



New QoS-based Decision Making Approach for Heterogeneous Networks

Ayman Khalil, Hussein Al Haj Hassan, Maryline H elard and Jean-Fran ois H elard

European University of Brittany (UEB)

INSA, 20 avenue des Buttes de Co esmes, 35708 Rennes, France

Institute of Electronics and Telecommunications of Rennes (IETR)

ayman.khalil@insa-rennes.fr

ABSTRACT

Next generation wireless networks will provide seamless high bandwidth connectivity with high quality of service (QoS) support to mobile users, where a mobile will be able to connect to several wireless access networks simultaneously. In this environment, heterogeneity of networks and heterogeneity of applications are presented as challenging problems, where an efficient architecture is needed to integrate different technologies. In addition, deciding the serving network is essential to avoid resource wastage and provide QoS for users. In this paper, we propose new integration architecture based on a virtual operator that rents or invests resources from a physical operator, and sells services to customers. The virtual operator will monitor the network's conditions and QoS parameters, and then provide the user requesting access with useful information to decide the serving network. In this context, we modify four well-known decision mechanisms (load balance, minimum distance, minimum price and minimum delay) by adding the QoS as a hard constraint. We also propose a new decision mechanism, where we consider new decision criteria that improves the overall system performance. Simulation results show that the new proposed mechanism outperforms the modified mechanisms, where the blocking probability is clearly reduced and a good level of QoS is provided to real-time applications or high-priority users.

Indexing terms/Keywords

Resource sharing; QoS; integration; virtual operator.

Academic Discipline And Sub-Disciplines

Engineering

SUBJECT CLASSIFICATION

Optimization

TYPE (METHOD/APPROACH)

Algorithms

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 10, No 10

editor@cirworld.com

www.cirworld.com, member.cirworld.com



INTRODUCTION

Heterogeneity has been introduced to be one of the most interesting features in mobile fourth generation [1,2]. In heterogeneous wireless networks, different Radio Access Technologies (RATs) are integrated together to complement each other in terms of coverage, bandwidth, services, and mobility support, thus providing benefits for both operator and users. On the other hand, heterogeneity is known also as a main characteristic of future applications, where each application has its own quality of service (QoS) constraints and requirements.

At present, there are two different ways to design integrated heterogeneous wireless access network architecture, known as tight coupling and loose coupling [3, 4]. In a tightly coupled architecture, networks are connected to a core network, usually 3G access network, in the same manner as other cellular radio access networks. The main advantage of this scheme is that, for mobility and QoS, the mechanism would be the same in all networks, which will lead to a relatively low handoff delay and less degradation of QoS. However, this scheme requires several modifications of access networks and protocols. Instead, networks are directly connected to the Internet for the loose coupling approach. In this scheme, networks (WMAN, 3G and WLAN, etc...) use different mechanisms and protocols to handle mobility and QoS. Although this kind of coupling will cause higher handoff delay, but it does not need significant modifications in the existing architecture or protocol.

In decision process, it is essential to decide where to put the decision unit. In centralized model, a central unit responsible of taking the decision has to collect different information concerning the networks' conditions. Then, when a new user wants to connect, the user sends a request to this central unit containing his application QoS requirements. Based on the latter information and the networks' information, the central unit decides the serving network and responds the user [5, 6, 7]. Instead, the new user terminal collects information concerning the networks' conditions, and then itself, i.e. the terminal, takes the decision in a distributed manner based on the collected information [8, 9]. This concept, where an intelligent terminal takes the decision of the serving network, is one of the most important issues proposed for 5G mobile phone concept.

Moreover, different approaches are adopted for the decision process in heterogeneous networks, such as individual, global and hybrid approaches. In the individual approach [9, 10], each user selfishly strives to improve its own performance such as delay, throughput, service time, etc. In the global approach [11, 12], the system manages the users association in order to enhance the global performance. Finally, the hybrid approach [13] combines the previous approaches, where the system manages the user association and benefits from user feedback or vice versa.

Effectively, there is a strong relation between the decision model, adopted approach and decision mechanism. For instance, the decision model will determine the way the system will exchange information, and thus the availability of information at each unit (user, network, central entity...). On the other hand, the adopted approach will determine the general aim of the decision mechanism and might introduce serious restrictions on its performance.

In literature, many decision mechanisms are introduced. In [8] the authors present an energy efficient radio access network (RAN) selection algorithm that selects a RAN with lower energy consumption to realize long terminal battery life while satisfying QoS. The selection mechanism also employs a penalty function that avoids discarded vertical handovers to reduce handover overhead and network loading. In [9] the authors consider a new network selection strategy based on the estimated Signal to Interference-plus-Noise Ratio (SINR) value in an integrated heterogeneous wireless network. This strategy allows the users to select the network that has the highest SINR value from all the available networks during its communication. This results in enhancing the physical user conditions and consequently his performance. In [10], the authors study the individual approach in an integrated WLAN and UMTS hybrid cell where users decide to join one of the two RATs so that their own cost is optimized. They consider the average service time of a mobile as the decision cost criteria and an incoming mobile connects to one of the two RATs depending on which of them offers the minimum average service time. In [5] the authors study the percentage of blocking of a user in a heterogeneous environment (WLAN, UMTS, and WiMax). They consider QoS demands for different types of applications, as they apply load balancing distribution and decision mechanisms based on user's preference. In [14] an algorithm for a context-aware network selection is proposed based on a modified weighted product method (WPM) for access network selection. The authors use weight distribution method based on sensitivity analysis of WPM for the most influential criteria on the state of user at a given time. In [11] the authors adopt the global approach. They try to decrease the complexity of the decision by proposing four heuristic algorithms based on distance, probability of distance, peak rate, and probability of peak rate. A comparison between the proposed algorithms and the optimal solution is presented. In [12] the authors study the global approach in comparison to the individual one, where the system load balances the downlink traffic of every user between two RATs (WiMAX and WiFi) in a way to improve the overall system performances. In [13] the authors develop a hybrid approach in a heterogeneous system composed of CDMA, GPRS, WLAN and satellite networks. It provides an overall performance of the whole system while taking into account user preferences. It is a two-level decision making scheme, where each user monitors and collects the dynamically varying network conditions for decision making and then a central controller finds the optimal user distribution for each individual network based on global observation and achieve load balancing for the whole system. In [15] a model to integrate a set of key performance indices (KPIs) of both the networks and the user into a single one is presented, by using a cost function based on weighted sum method (WSM) that includes these KPIs, providing for each network node a single evaluation parameter as output, and reflecting network conditions and common radio resource management strategies performance.

