Location Management

Location Management: Context

- **Mobility Management**: Enables users to support mobile users, allowing them to move, while simultaneously offering them incoming calls, data packets, and other services.
  - Types of mobility:
    1. **Terminal mobility**: ability of terminal to retain connectivity with the network so that all on-going communication services remain active despite terminal's migration.
    2. **Personal mobility**: disassociates user from the terminal (e.g. in GSM a mobile station = mobile terminal + smart card with subscriber identification module (SIM)).
    3. **Service mobility**: provides continuous service to mobile clients across multiple administrative domains.
  - Consists of:
    1. **Location management**: tracking mobiles and locating them prior to establishing incoming calls (deliverying pending messages).
    2. **Handoff management** (automatic link transfer): rerouting connections with minimal degradation of QoS.

Location Management Problem

- In static networks, a terminal's network address serves two purposes:
  1. End-point identifier
  2. Location identifier
  - Mobility prevents using a single address for both purposes
  - Both end-point identifier and location identifier are needed.
  - Location management keeps mapping between an end-point identifier and its location identifier
    - Basically a directory problem.
- Two primitive operations:
  1. Lookup (search/find/paging/locating) operation: is the procedure by which the network finds the location of the mobile.
     - required when a call (message) to a user is placed (to be delivered)
  2. Update (tracking/move/registration) operation: is the procedure by which the network elements update information about the location of the mobile.
     - required when a user changes its "location"
     - The information gathered during updating/tracking is used during the locating operation

Location Management: Issues

- More precise location needs to be maintained as cell size shrink:
  - Wide area cells are 10’s – 100’s km in diameter
  - Macro-cells: 1-10 km
  - Micro-cells: 100’s m
  - Pico-cells: under 10 m
- Database issues in tracking mobile users:
  - Maintaining update intensive location information
  - Strategies to reduce location query latency (such as replication) and traffic (such as caching)
  - Consistency between replicas; Cache management polices

Location Management Principles & Techniques

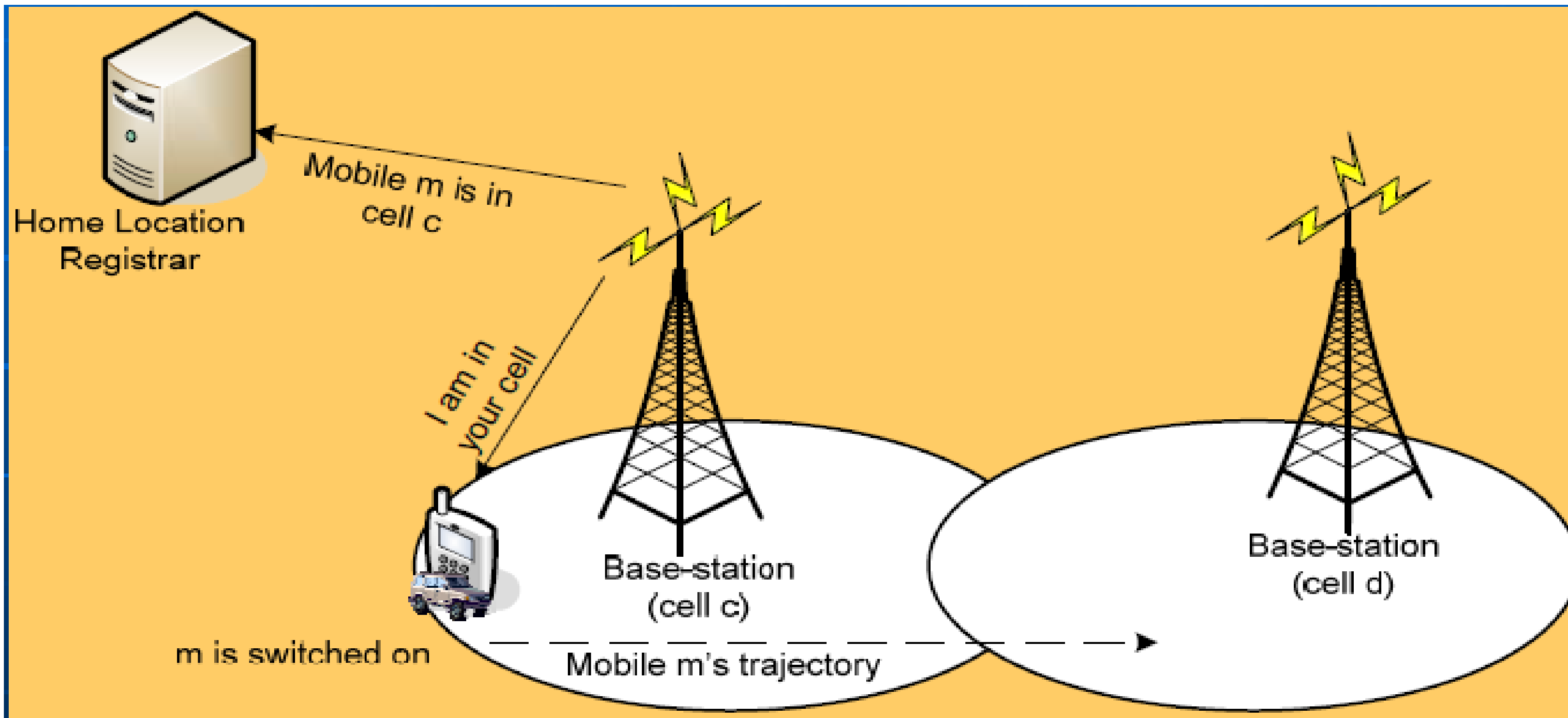
- Location Registrars (databases)
- Location Area
  - A set of base stations
  - Grouped for optimized signaling
- Search Operation
- Update operation
  - Static Update Schemes
  - Dynamic Update Schemes

Location Management: Schemes

- Several schemes have been developed which are motivated by fundamental trade-off between **search operation cost** and **update operation cost**.
  - Schemes which try to minimize one cost tend to increase the other cost
  - Try to optimize the aggregate cost or normalized cost.
- Categorization:
  1. Update Scheme: Static or Dynamic
     - Static update scheme: registration areas
     - Dynamic update scheme: distance/time/movement based strategy
  2. Locating Scheme: Static or Dynamic
     - Static location scheme: page all the cells in the network
     - Dynamic location scheme: expanding ring search centered at last reported location of the the user
  3. Database Architecture: Flat or Hierarchical

