Optimal Vertical Handover Control Policies for Cooperative Wireless Networks

Katerina Papadaki and Vasilis Friderikos

Abstract: **Inter-operability between heterogeneous radio access technologies (RATs), in the sense of seamless vertical handover (VHO) support with common radio resource management (CRRM) functionalities, has recently attracted a significant research attention and has become a prominent issue in standardization fora. In this paper, we formulate the problem of load balancing between cooperative RAT's as a mathematical program and by trading off a pre-defined delay tolerance per request we propose a vertical handover batch processing (VHBP) scheme. To quantify the performance of the proposed VHBP scheme we compare it with a baseline processing scheme, where each VHO request is processed independently under a number of different network scenarios. Numerical investigations reveal significant net benefits of the proposed scheme compared with the baseline, both in terms of achieved load balancing levels but also with regard to the acceptance rate of the VHO requests.**

Index Terms: **Batch processing, heterogeneous radio access technologies (RATs), linear integer programming, load balancing, vertical handover (VHO).**

I. INTRODUCTION

One of the main features of next generation mobile systems, usually referred to as systems beyond 3G, is the envisioned inter-operability support between different RAT's [1], [2]. Cooperative radio access technologies (RATs) would allow network operators to utilize the diversity of different access networks to increase the efficient use of their radio resources, while satisfying contracting quality of service (QoS) to the mobile users. Allocation of mobile users' traffic across different RATs will be mainly driven by the aim of increasing overall system performance by taking into account required QoS and terminal capabilities. This operation can be achieved by deploying common radio resource management (CRRM), which can be considered as an overlay entity laying above the traditional system-specific RRM functionalities. In that respect, vertical handover (VHO) requests would be handled by CRRM functionalities with the primary objective to optimize the system as a whole. Additionally, vertical handover mechanisms need to ensure service continuity, i.e., changing the access network technology system within a pre-defined time window.

Load balancing emerges as an important functionality in cooperative access networks since it can decrease the dropping rate of the requests by minimizing the dispersion around the average utilization across the systems. Clearly, this is of critical importance especially in geographical areas with high density of mobile users, where network congestion at peak hours of the day can be the norm rather than the exception. Additionally, since the heterogeneity of wireless access networks in the future is likely to increase, see table 1 for a non exhaustive list of current and emerging networks, load balancing naturally becomes a prominent issue for consideration. The current line of thought for handling VHO requests mainly assumes a first come first served (FCFS) processing scheme, where decision making is performed for each request individually. In this paper, by trading off a tolerable delay for each request, we propose a vertical handover batch processing (VHBP) scheme that can significantly increase the performance of the system as a whole. We discuss the conditions under which batch is performed together with implementation aspects on how to batch so that a maximum delay tolerance per request is not violated. For each batch, a mixed linear integer optimization problem is solved with the main objective to provide load balancing. The rational behind the proposed batch processing scheme is three fold. Firstly, the majority of the applications either degrade gracefully with increased delay (elastic applications) or can tolerate a variability in delay (streaming multimedia applications). Therefore, it is possible to delay a VHO request—up to a certain degree—in order to provide batch processing. Secondly, processing a batch of requests rather than handling each request individually leads to a significant degree of improvement regarding load balancing and acceptance rate. Thirdly, batch processing enables prioritization by introducing a weighting factor of each request to meet different QoS or service level agreement requirements.

We formulate the optimization problem for two different scenarios, which we call the targeted and untargeted scenarios. In the targeted scenario, we assume that the mobile host while connected to the origin access network can listen to only one other access network. In this case, the VHO request has a predefined origin destination pair. In the untargeted scenario, we assume that the mobile host can listen to all other available access networks. In this case, the VHO request has a predefined origin but several potential destinations. In addition to the radio resources of the targeted access networks the proposed optimization problem can also take into account the cost of the route for performing the VHO. This aspect is portrayed more clearly in Fig. 1, which depicts the case, where an untargeted VHO request from RAT 1 to RAT's 2 or 3 is considered. In this case, the optimization problem for the VHO request will not only consider the available resources in RAT's 2 and 3 but also the cost of the route for re-directing the user's traffic to systems 2 or 3. This operation is similar to the one proposed in [3], where the access point selection procedure for each user takes into account bandwidth constraints on the wireless link but also on backhaul links.

Manuscript received May 2, 2006.

K. Papadaki is with the Operational Research Department at London School of Economics (LSE), email: K.P.Papadaki@lse.ac.uk.

V. Friderikos is with the Centre for Telecommunications Research at King's College London, email: vasilis.friderikos@kcl.ac.uk.

1229-2370/06/$10.00 © 2006 KICS

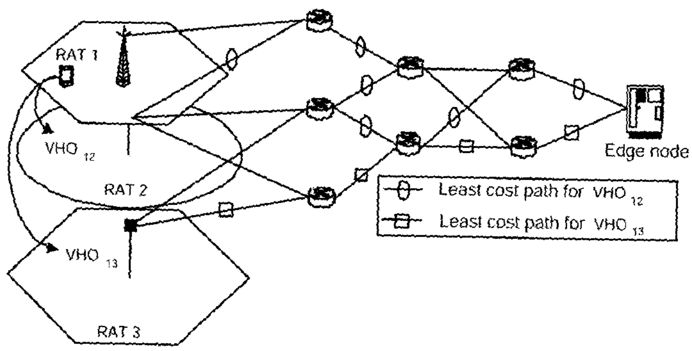


Fig. 1. Both radio resources but also route costs can be taken into account in the decision making of a VHO request from access network 1 to access networks 2 or 3.

Table 1. Heterogeneity of wireless connectivity based on the emerging and existing radio access networks.