In this paper, we propose new integration architecture based on a virtual operator. For decision making, we adopt the hybrid approach where we aim to enhance the global performance of the system and satisfy the user QoS requirements. In addition, we aim to provide priority of real-time application over non-real time ones by proposing a new decision mechanism that differentiates between each type of application. Our study will focus on integrating 802.11 WLAN and CDMA cellular networks.

The rest of the paper is structured as follows: Section II presents the proposed integration architecture. In section III, we provide four modified well-known decision mechanisms and propose a new one. Section IV presents the system model and the simulation results. Finally, the paper is concluded in Section V.

PROPOSED INTEGRATION ARCHITECTURE

As mentioned before, there are two types of integration architecture. In tight coupling, there is low handoff delay and less degradation of QoS but a significant modification of access networks and protocols. On the other hand, loose coupling has the advantage of no significant modification in protocols or networks but, suffers from relatively high handoff delay. In our work, we will propose a new architecture, based on the use of virtual operator, where we combine the advantages of both architectures and minimize their disadvantages.

The idea of virtual operator is not new, where Virgin Mobile (found in United Kingdom, USA, etc...) is known as a successful example of virtualizing. In general, the virtual operator does not possess its own resources (frequencies for example), but invests or rents from existing ones. However, it must guarantee the QoS for its customers based on a service level agreement (SLA). Moreover, the relation between the virtual operator and the physical operator (network operator) is based on another SLA. This flexibility presents the virtual operator as a strong candidate for integrating different technologies in a common platform. In addition, the integration of technologies can be done step by step based on the user requirements and the amount of investment. On the other hand, the virtual operator can be a very attractive vision for many customers, since the users will deal with only one company but profit from a wide band of services that need several companies (operators) to provide them.

Figure 1 presents the simplified architecture of the new integration architecture. The main elements are the networks, terminals and virtual operator. In our context, we suppose that each network corresponds to a specific technology (e.g. WiMAX, WLANs, UMTS, LTE...). With respect to the multi-interface terminal, we suppose that the user equipment can simultaneously connect to different networks (technologies). The virtual operator does not manage or control the QoS of transmissions inside each network directly, but only monitors their QoS parameters and provides the user with this information to decide the most convenient network.

The virtual operator has also several functions such as supporting AAA (Authentication, authorization and accounting), locating users connections and positions, coordination between networks, profile manipulation and interfering in case of overload or crash.

In a continuous manner, the virtual operator monitors the networks load and QoS (delay, jitter, packet loss...). Then when a new user wants to connect, it will discover the available networks and contact the virtual operator via any interface for authentication and authorization. The virtual operator checks the user account and replies by sending the information of each network and the access key via its interface. Then, the user launches a decision mechanism for choosing the serving network. It should be noted, that the virtual operator periodically sends the condition of each network and whenever it is triggered by a new user. Finally, a handoff mechanism (vertical) is launched in several cases such as degradation of QoS, energy issues, presence of new networks, etc...

In this architecture, there is no significant modification of the networks architecture or protocol. However, the virtual operator will have a wide knowledge of the networks conditions, which permits a virtual cooperation between networks.

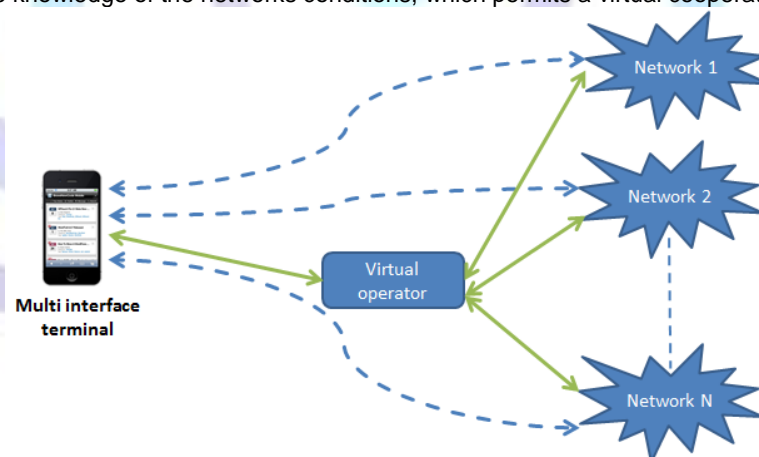


Fig 1: Simplified integration architecture

DECISION MECHANISMS

The decision process is one of the most important aspects of deploying an efficient ABC platform [5, 16]. We will adopt the distributed model, i.e. the user will choose the serving network based on certain parameters and considering QoS as a hard constraint. In the following, we will explain some modified well-known decision mechanisms. Afterwards, we will present the new proposed decision mechanism.

