Resource Management in Heterogenous Wireless Networks with Overlapping Coverage

Bin Bin Chen
Department of Computer Science
National University of Singapore
Email: chenbinb@comp.nus.edu.sg

Mun Choon Chan
Department of Computer Science
National University of Singapore
Email: chanmc@comp.nus.edu.sg

Abstract— **Development in new radio technologies and increase in user demands are driving the deployment of a wide array of wireless networks, ranging from 802.11 networks in the local area, to third generation data-only wireless networks in the wide area. With their complementary characteristics, these heterogeneous Radio Access Technologies (RATs) are expected to be integrated together to provide mobile users "Always Best Connections". Base Stations (BS) using different RATs will coexist and have arbitrary overlapping coverage without interfering with each other, and Mobile Stations (MS) equipped with multi-interfaces or "Software Defined Radio" technology can be associated with one or more BSs using different RATs.**

In this work, we study the Common Radio Resource Management (CRRM) problem, i.e., how should the network manager of these integrated heterogeneous wireless networks jointly allocate resources from various networks such that the scarce radio resources are used efficiently. We focus on applications with bandwidth reservation requirement, such as voice or video calls. We extend earlier works to consider the different increase in load when a single MS request is assigned to different BSs, due to independent channel condition and adaptive modulation/coding efficiency among different MS-BS pairs. We formalize CRRM problem as an online load balancing problem for temporary tasks with unrelated processors, and represent the input of problem using a weighted BS-MS graph, with a weighted BS-Region graph as its compact form. We study the computational complexity for the optimal solution. We also characterize the competitive ratio for general online algorithms by exploiting combinatorial properties of the weighted BS-Region graph. Cluster algorithm, which decomposes the whole area into disjoint clusters can potentially achieve a lower competitive ratio. However, its stochastic performance largely depends on the traffic distribution symmetry.

I. INTRODUCTION

Development in new radio technologies and increase in mobile users' demands for ubiquitous high speed data access are driving the deployment of a wide array of wireless networks, ranging from satellite networks, to Wireless Wide Area Networks (WWAN), Wireless Metropolitan Area Networks (WMAN), Wireless Local Area Networks (WLAN) and Wireless Personal Area Networks (WPAN), as shown in Figure 1. Each of these networks has different design tradeoffs in coverage, data rates, cost and many other network parameters. It is envisioned that these complementary Radio Access Technologies (RATs) will co-exist in the future and be integrated together to offer mobile users "Always Best Connections" [1].
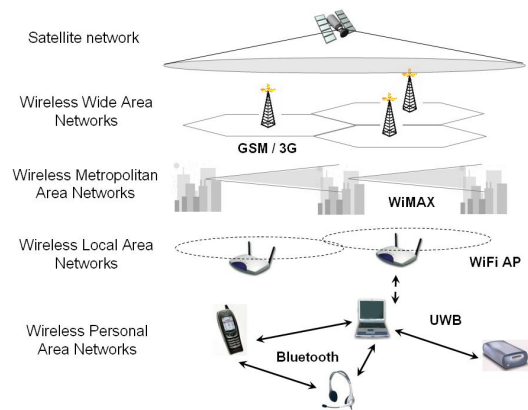

Fig. 1. Layers of heterogeneous wireless networks

In this work, we considered the case where different RATs are being deployed and a single wireless provider owns and operates these wireless networks. We study the *Common Radio Resource Management (CRRM)* problem, i.e., how should the network manager load balance Mobile Station (MS) requests among the various networks such that the scarce radio resources are used efficiently. Using 3GPP's definition [2], CRRM is a platform for coordination between heterogenous components. Our discussion will be based on a simplified heterogeneous wireless networks integration framework. As shown in Figure 2, the main components of a wireless network architecture are: MS, Radio Access Network (RAN) and Core Network (CN). Our model makes the following assumptions:

- The RAN as well as CN's bandwidth resource are over-provisioned. Therefore, the system's performance bottleneck is at the last hop, i.e. between Base Station (BS) and MS. Radio Resource Management (RRM) strategy is responsible for efficient use of scarce radio resource.
- BSs using different RATs can have arbitrary overlapping coverage, and they can simultaneously operate without interference with each other. We assume that there exists some mechanism to jointly manage their radio resource. For example, as proposed in [3], each RAT will expose a RRM interface for joint management, either loosely or tightly. Common RRM (CRRM) is logically defined
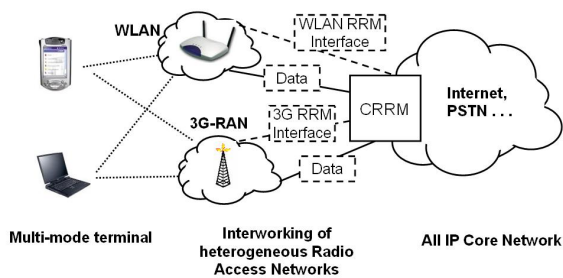
Fig. 2. Integrated Heterogeneous Wireless Networks

as the function to gather information from BSs and coordinate resource allocation among BSs to optimize system performance. CRRM can be implemented either in central or distributed way.

- MSs are equipped with multi-interfaces or "Software Defined Radio"(SDR) technology, so that they can be associated with one or many of the adjacent BSs.

We focus on applications requiring bandwidth reservation, such as voice or video calls, other than elastic traffic as considered in our previous work [4]. For such circuit-switch type applications, their requests will be blocked if available radio resource is not enough to meet their reservation requirement.

With this framework, the key issue we address is how available radio resource of overlapping BSs should be shared among multi-mode MSs. The performance measure used is *blocking rate*.

We made two contributions in this paper. First, We extend existing works to consider the different increase in load when a single MS request is assigned to different BSs, by formalizing CRRM problem as an online load balancing problem for temporary tasks with unrelated processors, and represent the input of problem using a weighted BS-MS graph to study the computational complexity for the optimal solution, which is shown to be impractical for implementation. Thus, study of efficient heuristic algorithm is a must. Our second contribution is to characterize the competitive ratio for general online algorithms by exploiting combinatorial properties of the weighted BS-Region graph, which is a compact form of weighted BS-MS graph. We show that cluster algorithm as proposed in [6], which decomposes the whole area into disjoint clusters, can potentially achieve a lower competitive ratio. However, competitive ratio only describes the worst case of an algorithm's performance, which may rarely happen under real situation. Complementarily, we also show that cluster algorithm's stochastic performance.

