International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:12, 2008

An Innovative Approach to the Formulation of Connection Admission Control Problem

Carlo Bruni, Francesco Delli Priscoli, Giorgio Koch, and Ilaria Marchetti

***Abstract*—**This paper proposes an innovative approach for the Connection Admission Control (CAC) problem. Starting from an abstract network modelling, the CAC problem is formulated in a technology independent fashion allowing the proposed concepts to be applied to any wireless and wired domain. The proposed CAC is decoupled from the other Resource Management procedures, but cooperates with them in order to guarantee the desired QoS requirements. Moreover, it is based on suitable performance measurements which, by using proper predictors, allow to forecast the domain dynamics in the next future. Finally, the proposed CAC control scheme is based on a feedback loop aiming at maximizing a suitable performance index accounting for the domain throughput, whilst respecting a set of constraints accounting for the QoS requirements.

***Keywords*—**Network Management, Quality of Service (QoS) requirements, Optimal Control.

I. Introduction

THE present challenge in the telecommunication arena is to enhance the Internet Protocols with respect to the following three basic features: (i) Quality of Service (QoS), (ii) Mobility, (iii) Security.

In this respect, several research projects (e.g. [1],[2],[3],[4]) are proposing to add a technology-independent layer between the IP layer and the technology dependent Underlying Network (UN) layers, herafter referred to as *Convergence Layer*, which is *transparent* with respect to both the IP layer and the UN layers, i.e. its insertion between the IP layer and the UN layers does not modify either the usual IP protocols or the usual UN protocols. This layer includes UN independent protocols aiming at the three above-mentioned enhancements. Clearly, only those elements of the telecommunication network in which the enhancements in question are actually necessary will be provided with the Convergence Layer.

In light of the above, the overall conceptual layering architecture of the present telecommunication network can be represented as in Fig. 1 in which, for the sake of simplicity, we have represented just four different domains with three technology dependent layers.

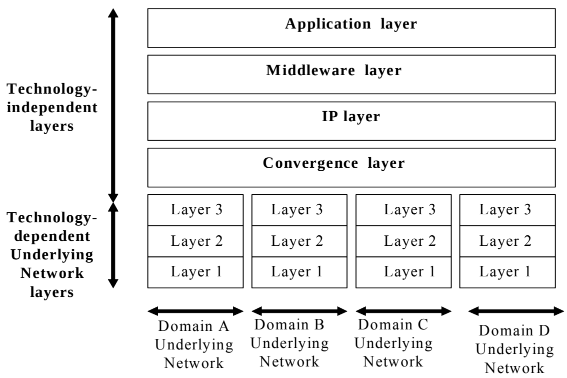

Fig. 1 Overall Layering architecture

It is important noting that, in each domain, some (even all) of the functionalities offered by the Convergence Layer can be disabled in case the Underlying Network already provides them in a satisfactory way.

In this context, the more research moves toward problem formulation and possible solutions which leave out of consideration the specific technology dependent issues, the more relevance is attributed to approaches and methodologies such as those of system theory.

Highlights of system theory are mathematical modelling of the relevant processes and formulation of management tasks as formal control problems. This may happen for various issues dealt with in the *Convergence Layer* such as all Resource Management procedures (Connection Admission Control, Congestion Control, Scheduling, Routing, Dynamic Capacity Allocation) aiming at maximizing the exploitation of network resources while keeping a satisfactory Quality of Service level.

The aim of this paper is to provide an abstract QoS modelling which allows to implement QoS related controls in a technology independent fashion, and in particular to formulate the Connection Admission Control (CAC) as an optimal control problem.

This paper is partly sponsored by the EuQoS (End-to-end Quality of Service support over heterogeneous networks) project and by the IMAGES (Integrated multiservice architecture for next generation services) project, sponsored by the European Union within the CELTIC and the IST programmes, respectively.

C. Bruni is Full Professor at DIS (Informatica e Sistemistica Department " A. Ruberti") of University of Rome "La Sapienza", Italy (e-mail: brunic@dis.uniroma.it).

F. Delli Priscoli is Full Professor at DIS (Informatica e Sistemistica Department " A. Ruberti") of University of Rome "La Sapienza", Italy (e-mail: dellipriscoli@dis.uniroma.it).

G. Koch is in charge of a research contract at DIS (Informatica e Sistemistica Department " A. Ruberti") of University of Rome "La Sapienza", Italy .

I. Marchetti is a PhD. Student at DIS (Informatica e Sistemistica Department " A. Ruberti") of University of Rome "La Sapienza", Italy (e-mail: ilaria.marchetti@dis.uniroma.it).

As detailed in the following, the CAC is a fundamental procedure aiming at maximizing the exploitation of the available capacity of a telecommunication network and, at the same time, assuring the respect of the QoS requirements of the various Service Classes.

Whenever a stand-by user requests a connection set-up, the CAC is in charge of deciding whether to accept or to reject the connection set-up. In the following, for the sake of brevity, a connection whose set-up is requested by a stand-by user will be simply referred to as *new connection*.

becomes *in progress* and remains in this status up to either its natural completion (in this case, the user decides to terminate the connection), or its forced dropping (in this case, as explained below, the CAC decides to terminate the connection); whichever is the reason, whenever a connection is terminated, the corresponding user comes back in the *stand-by* status.

