Optimization of the Assignment of Cells to Packet Control Units in GERAN

M. Toril, V. Wille and R. Barco

***Abstract*— In GSM EDGE Radio Access Network (GERAN), the assignment of cells to packet control units within a base station controller is of key importance to optimize the quality of service experienced by end users. In this paper, a novel methodology is presented to optimize this assignment by means of graph partitioning theory. The proposed algorithm aims at minimizing the number of cell re-selections between cells on different packet control units based on handover information. Results from the application of the method in a live network show its capability to improve end-user performance in terms of service delay and application throughput.**

I. INTRODUCTION

Over the past years, operators of GERAN-based networks have been actively launching a number of packet-data based services. To offer such services, several network elements must be added to an existing GSM network, among which are the packet control units (PCU) in the base station controllers (BSC). In addition, for a cell to offer such services, it has to be associated with a PCU.

The user experience of these services is strongly linked to the data rates offered by the network and the delays experienced by the applications. In current networks, application delays are mainly due to the change of serving cell. This event is called cell re-selection (CR) and is the counterpart of the handover experienced by circuit-switched (CS) traffic. Trial results in live networks have shown that the service break associated with CR ranges from 2 to 8 seconds [1]. The duration of the service break strongly depends on whether the CR occurs within a PCU or between PCUs. Long service breaks are associated with inter-PCU CR, while shorter service breaks are associated with intra-PCU CR. A shorter break in the intra-PCU CR is possible because data for the mobile is buffered in the PCU and transmission can be resumed as soon as the mobile is ready to receive additional data. By contrast, in the inter-PCU CR case, the referred data has to be moved to the new PCU, before resuming the data transmission. Because of this difference, it is important to ensure that cells that are closely ”connected” are assigned to the same PCU, i.e. the cell-to-PCU assignment process should minimize the number of inter-PCU CR. At the same time, the PCU plan should ensure that the load of all PCUs is within certain limits, i.e. not all cells can be connected to the same PCU.

Manuscript received December 13, 2004. The associate editor coordinating the review of the paper and approving it for publication was Prof. Christos Douligeris. This research was partially supported by TIC2003-07827 grant from the Spanish Ministry of Science and Technology.

M. Toril and R. Barco are with the Dpt. of Communications Engineering, University of Málaga, Spain (email: mtoril@ic.uma.es, rbm@ic.uma.es).

V. Wille is with Nokia Networks, Performance Services, Huntingdon, United Kingdom (email: volker.wille@nokia.com).

The Cell-to-PCU Assignment (CPA) problem can be formulated as a graph partitioning problem [2]. Although most formulations of the graph partitioning problem are known to be NP-complete, several heuristics have been proposed to find near optimal solutions [3][4][5]. In this work, a graph partitioning solution is proposed for the CPA optimization problem. The proposed algorithm provides for each PCU a cluster of cells, which is connected and bounded in size. As main contribution of this paper, the PCU planning problem is first formulated as a graph partitioning problem. Likewise, a novel assessment methodology based on drive test data is proposed to determine the impact on end-user performance. Section II introduces the PCU planning problem and the proposed methodology, while section III provides a discussion of the results obtained from a trial in a live network.

II. THE CELL-TO-PCU ASSIGNMENT PROBLEM

A. Formulation of the problem

To solve the CPA problem, the network area under optimization (i.e. a BSC) can be modeled as an undirected graph with weighted vertices and edges. The vertices of the graph represent the cells within a BSC, while the undirected edges between them represent the adjacencies defined by the operator for cell re-selection (CR) purposes. The weight associated to each vertex stands for its contribution to the load of the PCU in terms of the number of time slots (TSL) devoted to General Packet Radio Service (GPRS) traffic. In this work, the number of GPRS TSLs assigned to a PCU is limited to 256 by the hardware capabilities of the PCU. The weight of each edge denotes the number of users that experience a CR between the cells on their sides. Over this graph, the CPA problem can be formulated as a *bounded, connected, min-k cut* problem [3]. This problem calls for the grouping of vertices into a fixed number of disjoint subsets (or subdomains), such that the sum of the weights of the edges that connect vertices in different subdomains (referred as edge-cut) is minimum, provided that the sum of vertex weight in every subdomain is below a certain limit. It is worth noting that the exact formulation applied differed from the classical formulation due to the fact that connected partitions were required and a new constraint, the imbalance between subdomain weights, was added.

