Dynamic Resource Reservation in IEEE 802.16 Broadband Wireless Networks

Kamal GAKHAR, Mounir ACHIR, Annie GRAVEY
ENST-Bretagne
Computer Science Department
ENST-Bretagne CS 83818, 29200 Brest France
e-mail: {kamal.gakhar; mounir.achir; annie.gravey}@enst-bretagne.fr

***Abstract*— This paper presents a mechanism for dynamic resource management and discusses its relevance for traffic in IEEE 802.16 broadband wireless network. The basic goal is to minimize the amount of bandwidth being actually provisioned for committed bandwidth traffic while keeping the cost of MAC signalling to a minimum. In particular, this mechanism restricts the provisioned bandwidth to a predefined minimum when the actual offered load is significantly lower than the load that has been taken as a dimensioning objective. The proposed mechanism dynamically changes the amount of reserved resources between a small number of values (Two in the base model) depending on the actual number of active connections while limiting the number of transitions by imposing a hysteresis behaviour. In particular, it is not necessary to update the resource reservation whenever a traffic flow is activated or terminated. A Markov Chain model yields two performance parameters : the reserved bandwidth and the transition rate. A new parameter, noted θ, has been introduced in addition to the performance parameters discussed to minimize the global cost of the system. A generalization of this method to more than a single threshold is also proposed and discussed.**

I. Introduction

To ensure optimum reservation and utilization of resources while minimizing the costs involved for a network has always been a challenging problem. Our approach is based on dynamic reservation of the bandwidth following a mechanism of hysteresis. The focus of this paper is the dimensioning of IEEE 802.16 network [1] which needs an admission control policy (also called CAC: Call Admission Control). In fact, an admission control policy needs to be implemented in these networks as the data are transmitted in a "connection" mode. To explain it precisely, each subscriber station (SS) needs to establish a connection to the base station (BS) before it could transmit the packets. Although IEEE 802.16 standard recommends to implement a CAC policy (and support all necessary MAC mechanisms), no CAC policy is specified in the standard, and it is actually possible not to implement a CAC by setting up permanent connections when a SS is initialized.

Two broad types of traffics are supported in networks: Committed Bandwidth (CB) traffic and Best Effort (BE) traffic. CB traffic requires that a fixed amount of bandwidth be reserved; when less number of resources are available than needed, the QoS for the application degrades. Moreover, if CB traffic shares a globally reserved amount of bandwidth, that is not sufficient to support all the active CB flows, then all the active flows are likely to be degraded. CB traffic can be multimedia (e.g. ToIP or IPTV) but can also be pure data traffic for which a strict SLA is negotiated. It is expected that several classes of CB traffic be supported by network operators, who should dimension their system in order to limit the blocking probability, i.e. the probability that a new CB flow is not accepted due to lack of resources.

IEEE 802.16 standard offers three classes of connections to support CB traffic. These differ on the QoS that is required by the applications supported on these connections [1][3]. The present work does not refer to one specific class, but is applicable to any framework where the operator wishes to minimize the blocking probability for CB traffics, while limiting the signalling and configuration costs involved by modifying the amount of reserved bandwidth.

In general we can identify two families of policies for resource reservation. The first represents the mechanisms where resources are reserved in a semi-permanent or permanent manner also called permanent virtual circuit (PVC). The second family comprises of the procedures in which resources are reserved on demand also known as switched virtual circuit (SVC) [2]. The PVC method has a very low cost in terms of signalling and configuration but relies on overprovisionning for the supported traffic. This in turn may lead to excess blocking for other classes of CB traffics. On the other hand, the SVC method reserves only what is requested but incurs a high cost in terms of signalling and provisionning. This work presents a compromise between the PVC and the SVC methods by proposing a dynamic resource allocation mechanism of PVC type based on two state reservation. In this mechanism the reserved bandwidth varies depending on how much of it is being used and on a threshold. In the following section, we present our dynamic reservation framework. In section 3 we discuss the state of the art of some resource reservation mechanisms. In section 4, we see the hypothesis and the principles of proposed mechanisms. Following that, in section 5, we introduce and resolve respectively the system model based on Markov Chains. In section 6, we will see analytical results concerning modeling. A generalized approach to the mechanism is presented in section 7 with a conclusion in

1-4244-0476-2/06/$20.00 ©2006 IEEE.

section 8.

II. Dynamic Reservation Framework

The objective of this paper is to investigate methods that limit the cost of signalling and configuration while adapting to varying levels of offered load for a single CB class of traffic. Our study can be applied typically in hybrids WiFi/WiMAX networks [3], see Fig. 1. Note that in this architecture, the WIMAX network is a WAN that collects the traffic from primary (WiMAX) SS, while each primary SS directly connects one or several Access Points (either WiFi, or WiMAX) on which actual subscribers are directly connected through secondary SS. In this type of architecture, if the configuration of a primary SS has to be modified every time a new CB flow is activated at a secondary SS, the signalling load between the primary BS and SS would possibly be overly important. Therefore, it makes sense to limit the configuration load for primary SS by implementing a semi-static configuration that is seldom modified, but can still support CB traffic with good QoS, both in the command plane (by limiting the blocking probability) and in the transfer plane (by ensuring that enough resources are available for each non-blocked CB flow).

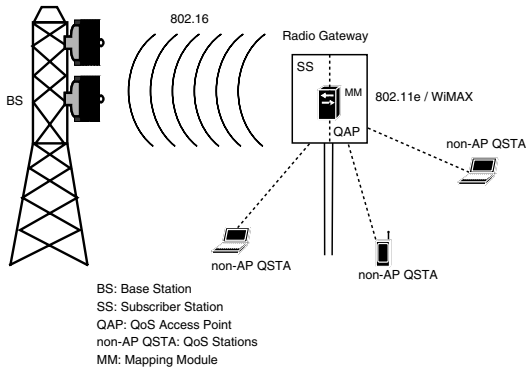

Fig. 1. WiMAX-WiFi hybrid network architecture

The above hybrid architecture has to be dimensioned for a target traffic class e.g. on the basis of the "busy hour" offered load of this traffic, but may actually operate at a significantly lower load. If the network operator prefers to use only permanent connections, he has to choose between reserving a large amount of resources (based on the "busy hour" offered load of the target traffic) and then risk blocking other types of CB traffic, or reserving a significantly smaller amount of resources for the target traffic and then risk a performance degradation for this traffic. The mechanism we propose here allows to operate with a minimum amount of reserved resources when the actual load is small, while seamlessly reserving the necessary amount of resources when the actual load significantly increases. The cost in term of signalling can be optimised together with the gain in reserved resources. However, as we shall see further, the selection of the optimum set of parameters depends on expected offered load, and on the respective costs of signalling and reserved resources. It is therefore up to the operator to select the appropriate set of parameters according to its cost structure and expected traffic profiles.

