Intelligence in Mobility Decisions

Nuno Coutinho, Susana Sargento, Vítor Jesus

***Abstract* - In the future, different technologies, such as Wi-Fi, UMTS, HSDPA and Wimax, will converge in a complementary manner forming a heterogeneous infra-structure. Moreover, the evolution of mobile terminals will allow them to connect simultaneously to several access networks and make use of multihoming capabilities. Thus, the concept of "always best connected" (ABC), consisting in using the devices and access technologies that best suit communications needs, users and networks, can now be implemented.**

In this paper we present a performance study of a context-aware and personalized network selection algorithm that enables the support of any criteria, quantitative and qualitative, including context and preferences, network and terminal characteristics, to determine the best access connection for each terminal and service. The performance results show the benefits of using such an algorithm in the network performance, and address the influence of specific criteria and constraints considered in the decision process.

I. INTRODUCTION

Over the last few years, various access technologies, such as WiFi, GPRS, UMTS, HSDPA and WiMAX, have been deployed and are available to mobile devices, which are increasingly equipped with more interfaces of different technologies (multihomed).

Due to these improvements, the next generation of mobile communications will be based on a heterogeneous infrastructure where the different technologies combine in a common platform to complement each other for different service requirements. This multiple technologies environment will also lead to high mobility scenarios increasing the expectations of the users and their Quality-of-Experience. Thus, each mobile terminal will be able to connect simultaneously to different technologies, which vary in bandwidth, delay, communication range, power consumption, security, reliability, end-user cost and several other aspects. Therefore, since the prime objective for the next generation of mobile systems is to integrate the different access technologies in a complementary manner, the concept of being always connected changes to the concept of *always best connected* (ABC) 1, enabling the choice of the best point of attachment to each user/services.

However, the definition of best may have different perspectives depending of the ABC actor 2. It depends on several aspects that may be subjective or objective. Characteristics like personal preferences, device capabilities, application requirements, network coverage and resources are strictly related with *best connectivity*. This paper focus on the access selection process, using any-constraint algorithm based on parameters related to context, preferences, and terminal and network characteristics, combining this knowledge to enable the optimization of both terminal and network point of view. [3].

In order to evaluate the efficiency of the implemented algorithm, different scenarios were simulated showing the benefits of the selection scheme. The simulations focus on the evaluation of the impact of one or more parameters in the decision process, in order to highlight the criteria flexibility, functionality and efficiency of the scheme.

The rest of the paper has the following structure: section II briefly discusses other proposals of network selection algorithms. Section III conceptually describes the algorithm implemented and its main features, and section IV presents the results of the network selection process. Finally, section V concludes the paper and introduces topics for further research.

II. RELATED WORK

In this section, we review related work, setting our requirements and briefly compare selected related work.

Some work has already been done which explores the described optimization problem. The most common approach is to center the selection process on radio signal considerations (e.g., 10). As discussed, despite its importance, we consider that it is only one between many criteria to be accounted for network selection. It is also common to consider link quality metrics (such as delay or bandwidth) – e.g., [11].

Song et al. [14] also use Grey analysis in a mix WLAN/UMTS environment, being the main metrics for decision the QoS desired of the user or application and the current conditions of each technology (WLAN or UMTS). As we said, we depart from this model of selection since, as discussed, we adopt a model of discrete services. Hence, we're closer to Gazis et al. [7] and Xing et al. [8] that model the problem of flow allocation as a knapsack problem: a user has applications to distribute across available PoAs. However, their work views mobility mainly as a resource problem, whereas we consider it to be only a part in a complete scheme for network selection.

Combined criteria not related only to signal strength or link quality have also emerged recently such as Iera et al. 5], McNair et al. [12] and Chen et al. [13] These authors propose schemes that are based on cost functions that contribute to an overall cost function which will, in the end, determine the best (under the selected criteria) PoA to handover to.

A. Key Requirements and Global Perspective of Related Work

We state now the requirements that have guided our design:

(i) strict service admission:
clear separation between runtime admission control and quality information.

(ii) easy plug-in of arbitrary criteria.
Any type of information should be possible to use, as long it is in a suitable format.

(iii) flow granularity.
A flow should be the basic element.

(iv) separation of powers.
Terminal and network should exchange information and not attempt unilateral decisions.

(v) fast environments.
Support to queues of events.