Modified Mechanisms

Figure 2 shows the flowchart of the load balance mechanism. The same flowchart describes the minimum distance, minimum delay and minimum price mechanisms with a small change in the operation marked with dotted rectangle. In

these mechanisms, the user gets the information of the available networks, then decides whether the network is a candidate or not, by checking its resources and QoS. Finally, the user decides the serving network based on certain criteria. In case of no candidate networks, the user will terminate the application. It must be noted that these decision mechanisms do not differentiate between different types of applications. In load balance mechanism, the network load is defined as the ratio of the number of occupied resources to the total number of the network resources (bandwidth for example)

$$load = \frac{used\ resources}{total\ resources} \tag{1}$$

The purpose of this mechanism is to provide fairness between networks by distributing the load among the networks. However, this mechanism does not take into account several parameters such as coverage of the network, number of resources of the network (capacity) and the user position and conditions with respect to the network. A network with more coverage area is expected to be loaded more than a network with smaller coverage area. In addition, a network with small number of resources would be loaded faster than a network with high number of resources. Moreover, the user position and conditions affect the consumed resources, and thus a user with bad conditions may consume resources similarly to several users with good conditions.

In minimum distance mechanism, the user chooses the network with minimum distance to the network access point. The purpose of this mechanism is to reduce the resources consumed by the user, and thus increase the number of accepted users. Although it is expected that the number of accepted users would increase, but this mechanism does not take into account the coverage, the number of resources of each network and the type of application.

In minimum delay mechanism, the number of users, type of users and the network itself affect the decision. The delay increases with the number of users. However, this increase depends on the type of the application and the network; for example, medium access delay is important in WiFi and negligible in UMTS. This mechanism leads to unfair distribution of users. It should be noted that the minimum price mechanism is implemented as a reference algorithm for comparing the price paid by the user using the new proposed mechanism.

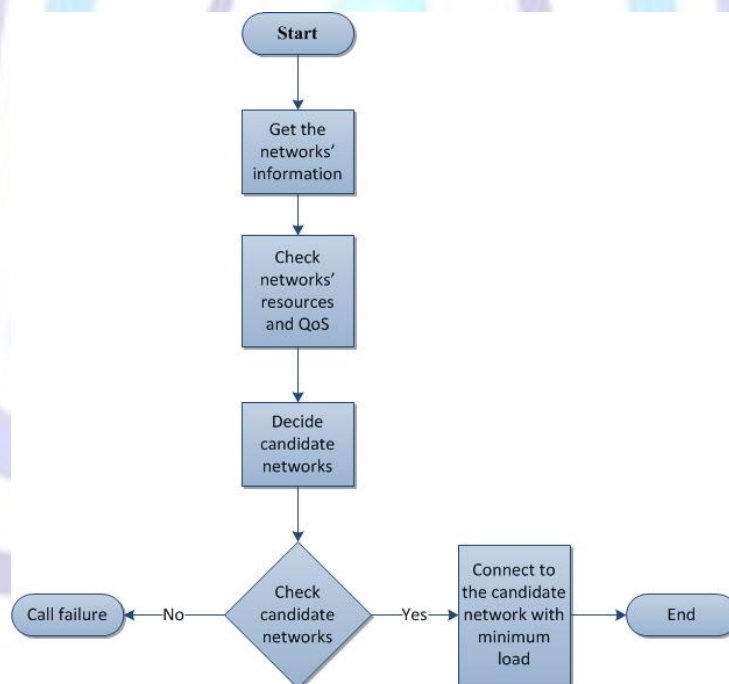


Fig 2: Load balance decision mechanism

Proposed Mechanism

The advantage of the previous mechanisms is their low complexity, which means that they are easy to implement and do not need high calculation complexity. However, each of them drops the effect of some parameters and conditions. Moreover, priority of real-time applications over non-real applications is not considered.

In our proposed mechanism, we adopt the hybrid approach where we aim to enhance both the global and individual performance at the same time. The fact that we consider the QoS as hard constraint means that the QoS users will be always satisfied, whenever they are served. On the other hand, we aim to decrease the percentage of blocked users, which is equivalent to enhancing the global performance of the system. Moreover, we aim to give priority to real-time applications over non-real time applications. The problem can be described as minimizing the blocking probability of users restricted by some constraints:

Capacity of network is limited, thus the sum of capacity of users served by network must be less than the capacity of the network.

The QoS (delay in our context) must be satisfied, thus the access delay of the networks must always be under the threshold.

To solve the aforementioned problem, we propose a new decision mechanism where we define different algorithms for each of the real and non-real time applications. Moreover, we define a new performance parameter based on the weighted sum and weighted product methods:

$$performance\ parameter = \sum_{i=1}^n w_i * F_i \tag{2}$$

where F_i represents the i^{th} factor and w_i is the weight of this factor. Every factor is composed of several related parameters and can be defined by:

$$F = \prod_{j=1}^m (P_j)^{w_j} \tag{3}$$

where P_j represents the j^{th} parameter and w_j represents the weight of this parameter. We note that the values of w_i and w_j may change with the type of application and that determining them is a subject of another study.

