Congestion Aware Dynamic User Association in Heterogeneous Cellular Network: A Stochastic Decision Approach

Longwei Wang, Wen Chen
Department of Electronic Engineering,
Shanghai Jiao Tong University, China
E-mail: {mikesjtu007,wenchen}@sjtu.edu.cn

Jun Li
School of Electrical and Information
Engineering, University of Sydney, Australia
Email:jun.li1@sydney.edu.au

***Abstract*—In this paper, we propose a novel distributed optimization method for dynamic user association in a downlink Heterogeneous cellular network (Hetnet). We aim at maximizing the utilization of the base stations (BSs) in such a network by jointly considering the effect of the channel gains and load balancing of different BSs. Specifically, we first formulate the user association as a combinatorial optimization problem, whose global optimal solution is difficult to obtain. This is due to the high complexity caused by the large network scale and prohibitive signaling overhead. To address this issue, we then consider the optimization problem under the stochastic decision framework, and propose a distributed heuristic algorithm to independently and dynamically associate each user with the best BS. By posing a price factor to the BS evaluation update, the convergence of the heuristics is guaranteed. Numerical results indicate that the proposed heuristics can perform better than the conventional best-SNR (signal-to-noise ratio) method in the presence of large number of users, and achieves the nearly optimal solution.**

I. Introduction

To improve the spectrum efficiency and energy efficiency of the mobile cellular networks, heterogeneous cellular network (Hetnet) is proposed as a promising technique, where macro base stations (BS) are overlaid with a set of low-power different types of BSs , including picocells (also called small cells in the literature), femtocells, and relay BSs [1][2]. The small cells can offload traffic from macro BSs and reduce the average distance between users and transmitters. The coexistence of different BSs with different power levels and different cell sizes can provide significant gains by improving QoS of cell edge users, creating spatial reuse, and enable more flexible self-organized network structure[3].

We find that cell association has been investigated for load balancing in the conventional homogeneous cellular networks. In [6], it considers the scenarios that the base station serve a single user at each time slot, and then centralized and distributed algorithms are derived. In [7], it formulates the cell association problem as a Boolean linear programming and consider the proportional fairness among the users. Then the optimal solution is derived. However, The more complicated heterogeneous cellular network is different from the conventional homogeneous scenarios. It works by incorporating different BS transmitting powers and the cell structure is asymmetric and hierarchical.

The problem of associating users to either the macro BSs or the small BSs is of crucial importance in the Hetnet. The standard approach to this problem in 3GPP Long Term Evolution networks is to associate a user to the BS that provides the best signal-to-noise ratio (SNR) which is the best-SNR heuristic. The main advantage of this approach is that it keeps a good signal-to-noise ratio (SNR) for each user, while the disadvantage is that users tend to be mainly associated to the macro cell [4] with high transmit power. It will lead to load imbalance and inefficient utilization of the assigned channel. To address this problem, Many association methods have been proposed. They can perform better than the conventional rule in Hetnet [2][12]. However, they are based on different assumptions that makes it difficult to determine which one is the best.

In this paper, we address the problem on how to distributively associate the base station by decision making in the user side. We want to maximize the total BS utilization by jointly considering the effect of the channel gains and load balancing of different BS. We formulate this as a combinatorial optimization problem. Then a stochastic decision framework based distributed heuristic algorithm is proposed to approximate the optimal solution which can be derived by the conventional exhaustive search method. Since associating a user with one BS will affect the throughput of the other users associated with the same BS, by the proposed approach, the user ends can automatically select the optimal BS to access with only local information without a prior knowledge of the global information of all other users' SNR and the association strategies of other users. Finally, numerical results are presented which indicate its superior performance.

II. Network Model and Problem Formulation

A. Heterogeneous Cellular Network Model

In the conventional homogeneous cellular networks, the BSs are all equiped with the same transmitting power. However, this is not the case in HetNets, since the BSs of different tiers have widely divergent transmit powers.

