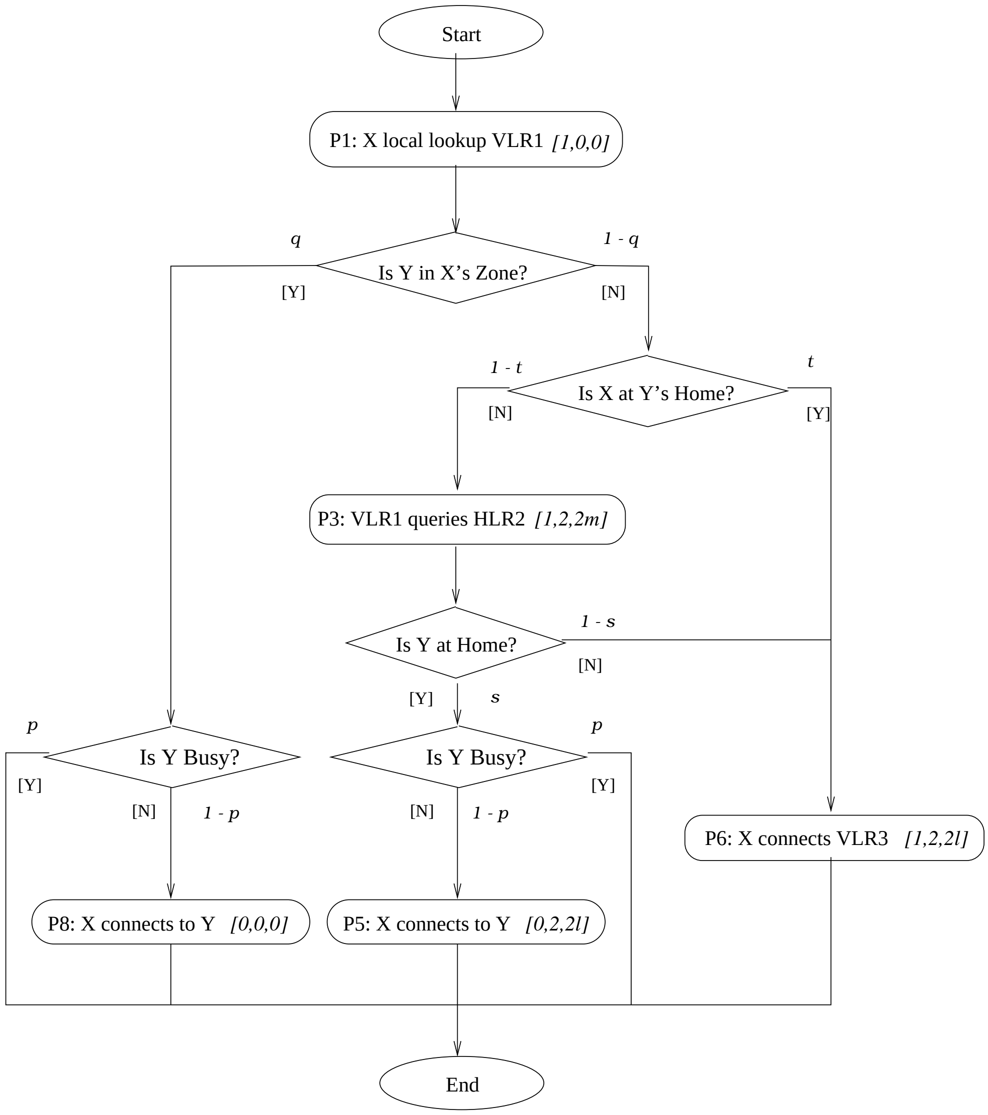

Figure 12: HLR/VLR in LiLLP

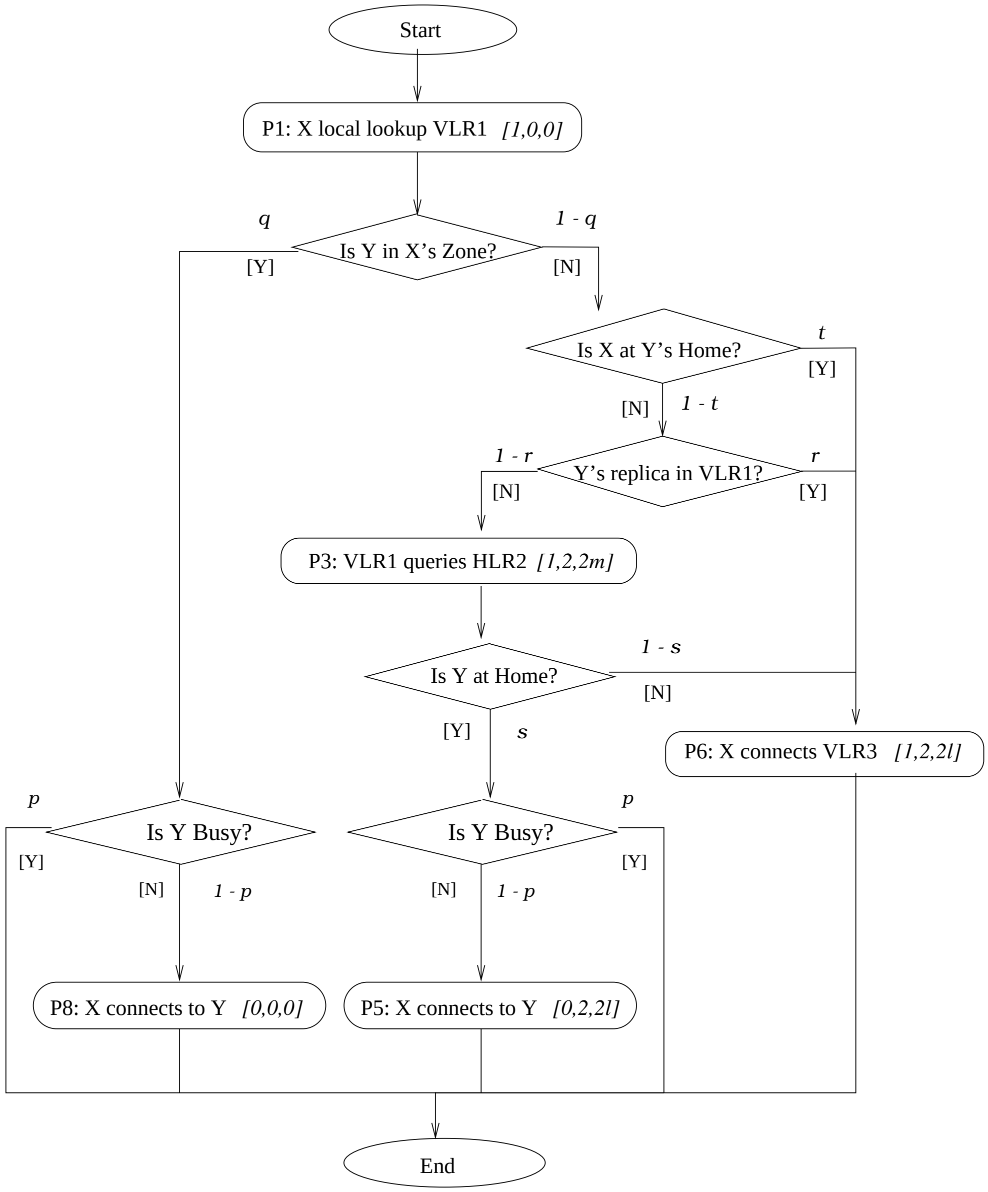

