Formulation of Optimization Problems for Access Selection in Next Generation Wireless Networks

Anastasios Kikilis* and Angelos Rouskas
Department of Information and Communication Systems Engineering
Karlovassi, 83200
Samos,Greece
{akikilis*,arouskas}@aegean.gr

ABSTRACT

Wireless Networks are rapidly evolving, constantly increasing both in coverage and offered bandwidth, while the vision for Next Generation Wireless Networks (NGWNs) encompasses a unified core network incorporating various Radio Access Technologies (RATs) in a unified and seamless manner. In such an environment providers would like to maximize their users' satisfaction, while at the same time avoid overloading their subsystems. In this paper we identify several aspects of access selection and resource allocation in NGWNs and provide different variations of optimization problems that seem to present significant interest in this context. All problem formulations seem to be NP-Hard, and our ongoing research focuses on designing approximation techniques for solving them.

Keywords

Access Selection, Optimization, Resource Allocation, Wireless Networks

1. INTRODUCTION

Cellular Networks have developed significantly in the past decade from voice-mainly 2G systems to voice and data 3G networks, mainly suitable for high coverage but relatively low to higher bandwidth. On the other hand, Wireless LAN technologies have emerged as an omnipresent technology, offering very high bandwidth compared to cellular networks but significantly lower coverage. Apart from the aforementioned technologies, other wireless systems have also emerged, mainly focusing to special needs, such as the DVB family for multicast and broadcast communications, or the 802.16 standard.

NGWNs seem to be characterized by even higher bandwidth, along with the coexistence of various Radio Access Technologies (RATs). 3GPP has already developed a series of documents to deal with the 3G-WLAN coexistence. In such a multi-RAT environment, various issues are about to arise mainly due to the coexistence. The concept of being Always Best Connected (ABC) has arisen [3], where the notion of "Best" can be represented by the satisfaction a user gains by using the network. In ABC networks, user satisfaction becomes an important variable to successful network operation. Additionally, the market deliberation and the introduction of the number portability service have made it much easier for a user to migrate from one provider to another. As a consequence, the users have increased expectation, aiming at high QoS with less commitments. Thus, the need for consideration of the user satisfaction when studying resource management in next generation wireless networks is evident. Plain or horizontal handover is now extended to include the transition from one RAT to another, also known as "Vertical Handover".

In this paper we examine the Access or Network Selection problem, which has been given a number of approaches lately. As its name implies, this problem deals with the assignment of the most suitable RAT for each terminal. In [4] users are assigned to subsystems, in order to minimize blocking probability and at the same time maximize the system capacity, while the formulation is done according to the Online Bin-Packing Problem. In a similar context, the authors in [2] study resource allocation in the context of ABC using the Knapsack Problem formulation. The overall goal is to maximize the users' utility, while taking their preferences and satisfaction into account, through a quality-to-utility mapping. Finally, the work in [6] utilizes Game Theory to provide the users with the data rate they have requested, while it is differentiated from the previous two, because the users are supposed to be using all available network interfaces in their devices.

We first define the system model of our study and introduce several possible cases for the objective functions and the constraints of the optimization problems. We then present some interesting cases of NGWNs operation where optimization can be applied to, and formulate corresponding variations of optimization problems for access selection and resource allocation in NGWNs.

2. SYSTEM MODEL AND FORMULATION

A representation of the entities involved in the scope of this paper is depicted in Figure 1. We assume that there is a specific server responsible for collecting all neccessary measurements and reaching the required decisions, such as the Common Radio Resource Management (CRRM) entity described in [1]. Upon arrival, the users' requests are for-

---

*Correspondence Author

Mobimedia'07 Month 8, 2007, Nafpaktos, Aitolokarnania, Greece

warded to the CRRM. The optimization module that is hosted there is responsible for assigning each user to an available RAT, or to allocate different portions of his requested rate to various RATs.