Simple Location Management Scheme

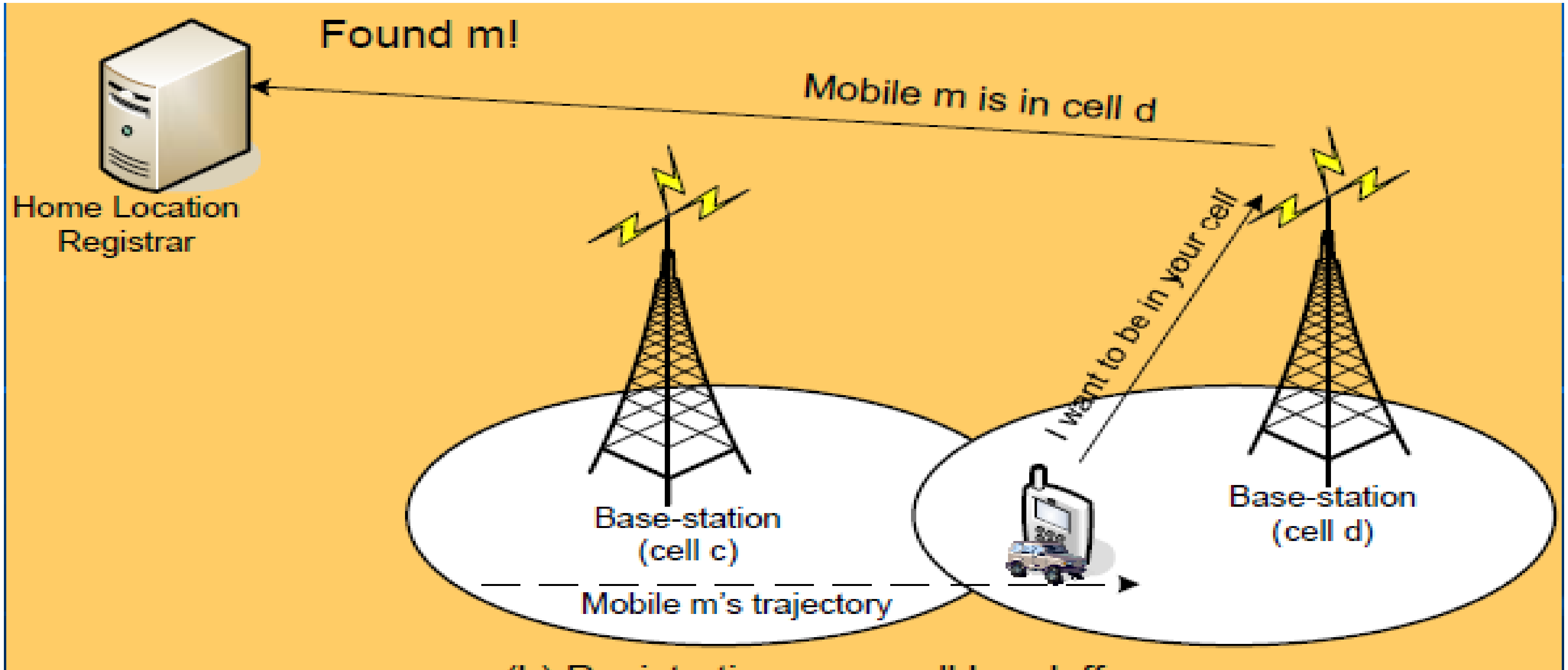
- Search and Update Operations (mobile node m is switched on) – Static Update



(a) Registration upon mobile switching on

Simple Location Management Scheme

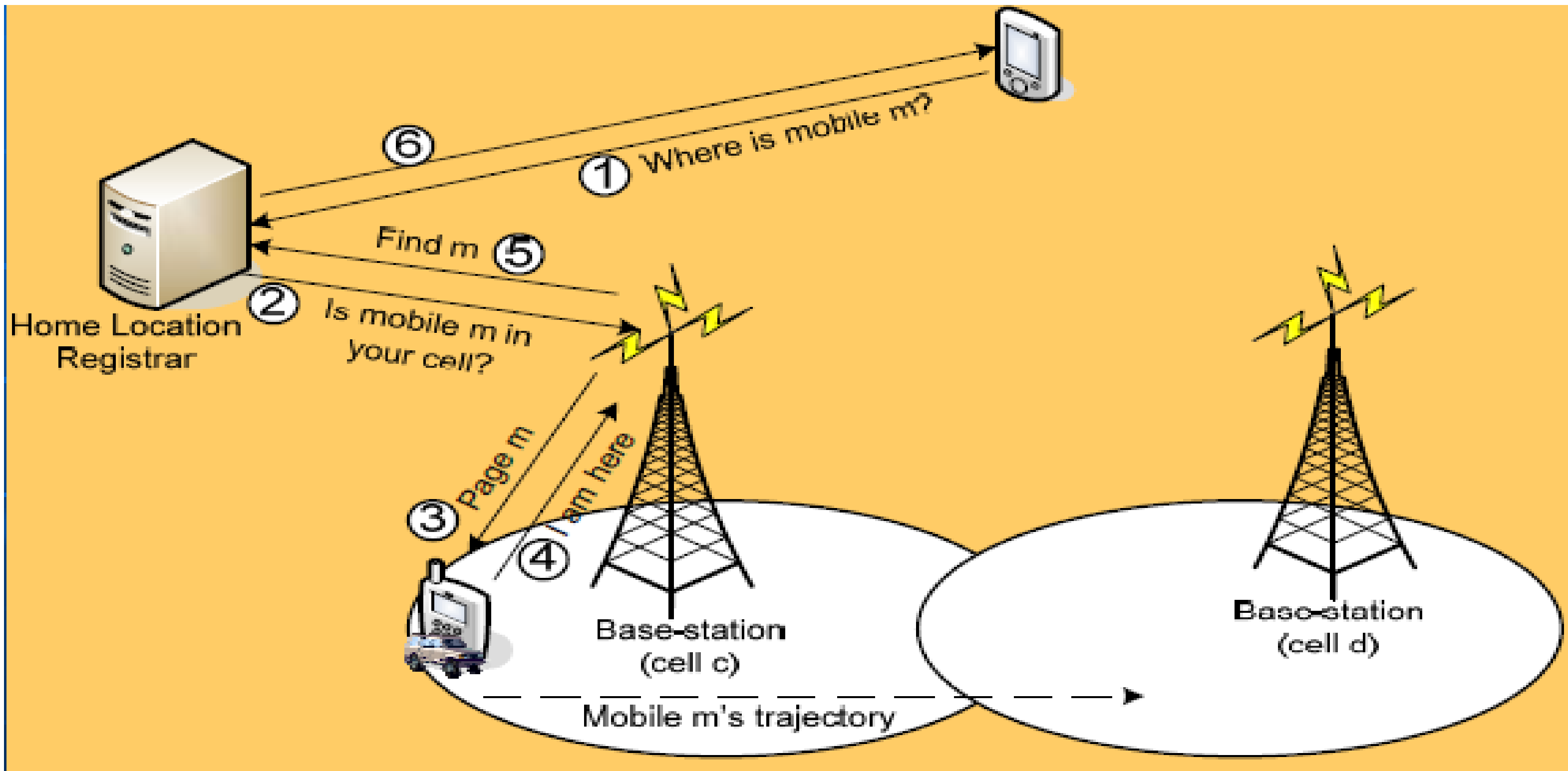
- Search and Update Operations (mobile node moves from cell c to cell d)



