ICT-MobileSummit 2009 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2009
ISBN: 978-1-905824-12-0

A Novel Autonomous RAT Selection Algorithm for Non Real Time Services

Jordi PÉREZ-ROMERO, Oriol SALLENT
Dept. Signal Theory and Communications (TSC)
Universitat Politècnica de Catalunya (UPC), c/ Jordi Girona, 1-3, 08034, Barcelona, Spain
Email: [jorperez, sallent]@tsc.upc.edu

Abstract: This paper proposes a new solution for autonomous decision-making in the RAT selection process in multi-service multi-access network scenarios. It is based on exploiting the time dimension in the RAT selection, so that, based on the current mobile terminal and network context and its future evolution, terminals can choose the appropriate initiation time to start a given transmission, waiting for the arrival to the coverage area of the most suitable RAT. The proposed strategy is claimed to have applicability for non-real time services without stringent deadline constraints (e.g. e-mails, downloading MP3 files, etc.). The algorithm is evaluated through simulations in different scenarios, revealing that it provides remarkable capacity increases, while at the same time ensuring that non real time transmissions do not exceed specific delay bounds.

Keywords: RAT selection, autonomic decision-making, JRRM.

1. Introduction

As new emerging technologies enter the wireless arena, operators and manufacturers are faced with the challenging task of jointly managing, as efficiently as possible, the pool of resources offered by several Radio Access Technologies (RATs) covering the same area. Each of these RATs, in turn, may provide different Quality-of-Service (QoS) levels for each of the supported services, thus adding a higher complexity to the problem of allocating a particular service to the RAT that is best-suited to support it. This network heterogeneity has indeed received a huge attention in the recent years [1][2]. The set of mechanisms devoted to the efficient and utmost use of radio resources is typically denoted as Radio Resource Management (RRM), term that has been conveniently redefined to Joint RRM (JRRM) in order to capture the necessity of managing the total amount of resources offered by the different RATs in a coordinated way rather than considering each RAT as a stand-alone entity. Benefits of JRRM have been assessed in a number of papers, see e.g. [3]-[6].

Traditionally, (J)RRM functions in wireless networks have been mainly centralized. This can be justified because a central network node may have a more complete picture of the radio access status than a particular node, so that decisions can be made with more inputs. However, a centralized (J)RRM implementation has some drawbacks in terms of increased signalling load or transfer delay of the (J)RRM algorithm inputs to the central node. This prevents an efficient implementation of short-term (J)RRM functions such as packet scheduling and explains why wireless cellular technology evolution (e.g. HSDPA) exhibits the trend towards implementing (J)RRM functions on the radio access network edge nodes (e.g. base stations). Indeed, there is a clear trend towards decentralized and autonomous (J)RRM functions in the mobile devices. This approach has claimed to be inefficient in the past because of the limited information available at the mobile device side (e.g. the mobile device does not know what is the cell load). Nevertheless, this can be overcome if the network is able to provide some information or guidelines to the mobile device assisting its decisions. Then, while a mobile-assisted centralized decision making process requires the inputs from many mobile devices to a single node, the network-assisted decentralized decision making process requires the input from a single node to the mobile devices, which can be significantly more efficient from a signalling point of view.

Some approaches to decentralized (J)RRM solutions can be found in the literature. In [7] a decentralized RAT selection scheme is presented for CDMA/TDMA networks. The mobile device takes the decision on which RAT to connect based on its own path loss measurements and the information broadcast through radio enabler IEEE P1900.4, in terms of a certain maximum path loss. In [8] the role of the Cognitive Pilot Channel (CPC) as support to JRRM is stressed, enabling the implementation of a decentralized fuzzy neural JRRM. Game theoretic approaches have also been considered in [9].