B. Current Solution Technique

The assignment of cells to PCU is currently a manual process, which is performed when the equipment associated with packet-data services is introduced into the network. During the operational stage, subsequent increments of GPRS TSLs on cells can lead to the extension of the number of

PCUs in a BSC. Considering the size of existing networks, comprising about 25000 cells and 500000 adjacencies, the creation of an optimal PCU plan is a very difficult and time-consuming process. Analysis of several networks has shown that sub-optimal solutions are commonly adopted for the sake of simplicity. This indicates the need for an automated method that provides a better solution to this complex problem.

C. Novel Methodology

The proposed methodology is the first time to apply graph partitioning theory to the CPA problem. An initial solution to the problem is computed by multilevel (ML) graph partitioning techniques [4]. Unlike traditional graph partitioning methods, ML techniques first coarsen the graph by collapsing vertices and edges, thus reducing the size of the graph. Rough solutions are efficiently computed on smaller versions of the graph and later uncoarsened to obtain the partition of the original graph. The solution proposed compares favorably in terms of quality (i.e. edge-cut) and run time. Concretely, the time complexity of the k-way ML partitioning algorithm used in this work is linear in the number of edges.

The classical ML approach aims at a perfect balance among subdomain weights (i.e. load of PCU) at the expense of increasing the edge-cut (i.e. inter-PCU CR). As perfect balance of subdomains is not necessary in the CPA problem, a refinement technique is subsequently used to enhance the quality of the solution. The refinement process is based on the variant of the Kernighan-Lin heuristic proposed in [5], whose time complexity is also linear in the number of edges. This greedy algorithm consists of a number of passes through the vertices, where the edge-cut reduction of every possible move of a vertex is computed and ranked by benefit. The top vertex that still maintains the load constraints is moved.

In theory, the smallest network entity assignable to a PCU is a single cell. However, there are several advantages from assigning all cells of a site to the same PCU. The validity of a site-based solution is easier to check in a map by maintenance personnel. Likewise, in a site-based solution, the combination of geographic and mobility issues increases the robustness of the solution against changes in the mobility and propagation environment, thus extending its validity period. Obviously, these benefits are obtained at the expense of a slight increase in edge-cut when compared to a cell-based solution.

D. Limitations of the proposed method

The weight of each edge in the graph should denote the number of CRs between the cells. Unfortunately, such information is not available for packet-switched (PS) services in the Network Management System (NMS). In the absence of such statistics, it is deemed most suitable to utilize HO statistics related to CS services. Assuming that the mobility of users in PS mode is similar to users in CS mode, the errors in such analysis should remain relatively small.

III. Field Trial

A. Trial Set-up

The above-described method was applied to a public GERAN-based network. The trial area consisted of one BSC with 139 cells. Based on CS-HO statistics from the previous nine days, a new PCU plan was created for the trial BSC. The execution time was about 30s on a 2.8GHz 1GByte-RAM laptop. Performance statistics were gathered before and after the new PCU plan was implemented in the network to assess the value of the method.

