

Connection Admission Control CAC and Differentiated Resources Allocation RA in a Low Earth Orbit LEO Satellite Constellation

Rima Abi Fadel^{1,2} and Samir Tohmé²

¹Ecole Supérieure d'Ingénieurs de Beyrouth
Mkalles, Mar Roukoz, BP 11514, Liban
{rima.abifadel@fi.usj.edu.lb}

²Ecole Nationale Supérieure des Télécommunications
46, Rue Barrault, 75013, Paris, France
{abifadel,tohme@inf.enst.fr}

Abstract. The Up/Down Link UDL of a Low Earth Orbit LEO satellite constellation is a scarce radio resource that needs to be shared efficiently between many users with different needs. A suitable Connection Admission Control CAC policy is required. In our study we assume that the network handles three types of calls: real time (voice) calls with strict constraints over the delay and the bandwidth, non real time (data) calls delay tolerant but with bandwidth guarantees requirements and Best Effort calls with no guarantees requirements. In order to ensure priorities are respected, we use an "enhanced trunk reservation policy" in order to derive the Resources Allocation RA. A differentiated RA scheme is proposed, associated with queuing for the lower priority calls. Different unity bandwidths are associated with calls depending on their requirements. The analytical markovian model is first derived, then the differentiated RA model is compared with two cases of bandwidth granularity choice. Impact of non markovian laws is studied using simulation.

Keywords: Access Networks, Performance Analysis, Satellite Communications.

1 Introduction

LEO Satellite Constellations are seen as a suitable mean for providing mobile users a global access service to terrestrial networks with the main advantage of a small propagation delay compared to the geographically stationary satellites. The general CAC problem within a multi-service LEO Satellite Constellation usually with Inter-Satellite Link ISL is very complex. It is common to split this problem into two components: the CAC associated with the Medium Access Layer MAC at the air interface level and the CAC associated with the establishment of the path within the LEO core network involving the ISL Routing problem.

In this paper we will consider only the CAC problem at the air interface level to the satellite constellation network. This case is found in two typical situations. The first corresponds to a constellation without Inter-Satellite Link ISL like SKYBRIDGE and the second is found in the case of integration of the satellite within the Universal

Mobile Telecommunications Service UMTS Satellite Radio Access Network USRAN [4]. A CAC policy based on an "enhanced trunk reservation technique" is used.

The classical trunk reservation technique and a number of its variants have been proposed and widely used in both contexts of routing and handoff. In the routing context, this technique was first applied to telephone networks with a non hierarchical circuit switched environment. By reserving a certain number of trunks for direct-routed traffic in a group, it has been seen as an effective way of stabilizing the network, and preventing performance degradation under overload [8]. In [10], it has been shown that a dynamic trunk reservation policy with a level depending on the traffic load yields better overload performance than a fixed trunk reservation scheme. Optimizing the network performance by a suitable choice of the trunk reservation parameters was studied in [12]. The approach was generalized from the classical telephony context to a multi-service network. In [9], blocking probabilities of multiple traffic streams due to trunk reservation in circuit-switched networks employing adaptive routing were studied. The technique was applied to elastic flows in [5], in a broadband network [6]. A dynamic control has also been shown to outperform the fixed one in a broadband call control admission context in [3]. Other variants also attribute a probability to the acceptance of call once the threshold is reached [13]. In the mobile networks context, the trunk reservation technique, also known as the guard channels technique, is also very popular. It has been used in classical voice cellular networks, as well as in ATM based mobile networks. It is seen as an appropriate way to prioritize calls experiencing handoff over newly generated calls in a cell [7]. Although the utility of using guard channels has been discussed, especially regarding the overall system utilization, and non-reserving schemes were proposed [15,16], the guard channels approach remains popular in the mobile context. In order to increase the system performance, the number of guard channels is varied dynamically using the information concerning ongoing calls in neighboring cells [17] with the mobility pattern [18,19,20]. In [14] the classical guard channels scheme is associated with queuing of originated calls in order to increase channel utilization. A fractional policy is proposed in [21]; it reserves a non integral part of guard channels for hand-off calls by rejecting new calls with some probability that depends on the current channel occupancy. A guard channel scheme is combined with queuing in an integrated voice/data wireless network in [22]. Priority is given to voice handoffs over data handoffs.

In [1], we propose an analytical solution for the "enhanced trunk reservation technique". It consists in admitting a lower priority service until a given threshold is reached, so far identical to the classical scheme, but it admits the blocked service back into the system only when the total number of occupied resources falls back to another threshold. This hysteresis introduction has two main advantages. First, oscillations of the system to and from the blocking state for the lower priority service are excluded. And since, in the satellites context, whenever a transition from a state to another occurs, the Network Control Center NCC, placed on the terrestrial part of the system is to be informed, this leads to less signaling traffic on the air interface.