(b) Registration upon cell handoff

Simple Location Management Scheme

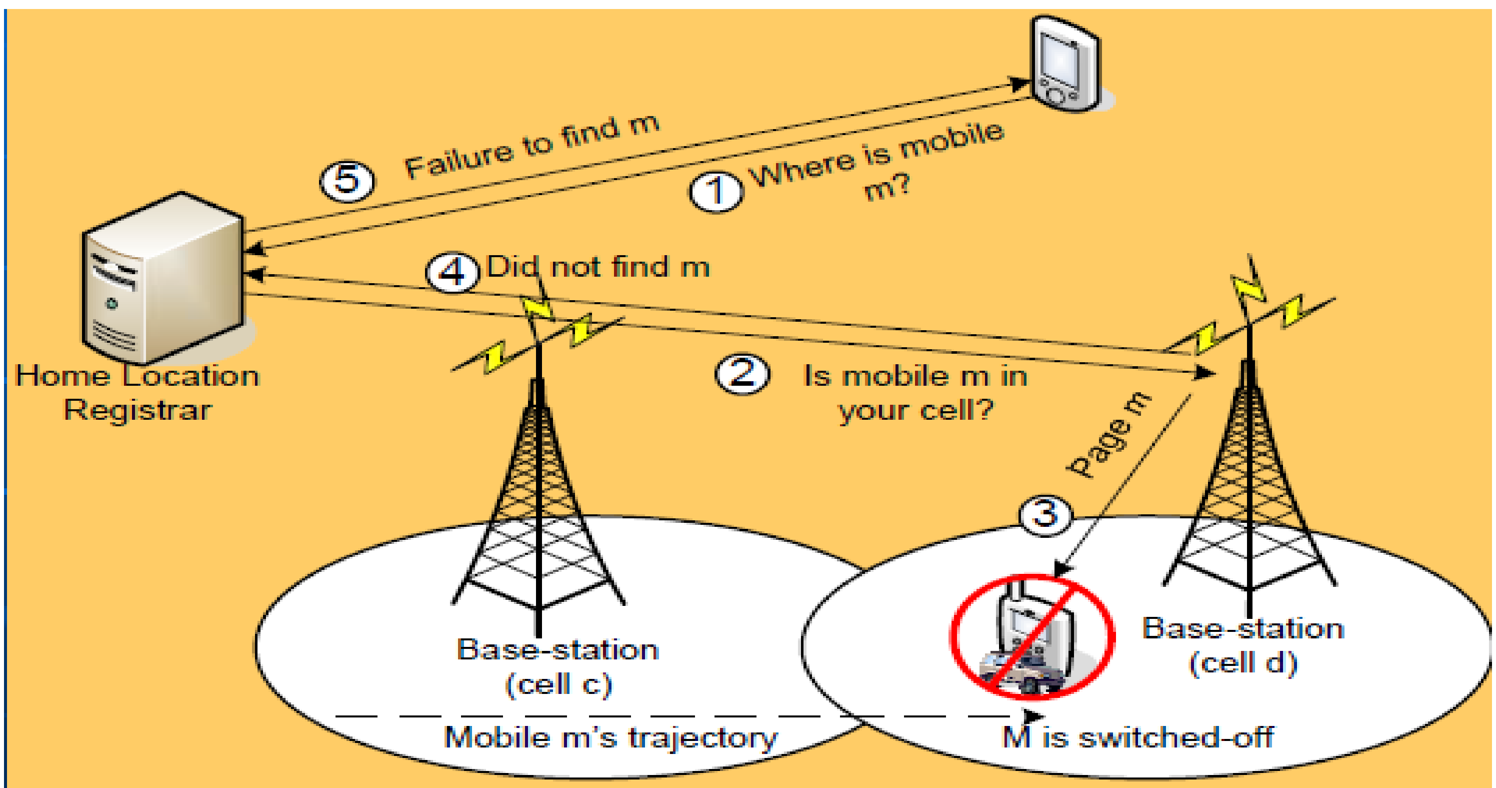
- Search and Update Operations (m in cell c & ON)



(c) Another mobile wants to find m – success case

Simple Location Management Scheme

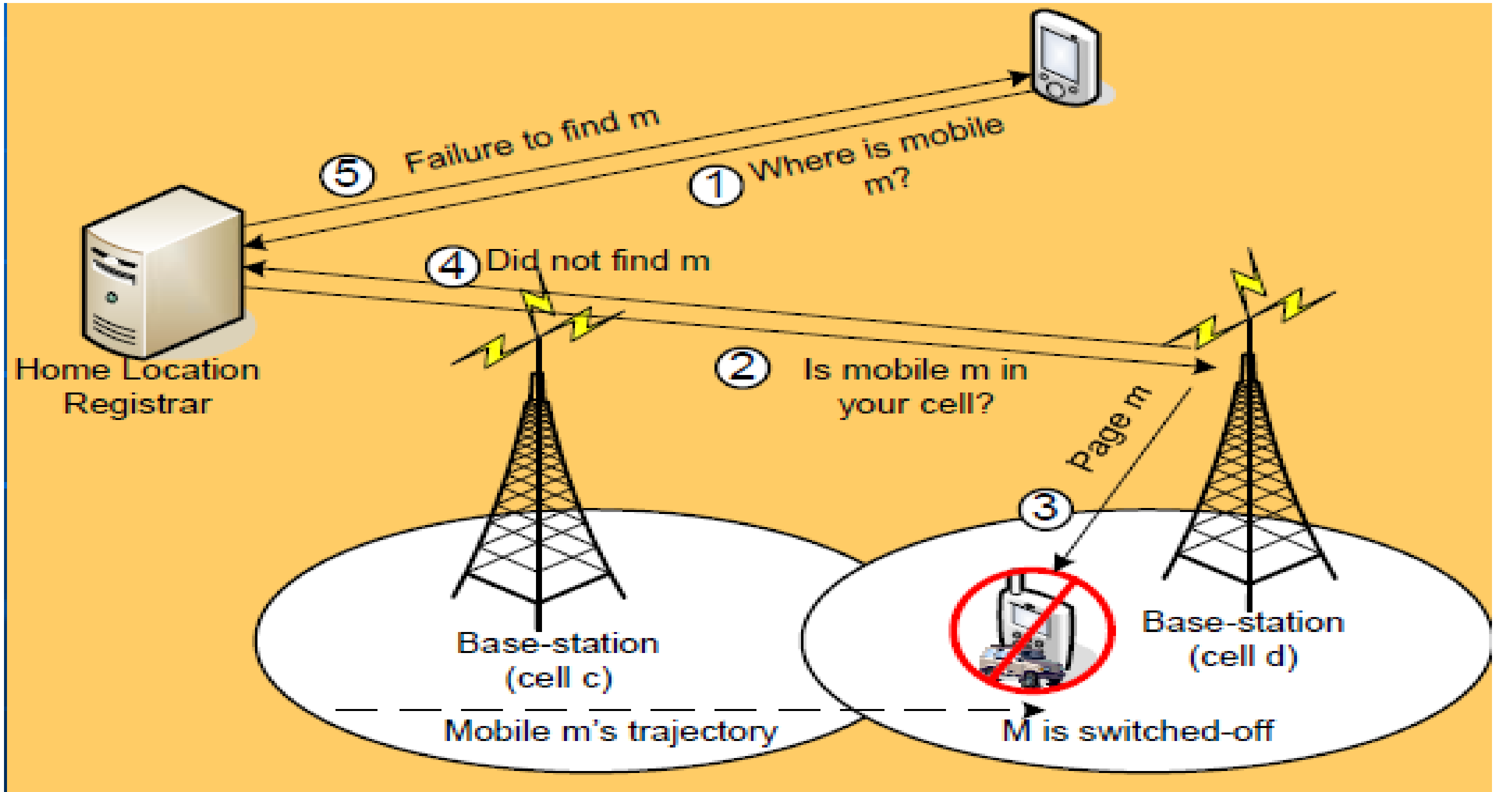
- Search and Update Operations (find m location; m is OFF)



