Exploiting User Profiles to Support Differentiated Services in Next-Generation Wireless Networks

Vijoy Pandey
Nortel Networks,
4655 Great America Parkway,
Santa Clara, CA 95054, USA
vijoy@nortelnetworks.com

Dipak Ghosal
Dept. of Computer Science
University of California,
Davis, CA 95616, USA
ghosal@cs.ucdavis.edu

Biswanath Mukherjee
Dept. of Computer Science
University of California,
Davis, CA 95616, USA
mukherje@cs.ucdavis.edu

ABSTRACT

In the next-generation wireless network, user profiles such as the location, the velocity (both speed and direction), and the resource requirements of the mobile device can be accurately determined and maintained by the network on a per-user basis. We investigate the design of a differentiated-services architecture which exploits user profiles to maximize the network efficiency and which supports differentiated services classes, each with different Quality-of-Service (QoS) guarantees. In this paper, we provide implementation details of such an architecture for the Third-Generation Partnership Project (3GPP) network. The key underlying primitive of the architecture is the use of user profiles to perform advance resource reservation in target cells of the wireless cellular network. We identify the design tradeoffs and present performance results for an architecture consisting of two service classes, namely (1) a higher-cost profiled service with higher QoS, and (2) a lower-cost non-profiled service with best-effort QoS. Our analysis indicates that a significant decrease in the dropping probability[1] — and, hence, higher QoS — can be guaranteed to users who subscribe to the profiled service. We examine the tradeoffs associated with some of the key system parameters including the reservation distance and the reservation granularity, and we determine their values which maximize the improvement in the dropping probability for all users.

I. INTRODUCTION

The next-generation wireless network [1], [2], [3] will support a rich set of multimedia applications similar to those available in wired networks. To achieve the goal of providing high-quality multimedia services to anyone, anywhere, and at any time [4], network designers will need to implement new techniques that can support Quality of Service (QoS) while accounting for limited bandwidth and for the delay and error characteristics of the wireless network [5]. To support and guarantee QoS, the next-generation wireless network must implement a differentiated-services architecture. This architecture would contain multiple service levels, each with a different QoS guarantee.

The mobility pattern of a user has a high degree of predictability due to temporal and spatial locality. Temporal locality refers to the fact that a mobile user typically takes predictable routes, which implies that a user will typically cross the same set of cells at predictable times in a wireless cellular network. For instance, a mobile user will typically follow the same path to work in the morning, and the reverse route back home in the evening. Spatial locality refers to the fact that user mobility is constrained along pathways and highways which results in a mobile user crossing the cells in an ordered sequence determined by the manner in which these pathways intersect the cellular coverage area.

The real-time and aggregate values of a user's mobility and resource requirements are known as the "user profile". The goal of our study is to investigate the design and implementation of a user-profile-based differentiated-services architecture for the next-generation wireless network. Specific implementation details have been provided for the Third-Generation Partnership Project (3GPP) [3] cellular network architecture, but should translate easily to most other wireless architectures as well. We have studied the performance benefits of our proposed approach in a network with two types of users - (1) profiled users who subscribe to a higher-cost profiled service which guarantees higher QoS and (2) regular (non-profiled) users who receive best-effort service. We observe that the network provides improved QoS to profiled users by significantly reducing their dropping probability through advanced reservation of cell resources along the path predicted by the user profile. There are optimal values of the reservation distance (which is the distance prior to a cell crossing when the reservation is attempted) and the reservation granularity (which is related to the frequency of the re-attempts when a reservation attempt fails) which result in the maximal improvement in dropping probability.

II. EXISTING AND EMERGING NETWORK INFRASTRUCTURE

Figure 1 shows the components of a cellular network architecture[2]. In this section, we will only dwell upon those components that relate directly to our architectural implementation. For an in-depth description of this architecture, the reader is referred to [3].

Wireless service providers are rapidly shifting focus from simple voice services to mobile LoCation Services (LCS) [6], [7] which utilize a user's position information to provide localized and personalized services. In this extended cellular architecture, the *Gateway Mobile Location Center (GMLC)* is responsible for interfacing with the external world, i.e., with the LCS clients who request the mobile's position. The *Serving Mobile Location Center (SMLC)* determines the geographical coordinates of the mobile, and the potential error. in accordance with the quality requested from the GMLC and the capability of the mobile. The *Location Measurement Unit (LMU)* helps the SMLC take synchronization measurements. Its key function is to determine the

[1] Dropping probability is the probability that an admitted call fails due to an unsuccessful handoff.