For real-time applications, the performance parameter would determine the serving network in case of existence of several candidate networks as illustrated in Figure 3. In case of lack of resources, the user triggers the virtual operator to initiate resource sharing algorithm. If resource sharing succeeds, then the virtual operator will contact the user to connect to the serving network. Otherwise, the virtual operator initiates the reallocation algorithm. The reallocated users will be contacted by the virtual operator in case of success. Otherwise, the user will be blocked.

Figure 4 shows the flowchart of the resource sharing algorithm. The virtual operator checks the networks that cover the new user and their QoS conditions. Then, it searches for existing candidates for resource sharing. One of these candidates is randomly chosen and is contacted by the virtual operator. The candidates of resource sharing must be non-real time users to avoid the performance degradation of real-time users. In addition, a user is considered as candidate if it did not undergo any sharing before, nor handed (or reallocated) before, and it can satisfy the resource required by the new user.

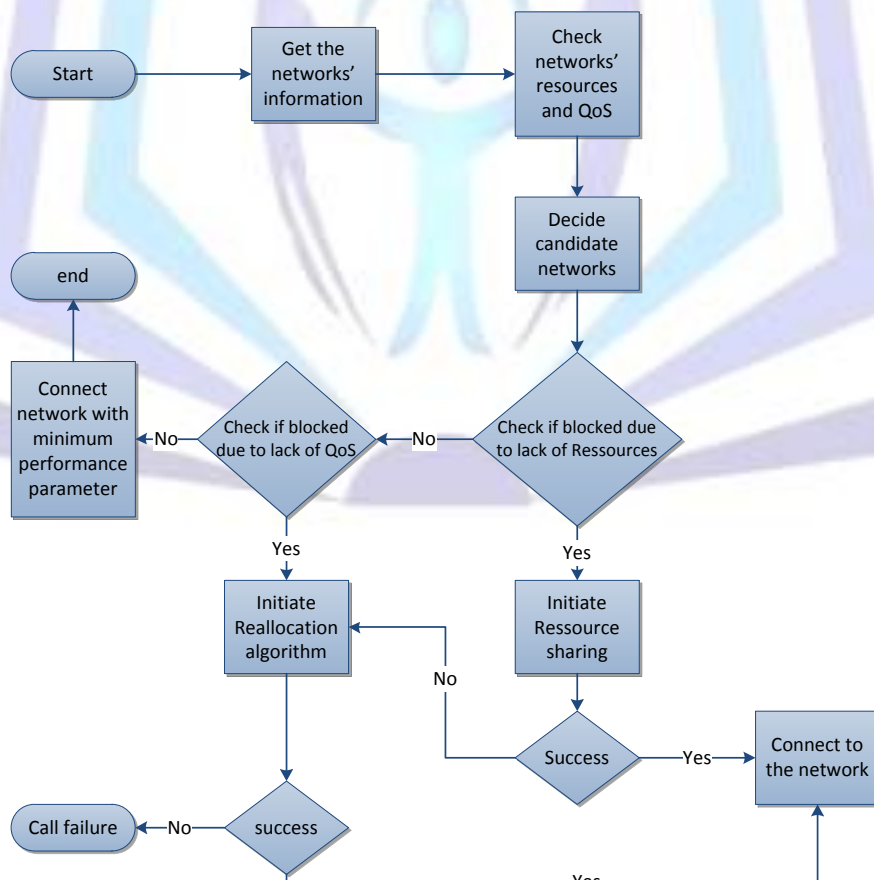


Fig 3: Decision mechanism for real-time applications

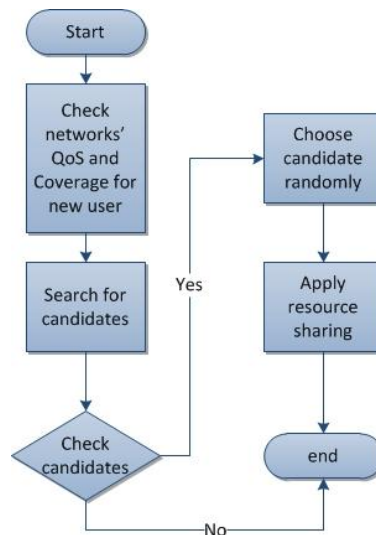


Fig 4: Resource sharing algorithm

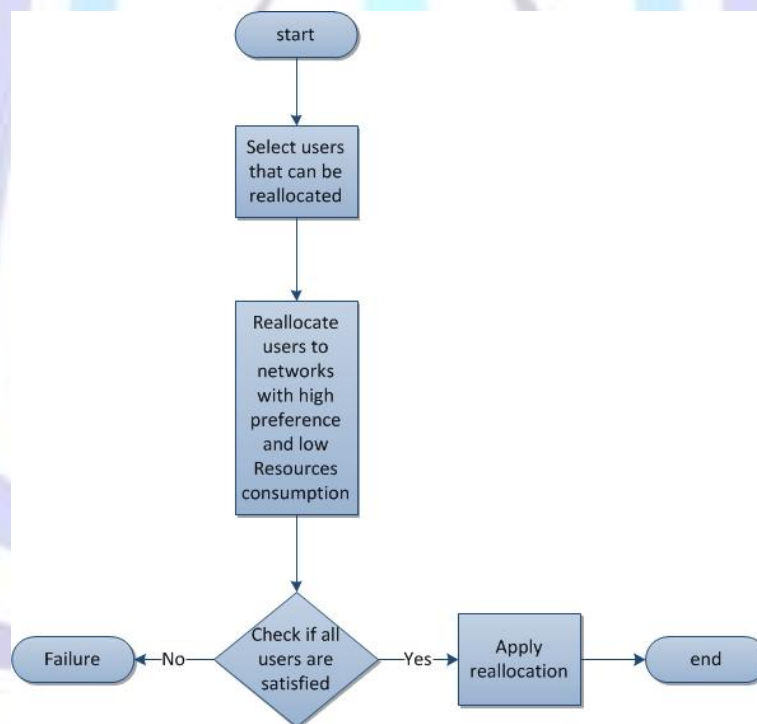


Fig 5: Reallocation algorithm