This paper proposes and evaluates a novel solution for autonomous JRRM, which is claimed to bring substantial gains in the overall system performance. The proposed approach is inspired on the Infostation concept, which has been proposed and analyzed in several papers in the literature [10]-[12]. An Infostation is seen as a small cell providing a high bandwidth radio link for data services, usually overlaying with other low bit-rate networks. While in the formulation of the concept back in late 90s the Infostation was seen as a specific technology to be deployed, this paper exploits the fact that the evolution of the wireless arena has brought a heterogeneous network scenario, where some of the technologies are capturing the capabilities envisaged for Infostations. For example, WiFi offers high bit rate with reduced coverage in a scenario where UMTS R99 offers lower data rates for extended coverage and mobility conditions. In this context, the strategy developed in this paper will let the mobile terminal take decisions, with some support from the network side, about the suitability to establish a data connection with one or another of the available RATs in a given heterogeneous scenario.

The rest of the paper is structured as follows. Section 2 presents the description of the proposed autonomous JRRM algorithm as part of a more general framework on how to exploit the time dimension in different functions. The simulation model and the scenarios in which algorithm performance is evaluated are presented in Section 3, while the obtained results are discussed in Section 4. Finally conclusions are summarised in Section 5.

2. Autonomous JRRM Algorithm

The considered scenario involves a number of RATs, offering diverse characteristics in terms of achievable bit rate and coverage area. Usually, a RAT selection algorithm is in charge of deciding “what” RAT to choose for a given call/session. In turn, the decision about “when” to set-up the connection is given for granted, since the connection is typically established on the selected RAT right after the service has been requested. In this sense, one of the main contributions of the proposed JRRM algorithm is that it performs a RAT selection strategy intending to exploit the flexibility of services without stringent delay requirements, so that the “when” dimension is questioned and the algorithm may decide to wait for a later time to activate the requested service. Therefore, the autonomous decision making process executed at the mobile terminal introduces an additional dimension in the decision making process by deciding the most appropriate instant to start data transmission (i.e. the Initiation Transmission Time, ITT) based on the current context information (i.e. the RATs and operators available) and the estimated future evolution of this context information.

A practical applicability example could be the following one: a user activates a certain non-real time (NRT) service (e.g. sending a number of e-mails, downloading a MP3 file, etc.) in a position where only a cellular system that offers a relatively low bit rate is available. By analyzing context information, the terminal may estimate that, after some time, it may reach the coverage area of a high bit rate RAT belonging to a different operator offering a cheaper price. Then, the decision process analyzes the different conditions and requirements (e.g. available bit rates in each technology, reliability in the context evolution prediction, estimated time until arriving to high bit rate RAT coverage area, etc.) in order to decide whether it is more convenient to start the transmission in the current position or to wait for a later time, with the expectation to reach the coverage area of the high bit rate RAT. In this case, the user could benefit from lower cost by deferring the transmission while meeting the QoS expectation (e.g. the user does not mind if the e-mails are kept in the outcome e-mail box for some time). Besides, in case that both RATs belong to the same operator, the operator itself could also benefit from this decision if lower interference is caused and, consequently, capacity is increased.

2.1 General Framework for the Inclusion of ITT in JRRM Functions

Assuming that the RAT selection algorithm is executed at instant t_0 and that there exists a maximum deadline T_e for the delivery of the information (note that, typically, for the type of NRT applications using this strategy the maximum deadline T_e will be relatively large, even several minutes), the decision taken by the algorithm should be based on the evolution of the context information in the interval $[t_0,T_e]$. This includes, on the one hand, the RATs available and the corresponding achievable bit rates, which in turn depend on the mobile position, speed and direction, and on the other hand the amount of information to be sent.

In general, when the algorithm is executed at t_0, the predictability on available operators, RATs and cells at a future time t_1 may be difficult, particularly if $t_1 >> t_0$. Therefore, predictions performed at t_0 as for the context estimated in t_1 should only consider a macroscopic view of the scenario. This means that it is probably sufficient to have a knowledge of which operators and RATs there will be available in a given region (i.e. the macroscopic view) while the details in terms of the specific cells and even intra-cell positions (i.e. the microscopic view) will only be considered once the terminal is close enough to the desired RAT. In this way, it is assumed that the context information follows a layered structure as depicted in Figure 1. In the following, a description of the proposed algorithm is given focusing only on the macroscopic layer, so that only the RAT and operator are selected.