The rest of the paper is organized as follow. In Section II, we present related works on radio resource management issues in integrated heterogeneous wireless networks. In Section

III, we formalize the *Common Radio Resource Management (CRRM)* Problem as an online load balancing problem for temporary tasks with unrelated processors and the computational complexity for optimal solution is given in Section IV. Section V characterizes the competitive ratio for general online algorithms, and shows that cluster algorithm can potentially achieve a lower competitive ratio. The stochastic performance of cluster algorithm is numerically evaluated using a Markov chain under different network settings in Section VI. Finally, we conclude in Section VII.

II. RELATED WORKS

A. Integration framework for Heterogeneous Wireless Networks

The idea of combining heterogeneous access technologies is not a new idea initiated for wireless networks. In wired networks, multi-homing allows user to access Internet using multiple access technologies, e.g., dial up, ADSL, cable, ethernet, etc. Integration of heterogeneous technologies is also found in wireless voice area. The DECT/GSM dual mode telephones allow the user to use the wide-area GSM cellular network whilst out of the office or away from home, while also using the same handset at home when in range of a cordless telephone BS.

However, the design of a network architecture to fulfill the objective of seamless and efficient integration of heterogeneous RATs such as 3G and WLAN is still a challenging task. Wireless LANs, originally targeted at enterprise and home networks, lack many of the capabilities which are essential in public environments, such as unified and universally accepted AAA mechanisms, the integration of mobility mechanisms with QoS and application-level services, the support for roaming agreements, security issues. Conversely, although these characteristics are present in the design of 3G networks, their implementation depends on specific wireless access architectures such as CDMA2000 or UMTS, and their extension to other wireless technologies such as 802.11 presents several compatibility issues. In addition to tradition cell-based roaming (horizontal roaming), fast and transparent roaming across the constituent networks (vertical roaming) raises a challenge for mobility management, with the requirement of switching to a different network interface. transmission protocol should also adapt to multiple interfaces accordingly [7].

The integration can be done in both tightly coupled or loosely coupled ways [8]. Generally speaking, the tighter the coupling, the more complex the interworking interface will be. Thus loosely coupling is suggested by both [8] and 3GPP's ongoing work on integrating WLAN [9] [10] because of its flexibility.

B. Common Radio Resource Management

Because of the scarcity of radio resource, Common Radio Resource Management (CRRM) will serve as a driving force and an important design factor in the integration of heterogeneous wireless networks. In EU's Sixth Framework Programme, the EVEREST project [3] addresses the objective

to devise and assess a set of specific strategies and algorithms leading to an optimized utilization of scarcely available radio resources for the support of mixed services within heterogeneous networks. Similar to our system model, EVEREST's reference architecture is based on several RATs interfacing a common core network. Multiple cells using different RATs will be overlapped in the same area. Services are delivered via the RAT that is most efficient for that service. Since multi-interface MS can be connected to different RATs, unless there is knowledge about each RAT it would be very difficult to optimize network performance and to manage resources efficiently. As a result, a logical function box named CRRM (Common Radio Resource Management) is defined by 3GPP [2] [11] and studied in EVEREST project to jointly optimize the overall system performance. CRRM is characterized by two functionalities:

- Input: CRRM collect information from different resource pools, i.e. heterogeneous RATs. One resource pool is characterized by having its own RRM functionality.
- Output: CRRM then direct users to the resource pool which is most suitable. What is most suitable may depend e.g. on the user's position, service requirements, and network aspects like load balancing, avoidance of needless handovers, etc.

Briefly speaking, this paper study the resource allocation strategy for CRRM.

C. Coordinated Proportional Fairness

In a previous related work [4], we consider a similar integrated heterogeneous wireless framework as this paper, except that the target application in that paper is elastic traffic. For elastic traffic, there is no pre-defined and strict bandwidth requirement, and fairness among MSs instead of block rate is a commonly used performance criterion. [4] extends Proportional Fairness Concept used in HDR to multiple wireless networks with overlapping coverage, striking a balance between system throughput and user fairness. [4] also shows that the assignment flexibility of MS to any one of adjacent BSs can greatly improve the system performance, even with an algorithm using only one interface at a time.

D. Channel Assignment in Overlapping Networks

Radio resource management problem addressing circuit switched applications in overlapping homogeneous networks has been studied by several works [5] [6]. The assumption of homogeneity simplifies their problem formalization in two aspects:

- New MS will increase BS's load by a fixed value, which does not depend on which BS is chosen. This capacity allocation unit is referred as channel, and the radio resource management problem is thus named as "Channel Assignment Problem" in existing works.
- Current results are largely based on the nice properties of regular topologies, such as hexagonal grid topology in which each BS has six adjacent BSs.

[5] defines the *Transceiver (BS)- Mobile (MS) Control Channel Graph* to model the off-line version of the channel assignment problem. The control channel corresponds to the BS-MS covering relationship. Since every MS in a particular cell segment , [5] proposes to aggregate MSs having the same set of control channels together to reduce *Transceiver-Mobile Control Channel Graph* to a compact *Cellular Assignment Graph*, which represents the problem with a smaller bipartite graph of fixed vertices and edges with changing weight. Their main result is the *Channel Assignment Theorem* characterizing the existence of feasible solutions. Since they assume all MS consume identical resource (one channel), network flow procedures provide a polynomial algorithm to find a solution if one exists.

[6] models online version of the channel assignment problem as an *online load balancing problem for temporary tasks with restricted assignment*, where a task of unknown duration can be assigned to a restricted subset of processors based on its task type. But within this subset, the increase of load is same no matter which processor is chosen. In the general case without network structure constraints, the subset of processors a task may be assigned can be any element of the power set of all processors. The optimal algorithm for this general case is $O(2\sqrt{n}+1)$-competitive [12], where n is the number of processors. As this bound is too loose, [6] exploits the combinatorial properties of regular network topology structure as represented in *Cellular Assignment Graph* to attain a better competitive ratio. Also, they propose *cluster algorithm* to decompose the whole area into disjoint clusters which are managed by different BSs separately. They show that cluster algorithm can achieve a competitive ratio strictly better than greedy algorithm, with appropriate decomposition of clusters in some regular topologies.

