Network Selection for Heterogeneous Multi-Service Wireless Networks

Jun Seok Kang, Seung-jae Han
Department of Computer Science,
Yonsei University,
Seoul, Korea
jun.kang@yonsei.ac.kr, sjhan@cs.yonsei.ac.kr

Abstract— **Wireless networks have been evolved to heterogeneous overlay structure, in which various wireless technologies provide partial or full coverage of the service area. Another evolution direction is multi-service wireless networks. In multi-service networks, traffic of various QoS classes is serviced concurrently. In this paper, we propose a scalable and flexible network selection scheme for multi-service heterogeneous wireless networks. In the proposed scheme, the profiles are generated for users or cells to represent their characteristics in three categories, which are traffic status, mobility and QoS. The best cell is determined for each user via profile matching. The proposed scheme can be tailored to utilize the information available from the underlying wireless technologies. Via extensive simulations, it is shown that the proposed scheme handles the interplay between multi-service traffic classes more effectively than existing schemes. We discuss how the proposed scheme can be applied to the 3GPP LTE networks.**

Keywords- network selection, handover decision, heterogeneous wireless networks, multi-service

I. INTRODUCTION

Wireless networks have been evolved to heterogeneous overlay structure, in which various wireless technologies provide partial or full coverage of the service area. Another evolution direction is multi-service wireless networks. In multi-service networks, traffic of various QoS classes is serviced concurrently by sharing the available network resources. Network selection in heterogeneous wireless networks is different from the traditional cell selection in homogenous networks, which solely relied on the received signal strength (RSS), in two aspects. First, the comparison of RSS from the cells of heterogeneous wireless technologies is not as obvious as it used to be. Second, certain wireless technologies are more suitable for certain types of services. Thus, new methods that embrace the network-specific or traffic-specific characteristics are needed for network selection in heterogeneous wireless networks.

There exist quite a few researches on the issue. In [7] and [8], more extensive surveys can be found. One approach is to design an optimal network selection scheme. In [1], an optimal and adaptive strategy for users with non-real time data traffic, based on a probabilistic modeling of user's characteristics, is presented. In [2], an optimal network selection of real time video streaming traffic was achieved by optimally minimizing the video frame drop ratio. In [3], an optimal session admission control scheme for multimedia traffic is presented, which uses a semi-Markov decision process (SMDP). In general, the optimal schemes are difficult to apply to the real world since they induce large computational overhead. An alternative approach is to use heuristic algorithms, such as the one proposed in [4]. It applies different static preference for different type of calls. More sophisticated heuristic schemes rely on existing mathematical tools. In [5], the analytic hierarchy process (AHP) and grey relational analysis (GRA) are used to derive the weights among various factors and to select a cell. However, using AHP has a drawback of heavy dependence on the manual assignment of relative weights among the decision factors, which is typically unclear.

In this paper, we take a more scalable and flexible approach. Our approach is based on a heuristic solution structure called Profile-based Network Selection Framework (PNSF). PNSF provides flexibility in terms of choosing the factors to be considered in network selection decision. Depending on the information availability for the decision making from the underlying wireless technologies, PNSF can be customized accordingly. In particular, on the basis of PNSF, we propose a network selection scheme for multi-service heterogeneous wireless networks. In the proposed scheme, multi-service traffic classes are handled under a single unified framework. The proposed scheme can be implemented in a distributed manner, and we address the implementation issues. The rest of the paper is organized as follows. The PNSF is described in Section II, and the proposed scheme which is designed on the basis of PNSF is presented in Section III. In Section IV, performance evaluation results are presented. The discussion of the implementation issues for 3GPP LTE networks and the conclusion are presented in Section V.

II. PROFILE-BASED NETWORK SELECTION FRAMEWORK

In PNSF, network selection is executed in two phases. At first, profiles are generated for users and cells. The user (cell) profile represents the relative status of a user (a cell) as compared to other users (cells), in such aspect as traffic demand (available capacity). Then, the best matches between users and cells are found by considering their profiles.

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2010-(C1090-1011-0004))

We organize the profiles in a hierarchy. For the purpose of explanation, let us consider only the user profile. A *factor* is an individual status parameter of a user. Factors are grouped into *categories*. A *profile* is composed of *n profile values,* each of which corresponds to a category. To represent the integral view of the status of a user, *profile space* is used, which is an *n*-dimensional space where each axis is mapped to each category. Currently, PNSF considers three categories, *traffic status*, *mobility*, and *QoS*, but more categories can be added. The categories are classified into two types, dynamic and static categories. The factors in dynamic categories are changed as a result of network selection decision while the ones in static categories are not. Let μ denote a user of interest. The list of currently accessible cells of the user μ is denoted by χ_μ, but for notational simplicity, μ will be omitted. A cell's capacity is the data throughput it can support collectively, and the available capacity, λ_{AC}, of a cell is the data throughput a cell can support after serving all the traffic of attached users.