Figure 13: HLR/VLR with Replication in LiLLP

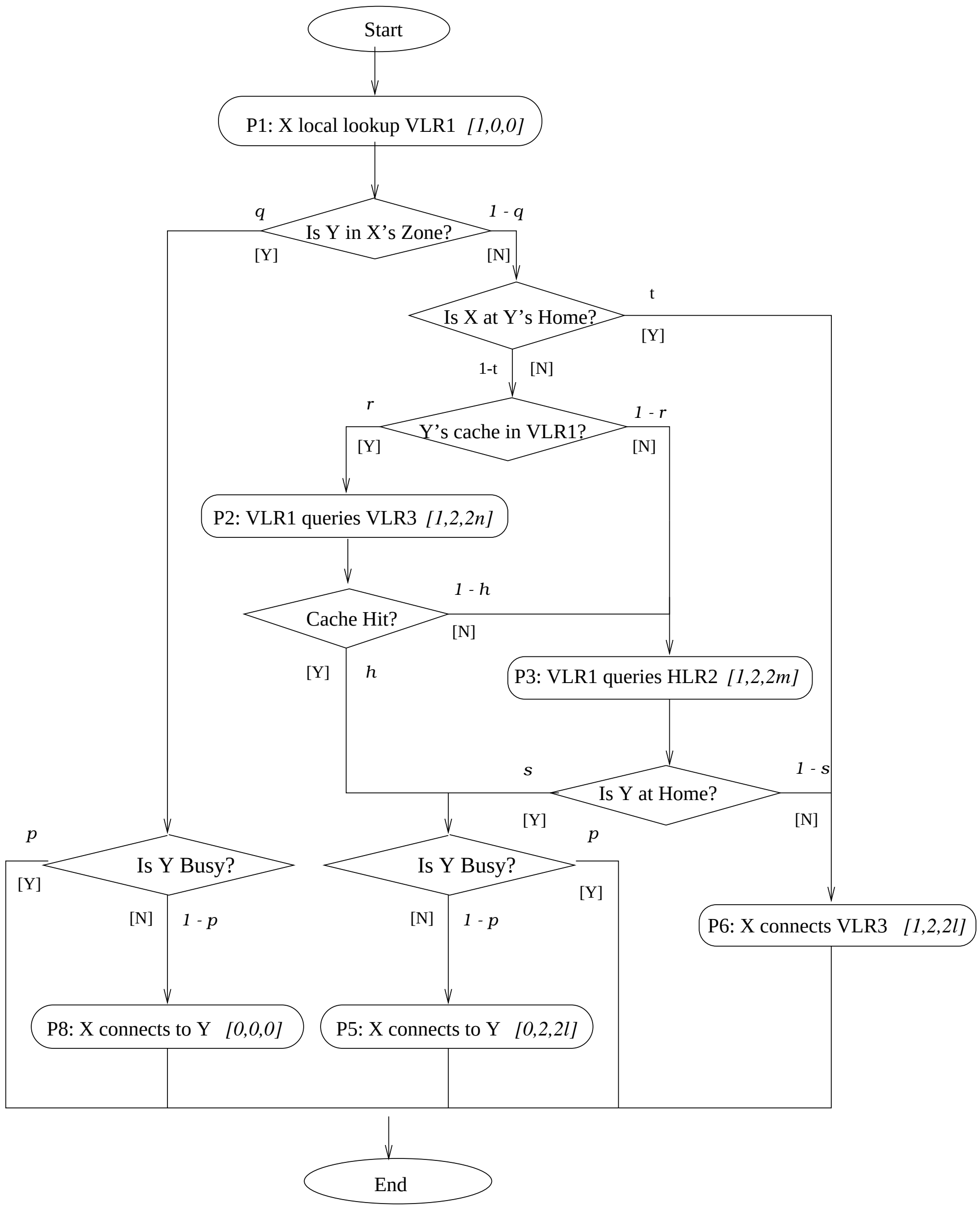