However, the two assumptions of homogeneous networks model do not hold in heterogeneous wireless networks. In fact, they are not valid even in some homogeneous wireless networks. On one hand, MS-BS pair supports different data rate with varying modulation/coding schemes. For example, 802.11b wireless networks support four transmission rate: 1Mbps, 2Mbps, 5.5Mbps and 11Mbps. CDMA2000 1xEV technology (also known as High Data Rate "HDR") system provides a peak data rate of 2.4Mbps and an average data rate of 600Kbps using one 1.25MHz CDMA carrier in the forward direction. The rate being used depends on the channel condition. If we assume MSs have fixed bandwidth reservation request, then a BS can support more MSs in better channel condition with the same resource. On the other hand, heterogeneous wireless networks can have arbitrary overlapping coverage with each other. For example, the location distribution of WLAN APs with respect to 3G BSs can not be assumed to follow any regular pattern. In following sections, we will extend results in [5] and [6] to formalize a general framework by relaxing these two assumptions.

III. Problem Formalization

If we map MS requests to tasks and BSs to processors, CRRM Problem can be formalized as an *online load balancing problem for temporary tasks with unrelated processors* [12] [13].

In the definition of online load balancing problem for temporary tasks, tasks (MSs) arrive at random time and their service duration is unknown. Each arriving task j has an associated load vector, $p(j) = (p_1(j), p_2(j), \ldots, p_n(j))$ where $p_i(j)$ defines the increase in the load of processor i if we were to assign task j to it. Load vector can be used to categorize load balancing problem into several classes: identical processors case, related processors case, restricted assignment case and unrelated processors case.

- In the *identical processors case*, all the coordinates of a load vector are the same.
- In the *related processors case*, $p_i(j) = w(j)/v_i$, where the "weight" $w(j)$ depends only on the task j and the "speed" v_i depends only on the processor i.
- In the *restricted assignment case*, each task j has a weight $w(j)$ and can be assigned only within a subset of processors; in terms of the load vector the coordinates are either $w(j)$ or ∞. [5] [6] formalize the channel assignment problem as this case.
- The *unrelated processors case* is the most general case. Here, $p_i(j)$ are arbitrary non-negative real number, or ∞ if task j is not allowed to be assigned to processor i. Similar to restricted assignment case, a task j's type is uniquely defined by the subset of processors to which this type of tasks can be assigned. Note that coordinates of the load vector for this processor set can be unrelated. Clearly, related processors and the restricted assignment are both special cases of the unrelated processors case. The identical processors case is a special case of the related processors case where all the speeds v_i are the same. It is also a special case of restricted assignment where every task can be assigned to every processor, i.e. all the coordinates of tasks j are $w(j)$.

Without the two assumptions presented in the previous section, CRRM problem in heterogeneous wireless networks falls naturally into the unrelated processors case. In wireless networks, BS and MS can transmit to each other with different data rate by utilizing different modulation and coding schemes. The specific scheme in use depends on the channel condition and RAT, which are unrelated for different BS-MS pair. Even for the same MS its data rates to different BSs, or for the same BS its data rates to different MSs, are not directly related.

Online load balancing problem formalization requires the load vector be specified at arrival and keeps unchanged after that. But in wireless networks, MS mobility and channel fading will change the load vector from time to time. In this paper, we assume load vector is fixed after MS arrival for simplicity. The impact of mobility and fading will be considered in future works.

Capacity is an abstract measure for BS's radio resource,

TABLE I

COMPETITIVE RATIO OF ONLINE LOAD BALANCING PROBLEM

Cases	Competitive Ratio
Identical Processors	2-o(1)
Related Processors	$\Theta(1)$
Restricted Assignment	$\Theta(n^{1/2})$
Unrelated Processors	Unkown

either they are in the form of time slots, frequencies, power, or others. The definition of one capacity unit can be different across BSs. By appropriate scaling, we can compare different BSs' capacity directly, even when they are using heterogeneous RATs. Different scaling factor is chosen separately for different BS, such that the goal of load balancing algorithm is to balance load uniformly across all BSs.

In online load balancing problem formalization, processors are assumed to have infinitive capacity. Online algorithms' competitive ratio is defined by comparing its maximum processor load with optimal off-line algorithm's. An online algorithm A's competitive ratio is at most r if under any tasks input the maximum processor load ever used by algorithm A is no larger than $r \times L + b$, where b is a constant and L is the maximum processor load ever used by optimal off-line algorithm during processing the same input. In contrast, CRRM problem assumes processors have finite capacity and the performance criterion is blocking rate other than maximum processor load. The relationship between the two different performance criteria can be interpreted as follows: if an algorithm A is r-competitive in the definition of classical online load balancing problem and the maximum processor load is L under optimal off-line algorithm, this implies that the optimal algorithm can achieve zero block rate with each processor having capacity L. Thus algorithm A can achieve zero block rate with capacity no more than $r \times L + b$. Because of this equivalence, we will directly use the definition of classical online load balancing problem when discussing algorithms' competitive ratio in CRRM problem. A closer investigation of the relationship between these two criteria will be conducted in our future works.

The optimal online algorithms' competitive ratio for various cases of classical online load balancing problem is summarized in Table I [13]. Reassignment of accepted MSs is not allowed in the considered family of online algorithms.

While formalization in [5] and [6] as the restricted assignment case has a $\Theta(n^{1/2})$-competitive optimal online algorithm, such a result even does not exist for our formalization as the unrelated processors. Thus, similar to [6], we need to exploit the combinatorial properties of the wireless network topology to get a competitive ratio bound for the studied online algorithms.

IV. COMPUTATIONAL COMPLEXITY FOR OPTIMAL SOLUTION

A. Weighted BS-MS Graph

CRRM problem can be represented using a weighted BS-MS graph $G_\tau(B \cup M_\tau; E_\tau)$ as shown in Figure 3.
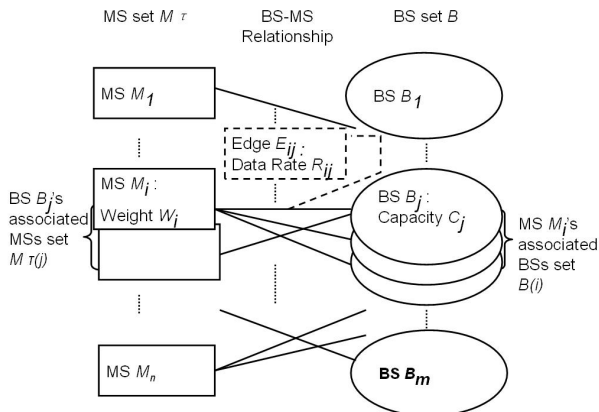

Fig. 3. Weighted BS-MS Graph

G_τ is an extension of the control channel graph in [5]. $B = \{B_i\}$ is the finite set of BSs covering the geographic area under our study, which is assumed to be invariant with time. $M_\tau = \{M_j\}$ is the finite set of active MSs distributed in the area at time τ. G_τ is a connected bipartite graph with vertices $V = B \cup M_\tau$ and edges $E_\tau = \{E_{ij} \mid B_i$ covers M_j at time $\tau\}$. Function $B(M_j)$ gives the set of BSs covering MS M_j, and function $M(B_i)$ gives the set of MSs covered by BS B_i.