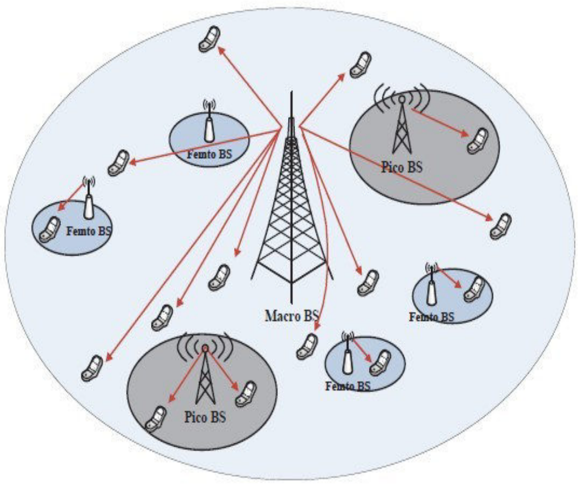

Fig. 1. The hierarchical hetnet scenario [2].

We consider the downlink communication in a simple HetNet hierarchical setting with one macro station and several pico and femto stations. There are overall M base stations with the mth base station, $m \in \{1, 2, \cdots, M\}$, transmitting at power level p_m. The various base stations transmit at various power levels $\mathbf{p} = [p_1, p_2, \cdots, p_M]$. There are totally N users distributed in the Hetnets. The nth user, $n \in \{1, 2, \cdots, N\}$, needs to choose the proper base stations to access. This scenario is illustrated in Fig. 1. The users are randomly distributed in the network area. Our goal is to find the optimal association method that maximizes the overall throughput, i.e. total latency or total throughput of all the users.

B. Problem Formulation

The M/M/C/C queuing model [8] is used to model the congestion among the users in the network. When a user requests access to a certain BS, but due to the limited channels assigned to each BS, the request is then blocked. We do not consider handovers and delay here. When the assigned channels are all occupied, the user do not wait until some channels are available [12]. The cell load is defined by all users' arrival rate and the number of channels assigned to each BS.

$$\Lambda = \frac{\rho}{c} = \frac{\lambda}{c\mu} = \frac{N'\lambda_u}{c\mu}, \tag{1}$$

where Λ is the traffic load, λ is the total arrival rate of users, N' is the number of users current requesting access, λ_u is the arrival rate of each user, μ is the service time of each user and c is the number of available channels for each BS. We can then give blocking (congestion) probability in each BS by using the Erlang-B formula [9] as:

$$P_{blocking} = \frac{\frac{\rho^c}{c!}}{\sum_{i=0}^{c} \frac{\rho^i}{i!}}. \tag{2}$$

Here we denote the blocking probability of base station m as P_m and the number of users served simultaneously by base station m as σ_m. P_m is a function of σ_m.

According to the Shannon capacity formula, the average transmission capacity for the channel between the mth BS and the nth user is

$$C_{mn} = E\{\log(1 + p_m h_{mn}/n_0)\}, \tag{3}$$

where h_{mn} and n_0 are the channel gains and noise, respectively. The average transmission capacity in (3) is thus multiplied by the factor $(1 - P_m)$.

We can find that each user's average capacity depends on other users' base station associaton. This motivates us to introduce an association indicator as

$$\alpha_{mn} = \begin{cases} 1, & \text{if user } n \text{ is associated with base-station } m, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

Then the optimization problem is formulated as

$$\max \sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_{mn}(1 - P_m) E\{\log(1 + p_m h_{mn}/n_0)\}, \tag{5}$$

$$s.t. \sum_{m=1}^{M} \alpha_{mn} \log(1 + p_m h_{mn}/n_0) \geq \beta_n, \tag{5a}$$

$$\sum_{m=1}^{M} \alpha_{mn} = 1, \forall n, \tag{5b}$$

$$\alpha_{mn} \in \{0, 1\}, \forall m, \forall n. \tag{5c}$$

where β_n is the QoS requirement of user n.

C. Computation Complexity

The above formulated problem is combinatorial due to the binary variable α_{mn}. To maximize the overall network throughput subject to a QoS and resource constraints, each user's association activity are coupled with each other. When the network size and user number are large, the problem is NP hard and can not be computed efficiently.