(vi) user optimizations.
For scalability reasons, a scheme should support local optimizations (single user) and global optimizations.

(vii) clear separation of entities based on self-contained properties.
We have identified the following three entities whose properties should not be dependent on each other's: network infrastructure (e.g., reliability of a PoA), the user/terminal (e.g., preference for low monetary cost) and resource constraints (available resources or not in a PoA).

The review of the literature shows that, to the best of our knowledge, no scheme copes with all requirements.

III. ANY-CONSTRAINT ALGORITHM DESCRIPTION

The network selection architecture implemented is based in the solution proposed in [3]. This section will briefly describe this scheme, presenting the main guidelines considered in the development of the solution, and the modeling of the several properties of each element in the network, providing an easier manipulation of the information. Finally, a description of the selection process will be given concerning triggers, the algorithm and the final handover decision made by the terminal.

A. Design Guidelines

The main objective of the network selection scheme proposed is to produce a ranked list of possible handovers that the terminal is allowed to perform after any event which triggers the selection process. The ranked list is composed by flow maps [9], each containing a possible distribution of the user's flows through the available access points. This schemes addresses the several requirements listed in A.

The events are the triggers of the architecture, caused by terminal requests, terminal movement, and any other possible change in the network that is relevant to the performance of its service provided. To support this, the scheme proposed must be able to deal with any type of trigger, being it classified as periodic, scheduled or based in context changes.

The ranked lists should be produced associated to values directly linked with QoE. However, due to the subjectivity of this metric and the difficult of associate it with the flow maps rank, it is necessary to model the main elements in the network according to their properties.

Regarding Points of Access (PoAs), there are two obvious properties: static priorities and resources available. Static priorities of a PoA could be reliability, monetary cost and mobility prediction. The resources of a PoA cannot be only related with bandwidth, but also with the capacity to provide different services to the user that wants to connect to it.

User properties can also be divided in static and real-time. The static properties of a user are related with all the context information that can be relevant in the handover process.

An important guideline, besides triggers and ranked lists, is that the resource management is totally independent of the ranked lists process. This means that only PoAs with resources available are allowed to enter in the flow maps calculation, making all feasible and reducing processing effort.

B. Entities Modeling

In order to be able to model any criteria to be used in the algorithm, we decided to format it in a matrix presentation form. This is a friendly and legible way of organize the different types of information of each entity.

We start by the following definitions:

- k is the index of a terminal belonging to the set of the K terminals able to perform a handover, $k \in K$ and $\#k = K$;
- M_0 is the set of all possible PoAs, $M_1^{(k)}$ the set of all detected PoAs by the k terminal;
- $M^{(k)}$ is the set of PoAs that are allowed to the k terminal, $\#M^{(k)}=M_k$;
- W is the number of the properties of a PoA that will enter in the ranking process;
- $F^{(k)}$ is the set of all running flows of terminal k, being $\# F^{(k)} = N_k$ the number of running flows;
- *Flow map* allows mapping each of the N_k flows of a terminal to one PoA out of the M_k possible, $FM^{(k)}: F^{(k)} \rightarrow M^{(k)}$.

In order to model the three basic and independent entities in the architecture scheme (PoAs, users and flow maps), a specific matrix was define for each. The PoA profiles cover all the properties and context information about each PoA specifically. User profile relies on user/terminal preferences and on non real-time activity of the user, being totally independent of the PoAs properties. Flow maps are related with user's flows and with the resources available, being a kind of bridge between the information of the PoA and the user personal preferences and status.

1. PoA Profiles

Regarding PoA profiles, they are defined in this form $AP = (AP_{ij})_{M \times W}$. This matrix keeps the PoAs properties and can be easily changed according to different criteria or preferences relevant in the mobility management decision. To keep the scheme architecture independent, we did not specify a method to set the mapping between numerical values and properties. However, a solution addressed by A. Iera et al. 5 is used. It is a simple analysis of each property setting an empirical numerical value to the criteria or being this value the result of a cost function.

The *AP* matrix is built based on all the specific properties of each PoA: taking this into account, its structure may be presented in three types of properties:

$$AP = \left(AP^{(user)} \middle| AP^{(static)} \middle| AP^{(real-time)}\right)_{M \times W}$$

The first substructure is set by information proceeding from the user, such as its preference for the PoA. The *static* part refers to the properties of the PoA that, first of all, are static and independent of context, users, or time. The third part is built regarding the information that comes from the network, like the current resource status of the PoA.