In the reallocation algorithm, illustrated in Figure 5, all non-shared and non-handed users are selected to enter the operation. The users are reallocated based on the preference of their applications and lowest resource consumption. The objective of such an algorithm is to satisfy the user demand, where some studies such as [5] consider the preference of the applications as the decision parameter, and decide to free the maximum number of resources by choosing the network where the user consumes the minimum number of resources. The virtual operator checks if the existing users and the new user are satisfied after reallocation. Otherwise, the reallocation algorithm is terminated and the user is blocked.

For non-real time applications, the user chooses to connect to the candidate network with minimum performance parameter as shown in Figure 6. The user receives the networks information, checks the networks QoS and resources, and then decides the candidate networks. For each candidate network, there is a blocking probability in function of resource consumption and network load.

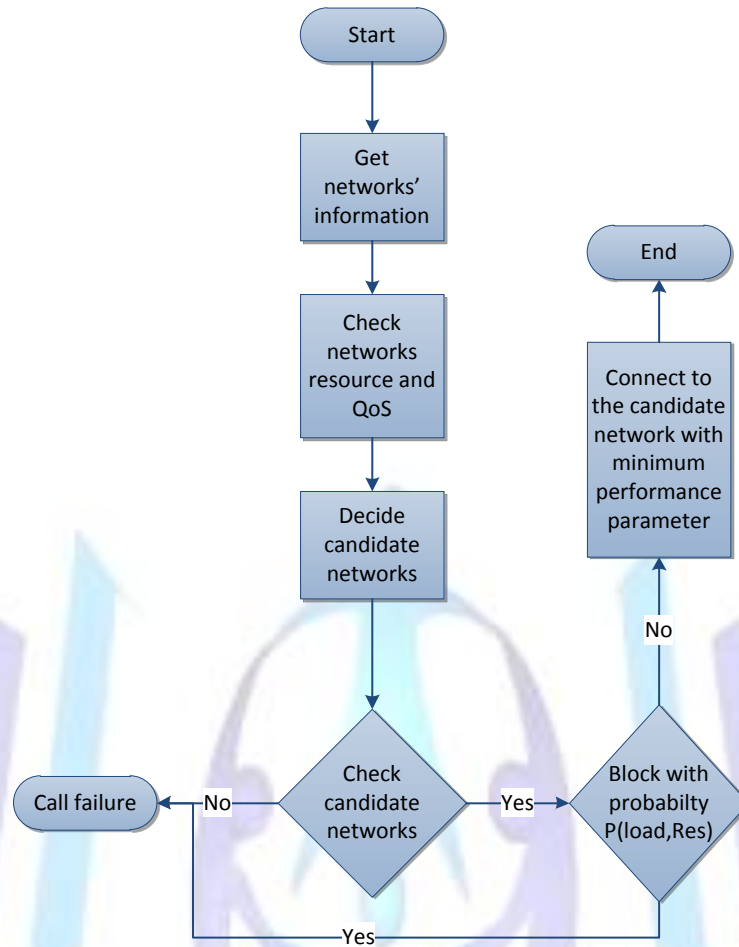


Fig 6: Decision mechanism for non-real time applications

SIMULATION RESULTS

System model

We consider a heterogeneous wireless access environment consisting of IEEE 802.11 WLAN and CDMA cellular network radio interfaces as shown in Figure 7. A mobile with multiple radio transceivers is able to connect to these radio access networks simultaneously.

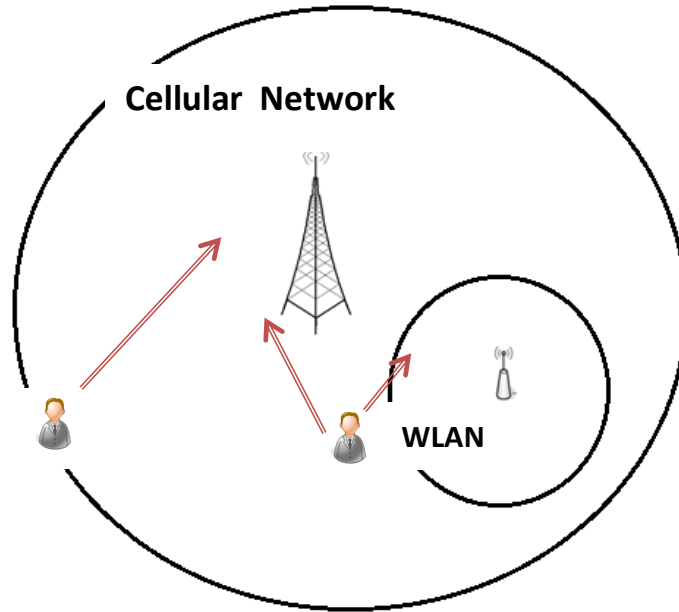


Fig 7: Service area of wireless networks

The geographical area is totally covered by cellular network and partially covered by WLAN (802.11 b). The price (arbitrary unit) of each network is presented in Table 1.