The complexity of the exhaustive search solution to the problem is $\Theta(m^n)$ [1], where m and n denote the number of BSs and the number of users, respectively. The computation complexity is impossible for even a not-so-large sized heterogeneous cellular network. To overcome this difficulty, we take advantage of the method of Fractional User Association (FUA) [2]. Users are allowed to be associated to more than one BSs. This is equivalent to the one where $\alpha_{mn} \in \{0, 1\}$ is replaced by $0 \leq \alpha_{mn} \leq 1$. This will allow us to come up with another solution to the problem.

III. Distributed Heuristic Approach for Dynamic User Association

To solve the optimization problem (5), we need to know the global information of channel gains and all users' association strategy. However, in most cases, the global information is

[1]it means that there are two positive numbers c_1 and c_2 such that $c_1 m^n \leq \Theta(m^n) \leq c_2 m^n$.

not available. Even they are available, the system overhead would be prohibitively high which could not be practically implemented. Each user's achievable rate is affected by the base station's transmission power and all the users' station association strategies. Since there is no information exchange among the users and no central control unit, in order to cope with the unknown and dynamic user association, we propose a stochastic decision based distributed heuristic algorithm by which the users can learn from local experience and finally search for the optimal base station to access.

To circumvent the difficulty, we study the problem under the framework of Stochastic Decision (SD) Process [10]. Then the distributed heuristic approach is presented to distributively select the proper base stations at the user ends.

A. Stochastic Decision (SD) Formulation for User Association

Stochastic Decision (SD) provides a framework for studying the sequential optimization of discrete time stochastic systems in the presence of uncertainty. As the network state is unknown, the users try different actions, and receive relative rewards. Then the rewards are used to update the estimation for each action so as to guide the future decision making.

The stochastic decision process comprises of several parts, such as the set of users and actions (base stations), reward, action selection policy, and BS estimation updates. It can be described as $\mathcal{SD} = \{\mathcal{N}, \{\mathbf{a}_n\}_{n\in\mathcal{N}}, \{r_{mn}\}_{n\in\mathcal{N}}, \{\pi_n\}_{n\in\mathcal{N}}\}$, where $\mathcal{N} = \{1, 2, \cdots, N\}$ is the set of users dispersed in the cellular network, $\mathbf{a}_n = \{1, 2, \cdots, M\}$ is the set of available actions (the set of base stations) of user n, and r_{mn} is the reward function of user n, which reveals the current reward of the selected BS m and π_n is the BS selection policy, which maps from each state to an action $\mathbf{a}_n$.

The reward function is defined as the current capacity achieved by each user. Suppose that the nth user chooses the mth base station for transmission. Denote the number of users who select the mth BS as σ_m. Then the reward function of user n is given by:

$$r_{nm}(\sigma_m) = (1 - P_m(\sigma_m))E\{\log(1 + p_m h_{mn}/n_0)\}, \quad (6)$$

The congestion is captured by the decreasing property of r_{mn} with respect to σ_m. In the process of BS selection, the users should well balance the tradeoff between accessing the high power BS and probably severe congestion caused by other users. Suppose that at stage t in the stochastic decision process , the total throughput of the network is

$$U^t = \sum_{n\in\mathcal{N}} \sum_{m\in\mathbf{a}_n} r_{nm}\alpha_{mn}, \quad (7)$$

In $\mathcal{SD}$, the network intends to maximize the overall time average system throughput, i.e.,

$$\Phi_\pi = \lim_{T\to\infty} \frac{1}{T} E_\pi \left[\sum_{t=0}^{T-1} U^t \right], \quad (8)$$

We want to find an optimal stationary policy π^* that maximizes the average throughput Φ_π, i.e.,

$$\pi^* = \arg\max_{\pi} \Phi_\pi. \quad (9)$$

B. Distributed Heuristics based Optimal Base Station Selection Policy

The heuristic approach is a local optimal solution in a realistic environment, where users arrive and depart dynamically. The user makes decision by itself. In accordance with the continuous relaxation mentioned in Section II, we adopt a probabilistic strategy π_n for each user n to select the BS over $\mathbf{a}_n$ in which the BS m is randomly chosen according to the probability vector that depends on the estimation of each BS station.

From the networks perspective, the objective of user association is to optimize the overall throughput. However, the users are only interested in optimizing their own performance. Therefore, the user association among users are competitive and they will crowd to the best BS to improve their own transmission throughput, which will lead to congestion in the high transmit power BS. To regularize the association behavior of the users, we poss a price on the reward function of each user.

