

Issues in the Design of Future Mobile Communication Systems

P.P. Demestichas, E.C.Tzifa, V.P. Demesticha, M.D.Louta, D. Karagiannis, M.E.Theologou

National Technical University of Athens, Computer Science Division,
9 Heroon Polytechniou Str, Zographou 15773 Athens, GREECE
Tel: + 30 1 772 2529, Fax: + 30 1 772 2534
email: pdemest@cc.ece.ntua.gr

Abstract : The success of third-generation mobile communication systems, e.g., UMTS, will depend on the Quality of Service (QoS) that the mobile users will perceive, and whether the QoS levels will be provided in a cost efficient manner. In this perspective, the importance of tools for “proper” planning of a mobile system becomes straightforward. In the context of this paper, we present such planning tools related to the design of future mobile communication systems. In addition, we present a platform that enables the assessment of the derived solutions by means of simulation.

I. INTRODUCTION

Future mobile communications systems, e.g., UMTS [1] or the future versions of GSM [2], have promised to offer a wide variety of over the widest possible service area. From the viewpoint of the users, the success of these systems will depend on the Quality of Service (QoS) that they will provide, and especially, on whether it will be comparable to that provided by fixed systems. From the providers perspective, the aim will be to provide QoS in the most cost efficient manner. Other important objective in the design of future mobile systems is introducing them by minimally impacting the existing fixed communication infrastructures, and by gradually altering (upgrading) legacy systems. It is believed that planning tools [3,4] may play a significant role in meeting these challenges. Such tools enable the operators to design from scratch (“green-field” scenario), or to expand and modify, their systems, by providing solutions to difficult combinatorial problems.

The starting point in this paper is the presentation of an overall planning process. The main focus, however, will be to discuss about advanced planning problem versions that may be incorporated in the process. This incorporation may be seen as a step towards the migration from legacy to third generation mobile systems. Before proceeding in this discussion we provide some basic assumptions. The system architecture may be depicted as in Figure 1.

Essential, is the separation of the architecture into the following three segments: (a) the *core network* segment that provides the switching and transmission functions required by the mobile system, (b) the *intelligent network* segment that comprises the logic that enables the provision of services to mobile users, and (c), the *access network* segment that enables interworking between the mobile unit and the fixed network. In this paper we will discuss about advanced planning problems of the access and intelligent segments. Our

reference system is UMTS. However, the practices proposed hereafter may be equally applicable to any other mobile system of the future.

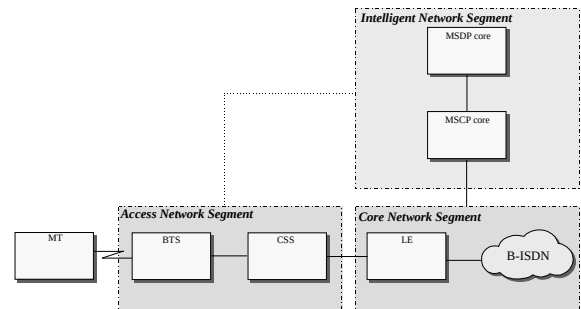


Figure 1. Reference configuration of a third generation mobile system

The network elements in the access segment (Figure 1) are the *Base Transceiver Stations* (BTSs), which provide radio link management, and the *Cell Site Switches* (CSSs), which provide switching functionality, as well as connection and call control. Interworking with the core network is done through the *Local Exchanges* (LEs), which provide the switching infrastructure that will support UMTS. The network elements in the intelligent segment are the *Mobility and Services Control Points* (MSCPs) and the *Mobility and Services Data Points* (MSDPs). MSCPs comprise the functionality for supporting service access, continuity, and mobility. The MSDPs are nodes of the system database, and hence, they contain the data needed for supporting the UMTS services.

The overall planning process and the pertinent tool follow the architecture separation presented above (Figure 2). The advanced problem versions addressed in this paper are related to the *Access* and *Intelligent Network Planning* components, and will be presented in section 0 and 0, respectively. The third component (*Simulator*) complements and validates the design of the previous two components and is presented in [5,6].

