Published in *Proceedings of the 16th AFCEA Europe Symposium,* AFCEA (Association for Communications, Electronics, Intelligence & Information Systems Professionals) / IEEE COMSOC / IEE, October, 18-20, Bruxels, Belgium, pp. 58-68, 1995.

Specification, Quantification and Provision of Quality of Service and Congestion Control for New Communication Services

Edmundo Monteiro*, Fernando Boavida*, Gonçalo Quadros*, and Vasco Freitas**

*Universidade de Coimbra
Departamento de Engenharia Informática
Pólo II - Pinhal de Marrocos
3030 Coimbra
Portugal
Tel: +351 39 7000000
Fax: +351 39 701266
e-mail: edmundo@dei.uc.pt

**Universidade do Minho
Departamento de Informática
Largo do Paço
4709 Braga Codex
Portugal
Tel: +351 53 604465
Fax: +351 53 612954

Abstract

Reservationless computer communication networks based on the best effort paradigm are no longer capable of providing the quality of service necessary for the new generation of computer communication services, such as voice, video and multimedia services. These new services and their users have very different communication needs, and must not be treated equal by the communication system.

Quality of service (QoS) specification on a per communication service basis, and quality of service guarantee by reservation-oriented communication systems are the key solutions for the support of new age information systems. QoS specification requires service characterization based on a set of QoS parameters that, on one hand, is rich enough to accommodate service and traffic dynamics and, on the other hand, is simple and flexible enough to allow the implementation of an efficient reservation policy. QoS guarantee by reservation-oriented communication systems can only be achieved by the deployment of new congestion control functions that can prevent service degradation and - based on a quantification scheme - be used to evaluate the network in terms of QoS performance and available capability.

Traditionally, congestion control functions were focused on the communication system itself. It is now evident that new congestion control functions must be focused on communication services, as the network may be in a congestion level that is perfectly adequate for a broad range of services and totally inadequate for other services. Thus, the congestion must be evaluated from the user (or service) point of view, in addition to the network point of view.

In this paper, the authors propose a Quality of Service Framework that provides the basis for QoS specification, quantification and guarantee, and thus constitutes a fundamental tool to the definition and establishment of new generation computer communication services. Firstly, a new congestion definition is proposed, focusing the effects of the communication system on the behavior of communication services. In line with the proposed definition, a model for QoS specification and characterization is presented. Secondly, a metric for the continuous evaluation of the congestion level of individual communication services is given. The metric is based on the deviation of traffic parameters from the values specified during the QoS negotiation phase. From individual service evaluation, an algorithm for the calculation of partial and global communication system congestion is then proposed. Finally, a multidimensional framework for congestion control characterization is presented, in which each dimension identifies a different view - or plane - of the problem.

1. CONGESTION CHARACTERIZATION

1.1 Congestion definition

A globally non-congested communication system may behave as a congested one for one or more of its users. Thus, there is the need for a congestion definition that is able to cope with congestion from a global as well as from an individual point of view. The following types of congestion definitions are commonly found in the literature:

- *"a communication system is congested if the transit delay is greater than X"* [Jain 89];
- "*a communication system is congested if the effective throughput is less than Y*" [Jacobson 88];
- a combination of the above definitions [Jain 84, Ramakrishnan 88].

These types of definitions are not precise, as they evaluate congestion on the basis of one or more of its effects (increase of transit delay and/or decrease of throughput) without taking into account the main cause of congestion: the *load increase*. In addition, the determination of the congestion thresholds X and Y is subjective and/or difficult, which results in the inability to quantify the congestion degree of the communication system. Another limitation lies in the fact that congestion is evaluated from a global perspective, without concern for individual applications. Thus, it is perfectly possible that in a congested system (according to one or more of the above presented definitions) certain applications may be able to perform within the throughput and delay limits that suite their needs.

In [Keshav 91] a congestion definition is proposed that overcomes some of the above mentioned limitations. According to this author, a communication system is under congestion (from a user point of view) if the system *usefulness* decreases due to an increase of system load. The concept of *usefulness* expresses the user preference by the communication resources, through a *usefulness function*. As an example, the author presents the usefulness function $\alpha T - (1-\alpha)\cdot RTT$, where α is a constant, T is the average throughput and RTT is the round trip time. This function enables the users to express their preference by the throughput factor $(\alpha \to 1)$ or by the delay factor $(\alpha \to 0)$. The Keshav definition evaluates congestion from the users point of view and characterizes congestion on a cause basis (as opposed to an evaluation base on the effects of congestion). Nevertheless, it does not support the quantification of the congestion situation, as the usefulness functions are user specific and, thus, usefulness values cannot be compared - in order to measure relative congestion among users - nor combined - in order to measure the global congestion of the communication system.