As the core of the assessment methodology, drive surveys were carried out to quantify the benefits in terms of data throughput and service break duration. Two different routes (referred as A and B) were defined, encompassing a total of 314 km (7 hours drive). The selected routes covered a variety of environments, ranging from dense urban to open rural. Consequently, the drive survey included different driving conditions, from almost pedestrian in the city center to very fast moving on the open freeways. During the drives, a mobile terminal was configured to repeatedly download a 10MB data file via FTP (File Transfer Protocol). This test set-up was selected to utilize a traffic source offering a steady traffic flow. Thus, measurement variability is reduced to the statistical variation of the service gap, hence simplifying interpretation of the results. The data logging equipment was used in a 3 Downlink (DL) + 1 Uplink (UL) TSL configuration. As main performance indicator, the DL throughput was collected with a resolution of 0.5s. Even though a maximum DL throughput of 3*12=36Kbps can be achieved by such a configuration with the coding schemes CS1-CS2 currently implemented in the trial network, the actual throughput depends on the propagation conditions experienced by the user and the overall traffic demand in the trial area. It is worth noting that these conditions were not controllable during the trial, since coordination with ordinary subscribers in the network was impossible for obvious reasons. Nonetheless, it should be pointed out that the traffic demand remained virtually unchanged during the trial. Concretely, the total number of DL temporary block flow (TBF) establishments in the trial area differed less than 0.5% between the before and after periods.

Finally, the impact of the new PCU plan on the data rate was analyzed. The data analysis focused on every CR event (and not for the entire route), since the influence of the new plan on data performance should be confined to the vicinity of these events. A time period of 20s around CR events was considered appropriate to reflect the influence on data performance. The mean DL throughput and the service break duration were computed for each CR.

B. Trial Results

Table I presents the main performance indicators collected during the drive tests, broken down by PCU plan and route. Key performance issues for the different types of CR are first compared, based on the figures at the bottom row of Table I. The average application throughput during intra-PCU and inter-PCU CR events was 18.36 and 11.05 Kbps, respectively. Similarly, the mean duration of the service gaps in the intra-PCU and inter-PCU cases was 4.55 and 9.32s, respectively. These figures clearly indicate that it is beneficial to increase the share of intra-PCU CRs, which is the aim of the PCU planning activity. It is worth noting that these observations remain valid, regardless of the PCU plan and drive route.

TABLE I

PERFORMANCE COMPARISON OF THE DIFFERENT PCU PLANS AND ROUTES DURING THE TRIAL.

PCU Plan	Route	Avg. throughput[Kbps]		Avg. break duration[s]		Nbr. of		Overall intra-PCU CR ratio [%]	Overall average throughput [Kbps]	Overall average break duration[s]
		Intra-PCU CR	Inter-PCU CR	Intra-PCU CR	Inter-PCU CR	Intra-PCU CR	Inter-PCU CR			
Old	A	18.71	9.94	4.42	11.12	31	59	34.4	12.96	8.81
Old	B	17.34	10.84	5.39	8.42	33	70	32.0	12.92	7.45
Old	A+B	18.00	10.43	4.92	9.68	64	129	33.2	12.94	8.10
New	A	18.34	13.20	4.3	7.71	61	31	66.3	16.61	5.45
New	B	18.81	11.59	4.4	9.42	58	26	69.0	16.58	5.95
New	A+B	18.55	12.47	4.35	8.49	119	57	67.6	16.58	5.69
Old+New	A+B	18.36	11.05	4.55	9.32	183	186	-	-	-

TABLE II

RESULTS FROM THE GRAPH PARTITIONING TECHNIQUES BASED ON NMS CS-HO STATISTICS IN THE TRIAL AREA.