The CAC problem will be formulated as an optimal control problem subject to a set of constraints. As a matter of fact, the proposed controller - modelling the CAC mechanism – is in charge of computing the above-mentioned decision variables so that (i) a set of proper constraints, which model the QoS requirements, are respected and (ii) a proper performance index, which models the exploitation degree of the available capacity, is maximized.

Basing on the above mentioned modelling, an innovative predictive approach is proposed. A forward model for the variables accounting for the QoS assessment, based on their past history, is used and the constraints in (i) as well as the performance index (ii) are expressed as probabilities and expected values over some future time interval.

While clearly all the above mentioned Resource Management tasks are expected to strongly interact with each other, it would however be unfeasible to formulate and to solve them in a unitary time and functional context. Our approach overcomes the above stall by a convenient feedback formulation of the CAC problem which includes the other interacting control problems without requiring their explicit solution.

The paper is organized as follows:

In Section 2 General Reference Architecture of a telecommunication network is provided, which reflects a system theory point of view.

In Section 3 the overall intra-domain resource management problem is formulated as the problem of maximizing a suitable performance index while respecting a proper set of constraints.

Section 4 is devoted to the description of the relationships between the CAC procedure and the other intra-domain Resource Management procedures.

Finally, Section 5 is focused on the CAC problem and suggests a suitable control strategy, following which a solution of the problem can possibly be achieved.

II. General Reference Architecture

The general reference architecture considered in this paper is sketched in Fig. 2. The figure shows, as an instance, four different domains: two Access Network domains (A and D), which could represent two wireless domains, and two Core Network domains (B and C).

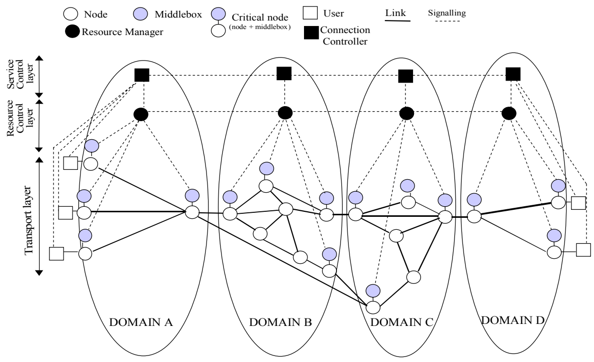

Fig. 2 General Reference Architecture

The reference architecture includes the following network entities:

- The *nodes* (indicated in Fig. 2 with white circles) represent the points of the telecommunication architecture where operations (e.g. switching, scheduling, etc.) impacting on resource management are performed.
- The *links* joining the nodes (indicated in Fig. 2 with continuous thick lines) represent wired or wireless connections between the nodes having fixed or variable capacity.
- The *users* (indicated in Fig. 2 with white square) represent communication sources or destinations; an user is always co-located with a node (namely, the terminal through which the user accesses to the communication network).
- In each domain a *Resource Manager* (indicated in Fig. 2 with a black circle) is present which (i) autonomously performs the intra-domain resource management procedures, (ii) in cooperation with the other Resource Managers (an ad hoc *Signalling* is exchanged among Resource Managers, indicated in Fig. 2 with continuous lines) performs the inter-domain resource management procedures.
- Some nodes, hereinafter referred to as *critical nodes*, are provided with so-called *Middleboxes,* indicated in Fig. 2 with grey circles . The middleboxes communicate with the Resource Manager handling the domain in which they are placed through a proper *Signalling* (indicated in Fig. 2 with dashed lines). The middleboxes have the following roles: (i) collecting performance (traffic, delays, losses) measurements at the associated critical nodes, (ii) carrying into operation the decisions of the Resource Manager. Both these roles are performed in a transparent way with respect to the underlying domain technology, i.e. without affecting the usual way of operating of the considered domain.
- In each domain a *Connection Controller* (indicated with a black square) is present which (i) in cooperation with the other Connection Controllers and with the users (an ad hoc *Signalling* is exchanged among Connection Controllers, indicated in Fig. 2 with dashed lines) handles the signalling relevant to connection set-ups and terminations, (ii) at

connection set-up extracts the connection QoS requirements, forwards them to the associated Resource Manager waiting for the acceptance/rejection decision of the Resource Manager (*Signalling* indicated in Fig. 2 with dashed lines).

In light of the above, it should be clear that the Resource Managers are key entities since they perform all the decisions related to the resource management procedures. The Resource Manager decisions are performed on the grounds of (i) performance (traffic, loss, delay) measurements in the considered domain taken by the middleboxes at the critical nodes), (ii) information concerning the occurrence of the connection set-up attempts/connection terminations (iii) information concerning the connection QoS requirements .

It is important noting that the Resource Manager of a considered domain is not necessarily located in a single network physical entity. Conversely, the Resource Manager functions of a considered domain can be distributed in several network physical entities.