Access Technology	User Preferences (user)	Monetary Cost (static)	Handover Effort (static)	Reliability (static)	Bandwidth Allocation (real-time)
UMTS	80	50	75	90	50
WiMAX	100	30	75	80	50
WLAN	70	80	100	40	50

Table 1: Possible PoA properties.

An example of PoA properties and their empirical values is presented in Table 1 (these values are the ones of the AP matrix). In this matrix, the values closer to 100 are the best ones in that specific criterion. Bandwidth allocation can be a good example of a real-time property of a PoA, since it is dynamic and a result of a simple cost function, where the more occupied is a PoA, lower will be this value.

2. Flow Maps

The flow maps map the distribution of the different flows that belong to the same user through the available and allowed PoAs. Its mathematical model definition is:

$$FM^{(k,l)} = (FM_{ij})^{(k,l)}_{N \times M}, FM^{(k,l)}_{ij} \in \{0,1\}, \forall\, i,j$$

The l index defines a specific flow map for a given terminal k. Since we defined flow as the minimum indivisible unit of resources

$$\sum_{j=1}^{M} FM^{(k,l)}_{i,j} = 1, i = 1, \ldots, N.$$

3. User Profile

The user profile is based on properties and information independent of the context and real-time activity of the network. In order to have the proper interaction between the PoAs and the users, the user profile matrix must be modeled concerning the PoA properties:

$$UP^{(k)} = (UP_{ij})^{(k)}_{W \times W}, (UP_{ij})^{(k)} = 0 \ if \ i \neq j.$$

The UP is thus a diagonal matrix whose elements are weights that measure the importance given by the user *k* to the respective PoA criterion. It is now possible to shape qualitatively and quantitatively users using various combinations of different weights for each of the properties of the PoAs. Making a simple example by following this logic, different user profiles may exist such as *business man*, *gamer* and *groupie*. As is understandable and common sense, different users have different needs, and these requirements may be quantitatively weighted by the values in the UP matrix. Given this, a weight distribution like the present in Table 2 is adequate to model the type of users and their requirements.

User Profile	User Preferences	Monetary Cost	Handover Effort	Mobility Prediction	Bandwidth Allocation
Business man	0.5	1.0	1.5	1.5	1.0
Gamer	1.5	0.5	1.5	1.5	1.0
Groupie	0.5	1.5	0.5	0.5	0.5

Table 2: Possible weight distribution for different user profiles.

As is readily apparent all these properties and values are easily configured and modified according to the criteria followed by the architecture designer, as planned in the design guidelines and requirements for the network selection scheme.

C. Network Selection Scheme

This section will briefly present The network selection scheme. It consists in four main phases performed sequentially: trigger management and processing, classification and prioritization, calculation of the ranked list of flow maps and, finally, handover initiation.

1. Trigger Management

In the developed architecture, three types of triggers were taken into account: real-time events, scheduled and periodic. The real-time events can be caused by the network, or by the terminal. As one of the main ideas of the scheme is the separation of powers, an exchange of information and perspectives is needed in order to provide the current state of the connectivity.

The further types of events are related with expected changes of context either regular or occasional scheduled to perform a specific action, as a global re-arrangement for instance.

2. Classification and Prioritization

As many triggers can occur, and not all have the same urgency in being served by the network selection scheme, it is necessary to classify and prioritize them in order to attend through an orderly manner according to their importance or their type of user. Another advantage of classifying triggers is that it also supports user differentiation. This ability is very important to differentiate the service of premium users offering them their first choices in access selection and preferences.

3. Flow Maps Calculation

The architecture considers two different flow map calculation processes. The local optimization is the first and simplest referring to just a single user. The second, global optimization, concerns to a large number of users, being an iterative process for all users, similar to what is done in the local optimization. The necessity of two different optimizations arises due to the heavy processing of the global scheme. To describe this process, it will be followed the local optimization method, Figure 1, where the global optimization is only a generalization of the simplest routine.

This process is a simple algebraic manipulation of the matrices described so far. The matrix CAP contains a preference value by each PoA, being already an indication of the preferred access of the terminal. If it is not possible to find a flow map, a global optimization should be performed so that the architecture attempts to find a solution considering all terminals. The following “PoA Allocation” matrix (APA) determines, for each flow map, how much used is a certain PoA. Finally the “Weights of flow maps” matrix contains the rank of each flow maps, which after normalized, turns into the matrix Q which indicates the quality value of a flow map.