Network	*Rate* (Mbps)	*Mobility*	*Coverage* (km)
GPRS	0.144	Vehicular	1–3
WCDMA	2	Vehicular	1–3
S-UMTS	0.144	Vehicular	Global
IEEE 802.16a	70	Pedestrian	6–9 *
IEEE 802.20	1–9	Vehicular	2–4 *
IEEE 802.11a	54	Pedestrian	0.1–0.3
IEEE 802.11b	11	Pedestrian	0.1–0.3
HIPERLAN 2	25	Pedestrian	0.1–0.3

*Depending on the carrier frequency used

The rest of the paper is organized as follows. In Section II, background information and related work are discussed covering the aspects of load balancing in mobile networks and current activities in standardization fora towards inter-operable access networks. Section III discusses the necessary conditions to perform batch processing together with implementation issues. In Section IV, the system model is described and the optimization problem for providing load balancing is derived. In Section V, the results of numerical investigations are provided that depict the performance gains of the proposed scheme. Finally, the conclusions together with some comments on avenues for future research end the paper in Section VI.

II. BACKGROUND AND RELATED WORK

A. *Standardization Activities for Inter-Operability between Access Networks*

From a standardization perspective, the third-generation partnership project (3GPP) has defined a series of six different scenarios with incremental levels of technical complexity (degrees of network coupling) for interworking between WLANs and UMTS [1]. The first scenario has no impact on either radio access networks but ensures that mobile users receive only one bill for using services from both networks. The second scenario envisions a 3GPP system-based access control and charging. In this case, a user can access the WLAN infrastructure through USIM based authentication. Scenario four implies a continuity of packet service between cellular and WLAN systems but not seamless handover. Scenario five introduces seamless handover support between the systems and scenario six allows a VHO of a VoIP call in WLAN to a cellular circuit-switched link. The assumption in this paper is the provision of seamless VHO support between multiple RATs, which is well beyond Release 6 standards. Within 3GPP, common radio resource management functionalities are considered by taking into account resource availability from individual RRM entities [4].

Concerning WLAN networks there are current ongoing standardization efforts to include functionalities similar to those of cellular networks such as roaming and resource management to provide ubiquitous coverage. In that respect, the IEEE 802.11f task group [5] was established to specify an inter access point protocol (IAPP) to support inter-operability, mobility, handover and coordination among WLAN access points. The IAPP, which is an application layer protocol, is responsible to convey access point management information related to the association, reassociation and disassociation of mobile nodes. The IEEE 802.11 task group k (TGk) develops radio resource measurements as an extension to the IEEE 802.11 standard for wireless local area networks. These extensions will specify what types of radio resource information to measure and the request/report mechanism through which the measurement demands and results are distributed among stations [6].

Finally, inter-operability issues between WMAN and mobile networks are currently considered within the IEEE 802 MAN standards committee. More specifically, the IEEE 802.21 working group [7] aims to develop standards for the establishment of access independent mechanisms for supporting optimal mobility between different RATs. To this end, the IEEE 802.21 main objectives are to facilitate handovers among heterogeneous access networks, such as IEEE 802.11/15/16, and mobile networks by devising the required functional models together with the related messages and interfaces.

B. *Load Balancing in Wireless Networks and Vertical Handover Support*

Efforts for providing load balancing in mobile networks can be traced back to the case of a homogeneous mobile environment, i.e., single RAT. The aim of load balancing on a single multicell RAT is to allow a flexible assignment of resources per cell to minimize call blocking/dropping probability due to spatial imbalance of traffic demand in different cells (see [8] and references therein). A different approach, that has been widely used, is to perform cell overlapping to increase utilization levels and balance network traffic [9]. In a more recent work, load balancing has been considered jointly with coordinated scheduling for CDMA networks [10]. The work mentioned above considered a homogeneous cellular environment and therefore can not be applied to provide load balancing in a heterogeneous wireless network environment. On the other hand, we should reckon that the issue of load balancing in a heterogeneous wireless access environment can be considered as a natural evolution from the homogeneous mobile network setting to perform network optimization. Despite this analogy, research efforts for the case of multi-RAT environment have mainly focused on the required VHO mechanisms and architectural elements to provide the necessary functionalities to accomplish inter-system mobility rather than perform optimization of network performance. The concept of VHO, i.e, a handoff scheme in presence of heterogeneous radio access technologies, was firstly introduced in [11], where different VHO schemes were proposed and evaluated. An overview of such research challenges including VHO issues have been articulated in [12] and [13]. Mobility management schemes that allow a seamless service provision for two RATs (i.e., WCDMA and WiFi) have been discussed in [14] and [15].

In a more close related work, the authors in [16] conducted performance analysis of degradation in user experience (in terms of throughput and handoff latency) caused by inter-technology handoff between WLAN and GPRS. They also proposed an optimization scheme that aims to minimize handoff delay in conjunction with other factors that can cause throughput degradation, such as packet loss (which cause retransmission), packet encapsulation delay, and number of active users in the cell. In [17], a simple multi-network optimization scheme is proposed that aims to achieve efficient usage of network resources and is based on a network elimination factor that characterizes network conditions in each RAT. In [18], a VHO decision algorithm is presented that takes into account the requirements from the application layer in order to ensure session continuity. A different approach for load balancing has been proposed in [19], which is based on the deployment of a set of relaying nodes in a cellular network to enhance capacity and reduce power consumption/interference.

The fundamental assumption of these previous works is that VHO requests are handled on a first come first served (FCFS) basis, where the emphasis is on delivering the required QoS to the mobile users without optimizing the system as a whole. On the contrary, this paper proposes a batch processing scheme with optimal allocation of the VHO requests. Our work shares many similarities with the batch processing scheme for admission control and related comparison with the FCFS discipline, which has been proposed and studied in [21]. The difference in our work is that [21] focuses on handoff prioritization in one access network, whereas this paper focusses on load balancing in different heterogeneous access networks.

