The 7th International Telecommunications Symposium (ITS 2010)

A Reinforcement Learning Based Joint Call Admission Control for Heterogeneous Wireless Networks

Rennan J. M. Silva, João C. W. A. Costa
Faculty of Electronic Engineering
Federal University of Pará - UFPA
Belém, Brazil
{rennanmaia, jweyl} @ufpa.br

Gláucio H. S. Carvalho
Faculty of Statistics
Federal University of Pará
Belém, Brazil
ghsc@ufpa.br

Abstract—**Currently, there are many wireless networks based on different radio access technologies (RATs). Despite this, new kind of networks will be developed to complement those already existing today. As there will be no RAT able to give users full service requirements with universal coverage, the next generation wireless networks will integrate multiple technologies, working jointly on a heterogeneous way. Heterogeneous networks necessitate joint radio resource management (JRRM) mechanism to enhance better resource utilization and give users better quality of service. Joint call admission controls (JCAC) are a kind of JRRM mechanisms. In this paper, we present a JCAC approach to heterogeneous wireless network management based on reinforcement learning to treat call admission and technology selection, enhancing the network's performance. The effectiveness of this approach is assessed in terms of blocking rate results obtained by two simulation scenarios.**

Joint call admission control, JCAC, resource allocation, reinforcement learning, heterogeneous networks.

I. INTRODUCTION

Nowadays, there are many wireless networks based on different Radio Access Technologies (RATs). Despite this, new kind of networks will be developed to complement those already existing today. As there will be no RAT able to give users full service requirements with universal coverage, the next generation wireless networks (NGWN) will integrate multiple technologies, working jointly on a heterogeneous way [1].

A heterogeneous wireless network (HN), as shown in Figure 1, is composed of more than one RAT, such as UMTS, WLAN, WiMax, satellite links, etc., which coexist in the same area [2], where a call started by a mobile terminal (MT) can be attended by any of the available technologies. In this kind of network, each technology has its own characteristics, such as coverage, bandwidth, security level, cost of service, and level of quality of service (QoS) offered by the operator.

On wireless networks, the study of radio resource management (RRM) is a key issue because radio resources are often scarce and expensive, which make their efficient use a constant research area [3]. Moreover, the coexistence of different RATs necessitates joint radio resource management (JRRM) for enhanced QoS provisioning and efficient radio resource utilization [4].

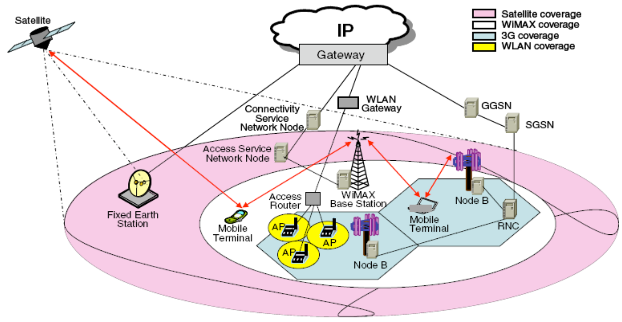

Figure 1. Example of heterogeneous wireless network [4]

When a call asks for resource on communications networks, it can be accepted or rejected depending on network conditions and of radio management policy. The mechanism that manages the accepting or blocking calls is known as CAC (Call Admission Control). The main purpose of CAC algorithms in wireless networks is the best use of available radio resources, by ensuring that the QoS requirements of all admitted calls are satisfied [4].

Several CAC algorithms have been developed to homogenous wireless networks [5] and a review of these works is showed at [6][7]. However, homogeneous CAC algorithms do not provide a simple and good solution for heterogeneous architecture, which are the kernel of next generation networks [8]. This limitation has driven the development of new specific algorithms to this kind of networks, called JCAC (Joint Call Admission Control) [4].

In JCAC algorithms, besides the task of accepts or not an incoming call, they have to decide which RAT is more suitable to accommodate this call.

Some algorithms are proposed to manage resources jointly on heterogeneous networks. Falowo et al. [4] presents an overview about JCAC algorithms and joint resource management highlighting the benefits of this approach. Wang et al. [2] propose an adaptive call admission control for integrated cellular and WLAN networks. In this approach, call admission decisions are based on requested QoS and availability of radio resources in the considered RATs.

A genetic algorithm based scheme call admission control for heterogeneous networks was proposed by Karabudak et al. [8] that aims to achieve maximum wireless network utilization, to guarantee mobile terminal's QoS requirements and significantly reduce handover latency. Lee et al. [9] presents a study that manages the vertical handover from a 3G to a WLAN network. The decision policy was probabilistically derived to avoid unnecessary downward vertical handover (from 3G to WLAN).