[2] We have chosen the 3GPP network architecture for our examples, but the concepts outlined in this investigation should translate easily to other wireless cellular architectures as well.

0-7803-7569-6/02/$17.00 ©2002 IEEE

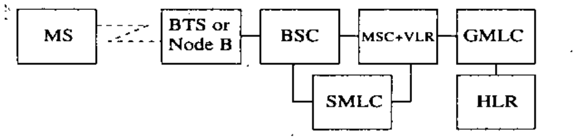

Fig. 1. Network elements needed to support profile-based channel reservation.

location of a mobile and convert it into meaningful coordinates X,Y. The LMUs can either be integrated into the Base Transceiver Station (BTS) or be independently distributed across the network.

III. USER PROFILE AND USER PROFILE REGISTER

We introduce an architectural component called the **User Profile Register (UPR)** which is a database similar to Home Location Register (HLR) and Visitor Location Register (VLR) [3]. The UPR contains user-profile information which can be queried for by the wireless network to provide differentiated services to its customers. A user profile consists of mobility patterns and services accessed by the mobile user, tabulated against the time of the day and the day of the week. It contains pointers to network elements which can provide real-time values of the user's location and velocity information. A UPR should have interfaces to external information sources — such as network information databases, described later — to aid in QoS management. The components of the UPR are outlined below.

- **User Location Interface:** This is a logical interface to devices such as the SMLC and the LMUs which can provide real-time user location values to the UPR.
- **User Velocity Interface:** This logical interface will query the network element responsible for real-time velocity estimation of a mobile user [8], [9].
- **User Path Table (UPT):** This table is an ordered list of the most probable paths a mobile user could traverse at any given time on any day of the week. A mobile path is a list of Cell-IDs, $< c_1, c_2, ...c_n >$, which a mobile user traverses. This path could contain a number of *hot-spots*, which is defined as a collection of cells within a geographical region where the mobile-user population density is very high, e.g., greater than a pre-defined threshold. City downtown regions, train stations, airports, or residential areas represent typical hot-spots. Hot-spots could be dynamic in nature, and can change depending upon the time of the day, traffic conditions and special events. The user could make a call and terminate it at any point along this path. If all the cells in a mobile user's path are contained within a single hot-spot, the path is considered *internal* to the hot-spot, and does not appear in the UPT. The UPT can either be specified by the user before usage, or it could be built up statistically, by prioritizing the paths that the user takes more often.
- **User Resource Table (URT):** This table is an ordered list of resources (services) a mobile user uses at any given time on any day of the week. As with the UPT, the URT can either be user-specified or it could be statistically constructed over time by tallying the services being accessed by the user.
- **Interfaces to External Information Systems:** This entry contains variables obtained from external information systems such as network architecture databases or GPS and Global Information Services (GIS) devices. For example, network architecture databases could supply information about cell layout and sizes. GPS and GIS devices could supply real-time changes in hot-spot definitions.

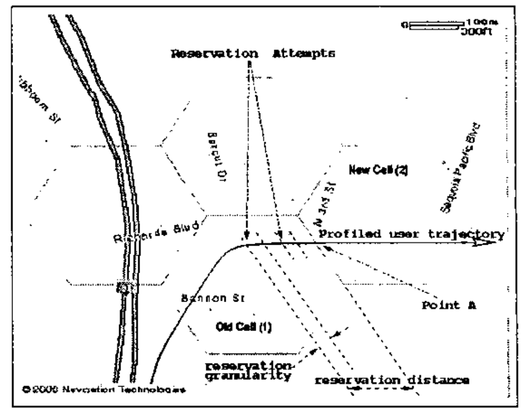

Fig. 2. Illustration of PARMA.

IV. USER-PROFILE-BASED DIFFERENTIATED-SERVICES ARCHITECTURE

In order to implement a differentiated-services architecture in a wireless network, the following two key components should be designed, namely (1) *a call-admission-control algorithm* and (2) *a call QoS control framework*. In this work, we mainly focus on a call QoS control framework, which is a set of algorithms and policies that perform resource management in the cellular network and that attempt to guarantee the required QoS to the users admitted into the network. User-profile-based call QoS control algorithms utilize the statistical profiles of the mobile user to manage and control the network resources in order to guarantee the negotiated QoS.