The congestion aware dynamic user association process is illustrated as follows:

- BS selection policy: At the user end, each user randomly choose a BS according to a stochastic policy. Instead of selecting a BS deterministically, the user n probabilistically choose a BS according to a probability vector $\pi_n = \mathbf{q}_n = (q_{n1}, q_{n2}, \cdots, q_{nM})$, where q_{nm} is the probability that the user n chooses BS m. If the BS can provide better performance, then higher probability will be given to the BS, or otherwise. In this way, the users will automatically turn to the best BS, respectively. The most common method of characterizing the probability vector components is to use a Boltzmann distribution [10]. The user choose the mth BS with probability

$$q_{nm} = \frac{\exp\{\gamma r_{nm}\}}{\sum_{k\in\mathbf{a}_n} \exp\{\gamma r_{nk}\}}, \quad (10)$$

  with $\sum q_{nm} = 1$ and γ is a positive parameter called the temperature.
- Price charging factor: It would cause severe congestion according to the best SNR association principle. The network's objective is to maximize the aggregate throughput of all users in (7). It will lead to poor performance without coordination. So here we charge a price on the BS m that the user selected, which is given by

$$p_{mn} \propto \sigma_m \quad (11)$$

  where p_{mn} is the price charged at the BS m when it has a total of σ_m users. So the reward can be given by $\tilde{r}^t_{mn} = r^t_{mn} - p_{mn}$. The price well captures the congestion effect among users and to some extent relieves it.
- BS estimation update: At the end of the stage t, each user only observes the immediate reward of the selected BS

Algorithm 1 : Heuristic Algorithm for Congestion Aware Dynamic User Association

Initialization: FOR each user $n \in \mathcal{N}$, each BS $m \in \mathbf{a}_n$.
Initialize $t = 0$ the BS station selection probability vector $q_{mn}^t = 1/M, \forall n \in \mathcal{N}, m \in \mathbf{a}_n$.

Repeat until convergent:

1: **BS selection:** at the beginning of each stage, each user selects a BS $\mathbf{a}_n$ according to its current BS selection probability vector. $\mathbf{q}_n$.
2: **Reward received:** after accessing the selected BS at the end of each stage, each user receives the reward given by (6).
3: **BS estimation update:** all the users update their BS selection probability according to the rule specified in (10) and prepare for next stage transmission.
4: **convergence condition:** stop until there exists a component of $\mathbf{q}_n$ approaching one; otherwise, continue update.

r_{mn}^t and updates its evaluation by

$$r_{mn}^{t+1} = \begin{cases} ((1-\epsilon)r_{mn}^t + \epsilon\widetilde{r}_{mn}^{t-1}, & if\ \mathbf{a}_n = m, \\ r_{mn}^t, & otherwise. \end{cases} \quad (12)$$

The proposed heuristic algorithm is summarized in **Algorithm 1**.

We now discuss the convergence of Algorithm 1. Due to the limited information obtained by each user and the user association are dynamic, we do not know whether the reward of each user are stable after certain slots. The following theorem provides the proof of convergence of the process.

Theorem 1: (**Convergence**) With $\gamma \rightarrow 0$, the term r_{mn} converges to r_{mn}^*. It can be calculated by the following differential equation:

$$\frac{\Delta r_{mn}}{\Delta t} = F(r_{mn}), \quad (13)$$

where $F(r_{mn})$ is the conditional expected function defined as

$$F(r_{mn}) = E[w_{mn}(r_n)|r_{mn}], \quad (14)$$

Proof: The (12) can be written as

$$r_{mn}^{t+1} - r_{mn}^t = \gamma(w_{mn} - r_{mn}^t), \quad (15)$$

where

$$w_{mn} = \begin{cases} \widetilde{r}_{mn}^{t-1}, & if\ \mathbf{a}_n = m; \\ r_{mn}^t, & otherwise. \end{cases} \quad (16)$$