The contributions of this paper are an artificial intelligence based JCAC scheme, developed using reinforcement learning (RL) technique, to enhance connection-level QoS in a generic heterogeneous wireless network that support multiple service classes, and the evaluation the performance of this JCAC scheme using call blocking rate comparing it with a random selection JCAC scheme.

The rest of the paper is organized as follows: in section II are showed the fundamentals of reinforcement learning and the system model. Numeric results are shown and described in section III and, finally, in section IV are presented the final remarks.

II. METHODOLOGY

A. Introduction

To address the described objectives, we identified the best solution modeling the CAC problem as a Semi Markov Decision Process (SMDP), using the reinforcement learning technique, and the communication system as a discrete-time event system.

A traditional approach to SMDP is Dynamic Programming (DP) [10] that is guaranteed to give optimal solutions to MDPs and SMDPs. The case is that, obtaining the theoretical model (transition probabilities, transition rewards and transition times) is often a difficult and tedious process that involves complex mathematics in real life problems. DP needs the values of all these quantities. RL has the potential to solve a MDP without to having to construct the theoretical model [11].

In SMDP based on RL the learner and decision-maker is called agent. The entity that it interacts with, comprising everything outside the agent, is called environment. These elements interact continually: the agent selecting actions and the environment responding to those actions and presenting new situations to agent. The environment also gives rise to rewards, special numerical values that the agent should try to maximize over time. This interaction scheme is illustrated on Figure 2.

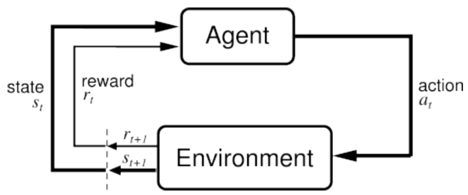

Figure 2. The agent-environment interaction in RL

Reinforcement learning is an artificial intelligence formalism that allows learning what to do – how to map situations to actions – so as to maximize a numerical reward signal [12]. There are three basic classes of methods to solve the RL problem [12]: Dynamic Programming, Monte Carlo and Temporal Difference (TD). We chose to solve our problem using a TD method because this method does not require explicit expression of the state transition probabilities and can handle SMDP problems with large state spaces efficiently. Also, it's incremental like DP and get convergence faster MC Method.

One of the algorithms based on TD method is the Q-Learning [11][12] algorithm. It has no control about which state-action pairs should be visited and updated, however, to ensure the method convergence to a near-optimal policy, it needs all state-action pairs are visited and updated continuously and, because this, it is known like off-policy method. It keeps a matrix, that maps and attributes a numeric value for the agent being in each state and performs a specific action on that state. From this value – called *Q(s,a)* – the agent finds what state-action pair is more suitable and so, what is the most appropriate action to be taken in the state *s*.

The *Q(s,a)* value is updated at each visit of the agent to a state-action pair, using the update rule (1), according with [11]. This approach was used because it was designed specifically to SMDP problems and guarantee faster converging to near-optimal policy on this kind of problems. This equation may change depending on the approach and the problem to be applied and, in this case, its elements can be described as follows: *r(s,a,s')* is the reward received for being in state *s*, performing the action *a* and, as result, go to the *s'* state; γ is a discount factor used in RL algorithms for measuring the influence of a future action on the current *Q(s,a)* value; *t(s,a,s')* is the transition time that the agent takes to go from state *s* to state *s'*; and finally *Q(s',a')* is the value by the agent to be in the next state *s'* and to perform the action *a'*.

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha\left[r(s,a,s') + e^{\gamma\ t(s,a,s')} \max_{a'\in A(s')} Q(s',a')\right] \quad (1)$$

$$\alpha = \frac{L}{V(s,a)} \quad (2)$$

The learning rate is represented by α and serves to indicate how a current reward value will influence on the future *Q(s,a)* value. One approach suggests that it may decline to allow faster convergence to an optimal policy [11]. This fallen is calculated using the number of visits to a state-action pair where a variable *L*, a value between 0 and 1, is initialized to make this calculation. In addition, a variable *V(s,a)* is incremented by one each time the state-action pair is visited. Thus, the learning rate is the result of the equation (2).