For this study, we have considered two classes of service, for two types of users — (1) profiled users who subscribe to the profiling service, expect better QoS, and hence pay more, and (2) non-profiled users who pay less and expect a "best-effort" service from the network. Below, we outline our user-profile-based resource-management algorithm.

A. PARMA: Profile-Assisted Resource-Management Algorithm

Figure 2 shows a part of the trajectory of a user commuting from the Arden Town suburb to Richards Boulevard, near downtown Sacramento. The steps a cellular network would take to reserve resources for a profiled user are outlined below; these steps form our profile-assisted resource resource-management algorithm (PARMA).

1. When a mobile user is *close* to a cell boundary, the network would consult its subscriber database to find out whether the user is a profiled customer.
2. If the user does not subscribe to the profiling service, then she travels to Point A and attempts a handoff to Cell 2. If there are no channels available in Cell 2 at this handoff instant, the user gets dropped.
3. If the user is a profiled customer, the network consults the UPR and extracts the user's resource requirements from the URT. For

the purpose of this discussion, let us assume that the top-most entry in the URT for this part of the mobile's path is voice services, and hence would simply need a channel reservation in the target cell.

4. The network tries to predict the target cell, based on the UPT and the current location and velocity of the mobile.

5. The network would then attempt to reserve a channel in advance for the user in Cell 2 when the user is at a distance r_d from the cell boundary, where r_d is known as the *reservation distance*.

6. If the reservation attempt succeeds, the user is handed off to Cell 2 on the reserved channel at Point A. If the reservation fails, the network re-attempts the reservation every r_g (*reservation granularity*) distance apart, till the reservation succeeds or till a handoff takes place at Point A.

7. If the reservation is unsuccessful till Point A (even after several attempts), then the user session gets dropped. By allowing multiple reservation attempts, the dropping probability of a profiled user can be substantially reduced.

As mentioned in Step 1, PARMA is initiated when the mobile user is *close* to a cell boundary. There are two key approaches to proximity evaluation, and both approaches are conceptually equivalent. One approach uses the signal strengths provided by a mobile for the purpose of a handoff decision, to estimate proximity and also the target BTS. As an alternative, the network could utilize the user's velocity (speed and direction), match it against the user trajectory in the UPR, and therefore deduce proximity information. In our study, we choose the first approach to measure a user's proximity to a cell boundary.

Though we have considered channel reservation to highlight our algorithm, PARMA is much broader in scope. Any user-specific resource, such as the browser cache [10] for mobile browsers, session and state information for data connections, application proxy states for thin clients running on mobile devices, as defined in the URT for the profiled user, can be allocated for in the target cell.

V. DESIGN AND IMPLEMENTATION ISSUES

We can leverage the location-measurement infrastructure described in Section II for obtaining current updates to the position of a profiled user. Furthermore, we have to modify some of the network elements to support PARMA.

- The SMLC should be modified to store a short history of the mobile's position instead of storing just the current position. The size of this history depends on the accuracy of the path-prediction algorithm.
- The UPR database should be implemented to include the profile tables and real-time values for each mobile user in her home area. Each UPR should also be able to accept and incorporate updates to user profiles available from the Mobile Switching Center (MSC).
- Logical interfaces should exist between the UPR and external information databases and systems, such as the HLR, network architecture databases, the LMUs, and the SMLC. The HLR interface would aid the UPR in gathering subscription information about a user. Network architecture databases would provide information on hot-spot definitions and cellular layouts. LMUs and the SMLC would provide current location information and would assist in user-velocity estimation.

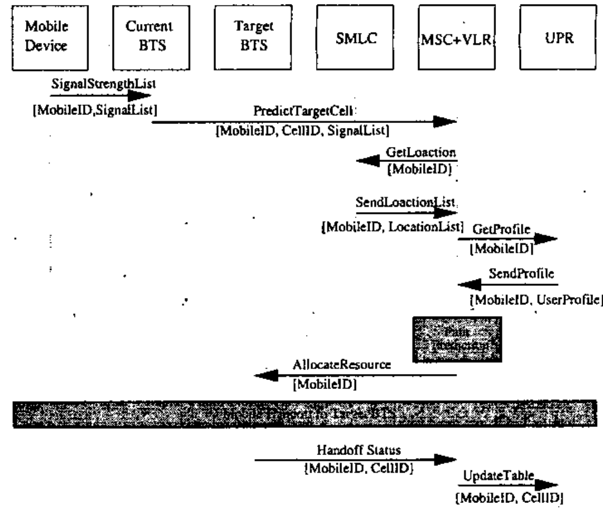

Fig. 3. Message Sequence Chart for PARMA.