III. The State of the Art

In current literature we find many references to mechanisms and methods which explain dynamic resource reservation in a system consisting of multiple traffic classes.

To refer to first work [7], the hypotheses in this paper consider multiple thresholds $(T_0, T_1, T_2, ..., T_k, ..., T_n)$. The reserved bandwidth varies from one to another threshold until the number of channels being used, $n(t)$, reaches a prefixed threshold. The reserved bandwidth can be expressed as : $BP(n(t)) = min(T_i) : T_i \geq n(t)$, whereas $n(t)$ is the number of channels being used. The equations for minimizing the bandwidth being used have been resolved and give the corresponding transition rate, R^* (number of changes in threshold per second). An extension for this mechanism has also been proposed inspired by the one proposed in [8]. This one proposed to use two thresholds around each reservation threshold T_k, a superior threshold noted as $u(k)$ and second inferior noted as $l(k)$. This extension insights to minimize transition rate. The proposition is to calculate $u(k)$ and $l(k)$ in a manner to obtain a transition rate approaching R^* which minimizes the reserved bandwidth. The results show that when the transition rate increases the reserved bandwidth decreases hence the blocking probability.

Another work of inspiration is [9] that gave a mechanism to optimize resources allocations in ATM networks. This mechanism combines set up and free connections processes, dynamic bit rate allocation and flows control. A markovian model is developed to evaluate the performance of the mechanism. This model establishes an exact analysis of delays and rejected request without numerical computations. This is a hysteresis mechanism. In fact, the passage of the bandwidth to a superior or inferior reservation is attained as a function of the state of the client queue at ATM switch. Our solution is closed to the one proposed in this work except that the change in the reserved bandwidth is done as a function of the number of active connections. The principles of our solution are discussed in section IV.

Other solutions of resource reservation are proposed but for the cellular networks [4]. This paper takes two traffic classes into account : the new calls and the handover calls. The idea is to vary a threshold dynamically which in fact represents the maximum number of channels above which all new calls will be rejected. This adaptive mechanism is characterized by four parameters : α_u, α_d, N and τ. The principle followed is such that if the blocking probability for the handover traffic is greater than a certain value, $P^h > \alpha_u.B^h$, the threshold for new calls, l^n, will be decremented by 1. If during N ($N = 10$) consecutive handovers we have $P^h < \alpha_d.B^h$, the threshold l^n will be incremented by 1. The τ (2 hours) corresponds to the estimated period of blocking probability of the handover traffic. In this approach we need to manage

multiple parameters whereas the time of convergence to obtain stationary values of blocking probabilities is relatively longer.

In order to ameliorate the time of convergence of this method and to reduce the number of parameters to manage, [5] proposes a less complex yet more efficient method. This one needs three parameters α_u, α_d and P_{dec} (the probability to reduce l^n). The principle used is globally the same as discussed in [4] and have its basis on estimation of the blocking probability of handover traffic P^h . The advantage is that it is less complex as it involves less number of parameters to be manipulated. In this method the blocking probability is calculated by memorizing all the blocked handovers and the threshold l^n is updated at each blocked handover which results in short convergence time compared to the mechanism in [4]. If the esteemed probability is inferior to $\alpha_d.B^h$ then the threshold l^n is incremented by 1. When $\alpha_d.B^h < P^h < \alpha_u.B^h$, the threshold l^n will be decreased with a probability P_{dec} and incremented with the probability $(1 - P_{dec})$ where as if $\alpha_u.B^h < P^h$, the threshold l^n is decreased with a probability P_{dec} and is incremented with a probability of $(1 - P_{dec})$. The simulation results show a gain of 50% for convergence time towards the targeted probabilities whereas the precision measure remains the same for two mechanisms.

The article [6] proposes a dynamic mechanism and compare its performances with the mechanisms presented in [4] and [5]. The general principle is very simple as the acceptation and the rejection of the calls are based on MGC (Multiple Guard Channel). A call is accepted if the number of occupied channels is inferior to the threshold $N^i < l^i$, otherwise it is rejected. The mechanism is adaptive and modify the threshold for each class after each connection demand. The desired blocking probability for the traffic class i is noted $Bi = b_i/o_i$ with b_i the number of rejected calls and o_i the total number of calls. The blocking probability is evaluated in varying the threshold l^i in the following manner: when a call from class i is accepted, the threshold is decreased as $l^i = l^i - 1$, with the probability $1/(o_i - b_i)$, and when a call is rejected, the threshold is incremented as $l^i = l^i + 1$ with the probability $(1/b_i)$. The simulation results show the efficiency of this mechanism as compared to the ones presented in [4] and [5] in terms of the load accepted by the network and the mean value of threshold.

To focus back, we base our approach for CB type traffic and propose our hypothesis and principles in the following section.

IV. Model Hypothesis and Principles

Our context involves a SS (subscriber station) which requests a certain amount of bandwidth reservation to the BS. The basic principle of our approach is : the reserved bandwidth will vary between a minimum and a maximum value as per the bandwidth utilized by the clients. We define the following parameters before moving on:

- C_M and C_m, maximum and minimum bandwidth respectively that could be reserved.
- T, a fixed threshold with a value inferior to that of C_m.
- $C(t)$, the instant bandwidth at a given t.