B. Assumptions and Modeling

In order to utilize RL algorithm on proposed problem, we need to identify system's states, actions and rewards. The agent of the system is the JCAC algorithm that performs the acceptance decision and can change system's states. Moreover, the system under consideration consists of a heterogeneous wireless network composed by *T* RATs, with finite capacity of B_t basic bandwidth unit (bbu), where the total HN capacity is the sum of each RAT bandwidth, like is shown in (3). These networks can carry *I* service classes representing multimedia service classes like VoIP or video streaming what characterizes constant bit rate.

$$B = \sum_{t=1}^{T} B_t \qquad (3)$$

The physical meaning of a unit of radio resources (such as time slots, code sequence, etc) is dependent on the specific technological implementation of the radio interface [13]. However, no matter which multiple access technology (FDMA, TDMA, or CDMA) is used, we could interpret system capacity in terms of effective or equivalent bandwidth [14][15]. Therefore, whenever we refer to the bandwidth of a call, we mean the number of bbu that is adequate for guaranteeing the desired QoS for this call.

For the sake of Markov modeling, incoming calls arrive in the system following Poisson processes mutually independents with parameters λ_1, λ_2... λ_I, where λ_I is the arrival rate of service class I. The service time of these classes are random variables exponentially distributed with parameters μ_1, μ_2,... μ_I, respectively.

$$Bo = \sum_{t=1}^{T} Bo_t \qquad (4)$$

$$Bo_t = \sum_{i=1}^{I} n_{t,i} * b_i \qquad (5)$$

If an incoming class I call is accepted, then it will receive a fixed amount of b_i bbu. Thus, we can model the total bandwidth used in the HN by equation (4), where Bo_t is the used bandwidth in technology t calculated by (5) and $n_{t,i}$ is the number of ongoing class i calls in the network t.

1) State Spaces

More formally, the proposed resource allocation is modeled as a SMDP, whose state is given by (6).

$$\Phi = \left\{ (m_{t,i}, e) : \sum_{i=0}^{I} n_{t,i} * b_i \le B_t \forall \;\; 1 < t \le T ; 0 < i \le I \right\} \qquad (6)$$

Where $m_{t,i}$ is a matrix containing the number of ongoing calls of all service classes in the HN, and e is the last event that happened, which identifies the arrival or departure of calls for each service class on the system.

2) Possible Actions

In Reinforcement Learning (Q-learning method), [12] the decision of what action to take is made by a stored value called *Q(s,a)*. This factor indicates the numeric value for the agent that is in state s to perform action a in a given time. Thus, the actions to be taken in this type of network are: 0 for reject the call; 1 for accept the call on the network 1; 2 for accept the call on the network 2; t for accept the call on the network t.

3) Reward Function

In our approach, it was generated two reward functions for each service class-network pair, one to acceptance and another to rejection, like in [18]. They are treated separately and generate a value based on used bandwidth on the network, rate of arrivals and time duration of all classes, and an assigned price to each service class, as follows in (7) and (8). The function of $r_{At,i}$ is the reward value received for accepting a call of class i on technology t, while the function $r_{Rt,i}$ represents the reward for rejecting a call by this same class on this network.

$$r_{At,i}\left(Bo_t, \vec{\lambda}, \vec{\mu}, \vec{\rho}\right) = f(Bo_t) - \Delta_i\left(\vec{\lambda}, \vec{\mu}, \vec{\rho}\right) \qquad (7)$$

$$r_{Rt,i}\left(Bo_t, \vec{\lambda}, \vec{\mu}, \vec{\rho}\right) = f(0) - r_{At,i}\left(Bo_t, \vec{\lambda}, \vec{\mu}, \vec{\rho}\right) \qquad (8)$$

Where *f(.)* is the contribution connected to the bandwidth, calculated accord to (9); Δ *i(.)* is the inversion contribute for the i^{th} service class and can be calculated as shown in (10); Bo_t is the used bandwidth by calls on the network t. Also, $\vec{\lambda} = (\lambda_1, \ldots, \lambda_I)$ is the vector containing the arrival rates of calls to all service classes, $\vec{\mu} = (\mu_1, \ldots, \mu_I)$ is the vector containing the holding time, being μ_I the average duration of a connection of the i^{th} service class; and $\vec{\rho} = (\rho_1, \ldots, \rho_I)$ is the vector containing the prices for each service class and may be based on the bandwidth occupied by the call, the time of channel usage, and so on.