The scheme proposed is applied to a network with an integration of three service classes. The highest priority class is the real-time service, carrying mainly voice over

the allocated channels. Data traffic is separated into two priority classes. The higher priority service has guaranteed bandwidth obtained by reservation that requires a CAC mechanism. The lowest priority data service called “Best Effort” service has no guarantees requirements.

In this paper, the trunk reservation technique is combined with a bandwidth reservation related to the type of call being served.

The paper is organized as follows. In section 2, we introduce the traffic models and parameters that have been used for both the analytical solution and the simulation. The control admission scheme is explained in section 3. In section 4 the analytical solution for the simplified differentiated RA model including queuing for the two data classes is presented. In section 5 this model is compared via simulation to two fixed choices of unity bandwidth allocation per call. Section 6 discusses blocking probabilities experienced by different services and delay experienced by data calls with guarantees when different queue lengths are considered for the differentiated RA scheme together with an enhanced trunk reservation mechanism. In section 7 non markovian service times are chosen and their impact on the above parameters is evaluated. Section 8 concludes the study.

2 Traffic Description

According to the French Réseau National de Recherche en Télécommunications RNRT Satellites Constellation Project [2], a mobile communication access service has been introduced for mobile users using a LEO satellite constellation. The total bandwidth serving pedestrian users on the uplink is of 72 kbits/s, with a granularity of 2.4 kbits/s. The cellular system that will be considered in this paper is based on cells fixed to the earth (similar to the Skybridge LEO Constellation case). Furthermore, the user mobility will be considered as negligible in comparison with the satellite mobility because the typical diameter of a satellite cell is of order of few hundred kilometers and because the only users considered are the pedestrian ones. Thus, we will not consider in our CAC model the issues associated with the Handover HO. Because the bandwidth on the downlink is usually higher than on the up link for pedestrian users, we will focus on the channel allocation on the up link. We assume also that any mobile user will always be able to see at least two satellites at the same time. Three call classes are considered. Voice calls are generated according to a Poisson process with an exponentially distributed call duration time of mean 120s. Channel allocation for voice is considered on the basis of the call level; the period of the ON time in an ON/OFF model being too small to justify the burst mode for voice. Data calls however can be served as bursts. Two types of non real time data calls will be considered. One corresponds to WEB applications, the other to FTP/SMTP data transfer. In all cases, we consider an allocation’s duration based on a request/response transaction scheme. In this paper, we consider different unity bandwidth’s allocation depending on the type of call being served. That is, a bandwidth of 2.4 kbits/s is to be allocated to a voice call entering the system while a 4.8 kbits/s bandwidth is to be allocated to each data or best effort call. This differentiated RA model will be compared to the all 2.4 kbits/s granularity as proposed by the french RNRT project [2] and to another proposition of

4.8 kbits/s granularity for all users. For the WEB applications, an exponentially distributed inter-arrival transactional time - usually known as thinking time - of mean 2 minutes or, equivalently, an arrival rate of 8.33×10^{-3} transaction arrivals/second for a given user is assumed. Assuming that a WEB page is represented by a load of 8000 bytes downloaded on a link of 2.4 kbits/s (respectively 4.8 kbits/s) as stated above, the average service time is thus equal to 26.67s (respectively 13.33s). The propagation delay is function of the satellite position covering the user and will be the same no matter what the bandwidth choice.

3 The Connection Admission Control Scheme

The lower priority data calls are accepted into the system until the number of allocated channels reaches a certain threshold s_1 . The (higher priority) real-time calls are accepted until all channels are occupied. The data service will be admitted back into the system only when the total number of occupied channels falls back to s_1 . Identically, Best Effort is admitted into the system until the number of occupied resources reaches s_2 , and is only be admitted back when the total number of occupied resources falls back to s_2 . Two separate queues are associated with respectively data calls with guarantees and best effort calls. Calls refused immediate service can still wait in queue until a certain number of channels becomes available.

The problem can be viewed as a generalization of a Birth-Death system with a number of channels equal to N , (30 or 15 channels depending on granularity), (λ_h, μ_h) , (λ_v, μ_v) and (λ_{be}, μ_{be}) the arrival rates and the inverse of the mean service times for respectively voice, data and best effort services.

Two main configurations for the respective positions of s_2 and s_3 are proposed.