The reserved bandwidth varies as per the following rules:

- The reserved bandwidth will be equal to C_m when the bandwidth being used, $C(t)$, is less than C_m.
- When $C(t)$ attains C_m , the reserved bandwidth passed on from C_m to C_M.
- When the bandwidth being used, $C(t)$, moves to a value less than that of threshold T, the reserved bandwidth gets reduced from C_M to C_m.

From the rules above, it is clear that the reserved bandwidth will be controlled, in a simple manner, by the bandwidth being actually used, that is, $C(t)$. This reserved bandwidth is obviously inferior to C_M, which in turn could lead to a non-negligible gain in terms of bandwidth. Further we will see an example quantifying this gain. Fig.2 represents the reserved bandwidth variation as a function of the bandwidth used by the clients.

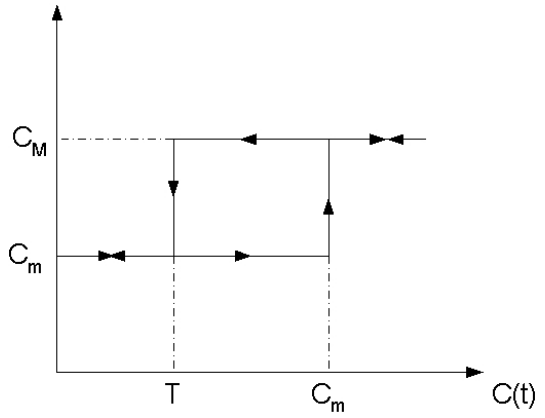

Fig. 2. The amount of reserved bandwidth varies according to the number of active flows

In addition to the reserved bandwidth, one more important criterion should be taken into account for the mechanism, namely, the choice of parameters C_m and T. In fact, the fluctuations of the reserved bandwidth from C_M to C_m, or reciprocally, should be the minimum possible as the signalling costs are involved with each of these.

V. System Characterization

This section discusses a Markov Chain model for the proposed mechanism. We consider the following hypothesis:

- The arrival of clients is a poisson process with a rate equal to λ.
- The connection duration is random and denoted by exponential parameter: μ.
- The bandwidth utilized by a connection: 1 channel.

The above hypothesis are common in any system where a large number of potential users initiate identical real time multimedia communications. The model then applies to the case where the target CB traffic is of this type (e.g. interactive voice or video). We assume that all traffic flows are identical.

The model extends easily to several types of CB traffic, without modifying the qualitative behaviour of the system. Let us define also the following variables:

- The maximum number of reserved channels: C_M.
- The minimum number of reserved channels: C_m.
- The threshold of passage from C_M to C_m : T.

It is easily seen that the pair (S, C) where S is the amount of reserved resources and C is the number of active flows is a simple Markov process illustrated below. The left hand branch of the graph is the set of states for which $S = C_m$, and the right hand branch is the set of states for which $S = C_M$.

Based on this hypothesis, we model the system using Markov Chains as shown in the Fig.3.

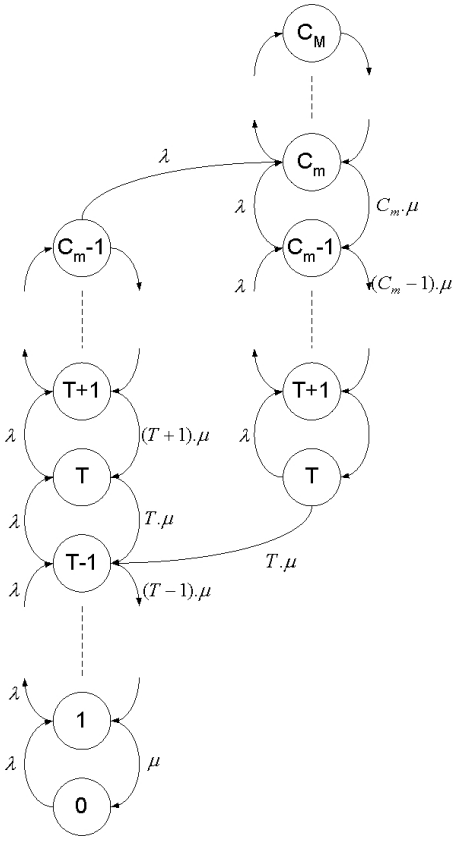

Fig. 3. One threshold system model based on Markov Chains

We resolved the state equations for the Markov Chains in order to calculate the mean reserved bandwidth. This Markov Chain is irreducible and non-periodic and thus allows a unique solution [10]. To avoid the complexity of mathematics involved in resolution we present only the probability states, the reserved bandwidth, and the transition rate. We notice:

- $P(i)$: the probability of the state i, for i between T and $C_m - 1$, given that the reserved bandwidth is equal to C_m.
- $Q(i)$: the probability of the state i, for i between T and $C_m - 1$, given that the reserved bandwidth is equal to C_M.

The expressions for $P(i)$ and $Q(i)$, for i between T and $C_m - 1$, are given as:

$$P(i) = R(T-1).\rho^{i-(T-1)}.\frac{D(C_m-(i+1))}{D(C_m-T)} \quad (1)$$

$$Q(i) = R(i) - R(T-1).\rho^{i-(T-1)}.\frac{D(C_m-(i+1))}{D(C_m-T)} \quad (2)$$

where as :

$$R(i) = \frac{\rho^i}{i!.\sum_{n=0}^{C_m}\frac{\rho^n}{n!}} \quad (3)$$

and :

$$D(k) = \sum_{n=0}^{k}\rho^{k-n}\prod_{j=0}^{n-1}(C_m-(k-j)) \quad (4)$$

For i between 0 and $T-1$ or between C_m and C_M, we have the probability of the state i equal to R_i, as in (3). Having obtained the probability states, we can calculate the mean reserved bandwidth as expressed in (V), and the transition rate by : (6).

$$BP = C_M - (R(0) + R(1) + ... + R(T-1) + P(T) + ... + P(C_m-1)).(C_M - C_m) \quad (5)$$

$$TR = \lambda.P(C_m-1) + \mu.T.Q(T) \quad (6)$$

VI. Results

We consider here a system in which the exact offered load is unknown, but is upper bounded by 20 Erlangs. This yields a C_M value of 30 (which corresponds to a blocking probability value of 10^{-2}). We discuss here the issue of selecting Cm and T values optimally.