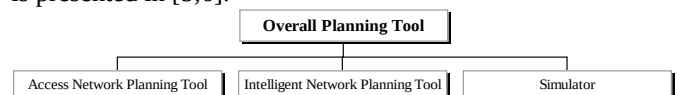


Figure 2. Structure of the overall-planning tool

II. ACCESS NETWORK SEGMENT

We assume that the access segment may be designed in the following four phases (Figure 3). The first aims at

finding the proper BTS deployment pattern [7]. The second is targeted to the efficient utilisation of the available radio spectrum [8,9]. The third is concerned with the grouping of cells into location and paging areas [10]. The last one aims at finding the proper CSS and LE deployment pattern given the BTS layout. In this paper we propose problems related to the last three phases.

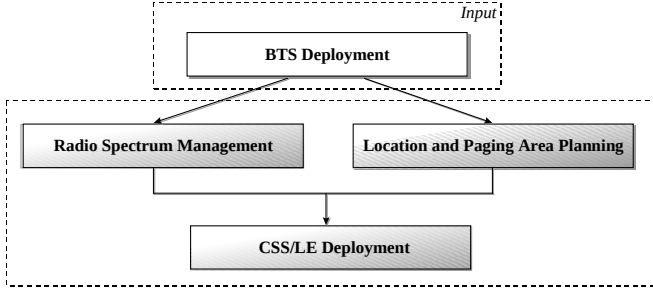


Figure 3. Problem areas in the access network segment. The shaded box depicts the problems addressed in this paper

A. Traffic adaptive spectrum allocation

In general, the *basic* version of the spectrum allocation problem applied that is applied in legacy cellular systems, expresses our aim to find the best possible allocation of frequencies to cells for a given distribution of cells and traffic load. The problem may be generally stated as follows: “Given the cell layout (e.g., in the form of a graph $G(V, E)$; V is the set of cells; edges in E connect neighbouring cells), the interference conditions in the system (e.g., in the form of a compatibility matrix I [11]), the available radio spectrum (collection of frequencies F) and the traffic and mobility volume in each cell (e.g., vector B ; each element being the load in a particular cell), find the best allocation of frequencies to cells (e.g., vector A ; each element being the set of frequencies in a particular cell), i.e., the one that optimises a given quality criterion (objective function $C(A)$), and at the same time preserves the constraints that derive from the traffic and interference conditions”. The quality criterion considered in this paper is the maximisation of the traffic that is carried by the system. However, our schemes can readily be combined with other (perhaps more elaborate or complex) criteria.

A limitation of the problem above is that it is targeted to a fixed traffic condition. Therefore, spectrum allocation can only be based on worst case estimates, which may lead to inefficient utilisation. An interesting alternative is to exploit the fact that the traffic demand is time-variant. Let's assume that it consists of a set of load vectors, each of which is valid during a particular time zone of a day. In this case, an alternative is to handle each of these vectors, and when the demand changes, to reconfigure the frequency allocation so as to adapt to the traffic variation. This, however, requires an *extension* to the basic spectrum allocation problem.

An additional objective in the extended spectrum allocation problem (handling time-variant traffic loads

through successive reconfigurations of allocation pattern) may be the acquisition of each new allocation $\tilde{A}$ by minimally disturbing the existing (already established) one A . This goal is complementary to the maximisation of the total carried traffic. The practical meaning of the new requirement is that the network should be charged with the minimum possible management effort. Hence, every new allocation should be implemented by performing the minimum number of alterations in the already existing allocation of frequencies to cells. The *extended* problem can be generally stated as follows: “Given the cell layout (graph $G(V, E)$), the interference conditions in the system (compatibility matrix I), the available radio spectrum (collection of frequencies F), the traffic and mobility volume in the current time-zone (vector $\tilde{B}$), and an existing allocation of frequencies to cells A (which was optimal for the previous time-zone), find a new best allocation of frequencies to cells $\tilde{A}$ that optimises the quality criterion (objective function $C(\tilde{A})$), by performing the minimum number of alterations on the already established allocation.

The problem above may be solved by a two-phase procedure. The first phase aims at the optimisation of the quality criterion, and hence, may be seen as an instance of the basic spectrum allocation problem. The second phase harmonises the allocation that derived from the first phase with the allocation that was established in the previous time zone. The objective is to find an allocation that is also optimal with respect to the objective function, and furthermore, may be obtained by minimally changing A . Simple instances of the second phase may be optimally formulated as a 0-1 quadratic programming problem (the formulations are omitted for brevity; they may be provided if requested) [9].

Computationally efficient solutions are based on the following ideas. Any known solution can be applied in the first phase. An approximate solution to the second phase may be based on an appropriate instance of the maximum flow-minimum cut problem (more details may be provided if requested).