A. Cell Profile

Cell profiles represent the relative status of the cells in χ. For cell profiling, each cell is projected on a multi-dimensional profile space, in which each axis is mapped to a profile category. For each category, one or many factors about the cell's current status are collected from all the cells in χ. The collected factor values are normalized into a range of [S_{min}, S_{max}] as shown in (1). Via normalization, only the relative magnitude relationships are kept. $p_{c,i}$ denotes the profile value of the cell i for category c. $\lambda_{c,i}$, and $\lambda_{c,\chi}$ represents the value of a factor of cell i or a set of them in χ, for category c. Min($\lambda_{c,\chi}$) and max($\lambda_{c,\chi}$) indicate the minimum and maximum values among $\lambda_{c,\chi}$. When there is more than one factor in a category, the summation of the profile values of each factor is computed while applying the weight that specifies the relationship among the factors.

$$p_{c,i} = (S_{\max} - S_{\min}) \times \frac{(\lambda_{c,i} - \min(\lambda_{c,\chi}))}{(\max(\lambda_{c,\chi}) - \min(\lambda_{c,\chi}))} + S_{\min} \qquad (1)$$

For some categories, the order of the values needs to be reversed as follows.

$$p_{c,i} = S_{\max} - (S_{\max} - S_{\min}) \times \frac{(\lambda_{c,i} - \min(\lambda_{c,\chi}))}{(\max(\lambda_{c,\chi}) - \min(\lambda_{c,\chi}))} \qquad (2)$$

B. User Profile

The user profile represents the relative status of the user μ against other users who can be connected to the cells in χ. The user profile values are also normalized by the same way, (1) and (2), from the cell profiling. The information on other users may be collected from the cells in χ or may be estimated by historical statistics.

C. Profile Matching

After the user and the cell profiles are generated, we calculate the distances between the user μ and the cells in χ after juxtaposing the user profile space and the cell profile space. To compute the distance between a user and a cell in the profile space, the *Manhattan distance metric* is used. The conventional Manhattan distance, also known as the city block

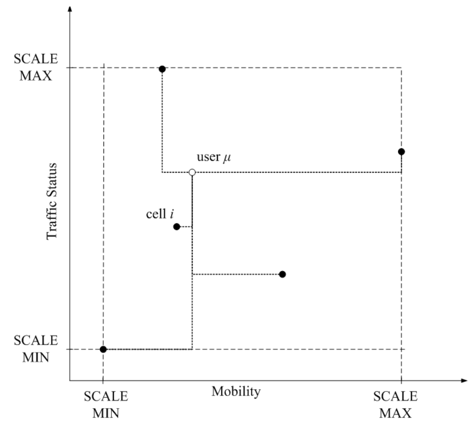


Figure 1. Manhattan profile distance

distance, is simply the sum of distances on each axis between two coordinates. The Manhattan distance is employed since it facilitates the use of a different distance evaluation method on each axis according to the category characteristics. The modified Manhattan distance metric used for profile matching allows negative distance values for both the axis distance (distance on an axis) and the final distance. This is in order to represent the directional relationship between the user profile and the cell profiles. An example of Manhattan distances from the profile matching is illustrated in Fig. 1, with only two categories are drawn for simplicity. For each axis, considering its dynamic/static property, we apply different methods to calculate the distance. On the axis of dynamic category, we favor a cell with the largest profile value in the category. The distance between the user μ and a cell is computed by subtracting the user profile value from the cell profile value. In contrast, for the static category, we favor the best-fit cell by giving the largest distance. A cell closer to the user gets a larger distance, whereas a cell farther from the user has a smaller distance value. Traffic status and QoS categories are dynamic categories and mobility category is a static category. The profile distance between the user μ and the cell i is the summation of the distances on each axis.

III. THE PROPOSED SCHEME

On the basis of PNSF, we design a network selection scheme for multi-service wireless networks. In multi-service wireless networks, both real-time (RT) traffic and non-real-time (NRT) traffic are serviced concurrently by sharing the available network resources. RT traffic, such as VoIP, requires throughput and delay guarantee, while NRT traffic, such as web browsing, does not require any guarantee in throughput or delay. In the proposing scheme, the coexistence of RT and NRT traffic users can be easily handled by making independent decision for each user according to their traffic type. The mutual interactions between two types of traffic are captured by the admission control policy and traffic scheduling policy, and the proposed scheme is independent from them. Note that the proposed scheme is one possible realization of PNSF and extendable in various ways.