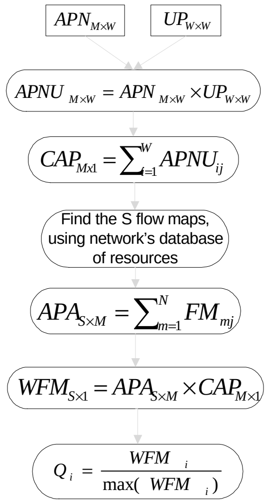

Figure 1: Local Optimization process.

4. Mobility Initiation

After the optimization algorithm, the best *p* flow maps must be sent to the terminal, so that it receives the set of flow maps ordered by rank. Also, in accordance with the separation of powers and the idea of having the most independent entities, the terminal is free to choose one of the flow maps according to its policies. However, the architecture was designed to deliver to the terminal a ranked list which already covers its preferences, network state and resources available. So the first flow map of the list is always the most appropriate for the terminal flows, unless there exists any other unknown reasons to the network selection scheme.

IV. SCENARIO AND RESULTS

This section presents a performance evaluation of the network selection scheme implemented through different scenarios. It also contains a study concerning the different parameters that can be configured in order to enhance the response of the global architecture.

A generic topology was developed for ns-2 2.31. This scenario is adjustable depending on the number of mobile terminals and PoAs, which are inputs of the topology file. The topology is based on a very simple wired-cum-wireless scenario. The topology dimension is calculated knowing the number of nodes in the network: distance of 700m between PoAs to avoid collisions and in order to emulate a multi-access technology scenario where the technologies do not interfere with each other. Links are defined to connect the fixed nodes, configured with 100Mb/s of bandwidth and a delay of 2ms. For each mobile node it is created a User Datagram Protocol agent (UDP) and a Constant Bit Rate traffic (CBR) generation agent, transport and application respectively. Regarding CBR traffic, it is defined a rate of 100kb/s for every terminal, and a packet size of 1000 bytes, with each terminal generating/requesting traffic.

.

A. *Load Balancing*

One of the real time properties of the PoAs is the resources availability (*Bandwidth Allocation*) at each moment in a specific access point. In the scheme implemented, this property is also considered, since it is expected that its utilization improves the performance of the architecture.

Introducing the maximum weight (1.5) for load balancing in the corresponding field of the user profile matrix, the global performance is the one shown in Figure 2. As the number of PoAs increase, better performance is achieved, since there is a wider range of possible accesses and more available resources (with no delay for 10 PoAs).

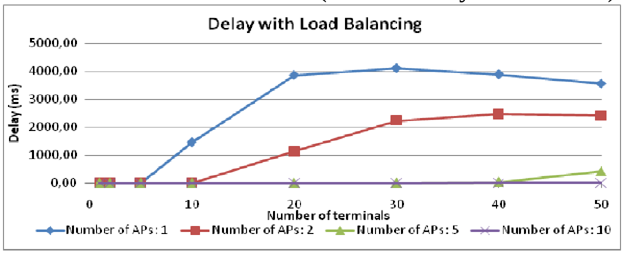

Figure 2: Mean delay of scenarios with load balancing.

B. *Resource Management*

The wireless channel in ns-2 2.31 is modeled to provide a maximum transfer rate of 1Mb/s, although in a real scenario this rate cannot be achieved without downgrading the quality of service provided. To evaluate this mechanism, we considered bandwidth thresholds for admission control ranging from 700kb/s to 1000kb/s, Figure 3 and Figure 4.

Figure 3 depicts the packets delay achieved for different bandwidth thresholds that a PoA may allocate (with 5 PoAs). As expected, there is a clear tradeoff between traffic in the network and the quality of service provided. In the curves corresponding to 900kb/s and 1000kb/s bandwidth threshold, the value of the delay gets significant and stabilizes due to admission control (Figure 4).