B. Location/Paging Area Planning Tool

The location update and paging procedures [10] impose a significant signalling load on the system, which is also transferred over the radio interface (a scarce resource). Location and paging area planning are methods for controlling this load.

The basic version of the location area-planning problem may be described as follows. We assume that we are given the cell structure of the system, the crossing rates among neighbouring cells (i.e., the number of mobile users that move from a certain cell to a neighbouring one, per time unit), and the rate of incoming calls per cell, i.e., the number of calls per time unit that are directed to users located in the cell (which may be obtained by the expected number of users

in the cell, and the rate of incoming calls per user). The crossings among cells that belong to different location areas result to location updates. The incoming calls to a certain cell result, in the worst case, to paging messages that are broadcasted in all the cells of the location area. The objective of our problem is to find the partition of cells to location areas that results to the minimisation of a cost function, associated with the number of location updates and paging requests (that are caused by the partitioning) and the cost of each type of message. A constraint of the problem is that each cell should be placed in one location area.

The problem above has the same drawback as the basic version of the spectrum allocation problem. More specifically, it is targeted to a fixed mobility and traffic condition. In principle, the adaptation of the location area layout to the changing traffic and mobility pattern will cause excessive signalling at the time of the transition. Hence, reconfigurations are practical only if the location area identifiers (of the new layout) derive from the application of special filtering mechanisms (Figure 4). The filtering process may be formulated as an instance of the weighted matching problem.

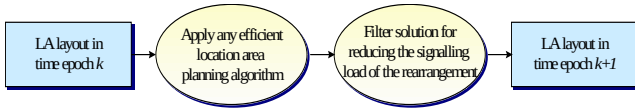


Figure 4. Approach for location area layout reconfigurations

Location area planning is the first method for controlling the signalling load that is due to location update and paging operations. A step further is to apply a multiple step-paging scheme, in order to further reduce the paging load. However, these schemes require paging area planning. In the sequel, we describe a version of the paging area planning problem, under the assumption that the paging scheme applied falls in the last interaction based paging (LIBP) category [12]. The scheme consists of two steps. During the first step the paging area selection is based on the cell of the location area from which the last interaction (namely, location update or call set-up) among the system and the terminal took place, and the mobility and traffic characteristics of the terminal class (to which the paged terminal belongs). In case the first step proves unsuccessful, the paging area in the second step is the complementary portion of the location area, which was excluded in the previous step.

An arbitrary location area l is described by a graph $G(V_l, E_l)$ (V_l comprises the cells; each edge of E_l connects neighbouring cells). An allocation of cells to paging areas $P(l) = \{P_u \mid \forall u \in V_l\}$, ($P_u \subseteq V_l$) is sought. Each set P_u ($u \in V_l$) comprises the cells of location area l in which a terminal (of a given class with known characteristics) should be paged, given that its last interaction with the system took place in cell $u \in V_l$. In other words, the paging area P_u should comprise the cells of the location area in which the

terminal is most likely to be found given the characteristics of the class in which it belongs, and that it was last seen in u .

The objective of the paging area planning problem is to have each P_u ($u \in V_l$, $P_u \subseteq V_l$) as small as possible, since this increases the possible savings in the paging load. The constraint of the problem refers to the probability of successfully paging the terminal in the selected subset of the location area, $p_s(P_u)$, which should exceed a given (by the operator or designer as input to the problem) threshold $p_{\min}$.

An approach for solving this problem is depicted in Figure 5. It builds on the assumption that $p_s(P_u)$ is the sum of probabilities $p_f(u, v)$, where $v \in P_u$. Probability $p_f(u, v)$ expresses the likelihood that the terminal is currently located in (i.e., will respond to the paging message from) cell $v \in V_l$, given that its last interaction was in cell $u \in V_l$. Some factors that may affect these probabilities are the following. Firstly, the structure of l , e.g., the set of possible routes in the location area that the terminal may have followed after its last interaction with the system. Secondly, the terminal (user) behaviour, which in general, may be characterised in terms of the (incoming and outgoing) call (traffic) volume and the mobility levels. These aspects may be exploited so as to determine the probabilities $p_f(u, v)$. In the sequel, cells $v \in V_l$ may be appended to P_u until the condition $p_s(P_u) \geq p_{\min}$ is met.