Figure 14: HLR/VLR with Caching in LiLLP

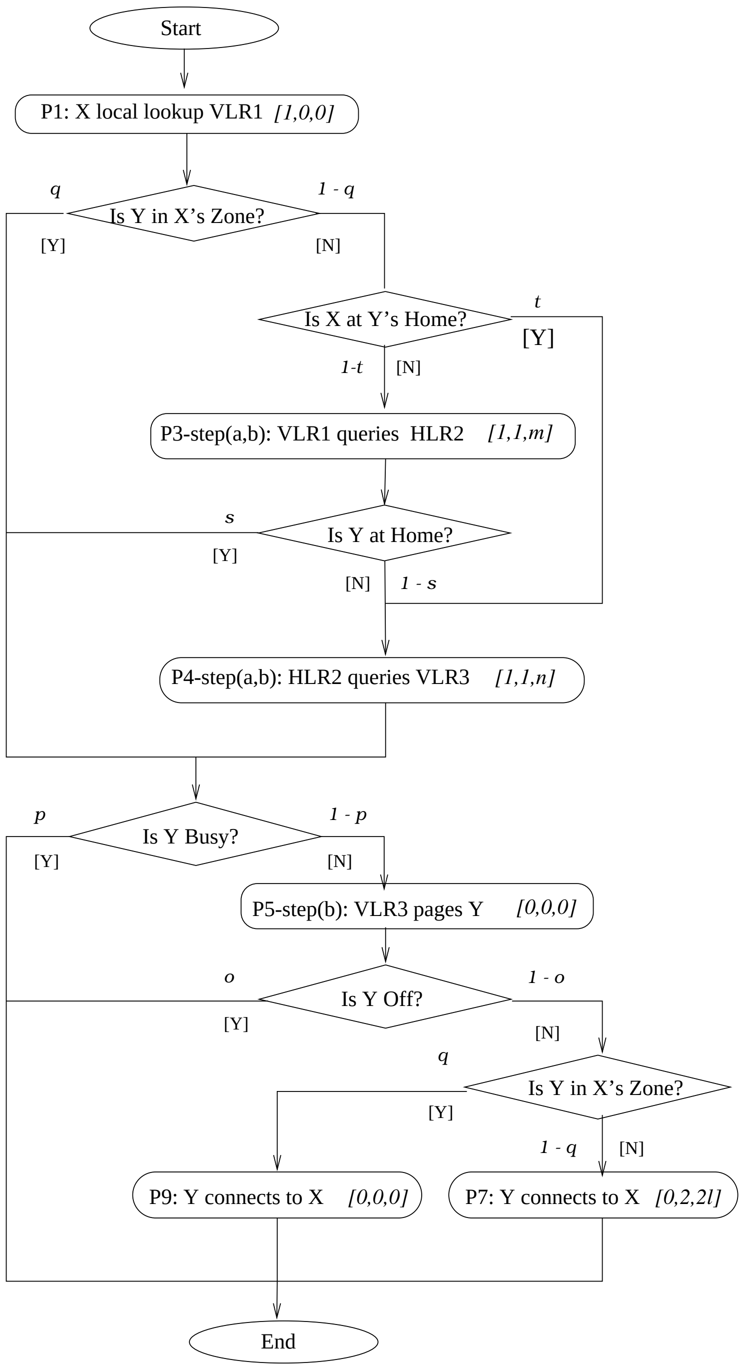