BS B_i has a capacity C_i, which is an abstract measure for B_i's total amount of resources with appropriate defined measure unit, so that BSs' capacities can be compared directly even when they are using different RATs.

Weighted BS-MS graph G_τ extends Control channel graph in [5] by defining the parameters associated with MSs and edges. Every MS M_j has a bandwidth requirement weight W_j, and every edge E_{ij} is associated with an available data rate R_{ij} between B_i and M_j. In [6], they consider the case where MS M_j can have different weight W_j, but they still assume uniform data rate between different BS-MS pairs.

The information used to generate and update G_τ can be collected using CRRM's input function. At time τ when a new MS request arrives and becomes an element in M_τ, the CRRM output function should come out with an assignment solution $A = \{A_{ij}\}$, where A_{ij} is the resource allocated by BS B_i to MS M_j.

In this paper, we focus on the family of algorithms which will not intentionally drop a new request if they can find a feasible assignment configuration in their solution space without dropping already accepted MSs. It is possible for CRRM schemes to block a new request even when they can find enough resource for it, or drop an accepted request to take a new one. We will investigate these situations in future

work. The following sections will first study the computation complexity of optimal algorithms if reassignment is allowed, i.e., an accepted MS request can be reassigned to another BS.

B. LP-Optimal Algorithm

If simultaneous assignment is allowed, i.e. MS can communicate with multiple BSs in the same time and achieve aggregate bandwidth, the optimal algorithm is to block a new request arrived at time τ only if no feasible solution exists for the following linear programming problem:

$$
\begin{cases}
\Sigma_{B_i \in B(j)} A_{ij} \times R_{ij} = W_j & \forall M_j \in M_\tau \\
\Sigma_{M_j \in M_\tau(B_i)} A_{ij} \leq C_{ij} & \forall i \in B \\
A_{ij} \geq 0 & \forall i \in B, j \in M_\tau
\end{cases}
\tag{1}
$$

We call this linear programming based algorithm LP-Optimal algorithm. Though linear programming problem can be solved in polynomial time, LP-Optimal algorithm is still not efficient enough for real time CRRM decision, especially in a large wireless network with tens of thousands of MSs. In addition, the solution of a LP can have simultaneous assignment of one MS to several BSs, because the solution matrix A of this linear programming problem may have multiple non-zero elements in a single row. Further more, several (potentially all) elements of A may change for a single new coming MS.

Though not practical for implementation, LP-Optimal algorithm will be used as the performance baseline for our following discussions.

C. GAP-Optimal Algorithm

It is of special interest to consider the family of algorithms using only one interface per MS because:

- From the application viewpoint, simultaneous use of multiple interfaces for circuit-switch type of application will cause synchronization difficulty and other complexity.
- From the network perspective, simultaneously using multiple interfaces will increase coordination overhead.
- From networking device's point, SDR allows only one interface to work at one time. Even when multiple interfaces are equipped, it will be desirable to use only the most appropriate one of them and turn off all others to save energy.

When simultaneous assignment is not supported, new constraint should be added to the linear programming problem, requiring that only one element A_{ij} in each row of A is allowed to be non-zero. With this non-linear constraint, this problem can be mapped to the **GAP (General Assignment Problem)** [14] [15].

GAP considers a pair $(\mathcal{B}, \mathcal{I})$, where $\mathcal{B}$ is a set of bins and $\mathcal{I}$ is a set of items. Each bin $j \in \mathcal{B}$ has a capacity $c(j)$, and for each item $i \in \mathcal{I}$ and bin $j \in \mathcal{B}$, we are given a size $s(i, j)$ and a profit $p(i, j)$. The objective of GAP is to find a set $\mathcal{U} \subseteq \mathcal{I}$ of maximum profit such that $\mathcal{U}$ has a feasible packing in $\mathcal{B}$.

In CRRM problem, BS corresponds to bin, and MS requests correspond to items. The solution of GAP corresponds to a star matching in weighted BS-MS graph G_τ, which is a

subset of edge $E^* \subseteq E$ where the induced subgraph is a forest with every component a star whose center is a BS. Size $s(i,j)$ depends on the data rate between BS B_i and MS M_j, and is unrelated among different (i,j) pair. Profit $p(i,j) = 1$ for minimization of block rate. If we consider other optimization criteria, $p(i,j)$ can be assigned with different values accordingly. GAP has been proved to be NP Complete if size $s(i,j)$ is unrelated with each, even when profits are uniform. [14].

If simultaneous assignment is not allowed, the optimal CRRM algorithm is to block a new request only if GAP's solution set $\mathcal{U} \neq \mathcal{I}$. We name this algorithm as GAP-Optimal algorithm. GAP-Optimal algorithm's solution space is a strict subset of LP-Optimal algorithm's. Though GAP-Optimal algorithm uses only one interface per MS, it may still require reassignment of MSs when a new one arrives. Formalization in [5] is a special case of GAP model. By assuming uniform size $s(i,j)$, [5]'s model employs a polynomial network flow algorithm for optimal solution.

No matter simultaneous assignment is allowed or not, both LP-Optimal and GAP-Optimal algorithms are impractical for real implementation, thus it is a must to study the performance of efficient heuristic online algorithms.