A. Real-time Traffic User Case

PNSF uses two profiles, cell profile and user profile. Let us first describe the derivation of cell profile for RT traffic user. For the traffic status category of the cell profile, we use the available capacity of a cell. When the profile value of a cell is closer to S_{max} than other cells, it indicates that the cell has more capacity left for the newly coming RT users than other cells. λ_{AC} will be determined by the current traffic load level of the cell and the admission control policy used. The profile value of traffic status category, $p_{t,i}$, is computed by (1) with $\lambda_{AC,i}$ and $\lambda_{AC,\chi}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, respectively.

For the mobility category of the cell profile, we use the size of the cell's coverage. The cell size is used since it determines the frequency of handover. Fast moving users would prefer the cells with the large mobility profile value. On the contrary, slow moving users would concern less on the mobility profile value. $\lambda_{Cov,i}$ and $\lambda_{Cov,\chi}$ denotes the coverage (i.e., cell radius) of cell i and that of the cells in χ respectively. The profile value for mobility category, $p_{m,i}$ is computed by (1) with $\lambda_{Cov,i}$ and $\lambda_{Cov,\chi}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, respectively.

For the QoS category of the cell profile, we use the expected data loss by either cell overloading or handover delay. Note that we assume that the requested throughput and average speed of each RT user are known. From [6], the user's mean cell boundary crossing rate η_b and the handover rate η_h are approximately the same. We estimate the handover rate of the user μ as $\eta_{h,i,\mu} \approx \eta_{b,\mu,i}=2V_\mu/\pi r_i$, where V_μ is the average velocity of the user μ and r_i is the radius of the cell i. Then the user's expected sojourn time becomes $1/\eta_{h,i,\mu}$. From this, we can estimate the amount of data loss due to the traffic overload at the cell i, $E_{OL,i}$, where $\tau_{RDR,\mu}$ is the requested data rate of the user μ, as follows.

$$E_{OL,i} = \frac{\tau_{RDR,\mu} - \lambda_{AC,i}}{\eta_{h,i,\mu}} \tag{3}$$

The expected data loss for the user μ due to its handover to a cell i, $E_{HO,i}$ is computed by the product of $\tau_{RDR,\mu}$ and the handover delay. The total expected data loss $\lambda_{E,i}$ is the sum of $E_{OL,i}$ and $E_{HO,i}$. Optionally, RSS may be considered in computing the QoS profile value. The cell with least expected data loss should have the highest profile value. Therefore the profile value of QoS category is computed by (2) replacing factors with $\lambda_{E,i}$ and $\lambda_{E,\chi}$.

For the user profile's traffic status category, the traffic status profile value $p_{t,\mu}$ is computed by (1) with $\tau_{RDR,\mu}$ and $\tau_{SDR,\chi}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, where $\tau_{RDR,\mu}$ is the requested data rate, the traffic load the user μ is going to create, and $\tau_{SDR,\chi}$ is a set of serviced data rates of the existing users in the list χ. The mobility category of the user profile uses the velocity of the users. $\tau_{v,\mu}$ denotes the user μ's velocity, and $\tau_{v,\chi}$ is a set of velocities of the existing users attached to the cells in χ. The profile value for the user μ's mobility axis, $p_{m,\mu}$ is computed by (1) with $\tau_{v,\mu}$ and $\tau_{v,\chi}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, respectively. The QoS user profile represents the user's QoS requirement level. For RT traffic, it means how much data loss the user can tolerate. Therefore, especially for RT users, it is natural to prefer smaller value.

B. Non-Real-time Traffic User Case

We assume that NRT traffic is generated by data transfer applications where the amount of the data is known a priori. Let us first describe the derivation of cell profile for NRT traffic user. For the cell's traffic status profile, the average service data rate per user, $\lambda_{AvgSDR,i,}$, is used. All the NRT traffic will share the capacity available for the NRT traffic in a cell, which is determined by the admission control policy of the cell. For given available capacity for NRT traffic, the average service data rate per user is determined by the NRT traffic scheduling policy (e.g., round-robin) at the cell. The profile value of the cell i's traffic status, $p_{t,i}$ is computed by (1) with $\lambda_{AvgSDR,i,}$ and $\lambda_{AvgSDR,\chi,}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, respectively. $\lambda_{AvgSDR,\chi,}$ denotes a set of average serviced data rates of the cells listed in χ. The mobility cell profile for NRT traffic is identical to that for RT traffic. For the QoS cell profile, we use the ratio of the expected service data rate against the cell's capacity. This ratio is denoted by $\lambda_{SoC,i}$, which indicates the expected share of the capacity (SoC) the user will take at a cell i. The QoS cell profile of a cell i is computed by (1) replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$ with $\lambda_{SoC,i}$, $\lambda_{SoC,\chi}$ respectively. It is optional to take RSS into account.