The choice of $s_2 < s_3$, a prudent admission policy, will penalize the data services, but will be much more helpful for the high priority real time voice service. Once blocked, the admission back into the system of the data calls is delayed. Choosing $s_2 > s_3$ on the other hand, represents a rather tolerant policy for data calls. Of course transition from one configuration to the other can be done through threshold s_2 variation. Simulation will thus represent different performance measures variations as functions of varying s_2 .

4 Exact Analytical Solution for the Simplified Differentiated RA Model

As explained above, we consider that a voice call arriving to the system will be allocated a 2.4 kbits/s bandwidth while a data call, or a best effort call, is to be served with a 4.8 kbits/s bandwidth. This means that if an arriving lower priority call, when its admission into the system is allowed, finds less available bandwidth than its 4.8 kbits/s requirement, it will be queued if the queue is not full. The non differential bandwidth allocation scheme has been analyzed, via simulation, in [11].

In order to simplify the analytical model derivation, thresholds s_1 and s_2 on one hand and s_3 and s_4 on the other, are assumed to be identical.

The process consists in a five dimensional, continuous time, discrete states Markov process. The solution is given under stationary conditions (when time tends to infinity). Although the equations, given in the appendix, are written for the case for which data and best effort calls require a double bandwidth than the one needed for voice calls, the approach remains the same when other granularities are chosen.

Figure 1.a) shows the evolution of the blocking experienced by data calls and voice calls as threshold s_1 varies, for different values of the data queue length. Best effort queue length is fixed to a value of 5. The blocking experienced by best effort calls is not represented because in the configuration studied, these calls represent a proportion of 10% of the total system load while voice calls are responsible for 50% and data calls for the remaining 40%.

Two important observations can be made. As s_1 increases, the blocking experienced by data calls decreases while the blocking probability for voice increases. Of course, increasing s_1 means that less channels are reserved for the exclusive use of voice. On the other hand, for a given value of s_1 , as the data queue length increases, data calls experience less blocking. In fact, when a queue exists, a data call is not completely rejected once threshold s_1 is reached. It will be allowed to wait until a sufficient number of resources becomes available. The longer the queue is, the more calls are admitted to wait, and the less blocking situations will occur. A data call is rejected if, upon its arrival into the system, it cannot directly access the system and the queue is full. However, the longer the queue is, the longer a call will have to wait. The delay aspect will be considered in the simulation.

Figure 1.b) represents the impact of the choice of threshold s_3 over the blocking probabilities of data and best effort calls. Numerical values show a very light impact over the blocking of voice calls, the latter is thus not represented here. This is due to the light load associated with best effort. The data queue length is fixed to a value of 5, and the best effort queue length is varied. The same discussion concerning the impact of data queue length over data calls can be made to explain the impact of the best effort queue length over best effort calls. Figure 1.b) also shows a very important sensibility of best effort blocking probability over s_3 .

5 Granularity Impact over Blocking Probabilities and Advantages of the Differentiated RA Scheme

Simulation uses the QNAP simulation environment. The program was first validated for special cases for which the exact solution is known. The parameters used for the simulation are described in section 2 above. Voice calls, data calls and best effort calls are assumed to be respectively responsible for 50%, 40% and 10% of the traffic carried by each of the systems considered.

Both the tolerant and the prudent policy are studied by varying threshold s_2 . The transition from a configuration to the other is made for $s_2 = s_3$. The other thresholds are fixed to the values of $s_1=87%$, $s_3=53%$, $s_4=27%$ of the total number of available channels. The percentages are deduced from the necessity of having integer values for these thresholds. We first consider a system without queuing.

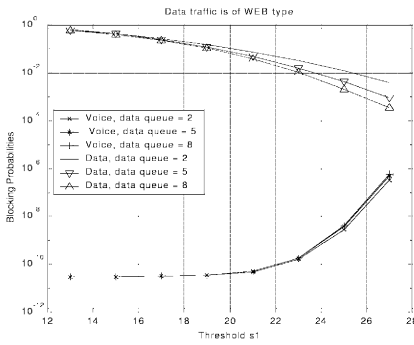


Fig. 1. a) Blocking probabilities for voice and data as functions of threshold s_1 .

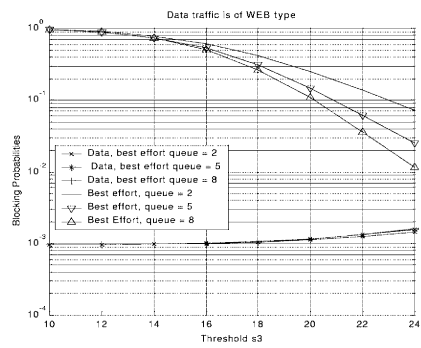


Fig. 1. b) Blocking probabilities for data and best effort as functions of threshold s_3 .