V. ONLINE ALGORITHMS' COMPETITIVE RATIO

A. Weighted BS-Region Graph

A typical BS-MS Graph may have a large number of vertices corresponding to the active MSs in a metropolitan area, and this graph keeps changing over time with MSs initiating and terminating their requests. To study the competitive ratio, we need a more compact form to represent the underline network structures. By aggregating MSs covered by the same set of BSs together, we can reduce the weighted BS-MS graph to a weighted BS-Region graph, which has small and fixed vertices set.
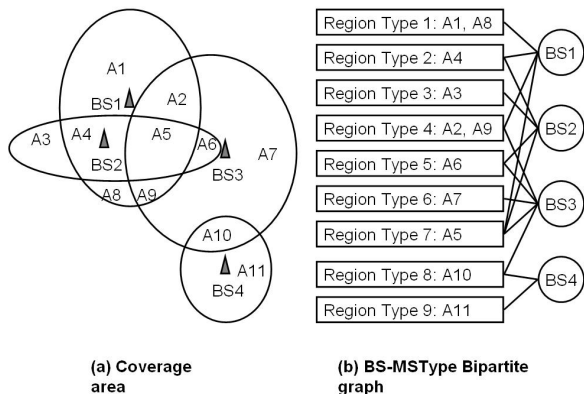

Fig. 4.   Regions

A region type is uniquely defined by the subset of BSs covering it. For example, in Figure 4, we can distinguish eleven different cell segments (A1 to A11) partitioned by the
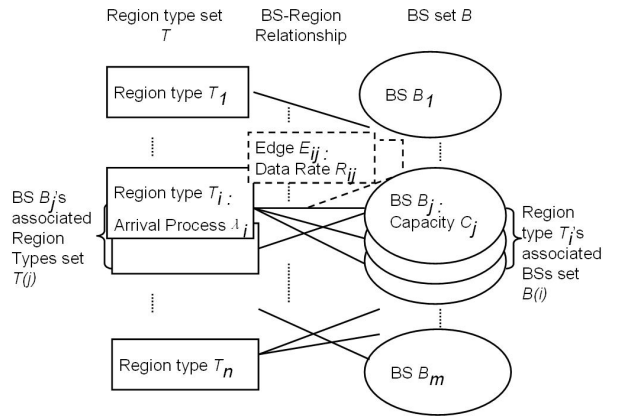

Fig. 5.   BS-Region Graph

coverage borderlines of four BSs, but there are only nine regions. The size n of region type set $T = \{T_1, \ldots, T_n\}$ is upper bounded by $n \leq 2^m$, where m is the size of BSs set $B = \{B_1, \ldots, B_m\}$. As shown in Figure 5, weighted BS-Region graph is a connected bipartite graph with vertices $V = B \cup T$ and edges $E = \{E_{ij} \mid B_i \text{ covers } T_j\}$. BS B_i has a capacity C_i, and Region R_j is associated with a stochastic MS arrival/departure process λ_j. Every edge connecting a BS-Region pair is associated with some efficiency information R_{ij} which shows the available data rates between BS B_i and MSs of type T_i. MSs of the same region type is not guaranteed to have same data rates, so R_{ij} can be interpreted as a random variable. We denote its maximal value and minimal value by $max(R_{ij})$ and $min(R_{ij})$ respectively, and $[min(R_{ij}), max(R_{ij})]$ is the variation range of R_{ij}. In addition, we denote the set of regions covered by BS B_i as B_i's associated regions set $T(B_i) = \{T_j \mid E_{i,j} \in E\}$, and the set of BSs identifying region type T_j is denoted as T_j's associated BSs set $B(T_j) = \{B_i \mid E_{i,j} \in E\}$. The set of BSs except BS B_j itself whose associated regions set have non empty intersection with the associated regions set $T(B_i)$ of BS B_i is defined as BS B_i's neighbor BSs set, i.e., $N(B_i) = \{B_j \mid \exists T_k, E_{ik} \in E \text{ and } E_{jk} \in E \text{ and } j \neq i\}$. For conciseness, we present BS B_i as i and region T_j and j if the context is clear.

B. Competitive Ratio for General Online Algorithm

For any online algorithm, we can characterize the upper bound for its competitive ratio by exploiting combinatorial properties of the given weighted BS-Region graph G.

Theorem 1: The competitive ratio of any online algorithm in a weighted BS-Region graph G is upper bounded by:

$$r = max_{i \in B}\{\Sigma_{j \in N(i)}(max_{k \in T(i) \cap T(j)}\{\frac{max(R_{jk})}{min(R_{ik})}\}) + 1\} \quad (2)$$

Proof: Assume that the maximal load of BS in G under optimal algorithm is L, then capacity L is enough to guarantee

zero block rate for optimal algorithm. We refer L as the optimal capacity. For any BS B_i, its load is upper bounded by MS requests generated in its associated regions set $T(i)$, which can only be allocated to $\{B_i\} \cup N(i)$, So the amount of these requests' load is upper bounded by the amount of capacity of BSs in $\{B_i\} \cup N(i)$. Because of different transmission efficiency among different BS-MS pairs, same amount of capacity in different BSs can support different load. $max_{k \in T(i) \cap T(j)}\{\frac{max(R_{jk})}{min(R_{ik})}\}$ is the upper bound for the relative efficiency between BS B_j's capacity and BS B_i's. The maximum capacity needed for a single BS to guarantee zero loss rate thus is a the weighted sum of the optimal capacity of $\{B_i\} \cup N(i)$. This weighted sum is $L \times r$, where r is as given in Theorem 1.

This bound is tight in the sense that there exists such online algorithms that have this competitive ratio r. We can relax this bound to get a simpler form.

Corollary 2: The competitive ratio of any online algorithm in a weighted BS-Region graph G is upper bounded by $N \times M + 1$, where N is the maximal number of neighbor BSs of any BS in G, and M is the maximal ratio between the data rate of any neighbor BSs.

For example in Figure 4, $N = 3$ as BS3 overlaps with all other three BSs, if we assume $M = 11$ as in the case of 802.11 network, then we can calculate the upper bound for any online algorithm in this topology, which is $N \times M + 1 = 34$.

The result of $r = N + 1$ in [6] is a special case of above theorem, where coefficients of one region's load vector for all its associated BSs are identical, i.e. $M = 1$. $r = N + 1$ is a tight bound for $M = 1$.

C. Cluster Algorithm

The cluster algorithm as defined by [6] for homogeneous wireless networks is to decompose the whole area into disjoint clusters which are managed by BSs separately. This algorithm can be extended to work in heterogeneous wireless networks. Cluster decomposition can be viewed in the weighted BS-Region graph G as two steps which help to minimize the competitive ratio. While step 2 is similar to [6], step 1 is new in our model, and only meaningful when multiple data rate is considered. Step 1 explains why greedily assigning MS to BS with higher data rate performs well in practice.