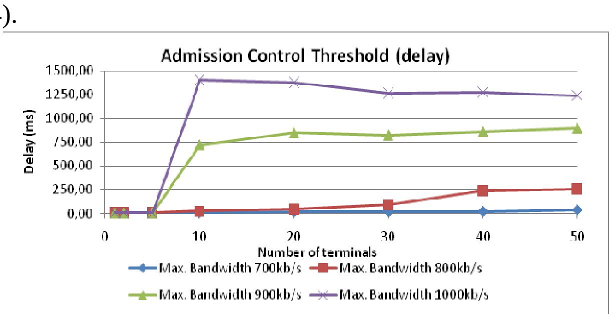

Figure 3: Admission Control Thresholds comparison for delay.

As shown, the optimization algorithm filters the PoAs totally occupied, forbidding the terminals to connect to them even if they are the preferred ones. The number of blocked flows starts to increase as soon as the resources are all occupied in all PoAs.

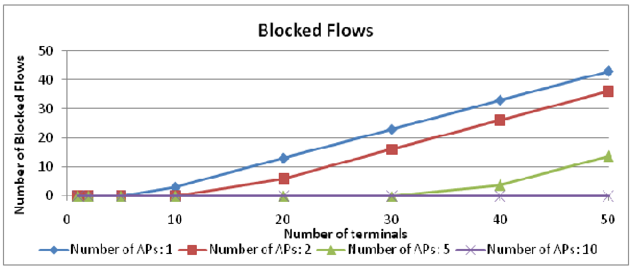

Figure 4: Blocked flows with admission control.

C. Triggers

As explained in section 0, the triggers are one main part of the optimization process, since they are the ones that initiate local or global optimizations. The decision on which optimizations should be performed may be configured through different criteria, besides the usual user requests that are considered a trigger.

To evaluate the effect of using triggers, simulations were performed based on the variation of the delay threshold used to trigger optimization. The scenarios tested were based on a threshold of maximum admissible bandwidth for admission control of 800kb/s, to be able to achieve significant delays and losses in order to trigger the optimization mechanism. These tests were performed using periodic QoS reports from the correspondent nodes at every second. From the results obtained in Figure 5 in scenarios with 10 PoAs, it is possible to observe the improvements obtained with the utilization of QoS triggers. For the maximum number of terminals in each scenario, the network is never saturated, existing always available candidates for each terminal. As expected, for all scenarios, as the value of the delay trigger threshold decreases, the overall delay of the network also decreases (and also losses, not shown here). As depicted in Figure 5, the differences between the curves are significant. One interesting result (not shown here due to space limitations) is the non-significant influence in the overhead of the optimization process.

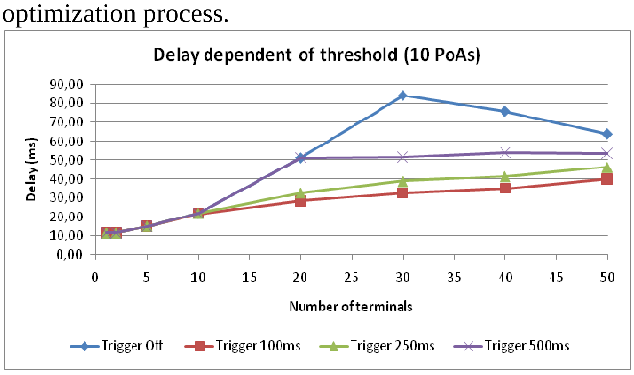

Figure 5: Delay dependent of trigger thresholds.

Directly related with triggers are the QoS reports from the correspondent node. These reports are sent periodically and are also responsible for the number of triggers in place during a simulation. We will now evaluate the impact of QoS reports rate in the network in scenarios with 10 PoAs available and with a delay threshold value of 250ms. As it is possible to conclude from the results in Figure 6, there is an impact of the QoS reports rate in the network. However, it is not as evident as in the threshold case. For instance, establishing reports at every 0.1sec is clearly better than configuring reports to each 5sec, as expected. However the difference between rates of 0.1sec, 0.25sec and 0.5sec is minimal. In this case, as opposed to the previous one, the difference in overhead is significant, because the reports can introduce large extra information in the network.

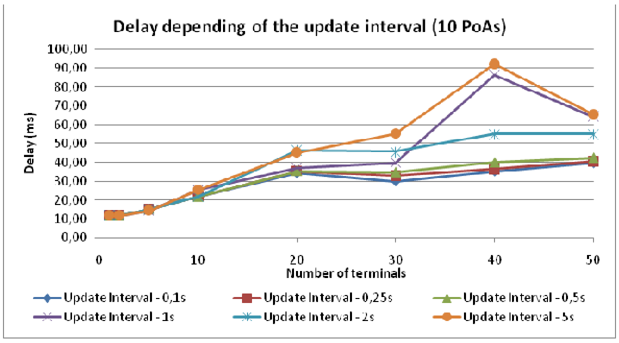

Figure 6: Delay depending of QoS reports rate.