Table 1. Network price

Network	WLAN (WiFi)	Cellular network
Price	1	4

The simulations are done with the variation of the rate of arrival of the users per minute ($\lambda=3, 5, 8, 10, 15, 20$). This variation allows us to examine the decision mechanisms performance with different networks load. The user positions are randomly chosen and their type is divided into real-time (Voice) and non-real time (Data) applications. The voice codec used is G726 and the data is simply modeled as Poisson process.

Results

The blocking probability of voice calls (as real-time applications) in terms of user arrival rate is shown in Figure 8. The minimum delay mechanism has the worst performance between all mechanisms, while the minimum distance, minimum price and load balance have approximately the same performance. However, it is clear that the proposed mechanism has the best performance and enhances the possibility of accepting voice calls. For Data, as non-real time application, all mechanisms have almost the same performance with respect to the percentage of call blocking as shown in Figure 9.

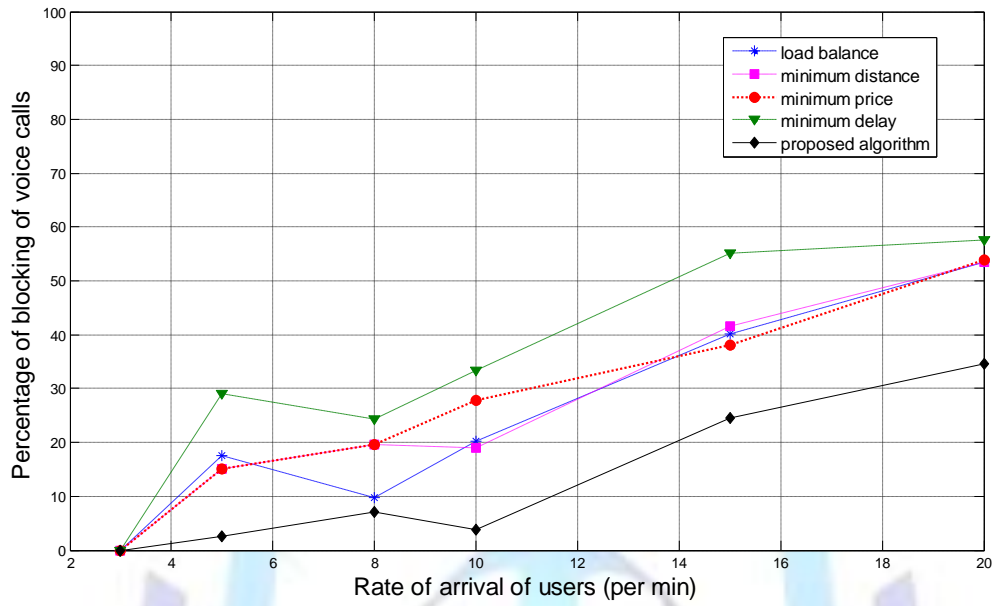


Fig 8: Percentage of blocking of voice calls in terms of rate of users arrival

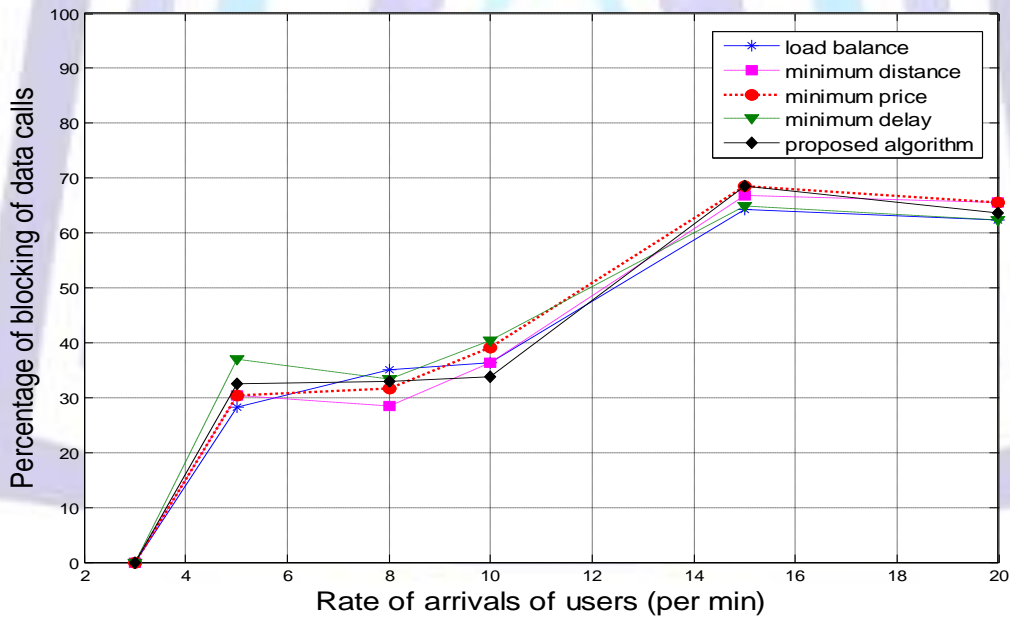


Fig 9: Percentage of blocking of data calls in terms of rate of users arrival

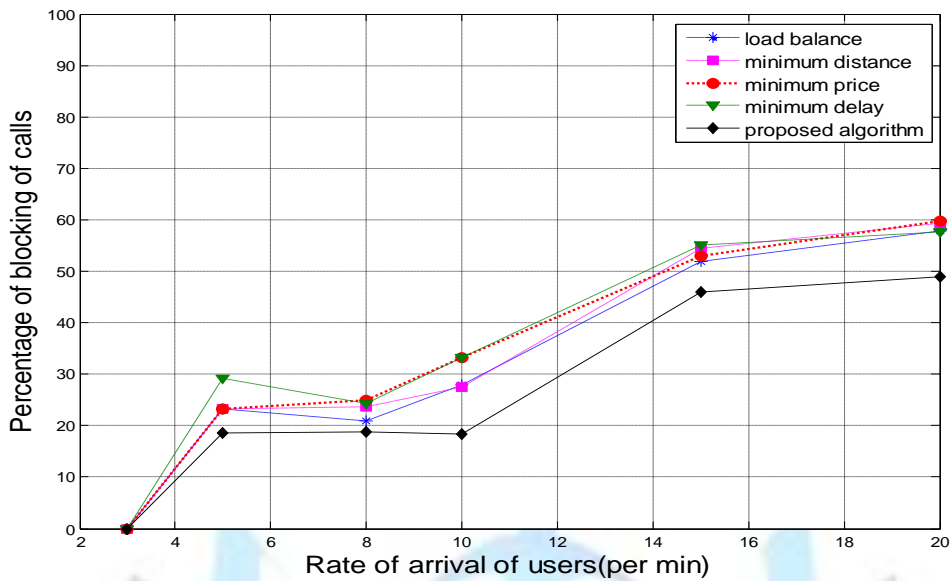


Fig 10: Percentage of blocking of calls in terms of users arrival

As an overall performance it is clear that the proposed mechanism outperforms the other mechanisms where the blocking probability of applications is significantly reduced in comparison with other mechanisms and for different user arrival rates as shown in Figure 10.

A small analysis of the blocking probability of calls shows that the proposed mechanism improves the overall performance of the system, where the system is able to serve more users. In addition, the proposed mechanism reduces the blocking probability of voice calls, as real-time applications, and preserves the same blocking probability for data calls, as non-real time applications. Moreover, it is obvious that the blocking probability of voice calls is less than that of data using the proposed mechanism. Thus, the proposed mechanism ensures priority to real-time applications over non-real time ones (voice users over data users). Note that the accepted users are fully satisfied, since QoS are considered as hard constraints.

Moreover, the average price paid by a user using the proposed mechanism is very close to that of a user using the minimum price mechanism as shown in Figure 11. However, the revenue of the operator using the proposed mechanism is the highest between all other mechanisms as shown in Figure 12. That is due to the fact that the proposed mechanism increases the number of accepted users, and thus increases the revenue of the operator.