Normally, communication systems and applications are dimensioned in such a way that the traffic alterations caused by the physical and technical limitations are within the limits tolerated by the applications and do not prevent their normal functioning. When the traffic characteristics are modified in a degree that decreases the performance of a given distributed application, the application is said to be affected by congestion. Thus, the congestion phenomenon can be characterized, from a communication service user point of view, by the following definition:

Definition 1 — Congestion of a communication system:

A communication system is congested whenever the functioning of communication services is affected in a way adversely percepted by their users.

This definition emphasizes the communication services users perspective and accommodates all the factors that may cause the refusal, interruption, or degradation of the communication services, in addition to the load increase factor. In fact, this factor independence is consistent with the user point of view, to whom service degradation is the only effect he/she is concerned with, regardless of the factors that cause it.

The proposed definition has - when compared to the previous definitions based on throughput variation and/or transit delay - the advantage of characterizing congestion from a *microscopic* point of view - for each of the supported services, and at each instant of time - as opposed to a *macroscopic* characterization based on the global throughput and transit delay. In relation to the Srinivasan Keshav definition, the proposed definition has the advantage of enabling a finer congestion evaluation, as it potentially includes all the congestion factors in its scope, and not only the effects on the expected *usefulness*.

1.2 Congestion metric

From a user perspective, a communication system without congestion is characterized by the ability to provide the *quality of service* (QoS) required by the active applications. Quality of service can be objectively defined by a set of operating parameters or, implicitly, by a set of values that are considered "normal" when the application is active. On networks that are based on the resource reservation paradigm - normally operating in the connection-mode - it is always possible to objectively establish a set of quality of service parameters, because these are the parameters that are necessary for the resource reservation at the time of connection establishment. On networks that are based on the best effort paradigm - normally operating in the connectionless-mode - there is no need to explicitly define the QoS parameters; nevertheless, they can be implicitly deducted from the evaluation of the behavior of applications.

In either case - best effort or resource reservation - it is always possible to determine a set of parameters - the QoS parameters - that are responsible for the characterization of the quality of service of the supported applications. Let P_{QoS}, defined in Expression 1, be the set of all of the QoS parameters supported by a given communication system. The description of the physical significance of each parameter and the identification of its respective units is considered to be associated with the definition of the P_{QoS} set.

$$P_{QoS} = \{q_1, q_2, q_3, \ldots, q_n\} \tag{1}$$

For each of the supported services s_i, and at discrete time instants t_k (or continuously in time) it is possible to measure or compute a set of values (one for each parameter belonging to the P_{QoS} set) that can be stored in a vector, as shown in Expression 2.

$$V_{QoS}(s_i)_{t_k} = \begin{bmatrix} v_1 & v_2 & v_3 & \ldots & v_n \end{bmatrix} \tag{2}$$

The specification of the quality of service necessary for each user application may be done, for each parameter, by the specification of an interval in which the parameter values imply no QoS degradation, and the specification of one lower threshold and one upper threshold beyond which the quality of service is unacceptable. The set of interval limits and degradation thresholds, specified for a given service s_i, may take the form of the matrix presented in Expression 3 - the QoS matrix, M_{QoS} - in which m_j and M_j are, respectively, the minimum and maximum values that parameter q_j may take without QoS degradation, and l_{m_j} and l_{M_j} are the thresholds that subtracted from m_j and added to M_j, respectively, define two operating zones with degraded - but still acceptable - quality of service.

$$M_{QoS}(s_i) = \begin{bmatrix} m_1 & l_{m_1} & M_1 & l_{M_1} \\ m_2 & l_{m_2} & M_2 & l_{M_2} \\ m_3 & l_{m_3} & M_3 & l_{M_3} \\ \vdots & \vdots & \vdots & \vdots \\ m_n & l_{m_n} & M_n & l_{M_n} \end{bmatrix} \tag{3}$$

The M_{QoS} elements m_j, M_j, l_{m_j} and l_{M_j} may define static intervals, in the case of services with deterministic QoS needs, or probability intervals, in the case of services with probabilistic QoS needs. At the limit, M_{QoS} elements may be random variables, with associated probability distribution functions, in order to deal with QoS parameters of stochastic nature.