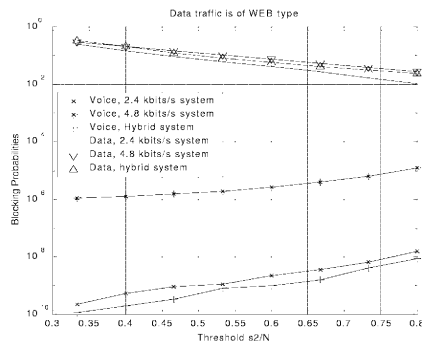


Fig. 2. Blocking probabilities for voice and data calls for two different granularity cases and for the differentiated RA scheme.

Figure 2 shows that data service blocking probability is very sensitive to the choice of s_2 , decreasing as s_2 increases, and thus resulting in an increase in the voice calls blocking probability. The situation corresponds to a case where data calls are, once blocked, admitted earlier back into the system. It can also be seen that for a prudent policy, voice calls experience considerable improvement in their blocking performance. This was the basic idea for which the prudent policy was thought of.

The best effort blocking performance is not represented. Numerical values however show a slight increase of the best effort blocking probability with s_2 . In fact, as s_2 increases, more data calls can be admitted into the system, both the two other service classes will be affected.

Figure 2 also shows that, when the two systems at 2.4 kbit/s and 4.8 kbit/s fixed granularities are equally loaded, the smaller granularity system, although carrying a higher traffic, presents better blocking performance regarding voice calls and data calls. This is due to the fact that the overall number of available circuits is bigger when granularity is smaller. The problem with the 4.8 kbit/s system is also that, although data calls are quicker served, and thus leave the system earlier, voice calls will occupy the unit resource for the same amount of time as in the 2.4 kbit/s system,

while occupying twice as much bandwidth. This situation naturally leads to more blocking.

The whole idea of introducing a differentiated RA scheme is to gain the advantages of both systems. Data calls, for the same amount of data to be transferred, once admitted into the system, will finish service and leave sooner, as in the 4.8 kbits/s granularity system, but without the drawback of wasting bandwidth over voice calls that are still served at a smaller unit bandwidth. We consider the same load factor for the differentiated RA scheme. Figure 2 shows that regarding voice calls, the blocking performance is even better than what was observed for the 2.4 kbits/s case. This is explained by the fact that data calls will leave the system sooner than for the 2.4 kbits/s case, leaving behind twice as much idle resources as in the 2.4 kbits/s system.

Regarding data calls however, the 2.4 kbits/s system still shows the best blocking situation. One must not forget that for a given value of the blocking threshold of data calls in the differentiated RA scheme, the total number of occupied resources must be at least two unities smaller than s_1 for the data call to be accepted. Otherwise, the available bandwidth is not sufficient to serve data. Admission of data calls is thus subject to more constraints in the differentiated RA scheme than in the 2.4 kbits/s system. The all 4.8 kbits/s system shows the most important blocking probability for the same relative value of the threshold s_2 , due mainly to the smaller number of available circuits for this system.

6 Impact of the Data Queue Length on Different Performance Measures

In this section, only the differentiated RA scheme is considered. Blocking probabilities of different traffic classes are evaluated as functions of the blocking out threshold of data calls s_2 , for different values of the data queue length. The best effort queue length is fixed to a value of 5. The simulation is run for the fixed values of $s_1 = 26$, $s_3 = 16$ and $s_4 = 8$, under the same relative contributions of the different services to the total traffic as in the previous section. Simulation results are given in figure 3.

Figure 3.a) represents the evolution of voice and data blocking probabilities with threshold s_2 and the data queue length. For a tolerant policy and a given $s_2 > s_3 = 16$, increasing the data queue length will be clearly advantageous for the blocking probability of data calls. In fact, when a data call does not have the possibility to be immediately served, under the queuing hypothesis, it will wait in queue until the admission conditions into the system are fulfilled. The call is rejected when the queue is saturated, a less frequent situation when the queue length increases. Of course, this means that voice calls will be penalized.

For a prudent policy, the advantage of queuing is less perceptible. The three curves representing the blocking experienced by data calls on one hand, and the three ones representing the blocking of voice on the other tend to converge for a very strict admission policy. When s_2 is very small, blocking out data calls will be considerably delayed. No matter what the queue length, it will very quickly be saturated and arriving calls will quickly start to be rejected anyway. The advantage of having more room

for queuing is not considerable when severe blocking constraints are applied to data calls.