III. CONDITIONS FOR IMPLEMENTATION OF BATCH PROCESSING

We assume that VHO requests arrive according to a Poisson process with arrival density λ. For each VHO request there is a maximum delay tolerance D. If at time t we have N_t number of requests already in the batch (assuming that at $t = 0$, N_t is equal to one) and the computational time to solve the optimization problem with this number of requests is T_{N_t}, then the stopping time t_s would be

$$t_s = \min\{t : D - t \leq T_{N_t+1}\}. \tag{1}$$

The above equation simply means that we stop buffering requests at time t_s, which is the instant where the remaining available time before the expiration of the first buffered request, $D - t$, is less than the computational time required to solve the optimization problem with $N_t + 1$ requests, T_{N_t+1}. This means that if we wait for one more request we will not have enough time to perform the computation for the $N_t + 1$ requests, and thus we stop buffering at that instance.

Another important decision is to decide whether or not to provide batch processing. After the first request arrives we decide whether to wait for more requests to perform batch processing or to process the first request on its own. Such decision policy can be based on the arrival density of the requests and the value of the delay tolerance, D, as follows: At time $t = 0$, we have $D - T_1$ remaining waiting time. If the probability of actually receiving one or more additional VHO requests within this time interval is below a pre-defined confidence level α then batch processing will not be performed. If we define by $X(t)$ a Poisson random variable that expresses the number of arrivals in a time interval of length t, then the criterion for batch processing can be written more formally as follows

$$\text{Prob}\,(X(D - T_1) \geq 1) \geq \alpha. \tag{2}$$

Since $X(t) \sim \text{Poisson}(\lambda t)$, batch processing will be performed when the following inequality regarding arrival densities holds

$$\lambda \geq \frac{\ln(1-\alpha)}{T_1 - D}. \tag{3}$$

In the next section, a description of the optimization problem for providing load balancing based on batch processing is described.

IV. PROBLEM FORMULATION & SYSTEM MODEL

In this section, we define the parameters and variables for the system model and use them to formulate the optimization problem for load balancing resources between RATs used by the VHBP scheme. We formulate mathematical programs for load balancing resources for targeted and untargeted VHO requests and then we extend these programs to include in the objective the path cost and the acceptance rates of VHO requests.

A. Targeted VHO Requests

We define the problem for a batch of N targeted VHO requests between systems in $\mathcal{S} = \{1, \cdots, K\}$, within a time interval of length t_s. Let Q be the set of all VHO requests. Each request $n \in Q$ has an associated origin system, destination system and service level. We assume that the service level is a specific bit rate which consumes different amounts of resources (in terms of bandwidth or power) in different systems. We let $f_k(b)$ to be the amount of resource in system k needed to accommodate requested bit rate b. For each request n, we let $b_n \in \mathcal{R}$ be the associated bit rate, where $\mathcal{R}$ is the discrete set of allowable bit rates. We let $Q_{ij} \subseteq Q$ be the set of requests with origin system i and potential destination system j.

Additionally, similarly to the radio network subsystem application part signaling [20], we assume that load information has to be signaled between radio network controllers. We denote by B_k the current utilized resource level and by C_k the resource capacity for system k. We let $x = (x_1, \cdots, x_N)$ be the vector of decision variables, where $x_n = 1$ when accepting request n and $x_n = 0$ otherwise. Then, the available resource for system

k after the allocation of a batch of requests is denoted by $\widetilde{B}_k$, and it can be expressed as follows.

$$\begin{aligned}\widetilde{B}_k(x) =& B_k + \sum_{i\in\mathcal{S}:i\neq k}\ \sum_{n\in Q_{ik}} x_n f_k(b_n) - \\ &\sum_{j\in\mathcal{S}:j\neq k}\ \sum_{n\in Q_{kj}} x_n f_k(b_n) \quad \text{for all } k\in\mathcal{S}.\end{aligned} \tag{4}$$

The new resource level $\widetilde{B}_k$, in (4), is derived by adding to the current resource level B_k the resources from accepted requests whose destination was system k, and by subtracting the resources from accepted requests whose origin was system k.

We want to perform load balancing of resource utilization levels across all the systems. The objective is to accept/reject requests to minimize the sum of absolute deviations of the system resource fractional utilization levels, $\widetilde{B}_k(x)/C_k$, from their mean, subject to not exceeding the system resource capacities. This can be written as

$$\begin{aligned}\min_x & \sum_{k=1}^{K}\left|\frac{\widetilde{B}_k(x)}{C_k} - \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i}\right| \\ \text{s.t.}\quad & \widetilde{B}_k(x) \le C_k \quad \text{for all } k\in\mathcal{S} \\ & x_n \in \{0,1\} \quad \text{for all } n\in Q.\end{aligned} \tag{5}$$

Problem (5) is an integer program with a nonlinear objective. We can linearize the objective by introducing the continuous variables y_k, for all $k\in\mathcal{S}$

$$y_k = \left|\frac{\widetilde{B}_k(x)}{C_k} - \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i}\right| \tag{6}$$

and adding the constraints

$$\begin{aligned}-y_k \le \frac{\widetilde{B}_k(x)}{C_k} - \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i} \le y_k \\ y_k \ge 0.\end{aligned} \tag{7}$$

By minimizing y_k, the constraints (7) guarantee that y_k will take the value of the right hand side of (6). The optimization problem becomes

$$\begin{aligned}\min_x & \sum_{k=1}^{K} y_k \\ \text{s.t.}\quad & \frac{\widetilde{B}_k(x)}{C_k} - \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i} \le y_k \quad \text{for all } k\in\mathcal{S} \\ & -\frac{\widetilde{B}_k(x)}{C_k} + \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i} \le y_k \quad \text{for all } k\in\mathcal{S} \\ & \widetilde{B}_k(x) \le C_k \quad \text{for all } k\in\mathcal{S} \\ & x_n\in\{0,1\},\ y_k \ge 0 \quad \text{for all } n\in Q,\ k\in\mathcal{S}.\end{aligned} \tag{8}$$

Problem (8) is a linear mixed integer program with binary and continuous variables.