It is easy to identify, in the reference architecture shown in Fig. 2, (i) a *transport layer* namely the layer including the users, the nodes and the links which correspond to the technology dependent layers represented in Fig. 1, (ii) a *resource control layer* namely the layer including the middleboxes and the Resource Managers which can be considered as a sublayer of the Convergence Layer represented in Fig. 1 (this last also includes security and mobility functions) and (iii) a *service control layer,* namely the layer including the Connection Controllers which can be considered as a sublayer of the Middleware Layer represented in Fig. 1.

Some basic definitions have to be weighed up to correctly understand the following sections of the paper.

In each considered domain a set of different *Service Classes* is defined. Different domains can be characterized by different sets of Service Classes. Let k denote the generic Service Class in the considered domain and K denote the total number of Service Classes in the same domain. In the considered domain, a given Service Class k is characterized by a set of *QoS parameters* (detailed in the following Section).

Indeed, according to the most recent trends, the QoS control is performed on a *per flow basis* where a flow refers to the packets entering the domain at a given ingoing node i, going out of the domain at a given outgoing node j and relevant to in progress connections belonging to a given Service Class k. In the following, by "*flow* (k,i,j)" we mean the flow of packets entering the domain at a given ingoing node i (i =1, 2,…,I), going out of the domain at a given outgoing node j (j =1, 2,.., J) and relevant to in progress connections belonging to a given Service Class k.

III. Intra-Domain QoS Parameters and Performance Index

Here and in the following we will consider a given domain.

We will introduce the following notations:

- $R_{off}(k,i,j,t)$ (*offered bit rate*) denotes the bit rate, relevant to the flow (k,i,j), which, at time t, is offered to the domain;
- $R_{loss}(k,i,j,t)$ (*loss bit rate*) is the part of $R_{off}(k,i,j,t)$ which is unintentionally lost, from the convergence layer point of view, in the run from the ingoing node i up to the outgoing node j. These losses can be caused by technology dependent layer issues such as interference phenomena occurring on the medium supporting a link.
- $R_{disc}(k,i,j,t)$ (*discarded bit rate*) denotes the part of $R_{off}(k,i,j,t)$ which is intentionally discarded by the convergence layer by means of the *congestion control* procedure at any critical node crossed by the flow (k,i,j).
- $M(k,i,j,t)$ denotes the total number of in progress connections in the considered domain, at time t, relevant to the flow (k,i,j).

By considering users and operators requirements, as well as ITU-T (International Telecommunications Union-Telecommunications) recommendations, we have identified the following set of QoS parameters (the selected symbols and the relevant measurement units are indicated):

- $\Pi_b(k)$: Blocking frequency for the Service Class k;
- $\Pi_d(k)$: Dropping frequency for the Service Class k;
- $L_A(k,i,j)$: Link availability relevant to the flow (k,i,j), accounting for:
  - $D(k,i,j,t)$: Transfer delay (sec) of the traffic, relevant to the flow (k,i,j) at time t,
  - $R_{adm}(k,i,j,t)$: Admitted bit rate (bps) of the traffic, relevant to the flow (k,i,j) at time t,
  - $BER(k,i,j,t)$: Fractional evaluation of loss bit rate.

The above-mentioned parameters are quantitatively assessed as follows.

With reference to a given Service Class k, $\Pi_b(k)$ is the ratio between the number of blocked connection attempts and the total number of connection attempts, computed during a suitable time interval (e.g. a busy hour).

With reference to a given Service Class k, $\Pi_d(k)$ is the ratio between the number of dropped connections (i.e. connections which do not terminate in a natural way) and the total number of terminated connections, computed during the selected time interval.

With reference to a given flow (k,i,j), $L_A(k,i,j)$ is the percentage of time, computed during the selected time interval, during which the path from the ingoing node i to the outgoing node j is *available*. Such path is considered available whenever the performance in terms of transfer delay, admitted bit rate, loss bit rate is satisfactory (see below).

$D(k,i,j,t)$ is the delay experienced by the packets which leave the considered domain at the outgoing node j at time t, that is the difference between t and the time they entered the considered domain at the ingoing node i.

$R_{adm}(k,i,j,t)$ is the bit rate of the traffic which is admitted, at time t, at the ingoing node i for being carried up to the outgoing node j. More precisely, it is the difference between the bit rate $R_{off}(k,i,j,t)$ and the bit rate $R_{disc}(k,i,j,t)$ relevant to the flow (k,i,j).

$BER(k,i,j,t)$ is the ratio between the loss bit rate $R_{loss}(k,i,j,t)$ and the admitted bit rate $R_{adm}(k,i,j,t)$.

In each domain, the relevant operator establishes a set of *QoS thresholds* for the above-mentioned QoS parameters. For a given Service Class k, the relevant QoS thresholds are the following (the selected symbol is indicated):

- $\Pi_{b\text{-}max}(k)$: Maximum blocking frequency for the Service Class k,
- $\Pi_{d\text{-}max}(k)$: Maximum dropping frequency for the Service Class k,
- $L_{A\text{-}min}(k)$: Minimum link availability (% of time) for the Service Class k,
  - $D_{max}(k)$: Maximum transfer delay (sec) for the Service Class k,
  - $D_{min}(k)$: Minimum transfer delay (sec) for the Service Class k,
  - $R_{adm\text{-}min}(k,i,j,t)$: Minimum admitted bit rate (bps) relevant to the flow (k,i,j) at time t.
  - $BER_{max}(k)$, Maximum fraction of loss bit rate for the Service Class k.