Figure 3.b) shows the waiting time in queue of data calls evolution as a function of the threshold s_2 and for different values of the queue length. It can be seen that for a given s_2 , the bigger the queue length, the higher the delay. This is even more visible for prudent policies with small values of s_2 . This is the drawback of having less blocking situations for the data calls. Another observation is the very important delay decrease when s_2 increases. This is due to the fact that, as s_2 increases, data calls are blocked over a smaller interval of the occupied resources. This means that, after s_1 is reached, threshold s_2 will be reached much faster when it increases, and so it will be possible to serve the clients in queue that in turn, experience less waiting time.

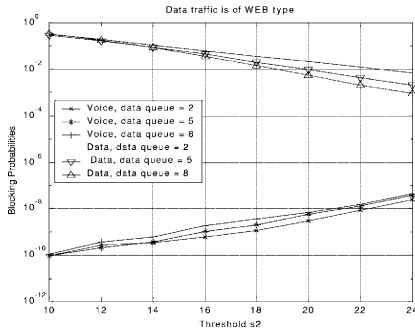


Fig. 3. a) Blocking probabilities of voice and data calls.

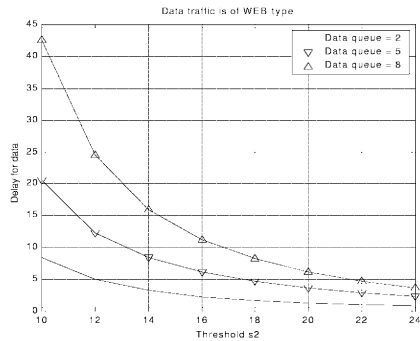


Fig. 3. b) Delay for data calls in seconds.

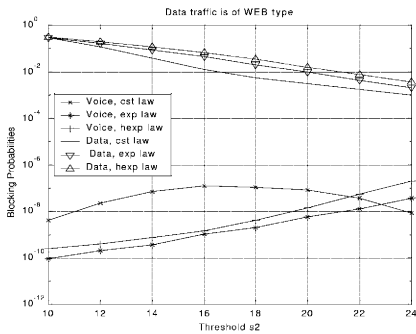


Fig. 4. a) Blocking probabilities of voice and data calls.

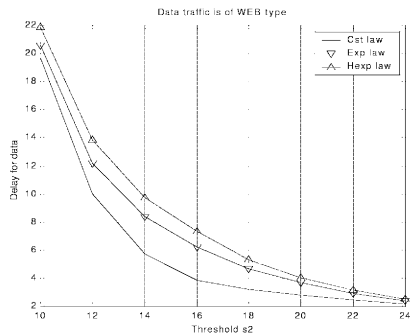


Fig. 4. b) Delay for data calls in seconds.

An important conclusion is that if a strict control policy is chosen for data admission control, increasing the queue length will be more costly, will induce a very high waiting time, without much improvement in the blocking performance of data.

Due to the very strict admission control associated with best effort, its blocking probability is high and almost insensitive at all to the choice of s_2 . The waiting time in the queue follows the severe blocking conditions with a high value that can attain an order of few minutes.

7 Impact of Different Service Time Distributions for the Data Classes on Different Performance Measures

In this section, the differentiated RA scheme blocking performance as well as the delay of data calls are evaluated for different distributions of the service time of the two lower priority data services. The service time for voice calls remains however exponentially distributed.

Two distributions are considered together with the exponential hypothesis of the previous sections. All distributions are assumed to have the same mean. We consider constant, exponential and hyper-exponential distributions respectively characterized with a variation coefficient of 0,1 and a chosen value of 2.

These laws are analyzed in order to see if the conclusions of the previous section still hold when more general, and thus realistic hypothesis are considered. One conclusion to be discussed is the impact of the choice of s_2 over the blocking probabilities of voice and data calls and over the waiting time for data. The previous section has shown that a good choice of s_2 should be done in the perspective of a compromise between the blocking probabilities of voice and data and the delay of data. The latter decreases when s_2 increases. Simulation results are given in figure 4.

The conclusion made for the delay still hold for more general laws. Increasing s_2 appears to be advantageous for the three cases of the variation coefficient being considered. This can be seen in figure 4.b). Another observation is the natural increase of the delay, for a given s_2 , with the variation coefficient. The best delay is observed for the constant law, with a null variance.

Regarding the blocking performance for voice and data calls, we can see that the conclusion concerning the choice of s_2 , still holds when the two lower priority services are served with a hyper-exponential distribution for the service time. When this service time is constant however, figure 4.a) shows that the choice of a very tolerant policy, $s_2 \approx s_1$, slightly penalizes voice calls compared to a very strict control policy for data calls, while considerably improving the blocking of data calls. An appropriate decision in this case would be to choose a very tolerant policy for data.