(d) Another mobile wants to find m – a failure case

Simple Location Management Scheme

- Search and Update Operations (find m location; m is OFF)



(d) Another mobile wants to find m – a failure case

Selection of LM Schemes

- Cost of location updates and lookups
- Maximum service capacity of each location database =
  - the maximum rate of updates and lookups that each database can service
- Space restrictions (size of the location database)
- Type and relative frequency of call to move operations (call-to-mobility ratio (CMR))

One-Tier Scheme

- A home database, called Home Location Register (HLR) is associated with each mobile user.
- The HLR of a user x maintains the current location of x as part of x’s profile.
- To locate a user x, x’s HLR is identified and queried.
- When a user x moves to a new cell, x’s HLR is updated.

Two-Tier (Basic) Scheme

- Visitor Location Registers (VLRs) are maintained in each zone (registration area).
- VLR in a zone stores copies of profiles of users not at home and currently located in that zone.
- When a call is placed from cell i to user x, the VLR at cell i is queried first, and only if the user is not found there, is x's HLR contacted.
- When user x moves from cell i to j, in addition to updating x's HLR, the entry of x is deleted from VLR at cell i, and a new entry for x is added to the VLR at cell j.

Registration Area-based Location Management

- Current Personal Communication Service (PCS) networks (i.e., cellular networks) use RA-based basic HLR-VLR schemes:

–The service coverage area is divided into registration areas (RAs), each with a visitor location register (VLR)

–Each RA covers a group of base stations (cells).

–A user has a permanent home location register (HLR)

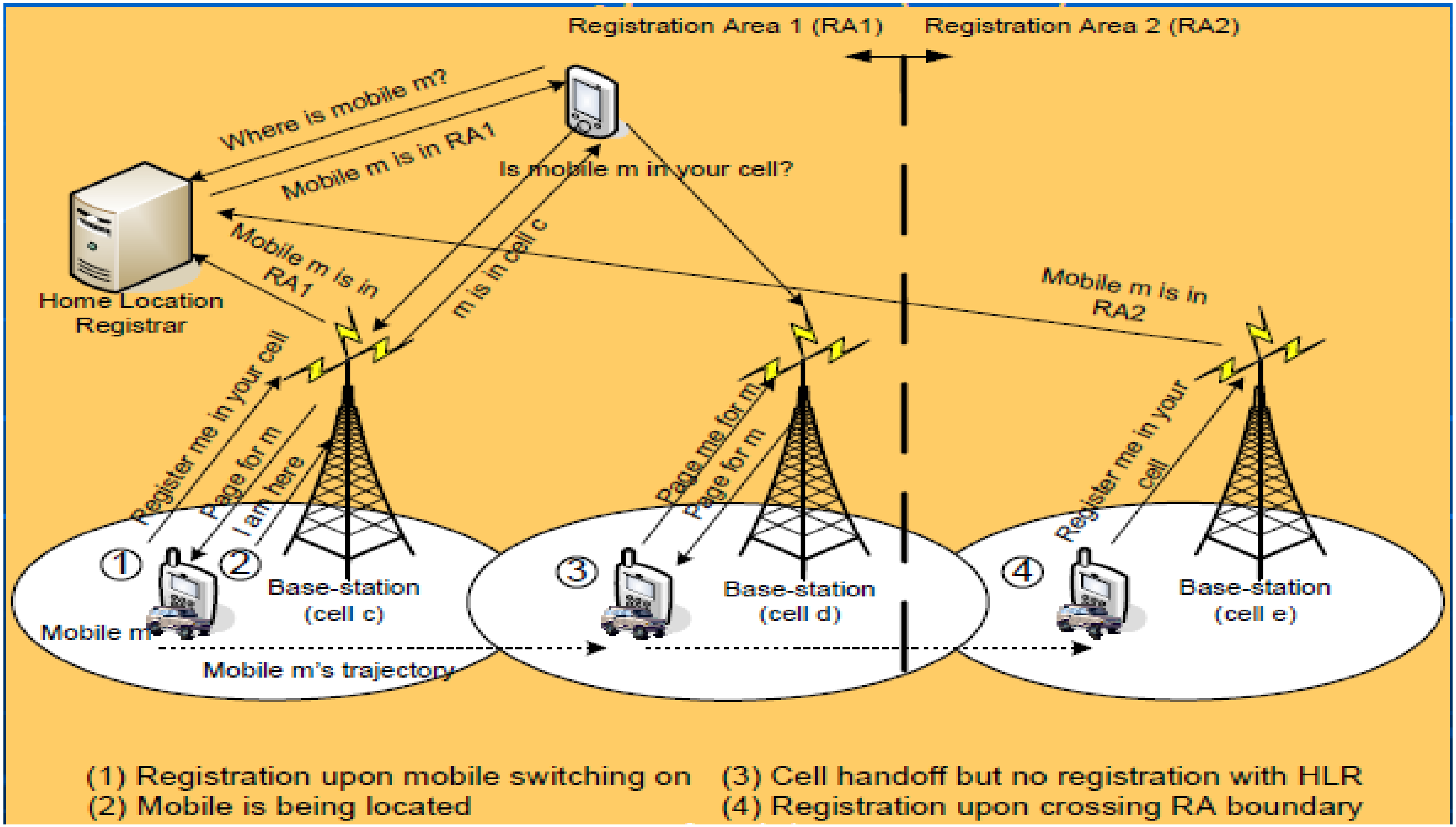
–Base stations within the same RA broadcast their IDs

–If ID is sensed different by the mobile terminal, then a cell boundary is crossed and a location update is sent to the VLR of the current RA.

–When crossing a RA boundary, an update is sent to the HLR.