The QoS matrix expresses the *service contract* - whether explicitly negotiated or implicitly assumed - between the user and the service provider, for a given communication service. In the QoS matrix of a specific service only the values (limits and thresholds) for the QoS parameters that are of concern for that service are specified. This parameter set is, normally, a subset of the system QoS parameters. For the parameters that are of no concern for a given service, the QoS matrix will hold the values $-\infty$ and/or $+\infty$. This is also the case for the parameters for which it is only required to specify the upper or the lower limit and threshold.

The congestion state of a communication service can be qualitatively evaluated at each instant of time - in light of the Definition 1 presented in the previous section - by the deviation of the parameter values with respect to the limits defined in the respective service QoS matrix. Consider the following definition:

Definition 2 — Deviation index (Id):

Let q_j be a QoS parameter for the communication service s_i, $v_j(t_k)$ its value at instant t_k, m_j and M_j its normal variation limits, and l_{m_j} and l_{M_j} its QoS degradation thresholds.

Then the QoS parameter deviation index, Id, is given by:

$$Id_{s_i,q_j}(t_k) = \begin{cases} 0 & \text{for} \quad m_j \le v_j \le M_j \\ 1 - 10^{-\left(\frac{m_j - v_j(t_k)}{l_{m_j}}\right)} & \text{for} \quad v_j < m_j \\ 1 - 10^{-\left(\frac{v_j(t_k) - M_j}{l_{M_j}}\right)} & \text{for} \quad v_j > M_j \end{cases}$$

The *deviation index* (Id), proposed in the above definition, enables the quantification of the divergence of the value of a QoS parameter in relation to the limits and thresholds defined in M_{QoS}, in any point of the communication system. It can take all the values between 0 (null deviation) and 1 (maximum deviation), being 0 for all the parameter values inside the interval defined by the m_j *and* M_j (normal variation limits). Outside this interval, the value of Id is a function of the divergence in relation to the normal variation interval and of the QoS degradation thresholds.

The deviation index identifies five different operating zones, as illustrated in Figure 1. If the QoS parameter value is between m and M, the Id index is 0, and this identifies the normal operating zone. The intervals $[m - l_m, m]$ and $[M, M + l_M]$ correspond to two operating zones with QoS degradation, in which the Id index rises up to 90% of its maximum value. For Id

values greater than 90% the parameter values are such that the quality of service is unacceptable. Figure 1 shows a particular situation in which the QoS parameter has symmetric, and finite, variation intervals.

The worsening of the deviation index for a QoS parameter - which corresponds to a quality of service degradation - is a sign of congestion, in light of Definition 1. Congestion of a given communication service may be quantified using the respective QoS parameter deviation indexes, evaluated at the output of the communication system, immediately before the receiving communication services. Nevertheless, we must take into account the possibility of the incoming traffic having an Id not equal to zero (traffic not conformant with the service contract expressed by the QoS matrix). In this case, a non null Id at the system output is not exclusively due to congestion in this communication system.