- The software in the MSC should be enhanced to statistically update the UPT and the URT as described in Section III. The MSC should be able to make corrections to the user profile depending upon the success or failure of the user-profile-based path-prediction process and the services which the user has accessed, by increasing the "rank" of successfully predicted paths and resource requirements in the UPT and URT, respectively. The MSC should then provide this feedback to the UPR database.

Figure 3 shows the message sequence chart for PARMA. In current networks, a mobile device periodically sends a list of BTSs and their signal strengths to the current BTS for the purpose of a handoff, using a *SignalStrengthList* message. We can modify the BTS software to trigger a *PredictTargetCell* signal to the MSC whenever the signal strength of another BTS comes within a trigger threshold (S_T) of the signal from the current BTS. We study the impact of this threshold on network performance in Section VII.

On receiving the *PredictTargetCell* signal, the MSC sends out a *GetLocation* message to the SMLC requesting the past few coordinates (positional history) of the mobile user. The SMLC replies using *SendLocationList*. For performing path prediction, the MSC also requires the UPT and the current velocity of the user. It also requires the URT for gauging the resource requirements of the mobile in the target cell. The *GetProfile* and *SendProfileList* messages accomplish this task. The MSC now performs path prediction and informs the most probable target BTS through *AllocateResource* to reserve resources for the mobile depending on the most probable services accessed by the customer. After handoff, the new BTS sends a *HandoffStatus* to the MSC. The MSC checks this new cell's Cell-ID to confirm correctness of the target cell. *HandoffStatus* also verifies whether the resource reservation was sufficient. Finally, the MSC updates the UPT and the URT at the UPR based on these status results.

Though we have shown nine messages for implementing PARMA (see Figure 3), the overheads are minimal as most messages can be piggy-backed on existing signals, as described in [11]. The additional signals in PARMA include the *PredictTargetCell* and *UpdateTable* messages. Given the, relatively small overhead in executing the path-prediction algorithm at the MSC, PARMA should be quite lightweight.

There are several issues to consider when implementing PARMA in a wireless network [11]. When implementing such a network architecture, the network designer could utilize a user's speed to dynamically determine the length of the reservation granularity, and hence could fine-tune the number of reservation attempts. By using direction information coupled with the network's information of cell boundaries, PARMA can employ techniques such as hysteresis and signal thresholds to reduce unwanted reservation re-attempts. The designer could utilize an available macrocell tier to temporarily "hold" the session (and resources) of a profiled user on an unsuccessful reservation attempt, while PARMA keeps re-attempting the reservation requests in future target microcells. Since intra-hot-spot paths are more difficult to predict as compared to inter-hot-spot paths [11], the designer may opt to employ path prediction for the latter, while switching over to velocity-estimation algorithms and location-measurement technologies for estimating intra-hot-spot target cells. For the lack of space, we have barely skimmed the surface of many of the design issues that a network architect would face.

VI. A QUANTITATIVE ANALYSIS

We have simulated a single-tier cellular network architecture, whose modeling parameters, along with their default values, have been discussed below. The primary resource accessed by cellular users in this network are channels. There are a limited total number of channels, $C\{400\}$, available to the network with a static reuse pattern with a reuse distance ratio of $R\{2\}$ [12]. We assume that new-calls arrivals into the network follow a Poisson distribution with parameter $\lambda\{0.025\}$ calls/sec. Call-holding time is assumed to follow an exponential distribution with a mean of $1/\mu\{120\}$ seconds. We employ a hexagonal cell structure with a cell radius of $c\{0.5\}$ km. We model a limited user population of $U\{10000\}$ users at any given time in the network, out of which a fraction $p\{0.5\}$ of the users are profiled. We assume a circular hot-spot with a radius of $h\{3\}$ km. Each user can be plotted as the co-ordinate (ρ, δ), where ρ is uniform between 0 and h, and δ is uniform between 0 and 2π. The user density thus obtained closely approximates the characteristics of a hot-spot. There are $H\{3\}$ hot-spots in the region covered by the network. We assume that each user has a different direction of movement (D) and speed ($V\{25-40mph\}$) which is chosen uniformly between a specified speed range. For the purpose of this study, we assume that we can accurately predict a user's trajectory at every given point in time.