–A search goes by HLR->VLR->cell->paging (by the base station)

Registration Area-based Location Management



Registration Area-based Location Management

- Cell c & d – in RA1
- Cell e – in RA2
- Node m moves from cell c to d
    - Average update cost is reduced, because the HLR is not informed when handoff involves cells belonging to same RAs
    - Search cost is increased, because all the cells in the RA have to be contacted for the exact location of the mobile node

Registration Area-based Location Management

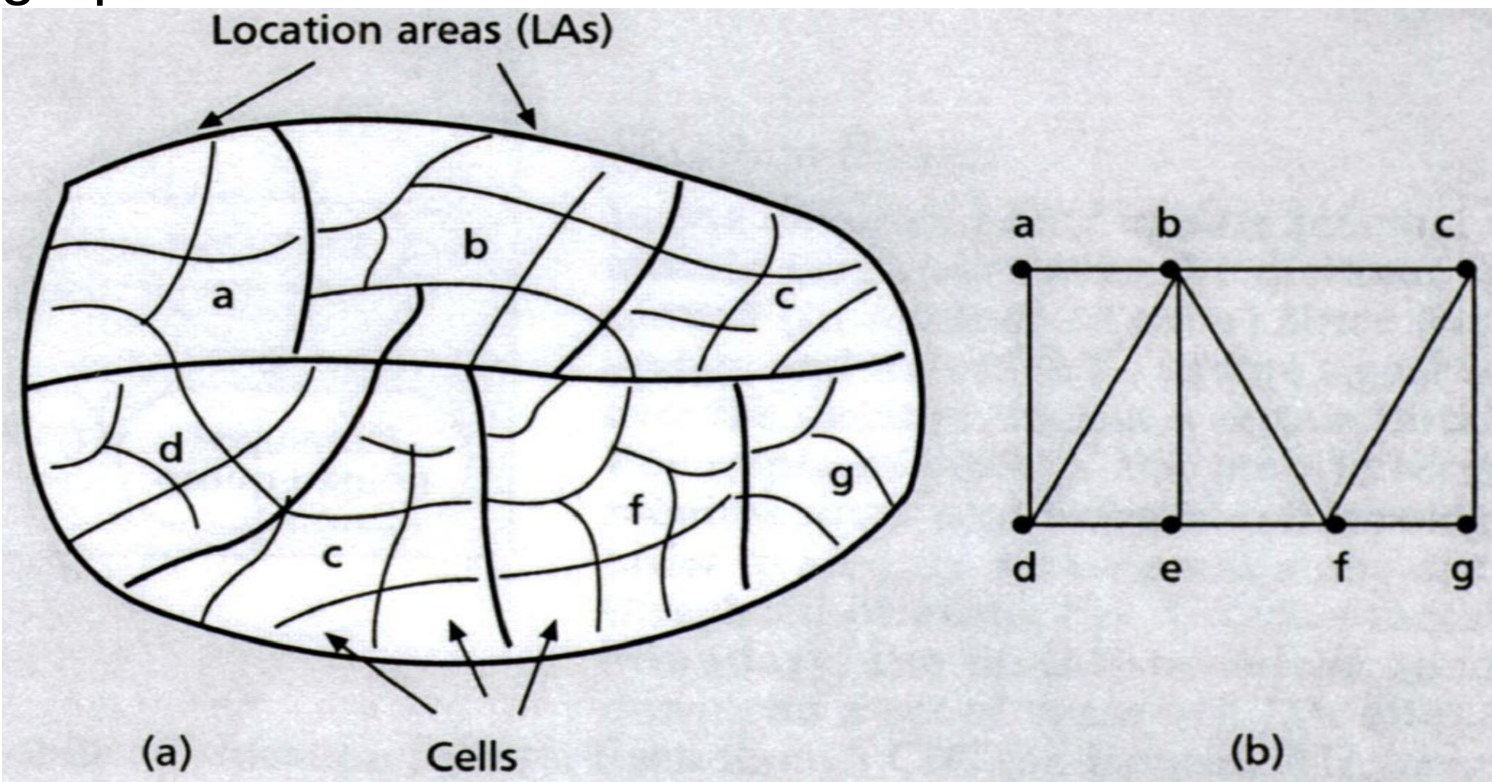
- 2-Level Hierarchy of Location Registrars
  - Local Location Registrars
  - Remote Location Registrars
- Used in GSM to avoid contacting all the cells in the RA to locate a mobile node
  - One Location Registrar ↔ 1 RA
  - One Location Registrar ↔ several RAs (in practice)
  - N Registration Areas (RA1, RA2, …, RAn)
  - N Local Location Registrars (LR1, LR2, …, LRn)
  - LRi is the Local Location Registrars of RAi
  - All others location registrars as Remote Location Registrars

Registration Areas in a PCN

- The figure shows a PCS network

–RA topology

–RA graph model



Per User Location Caching

- Basic Idea:
  - Every time user x is called, x’s location (or a pointer to this location) is cached at the VLR in the caller’s zone.
  - Any subsequent call to x originating from that zone can use this information:
    - Upon call origination the cache at the VLR of the caller’s zone is checked before querying the callee’s HLR.
- Issues:
  - cache replacement schemes (LRU can be used)
  - cache invalidation schemes
    - eager caching or lazy caching

Eager and Lazy Caching

- Eager Caching
  - Every time a user moves to a new location, all cache entries for this user's location are updated.
  - The cost of move operations increases for those users whose address are cached.
- Lazy Caching
  - the cached pointer for any given user is updated only on a cache miss
  - for lazy scheme to work better than basic scheme $p \geq C_H/C_B$ where p is the hit ratio, C_H is the cost of a lookup when there is a hit and C_B is the cost of lookup in the basic scheme.

Replication

- To reduce the lookup cost, the location of specific users is replicated at selected sites.
- Let
  - α : cost savings when local lookup succeeds as opposed to a remote query,
  - β : cost of updating a replica,
  - $C_{i,j}$: expected number of calls made from cell j to i in a unit time.
  - U_i: expected number of moves by i in unit time

Then a replication of the location of user i at cell j is judicious if

$$\alpha * C_{i,j} \geq \beta * U_i.$$

Per User Profile Replication

- Objective: to minimize the total cost of moves and calls, while maintaining
  - Constraint 1: a maximum of r_i replicas for user i, and
  - Constraint 2: a maximum of p_j replicas in the database of cell j.
- Replication assignment problem: The profile of user i is replicated at all cells in set R(i) such that the system cost

$$\Sigma^{N}_{i=1}\Sigma^{M}_{j=1, j\in R(i)} (\beta * U_i - \alpha * C_{i,j})$$

is minimized, where N is the number of users and M is the number of cells., and constraints 1 and 2 above are met

Working Set (WS) Replication

- Relies on the observation that each user communicates frequently with a small number of sources called its working set.
- Copies of location are maintained at the members of its working set.
- No constraints are placed on database storage capacity or on number of replicas per user.
- Hence, the decision to provide the information of the location of a mobile unit i to zone j can be made independently for each user.
- Adapts to user’s call and mobility patterns

Working Set Adaptation

- The inequality Q: $\alpha * C_{i,j} \geq \beta * U_i$. is evaluated locally at a mobile unit I each time:
  1. a call is set up,
  2. the mobile unit moves.
- In case 1, Q is evaluated only if the caller's site is not a member of mobile's WS:
  - If the inequality holds then the caller's site becomes member of the callee's working set.
- In case 2, the Q is evaluated for every member of WS; the members for which Q no longer holds are dropped from WS.