B. Untargeted VHO Requests

We consider a batch of N untargeted requests that have a predefined origin system and all other systems in $\mathcal{S} = \{1,\cdots,K\}$ are potential destinations. We use the notation of Section IV-A and let $Q_i \subseteq Q$ be the set of requests with origin system i. For each request n with origin i, we define the decision variables $x_n^1,\cdots,x_n^K$ such that $\sum_{j=1}^{K} x_n^j = 1$, where $x_n^j = 1$ for $j\neq i$ indicates that request n is accepted to system j and $x_n^i = 1$ indicates that the resource of request n remains in system i.

Then the available resource for system k after the batch allocation, $\widetilde{B}_k$, is as follows

$$\begin{aligned}\widetilde{B}_k(x) =& B_k + \sum_{i\in\mathcal{S}:i\neq k}\ \sum_{n\in Q_i} x_n^k f_k(b_n) - \\ &\sum_{i\in\mathcal{S}:i\neq k}\ \sum_{n\in Q_k} x_n^i f_k(b_n) \quad \text{for all } k\in\mathcal{S}.\end{aligned} \tag{9}$$

As in Section IV-A, after linearizing the objective, the optimization problem for the untargeted case can be written as follows

$$\begin{aligned}\min_x & \sum_{k=1}^{K} y_k \\ \text{s.t.}\quad & \frac{\widetilde{B}_k(x)}{C_k} - \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i} \le y_k \quad \text{for all } k\in\mathcal{S} \\ & -\frac{\widetilde{B}_k(x)}{C_k} + \frac{1}{K}\sum_{i=1}^{K}\frac{\widetilde{B}_i(x)}{C_i} \le y_k \quad \text{for all } k\in\mathcal{S} \\ & \widetilde{B}_k(x) \le C_k \quad \text{for all } k\in\mathcal{S} \\ & \sum_{j=1}^{K} x_n^j = 1, \\ & x_n\in\{0,1\},\ y_k \ge 0 \quad \text{for all } n\in Q,\ k\in\mathcal{S}.\end{aligned} \tag{10}$$

C. Extensions

In this section, we provide extensions to problems (8) and (10) that include in the objective, terms that take into account acceptance rates of VHO requests and path cost. Without loss of generality, we provide the extensions for the targeted problem (8) and omit the untargeted case.

The objective of problem (8) can be rewritten to take into account the acceptance rates of VHO requests as follows

$$\min_x \sum_{k=1}^{K} y_k - \alpha\sum_{n=1}^{N} x_n w_n. \tag{11}$$

The second term in the objective attempts to maximize the weighted number of accepted requests. The acceptance coefficient, denoted by α, defines the level at which acceptance decisions are either traded off with load balancing decisions or become a second priority to the objective function. When α is several orders of magnitude smaller than the load balancing term, then it serves as a tool to encourage swaps of VHO requests between systems that do not affect the load balancing of the system. When α is of similar magnitude to the load balancing term, then there is a trade of between load balancing and acceptance rates. The value of α can be set according to network operator policies. The coefficients w_n can be equal or they can be set to the bit rate b_n to prioritize acceptance of VHO requests with high bit rates.

Let c_n^o be the current path cost of request n at its origin system and let c_n^d be the path cost of request n at the potential destination system. Then, the objective of problem (8) can be rewritten to take into account the path costs as follows

$$\min_{x} \sum_{k=1}^{K} y_k + \beta \sum_{n=1}^{N} \left(c_n^d - c_n^o\right) x_n. \qquad (12)$$

The second term in the objective attempts to minimize the increase in the path cost by accepting requests. The coefficient β determines the trade off level between load balancing and path cost decisions.

V. NUMERICAL INVESTIGATIONS

The aim of the conducted experiments, which are discussed in this section, is to quantify the performance gains of the proposed VHBP scheme in an environment with four cooperative access networks under a number of different network conditions. We focus on the optimization problem with objective (11), where both load balancing and acceptance rate terms are taken into account. The acceptance coefficient is taken as $\alpha = 10^{-10}$, which, as explained in Section IV-C, gives priority to load balancing decisions. The optimization problems are solved using AMPL-CPLEX 9.0. With pre-defined maximum delay tolerance per VHO request equal to $D = 500$ msec the maximum batch size is 35 for the targeted case and 14 for the untargeted. For the experiments, we assume that system resources are simply bit rates, i.e., $f_k(b) = b$ for all $k \in \mathcal{S}$. We explore the effect of the proposed scheme in scenarios with different initial load balancing objective value of 0.07, 0.35 and 0.6, 1.0. These levels depict the degree of unbalance between systems: For four systems, the worst case scenario in terms of load balancing objective value is 2.0 (two systems fully utilized and two systems unutilized). The bit rates in kbps of the VHO requests, b_n, are taken from the discrete set $\mathcal{R} = \{8, 16, 32, 64, 128, 256, 512\}$ with respective probabilities $\{0.07, 0.07, 0.14, 0.14, 0.14, 0.22, 0.22\}$. The origin destination pairs for the targeted VHO requests and the origins of the untargeted requests are selected uniformly.

The VHBP scheme is compared with an enhanced vertical handover first come first served (VHFCFS) scheme. In the simple FCFS scheme, a request is accepted if it results in a better load balancing objective value. Because this scheme will always reject a request when the system is balanced, we consider the following enhanced scheme for comparison. In the VHFCFS scheme, requests are also accepted when the new objective value is less than 0.01 (a case of an almost balanced system). Further, the scheme also accepts requests when the current objective value is relatively low (less than 0.02) and the new objective value is within a 50% percent deviation from the current one. These enhancements allow the FCFS scheme to increase the acceptance rate when the system is in an almost balanced state.