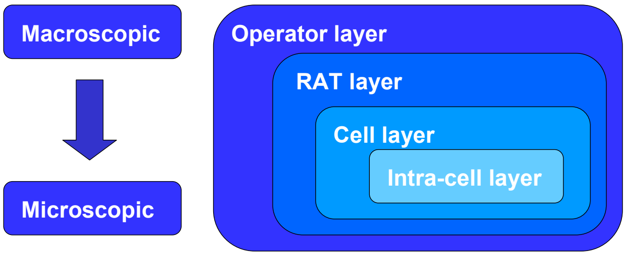


Figure 1: Layered Contextual Information

2.2 Algorithm Inputs

According to the aforementioned framework, it is considered that the terminal takes RAT selection decisions based on inputs obtained from the user and application, from its own location information and from the network, as detailed in the following:

User and application related inputs:

- User preference for operator *i* (Φ_i): This would be mainly related to how the user values the trade-off between price and QoS offered by the *i*-th operator.
- User preference for RAT *j* (ξ_j): Again, this would be related with the trade-off between price and QoS offered by a certain technology, although in this case also other

considerations could be included (e.g. battery life when connected to one RAT or the other, etc.)

- Maximum deadline T_e to deliver the information: Notice that, assuming non real time services (e.g. sending emails, downloading MP3 files, etc.) this deadline could take relatively high values.
- Buffer size (L): It is the total amount of data to be transmitted by the application at a given time (e.g. the set of emails to send, the file to download, etc.).

Location-based inputs:

- Distance to reach RAT j (D_j): It is the distance between the current position of the mobile terminal and the coverage area of a given target RAT j.
- Coverage area of RAT j (R_j): It is the expected distance in which the terminal, following the current trajectory, will have coverage of RAT j.
- Mobile speed (v) and direction

Network-related input:

- Achievable bit rate in RAT j ($R_{b,j}$): It is an estimation of the bit rate that the mobile terminal can obtain when transmitting in RAT j.

From the above inputs, the terminal can compute the following metrics, which will be used in the feasibility conditions described in the next subsection:

- Time needed to reach the coverage area of RAT j: $\tau_j=D_j/v$
- Transit time in the coverage area of RAT j: $\varepsilon_j=R_j/v$
- Time needed to transmit all the data in RAT j: $\delta_j=L/R_{b,j}$

In addition, operator and RAT preferences (Φ_i,ξ_j) can be combined into a *preference metric* for operator i and RAT j built as a function $P_{ij}=f(\Phi_i,\xi_j)$.

2.3 Feasibility Conditions

Assuming that the algorithm is evaluated at time t_0, the feasibility conditions to ensure that the data will be delivered before the deadline T_e can be summarized as:

- Condition i): This condition assesses if the total time to transmit the data is expected to fulfill the deadline, and is formulated as:

$$\alpha\left[\tau_j(t_0)+\delta_j(t_0)\right]\leq\left[T_e-t_0\right] \qquad (1)$$

where the time $\tau_j+\delta_j$ represents the total time to deliver the data if the terminal waits for reaching the coverage area of RAT j, T_e-t_0 is the remaining time and α is a parameter of the algorithm.

- Condition ii): This condition assesses if, once arrived to the RAT j coverage area, the transit time ε_j is expected to be enough to transmit all the data there.

$$\varepsilon_j(t_0)\geq\beta\times\delta_j(t_0) \qquad (2)$$

where β is a parameter of the algorithm.