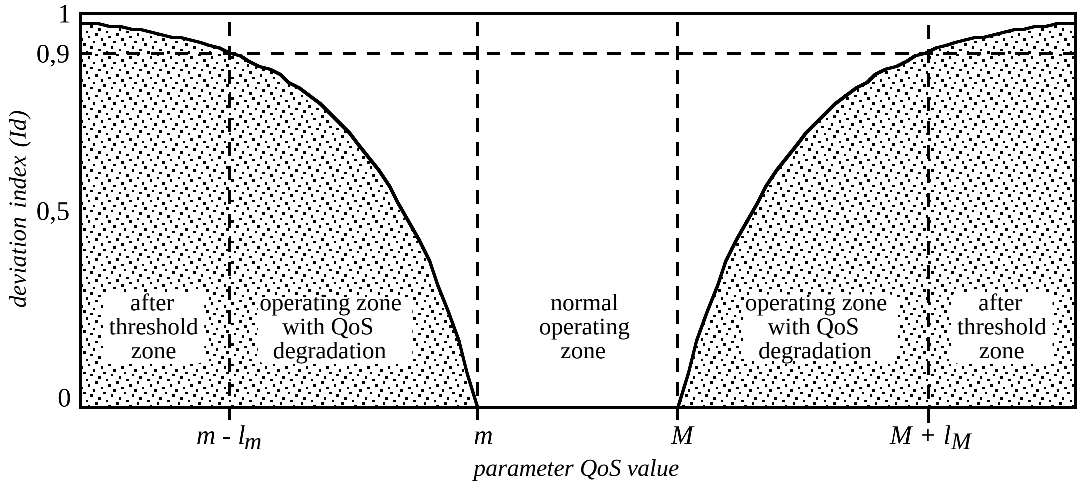

Figure 1 — QoS parameter deviation index.

In order to correctly evaluate the congestion introduced in a service by a communication system, it is necessary to relate the QoS parameter deviation indexes at the input and at the output of that communication system. Consider the following definition:

Definition 3 — Congestion index (Ic):

Let q_j be a QoS parameter for the communication service s_i and, $Id^{in}_{s_i,q_j}$ and $Id^{out}_{s_i,q_j}$ its deviation indexes at the communication system input and output, respectively, at instant t_k.

Then the QoS parameter congestion index, (Ic), is given by:

$$Ic_{s_i,q_j}(t_k) = Id^{out}_{s_i,q_j}(t_k) - Id^{in}_{s_i,q_j}(t_k)$$

The above defined *congestion index* measures the congestion introduced by the communication system in a given QoS parameter of a given service, and it can take values between -1 and 1. When the input traffic is in conformity with the normal variation limits, congestion is simply the value of the QoS parameter Id index at the system output. When the input traffic violates the service contract and the parameter deviation index is worsened in the communication system, congestion is evaluated by the difference between the output and input deviation indexes. Finally, whenever the deviation index is reduced by the communication system, congestion is negative. Negative congestion situations correspond to an active participation of the communication system in the improvement of the QoS characteristics.

Using the congestion indexes (Ic) of the set of QoS parameters of a given communication service s_i, congestion can be evaluated, at instant t_k, and from the user perspective, by the *service congestion index* — $C_{s_i}(t_k)$, — which is given by the weighted average of all the QoS parameter congestion indexes, as shown in Expression 4. The c_j constants are defined by the system management, for each service, and are used to weight the QoS parameters according to their relative importance for the communication service.

$$C_{s_i}(t_k) = \frac{\sum_{j=1}^{n} Ic_{s_i,q_j} \cdot c_j}{\sum_{j=1}^{n} c_j} \tag{4}$$

The *global system congestion* at instant t_k, $Cg(t_k)$, can be obtained by the average of the service congestion indexes of all of the active services, affected by the *service success rate*, as indicated by Expression 5, where N is the number of active services at instant t_k. The *service success rate* measures the probability of success of a given service, taking into account the number of times the service was aborted - not established by lack of resources or abruptly terminated - (Na), and the number of times it was successfully accomplished, (Ns), since the system activation until instant t_k.

$$Cg(t_k) = \frac{Ns}{Ns + Na} \cdot \frac{1}{N} \sum_{i=1}^{N} C_{s_i}(t_k) \tag{5}$$

Using Expression 5, one can obtain global system congestion averages by integration of the instantaneous values of Cg over a period of time (Expression 6) or by averaging the Cg samples at discrete - and regular - time instants $t_1, t_2, \ldots, t_m$ (Expression 7).

$$\overline{Cg} = \frac{1}{t}\int_0^t Cg(t) \tag{6}$$

$$\overline{Cg} = \frac{1}{m}\sum_{k=1}^{m} Cg(t_k) \tag{7}$$

The expressions presented in this section enable the quantification of the congestion of a particular communication service, the comparison of the congestion degree of two or more services, and the quantification of the global system congestion. They have the benefit of enabling an objective measurement of the congestion state of a service or of the whole communication system, based on the definition of a set of system QoS parameters, of their normal variation limits and of their degradation thresholds.

In the case there is the need to evaluate the congestion caused by a specific communication system component (e.g., an intermediate system) the proposed expressions can still be applied. For that, it is sufficient to consider the deviation indexes at the input (Id^{in}) and output (Id^{out}) of the particular component, as long as the normal variation limits (m and M) and the degradation thresholds (l_m and l_M) are established for that component (instead of being established for the whole system). This fact brings out an interesting issue - that will be addressed in the next section - concerning the congestion variation along the path constituted by the various system components, and the determination of the limits and thresholds for the parameter values in each component that guarantee a given global congestion degree.