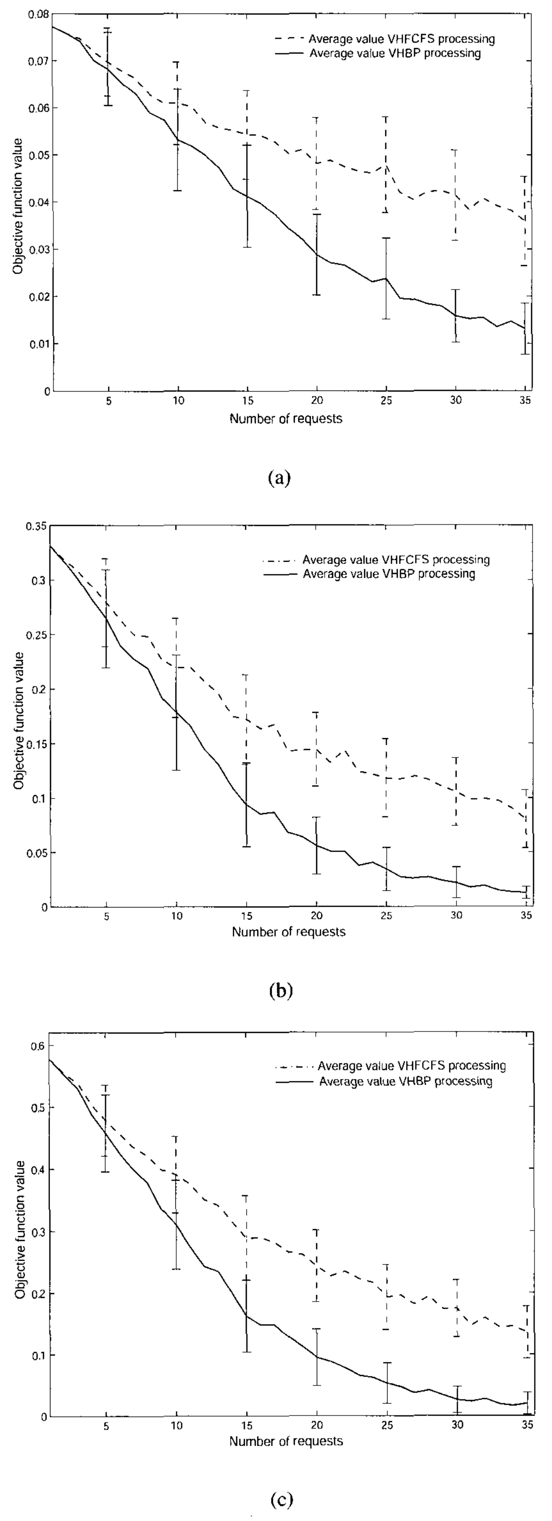
We perform two types of experiments. The first type evaluates performance gains of the VHBP scheme after a single batch of VHO requests. Figs. 2 (targeted scenario) and 3 (untargeted scenario) show the average objective value and standard deviation for different batch sizes for the VHBP and VHFCFS schemes for 100 Monte Carlo simulations. As can be seen from Figs. 2 and 3 the VHBP scheme significantly outperforms the VHFCFS scheme for both targeted and untargeted scenarios. For example, in the targeted scenario and batch size of 20 requests the average gain of the VHBP scheme is approximately 54%. Similarly, for the untargeted scenario and for a batch size of 10 requests the average gain of the VHBP scheme is approximately 43%. Also, the performance gains of the VHBP scheme increase as the batch size increases, since more information is available to the optimization problem for decision making.

The acceptance rate of the two schemes for the same set of experiments is shown in Figs. 4 (targeted scenario) and 5 (untargeted scenario). The figures depict the percentage of accepted requests over the total number of requests for different batch sizes. For the VHFCFS scheme it is interesting to observe that the acceptance rate decreases with the number of requests. This is because as the number of requests increases the load balancing objective value decreases (see Figs. 2 and 3), therefore with a more balanced system it accepts less number of requests. On the other hand, in the VHBP scheme the acceptance rate significantly increases with the batch size. This is because increased batch sizes allows the VHBP scheme to perform a larger number of swaps between the VHO requests, without affecting the load balancing objective. For a batch size of 20 requests in the targeted scenario the VHBP scheme provide an average improvement of 61% and in the untargeted scenario for a batch size of 10 requests the average improvement is 60%.

In the second type of the experiments, we want to show how the system evolves over time with different initial load balancing objective values. We evaluate the performance gains of the VHBP scheme after 100 consecutive batches of VHO requests of equal size. Tables 2 and 3 summarize the results for the targeted scenario. The performance gains of the VHBP scheme regarding the load balancing across different gains averages to 42%. At the same time, the acceptance rates of VHBP scheme are on average 4.8 times better from the VHFCFS. The results of the untargeted scenarios are summarized in Tables 4 and 5. The average performance gains of the VHBP scheme in this case are 38% on load balancing and 11.5 times improvement in acceptance rate.

VI. CONCLUSIONS

In this paper, we propose the vertical handover batch processing (VHBP) scheme for performing load balancing between heterogeneous cooperative radio access networks. This proposed scheme can be used both in urban hot spot areas where there is a high concentration of mobile users but also in cases where the VHO requests arrive by themselves as batches as in the case of moving networks (i.e., buses, trams, etc.). Since the proposed optimization approach allows batch handling of VHO requests, it provides significant performance gains with respect to traditional approaches where VHO requests are handled individually. For example, as we have demonstrated in the numerical investigations, if VHO requests are batch processed rather than