2.4 Algorithm Procedure

Based on the above inputs and feasibility conditions, the proposed procedure to decide the time in which the transmission should start is shown in Figure 2. Once a given NRT service needs to be started, the first step consists in building the preference list of RAT/operator combinations in accordance with the previously defined preference metric P_{ij}. Then, the combination with the highest preference is selected. At this point of time, in case the selected RAT/operator is already available, transmission is started without waiting any more. On the contrary, if the RAT is not yet available, feasibility conditions 1 and 2 are checked to see if the selected RAT can be reached in a reasonable amount of time without compromising the maximum deadline. If one of the two conditions does not hold, then the next combination in order of preference is selected and the conditions are checked again. In turn, if both conditions hold, it means that it is worth to wait for the arrival to the coverage area of the selected RAT and therefore transmission is postponed. In this case, the algorithm simply checks again the two conditions periodically every ΔT, in order to account for possible context variations (e.g. changes in mobile direction, etc.). In this respect, even if the transmission has already started in a given RAT/operator, depending on these context variations, it could be possible to decide to stop it in order to resume it after some time in another more convenient RAT.

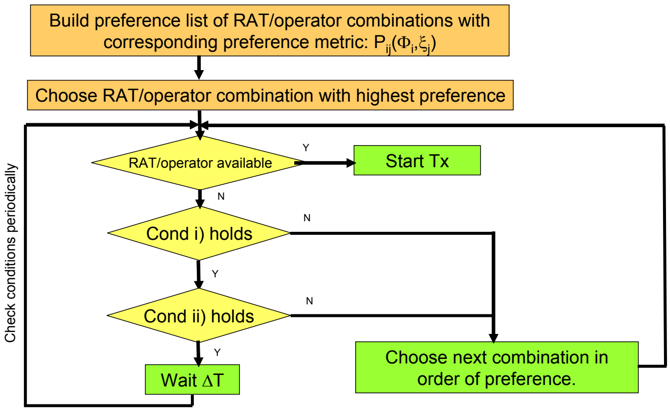


Figure 2: Algorithm procedure

3. Scenarios for Performance Evaluation

3.1 Scenario 1: Main Road

This scenario assumes a main road with 10 base stations equally distributed with separation 1 km and the terminals moving at a constant speed. All base stations support RAT1 being UMTS Release 99 (R99) and, in addition, central base station also supports RAT2 being HSUPA (High Speed Uplink Packet Access) in a different carrier than R99. A single operator is assumed. Note that this scenario allows a very accurate prediction of the trajectories and the instants when the terminals will enter the coverage area of the desired RATs, and consequently performance here can be regarded as an upper bound of the performance that can be obtained in other situations.

Two types of users are considered:

- Non Real Time (NRT) users: They try to deliver files with size L bytes in uplink direction with maximum deadline 300s. After successful delivery they randomly start a new transmission after an average time of 100s. They move with speed 50 km/h following the straight road in one of the two directions, randomly chosen at the beginning. When they are connected to UMTS R99 they get a maximum bit rate of 384 kb/s while when they are connected to HSUPA they get a maximum bit rate of 5.44 Mb/s. Note that, depending on load and propagation conditions, instantaneous bit rate can be lower. The preference metric of the proposed algorithm is such that, for these users RAT2 (i.e. HSUPA) is preferred.
- Real Time (RT) users: They generate conversational calls with average call duration 3 min and average interarrival time 3 min. A total of 100 users are scattered through the scenario and move randomly at 3 km/h. They are only connected to RAT1 (UMTS R99) getting a constant bit rate of 64 kb/s in both uplink and downlink.

The maximum transmitted power by the mobile terminals is 21 dBm, assuming perfect power control, and the thermal noise power at the base station is -106 dBm. Propagation loss at distance d is given by L_p (dB)=128.1+37.6log d(km). Furthermore, algorithm parameters are $\alpha=\beta=1$, ΔT=1s. On the other hand, it is assumed that terminals are able to acquire a full knowledge of the available bit rates in each RAT.