The outcome of (13) can be determined by

$$\frac{\Delta r_{mn}}{\Delta t} = E(w_{mn}|r_{mn}), \quad (17)$$

Based on the dynamics of (17), refer to [11] to obtain the convergence. ■

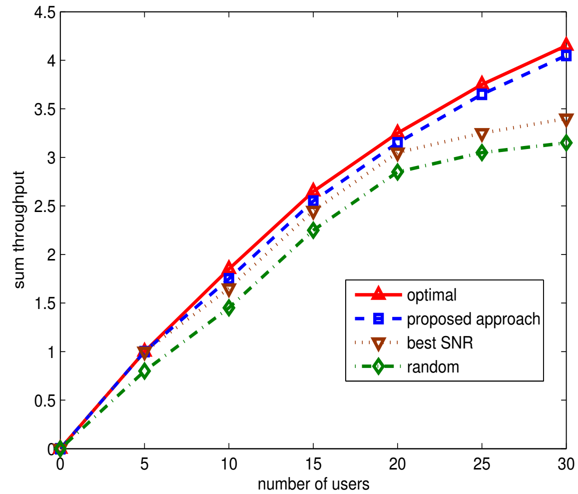

Fig. 2. Comparison between different methods.

IV. Numerical results and performance analysis

A. Performance Comparison

In a simulation-based scenarios, we consider a heterogeneous cellular network within a macro-cell overlayed with one pico-cell and a femto-cell. The transmit power of the macro BS is 40W and the power of the pico BS and femto BS are 1W and 0.1W, respectively. The users are are randomly dispersed in the area. As proposed in [9], the average channel gain between the BS m and a user n is given by $h_{mn} = -(128.1 + 37.6\log(d_{mn}))$. The channel gain is a function of the distance between two points. The users randomly arrive according to poisson process, and stay in the BS according to the exponential distribution.

We compare the sum throughput achieved by the proposed heuristic algorithm with that obtained by the best SNR principle, the random selection method with equal probability over all BSs, and the optimal solution with global information by exhaustive search. The achievable capacity at each time step is assumed as a time averaged transmission capacity in this simulation.

With the increasing time steps, it is shown in Fig. 2, that the system throughput achieved by the proposed heuristic algorithm outperforms that of the best SNR approach and the random selection method. This is because that the heuristic process incorporates the congestion factor. Because of congestion aware BS selection, the users could avoid crowding into the high SNR BSs, and the low SNR BSs can be well utilized which will make the small cell BSs be fully utilized. Thus, the overall system throughput is largely improved. Note that when the number of users is small, the advantage of the proposed method is not obvious since there is no congestion in the BSs.

B. Analysis of the Algorithm

To capture congestion case, we consider a complex scenario involving 50 users and 3 BSs. We focus on the association

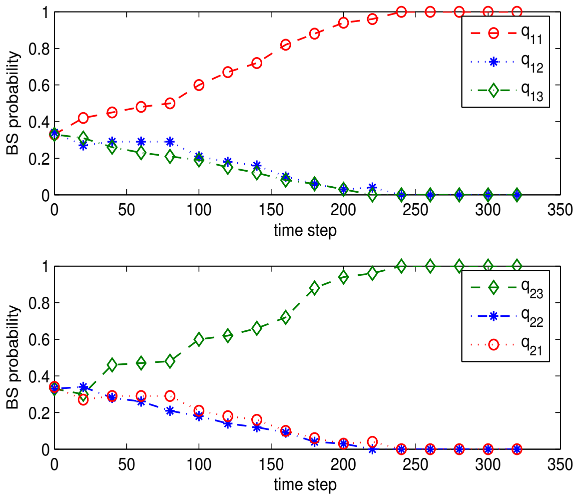

Fig. 3. Convergence performance.

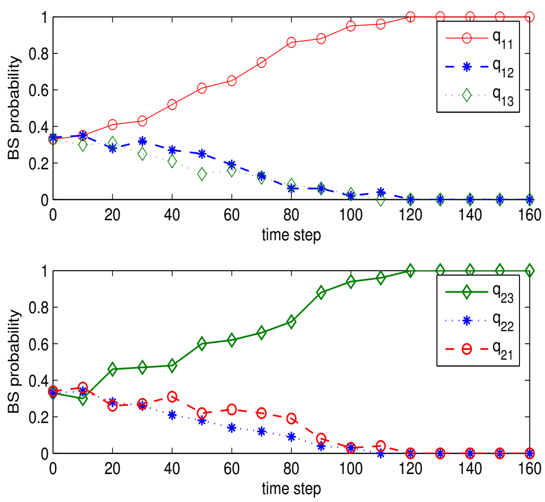

Fig. 4. The convergence rates when the channel gains exhibit high diversity.

behavior of three users involving in the process. According to Fig. 3, the symbol $\circ$, $*$, and $\diamond$ stand for BS 1, 2 and 3, respectively. It is shown clearly that each user selects a BS with almost probability 1 which indicates the convergence of the heuristic process.