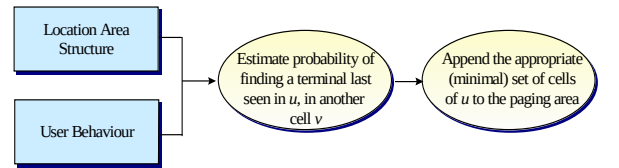


Figure 5. Approach for obtaining the paging area layout

C. CSS/LE deployment

The aim in this phase of the access segment design is finding the minimum cost deployment pattern of CSSs and LEs that satisfies a set of performance constraints. Due to space limitations we will limit our attention to the allocation of BTSs to CSSs. A general problem statement is the following. Given the BTS layout, the load originating from each BTS, the crossing rates among neighbouring BTSs, a set of candidate CSS sites, the cost of each CSS site, and the cost of connecting (cabling) a BTS with a CSS, find the minimum cost allocation of BTSs to CSSs (in terms of the number of CSSs that need to be deployed, the cost of inter-connecting BTSs to CSSs, and the of the handovers among BTSs that are connected to different CSSs [13]), subject to a set of constraints (each BTS connected to a CSS, and preservation of the capacity constraints of the CSSs).

The problem may be formulated as a 0-1 linear programming problem. An efficient solution may be based on

the *simulated annealing* technique [14]. Important extensions to the problem is the inclusion of various backbone technologies (e.g., ATM [15] or Internet).

III. INTELLIGENT NETWORK SEGMENT

Two problems encountered in this phase. First, the association of LE areas to MSCP areas (MSCP level problem), and second, the association of MSCP areas to MSDP areas (MSDP level problem). In the rest of this section we describe principles associated with the design of these problems. Our presentation abstracts from legacy technologies (e.g., current SCP technological status, or database technologies, i.e., two-tier VLR/HLR [2] or hierarchical [1])

At the MSCP level the objective is to find the minimum cost allocation of LE areas to MSCP areas, subject to a set of constraints. We assume that an MSCP may span over a set of LE areas. Factors affecting the overall cost are, the number of MSCPs that need to be deployed, the cost of connecting an LE to an MSCP (classically this is known as the cabling cost), and the cost of connecting an LE pair to different MSCPs. In order to model the latter factor, we use the traffic attraction volume $k_L(i, i')$, indicating for each LE area pair (i, i') the volume of connection requests that are generated in area i , and are directed towards the area i' . One of our objectives is to favour the connection of "related" LE areas under the same MSCP area. The constraints of the problem derive from the capabilities of the MSCPs, which are related to the amount of signalling load that may be handled (while keeping performance within the acceptable range), and probably, the maximum number of LEs they can serve. Hence, a general problem statement is the following. Given a set of LEs, the load originating from each LE and the traffic attraction volume among the LE pairs, a set of MSCPs, and their capabilities, find the minimum cost allocation of LEs to MSCPs, subject to the capacity constraints associated to each MSCP, and the overall performance requirements.

The problem at the MSDP level bears significant resemblance to the problem at the MSCP level. More specifically, an MSDP is a node of the UMTS distributed database. An MSDP may span over a set of MSCP areas, while each MSCP is controlled by one MSDP node. The objective is to find the minimum cost interconnection of MSCPs to MSDPs. In analogy to the MSCP level case, the three factors affecting the overall cost are, the number of MSDPs deployed, the cost of connecting an MSCP to an MSDP (cabling cost), and the cost of connecting an MSCP pair to different MSDPs. However, at the MSDP level the latter factor should not merely refer to the traffic attraction volume among MSCP areas. Another factor that should be incorporated are the user movements among MSCP areas that are controlled by different MSDPs. This is further explained in the sequel, and it is the point where the problems at the MSCP and the MSDP level are differentiated.

An MSDP distinguishes two major user categories. First, the resident users, i.e., those that have their data permanently stored in the respective node. Second, the visiting users, i.e., those passing through the area controlled by the MSDP node (and are resident somewhere else). A general performance criterion in database theory, namely locality of reference, implies that MSDP nodes should also contain data for the visiting users. Hence, an additional objective of the optimisation at the MSDP level, is to relate MSCP to MSDP areas, taking into account the mobility attraction volume of the MSCP areas (in addition to the traffic attraction volume among MSCP areas). This approach has the following advantage. It reduces the number of visiting users per MSDP area, and consequently, the data movements among the MSDP areas.