3.2 – Scenario 2: suburban area

This scenario assumes a hexagonal layout with 2 km separation between base stations. 11 omnidirectional base stations are deployed in a rectangular area of 6 x 4.6 km^2. Like in the previous case, all the base stations support UMTS R99 as RAT1 and, in addition, the central cell also supports HSUPA as RAT2 in a separate carrier. The same types of RT and NRT users as in scenario 1 are considered. However, the mobility model for NRT users is now a random walk model, in which they follow straight trajectories and, after every km, the direction is randomly changed in an angle uniformly distributed between -θº and θº. Mobile speed is also 50 km/h. Transmitted powers, propagation models and algorithm parameters are the same as in scenario 1. Note that, as a difference from previous scenario, in this case trajectories cannot be accurately predicted, so the purpose here is to study the robustness of the algorithm with respect to this inaccurate knowledge of the context information.

4. Results

The proposed RAT selection algorithm is compared against a reference algorithm that assumes NRT users initiate the transmission of the file at generation time in the best RAT available at the moment. Then, if during transmission they reach the coverage area of RAT2, they switch to this RAT.

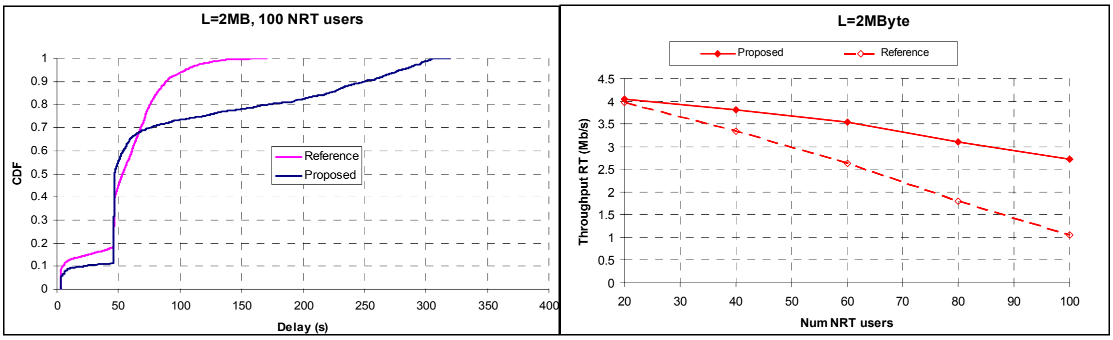
Figure 3(a) plots the Cumulative Distribution Function (CDF) of the delay in the case of transmitting files of L=2 MByte and with 100 NRT users in scenario 1. The effect of the algorithm is to increase the transmission delay due to the instants in which the transmission is postponed waiting for the coverage area of RAT2. However, the CDF also shows that the algorithm is able to take the appropriate decisions so that the maximum delay of 300s to deliver the information is not exceeded. In particular, the percentage of file transmissions exceeding the deadline of 300s is very small, around 1% with the proposed algorithm.

From the point of view of RT users, Figure 3(b) plots the total aggregated throughput of RT users in scenario 1 for the two algorithms. It can be observed that a very significant throughput increase is achieved thanks to the proposed algorithm, and this increases when increasing the number of NRT users in the scenario (e.g. up to 150% of improvement for the case with 100 NRT users). The reason for this throughput improvement is the fact that, with the proposed algorithm, NRT users are postponing their transmissions while transiting through the coverage area of RAT1 (UMTS R99). As a result, on the one hand they are not generating interference, which turns into a reduction in the packet error rate and, on the other hand, NRT users are not consuming resources from the RAT1 cells, and thus more RT users can be connected to them.

The impact of the file size L transmitted by NRT users is plot in Figure 4(a) in terms of RT throughput improvement achieved by the algorithm in scenario 1. A significant increase is observed, being particularly high for high values of L such as 5 MByte (up to 300% for the case of 100 NRT users). The reason is that, when increasing L for a given number of NRT users, the total load from these users increases, so that by postponing their transmissions more room is left for RT users in RAT1. In turn, Figure 4(b) compares the throughput increase in scenarios 1 and 2 (with θ=10º) for the case L= 2 MByte. The throughput improvements are smaller in scenario 2 because, with the random mobility model the algorithm decides in many situations to start the transmission in RAT1 since