The evolution of reserved bandwidth and transition rate has been observed in respect to various parameters that compose the system and is presented in (4) (varying T) and (6) and as in (5) (varying C_m) and (7). The load varies from 0 to ρ_{max}.

Fig.4 and 6 are obtained using the following parameters: $C_M = 30$, $C_m = 15$ and T varying between 1 et (C_m-1). We notice that in Fig. 4 the bandwidth increases very rapidly for small values of T. This happens when $C(t)$ surpasses (C_m-1) the bandwidth changes from C_m to C_M and stays at this value for a considerable time as the threshold T is small. On the other hand, when T gets close to C_m, the reserved bandwidth is minimum because in this case the reserved bandwidth varies from C_M to C_m, and inversely, each time that $C(t)$ surpasses or descends below $C_m - 1$. In Fig.5 and Fig.6, we have the mean reserved bandwidth and the transition rate in varying the C_m from $T + 1$ to $C_M - 1$.

In Fig.5, the mean reserved bandwidth increases when C_m increases. This gives us maximum reserved bandwidth for a maximum of C_m, this means up to $C_M - 1$. In Fig.6, the transition rate has a maximum for a value of T equal to C_m-1. In fact, as we notice here, when T is close to $C_m - 1$, the bandwidth varies from C_M to C_m and inversely with a greater

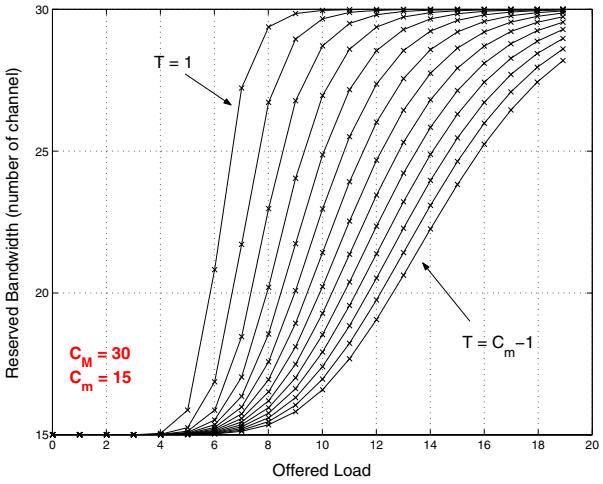

Fig. 4. Bandwidth reserved as a function of offered load

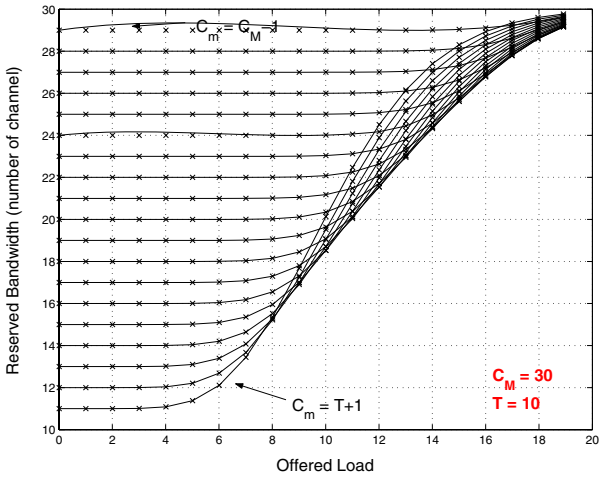

Fig. 5. Bandwidth reserved as a function of offered load

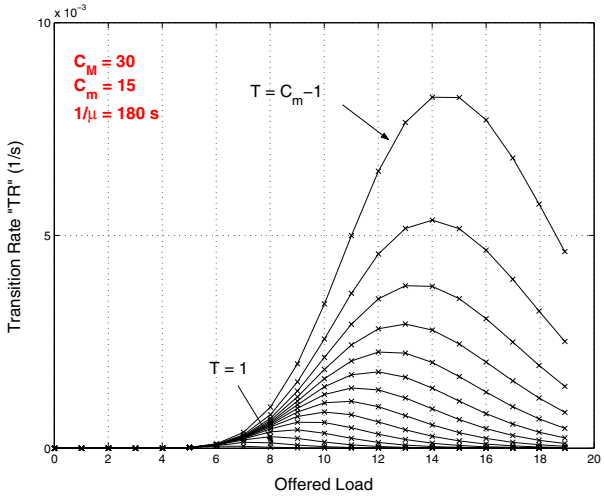

Fig. 6. Transition rate as a function of offered load

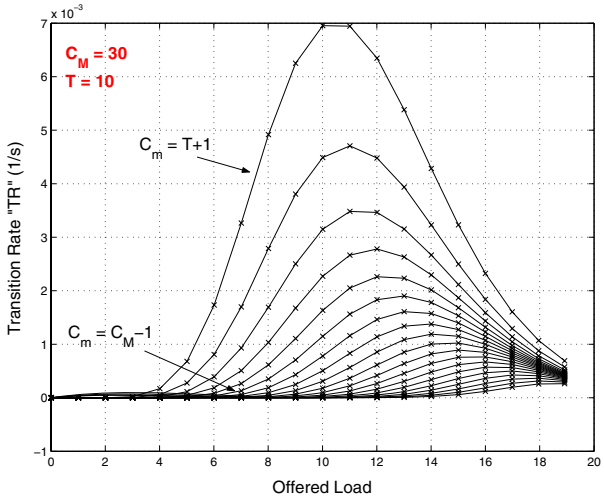

Fig. 7. Transition rate as a function of offered load

frequency which signifies an increase in the transition rate. For a given T, the transition rate is maximum when the mean number of channels is between T and $C_m - 1$. In the same manner for Fig.6, in Fig.7 we have a transition rate which is maximum for a C_m closer to T.