D. User Preferences and Profile

In order to study the impact of the user preferences and profiles in the optimized decisions, a new metric was considered. It is defined as the ratio between the handovers performed to preferred PoAs and the total number of performed handovers.

The following results, depicted in Figure 7 and Figure 8, introduce a new parameter to the simulation which is the user profile, considering also different preferences of the terminal by each PoA. It is referred in the first figure that the results are the same for business and groupie profiles because the weight corresponding to the user preferences property given in their user profile matrices (UP) are the same. Both figures describe the impact of the load balancing weight in the preferred handovers ratio. For a null weight, the results are equal for both profiles, since the remaining parts of the APN matrix stay constant; the unique parameter that changes is the user preferences, irrespectively of the weight given in the UP matrix (0.5 or 1.5) of each profile, because it will immediately determine the ranked list in function of this parameter.

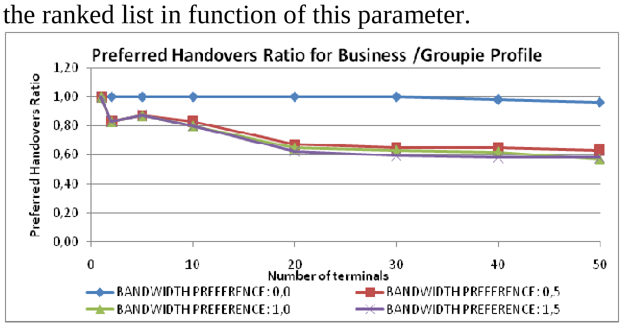

Figure 7: Preferred Handovers Ratio for Business/Groupie Profile.

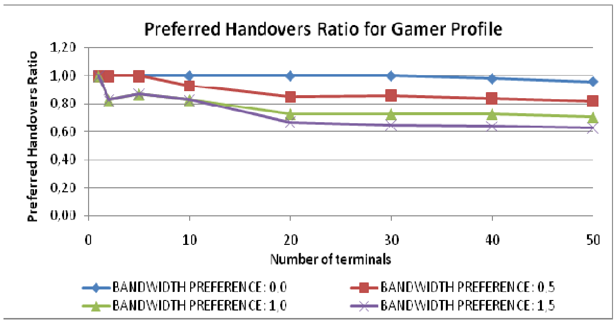

Figure 8: Preferred Handovers Ratio for Gamer Profile.

However, as the load balancing weight increases, the ratio decreases, but in the gamer profile case, this decline is not as pronounced as in the business or groupie profile. This situation occurs due to the difference of weights given in the UP matrices and that influence the final ranked list of flow maps. As was previously mentioned, our mechanism enables the presence of this type of criteria in a seamless way.

E. Global Optimization

Using an approach where the local optimizations just concern with resources available and terminal preferences, global optimizations can be used to re-organize the network. This process takes into account not only terminal's priority but also the state of current network resources through the load balancing feature. The results of this approach are present in Figure 9. The scenarios were evaluated using a maximum bandwidth allocation per PoA of 800kb/s. It was also considered that all terminals have the same preferences and profile in order to be more evident the impact of global optimizations. The "After Global Optimization" situation occurs after a unique global optimization is performed after all flows are distributed. The periodic optimizations are scheduled to be executed in intervals of 5 sec after the simulation starts.

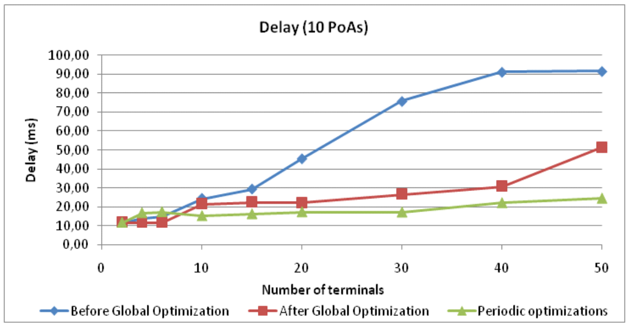

Figure 9: Impact of Global Optimizations in scenarios with 10 PoAs.