Step 1: For BSs B_j and B_i, which belong to region T_k's associated BS set $B(k)$, cluster algorithm can reduce $\frac{max(R_{jk})}{min(R_{ik})}$ by adjusting R_{jk} and R_{ik} with further assignment constraints. For example, for the overlapping region T_k of a HDR BS B_i and an 802.11 AP B_j, we assume $R_{ik} = 1Mbps$ while R_{jk} can be $1Mbps$, $2Mbps$, $5.5Mbps$ and $11Mbps$ with $max(R_{jk}) = 11Mbps$. Thus $\frac{max(R_{jk})}{min(R_{ik})} = 11$ without further assignment constraints. If the cluster algorithm always assigns MS in this region with $R_{jk} > 1Mbps$ to B_j, then these MSs are viewed to appear in region which is only covered by B_j instead of in region T_k. By imposing this assignment constraints, $R_{jk}^* = 1Mbps$ and $\frac{max(R_{jk}^*)}{min(R_{ik}^*)} = 1$. Decreasing efficiency ratio between overlapping BSs can potentially reduce the value of r in theorem 1. r will not increase in this step.

Step 2: After reducing the variation range of R_{ik} for edges in G, cluster algorithm will get a star matching $E^* \subseteq E$ with the induced subgraph $\langle E^* \rangle$ is a forest with every component a star whose center is a BS, and $T \subset V(\langle E^* \rangle)$, i.e., every region is associated with one and only one BS. By deleting edges from E, BS B_i's neighbor set is $N^*(i) = \{B_j \mid \exists T_k, E_{ik} \in E^* \text{ and } E_{jk} \in E \text{ and } j \neq i\}$. $N^*(i) \subseteq N(i)$ because $E^* \subseteq E$. Like step 1, step 2 can potentially decrease the value of r with no risk of increase r.

For example as shown in Figure 6 (a), BS A covers areas 1,2 and 3, BS B covers areas 3,4 and 5, while BS C covers areas 5,6 and 1. Both dashed and solid lines shown in Figure 6 (b) belong to E, and every BS has two neighbors, e.g. BS A shares area 1 with BS C and area 3 with BS B, so $N = 2$. To focus on the impact of step 2, we assume the maximal ratio between any BSs $M = 1$ for simplicity. According to corollary 1, there exists r-competitive online algorithms where $r = N + 1 = 3$. As $M = 1$, this bound is tight. On the other hand, a possible cluster algorithm is to assign area 1 and 2 to BS A, area 3 and 4 to BS B, area 5 and 6 to BS C. This decomposition is shown in Figure 6 (b) by removing all dashed lines to get the star matching E^* consists of only solid lines. According to $N^*(i)$'s definition, BS A only has neighbor BS C, BS B only has neighbor BS A, and BS C only has neighbor BS B, so $N = 1$. This cluster algorithm is at most 2-competitive according to corollary 1.
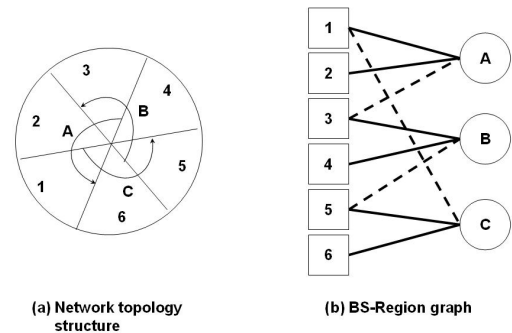

Fig. 6. Cluster algorithm's competitive ratio

Combining these two steps together, a cluster algorithm will derive from the original weighted BS-Region graph G a sub-graph $\langle E^* \rangle$ with $N^*(i) \subseteq N(i)$ and $[min(R_{ij}^*), max(R_{ij}^*)] \subseteq [min(R_{ij}), max(R_{ij})]$. By applying Theorem 1 to this subgraph $\langle E^* \rangle$, we can show that cluster algorithm's competitive ratio is at most r^*-competitive, where

$$r^* = max_{i \in B}\{\Sigma_{j \in N^*(i)}(max_{k \in T(i) \cap T(j)}\{\frac{max(R_{jk}^*)}{min(R_{ik}^*)}\}) + 1\} \quad (3)$$

r^* will be less than or at most equal to the competitive ratio r which we derived for the original graph G.

The cluster algorithm presented is based on [6]. There are many other interesting heuristics that do not reply on clustering. In the next section we describe the **uniform algorithm**, an algorithm that does not use clustering.

D. Uniform algorithm

In uniform algorithm, an MS j's load is shared simultaneously among its associated BSs set $B(j)$ by allocating same amount of resource from each associated BS. We choose uniform algorithm mainly because: (a) it demonstrates a good worst case competitive ratio (equal to cluster algorithm) in our chosen topology as we will show below; (b) it is easy for analysis, because the resource allocation decision does not depend on the current load of BSs, but only based on the task type of MS. To summarize, uniform algorithm and cluster algorithm have the same worst cast bound and use the same information when making decision. Cluster algorithm takes the approach of restraining assignment, while uniform algorithm exploits multiplexing by doing simultaneous assignment. The uniform algorithm is interesting because it has the same competitive ratio, its average performance is better as we will show in the next section.

There are many other resource allocation heuristic algorithms with interesting properties, such as **Greedy minimum load algorithm** which always assigns MS to the BS with currently minimum load. Greedy minimum load algorithm performs extremely good and robust in our simulation, though it has a higher worst cast competitive ratio. Because resource allocation decision of greedy minimum load algorithm depends on the current load of BSs, it requires the development of new numerical analysis techniques. We will investigate these heuristic algorithms in our future works.

VI. STOCHASTIC PERFORMANCE EVALUATION

Competitive ratio only describes the worst case of an algorithm's performance, which may rarely happen under real situation. Complementarily, we study the stochastic performance of cluster algorithm. Our main observation is that: *cluster algorithm's stochastic performance depends largely on the symmetry of traffic (MS requests) distribution across clusters. If the decomposition of cluster helps improve the traffic distribution symmetry, cluster algorithm demonstrates a good performance compared to non-cluster algorithm. Otherwise, cluster algorithm will have poor stochastic performance.*

A. Assumptions and Analysis Approach

For simplicity, we consider the network topology and cluster decomposition as shown in Figure 6. We assume that the arrival process of MS requests in each area $i \in \{1, \ldots, 6\}$ follows a Poisson distribution with arrival rate λ_i and is independent each other. We denote the arrival rate vector as $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$. All requests' lifetime follows the same exponential distribution with departure rate μ. We assume all BSs have infinitive capacity, and MS request is unitary-weighted, which can be served using one unit of capacity by any BS in its associated BSs set. These assumptions

reduce our problem to the channel assignment problem as defined in [5]. The following discussion can be extended to our general CRRM formalization with similar result.