We have studied the *Improvement in Dropping Probability*, γ, defined as the reduction in dropping probability of a profiled user as compared to a non-profiled user. If P_{dp} is the dropping probability for profiled users and P_{dn} is the dropping probability for non-profiled users, then γ is defined as: $\gamma = \frac{P_{dn}-P_{dp}}{P_{dn}} * 100\%$

VII. RESULTS AND DISCUSSION

Figure 4 shows the improvement in dropping probability, γ, experienced by profiled users as the new-call arrival rate is increased from 0.005 calls/sec to 0.05 calls/sec for $C = 300$ and $C = 400$ channels. At the left extreme of the figure, when λ is small (e.g., when $\lambda = 0.005$ calls/sec), the load to the network is very light. Hence, very few handoff attempts of both profiled and non-profiled users get dropped resulting in a small γ. At the other extreme, when λ is large (e.g., when $\lambda = 0.05$ calls/sec in this example), the network load is high. A significant number of the channel-reservation requests for profiled users get blocked. This causes the dropping probability for both non-profiled and profiled users to be close to each other, again resulting in a small γ. When the network load is moderate, the profiled users obtain the most benefit from channel reservation. The dropping probability for non-profiled users increases [11], while the dropping probability for profiled users flattens out, benefiting from the reservations. This results in a substantial improvement in dropping probability, with a peak occurring at $\lambda = 0.025$ calls/sec (for $C = 400$ channels), when we observe an improvement of 37%. For $C = 300$ channels, γ peaks at 34% for $\lambda = 0.02$ calls/sec.

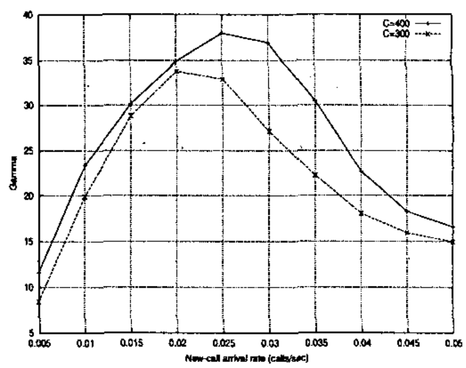

Fig. 4. γ for different new-call arrival rates.

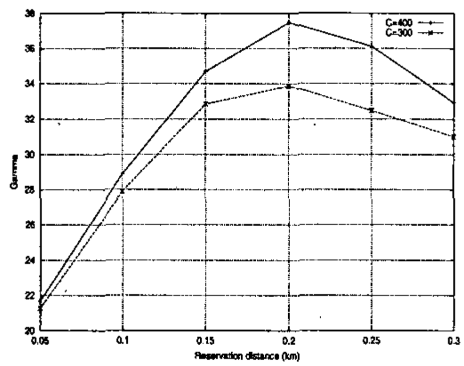

Fig. 5. Impact of reservation distance on γ.

Figure 5 shows the dependence of γ on the reservation distance, r_d. It should be pointed out that the region of the figure where $r_d > 0.3$ is not practical, since the network should not start reserving channels for a profiled user when that user is half a cell radius away from the cell boundary, and hence has not been shown in the figure. We observe that the there is an optimal value of r_d (at $r_d = 0.2$ km) which results in the maximum improvement in dropping probability. When r_d is small (e.g., at $r_d = 0.05$ km), the reservation attempts are made too close to the cell boundary.

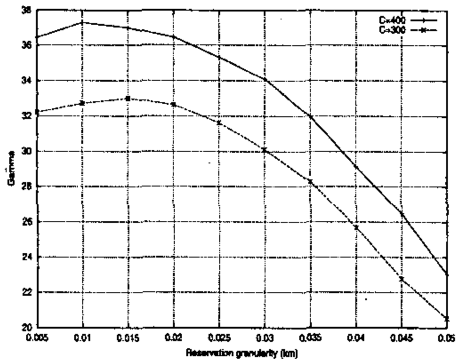

Fig. 6. Impact of reservation granularity on γ.