Now let us explain the derivation of user profile for NRT traffic user. For the traffic status user profile, we use the amount of remaining data to transfer. Let us denote the remaining data sizes of the user μ and other users in χ by $\tau_{LeftFS,\mu}$ and $\tau_{LeftFS,\chi}$ respectively. The remaining data size can be thought as the traffic load that the user is going to create. The traffic status profile value of NRT user, $p_{t,\mu}$ is computed by (1) with $\tau_{LeftFS,\mu}$ and $\tau_{LeftFS,\chi}$ replacing $\lambda_{c,i}$ and $\lambda_{c,\chi}$, respectively. The mobility user profile for NRT traffic is identical to that for RT traffic. For NRT traffic, the QoS user profile means how much share of capacity that this user will occupy in the cell to be connected. This is related with the NRT traffic scheduling policy and will determine the data transfer completion time of a NRT traffic user.

IV. PERFORMANCE EVALUATION

A. Simulation Setup

Simulations were conducted on a 4km by 4km map with 261 cells. More cell parameter information is shown in Table I. The vertical handover delay was assumed 1000ms and the horizontal handover delay was assumed 100ms. The random waypoint model was selected as a mobility model. Travel speeds and pause times were uniformly distributed in the range of 1~20m/s and 1~30s, respectively. The requested data rates and file sizes were distributed from 100kbps to 300kbps and 100MB to 300MB, uniformly. While we used an abstract model for wireless link, we conducted packet-level simulations. The traces of the user movements and the network selection decisions were generated by abstract simulations. Then the generated traces were fed into *ns-2* simulator for packet level simulations. All simulation result values are averaged over 50 simulation runs while a run is 1500 seconds long. More details of the simulation setup are omitted due the space limit.

B. Simulation Results

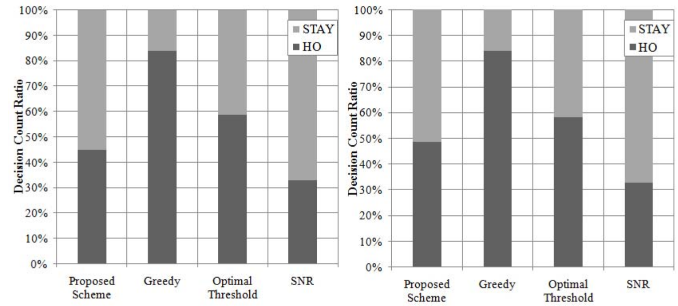
For performance comparison, three other schemes are used: *SNR-based scheme*, *Optimal Threshold scheme*, and *Greedy scheme*. In the SNR-based scheme, the cell with the strongest RSS is chosen. In the Optimal Threshold scheme, two static thresholds are used to select a cell for the user. The optimal thresholds are derived via the exhaustive search simulations. In

	LTE	*WCDMA*	*WLAN*
Number of cells	4	7	250 (10 hotspots)
Cell radius	1km	2km	100m
Cell capacity	10Mbps	4.5Mbps	6Mbps

the Greedy scheme, the cell with largest available capacity (RT) or average service data rate (NRT) is chosen.

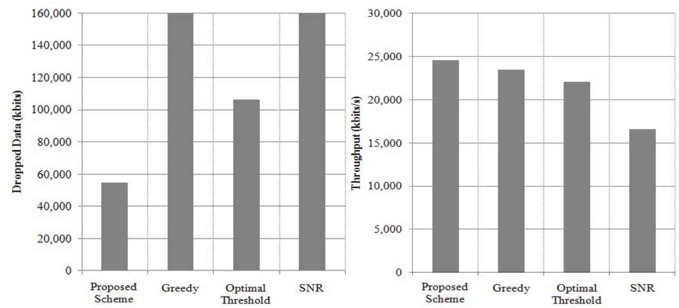
In the RT traffic only case, the proposed scheme outperforms the other three schemes at least 36% (100 users) or 12% (200 users) in terms of the data loss, caused by handover delays and overloaded cells. In NRT traffic only case, the performance impact by the frequent handover is less critical than it was in RT traffic case. The data dropped during the handover will be retransmitted, and the better throughput of the new cell can compensate the delay. Hence, the performance gains of system-wide throughput against the other schemes are less than those of the RT traffic only case (5% in case of 100 users and 8% in case of 200 users.) Due to the limit of space, we omit more details and the related figures for these cases.