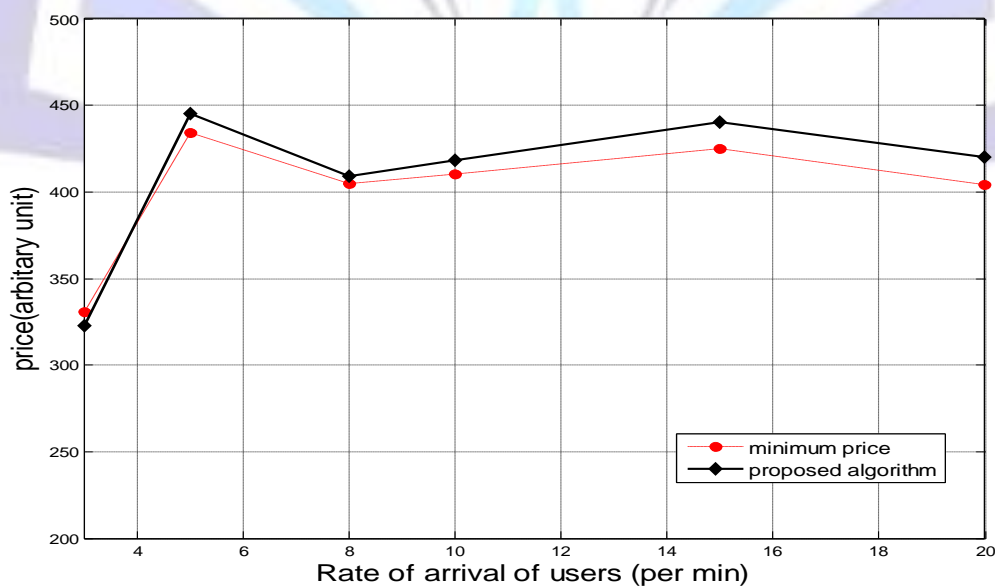


Fig 11: Average price paid by a user in terms of rate of users arrival

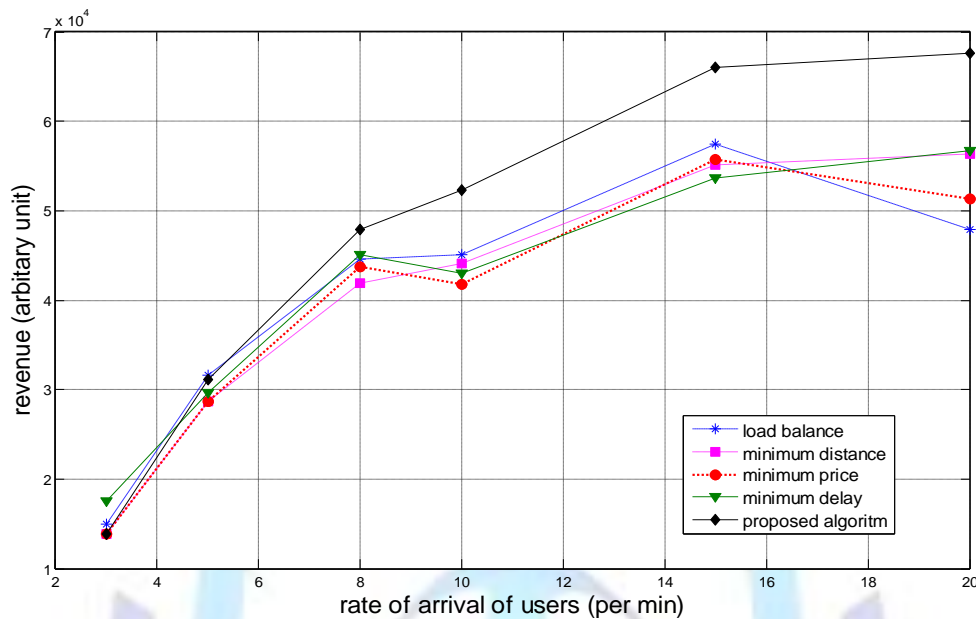


Fig 12: Revenue of the operator in term of rate of users arrival

CONCLUSION

The study of heterogeneous networks is a need for future communication systems, where heterogeneity is one of the challenging features of future networks and applications. In this paper, we have presented a new integration architecture based on the use of a virtual operator. In this architecture, the user takes the decision of the serving network. Meanwhile, a central unit would be involved in some cases in order to manage the allocation in efficient way. For the decision mechanisms, we have modified four existing decision mechanisms, where the QoS is added as a hard constraint in order to offer the best satisfaction to the end users. In addition, new QoS-based heterogeneous decision mechanism has been proposed. This mechanism has offered a good compromise in terms of user satisfaction and network performance. The performance of the proposed mechanism has been compared with that of the modified mechanisms where the simulation results have shown that the proposed mechanism outperforms the implemented mechanisms in terms of blocking probability, price and revenue of the operator.

REFERENCES

- [1] C. Luo et al., "Utility-based Multi-service Bandwidth Allocation in the 4G Heterogeneous Wireless Access Networks," *IEEE Wireless Communications and Networking Conference (WCNC'09)*, April 2009.
- [2] J. Perez-Romero et al., "Resource Management for Heterogeneous Wireless Access Networks," *IEEE International Conference on Wireless Information Technology and Systems (ICWITS'10)*, August 2010.
- [3] A. K. Salkintzis, "Interworking Techniques and Architectures for WLAN/3G Integration toward 4G Mobile Data Networks," *IEEE Wireless Communications*, vol.11, no.3, pp.50-61, June 2004.
- [4] M. Buddhikot et al., "Integration of 802.11 and Third-Generation Wireless Data Networks," *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies (INFOCOM'03)*, pp.503-512, April 2003.
- [5] J. Haydar, "Prise de Decision Orientee QoS dans les Reseaux Heterogenes ," *PhD thesis, Universite Pierre et Marie Curie*, 2010.
- [6] U. Toseef et al, "User Satisfaction Based Resource Allocation in Future Heterogeneous Wireless Networks," *Communication Networks and Services Research Conference (CNSR'11)*, pp.217-223, May 2011.