Performance of WS Replication

- Computation overhead:
  - in case 1 all four terms of Q need to be reevaluated
  - in case 2 only the number of moves (U_i) needs to be reevaluated.
- Adaptability:
  - when call-to-mobility ratio (CMR) value is low the WS scheme performs like a scheme without replication.
  - when CMR value is high, the scheme behaves like a static scheme in which the WS for a user is fixed.
- Performance is mainly dependent upon the CMR of individual users (not on num. of users).

Forwarding Pointers

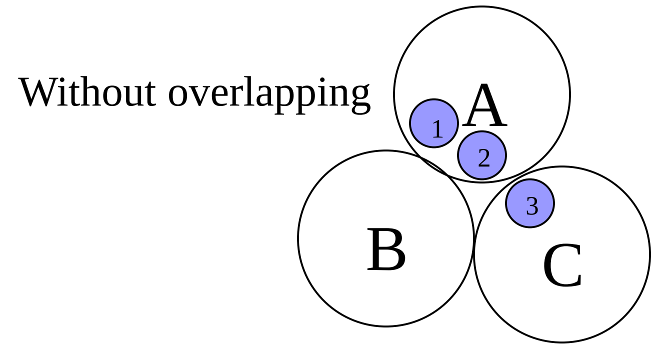
- Each time a mobile unit x moves to a new location, a forwarding pointer is set up to its previous VLR to point to the new VLR.
- To establish a call, the HLR of callee is queried to find the first VLR in the forwarding pointer chain. This chain is followed to get to the current VLR of the callee.
- To bound the time taken for lookup procedure, the length of the chain is bound to a max value of K.
- Pointer compression is used to eliminate loops.

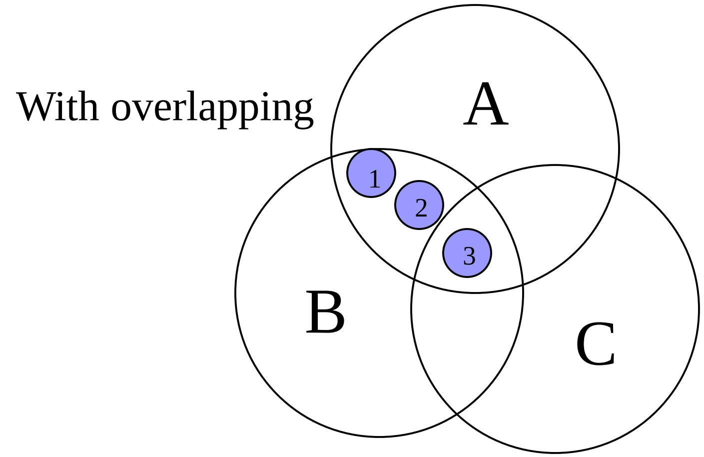
Forwarding Pointers (Cont.)

- Mobile IP protocol includes pointer forwarding in conjunction with lazy caching.
- The forwarding pointer strategy is useful for those users who receive calls infrequently relative to the rate at which they change registration areas.
- Benefits of forwarding depends also upon the cost of setting up and traversing pointers relative to the costs of updating the HLR.

Overlapping Registration Areas

- Inter-RA hand-off: a user changes cells and RAs
- Intra-RA hand-off: a user changes cells within an RA.
- Inter-RA hand-off doesn't happen as long as the hand-off can be intra-RA.
- Inter-RA call is when caller and callee are in separate RAs
- Intra-RA call is when caller and callee are in same RA.
- A non-overlapping cell is serviced by one LR.
- A overlapping cell is serviced by multiple LRs.
  - Reduction of inter-RA hand-offs.





Overlapping RAs (cont.)

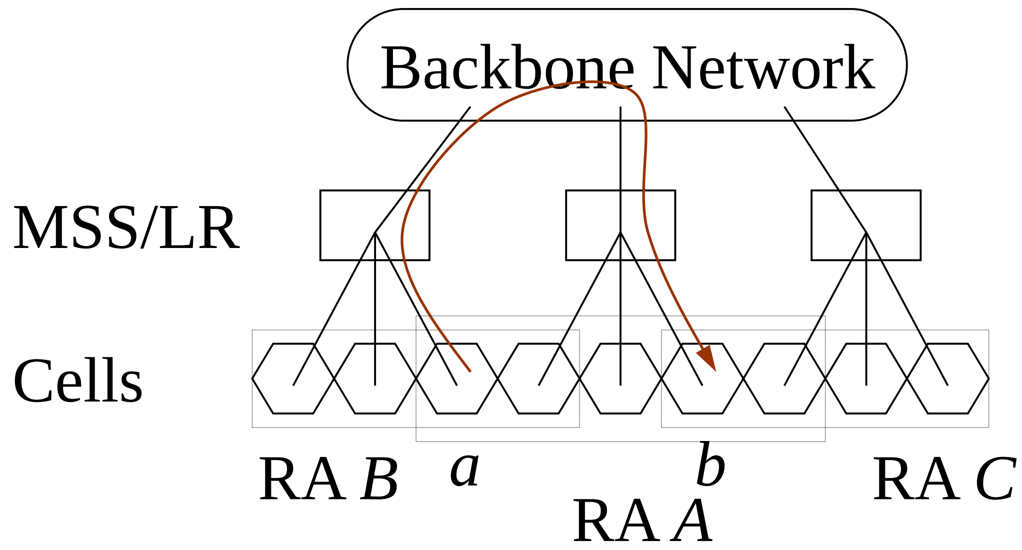
- Advantages:
  - Each RA can provide service to more mobiles within their covered area.
  - Reduces the number of inter-RA handoffs
  - Reduce the load to update mobile’s HLR.
- Disadvantages:
  - the communication overhead for call-delivery and intra-RA handoff is increased.
  - the increase in overhead depends upon the underlying network topology.
  - If this overhead is ignored then the extreme configuration in which each RA has all the cells in the system becomes the “optimal” configuration.

Overlapping RAs (cont.)

- Dynamically Resizing RAs
  - We need to find optimal configuration (allowing overlapping RAs) i.e. configuration which minimizes load on MSSs.
  - When move and call patterns periodically change, a static scheme may not provide a good solution
  - Our Approach: Allow RAs to be dynamically adapted.
  - Periodically resize RAs to minimize MSS load:
    - Resizing criterion: load reduction due to lesser number of inter-RA handoffs > increase in load due to more expensive call delivery and intra-RA handoffs.
    - If resizing criterion is ignored then each RA will grow to maximum size.

Overlapping RAs (cont.)