Based on above assumptions, the number of MS requests in area i can be modelled as an $M/M/\infty$ queue. The probability that area i has k_i requests is:

$$p(k_i) = [(\lambda_i/\mu)^{k_i} e^{-\lambda_i/\mu}]/k_i! \qquad k = 0, 1, 2, \ldots \quad (4)$$

As we assume that arrival processes in different areas are independent to each other, the probability that the traffic distributions in all 6 areas is $K = (k_1, \ k_2, \ k_3, \ k_4, \ k_5, \ k_6)$ is:

$$p(K) = \prod_{i=1}^{6} p(k_i) \quad (5)$$

For a given state K, the maximum processor load in this state under cluster algorithm is:

$$L_{cluster}(K) = max\{k_1 + k_2, k_3 + k_4, k_5 + k_6\} \quad (6)$$

The maximum processor load in this state under uniform algorithm is:

$$L_{uniform}(K) = max\{k_1/2 + k_2 + k_3/2, \\ k_3/2 + k_4 + k_5/2, k_5/2 + k_6 + k_1/2\} \quad (7)$$

The maximum processor load in this state under LP-optimal algorithm can be derived using the Channel Assignment Theorem [5]:

$$L_{optimal}(K) = max\{k_2, k_4, k_6, (k_2 + k_3 + k_4)/2, \\ (k_4 + k_5 + k_6)/2, (k_6 + k_1 + k_2)/2, \\ (k_1 + k_2 + k_3 + k_4 + k_5 + k_6)/3\} \quad (8)$$

We compare the stochastic performance of cluster algorithm and uniform algorithm by numerically calculate the distribution of

$$r_{cluster}(K) = \frac{L_{cluster}(K)}{L_{optimal}(K)} \quad (9)$$

and

$$r_{uniform}(K) = \frac{L_{uniform}(K)}{L_{optimal}(K)} \quad (10)$$

where state K is distributed as shown in Equation (5), which is decided by the arrival rate vector λ and departure rate μ. Thus, by fixing λ and μ, the distribution of $r_{cluster}(K)$ and $r_{uniform}(K)$ can be numerically calculated. As element of K increases, the probability that system is operating in state K decreases exponentially. Also, $r_{cluster}(K)$ and $r_{uniform}(K)$ are upper-bounded by a constant. This allows us to approximate the infinite number of states by the finite states set $L = \{K | K_i < K_{thr}, \forall i\}$. K_{thr} is chosen such that $\sum_{K \in L} p(K) > 0.99$.

B. Worst Case Competitive Ratio

We have shown in section V that the competitive ratio of cluster algorithm $r_{cluster} \leq 2$.

In fact, this is the lower bound in topology Figure 6 for any online algorithm without MSs reassignment.

Lemma 3: For any online algorithm without MSs reassignment in topology Figure 6, its competitive ratio $r \geq 2$.

Proof: Consider the following input instant: N amount of requests are generated at $A \cap B$, $B \cap C$, $C \cap A$ separately. As the total amount of request is $3N$, for any online algorithm, there must be one BS with load no less than N. Without loss of generality, we assume BS A has load no less than N. Then the input instant remove requests generated at $B \cap C$, and then generate N new requests at region covered only by BS A. With the capacity of reassignment, the optimal algorithm can always keep its maximal processor load no larger than N, while the online algorithm has to assign the N new requests to BS A, which results in totally no less than $2N$ load in BS A. So $r \geq 2$

From lemma 3 and $r_{cluster}(K) \leq 2$, we get $r_{cluster} = 2$. We can proof that uniform algorithm also achieves this lower bound.

Lemma 4: Uniform algorithm's competitive ratio $r = 2$.

Proof: From lemma 3, $r_{uniform} \geq 2$. We assume that the maximum processor load under optimal algorithm is L, thus optimal can support zero loss rate if each BS has L capacity. Without loss of generality, we assume that BS B has the maximum load under uniform algorithm. Load arriving from area 4 and area 5 together should not be more than $2L$, because the traffic from area 4 and area 5 can only be assigned to BS B and BS C whose total capacity is $2L$.

$$L_4 + L_5 \leq 2L \tag{11}$$

Similarly, we have:

$$L_5 + L_6 \leq 2L \tag{12}$$

The load assigned by uniform algorithm to BS B is

$$L_4/2 + L_5 + L_6/2 =$$
$$[(L_4 + L_5) + (L_5 + L_6)]/2 \leq 2L \tag{13}$$

Thus the competitive ratio of uniform algorithm is upper bounded by $r_{uniform} \leq 2$. So $r_{uniform} = 2$.

C. Results

We fixed $\mu = 1$. Given a rate vector $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$, the request arrival rate to BSs A, B, C are

$$(\lambda_1 + \lambda_2, \lambda_3 + \lambda_4, \lambda_5 + \lambda_6) \tag{14}$$

under cluster algorithm, and:

$$(\lambda_1/2 + \lambda_2 + \lambda_3/2, \lambda_3/2 + \lambda_4 + \lambda_5/2, \lambda_5/2 + \lambda_6 + \lambda_1/2) \tag{15}$$

under uniform algorithm.

To study the impact of traffic distribution symmetry, we consider three different network settings. Figure 7 (a)
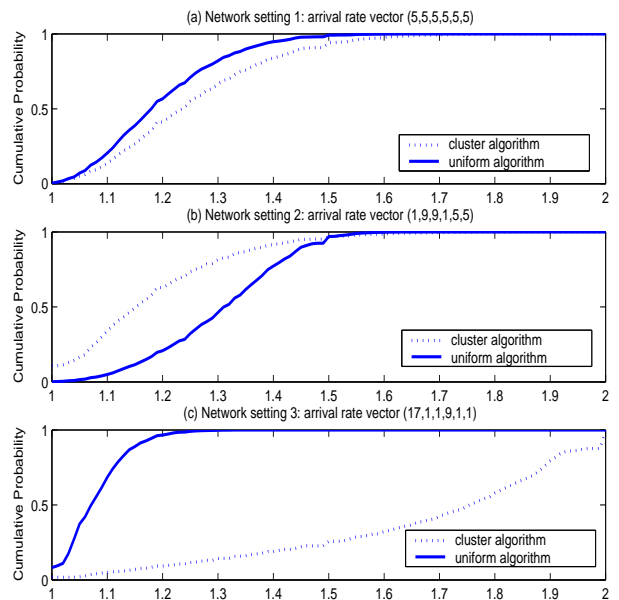

Fig. 7. Stochastic performance of online algorithms

shows the cumulative probability function of $r_{cluster}(K)$ and $r_{uniform}(K)$ under arrival rate vector $\lambda = (5, 5, 5, 5, 5, 5)$, which is a symmetrical traffic distribution setting for both cluster algorithm and non-cluster algorithm, which both have arrival vector for BSs as $(10, 10, 10)$. Non-cluster algorithm has a slightly better stochastic performance than cluster algorithm because it employs a higher degree of multiplexing.