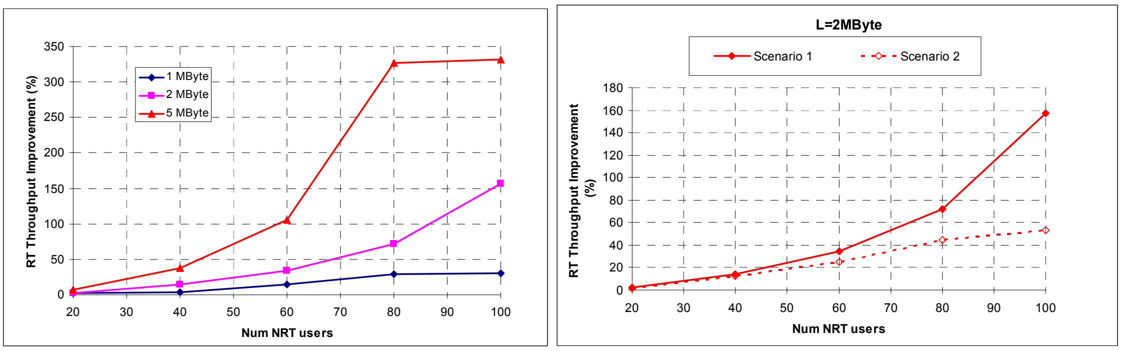
RAT2 is not in the expected trajectory. However, since decisions are reconsidered every ΔT=1s, the algorithm can react to changes in trajectory and this ensures that it can still provide a significant gain up to 50%.

Figure 5 plots the performance in scenario 2 for different values of the angle θ in which direction is changed and with L=2 MBytes. A similar trend as in the scenario 1 is observed, namely the proposed algorithm increases the delay with respect to the reference algorithm but it increases the RT throughput. However, note that the effect of θ is negligible and performance almost no varies. The reason is the reactivity of the algorithm to modify decisions when the trajectory has changed, so that, regardless of θ, what really finally matters is the number of users in the scenario that are approaching at a given time the coverage area of RAT2. This reveals the robustness of the algorithm in front of random variations in the trajectories.



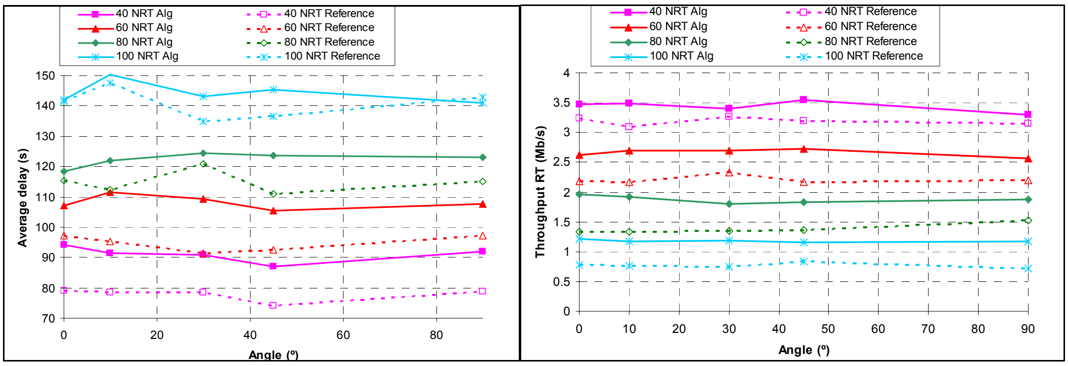
(a) (b)

Figure 3: (a) CDF of the total delay for NRT users and (b) total network throughput of RT users



(a) (b)

Figure 4: (a) Network throughput improvement with the proposed algorithm for different file sizes L in scenario 1, (b) Comparison between scenario 1 and scenario 2



(a) (b)

Figure 5: Performance in scenario 2 as a function of the angle θ in terms of (a) Average delay of NRT users and (b) Network throughput of RT users

5. Conclusions

This paper has proposed a novel solution for autonomous RAT selection in multi-service multi-access scenarios. It is based on the exploitation of the time dimension in the RAT selection process, so that, according to the current context (i.e. RATs and operators available) and its future evolution, the decision making can choose to postpone a transmission until reaching the coverage area of a more suitable RAT. This strategy is claimed to have applicability for NRT services without stringent deadline constraints (e.g. sending a number of emails, downloading MP3 files, etc.). The algorithm has been formulated in terms of the required inputs and the feasibility conditions to meet specific delay deadlines, while accounting for the RAT availability at the different positions.