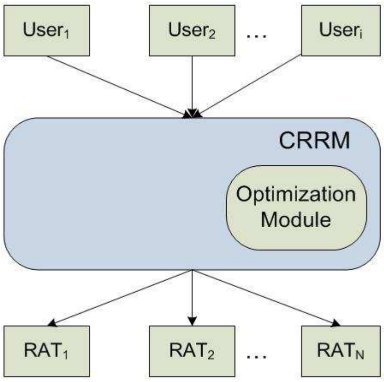

Figure 1: Optimization Procedure in CRRM

More formally, we assume one or multiple providers offering network services, through N RATs. We denote the set of RATs as:

$$S = \{RAT_1, RAT_2, \cdots, RAT_N\}$$

where each RAT has total data rate capacity of $C_j,\ j \in S$. During the operation of this system, users arrive and are allocated resources for a limited period of time, and then depart, freeing the occupied resources. We assume that each user i upon arrival declares its $R_{D,i}$, the data rate he wishes to get, along with $S_{R,i} \subseteq S$, the set of RATs the user prefers to connect to, mainly to denote those RATs that the user *does not* want to connect to. When admitted, the user may acquire a different rate from each RAT where the sum of these rates is equal to the rate assigned to this user. A $1 \times N$ allocation vector $R_{A,i} = [R_{A,i}(j)]$ denotes the rate allocated to user i in RAT j. Moreover, each user i has a level of satisfaction from the provider, represented by a Utility Function U_i.

In the next sections we will discuss and address different cases for the objective functions as well as the constraints that need to be imposed in order to subsequently formulate the optimization problems.

2.1 Objectives

2.1.1 Utility Maximization

The primary target for the provider is to maximize the total satisfaction of all users in all RATs, and that is the objective function of the optimization problem we intend to study:

$$maximize \sum_i U_i \tag{1}$$

Apart from the utility-based optimization, that can incorporate a series of parameters as we will in the following, a number of other alternative objectives can be achieved through the formulation of the optimization problems.

2.1.2 Load Balancing

A simple but important requirement for a provider is the well-balanced allocation of its resources in the various areas of its coverage. In such multi-RAT environment this can be translated into having similarly loaded RATs. This requirement can also be imposed as a constraint when using the utility-based objective function, thus providing for a utility optimization but at the same time ensuring that the solution also satisfies load balancing criteria, without overloading some RAT in order to gain utility, while leaving other RAT(s) lightly loaded.

2.1.3 Utility Balancing

Another requirement for the provider, in order to achieve a fair and balanced behavior in the case of traffic classes, is the smooth distribution of utility to the users of the same class. That means that the provider should gain approximately the same amount of utility from each user of the same class. In terms of the optimization problem this can be formulated by aiming to minimize the variation of utility among the users of the same class.

2.2 Utility Functions

A generic utility function for the problem studied here, would have the following form:

$$U_i = \sum_{j \in S_{R,i}} f_{ij} \tag{2}$$

Each RAT has a different function f_{ij} that maps various parameters of the network operation to a utility depicting the satisfaction of the user from these parameters. For example, different packet delay bounds or throughput offered to users provide different users' satisfaction levels.

In general, there exist K distinct classes of users arriving in the system, where all the users of a class behave identically. In this way, $K \times N$ functions are necessary to fully describe the way users will respond to the network parameters during their stay in the system. The index i denotes a user of class i. Two special cases are identified: (a) At one extreme, all users behave identically concerning their satisfaction from the parameters affecting utility function, i.e. the index i for distinguishing the users' utility is simply dropped. (b) At the other extreme, each user $i \in \{1, \cdots, K\}$ belongs to a different class and there are K classes available.

Parameters in Utility functions

In this multi-RAT environment, monitoring and managing the network requires keeping track of similar performance indicators in each RAT, to use them in resource allocation operations. Admission control and Load Balancing for example, could be based on measurements about the amount of traffic carried in each RAT, in order to admit a new session, or redistribute traffic respectively. However, in most cases these measurements are not directly comparable. For example, 384Kbps is the maximum rate achievable in UMTS, while this rate is easily attained in WLAN. Another similar case arises when the overload thresholds have to be calculated with respect to the RAT load. It is obvious that there is a need for a way to transform the various measurements to a common scale. On the other hand, a provider's primary goal should be to provide a service that satisfies the QoS requirements requested by the users. Utility functions are a suitable tool to depict user satisfaction from the QoS offered, with respect to some predefined QoS requirements.

As discussed earlier, user satisfaction is an important parameter for a provider, because in NGWNs users will have more options and it will be easier for them to roam to another provider if the QoS is not adequate or deteriorating.