The rationale of the above-mentioned thresholds is as follows.

The issue of imposing a maximum threshold $\Pi_{b\text{-}max}(k)$ for the blocking frequency for the Service Class k derives from the natural user requirement to avoid too many failures of the connection set-up procedure due to the fact that the network is busy.

The issue of imposing a maximum threshold $\Pi_{d\text{-}max}(k)$ for the dropping frequency for the Service Class k derives from the natural user requirement to avoid too many droppings of the connections in progress.

The issue of imposing a minimum threshold $L_{A\text{-}min}(k)$ for the link availability for the Service Class k derives from the natural user requirement to avail for a sufficiently high percentage of time of a "good" connection quality in terms of delays, guaranteed bit rate, and bit losses. In particular:

- as concerns the delays, thresholds $D_{max}(k)$ and $D_{min}(k)$ are imposed on the transfer delay in order to meet the natural user/application requirement of avoiding too long waiting times and in order to avoid too high transfer delay ranges (i.e. a too high delay jitter) which could cause problems in the dimensioning of the buffers at the receiving user side;
- as concerns the guaranteed bit rate, a proper threshold function $R_{adm\text{-}min}(k,i,j,t)$ is imposed in order to meet the natural user requirement to avail of a minimum guaranteed bit rate. Such function should be properly selected in order to account for the number of connections in progress and the burstiness of the offered traffic.
- as concerns the bit losses, a thresholds $BER_{max}(k)$ is imposed in order to meet the natural user requirement of avoiding too many losses in the transmitted traffic.

In view of the above, the following QoS constraints must be satisfied:

$$\Pi_b(k) \le \Pi_{b\text{-}max}(k)\ \forall\ k \qquad (3.1a)$$

$$\Pi_d(k) \le \Pi_{d\text{-}max}(k)\ \forall\ k \qquad (3.1b)$$

$$L_A(k,i,j) \ge L_{A\text{-}min}(k)\ \forall\ (k,i,j) \qquad (3.1c)$$

in which the link availability $L_A(k,i,j)$, previously introduced, is now more precisely defined as the percentage of time, computed during the fixed time interval, in which the following inequalities are simultaneously verified:

$$D_{min}(k) \le D(k,i,j,t) \le D_{max}(k) \qquad (3.2a)$$

$$R_{adm}(k,i,j,t) \ge R_{adm\text{-}min}(k,i,j,t) \qquad (3.2b)$$

$$BER(k,i,j,t) \le BER_{max}(k) \qquad (3.2c)$$

The above inequalities are conventionally assumed verified whenever $R_{off}(k,i,j,t)=0$.

The operator handling a given domain is interested in maximizing the weighted throughput carried by its domain. As a matter of fact, in any reasonable billing system, higher throughput means higher revenue. The weights $w(k)$ account for the fact that different Service Classes can be billed under different fares. So, the target of intra-domain resource management is to maximize the following performance index, computed during the selected time interval $[t, t+\Delta]$:

$$J = \sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J}\int_{t}^{t+\Delta} w(k)\ R_{adm}(k,i,j,\tau)d\tau \qquad (3.3)$$

while respecting the constraints (3.1).

IV. Intra-Domain Resource Management Procedures

As above illustrated, the target of intra-domain resource management is to maximize the performance index (3.3), while respecting the QoS constraints (3.1).

In the light of the analysis made in the previous section, it is easy to understand that the solution of resource manager problem is not an effortless task, especially in the technology independent framework; for this reason this problem is generally decomposed into a set of dedicated procedures (Routing, Connection Admission Control, Congestion Control, Dynamic Resource Assignment).

The intra-domain procedures cooperate in order to achieve the above-mentioned target.

At each connection set-up attempt, the intra-domain routing procedure identifies the path, which will support the connection in case the latter is accepted by the CAC procedure.

Basing on this information and on possible performance measurements (see below), at each connection set-up attempt relevant to the flow (k,i,j), the Resource Manager handling the considered domain performs the Intra-Domain CAC procedure in order to decide:

(i) whether to accept or to reject the new connection in the flow (k,i,j)

(ii) the possible forced dropping of one or more in progress connections.

Therefore, the CAC decides about the *blocked connections* and the *forcedly dropped connections*. The mobility (if any) and the connection dynamics, controlled by the CAC algorithm, determine the total number of in progress connections $M(k,i,j,t)$.

These connections generate packets according to certain dynamics (*packet dynamics*) which are outside the Resource Manager control. Nevertheless, the packet traffic flow is controlled by the other Resource Management procedures: Dynamic Resource Assignment, Congestion Control and Scheduling.

The Dynamic Resource Assignment procedures consist of (i) Dynamic Capacity Allocation procedure, (ii) Delay Range Assignment procedure, (iii) Minimum Admitted Bit Rate Assignment procedure and (iv) Maximum Fraction of Loss Bit Rate Assignment procedure.