In Fig. 8 and 9, we have three-dimensional representation of the reserved bandwidth and the transition rate as functions of C_M and T. These clearly show us that the bandwidth and the transition rate have a minimum and a maximum respectively for the same values of (C_m, T). Thus it is clear that minimizing the bandwidth and the transition rate will need a compromise. This brings us to the task of obtaining a cost function which depends on the bandwidth, the transition rate, and a parameter θ. We have opted for a simple but efficient function in order to determine the optimum coupled value of (C_m, T) and is given as :

$$f_c = \frac{BP}{max(BP)} + \theta.\frac{TR}{max(TR)} \qquad (7)$$

where as :

- max(BP), the maximum value of reserved bandwidth
- max(TR), the maximum value of transition rate

θ represents the relative importance of blocking for other classes of CB traffic versus the signalling and configuration costs. A network operator should select the value of theta corresponding to its policy, and then use e.g. figure 10 to select optimal C_m and T values.

We obtain the curves of C_m and T as in Fig.10 as function of parameter θ. It is easy to see that for small values of θ we have a C_m which is close to T. It was expected for a small value of θ, the cost function can be estimated by $\frac{BP}{max(BP)}$. We would like to minimize the reserved bandwidth which in turn gives us (Cm, T) such that $T \approx Cm$. When θ is large, we obtain a value of (C_m, T) such that $C_m = C_M - 1$ and $T = 1$. In fact, in this case, the cost function can be estimated by $\theta.\frac{TR}{max(TR)}$ and the minimization of this gives us the value

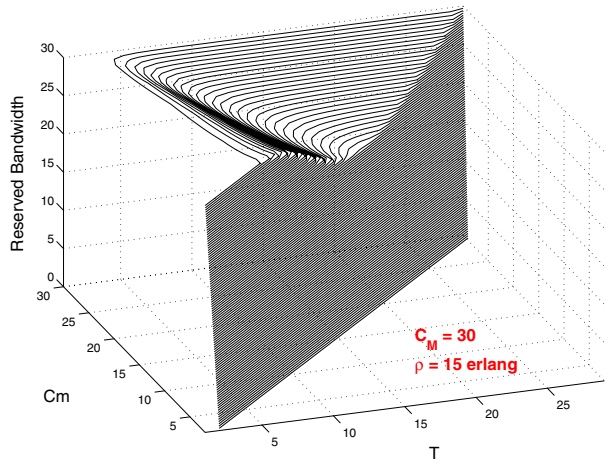

Fig. 8. Reserved Bandwidth as a function of T and C_m

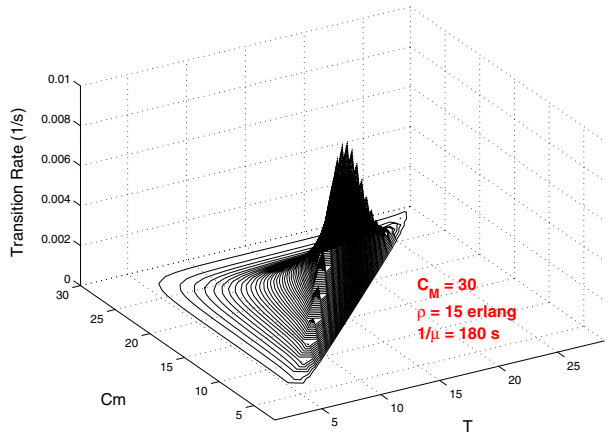

Fig. 9. Transition rate as a function of T and C_m

which minimizes the transition rate.

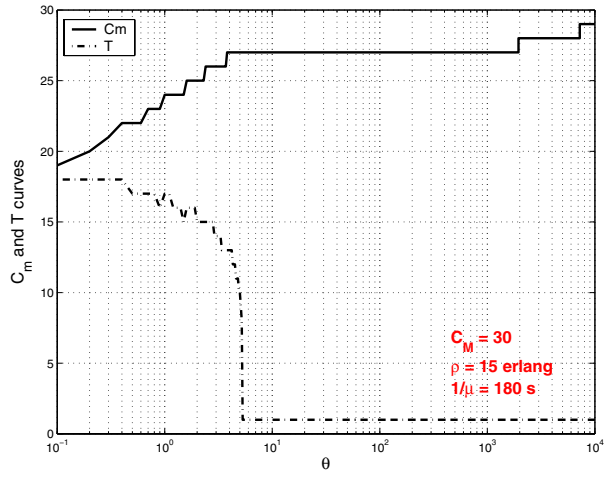

Fig. 10. Optimal C_m and optimal T as function of θ.

Fig.11 illustrates optimal C_m and T which minimize the cost function f_c versus total trafic load (ρ). For example if one dimensions a network for $\rho = 10$ Erlang, then C_m will be equal to 18 and T will be equal to 14.

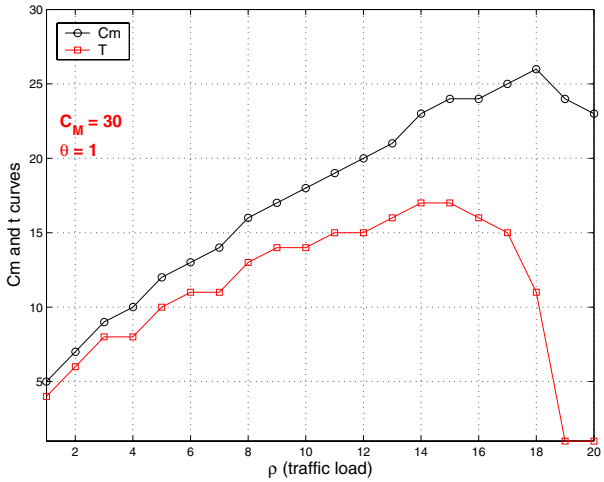

Fig. 11. Optimal C_m and optimal T as function of ρ.

When the offered load is near the maximum, C_m is close to C_M, and the system shall almost always reserve the C_M channels, since T is then close to 0. However, the interesting part of the figure is when the offered load is significantly smaller than the maximum (i.e. between 7 and 14). It is seen there that taking C_m between 15 and 20 and T equal to $C_m - 4$ is close to the optimum. On the other hand, if the offered load significantly varies between a low value (i.e. 5) and a high value (i.e. 10), there is no real optimal choice. This observation justifies considering more than one threshold, when the activity presents several clearly identified levels (i.e. office hours versus out of office hours). This generalisation is addressed in the following section.