In this work, utility functions are based on the Rate allocated, i.e. $f_{ij} = U(R_{A,i}(j))$ where f_{ij} appears in (2). However, utility functions can incorporate other parameters as well. Utility can be a function of the Signal to Interference Ratio (SIR) measured at the Mobile Terminal, the Received Signal Strength (RSS), or the delay experienced by the received packets. Utility functions can also depend on the monetary cost of the service, since it can be a significant parameter influencing users' decision on favoring or leaving a provider.

2.3 Constraints

The primary constraint of provider is capacity limits expressed in data rate. The provider cannot allocate more data rate in RAT j than the RAT total capacity C_j:

$$\sum_i R_{A,i}(j) \leq C_j, \quad \forall j \in S_{R,i} \tag{3}$$

As with the utility function, we can also distinguish special cases according to the relation between $R_{D,i}$ and $R_{A,i}(j)$:

2.3.1 Non-Elastic Traffic

In this case, only the rate requested can be assigned to users:

$$R_{D,i} = \sum_{j \in S_{R,i}} R_{A,i}(j) \tag{4}$$

We can formulate $R_{A,i}$ as a $K \times N$ matrix, or $1 \times N$ vector respectively, according to the categorization described in the previous section

2.3.2 Elastic Traffic

In this case, we allow the provider to assign less rate than requested, possibly admitting users that could not be admitted otherwise:

$$R_{D,i} \neq \sum_{j \in S_{R,i}} R_{A,i}(j)$$

. This case occurs when users use applications that can function even when the rate requirements are not met. Upon each connection request, the user rate limits $R_{D,i}^{MIN}$ and $R_{D,i}^{MAX}$ are defined. In this situation, utility has to be a function and not just a matrix, so as to capture the reduced satisfaction, resulting from the decrease in rate allocation. Additionally, we have to ensure that the rate allocation assigns rates within the predefined limits: Therefore, we should additionally enforce the respective constraints on rate:

$$R_{D,i}^{MIN} \leq \sum_j R_{A,i}(j) \leq R_{D,i}^{MAX} \tag{5}$$

Another constraint that can be applied is that the user still benefits from the service, which can be achieved through the utility function:

$$U_i \geq 0 \tag{6}$$

2.3.3 Single - Multiple Interfaces Open

In an environment where multiple RATs are present in the area, and users being able to sense them all, two cases arise:

Single Interface : In this case, even though the user's terminal is capable of sensing all RATs, it is not capable of transmitting and receiving in more than one interface at the same time. This practically reduces $R_{A,i}$ to a single value, and in order for our optimization problem to be complete, we need to add another series of constraints:

$$\sum_{j \in S_{R,i}} x_{ij} \leq 1, \qquad \forall i \tag{7}$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } R_{A,i}(j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Multiple Interface : In this case, we assume that users not only possess multi-mode terminals, but additionally that these terminals are capable of sending or receiving data using more than one interface simultaneously. If we assume that terminals can communicate with up to r RATs simultaneously ($r \leq N$), (7) is replaced by:

$$\sum_{j \in S_{R,i}} x_{ij} \leq r, \qquad \forall i \tag{8}$$

3. INTERESTING OPTIMIZATION PROBLEMS

In the following we describe the formulation of optimization problems for access selection and resource allocation.

3.1 Batch Processing for Network Selection

We partition time into identical time periods, $(t_{n-1}, t_n]$ being the n-th time interval. We assume that during the n-th time period, a_n users arrive into the system, submitting their service requests, along with their respective $S_{R,i}$ and $R_{D,i}$. i, similarly to the previous formulation, denotes the i-th user of this time interval, $i \in \{1, \cdots, a_n\}$. We assume however, that the requests are not examined before the end of the time period, t_n. We also assume that d_n users leave the system during the last time period, and their data rates have been added to the available capacity of each RAT system. The available capacities at time t_n are C_j^A , $j \in S$. At t_n the admission algorithm is executed to decide which users should be assigned to each RAT, so that Utility gained is maximized, while the allocation does not exceed the available capacities of the RATs.

In this case of Network Selection, only users with non-elastic traffic in the single-interface-open scenario are considered, and adding the respective constraints discussed in the previous section yields the following problem:

$$maximize \sum_{i=1}^{a_n} U_i$$

subject to (4),(7), and a modified (3):

$$\sum_{i=1}^{a_n} x_{ij} \cdot R_{A,i}(j) \leq C_j^A, \quad j = 1, \ldots, N \tag{9}$$

3.2 Multi-Interface Elastic Traffic

In this case, we assume that the terminals have the capability of simultaneous transmission and reception in different

RATs and also assume that the traffic is elastic. Thus, the maximization problem of the previous section becomes:

$$maximize \sum_{i=1}^{a_n} U_i$$

subject to (5),(8),and (9).