- The first procedure is in charge of rearranging the capacity of the links (in the domains in which the links have not a fixed capacity) aiming to move capacity from the idle links to the congested ones.
- The Delay Range Assignment procedure is in charge of assigning to each link (connecting two generic nodes m and n) and for every Service Class k a *link minimum* $D_{min}(k,m,n,t)$ and a *link maximum* $D_{max}(k,m,n,t)$ *transfer delay*, namely the minimum and maximum transfer delays which a packet relevant to the Service Class k can experience from the time at which it enter the node m to the time t at which it arrives at the node n; so this delay, includes the possible queuing delay at the node m and the propagation delay of the link (m,n).
- The Minimum Admitted Bit Rate Assignment procedure is in charge of allocating to each link (connecting two generic nodes m and n) and for every Service Class k a so called *Minimum Admitted Bit Rate function* $R_{adm\text{-}min}(k,m,n,t)$, namely the minimum bit rate which, at time t, is guaranteed for the connections belonging to the Service Class k over the considered link.
- Finally, the Maximum Fraction of Loss Bit Rate Assignment procedure is in charge of assigning to each link (connecting two generic nodes m and n) and for every Service Class k a *maximum link fraction of loss bit rate* $BER_{max}(k,m,n,t)$, namely the maximum fraction of loss bit rate which the packets relevant to the Service Class k can experience over the link.

This means that the congestion control and the scheduling procedures (see below) must work respecting, at each time, the values fixed by the four dynamic resource assignment procedures (Capacity, $D_{min}(k,m,n,t)$, $D_{max}(k,m,n,t)$, $R_{adm\text{-}min}(k,m,n,t)$, $BER_{max}(k,m,n,t)$).

In particular, congestion control procedure decides the portion of the offered traffic $R_{off}(k,i,j,t)$ which can be actually admitted $R_{adm}(k,i,j,t)$ and the traffic which has to be discarded $R_{disc}(k,i,j,t)$.

At any critical node, the traffic admitted by the Congestion Control may be scheduled in order to decide (for every link that enters or goes out of this node) the priority according to which the packets presently stored in the critical node have to be forwarded over the link.

Note that the Dynamic Resource Assignment, the Congestion Control and the Scheduling procedures determine the actual admitted bit rates, delays and fractions of loss bit rates which, in the proposed CAC approach, are measured by the middleboxes, forwarded to the Resource Manager and used as key inputs by the CAC procedure; this last being in charge of aiming at maximizing the cost index (3.3), while respecting the QoS constraints (3.3). From this discussion, the key role played by the CAC procedure is evident; for this reason, in the following of this paper, we will focus our interest on this procedure.

V. Intra-Domain Connection Admission Control (CAC) Procedure

A. Basic Design Concepts

As pointed out in the previous sections, the target of the CAC procedure is to contribute to maximize the performance index (3.3), provided that the QoS constraints (3.1) are met.

In this respect, note that an accepted connection will offer additional traffic which possibly can be admitted into the domain, thus eventually increasing the performance index (3.3) and favouring the satisfaction of the blocking frequency constraint (3.1.a). Nevertheless, to accept new connections entails a worsening in the link availability, thus causing the possible violation of the Link Availability constraint (3.1c)-(3.2) and consequently jeopardizing the satisfaction of the dropping frequency constraint (3.1.b).

In addition, in the perspective of maximizing the performance index (3.3), connection acceptance/rejection at a given time t can be performed by considering what is expected to happen at times next to t. So, for instance, a connection can be blocked at a time t in order to assure the acceptance (without violating the Link Availability constraint) of another connection which is expected to be set up at a time next to t with a higher weight, and/or is supposed to generate a higher amount of traffic.

It should be noted that, as the number of connections in progress in the considered domain increases, the necessity to still guarantee the respect of the Link Availability constraint implies the increase of the number of connection attempt blocks and/or connection forced droppings, up to a limit at which the constraints (3.1a) and/or (3.1b) are no longer met: this limit can be somehow regarded as the domain capacity limit.

The proposed CAC approach is based upon the following four innovative ideas:

1. i. the CAC should be technology independent;
2. ii. the CAC should be decoupled from the other Resource Management procedures;
3. iii. the CAC behaviour should not just account for the instantaneous situation, but rather forecast the domain dynamics in the next future;
4. iv. the CAC should be implemented as a feedback control on the basis of a convenient optimality criterion.

I. As far as issue (i) is concerned, we stress the fact that the proposed formulation for the QoS parameters on the one hand features a precise operative definition, being on the other hand independent of any specific technology (i.e. the proposed approach applies to any variety of wireless and wired systems such as UMTS, wireless, LAN's, xDSL, ad hoc networks,…).

II. As far as issue (ii) is concened, it is important to point out a difference between the constraints widely described in Section 3. Indeed, while all the resource manager procedures cooperate in order to guarantee the QoS requirements, the

CAC is the only one in charge of guaranteeing the constraints (3.1a) and (3.1b). Conversely, for the satisfaction of constraint (3.1c), the CAC has to work together with the other procedures (congestion control, scheduling, dynamic resource assignment), without knowing their modus operandi and the control criteria adopted by them.