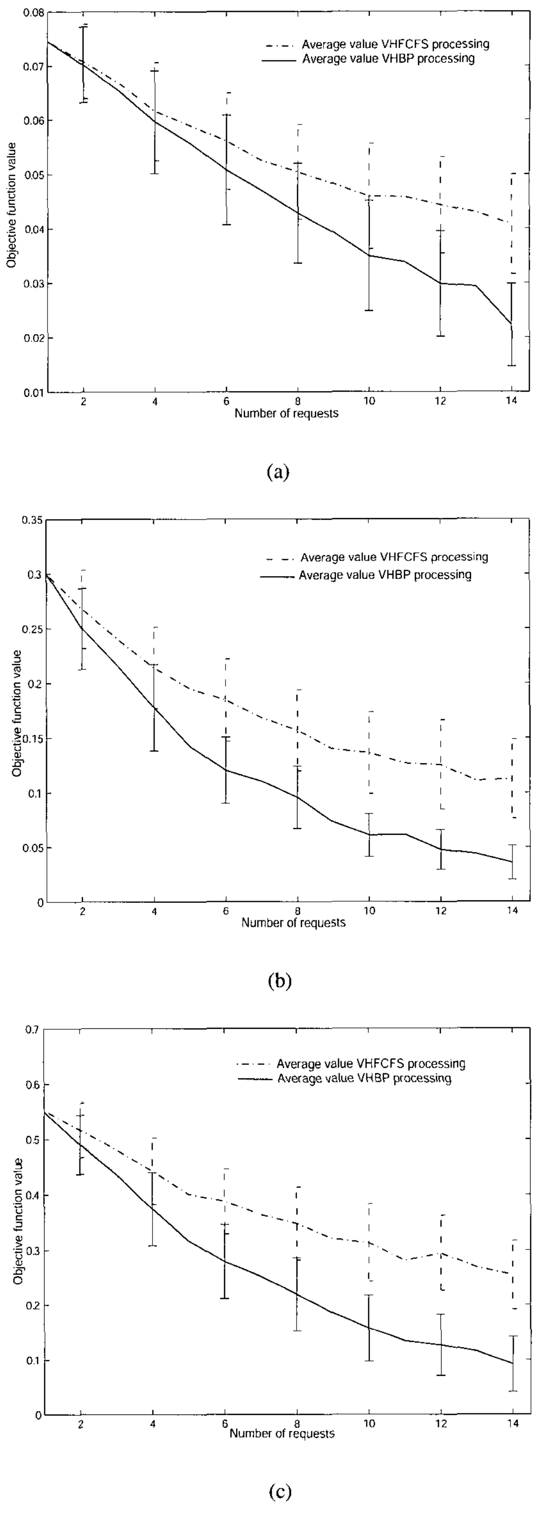


(a)

(b)

(c)

Fig. 2. Objective values vs. different batch sizes for the VHBP scheme and for the VHFCFS processing scheme. Figs. (a), (b), and (c) are for the targeted scenarios with initial load balancing objective value of 0.07, 0.35, and 0.60, respectively.



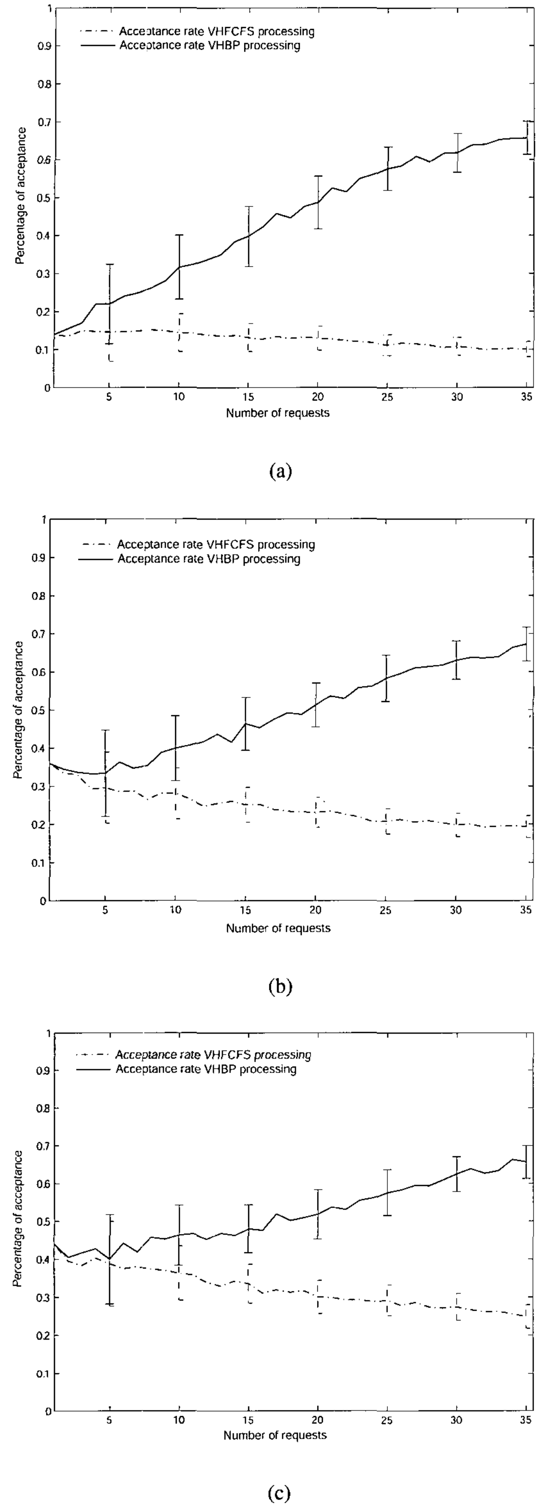
(a)

(b)

(c)

Fig. 3. Objective values vs. different batch sizes for the VHBP scheme and for the VHFCFS processing scheme. Figs. (a), (b), and (c) are for the untargeted scenarios with initial load balancing objective value of 0.07, 0.35, and 0.60, respectively.

handled individually the average performance gains in load balancing and acceptance rate is 43% and 60%, respectively. Since the proposed optimization framework can take into account load balancing, acceptance rate and route path cost on the backhaul network, an operator can weigh these objectives according to operational needs.