- [7] M. Kassar et al, "An overview of vertical handover decision strategies in heterogeneous wireless networks," *Computer Communications*, vol. 31, pp. 2607-2620, June 2008.
- [8] I. Kanno et al, "Adaptive Energy Centric Radio Access Selection for Vertical Handover in Heterogeneous Networks," *IEEE Wireless Communications and Networking Conference (WCNC'10)*, pp.1-6, April 2010.
- [9] A. Jabban et al., "SINR Based Network Selection Strategy in Integrated Heterogeneous Networks," *IEEE International Conference on Telecommunications (ICT'12)*, pp.1-6, April 2012.



- [10] D. Kumar et al., "User-Network Association in an 802.11 WLAN & 3G UMTS Hybrid Cell: Individual Optimality," *IEEE Sarnoff Symposium*, pp.1-6, April 2007.
- [11] F. Moety et al., "Distributed Heuristic Algorithms for RAT Selection in Wireless Heterogeneous Networks," *IEEE Wireless Communications and Networking Conference (WCNC'12)*, pp.2220-2224, April 2012.
- [12] K. Khawam et al., "Individual vs. Global Radio Resource Management in a Hybrid Broadband Network," *IEEE International Conference on Communications (ICC'11)*, pp.1-6, June 2011.
- [13] I. F. Akyildiz, "A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems," *IEEE Communications Magazine*, vol. 43, no .6, pp.S29-S36, June 2005.
- [14] P. TalebiFard et al., "A dynamic context-aware access network selection for handover in heterogeneous network environments," *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS'11)*, pp.385-390, April 2011.
- [15] A. Serrador, "A model for heterogeneous networks management and performance evaluation", *IEEE Network Operations and Management Symposium (NOMS'08)*, pp.690-693, April 2008.
- [16] E. Gustafsson, "Always Best Connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49-55, February 2003.



Ayman Khalil received the Ph.D. degree in Telecommunications from the National Institute of Applied Sciences (INSA), France, in 2010. During his PhD, he was with the Electronics and Telecommunications Institute of Rennes (IETR), where he worked on the cross-layer optimization of the high-rate wireless personal area networks. He worked on a European project (OMEGA) for three years in developing solutions and protocols for next generation home networks. Since 2010, he has worked in Lebanese universities and conducted research projects with IETR/INSA. His research interests lie in wireless communication systems, and particularly focus on topics related to all aspects of PHY, MAC and NET layers, cross-layer designs and heterogeneous networks.