1.3 Congestion variation along the communication path

Consider the communication system model presented in Figure 2, composed of *N* consecutive modules. The modules can correspond to a functional block (e.g., a protocol layer), to a whole or part of a physical system (e.g., physical medium, switch, bridge, router, end-system), or to the concatenation of multiple physical systems forming a communication subsystem.

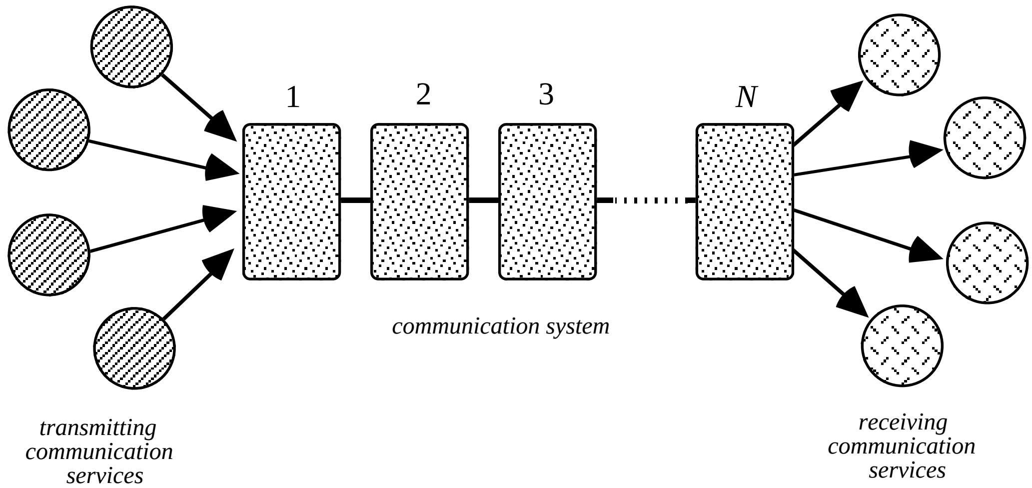

Figure 2 — Model of an *N*-module communication system.

In order to relate the end-to-end congestion with the congestion in each of the communication system modules, it is necessary to identify the influence of each module in the QoS parameters of each communication service.

Knowing the normal variation limits m_j^p and M_j^p, and the degradation thresholds $l_{m_j}^p$ and $l_{M_j}^p$, that characterize each module *p,* for each of the q_j QoS parameters, it is possible to determine - by the use of Definition 2 - the *deviation indexes* at the input and output of the module ($Id^{p_{in}}$ and $Id^{p_{out}}$, respectively). Using these values in Definition 3, one can obtain the *congestion index* of the q_j QoS parameter, concerning service s_i and the module p, for instant t_k (Expression 8).

$$Ic_{s_i,q_j}^p(t_k) = Id_{s_i,q_j}^{p_{out}}(t_k) - Id_{s_i,q_j}^{p_{in}}(t_k) \qquad (8)$$

The influence of the congestion indexes of each communication system module on the *global system congestion* mainly depends on the nature of the QoS parameters. One can identify two types of parameters, with regard to the way in which their variation influences the end-to-end degradation. Some parameters suffer a continuous and progressive degradation, in a way that the resulting end-to-end degradation is the sum (*cumulative*) of the partial degradation along the communication path. Others suffer a discontinuous and abrupt (*non-cumulative*) degradation, with the total (end-to-end) degradation being determined by the greater partial degradation suffered in the modules along the path.

For *cumulative* QoS parameters (i.e., QoS parameters whose end-to-end degradation is determined by the sum of the partial degradations suffered in each communication system module).

$$\forall\, q_j \in \; \{cumulative\;\; QoS\;\; parameters\;\; of\;\; service\;\; s_i\}$$

$$\Rightarrow \qquad M_{QoS}(s_i)_{q_j} = \begin{cases} m_j \leq \sum_{p=1}^{N} m_j^p \\ l_{m_j} \geq \sum_{p=1}^{N} l_{m_j}^p \\ M_j \geq \sum_{p=1}^{N} M_j^p \\ l_{M_j} \geq \sum_{p=1}^{N} l_{M_j}^p \end{cases} \tag{9}$$

$$\Rightarrow \qquad Ic_{s_i,q_j}(t_k) \leq \sum_{p=1}^{N} Ic_{s_i,q_j}^p(t_k) \tag{10}$$