PCU Plan	PCU id.	Nbr. of Intra-PCU HO	Nbr. of Inter-PCU HO	Intra-PCU HO Ratio[%]	Nbr. of cells	Nbr. of GPRS TSLs
Old	1	786059	604490	56.5	44	253
	2	562922	558241	50.2	42	254
	3	384128	460415	45.5	35	214
	4	3732	45637	7.6	6	36
	5	114	16116	7.0	5	42
	6	1737	37809	4.4	7	48
	Overall	1738692	1722708	50.2	139	847
PCU Plan	PCU id.	Nbr. of Intra-PCU HO	Nbr. of Inter-PCU HO	Intra PCU-HO Ratio[%]	Nbr. of cells	Nbr. of GPRS TSLs
New	1	117649	739051	86.3	20	112
	2	66771	175231	72.4	18	108
	3	115178	530676	82.2	24	148
	4	44777	459687	91.1	23	147
	5	52179	374230	87.8	21	129
	6	125656	660315	84.0	33	203
	Overall	522210	2939190	84.9	139	847

From Table I, it is evident that the share of intra-PCU re-selections was doubled with the new PCU plan (i.e. 33.2% to 67.6%). Results for end-user throughput provide evidence of the benefit from the optimized PCU plan. Thus, the overall average throughput increased from 12.94 to 16.58 kbps (i.e. 28%). Likewise, the average service break duration decreased from 8.10 to 5.69s (i.e. 30%). This performance enhancement was mainly obtained because a large number of inter-PCU CR were converted to intra-PCU CR. It is worth noting that the quoted throughput benefit is for the 20s period around the CR event and not for the entire drive route. The gain obtainable on the entire drive route is obviously lower, and strongly depends on the ratio of time spent in CR to the time outside CR. However, once the focus is on CR events, this increase in data throughput is independent of the drive route.

Although the selected drive routes cover a large part of the BSC area, they can still only provide a rough indication of the change in the ratio of intra-PCU to inter-PCU CR for the whole BSC. To provide such information, Table II shows the estimations from the graph partitioning techniques based on NMS CS-HO statistics in the entire trial area. The intra-PCU HO ratio increases after the optimization process from 50.2% to 84.9% (i.e. an increase of 34.7% in absolute terms, 69% in relative terms). While the intra-PCU HO ratio varied significantly from PCU to PCU in the old plan, small variations are observed in the new plan. Likewise, the number of GPRS TSLs is more evenly balanced in the new plan, which translates into a reduction of the maximum ratio between subdomain weights from 6.05 (=254/42) to 1.88 (=203/108), i.e. a threefold reduction. Thus, the need for additional PCUs in the future is minimized. From this data, it can be concluded that the same trends observed in the drive tests are also seen in the NMS-based estimations.

IV. CONCLUSIONS

In this paper, the problem of assigning cells to PCUs in GERAN was solved by classical graph partitioning techniques. The proposed method enhances the quality-of-service of real-time services, delays the need for additional PCUs and reduces the workload of the re-planning tasks. A drive-test study over a limited area showed that the average service break duration associated to a CR could be decreased by up to 30%. Due to a lack of PS-HO statistics for the entire trial area, it is difficult to determine the actual impact for the entire area. Nevertheless, based on CS-HO statistics, it is shown that intra-PCU cell re-selection ratio can be increased by about 69%.

REFERENCES

[1] V. Rexhepi, M. Moissio, S. Hamiti, and R. Vaittinen, “Performance of streaming services in GERAN A/Gb mode,” in *Proc. IEEE 60th Vehicular Technology Conference*, vol. 6, 2004, pp. 4511–4515.

[2] K. Schloegel, G. Karypis, and V. Kumar, “Graph partitioning for high performance scientific simulations,” in *CRPC Parallel Computing Handbook*, J. D. et al, Ed. Morgan Kaufmann, 2000.

[3] R. Krishnan, R. Ramanathan, and M. Steentrup, “Optimization algorithms for large self-structuring networks,” in *Proc. INFOCOM ’99*, vol. 1, 1999, pp. 71–78.

[4] C. Walshaw, “Multilevel refinement for combinatorial optimisation problems,” *Annals of Operations Research*, vol. 131, pp. 325–372, 2004.

[5] C. Fiduccia and R. Mattheyses, “A linear time heuristic for improving network partitions,” in *Proc. 19th ACM/IEEE Design Automation Conference*, 1982, pp. 175–181.