We assume that each MSCP area is associated with a number of users that will be resident in the controlling MSDP node. In order to reflect mobility (user movements among various areas) we define the mobility attraction volume $a_C(i, i')$ for each MSCP area pair (i, i') . This measure indicates the volume of movements of the resident users of MSCP area i towards MSCP area i' . Based on the above a general problem statement may be the following. Given a set of MSCPs, the number of resident users and the mobility and traffic attraction volumes among the MSCP area pairs, a set of MSDPs, and their capabilities, find the minimum cost allocation of MSCPs to MSDPs, subject to a set of capacity and performance constraints.

IV. DISCUSSION AND CONCLUSIONS

In the context of this paper, we defined and described planning tools for the solutions of some important design problems related to the design of the access and intelligent network segments of a future mobile network. In an extended version of this study we may provide the following. First, details on the optimal formulations or the computationally efficient, heuristic solutions, that solve the problems identified herein. Second, the overall framework (software tool) that comprises the practical solutions of the access and intelligent network segment planning processes. Third, results from comparative experiments, that will include the solutions obtained by the optimal formulations, the simulation platform and the heuristic techniques.

REFERENCES

- [1] J. Rapeli, "UMTS: Targets, system concept, and standardisation in a global framework", *IEEE Personal Commun.*, Vol. 2, No. 1, pp 20-28, Feb. 1995
- [2] M. Mouly, M.-B. Pautet, "*The GSM system for mobile communications*", published by the authors, Palaiseau, France, 1992
- [3] S. Dravida, H. Jiang, M. Kodialam, B. Samadi, Y. Wang, "Narrowband and broadband infrastructure design for wireless networks", *IEEE Commun. Mag.*, Vol. 36, No. 5, May 1998
- [4] E. Berruto, M. Gudmundson, R. Menolascino, W. Mohr, M. Pizarroso, "Research activities, on UMTS radio interface, network architectures and planning", *IEEE Commun. Mag.*, Vol. 36, No. 2, Feb. 1998

- [5] P.P.Demestichas, C.Bradley, V.P.Demesticha, D. Karagiannis, D.Mirra, F.Pascali, M. C. Ronchini, "Tools for Designing Third Generation Mobile Networks with Optimised Performance and QoS", in Proc. 3rd ACTS Mobile Communications Summit 98, Rhodes, Greece, June 1998
- [6] D.Mirra, F.Pascali, V.P.Demesticha, D.Karagiannis, P.P.Demestichas, M.E.Theologou, "Network performance assessment platform for third generation mobile systems", In Proc. IEEE International Conference on Universal Personal Communications (ICUPC'98), Florence, Italy, Oct. 1998
- [7] Q. Hao, B.-H. Soomg, E. Gunawan, J.-T. Ong, C.-B. Soh, Z. Li, "A low-cost cellular mobile communication system: A hierarchical optimisation network resource planning approach", IEEE J. Select. Areas Commun., Vol. 15, No. 7 Sept. 1997
- [8] I. Katzela, M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunications systems: A comprehensive study", IEEE Personal Commun., Vol. 3, No. 3, pp.10-31, June 1996
- [9] M.G. Kazantzakis, P.P. Demestichas, E.C. Tzifa, M.E. Anagnostou, "Optimal adaptive spectrum utilisation in cellular communications systems", accepted for publication in the IEEE Trans. on Commun.
- [10] B. Samadi, W.S. Wong, "Optimisation techniques for location area partitioning", 8th ITC Specialist Seminar on Universal Personal Telecommunications, Geneva, Oct. 1992
- [11] K.N. Sivarajian, R.J. McEliece, J.W. Ketchum, "Channel assignment in cellular radio", in Proc. IEEE Trans. Veh. Technol., 1989
- [12] J.G. Markoulidakis, E.D. Sykas, "Performance bounds of a multiple-step paging strategy in future universal mobile telecommunication systems", Int. J. of Wireless Information Networks, Vol. 2, No. 3, 1995
- [13] A. Merchant, B. Sengupta, "Assignment of cells to switches in PCS networks", IEEE/ACM Trans. on Networking, Vol.3, No.5, Oct 1995
- [14] E. Aarts, J. Korts, "Simulated annealing and the Boltzmann machines", New York: Wiley, 1989
- [15] F.Badini, P.P.Demestichas, V.P.Demesticha, R. Menolascino, "Design of an ATM-based access segment for third generation mobile systems", In Proc. IEEE International Conference on Universal Personal Communications (ICUPC'98), Florence, Italy, Oct. 1998