8 Conclusion

In this paper an enhanced trunk reservation mechanism associated with a differentiated RA scheme has been studied.

Simulation has shown that, compared to systems with fixed bandwidth allocation of 2.4 kbits/s and 4.8 kbits/s respectively for each call, no matter what the class of call, the differentiated scheme offers the best blocking performance for the high priority voice calls. Another advantage is that the faster transmission of data, provided by the 4.8 kbits/s system, is still possible with the differentiated RA scheme.

Queuing has proven to be more advantageous regarding the blocking of data calls. Waiting time certainly increases with the buffer size. An increase that is no longer justified when data is subject to severe blocking constraints. The improvement of the blocking probability when increasing the buffer size is not of importance for this case.

A suitable dimensioning of the buffer size is however a must when a tolerant admission policy scheme is used.

Finally, simulation has also shown that s_2 has to be chosen in a way as to have a suitable blocking probability for data calls without much penalizing voice: s_2 is to have an intermediate value between the very tolerant and the very strict admission control. However, numerical values show that even for very tolerant policies, blocking probability for voice remains considerably small. On the other hand, the choice of a tolerant policy is more suitable when data and best effort bursts have constant lengths. This means that in all cases, a tolerant policy with a high value of s_2 , appears to be a fair scheme. The impact of such a decision over the signaling traffic must however be evaluated in order to determine the exact values of the threshold parameters.

References

- [1] Rima Abi Fadel, Samir Tohmé. *Connection Admission Control CAC and Resources Allocation RA on the Up/Down Link UDL in a Low Earth Orbit LEO Satellite Constellation*. Proc. IEEE Int. Symp. Comp. and Commun., Hammamet, Tunisia, Jul.2001.
- [2] Poethi Boedhihartono, Gérard Maral. *Contribution to the Description of Mission Scenario for Handover Studies*. ENST June 2000 – RNRT Satellite Constellation Project. WP 2.2 on Handover and Associated Signalling
- [3] Heba Koraitim, Samir Tohmé. *Resource Allocation and Connection Admission Control in satellite Networks*. IEEE J. Select. Areas Commun., Vol. 17, No. 2, Feb. 1999.
- [4] H.Koraitim, G.Schäfer, S.Tohmé. *Quality of Service Aspects of Transport Technologies for the UMTS Radio Access Network*. IFIP Personal Wireless Communication Conference PWC'2000, Published by Kluwer, September 2000, Gdansk, Poland.
- [5] Sara Oueslati. *QOS Routing of Elastic Flows in Multi-service Networks*, PhD Dissertation, Ecole Nationale Supérieure des Telecomm. Paris, Nov. 2000.
- [6] P.Tran-Gia, F.Hubner. *An Analysis of Trunk Reservation and Grade of Service Balancing Mechanisms in Multiservice Broadband Networks*. IFIP TC6 Modelling and Performance Evaluation Workshop. La Martinique, 1993.
- [7] B.Tripathi, A.Kumar. *Performance analysis of microcellisation with channel reservation for supporting two mobility classes in cellular wireless networks*. Proc. IEEE Int. Conf. on Pers. Wireless Commun. ICPWC, India, 1997.
- [8] R.S. Krupp. *Stabilization of alternate routing network*. IEEE Int. Comm. Conference, Philadelphia, PA, 1982.
- [9] M. Rajaratnam, F. Takawira. *Modelling Multiple traffic streams in Circuit-Switched Networks*. IEEE Global Telecomm. Conf. GLOBECOM '96. 'Commun.: The Key to Global Prosperity', Vol. 1.
- [10] Ren P. Liu, Peter J. Moylan. *Dynamic Trunk Reservation for teletraffic links*. IEEE Proc, Global Telecomm. Conf., GLOBECOM '95., Vol. 1.
- [11] Rima Abi Fadel, Samir Tohmé. *Hybrid Connection Admission Control CAC in a Low Earth Orbit LEO Satellite Constellation*. IFIP Workshop on IP and ATM Traffic Management, WATM'01, Paris, September 2001.
- [12] V. Anantharam, M. Benchekroun. *Trunk Reservation based control of circuit switched networks with dynamic routing*. IEEE Proc. of the 29th Conf. on Decision and Control. Honolulu, Hawaii. Dec. 1990.
- [13] T. Oda, Y. Watanabe. *Optimal Trunk Reservation for a Group with Multislot Traffic Streams*. IEEE Trans. Commun., Vol. 38, no. 7, July 1990.

[14] Roch Guérin. *Queuing - Blocking System with Two Arrival Streams and Guard Channels*. IEEE Trans. on Commun., Vol. 36, no. 2, Feb. 1998.