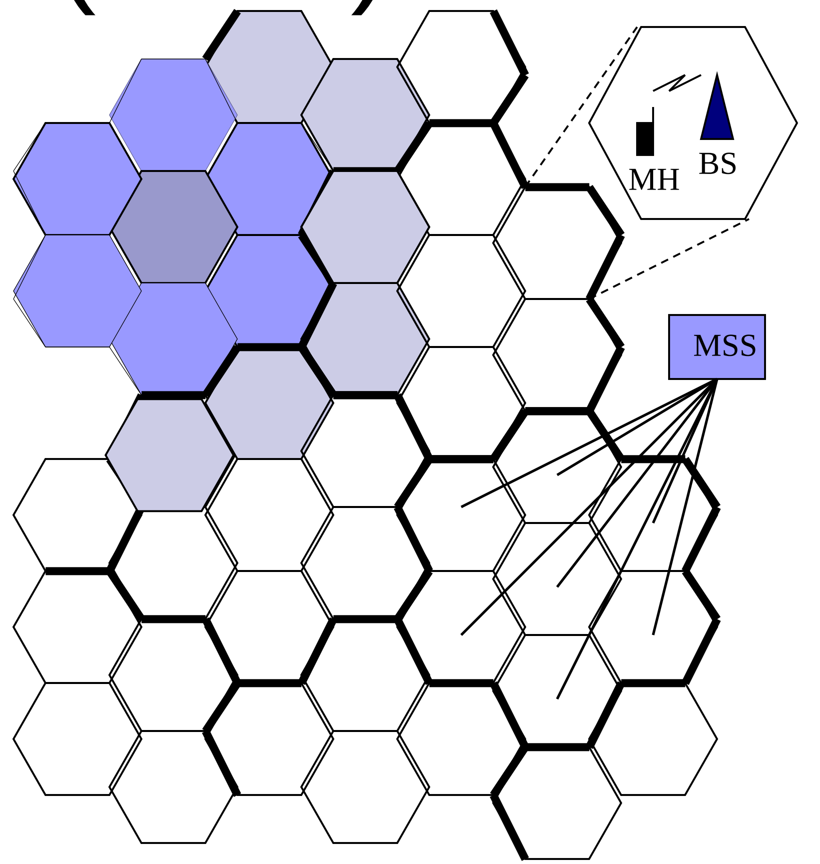
- Negative effect of underlying conventional star topology on signaling overhead under overlapping RAs



Even though mobiles *a* and *b* belong to the same RA, any calls between them would need to go through two MSSs.

Overlapping RAs (cont.)

- Inclusion and Exclusion Boundary
  - In order to facilitate orderly growth and shrinking of RAs, an MSS only includes and excludes cells from its RAs current boundary.
  - Two types of boundary:
    1. Internal Boundary
    2. External Boundary



Overlapping RAs (cont.)

- Inclusion/Exclusion Decision
  - The decision to include or exclude a candidate cell is based on whether the resulting configuration will have a lower expected load on MSS.
  - For a given system configuration A, mobility pattern M, and call C, SystemLoad(A,M,C) is the combined signaling load (in terms of message time complexity) as a result of all the handoffs due to M and call-deliveries due to C: SystemLoad(A,M,C) = $\sum$Load(k,M,C).
  - In case of inter_RA handoffs and call-deliveries we spilt the signaling overhead equally between the two MSSs involved.

Overlapping RAs (cont.)

- What changes when cell *x* is included in RA *r?*
  - Handoffs to cell *x* from cells of RA *r* become intra-RA handoffs.
  - Handoffs from cell *x* to rest of RA *r* performed by users already registered in *r* become intra-RA handoffs.
  - Calls to *x* from cells of *r* are now intra-RA calls.
  - Calls from users of *r* that are in *x* to rest of *r* are now intra-RA calls.
  - Mobility of users in *r* that move out of cell *x* into a new RA is now inter-RA mobility.
  - Inter-RA calls of users in r that call from cell x is inter-RA call loading to r.
- Call the decreasing part of the load $Cost_{in}(x,r)$ and the increasing part $Cost_{ex}(x,r)$.
  - At intervals T each MSS/LR *r* computes I_Boundary(*r*) and E_Boundary(*r*) and for each cell *x* in the two sets computes $Cost_{in}(x,r)$ and $Cost_{ex}(x,r)$. By comparing the two values, it decides if it is worth keeping excluded, keeping included, including or excluding the cell *c*.

Static vs Dynamic Update Schemes

- Static Update Scheme

• RA-based Location Update

• Ignore dynamic behavior of mobile nodes

• Boundaries of RAs are predetermined (static)

• Cost: a lot of location update due to mobile nodes moving between two adjacent RAs in quick succession

- Dynamic Update Schemes

• **Time-based (periodic) Updates:** A mobile terminal updates in every T time units

• **Movement-based Updates:** A mobile terminal counts the number of boundary crossings and performs the update when a threshold is exceeded (e.g. M=6)

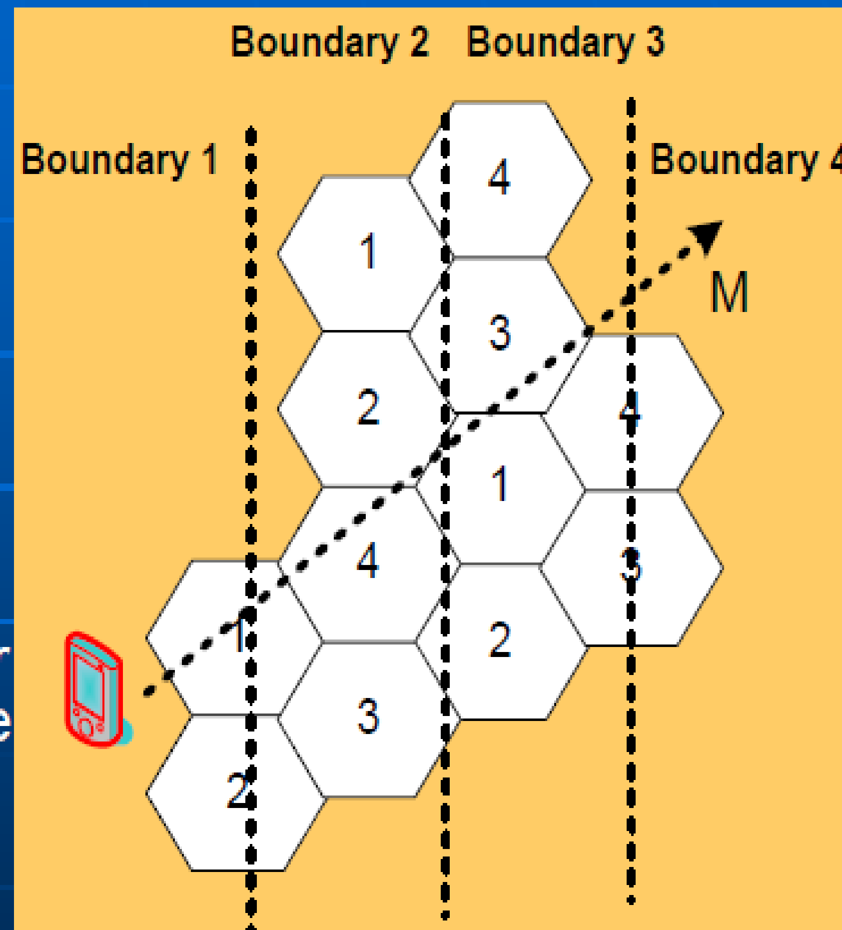
• **Distance-based Updates:** A mobile terminal tracks the distance (in terms of RAs) it has moved since the last update