Figure 15: HLR/VLR in ReCS

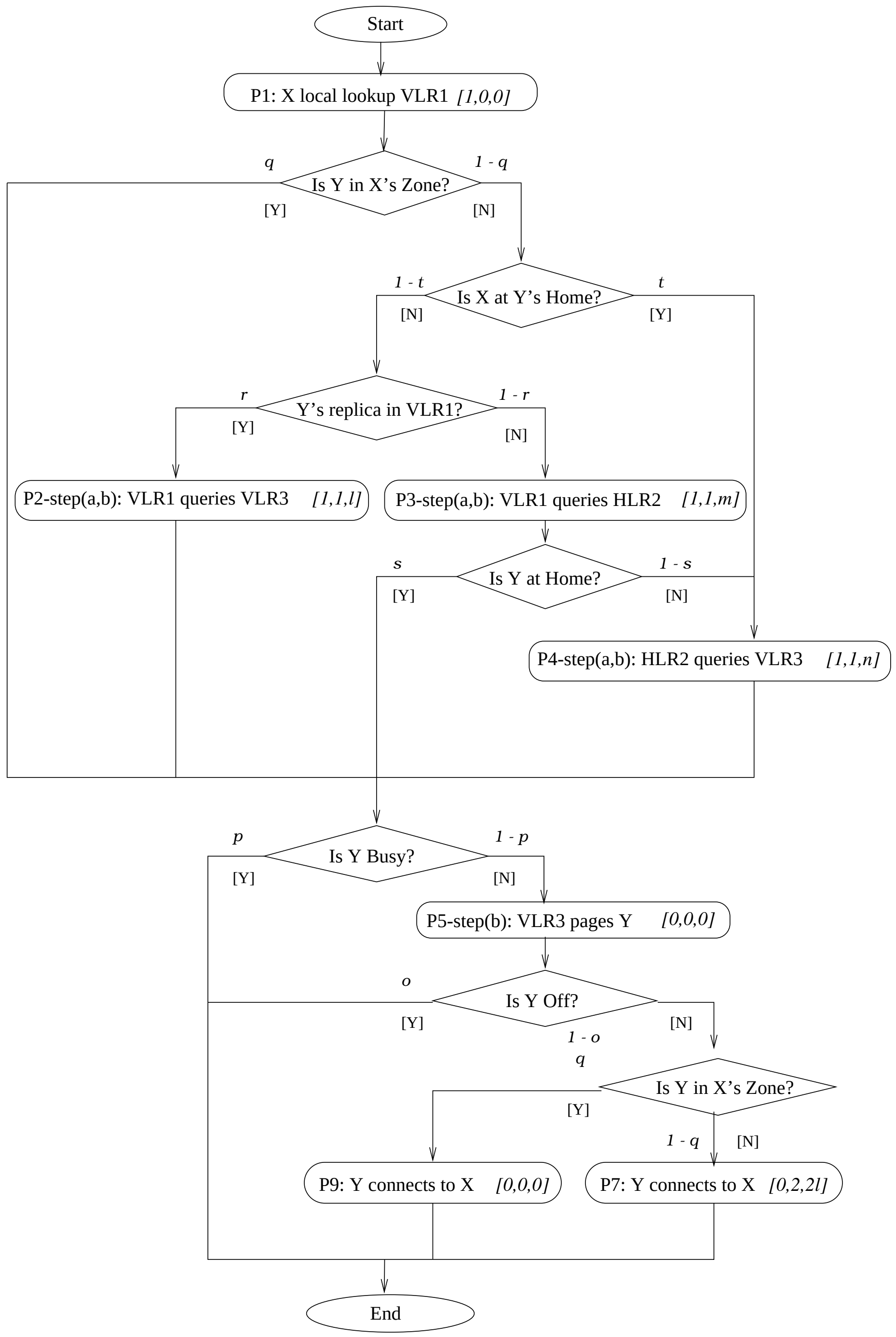

Figure 16: HLR/VLR with