VII. Hysteresis System With Multiple Threshold

This section presents the generalization of previously discussed mechanism. It is obtained by using many levels of reservation. In fact, such a reservation certainly allows us to optimize the reserved bandwidth as we will see in the results.

A. Hysteresis system with two thresholds

Fig.12 and 13 show us, respectively, the temporal curves of the reserved bandwidth and the number of channels being used for the hysteresis mechanisms with one and two thresholds. In Fig. 12, we consider a threshold of $C_m = 15$ to pass on from a reserved bandwidth of 15 to a reserved bandwidth equal to 30 where the threshold T equals 10. In Fig. 13, we observe the mechanism of hysteresis with two thresholds. The values of the thresholds are the following: $C_{m,1} = 20$, $T_1 = 16$, $C_{m,2} = 14$ and $T_2 = 10$. The reserved bandwidth changes from $C_{m,2} = 14$ to $C_{m,1} = 20$ when the number of occupied channels move to $C_{m,2} = 14$ and comes back to 14 when the number of occupied channels passed below $T_2 = 10$. In the

same manner, the reserved bandwidth passes on from $C_{m,1} = 20$ to $C_M = 30$ when the number of occupied channels attains $C_{m,1}$ and moves back to 16 when the number passed below 14.

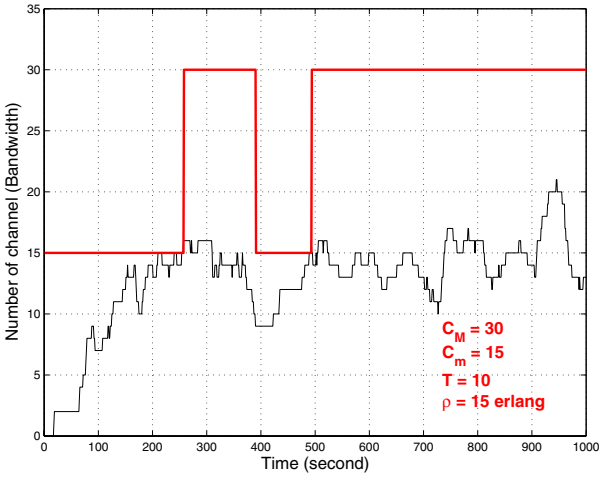

Fig. 12. Reserved bandwidth with hysteresis along one threshold

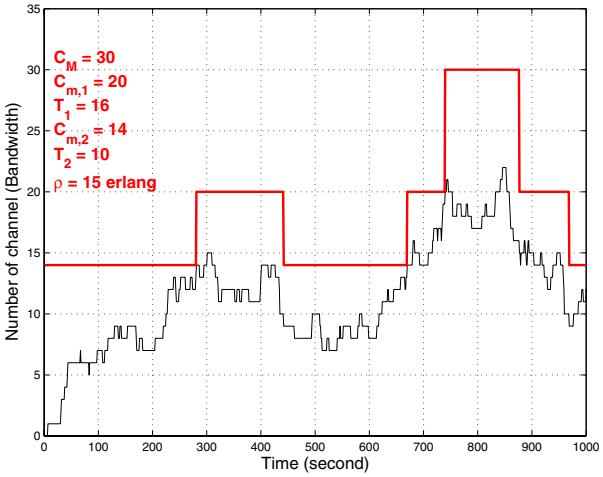

Fig. 13. Reserved bandwidth with hysteresis along two thresholds

Fig. 14 illustrates the Markov Chains for the system which is modeled using the functioning of a system at hysteresis with two thresholds. The analytical solution for this Markov Chain is complex but could be obtained intuitively using the system resolution presented earlier. The following are the probability state equations of the Markov Chains illustrated by the fig. 14.

- $P_1(i)$: the probability of the state i given the reserved bandwidth equal to $C_{m,1}$
- $Q_1(i)$: the probability of the state i given the reserved bandwidth equal to C_M
- $P_2(i)$: the probability of the state i given the reserved bandwidth equal to $C_{m,2}$
- $Q_2(i)$: the probability of the state i given the reserved bandwidth equal to $C_{m,1}$.

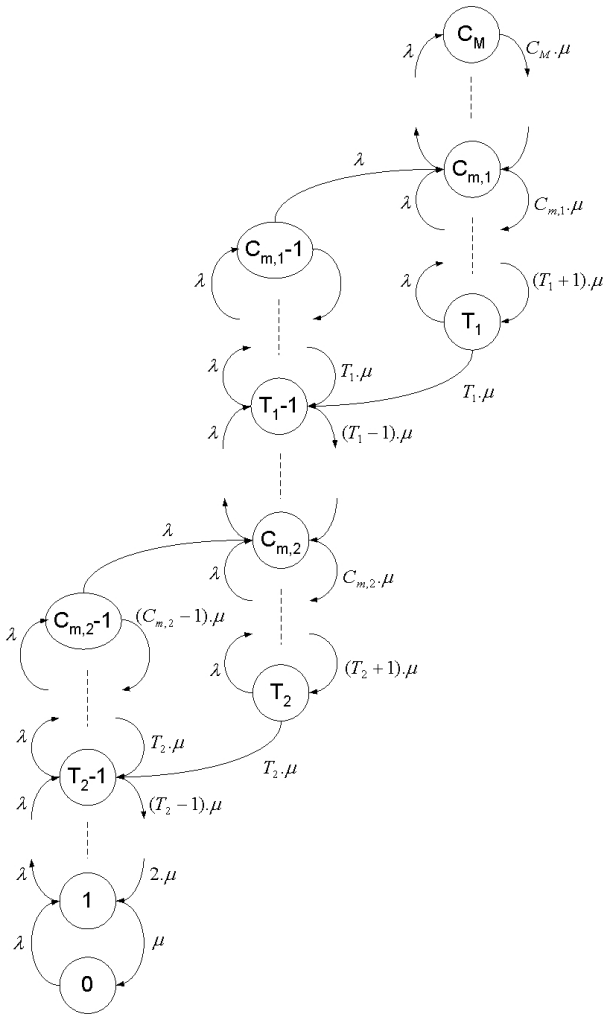

Fig. 14. Two thresholds system model based on Markov chains

The expressions for $P_1(i)$, $Q_1(i)$, $P_2(i)$ and $Q_2(i)$ are thus given by:

$$P_1(i) = R(T_1 - 1).\rho^{i-(T_1-1)}.\frac{D(C_{m,1} - (i+1))}{D(C_{m,1} - T_1)} \quad (8)$$

$$Q_1(i) = R(i) - R(T_1 - 1).\rho^{i-(T_1-1)}.\frac{D(C_{m,1} - (i+1))}{D(C_{m,1} - T_1)} \quad (9)$$

$$P_2(i) = R(T_2 - 1).\rho^{i-(T_2-1)}.\frac{D(C_{m,2} - (i+1))}{D(C_{m,2} - T_2)} \quad (10)$$

$$Q_2(i) = R(i) - R(T_2 - 1).\rho^{i-(T_2-1)}.\frac{D(C_{m,2} - (i+1))}{D(C_{m,2} - T_2)} \quad (11)$$

where as, for i between : T_1 and $C_{m,1} - 1$

$$R(i) = \frac{\rho^i}{i!} \sum_{n=0}^{C_M} \frac{\rho^n}{n!} \quad (12)$$

$$D_1(k) = \sum_{n=0}^{k} \rho^{k-n} \prod_{j=0}^{n-1} (C_{m,1} - (k-j)) \quad (13)$$

and for i between : T_2 and $C_{m,2} - 1$

$$R(i) = \frac{\rho^i}{i!} \sum_{n=0}^{C_M} \frac{\rho^n}{n!} \quad (14)$$

$$D_2(k) = \sum_{n=0}^{k} \rho^{k-n} \prod_{j=0}^{n-1} (C_{m,2} - (k-j)) \quad (15)$$

For i between 0 and $T_2 - 1$, between $C_{m,2}$ and $T_1 - 1$ or between $C_{m,1}$ and C_M, we have the probability of the state i : $R(i)$, given by the following equation :

$$R(i) = \frac{\rho^i}{i!} \sum_{n=0}^{C_M} \frac{\rho^n}{n!} \quad (16)$$

Given the probability states, we can calculate the mean reserved bandwidth given by (VII-A), and the transition rate given by (VII-A).

$$BP = C_M - \left(\sum_{i=0}^{T_2-1} R(i) + \sum_{i=T_2}^{C_{m,2}-1} P_2(i) \right) .(C_M - C_{m,2}) - (C_M - C_{m,1}) \cdot \left(\sum_{i=T_2}^{C_{m,2}-1} Q_2(i) + \sum_{i=C_{m,2}}^{T_1-1} R(i) + \sum_{i=T_1}^{C_{m,1}-1} P_1(i) \right) \quad (17)$$

$$TR = \lambda \left(P(C_{m,1} - 1) + P(C_{m,2} - 1) \right) + \mu \left(T_1 . P(T_1) + T_2 . P(T_2) \right) \quad (18)$$

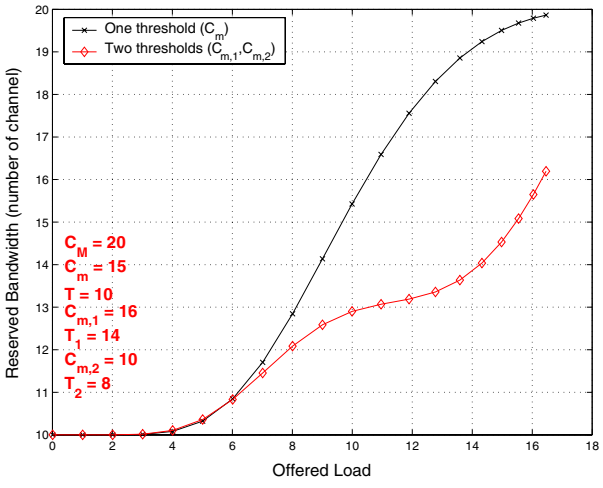

Fig. 15. Reserved Bandwidth as a function of offered load

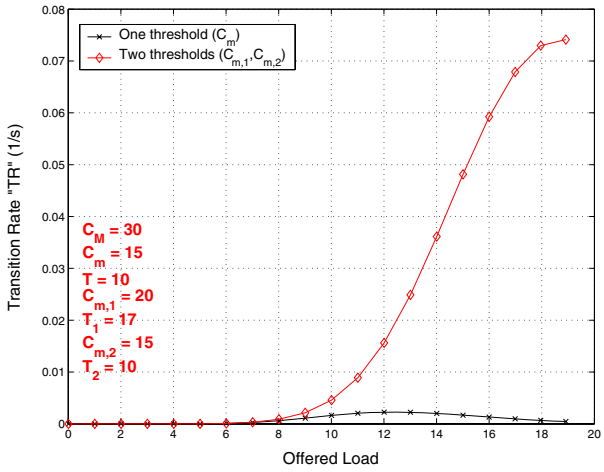

Fig. 16. Transition rate as a function of offered load

Fig. 15 shows the curves for the bandwidth reserved with one and two thresholds. We observe that the bandwidth reserved with two thresholds (near x-axis) clearly needs smaller values than with the one obtained using one threshold. However, as shown in the fig. 16, the transition rate for multiple thresholds increases brusquely.

B. General Case: system with multiple thresholds

We generalize the hysteresis mechanism with multiple thresholds. These multiple thresholds could be given as : C_M, $C_{m,1}$, $C_{m,2}$, $C_{m,3}$, ... , $C_{m,l}$, ... , $C_{m,n}$ and T_l, T_2, T_3, ... , T_l, ... , T_n. We notice that :

- $P_l(i)$: the probability of the state i given that the reserved bandwidth is equal to $C_{m,l}$
- $Q_l(i)$: the probability of the state i given that the reserved bandwidth is equal to $C_{m,(l+1)}$.