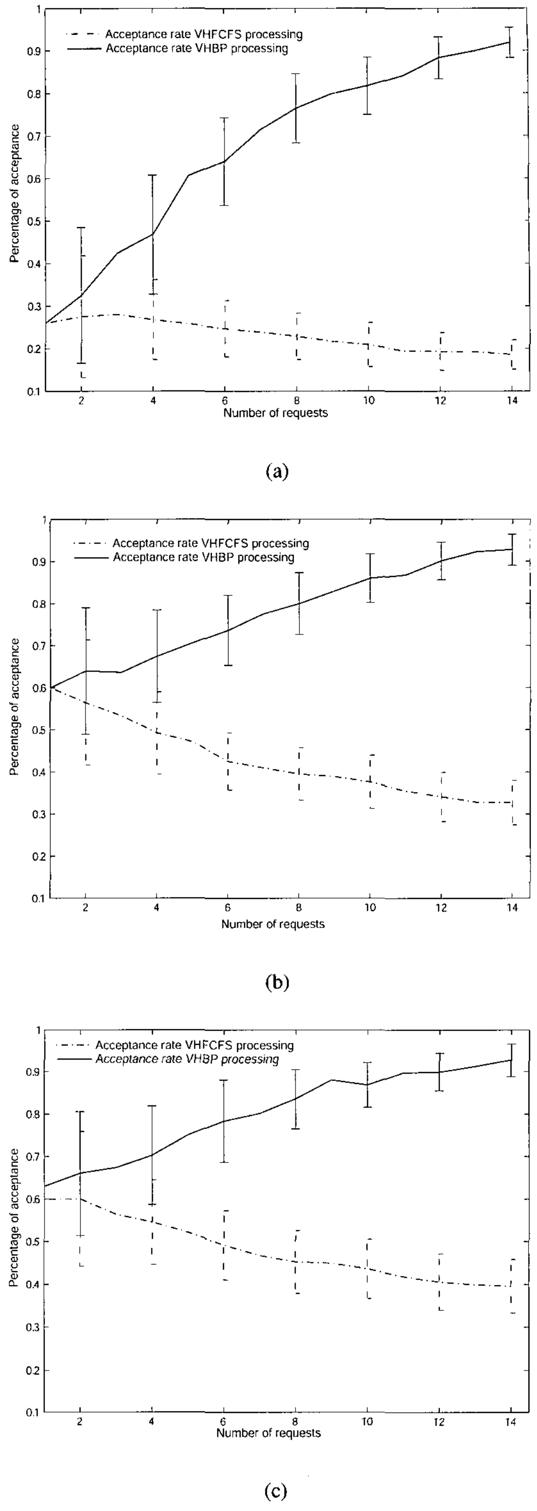


(a)

(b)

(c)

Fig. 4. VHO acceptance rates for the VHBP scheme and for the VHFCFS processing scheme. Figs. (a), (b), and (c) are for the targeted scenarios with initial load balancing objective values of 0.07, 0.35, and 0.60, respectively.



(a)

(b)

(c)

Fig. 5. VHO acceptance rates for the VHBP scheme and for the VHFCFS processing scheme. Figs. (a), (b), and (c) are for the untargeted scenarios with initial load balancing objective values of 0.07, 0.35, and 0.60, respectively.

REFERENCES

[1] "3GPP system to wireless local area network (WLAN) interworking, system description (release 6)," 3rd Generation Partnership Project, 3GPP TS 23.234 v6.0.0.

[2] IEEE P802.11 Wireless Interworking with External Networks (WIEN) Study Group, http://grouper.ieee.org/groups/802/11/Reports/wienupdate.htm.

[3] Y. Bejerano, S.-J. Han, and L. (Enran) Li, "Fairness and load balancing in wireless LANs using association control," in *Proc. MOBICOM 2004*,

Table 2. Performance evaluation of the VHBP and VHFCFS schemes in terms of the average objective value attained for the targeted scenario and for 100 consecutive batches of size N.

	Avg. objective value VHBP	Avg. objective value VHFCFS	Gain
Initial load balancing: Objective value 0.07			
$N=6$	0.0052	0.0097	0.46
$N=10$	0.0047	0.0083	0.43
$N=20$	0.0041	0.0072	0.43
Initial load balancing: Objective value 0.35			
$N=6$	0.0088	0.0151	0.41
$N=10$	0.0055	0.0129	0.57
$N=20$	0.0033	0.0127	0.74
Initial load balancing: Objective value 0.60			
$N=6$	0.0154	0.0203	0.24
$N=10$	0.0058	0.0135	0.57
$N=20$	0.0038	0.0097	0.61
Initial load balancing: Objective value 1.00			
$N=6$	0.0576	0.0604	0.05
$N=10$	0.0315	0.0391	0.19
$N=20$	0.0151	0.0230	0.34

Table 3. Performance evaluation of the VHBP and VHFCFS schemes in terms of VHO requests acceptance rates for the targeted scenario and for 100 consecutive batches of size N.

	acceptance rate VHBP	acceptance rate VHFCFS	Gain
Initial load balancing: Objective value 0.07			
$N=6$	0.0933	0.0417	1.24
$N=10$	0.1730	0.0330	4.24
$N=20$	0.4545	0.0235	18.34
Initial load balancing: Objective value 0.35			
$N=6$	0.1050	0.0667	0.58
$N=10$	0.1750	0.0590	1.97
$N=20$	0.4555	0.0545	7.36
Initial load balancing: Objective value 0.60			
$N=6$	0.1217	0.0733	0.66
$N=10$	0.1810	0.0520	2.48
$N=20$	0.4545	0.0330	12.77
Initial load balancing: Objective value 1.00			
$N=6$	0.1367	0.1033	0.32
$N=10$	0.2000	0.0820	1.44
$N=20$	0.4550	0.0630	6.22

Table 4. Performance evaluation of the VHBP and VHFCFS schemes in terms of the average objective value attained for the untargeted scenario and for 100 consecutive batches of size N.