In (9) a computation is made, ensuring that the agent accepts a call only if there is sufficient bandwidth available to afford it, otherwise it is rejected.

$$f(Bo_t) = \frac{1}{1 + e^{(Bo_t - B_t)/B0}} \qquad (9)$$

B_t is the total bandwidth supported by the network t and B0 is a tuning parameter. On Figure 3 is shown the pair of functions for building a generic class, showing how B0 influences in the curve of this function, which may be more "crushed" or not, but the 0.5 value is always reached for $Bo_t / B_t = 1$. We call this point the "inversion point", i.e. the point where $r_{At,i}$ = $r_{Rt,i}$: before there is a greater reward accepting calls, after rejecting them.

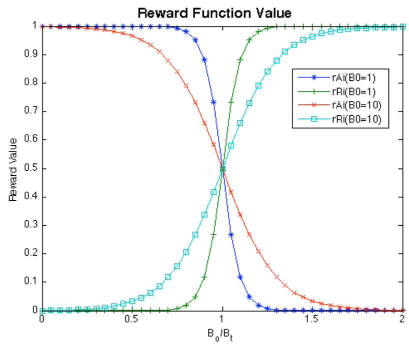

Figure 3. Reward Function value to a generic service class with (B0=1 and B0=10)

The second term in (7), the inversion term, comes from the need of the service provider to maximize its profits in the long term. The value of Δ *i(.)* in (10) should be as high as the service class i is not considered appropriate by the network operator, and as minor as deemed appropriate.

$$\Delta_i = c \cdot g_i(.) \qquad (10)$$

In this case, c is a free parameter that can be used to obtain an optimal tuning of the algorithm and $g_i(.)$ is called inversion function. It takes into account the desirability of accepting the other class of service requested in relation to class i as in (11).

$$g_i(\vec{\lambda}, \vec{\mu}, \vec{\rho}) = \frac{\sum_{k \neq i}^{N} \lambda_k \cdot \mu_k \cdot \rho_k}{\sum_{k=1}^{N} \lambda_k \cdot \mu_k \cdot \rho_k} \qquad (11)$$

Thus, according to the inversion, a certain service class *i* is more convenient if:

- The price is higher;
- Its frequency (λ_i) is greater;
- Its duration (μ_i) is greater;

III. RESULTS

To evaluate the effectiveness of the proposed JCAC, we simulate it like a heterogeneous network composed of two distinct technologies: Networks 1 and 2, with a bandwidth capacity of 88 bbu and 160 bbu, respectively; which accepts two service classes (classes 1 and 2). In the simulation were generated flows to new call requests, through discrete event, according to Poisson processes mutually independent, with an average attendance exponentially distributed.

The proposed algorithm was compared with a JCAC that accept calls while it has available bandwidth in heterogeneous network and make its technology selection randomly. The performance metric used to evaluate the algorithm was the blocking rate of calls to the two service classes, and it was calculated by the ratio of blocked calls in relation to the total number of calls that arrived into the system.

The simulation scenario was implemented using the Java programming language because it is a powerful language, relatively platform independent and permit to do interface our approach with other machine learning implementations like Weka [16]. However, to check results reliability, the random selection JCAC algorithm was implemented and compared with a model made at Arena Simulator [17]. The obtained results were identical. So, checked simulator reliability, we insert our decision policy, generated by RL algorithm, and we earn the results shown in this section.

Two simulation scenarios were created. In both scenarios, the arrivals rate of calls for class 2 (less intrusive) is fixed, but changes from one scenario to another, and class 1 arrivals rate of calls varies according to data from Table I in each of the mentioned scenarios.

TABLE I. EVALUATION SCENARIOS DESCRIPTION AND DETAILS

Parameters	Scenario A		Scenario B	
	Class 1	Class 2	Class 1	Class 2
Bandwidth	8 bbu	1 bbu	8 bbu	1 bbu
Average duration of a call (in seconds)	5400 sec	120 sec	5400 sec	120 sec
Price	8	1	8	1
Arrive rate (Calls per second)	From: 0,00027 to 0,04166	0,00278	From: 0,00027 to 0,04166	0,0278

The time of decision is always the arrival time of any type of call (class 1 or 2) and possible actions in these moments are:

- 0: Reject the call;
- 1: Accept the call on Network 1;
- 2: Accept the call on Network 2;

Based on cited data, it was performed the Q-learning method of reinforcement learning technique to obtain an optimal policy for call admission and RAT selection to creation of the Joint Call Admission Control.