The algorithm has been evaluated through simulations in different scenarios, with different levels of predictability in the mobile trajectories. It has been shown that the algorithm can significantly reduce the interference in some RATs, with the corresponding important capacity increase for RT users, while ensuring the NRT data is delivered within the desired delay bounds. Furthermore, it has also been shown that the algorithm is robust in front of errors in the predictability of the trajectories followed by the terminals.

Acknowledgements

This work was performed in project E^3 which has received research funding from the Community's Seventh Framework program. This paper reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein. The contributions of colleagues from E^3 consortium are hereby acknowledged. This work has also been supported by the Spanish Research Council and FEDER funds under COGNOS grant (ref. TEC2007-60985).

References

[1] Y. Kim, B. J. Jeong, J. Chung, C-S. Hwang, J. S. Ryu, K-H. Kim, Y. K. Kim, “Beyond 3G: vision, requirements, and enabling technologies,” IEEE Comms. Mag., vol. 41, n. 3, pp. 120–124, March 2003.

[2] S. Uskela, “Key concepts for evolution toward beyond 3G networks,” IEEE Wireless Communications, vol. 10, no. 1, pp. 43–48, Feb. 2003.

[3] A. Tölli, P. Hakalin, and H. Holma, “Performance evaluation of common radio resource management (CRRM),” in Proc. IEEE International Conference on Communications (ICC’02), vol. 5, New York, NY, USA, 2002, pp. 3429–33.

[4] E. Gustafsson and A. Jonsson, “Always best connected,” IEEE Wireless Communications, vol. 10, no. 1, pp. 49 – 55, Feb. 2003.

[5] A. Furuskar and J. Zander, “Multiservice allocation for multi-access wireless systems,” IEEE Transactions on Wireless Communications, vol. 4, no. 1, pp. 174–84, Jan. 2005.

[6] R. Chakravorty et al., “Performance issues with vertical handovers -experiences from GPRS cellular and WLAN hot-spots integration,” in Proc. Second IEEE Annual Conference on Pervasive Computing and Communications, Orlando, FL, USA, Mar. 2004, pp. 155 – 64.

[7] J. Pérez-Romero, O. Sallent, R. Agustí, J. Nasreddine, M. Muck “Radio Access Technology Selection enabled by IEEE P1900.4”, Procs. IST Summit, Budapest, June, 2007.

[8] O. Sallent, R. Agustí, J. Pérez-Romero, L. Giupponi “Decentralized Spectrum and Radio Resource Management Enabled by an On-demand Cognitive Pilot Channel”, Annals of Telecommunication, Vol. 63, No. 5-6, pp.281-294, June 2008.

[9] N. Halder, J. B. Song “Game Theoretical Analysis of Radio Resource Management in Wireless Networks: A Non-Cooperative Game Approach of Power Control”, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007.

[10] D.J. Goodman, J. Borras, N.B. Mandayam, R.D. Yates, “INFOSTATIONS: a new system model for data and messaging services”, Vehicular Technology Conference, 1997 IEEE 47th, Volume 2, 4-7 May 1997 pp. 969 - 973

[11] J. Irvine, D. Pesch, D. Robertson, D. Girma, “Efficient UMTS data service provision using Infostations”, 48th IEEE Vehicular Technology Conference, May, 1998. Vol. 3, pp. 2532 - 2536.

[12] J.C-P. Wang, H. El Gindy, J. Lipman, “On Cache Prefetching Strategies For Integrated Infostation-Cellular Network”, 31st IEEE Conference on Local Computer Networks, Nov. 2006, pp. 185 – 192.