Starting from this consideration, the authors suggest to disjoin the CAC procedure from the other ones, by providing a new formulation for the Resource Management problem in which, as far as the CAC is concerned, the constraint (3.1c) is replaced by a different "nominal" constraint, while the satisfaction of the actual Link Availability constraint is demanded to the other resource management procedures. This new constraint is formulated with the purpose of guaranteeing that, when it is satisfied, the other procedures are certainly capable to find at least one solution for the Resource Management problem.

The proposed CAC approach is based on the following principle: any flow has the "right" to feed the domain with a traffic equal to the minimum nominal guaranteed bit rate, denoted by $Rguar\text{-}min(k,i,j,t)$, with $Rguar\text{-}min(k,i,j,t) \geq Radm\text{-}min(k,i,j,t)$. A possible instance of such function is:

$$Rguar\text{-}min(k,i,j,t) = M(k,i,j,t)\, Rguar\text{-}min(k)$$

where $Rguar\text{-}min(k)$ is a minimum nominal guaranteed bit rate for each connection belonging to the Service Class k.

In the light of the above, the proposed CAC will strive to maximize the performance index (3.3), by admitting a new connection, or dropping a connection in progress, at time t, taking into account constraints (3.1a), (3.1b) and the nominal Link Availability principle. This latter can be stated as follows:

"If the domain is fed with the minimum guaranteed bit rates relevant to all the connections in progress in the domain at time t, then the Link Availability constraint (3.1c) *is respected".*

As far as the nominal Link Availability constraint is concerned, the CAC behaviour is not independent of the real traffic situation, nor of the behaviour of the other Resource Management procedures. On the contrary, the CAC task will be to monitor the performance of the actual traffic as it result from the offered traffic and from the processing of the offered traffic itself by the other Resource Management procedures. This task is performed in order to figure out whether, in case the traffic assumes the level fixed by the $Rguar\text{-}min(k,i,j,t)$, the Link Availability constraint (3.1c) is still fulfilled.

Under the above mentioned behaviour of the CAC, the other procedures will strive to maximize (3.3) to the best they can, with the constraint (3.1c) and the awareness that, just because of the working framework established by the CAC, they will always be able to find an admissible solution (at the worst, they will behave as to fix $Radm(k,i,j,t)$ to $Rguar\text{-}min(k,i,j,t)$).

Let $BER^{(R_{\text{guar-min}})}(k,i,j,t)$ and $D^{(R_{\text{guar-min}})}(k,i,j,t)$ respectively denote the percentage loss bit rates and the delays experienced by the various flows at time t in case the domain is fed with the minimum guaranteed bit rates $Rguar\text{-}min\ (k,i,j,t)$ for all the connections presently in progress.

Considering (3.2), the Link Availability $L_A^{(R_{\text{guar-min}})}(k)$ relevant to the Service Class k corresponding to the minimum guaranteed bit rates relevant to all the connections in progress in the domain is the percentage of time, computed during the selected time interval, in which the following conditions are met:

$$Dmin(k) \leq D^{(R_{\text{guar-min}})}(k,i,j,t) \leq Dmax(k) \quad \forall\ (i,j) \qquad (5.1a)$$

$$BER^{(R_{\text{guar-min}})}(k,i,j,t) \leq BERmax(k) \quad \forall\ (i,j) \qquad (5.1b)$$

Note that the constraint (3.2b) is always satisfied, due to the very definition of $Rguar\text{-}min(k,i,j,t)$.

The proposed CAC approach is based on the idea of admitting a new connection at time t, *if and only if the constraints* (3.1a), (3.1b) *and the following nominal Link Availability constraint* (5.2) (mathematical formulation of the nominal Link Availability principle):

$$L_A^{(R_{\text{guar-min}})}(k) \geq LA\text{-}min(k) \quad \forall\ k \qquad (5.2)$$

are respected during a suitable future time interval $[t, t']$ (*with* $t'>t$).

III. As far as issue (iii) is concerned, we note that as a matter of fact the quantities $BER^{(R_{\text{guar-min}})}(k,i,j,t)$ and $D^{(R_{\text{guar-min}})}(k,i,j,t)$ are not available. Thus the CAC should include a suitable algorithm for their estimation by exploiting available data. The estimates will be denoted by $B\hat{E}R^{(R_{guar-\min})}(k,i,j,t)$ and $\hat{D}^{(R_{aguar-\min})}(k,i,j,t)$ respectively.

Furthermore, the CAC will provide to satisfy the QoS constraints in a probabilistic sense over $[t, t']$; that is, for each t, it will guarantee that constraints (3.1a), (3.1b), (5.2) will be satisfied over (t, t') with a sufficiently high probability. This in turn will require a proper stochastic modelling for the behaviour of the domain and suitable prediction algorithms.

V. As far as issue (iv) is concerned, the CAC we are going to propose processes available data related to the actual behaviour of the domain.

As well known, this entails the introduction of a feedback loop with the classical advantages it implies (robustness, insensitivity to noise disturbances and/or parameter variations …).