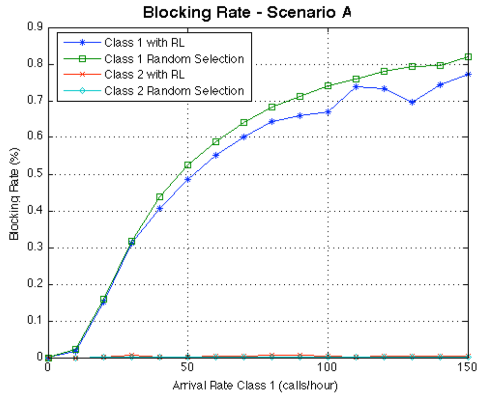

Figure 4. Blocking Rate of Calls on Scenario A

The results of Figures 4 and 5 show the performance in terms of blocking rate of calls for the RL-based JCAC and random selection-based JCAC. All the simulations were executed, at the very least, thirty times, admitting themselves a confidence interval of 95%.

In scenario A (Figure 4), where calls arrival rate of class 2 is smaller (0.00278 calls/second), we note that the blocking rate of service class 1 in JCAC algorithm based on RL performs better as its arrival rate increases, causing more network intrusion; what suggest that in a rather busy network both algorithms (RL-based JCAC and random selection-based JCAC) running well, since the blocking rate is very low.

When the focus of analysis is the service class 2, both algorithms have similar performance.

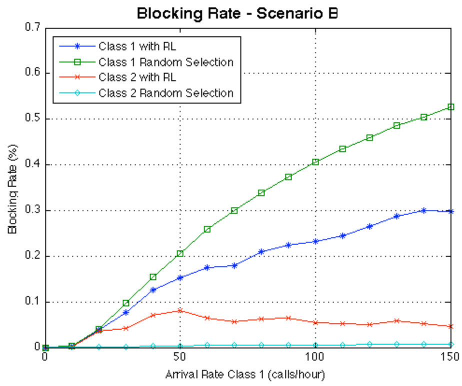

Figure 5. Blocking Rate of Calls on Scenario B

In scenario B, where class 2 calls arrival rate is one hundred times higher (0.0278 calls/second) than Scenario A, there is an increasing intrusion on the network and, thereby, the performance that the algorithms shows is differentiated. In this case, just as in scenario A, the difference in performance will appear as the arrival rate of calls of class 1 (Figure 5) is increased, however, this difference is more evident than in scenario A.

For class 2, the results show that the random selection JCAC algorithm remains always with better performance than RL based JCAC algorithm. This happens because the algorithm gives priority to traffic class 1, due its arrival rate and price. However, this treatment does not compromise significantly the QoS level for this service.

On Figure 6 is shown network utilization rate results to Scenario B. We can note the RL-based JCAC has always better performance compared to random selection-based JCAC. This happen because the JCAC reserves an amount of the network capacity to prioritized calls and, consequently, decreasing its blocking rate.

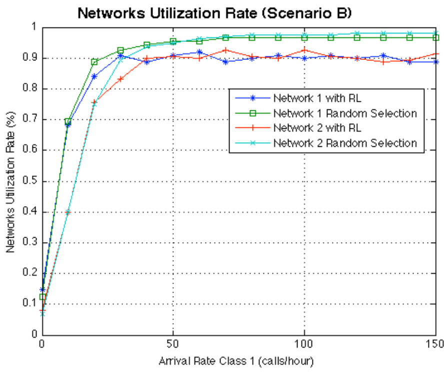

Figure 6. Networks Utilization Rates on Scenario B

These results clarify that, as the occupation of the network increases, the algorithm will improve its performance for the service class that is prioritized.

IV. CONCLUSIONS

Mechanisms of Call Admission are key elements in radio resources management of wireless networks. In next generation wireless networks, where the environment is heterogeneous and composed of more than one technology, it is necessary that this management be done jointly, in order to obtain better quality of service, providing more user satisfaction.

CAC algorithms for this kind of network have, besides the task of call accepting, the function of selecting which technology will receive an incoming call, those are the two main tasks of a JCAC.

This work presented an artificial intelligence-based JCAC model that uses the reinforcement learning technique to help on its decision. The results were compared with a JCAC algorithm that has no admission policy and selects randomly the technology. It was demonstrated the performance as satisfactory, in relation to random-selection technology JCAC, as the network gets busy, the algorithm shows a model of that reduce the blocking rate for classes of calls with higher priority.

Further studies need to be made to analyze the impact of this algorithm when handoff calls arrive at a specific network. In addition to parameters like bandwidth, arrival and service times, others may be added to the model, in order to make further performance analysis, and orientate the design of the network for specific purposes, such as greater user satisfaction in certain environments.