The multi-service case has same number of RT and NRT users. In Fig. 2, the handover frequencies are compared for 200 RT and NRT users. The performance results for RT users and NRT users are depicted in Fig. 3. Regardless of network load level, the proposed scheme consistently shows the best performance among the four schemes under comparison. For NRT users, the performance gain of the proposed scheme remains about the same as the one achieved in the NRT traffic only case. For RT users, however, the performance gain of the proposed scheme is significantly expanded. The proposed scheme outperforms the second best scheme by reducing the loss about 45% in case of 100 users and about 49% in case of 200 users. Greedy scheme performs poorly in the multi-service networks, particularly under heavy traffic condition.



(a) RT traffic (b) NRT traffic

Figure 2. Handover frequency comparison for multi-service case (200 users)



(a) RT traffic (b) NRT traffic

Figure 3. Total data loss and system-wide throughput of RT and NRT users

V. Discussion on Implementation

Now, let us discuss how the proposed scheme can be implemented in the 3GPP LTE networks. In the LTE's handover procedure, eNB is defined to make a handover decision based on various information. Therefore it is natural to implement the proposed scheme within eNB. Since eNB is already designed to make handover decisions, most of the information required by the proposed scheme is already available in eNB. Also, most of the unavailable information can be easily acquired after minor changes to the current standard, such as an addition of messages in the user measurement report. For example, the measurement report from UE already has the accessible cell list and the actual velocity information can be incorporated into the measurement report. Also it can be replaced by the mobility level, already defined in the standard. The information such as requested data rate and serviced data rate are already available, in the form of the GBR (guaranteed bit rate) information of the user's bearer QoS information. The neighbor's status information can be gathered from neighbor cells by using the standard RRC (Radio Resource Control) protocol through X1 interface or through MME. Therefore, the proposed scheme can be implemented on LTE with minor changes to the current system.

In this paper, we proposed a scalable and flexible network selection scheme for multi-service heterogeneous wireless networks. The profiles are generated for users and cells to represent their characteristics in three categories. The best cell is determined via profile matching. The proposed scheme is practical as it has a simple but flexible structure. The network selection for multi-service traffic classes is effectively handled thanks to the flexible structure of the proposed scheme. While the optimality is not guaranteed, the computational complexity of the proposed scheme is very low, which will warrant high scalability.

References

[1] Thierry E. Klein and Seung-Jae Han, "Assignment strategies for mobile data users in hierarchical overlay networks: performance of optimal and adaptive strategies," Selected Areas in Communications, IEEE Journal on , vol.22, no.5, pp. 849-86, June 2004.

[2] Joon-Sang Park and Seung-Jae Han, "Load balancing for video streaming services in hierarchical wireless networks," Computer Networks, vol.52, issue 1, pp. 259-274., January 2008.

[3] Fei Yu and Krishnamurthy, V., "Optimal Joint Session Admission Control in Integrated WLAN and CDMA Cellular Networks with Vertical Handoff," Mobile Computing, IEEE Transactions on , vol.6, no.1, pp.126-139, Jan. 2007.

[4] Wei Song and Weihua Zhuang, "QoS provisioning via admission control in cellular/wireless LAN interworking," 2nd International Conference on Broadband Networks, vol.1, pp. 543-550, Oct. 2005.

[5] Q. Song and A. Jamalipour, "A network selection mechanism for next generation networks," in Proc. of IEEE ICC'05, Seoul, Korea, pp. 1418-1422, May 2005.

[6] Yeung, K.L. and Nanda, S., "Channel management in microcell/macrocell cellular radio systems ," Vehicular Technology, IEEE Transactions on, vol.45, no.4, pp.601-612, Nov 1996.

[7] M. Kassar, B.Kervella, G. Pujolle, "An overview of vertical handover decision strategies in heterogeneous wireless networks," Computer Communications, vol.31, Issue 10, pp. 2607-2620, 2008.

[8] X. Yan, Y. Ahmet Şekercioğlu, Sathya Narayanan, "A survey of vertical handover decision algorithms in Fourth Gneration heterogeneous wireless networks.", Computer Networks , vol.54, Issue 11, pp.1848-1863, Aug. 2008.