The expressions for $P_l(i)$ and $Q_l(i)$ are given by the following equations :

$$P_l(i) = R(T_l - 1).\rho^{i-(T_l-1)}.\frac{D(C_{m,l} - (i+1))}{D(C_{m,l} - T_l)} \quad (19)$$

$$Q_l(i) = R(i) - R(T_l - 1).\rho^{i-(T_l-1)}.\frac{D(C_{m,l} - (i+1))}{D(C_{m,l} - T_l)} \quad (20)$$

where as, for i lying between T_l and $C_{m,l} - 1$:

$$R(i) = \frac{\rho^i}{i!} \sum_{n=0}^{C_M} \frac{\rho^n}{n!} \quad (21)$$

$$D_l(k) = \sum_{n=0}^{k} \rho^{k-n} \prod_{j=0}^{n-1} (C_{m,l} - (k-j)) \quad (22)$$

For i between 0 and $T_n - 1$, between $C_{m,l+1}$ and $T_l - 1$ and between $C_{m,1}$ and C_M, we have the probability of the state i : $R(i)$, given by the following equation :

$$R(i) = \frac{\rho^i}{i!} \sum_{n=0}^{C_M} \frac{\rho^n}{n!} \quad (23)$$

Having identified the probability states, we can calculate the transition rate and the mean reserved bandwidth as following:

$$TR = \lambda \left(\sum_{l=1}^{n} P(C_{m,l} - 1) \right) + \mu \left(\sum_{l=1}^{n} T_l . P(T_l) \right) \quad (24)$$

$$BP = C_M - \left(\sum_{i=0}^{T_n - 1} R(i) + \sum_{i=T_n}^{C_{m,n}-1} P_n(i) \right) . (C_M - C_{m,n}) - \sum_{l=1}^{n-1} \left(\sum_{i=T_{l+1}}^{C_{m,l+1}-1} Q_{l+1}(i) + \sum_{i=C_{m,l+1}}^{T_l - 1} R(i) + \sum_{i=T_l}^{C_{m,l}-1} P_l(i) \right) . (C_M - C_{m,l}) \quad (25)$$

These two global equations (24 and 25) give a relation between the network parameters: C_M, ρ and the mechanism parameters $C_{m,l}$ et T_l. A global optimization of the cost function (see equation 7) is done by determining all the mechanism parameters and the total offered load (ρ). This implies an optimization with multiple parameters.

VIII. Conclusion

This work studies optimum reservation and utilization of bandwidth for Committed Bandwidth (CB) type traffic while minimizing MAC signalling costs. Our solution can be applied to any network that supports several classes of CB traffic, and in which signalling and configuration costs should be limited. This paper presents a comprehensive approach to model such a system with multiple thresholds with the help of Markov Chains. Many numerical examples and curves permit to have a quantitative appreciation of the mechanism behaviour. The results show that it is possible to minimize signalling (and thus transition rate) costs. However a compromise has to be reached to ensure optimum reservation of bandwidth in terms of signalling costs involved.

A general use of this mechanism is presented in the section VII. A dynamic allocation with multiple thresholds is modelled by a Markov chain. With this model we obtain global equations for the used bandwidth and the transition rate which represent the total signalling quantity. The proposed mechanism enters in the context of a WiMAX access network dimensioning for a long term load objective but which operates often at a significantly lower offered load.

The operator can than, at a minimal signalling and configuration cost, dynamically modify the amount of resources that is reserved for the target traffic, while ensuring that this traffic shall be offered a good performance as long as the offered load is lower than the long term load. The system operates in a semi-permanent mode, in which there is no need to update the amount of reserved resources for each activated or terminated flow, but only to modify the reservation when the number of active flow becomes larger than a first threshold, or smaller than a second one. We have shown how to optimally select the threshold values, depending on the operator cost structure, and on the expected load. Moreover, if the load is expected to operate at several well identified values, the generalized system can take into account several thresholds in order to decrease the global operational costs.

References

[1] IEEE 802.16 standard - Local and Metropolitan Area Networks, 2004.

[2] K. Wongthavarawat, A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems", International Journal of Communication Systems, 2000.

[3] K. Gakhar, A. Gravey, A. Leroy, "IROISE: A New QoS Architecture for IEEE 802.16 and IEEE 802.11e Interworking", 2nd IEEE/Create-Net International Workshop on Deployment Models and First/Last Mile Networking Technologies for Braodband Community Networks, Oct. 2005.

[4] Y. Zhang, D. Liu, "An adaptive algorithm for call admission control in wireless networks", in Proceeding of the IEEE Global Communications Conference (GLOBECOM), Nov. 2001.

[5] X.-P. Wang, J.-L. Zheng, W. Zeng, G.-D. Zhang, "A probability-based adaptive algorithm for call admission control in wireless network", in Proceedings of the International Conference on Computer Networks and Mobile Computing (ICCNMC), Oct. 2003.

[6] D. Garcia-Roger, M. Jose Domenech-Benlloch, J. Martinez-Bauset, V. Pla, "Comparative evaluation of adaptive trunk reservation schemes for mobile cellular networks", Third International Working Conference, Performance Modelling and Evaluation of Heterogenous Networks, HET-NET 2005.

[7] H. Levy, T. Mendelson, G. Goren, "Dynamic allocation of resources to virtual path agents", IEEE/ACM Transactions on networking, vol. 12, N. 4, August 2004.

[8] A. Orda, G. Pacifici, D. E. Pendarakis, "An adaptive virtual path allocation policy for broadband networks", in Proceeding of the IEEE Computer Societies conference (INFOCOM), Mars 1996.

[9] S. Halberstadt, D. Kofman and A. Gravey, "A congestion control mechanism for connectionless services offered by ATM networks", in Proc. of IFIP TC.6 3d Workshop on Performance Modeling and Evaluation of ATM networks, 1996.

[10] B. Baynat, Théorie des files d'attente : des chaînes de Markov aux réseaux à forme produit, Edition Hermes 2000.