–Update is performed when a distance threshold is exceeded

–Mobile terminal needs some knowledge of the network topology

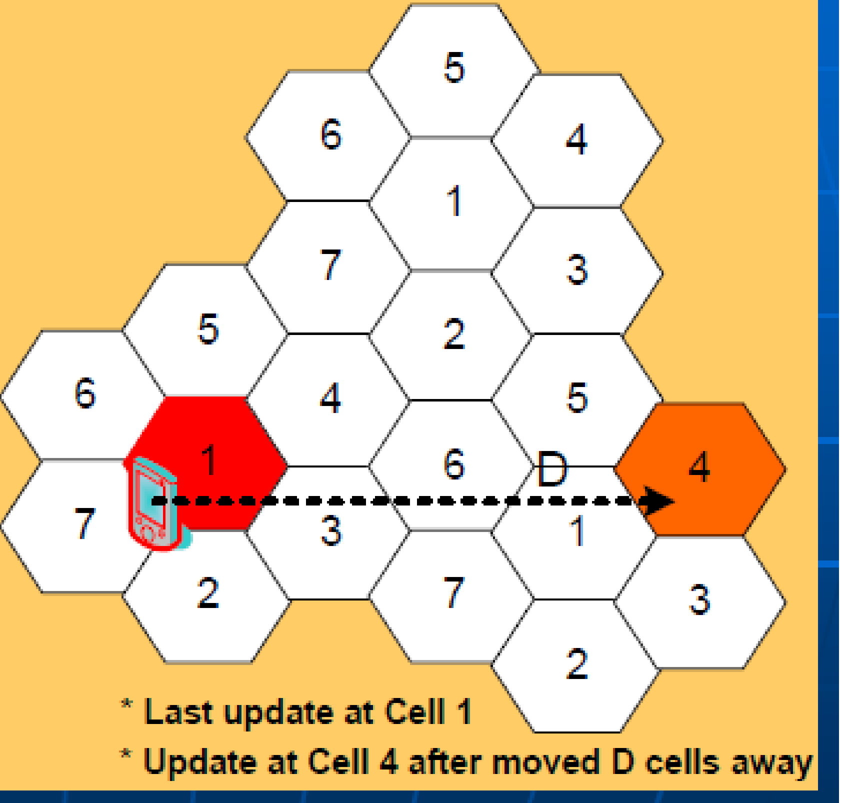
Dynamic Update Schemes

- Movement-based Updates
  - A mobile node update its location
  - When?
    - It crosses a certain number of cell boundaries M since it last registered
  - Mechanism
    - Counting the number of Handoffs since the last update
  - Suitable for stationary users



Dynamic Update Schemes

- Distance-based Updates
  - A mobile node updates its location
  - When?
    - It moves a certain number of cells D away from the last cell at which it last updated its location
  - Need to know the topology of cellular network
  - Difficult to implement
  - Suitable for mobile user who moves within a locality



Dynamic Update Schemes

- **Update Time Interval for Time-Based Schemes**
- Design parameter: T (the time interval for performing a location update)
- Assume a search operation performs an expanding ring search
- Let l be the call arrival rate and s be the mobility rate.
- The maximum area to be searched for is a circle with the radius of s * min(1/l, T) cells.
- Normalize each search operation with a cost of s * min(1/l, T)
- Normalize the cost of each update operation with a cost of 1
- The cost of time-based management per unit time is:

$$C = l * s * \min(1/l, T) + 1/T$$

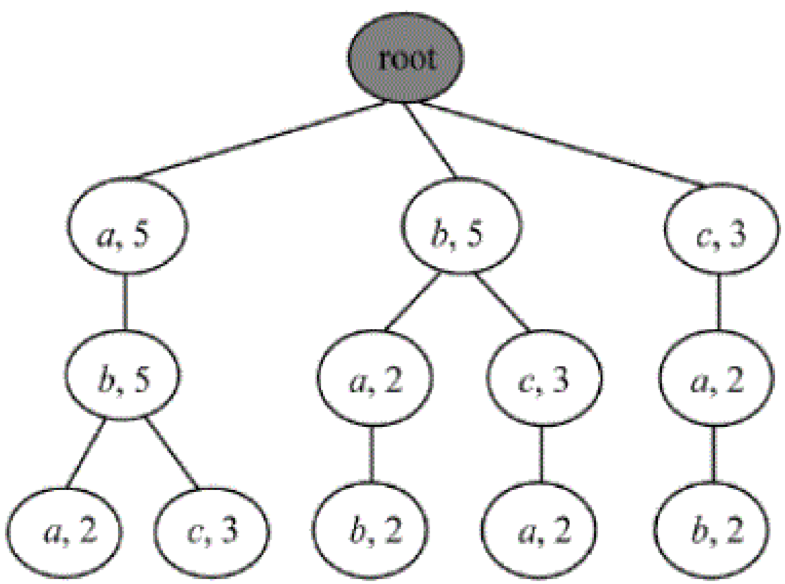
- When 1/l < T (frequent arrivals), C = s + 1/T ≈ s
- When 1/l > T (infrequent arrivals), C = l * s * T + 1/T. Take dC/dT=0, Topt = 1/sqrt(sl), implying that when either s or l decreases, Topt increases

LeZi Update

- Based on a compression algorithm by Ziv and Lempel
- •LeZi is a **path-based** update algorithm by which the movement history, not just the current location, is sent in an update message from the mobile user to the location database server
- –The history has a list of IDs of the zone (RA or cell) the mobile terminal has crossed
- –The location database keeps the history in compact form by means of a search tree structure (called a **LeZi trie**)
- •Can be part of the user's profile
- –On a call arrival, prediction of the current cell is given and selective paging is performed

Location Prediction Example

- Input sequence: … ababcabcababc (c is the current cell)
  - The LeZi Trie so far:



- Estimate the probability of the next cell using 0, 1 and 2 last cells history:
- P_{0a}=5/13, P_{0b}=5/13, P_{0c}=3/13
- P_{1a}=1, P_{1b}=0, P_{1c}=0 (after c)
- P_{2a}=1, P_{2b}=0, P_{2c}=0 (after bc)

$$P_a = w_2 \cdot P_{2a} + w_1 \cdot P_{1a} + w_0 \cdot P_{0a}$$
$$P_b = w_2 \cdot P_{2b} + w_1 \cdot P_{1b} + w_0 \cdot P_{0b}$$
$$P_c = w_2 \cdot P_{2c} + w_1 \cdot P_{1c} + w_0 \cdot P_{0c}.$$

  - What's the next cell?
    - P_a: Probability that the next cell is *a*