[15] C.H. Yoon, C.K. Un. *Performance of personal portable radio telephone systems with and without guard channels*. IEEE J. Select. Areas Commun., vol. 11, issue 6 , Aug. 1993.

[16] B. Narendran, P. Agrawal, D.K. Anvekar. *Minimizing Cellular Handover Failures without Channel Utilization Loss*. Proc. IEEE Global Telecomm. Conf., 1994. GLOBECOM '94.

[17] O.T.W. Yu, V.C.M. Leung. *Self-tuning prioritized call handling mechanism with dynamic guard channels for mobile cellular systems*. Proc. IEEE 46th Vehic. Technol. Conf., 1996. Mobile Technology for the Human Race, vol. 3.

[18] A. L. Beylot, S. Boumerdassi, G. Pujolle. *A new prioritized handoff strategy using channel reservation in wireless PCN*. Proc. IEEE Global Telecomm. Conf., GLOBECOM 1998. The Bridge to Global Integration, vol. 3.

[19] O.T.W Yu, V.C.M. Leung. *Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN*. IEEE J. Select. Areas Commun., Sept. 1997, vol. 15, iss. 7.

[20] M.H. Chiu, M.A. Bassiouni. *Predictive schemes for handoff prioritization in cellular networks based on mobile positioning*. J. Select. Areas Commun., Mar. 2000, vol. 18, iss. 3.

[21] R. Ramjee, R. Nagarajan, D. Towsley. *On optimal call admission control in cellular networks*. Proc. IEEE INFOCOM '96, vol.1, pp. 43 -50.

[22] D. Calin. *A probabilistic model for handling voice and data traffic in wireless networks*. IEEE Int. Conf. on Universal Personal Commun., 1998. ICUPC '98, vol. 1.

Appendix

i, j and k are respectively the number of voice calls, data calls and best effort calls being served. l and m are respectively the number of data calls and best effort calls being queued. N represents the total number of available channels. sup_l and sup_m denote respectively the capacity of the data queue and the best effort queue. $I_{condition}$ is a variable set to 1 when the condition holds true, and is equal to 0 otherwise.

Equations of the analytical solution, numerically under C++, are the following.

For k varying from 0 to $sup_k = \text{int}(s_3/2)$
 For j varying from 0 to $sup_j = \min(\text{int}(s_3/2), \text{int}((N-2*k)/2))$
 For i varying from 0 to $sup_i = (N-2*j-2*k)$:

If $(i+2*j+2*k) \leq s_3 - 2$:

$$\begin{aligned}
 & (\lambda_h + \lambda_l + \lambda_{be} + i\mu_h + j\mu_l + k\mu_{be})\pi_{i,j,k,0,0} = I_{i>0}\lambda_h\pi_{i-1,j,k,0,0} + I_{j>0}\lambda_l\pi_{i,j-1,k,0,0} \\
 & + I_{k>0}\lambda_{be}\pi_{i,j,k-1,0,0} + I_{i<sup_i}(i+1)\mu_h\pi_{i+1,j,k,0,0} + I_{j<sup_j}(j+1)\mu_l\pi_{i,j+1,k,0,0} \\
 & + I_{k<sup_k}(k+1)\mu_{be}\pi_{i,j,k+1,0,0}
 \end{aligned} \tag{1}$$

If $(i+2*j+2*k) = s_3 - 1$ or $(i+2*j+2*k) = s_3$:

$$\begin{aligned}
 & (\lambda_h + \lambda_1 + 1_{m < \sup m} \lambda_{be} + i\mu_h + j\mu_1 + k\mu_{be}) \pi_{i,j,k,0,m} = 1_{m=0} 1_{i>0} \lambda_h \pi_{i-1,j,k,0,m} \\
 & + 1_{m>0} 1_{i>0} 1_{(i+2j+2k)=s_3} \lambda_h \pi_{i-1,j,k,0,m} + 1_{m=0} 1_{j>0} \lambda_1 \pi_{i,j-1,k,0,m} + 1_{m=0} 1_{k>0} \lambda_{be} \pi_{i,j,k-1,0,m} \\
 & + 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j,k,0,m} + 1_{k>0} 1_{(i+2j+2k)=s_3} 1_{m < \sup m} 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j,k-1,0,m+1} \\
 & + 1_{j < \sup j} (j+1)\mu_1 \pi_{i,j+1,k,0,m} + 1_{k>0} 1_{(i+2j+2k)>s_3-2} 1_{m < \sup m} 1_{j < \sup j} (j+1)\mu_1 \pi_{i,j+1,k-1,0,m+1} \\
 & + 1_{m>0} \lambda_{be} \pi_{i,j,k,0,m-1} + 1_{k < \sup k} (k+1)\mu_{be} \pi_{i,j,k+1,0,m} + 1_{m < \sup m} k\mu_{be} \pi_{i,j,k,0,m+1}
 \end{aligned} \tag{2}$$