3.3 Load Balancing

In this case, we keep all assumptions made in the previous section, about partitioning time, and batch processing of requests. However, we assume that the provider wants to enforce specific levels of load in its RATs. This can be realized in two different ways.

In the first one, the provider wishes for a specific RAT's remaining capacity C_j^A to remain within predefined limits. Such a case imposes different constraints on a specific RATs capacity. For example, if a provider wishes to have $\alpha\%$ of RAT k capacity always available, where $k \in S$, this means that the rates allocated to RAT k must be less than this threshold, and the optimization problem would subtitute (9) for:

$$\sum_{i=1}^{a_n} x_{ik} \cdot R_{A,i}(k) \leq C_j^A - \alpha \times C_k$$

assuming that $C_j^A - \alpha \times C_k > 0$. Otherwise, the user cannot be allocated to RAT k.

In the second case, the provider wishes to enforce a specific distribution for the available capacities in its system. Such a distribution could be that the available capacity in each RAT every moment, compared to the RAT's total capacity should be virtually the same. Such a request, as with the previous case, has an effect to the rates allocated in each RAT and is formed like:

$$|LF_k - LF_l| \leq \beta, \quad \forall k, l \in S$$

where β is a constant denoting how close the allocations should be to one another, and

$$LF_j = \frac{C_j - C_j^A + \sum_{i=1}^{a_n} x_{ij} \cdot R_{A,i}(j)}{C_j}$$

is the Load Factor of RAT j.

3.4 User-Oriented Access Selection

In all previous cases, the resource allocation and access - network selection was a network driven procedure. According to that scenario, the provider is capable of collecting every measurement needed, and does not have any constraints on the processing power required for the access selection procedure processing. When user satisfaction is considered however, user-oriented access selection becomes an option. Terminal or user driven access selection, significantly lowers the processing requests on the provider side along with the handover signalling overhead, while at the same time allowing for a distributed scheme for access selection. An interesting case of User-Oriented Access Selection occurs when we examine the "Multi-Interface Elastic Traffic" case. The associated problem is a specialization of the problem in Section 3.2 where $a_n = 1$, whereas in this case, as expected due to the distributed nature of the approach, the complexity of the problem is much lower.

4. CONCLUSIONS

In this paper, we identified a number of optimization problems in the area of access selection and resource allocation in NGWNs. We analyzed the parts constituting such optimization problems , along with the possible variations in the kind of parameters involved, and we combined this variations in some interesting cases. All these problems are quite similar to well-known NP-hard problems such as the Knapsack and the Generalized Assignment Problem [5], so it seems that they will also be NP-hard, and our future work will target in devising heuristic algorithms for them.

5. ACKNOWLEDGMENTS

The paper describes work undertaken in the context of the Specific Targeted Research Project (STREP) project supported by the European 6th Framework Programme, with the contract number: 026906 and name UNITE, which stands for 'Virtual Distributed Testbed for Optimisation and Coexistence of Heterogeneous Systems', see www.ist-unite.org

6. REFERENCES

[1] 3GPP. Improvement of RRM across RNS and RNS/BSS. ts 25.981 v0.3.0 - release 6. Technical report, February 2003.

[2] V. Gazis, N. Houssos, N. Alonistioti, and L. Merakos. On the complexity of 'always best connected' in 4g mobile networks. *IEEE 58th Vehicular Technology Conference, 2003*, 4:2312–2316, October 2003.

[3] E. Gustafsson and A. Jonsson. Always best connected. *IEEE Wireless Communications*, 10(1):49–55, Feb. 2003.

[4] D. Mariz, I. Cananea, D. Sadok, and G. Fodor. Simulative analysis of access selection algorithms for multi-access networks. In *International Symposium on a World of Wireless, Mobile and Multimedia Networks, 2006*, 2006.

[5] S. Martello and P. Toth. *Knapsack problems: algorithms and computer implementations.* John Wiley & Sons, Inc., New York, NY, USA, 1990.

[6] D. Niyato and E. Hossain. A cooperative game framework for bandwidth allocation in 4g heterogeneous wireless networks. In *IEEE International Conference on Communications*, pages 4357 – 4362, June 2006.