Moreover, among all possible solutions, the CAC algorithm will be addressed toward the one which optimizes a suitable performance index, once the other Resource Management procedures are given.

B. Proposed CAC structure

The operations that characterize the proposed CAC approach, described in the previous sub-section, are represented in the block diagram of Fig. 3.

Note that our CAC approach requires, in addition to the quantities introduced in Section 3, the following information:

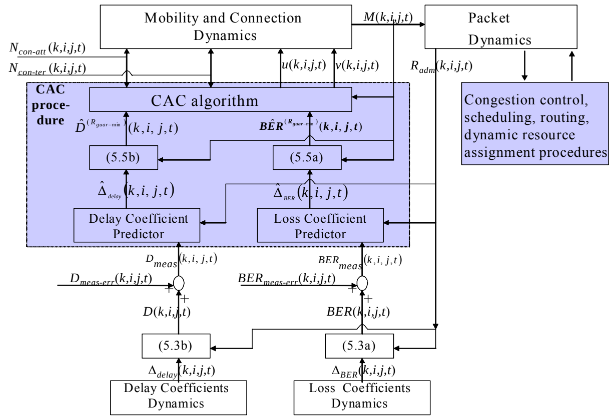

Fig. 3 Proposed CAC procedure approach

- *Number of attempted connection set-ups per flow Ncon-att*(*k*,*i*,*j*,*t*) occurred at time *t*.
- *Number of connection terminations per flow Ncon-ter*(*k*,*i*,*j*,*t*) occurred at time *t*.

Fig. 3 shows that the CAC algorithm is the core of the CAC procedure performed by the Resource Manager. Such CAC algorithm is based on (i) the estimates $\hat{BER}^{(R_{guar-\min})}$ (*k*,*i*,*j*,*t*) and $\hat{D}^{(R_{guar-\min})}$ (*k*,*i*,*j*,*t*), (ii) the numbers of the connection attempted set-ups and connection terminations (i.e. *Ncon-att*(*k*,*i*,*j*,*t*) and *Ncon-ter*(*k*,*i*,*j*,*t*)), (iii) the number of in progress connections *M*(*k*,*i*,*j*,*t*).

The CAC algorithm, basing on the above-mentioned information, computes the two following sets of control variables:

- *u*(*k*,*i*,*j*,*t*): the (binary) *acceptance* control for the flow (*k*,*i*,*j*) at time *t*. If *u*(*k*,*i*,*j*,*t*)=1 a new connection set-up attempt occurring at time *t* and relevant to the flow (*k*,*i*,*j*) is accepted; if *u*(*k*,*i*,*j*,*t*)=0 such a connection set-up is blocked.
- *v*(*k*,*i*,*j*,*t*): the *dropping* control for the flow (*k*,*i*,*j*) at time *t*. This control variable denotes the number of connections, relevant to the flow (*k*,*i*,*j*), which has to be forcedly dropped at time *t*. The dropping control *v*(*k*,*i*,*j*,*t*) is a non negative integer not exceeding *M*(k,i,j,t).

Due to the control actions of the other procedures, the admitted bit rates Radm(k,i,j,t) affect the actual losses and delays experienced in the domain. We here assume instantaneous (possibly non linear) relationships of the form:

$$BER\,(k,i,j,t) = \Delta BER\,(k,i,j,t)\ \psi BER\,[Radm(1,i,j,t), Radm(2,i,j,t), \ldots, Radm(K,i,j,t)] \quad (5.3a)$$

$$D\,(k,i,j,t) = \Delta delay(k,i,j,t)\ \psi Delay\,[Radm(1,i,j,t), Radm(2,i,j,t), \ldots, Radm(K,i,j,t)] \quad (5.3b)$$

where Δ*BER*(*k*,*i*,*j*,*t*) and Δ*delay*(*k*,*i*,*j*,*t*) are respectively the loss and delay coefficients of the flow (*k*,*i*,*j*) at time *t* in the considered domain, and ψ*BER* and ψ*Delay* are suitably selected functions.

As stated above, the losses and delays experienced in the considered domain are measured by the Middleboxes placed at the edge nodes; these measurements are periodically reported from the Middleboxes to the Resource Manager which performs the CAC procedure. So, these measurements are subject to a measurement error *BERmeas-err*(*k*,*i*,*j*,*t*) and *Dmeas-err*(*k*,*i*,*j*,*t*), respectively. Hence, the measured losses and delays (indicated as BER_{meas} (*k*,*i*,*j*,*t*) and D_{meas} (*k*,*i*,*j*,*t*), respectively) are given by the following relationships:

$$BERmeas(k,i,j,t) = BER(k,i,j,t) + BERmeas\text{-}err(k,i,j,t) \quad (5.4a)$$

$$Dmeas(k,i,j,t) = D(k,i,j,t) + Dmeas\text{-}err(k,i,j,t) \quad (5.4b)$$

The above-mentioned measurements, along with the knowledge of the admitted bit rates Radm(k,i,j,t), may be used by the Resource Manager in order to deduce, through a proper prediction process, the estimates of the loss and delay coefficients (denoted by $\hat{\Delta}_{BER}$ (k,i,j,t) and $\hat{\Delta}_{delay}$ (k,i,j,t), respectively). These estimates, in turn, may be used in order to compute $\hat{BER}^{(R_{guar-\min})}$ (k,i,j,t) and $\hat{D}^{(R_{guar-\min})}$ (k,i,j,t), as:

$$\hat{BER}^{(R_{guar-\min})}(k,i,j,t) = \hat{\Delta}_{BER}\,(k,i,j,t)\psi BER[Rguar\text{-}min(l,i,j,t), Rguar\text{-}min(2,i,j,t), \ldots, Rguar\text{-}min(K,i,j,t)] \quad (5.5a)$$

$$\hat{D}^{(R_{guar-\min})}(k,i,j,t) = \hat{\Delta}_{delay}\,(k,i,j,t)\psi BER[Rguar\text{-}min(1,i,j,t), Rguar\text{-}min(2,i,j,t), \ldots, Rguar\text{-}min(K,i,j,t)] \quad (5.5b)$$

VI. Concluding Remarks

The aim of this work is not (yet) to provide a solution to the CAC problem. Rather, we felt necessary to investigate the basic functional relationships among the various components of a telecommunication network, and to deduce the essential tasks that each of them has to face.This investigation is a qualifying unavoidable starting point to arrive at a formulation of the CAC problem, along with an approach to its solution, which differently from other pertinent works appeared in the literature ([5][6][7][8]), features the innovative ideas stressed in Section 5.1.

Future research work will be concerned with the definition of the general CAC procedure as designed in Fig. 3.

With reference to that, we stress that the actual content of the blocks denoted by Packet Dynamics and the other (non CAC) Resource Management procedures (Congestion Control, Scheduling, Routing, Dynamic Resource Assignment) does not need to be detailed, since our CAC procedure is decoupled from them, taking at the same time their behaviour into account.

The Mobility and Connection Dynamics may be filled with any convenient connection and mobility model. In that respect, our reference starting point (to be further developed and/or enriched) is the model discussed in [9], [10].

As far as Loss and Delay Coefficient Dynamics, as well as their predictors, guidelines are already provided by well established theories in stochastic modelling and prediction for (linear) dynamic systems.

The block which deserves most attention and investigation in future research work is the CAC algorithm. This will have to provide control $u(k,i,j,t)$ and $v(k,i,j,t)$ as solutions of an integer-valued optimal control problem, with suitable constraints corresponding to the QoS requirements.

REFERENCES

[1] L. Munoz, M. Garcìa, P. Mahonen, Z. Selby, D. Melpignano, G. Orphanos, "*Wireless IP based on WAL concept: WINE*", Multiradio Multimedia Communications, Berlin, November 2000.

[2] Olivier Dugeon, Donal Morris, Edmundo Monteiro, Wojciech Burakowski, Michel Diaz, "*End to End Quality of Service over Heterogeneous Networks* (*EuQoS*)", Proceedings of the Conference on Network Control and Engineering for QoS, Security and Mobility Net.

[3] L. Becchetti, F. Delli Priscoli, P. Mahonen, L. Munoz, T. Inzerilli, "Enhancing IP Service Provision over Heterogeneous Wireless Networks: a Path Towards 4G", *IEEE Communications Magazine, Special Issue on "Life after Third Generation Mobile Communications", IEEE Communications Society (U.S.A.), Vol. 39 No. 8, August 2001, pp. 74-81.*

[4] F. Delli Priscoli, A. Vernucci, "ATA: A Novel Call Admission Control Approach for 3G Mobile Networks based on W-CDMA", *Proc. of the 5th IST Mobile Communication Summit 2001,* Lione (France), September 2004.

[5] Z.Liu, M. El Zarki, "SIR-Based Call Admission Control for DS-CDMA Cellular Systems", IEEE Journal on Selected Areas in Communications, Vol. 12, No. 4, May 1994.

[6] F.Y. Li, N. Stol "A Priority-oriented Call Admission Control Paradigm with QoS Re-negotiation for Multimedia Services in UMTS" Proc. IEEE Vehicular Technology Conference 2001, pp 2021-2025.

[7] J.Sanchez, J.Perez-Romero, O.Sallent, R.Agusti *"Mixing Conversational and Interactive Traffic in the UMTS Radio Access Network"* Mobile and Wireless Communications Network, 2002. 4th International Workshop on , 2002, Page(s): 597 –601.

[8] W. Burakowski, M. Diaz, O. Dugeon, A. Pietrabissa, F. Racaru, G. Santoro, H. Tarasiuk, "*On Multi-Domain Connection admission Control in the EuQoS System*", IST Summit 2006 Mykonos.

[9] C. Bruni, F. Delli Priscoli, G. Koch, I. Marchetti, "*An Optimal Approach to the Connection Admission Control Problem*", International Journal of Control, vol.79, No 10, October 2006, 1237-1250.

[10] C. Bruni, F. Delli Priscoli, G. Koch, I. Marchetti, "*Connection Admission Control in Cellular Networks: a Discrete Time Optimal Solution*", INFOCOM'06, Barcelona, April, 23-29,2006.