If $s_3 < (i+2*j+2*k) \leq s_1 - 2$:

$$\begin{aligned}
 & (\lambda_h + \lambda_1 + 1_{m < \sup m} \lambda_{be} + i\mu_h + j\mu_1 + k\mu_{be}) \pi_{i,j,k,0,m} = 1_{i>0} \lambda_h \pi_{i-1,j,k,0,m} \\
 & + 1_{j>0} \lambda_1 \pi_{i,j-1,k,0,m} + 1_{m>0} \lambda_{be} \pi_{i,j,k,0,m-1} + 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j,k,0,m} \\
 & + 1_{j < \sup j} (j+1)\mu_1 \pi_{i,j+1,k,0,m} + 1_{k < \sup k} (k+1)\mu_{be} \pi_{i,j,k+1,0,m}
 \end{aligned} \tag{3}$$

If $(i+2*j+2*k) = s_1 - 1$ or $(i+2*j+2*k) = s_1$:

$$\begin{aligned}
 & (\lambda_h + 1_{l < \sup l} \lambda_1 + 1_{m < \sup m} \lambda_{be} + i\mu_h + j\mu_1 + k\mu_{be}) \pi_{i,j,k,l,m} = \\
 & 1_{l=0} 1_{i>0} \lambda_h \pi_{i-1,j,k,l,m} + 1_{l>0} 1_{i>0} 1_{(i+2j+2k)=s_1} \lambda_h \pi_{i-1,j,k,l,m} \\
 & + 1_{l=0} 1_{j>0} 1_{(i+2j+2k)>s_3} \lambda_1 \pi_{i,j-1,k,l,m} + 1_{l>0} \lambda_1 \pi_{i,j,k,l-1,m} + 1_{m>0} \lambda_{be} \pi_{i,j,k,l,m-1} \\
 & + 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j,k,l,m} + 1_{j>0} 1_{(i+2j+2k)=s_1} 1_{l < \sup l} 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j-1,k,l+1,m} \\
 & + 1_{j < \sup j} (j+1)\mu_1 \pi_{i,j+1,k,l,m} + 1_{l < \sup l} 1_{(i+2j+2k)>s_1-2} j\mu_1 \pi_{i,j,k,l+1,m} \\
 & + 1_{k < \sup k} (k+1)\mu_{be} \pi_{i,j,k+1,l,m} \\
 & + 1_{j>0} 1_{k < \sup k} 1_{l < \sup l} 1_{(i+2j+2k)>s_1-2} (k+1)\mu_{be} \pi_{i,j-1,k+1,l+1,m}
 \end{aligned} \tag{4}$$

If $s_1 < (i+2*j+2*k) < N$:

$$\begin{aligned}
 & (\lambda_h + 1_{l < \sup l} \lambda_1 + 1_{m < \sup m} \lambda_{be} + i\mu_h + j\mu_1 + k\mu_{be}) \pi_{i,j,k,l,m} = \\
 & 1_{i>0} \lambda_h \pi_{i-1,j,k,l,m} + 1_{l>0} \lambda_1 \pi_{i,j,k,l-1,m} + 1_{m>0} \lambda_{be} \pi_{i,j,k,l,m-1} \\
 & + 1_{i < \sup i} (i+1)\mu_h \pi_{i+1,j,k,l,m} + 1_{j < \sup j} 1_{(i+2j+2k) < N-1} (j+1)\mu_1 \pi_{i,j+1,k,l,m} \\
 & + 1_{k < \sup k} 1_{(i+2j+2k) < N-1} (k+1)\mu_{be} \pi_{i,j,k+1,l,m}
 \end{aligned} \tag{5}$$

For $(i+2*j+2*k) = N$:

$$\begin{aligned}
 & (1_{l < \sup l} \lambda_1 + 1_{m < \sup m} \lambda_{be} + i\mu_h + j\mu_1 + k\mu_{be}) \pi_{i,j,k,l,m} = \\
 & 1_{i>0} \lambda_h \pi_{i-1,j,k,l,m} + 1_{l>0} \lambda_1 \pi_{i,j,k,l-1,m} + 1_{m>0} \lambda_{be} \pi_{i,j,k,l,m-1}
 \end{aligned} \tag{6}$$

The equation written for $i=N, j=k=l=m=0$ is replaced by the equation stating that the sum of all state probabilities of the system is equal to one, in order to ensure that the numerical problem has only one solution.