Replication in ReCS

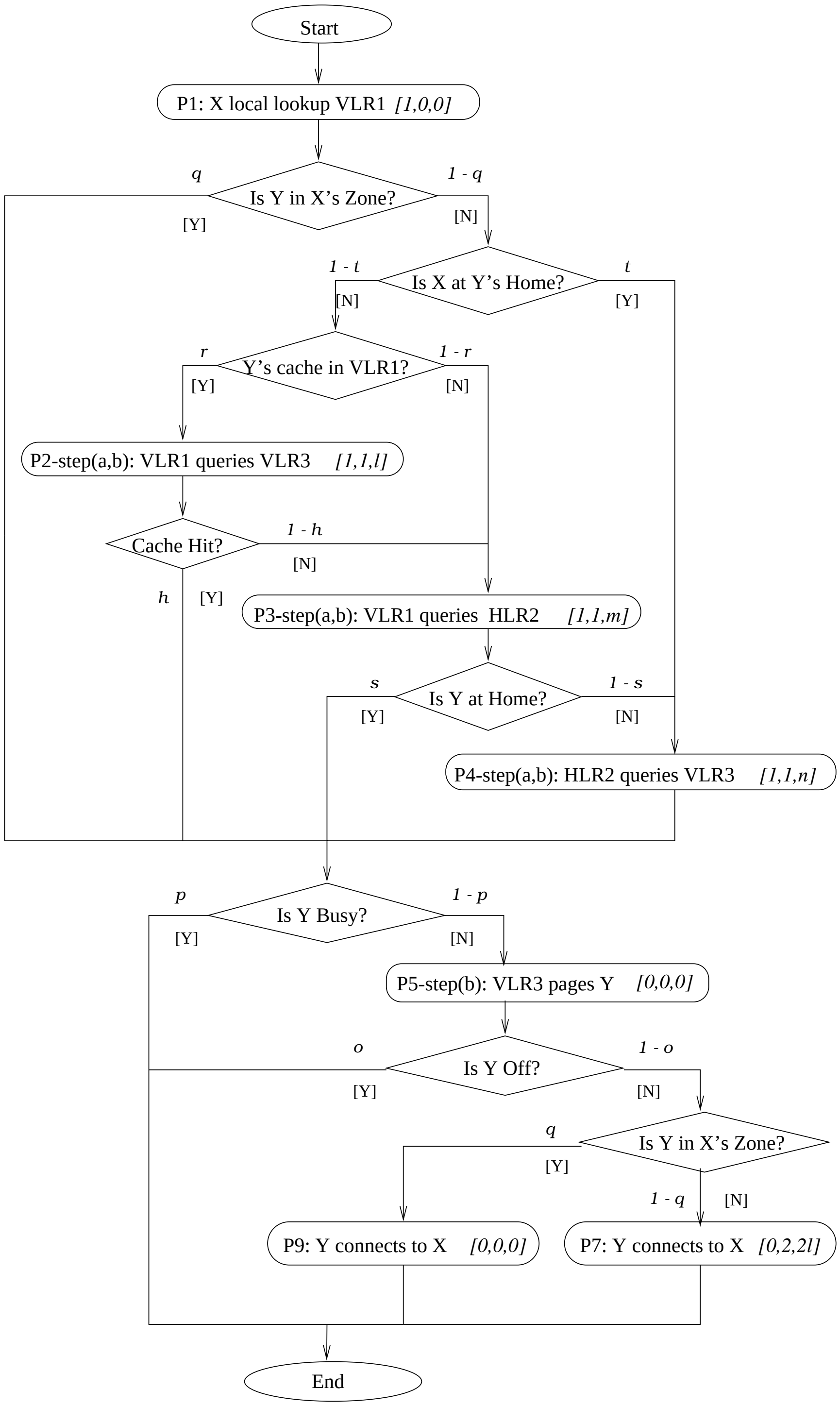

Figure 17: HLR/VLR with Caching in ReCS