Hence, there is not enough time to recover from a failed reservation attempt before the handoff occurs. Therefore, the dropping probabilities of non-profiled and profiled users differ by a small margin resulting in a small γ. When r_d is too large (e.g., at 0.25 km and beyond), the channel-holding time of profiled users is inflated by a large amount which causes the overall load in a cell to increase. This results in large dropping probabilities for both profiled and non-profiled users. Thus, we observe a small improvement in the dropping probability.

Figure 6 shows the variation in γ with respect to the reservation granularity, r_g. Again, we observe an optimal value of r_g ($r_g = 0.01$ for $C = 400$) which results in the maximum improvement in dropping probability. The reason for this optimality is very similar to the one presented above. If we keep r_g small, the network makes a large number of reservation attempts on behalf of the profiled user. Though this should improve the dropping probability of profiled users, a very small value of r_g results in higher load to a cell, and hence a large dropping probability for non-profiled as well as profiled users. This results in a small γ as can be seen in the left region of the Figure 6. When r_g is large, there are not enough reservation re-attempts for profiled users. Hence, there is very little difference in dropping probabilities between non-profiled and profiled users.

VIII. CONCLUSION

With the continuing deployment of intelligent network components, it is becoming easier to collect and maintain accurate real-time data on the location, the velocity, and the resource requirements of a mobile user. These data can be used to develop user profiles, and they can also be aggregated to develop mobility and resource-requirement patterns of users in a region. We have made the following contributions in this work: (1) We have described the design and implementation of a scheme, called PARMA, which utilizes user profiles to provide better QoS to mobile users in a wireless network. Specific implementation details have been proposed for the 3GPP network architecture, though the concepts would be broadly applicable to most wireless network architectures. (2) There are numerous challenges and design issues in implementing such a scheme for the next-generation wireless networks, and we attempt to resolve some of these design issues. (3) Through detailed simulation, we have studied the benefit of user profiles in improving the QoS of cellular customers. We have studied a resource-allocation scheme using the concept of reservation distance and reservation granularity. We have shown that this concept can produce significant improvement in dropping probability of profiled users over their non-profiled counterparts. We showed that there are optimal values of the reservation distance and the reservation granularity parameters which result in maximal improvement in dropping probability.

REFERENCES

[1] UMTS Forum, http://www.umts-forum.org, *Universal Mobile Telecommunications System.*

[2] International Telecommunication Union, http://www.imt-2000.org, *International Mobile Telecommunication - 2000.*

[3] 3rd Generation Partnership Project, *Network Architecture (TlTRQ3GPP 23.002-330),* Standards Committee T1 Telecommunications, version 3.3.0 edition, 2000.

[4] V. O. K. Li and X. Qiu, "Personal communication systems (PCS)," *Proceedings of the IEEE,* vol. 83, no. 9, pp. 1210–1243, September 1995.

[5] M. Naghshineh, M. Schwartz, and A.S. Acampora, "Issues in wireless access broadband networks," *Wireless Information Networks, Resource Management, and Mobile Data, edited by J.M. Holtzman,* 1996, Kluwer Academic Publishers.

[6] M-A. Dru and S. Saada, "Location-based mobile services: The essentials," Tech. Rep., Alcatel Telecommunications Review, 1st Quarter 2001.

[7] Nokia, http://www.nokia.com, *Mobile Location Services,* 2001.

[8] B. Jabbari and W. Fuhrmann, "Teletraffic modelling and analysis of flexible hierarchical cellular networks with speed sensitive handoff strategy," *IEEE Journal on Selected Areas in Communications,* vol. 15, no. 8, pp. 1539–1548, October 1997.

[9] K. L. Yueng and S. Nanda, "Channel manangement in microcell/macrocell cellular radio systems," *IEEE Transactions on Vehicular Technology,* vol. 45, no. 4, pp. 601–612, November 1996.

[10] S. Hadjiefthymiades and L. Merakos, "Using proxy cache relocation to accelerate web browsing in wireless/mobile communications," in *Proc., World Wide Web Conference,* Hong Kong, May 2001, vol. 10.

[11] V. Pandey, D. Ghosal, and B. Mukherjee, "Exploiting user profiles to support differentiated services in next-generation wireless networks," Tech. Rep. CSE-2002-37, Department of Computer Science, University of California, Davis, November 2002.

[12] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Personal Communications,* pp. 10–31, June 1996.

Due to space limitations, we have limited our references. A detailed study of related literature, a more comprehensive list of references is presented in our technical report [11].