When the number of BSs are very large, to reduce the complexity of heuristic process, the users can select among the BSs that can provide the SNR which could satisfy the its QoS requirement.

Also, we consider the effect of diversity among the SNR of the users for a given BS. When the SNR between users are much different from each other. The convergence rate of the heuristic algorithm becomes much faster, which is shown in Fig. 4. The convergence time is less than half of the previous one. This indicates a potential direction to improve the proposed approach in Fig. 3.

V. Conclusions

In this paper, we investigate the problem of user association in heterogeneous cellular network with hierarchical transmit power settings. When the number of users and BSs are large, it is computationally prohibitive. To alleviate the congestion caused by the traditional best SNR principle, we propose a distributed heuristic algorithm that makes the users autonomously and dynamically associate with different BSs. By posing a price to the BS evaluation update, the convergence of the heuristic process can be guaranteed. Also, the coordination information requirement is kept minimal. However, we just make an initial investigation in this paper and there are further problems to be studied. For example, here we assume that the BSs are with fixed transmit power, while in practice, the BSs may be dynamically switching. So for future work, more practical scenarios will be studied, including imperfect BS evaluation estimation, and overhead performance tradeoff.

Acknowledgment

This work is supported by the National 973 Project #2012CB316106, by NSF China #61161130529 and #61328101, by STCSM Science and Technology Innovation Program #13510711200, and by SEU National Key Lab on Mobile Communications #2013D11.

References

[1] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, ”Femtocells: past, present, and future,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497C508, Apr. 2012.

[2] J. G. Andrews, Sarabjot Singh, Qiaoyang Ye, Xingqin Lin, and Harpreet Dhillon, ”An Overview of Load Balancing in HetNets: Old Myths and Open Problems”, Submitted to IEEE Wireless Comm. Mag, July 2013

[3] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, ”A survey on 3GPP heterogeneous networks,” *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10C21, June 2011.

[4] X. Wu, B. Murherjee, and D. Ghosal, ”Hierarchical architectures in the third-generation cellular network,” *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 62C71, June 2004.

[5] S. Elayoubi, E. Altman, M. Haddad, and Z. Altman, ”A hybrid decision approach for the association problem in heterogeneous networks,” in *IEEE INFOCOM*, pp. 1-5, Mar. 2010.

[6] S. Das, H. Viswanathan, and G. Rittenhouse, ”Dynamic load balancing through coordinated scheduling in packet data systems,” in *Proc. IEEE INFOCOM 2003*, march-april 2003, pp. 786 C 796 vol.1.

[7] K. Son, S. Chong, and G. Veciana, ”Dynamic association for load balancing and interference avoidance in multi-cell networks,” *IEEE Trans. on Wireless Commun.*, vol. 8, pp. 3566 C 3576, July 2009.

[8] Sheldon Ross, ”Stochastic Processes,” 2nd Edition.

[9] Dimitri Bertsekas and Robert Gallager , ”Data Networks,” 2nd Edition.

[10] John Rust, ”Stochastic Decision Processes: Theory, Computation, and Empirical Applications,” Mini-course at Tel Aviv University, March 2000.

[11] P.Sastry, et al, ”Decentralized learning of nash equilibria in multi-person stochastic games with incomplete information,” *IEEE Trans. Syst., Man, Cybern. B*, vol. 24, no. 5, pp. 769-777, May 1994.

[12] Omar A. Elgendy, Mahmoud H. Ismail and Khaled Elsayed, ”On the Relay Placement Problem in a Multi-cell LTE Advanced System with Co-channel Interference ,” *The 8th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2012).*