	Avg. objective value VHBP	Avg. objective value VHFCFS	Gain
Initial load balancing: Objective value 0.07			
$N=6$	0.0045	0.0059	0.24
$N=10$	0.0042	0.0056	0.25
$N=12$	0.0037	0.0052	0.29
Initial load balancing: Objective value 0.35			
$N=6$	0.0041	0.0131	0.69
$N=10$	0.0033	0.0113	0.71
$N=12$	0.0030	0.0104	0.71
Initial load balancing: Objective value 0.60			
$N=6$	0.0049	0.0056	0.12
$N=10$	0.0040	0.0052	0.23
$N=12$	0.0038	0.0053	0.28
Initial load balancing: Objective value 1.00			
$N=6$	0.0324	0.0478	0.32
$N=10$	0.0186	0.0301	0.38
$N=12$	0.0176	0.0267	0.34

Table 5. Performance evaluation of the VHBP and VHFCFS schemes in terms of VHO requests acceptance rates for the untargeted scenario and for 100 consecutive batches of size N.

	acceptance rate VHBP	acceptance rate VHFCFS	Gain
Initial load balancing: Objective value 0.07			
$N=6$	0.5567	0.0517	9.77
$N=10$	0.7700	0.0420	17.33
$N=12$	0.5050	0.0225	21.44
Initial load balancing: Objective value 0.35			
$N=6$	0.5583	0.0950	4.88
$N=10$	0.7710	0.0890	7.66
$N=12$	0.5035	0.0495	9.17
Initial load balancing: Objective value 0.60			
$N=6$	0.5567	0.0467	10.93
$N=10$	0.7710	0.0390	18.77
$N=12$	0.5055	0.0210	23.07
Initial load balancing: Objective value 1.00			
$N=6$	0.5667	0.1417	3.00
$N=10$	0.7690	0.1180	5.52
$N=12$	0.5015	0.0635	6.90

Sept. 26–Oct. 1, Philadeplhia, Pennsylvania, USA, 2004.

[4] "Technical specification group radio access network improvement of RRM across RNS and RNS/BSS - release 5," 3rd Generation Partnership Project, 3GPP TR 25.881 V5.0.0.

[5] "IEEE trial-sse recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting IEEE 802.11 operation," IEEE Std. 802.11F, 2003.

[6] IEEE 802.11k Task Group, Radio Resource Measurement Enhancements, http://grouper.ieee.org/groups/802/11/index.html.

[7] IEEE 802.21 Media Independent Handover Interoperability, http://www.ieee802.org/21/.

[8] S. K. Das, S. K. Sen, and R. Jayaram, "A novel load balancing scheme for the tele-traffic hot spot problem in cellular networks," *ACM Monet Wireless Networks*, vol. 4, pp. 325–340, 1998.

[9] T. P. Chu and S. R. Rappaport, "Overlapping coverage with reuse particioning in cellular communications systems," *IEEE Trans. Veh. Technol.*, vol. 46, no. 1, pp. 41–54, 1997.

[10] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *Proc. INFOCOM 2003*, San Francisco, Apr. 1–3, 2003.

[11] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," *ACM Mobile Networks and Applications*, pp. 335–350, vol. 3, no. 4, 1998.

[12] S. Yu Hui and K. H. Yeung, "Challenges in the migration to 4G mobile systems," *IEEE Commun. Mag.*, Dec. 2003.

[13] K. Pahlavan, "Handoff in hybrid mobile data networks," *IEEE Pers. Commun.*, Apr. 2000.

[14] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, "Efficient mobility management

for vertical handoff between WWAN and WLAN," *IEEE Commun. Mag.*, vol. 41 , issue 11, pp. 102–108, Nov. 2003.

[15] V. K. Varma, S. Ramesh, K. D. Wong, M. Barton, G. Hayward, and J. A. Friedhoffer, "Mobility management in integrated UMTS/WLAN networks," in *Proc. IEEE ICC 2003*, vol. 2 , May 11–15, 2003, pp. 1048–1053.

[16] M. Ylianttila, M. Pande, J. Mkel, and P. Mhnen, "Optimization scheme for mobile users performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks," in *Proc. IEEE GLOBECOM 2001*, San Antonio Texas, 2001.

[17] J. McNair and F. Zhu, "Vertical handoffs in fourth-generation multinetwork environments," *IEEE Wireless Commun.*, pp. 8–15, June 2004.

[18] H. S. Park, S. H. Yoon, T. H. Kim, J. S. Park, M. S. Do, and J.-Y. Lee, "Vertical handoff procedure and algorithm between IEEE 802.11 WLAN and CDMA cellular network," in *Proc. CDMA Int. Conf.*, 2002, pp. 103–112.

[19] E. Yanmaz and O. K. Tonguz, "Dynamic load balancing and sharing performance of integrated wireless networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 5, June 2004.

[20] "Technical specification group radio access network, UTRAN I-ur interface RNSAP signaling (release 6)," 3rd Generation Partnership Project (3GPP), TS 25.423, version 6.2.0, June 2004.

[21] "D. Ayyagari and A. Ephremides, "Admission control with priorities: Approaches for multi-rate wireless systems," *Mobile Networks and Applications*, vol. 4, pp. 209–218, 1999.

Katerina Papadaki graduated from the University of California at Berkeley, with a double major in Pure Mathematics and Statistics in 1994. She completed the M.Sc. in Operational Research at the London School of Economics in Operational Research in 1996. She completed her Ph.D. at Princeton University (2001) in the Department of Operations Research and Financial Engineering, where she developed approximate dynamic programming algorithms for transportation problems. She is currently a lecturer at the Operational Research Department at London School of Economics (since 2002). Her research interests are resource allocation problems in wireless telecommunications.

Vasilis Friderikos graduated from the Aristotle University of Thessaloniki, Greece, Department of Electrical and Computer Engineering with major in Telecommunications in 1998. He completed the M.Sc. by Research in Telecommunications (with distinction) at the Centre for Telecommunications Research London) in 1999. During his Ph.D., he was working as a research associate in Mobile-VCE Core-2 research programme on algorithmic aspects of QoS enabled pure IP based mobile/wireless networks. He is currently a lecturer at King's College London and his research interests revolve around cross layer optimization algorithms with emphasis on scheduling for single or multi hop wireless packet transmission.