We observe in Figure 9 that the benefits of performing global optimizations with load balancing are higher as the number of terminals increase. The curve corresponding to the mean delay before the optimizations, as expected, has higher values as the number of terminals increase. However, for scenarios with a low number of terminals, it tends to be nearer the other curves. Comparing the difference in the results between a unique optimization and periodic optimizations, the difference is not very sharp, with better results for the curve of periodic optimizations. However, as a global optimization always involves many handovers and re-allocations of flows, this solution may not be always the better for the user. This depends on many factors, and one of them is the efficiency and seamlessness of the mobility mechanism in place.

V. CONCLUSIONS

This paper presented an implementation and evaluation of the network selection algorithm proposed by [3] under the future paradigm of ABC in next heterogeneous networks. The developed architecture allows the network to manage its devices connectivity using intelligent elements and decision algorithms. It is able to process a decision based on different types of criteria, context, resources availability, QoS state, user profile and preferences, through a matrices formalism and a sequential process of algebraic manipulations to provide a ranked list with the best maps of flow's distribution through the available and allowed access technologies.

Through the performed experiments, we can take the following important conclusions: the performance of the overall network is improved and it is able to provide the best selection both for the network and the users; the optimization scheme is indeed able to integrate any criteria, being a first step for the inclusion of context information and pervasive sensing environment characteristics, in future internet and networks.

As future work, we plan to integrate different types of networks in this mechanism, such as mesh and moving networks, different types of services, such as multicast, and improve the context and sensing criteria for the support of pervasive environments.

REFERENCES

1. E. Gustaffsson, A. Jonsson. Always Best Connected. IEEE Wireless Communications, Vol. 10, No. 1, pp. 49-55, Feb. 2003
2. M. O'Droma et al., "Always Best Connected Enabled 4G Wireless World", IST Mobile Communications Summit 2003
3. V. Jesus, S. Sargento, R. L. Aguiar, *Any-Constraint Personalized Network Selection*. In IEEE Intl Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, accepted for publication, September 2008.
4. Prehofer C et al. *A framework for context-aware handover decisions*. In IEEE Intl Symposium on Personal, Indoor and Mobile Radio Comm, Beijing, Sept03.
5. Iera et al., *An Access Network Selection Algorithm Dynamically Adapted to user Needs and Preferences*, Proceedings of IEEE Intl Symposium on Personal, Indoor and Mobile Radio Communications, Helsinki, Sept06.
6. A Furuskär, J Zander, *Multiservice Allocation for Multiaccess Wireless Systems*, IEEE Trans on Wireless Communications, vol 4, no. 1 Jan 2005.
7. V.Gazis, N.Alonistioti, and L.Merakos, *Toward a Generic "Always Best Connected" Capability in Integrated WLAN/UMTS Cellular Mobile Networks (and Beyond),* IEEE Wireless Comm, vol. 12, no 3 (Jun), pp. 20-28, 2005.
8. B Xing and N Venkatasubramanian,*Multi-Constraint Dynamic Access Selection in Always Best Connected Networks*, IEEE MobiQuitous'05, CA, July05.
9. Vítor Jesus, et al., *Mobility with QoS Support for Multi-Interface Terminals: Combined User and Network Approach*, IEEE Symposium on Computers and Communications, July'07, Aveiro, Portugal.
10. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J.P. Makela, R. Pichna and J. Vallstron, *Handoff in Hybrid Mobile Data Networks*, Personal Communications, vol. 7, pp. 34-47, April 2000.
11. E. Stevens-Navarro and V. W.S. Wong, *Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks*, in Proc. of IEEE Vehicular Technology Conference (VTC-Spring'06), Australia, May06.
12. J. McNair and F. Zhu, *Vertical handoffs in fourth-generation multinetwork environments*, IEEE Wireless Communications, vol. 11, no. 3, pp. 8–15, 2004.
13. W.-T. Chen and Y.-Y. Shu, *Active application oriented vertical handoff in next-generation wireless networks*, in Proc of IEEE Wireless Communications and Networking Conference, New Orleans, USA, Mar05.
14. Q Song, A Jamalipour, Network Selection in an Integrated Wireless LAN and UMTS Environment using Mathematical Modeling and Computing Techniques, IEEE Wireless Communication 12(3):42-48, June 2005.