Figure 7 (b) illustrates the case where cluster algorithm helps to achieve a better traffic distribution symmetry. In the network setting of $\lambda = (1, 9, 9, 1, 5, 5)$, cluster algorithm still has arrival vector for BSs as $(10, 10, 10)$, while for uniform algorithm the average arrival rates for the three BSs are $(14, 8, 8)$ respectively. Cluster algorithm has a better stochastic performance than non-cluster algorithm under this network setting.

On the other hand, Figure 7 (c) shows the network setting where cluster algorithm deteriorates due to traffic distribution asymmetry. By setting $\lambda = (17, 1, 1, 9, 1, 1)$, the three BSs' average arrival rate under cluster algorithm are $(18, 10, 2)$ respectively, while for uniform algorithm the average arrival rates for BSs are $(10, 10, 10)$. Non-cluster algorithm has a superior stochastic performance than cluster algorithm under this network setting.

This shows that cluster algorithm's stochastic performance depends largely on the symmetry of traffic (MS requests) distribution across clusters.

Figure 8 shows the distribution of $r_{cluster}$ and $r_{uniform}$ when λ_i for every area i follows uniform distribution in $[1, 20]$, and all λ_i are independent with each other. We get the result by integration over all the possible points in the range. As shown in the figure, uniform algorithm has a better stochastic performance than cluster algorithm in this range of network settings. It is harder to drive uniform algorithm to perform poor

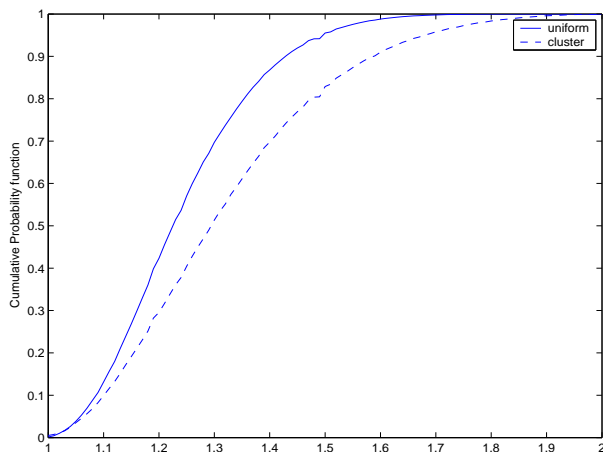than cluster algorithm, because uniform algorithm employs higher degree of multiplexing.

Fig. 8.   Stochastic performance of online algorithms II

VII. CONCLUSION

In this paper, we formalize the Common Radio Resource Management (CRRM) problem as an online load balancing problem for temporary tasks with unrelated processors, and study the computational complexity for the optimal solution. We also characterize the competitive ratio for general online algorithms by exploiting combinatorial properties of the weighted BS-Region graph. We describe how cluster algorithm can be used to achieve a lower worst case competitive ratio. We also show that cluster algorithm's stochastic performance depends largely on the traffic distribution symmetry.

REFERENCES

[1] M. Frodign, S. Parkvall, C. Roobol, P. Johansson and P. Larsson, "Future-Generation Wireless Networks," *IEEE Personal Communications*, vol. 8, no. 5, pp. 10–17, October 2001.

[2] 3GPP, "Improvement of RRM across RNS and RNS/BSS (Release 5)," *3GPP TR 25.881 V5.0.0*, December 2001.

[3] EVEREST, "Deliverable 11: First report on the evaluation of RRM/CRRM algorithms," November 2004.

[4] B. Chen and M. Chan, "Coordinated Proportional Fairness in Wireless Networks with overlapping Coverage," *Tech Report, Communication and Internet Research lab, National University of Singapore*, June 2005.

[5] M. Iridon, D. Matula and C. Yang, "A graph theoretic approach for channel assignment in cellular networks," *Wireless Networks*, vol. 7, no. 6, pp. 567–574, 2001.

[6] P. Crescenzi, G. Gambosi and P. Penna, "On-line algorithms for the channel assignment problem in cellular networks," *Discrete Applied Mathematics*, Vol. 137, no. 3, pp. 237–266, March 2004.

[7] H. Y. Hsieh and R. Sivakumar, "pTCP: An end-to-end transport layer protocol for striped connections," in *IEEE ICNP*, Paris, Fance, November 2002.

[8] M. Buddhikot, G. Chandranmenon, S. J. Han, Y. W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," in *IEEE INFOCOM*, San Francisco, CA, USA, March 2003.

[9] 3GPP, "Feasibility study on 3GPP system to wireless local area network (WLAN) interworking," *3GPP TR 22.934 V6.2.0*, September 2003.

[10] 3GPP, "Requirements on 3GPP system to Wireless Local Area network (WLAN) interworking (Release 6)," *3GPP TS 22.234 V6.0.0*, March 2004.

[11] 3GPP, "Improvement of RRM across RNS and RNS/BSS (Release 6)," *3GPP TR 25.891 V0.3.0*, February 2001.

[12] Y. Azar, B. Kalyanasundaram, S. Plotkin, K. Pruhs, O. Waarts, "Online load balancing of temporary tasks", *J. Algorithms* vol. 22, no. 1, pp. 93–110, 1997.

[13] Y. Azar, "On-line load balancing," in: A. Fiat, G. Woeginger (Eds.), "On-line Algorithms-The State of the Art," Springer, Berlin, 1998.

[14] D. B. Shmoys and É. Tardos, "Scheduling unrelated machines with costs," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 448–454, Austin, Texas, USA, 1993.

[15] C. Chekuri and S. Khanna, "A PTAS for the multiple knapsack problem," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 213–222, San Francisco, USA, January 2000.