ACKNOWLEDGMENT

The authors of this paper would like to thank to CNPq that supports part of this work.

REFERENCES

[1] Per Magnusson, Johan Lundsjo, Joachim Sachs, Pontus Wallentin. (2004) “Radio Resource Management Distribution in a Beyond 3G Multi-Radio Access Architecture”. In: Global Communications Conference, Exhibition and Industry Forum (GLOBECOM), November 29 – December 3, Dallas, United States.

[2] Wang, X. G., Min, G., Mellor, J., Al-begain, K. (2004) “A QoS-based bandwidth management scheme in heterogeneous wireless networks” In: International Journal of Simulations Systems, Science and Technology, ISSN: 1473-8031, Volume 5, Pages 9 a 17.

[3] Sousa, V. A. de, Jr, R. A. de O. Neto, Chaves, F. de S., Cardoso, L. S., Pimentel, J. F., Cavalcanti, F. R. P. (2005) “Performance of Access Selection Strategies in Cooperative Wireless Networks using Genetic Algorithms” In: Proceeding of the 15th World Wireless Research Forum Meeting (WWRF’15), 8 e 9 de December, Paris, França.

[4] Falowo, E. O. e Chan, H. A. (2008) “Joint call admission control algorithms: Requirements, approaches, and design considerations”, In: Computer Communications, Volume 31, Pages: 1200-1217.

[5] Augusti, R. Sallent, O. Pérez-Romero, J. Giupponi, L. (2004) “A Fuzzy-Neural Based Approach for Joint Radio Resource Management in a Beyond 3G Framework”. In: First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE’04). Dallas, United States.

[6] Rao, R. M., Comaniciu, C., Lakshman, T. V., Poor, H. V. (2004) “Call Admission Control in Wireless Multimedia Networks”, In: IEEE Signal Processing Magazine, Volume 21, Pages: 51-58.

[7] Ahmed, M. H. (2005) “Call admission control in wireless networks: a comprehensive survey” In: IEEE Communications Surveys & Tutorials”, Volume 7, Pages: 49-68.

[8] Karabudak, D., Hung, C., Bing, B. (2004) “A call admission control scheme using genetic algorithms” In: Sysmpoium on Applied Computing, March 14 - 17, New York, USA.

[9] Lee, S., Kim, K., Hong, K., Griffith, D., Kim, Y. H., Golmie, N. “A Probabilistic Call Admission Control Algorithm for WLAN in Heterogeneous Wireless Environment”. In: IEEE Transactions on Wireless Communications, Vol. 8, Nro. 4, April 2009.

[10] D. P. Bertsekas, Dynamic Programming deterministic and stochastic models. Prentice-Hall, Inc, Englewood cliffs, 1987.

[11] Govasi, Abhijit. (2003). “Simulation-Baesd Optimization: Parametric Optimization Techniques and Reinforcement Learning”. Kluwer Academic Publishers.

[12] Sutton, R. S. and Barto, A. G., (1998) “Reinforcement Learning: An Introduction”. MIT Press, Cambridge, Massachusetts, 1998.

[13] V. Pla, J. M. Giménez-Guzmán, J. Martínez, and V. Casares-Giner, “Optimal admission control using handover prediction in mobile cellular networks,” Proceedings of Second International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks, West Yorkshire, UK, 2004

[14] G. Kesidis, j. Warlrand, and C. Chang, “Effective bandwidths for multiclass Markov fluides and other ATM sources,” IEEE Transactions on Networking, vol. 1, no. 4, Aug. 1993, pp. 424-428.

[15] N. Nasser and H. Hassanein, “Adaptive call admission control for multimedia wireless networks with QoS provisioning,” Proceedings of the 2004 International Workshop on Mobile and Wireless Networking, Montreal, Canada, August 2004, pp. 30-37.

[16] H. Witten, E. Frank “Data Mining: Practical machine learning tools with Java implementations”, Morgan Kaufmann. 2000. Software available in: http://www.cs.waikato.ac.nz/ml/weka

[17] Drevna, M. e Kasales, C. “Introduction to Arena”. In: Procedings of the 1994 Winter Simulation Conference, Ed. J.D. Tew, M.S. Manivannan, D.A. Sadowski, e A.F. Seila, páginas: 431-436. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

[18] Mignanti, S. Giorgio, A. D. Suraci, V. (2009). “A Model Based RL Admission Control Algorithm for Next Generation Networks”. In: Eighth International Conference on Networks. March 1 - 6, Cancun, Mexico.