For *non-cumulative* QoS parameters (i.e., QoS parameters whose end-to-end degradation is exclusively determined by the module where the parameter suffers the greater degradation).

$$\forall q_j \in \; \{non-cumulative\;\; QoS\;\; parameters\;\; of\;\; service\;\; s_i\}$$

$$\Rightarrow \qquad M_{QoS}(s_i)_{q_j} = \begin{cases} m = \min\{m_p\} \\ l_m = \max\{l_{m_p}\} \\ M = \max\{M_p\} \\ l_M = \max\{l_{M_p}\} \end{cases} \qquad p = 1,\, 2,\, 3,\, \ldots,\, N \tag{11}$$

$$\Rightarrow \qquad Ic_{s_i,q_j}(t_k) = \max\left\{Ic_{s_i,q_j}^p(t_k)\right\} \tag{12}$$

The results contained in Expressions 10 and 12 are important, as they enable the decomposition of the end-to-end congestion control problem in a set of more restricted congestion problems, each pertaining to a communication system module. Using Expressions 4 to 7, 10 and 12, it is possible to relate the *service congestion indexes* $, C_{s_i}(t_k)$, and the *global system congestion*, $Cg(t_k)$, with the congestion indexes of each of the communication system modules, which makes it possible to determine the influence of a given individual module in the global system congestion, as well as to control the congestion in the system modules knowing the end-to-end constraints imposed by the service.

2. CONGESTION ARCHITECTURAL FRAMEWORK

The complexity associated with the definition of the functions and mechanisms for congestion control requires an analysis based on a multi-dimensional model, in which each dimension - or plane - represents a specific view of the problem. The congestion planes are a result of the identification of the intervening functional elements, followed by their grouping by affinity, having in mind the minimization of the interaction between planes. For each plane, there is the need to identify the functions and mechanisms necessary for their implementation. The identification of the various congestion control planes, complemented by the characterization of their respective functions and mechanisms, including the timescales in which they actuate, constitutes an *architectural framework* for congestion control in communication systems. In this paper we will only briefly present the proposed architectural planes, and identify their main functions and mechanisms.

Plane	Functions	Mechanisms
Services	Service characterization and establishment	Characterization of the P_{QoS} set Characterization of the service QoS matrixes, M_{QoS} Service establishment
	Service monitoring	Algorithm for the evaluation of $I_d(t_k)$ and $I_c(t_k)$ Algorithm for the evaluation of $C_{s_i}(t_k)$ Algorithm for the evaluation of $C_g(t_k)$
	Service control	Traffic shaping
Resources	Resource planning and creation	Resource planning and installation Resource configuration
	Resource control	Resource calculation Resource reservation Resource control
	Resource monitoring	Resource monitoring and accounting
Protocols	Service acceptance control	Service acceptance/refusal Service QoS matrix verification Routing control
	Traffic monitoring	Traffic monitoring Algorithm for the evaluation of $Ic^p_{s_i,q_j}(t_k)$
	Traffic parameter control	Transmission scheduling Traffic policing

Table 1 — Summary of the congestion architectural framework.

2.1 Planes of the congestion control architecture

Congestion control activities can be classified into three main groups: the group concerning the communication services, the group concerning the communication resources, and the group concerning the communication protocols and operation. Each of these groups corresponds to an architectural congestion control plane described below:

• the *service plane* — that groups the activities related to service characterization. This plane contains all the aspects of congestion control that relate to communication services or applications;

• the *resource plane* — that groups the activities related to resource planning, resource reservation, and resource monitoring. This plane contains all the aspects of congestion control that concern the communication resources. The functions of this plane are integrated with or associated to systems management functions because, according to the OSI reference model [ISO 89] resource management is one particular aspect of communication systems management;

• the *protocol plane* — this plane contains all the congestion control aspects that are related to the operation of the communication protocols. Namely, it contains the congestion control activities related to traffic transportation and traffic monitoring. The functions of this plane are integrated with or associated to the communication protocols.

Table 1 presents a summary of the congestion control architecture, that enumerates the functions and mechanisms of each architectural plane.

This multi-dimensional approach to congestion control can also be found in other proposals [Woodruff 88, Ramamurthy 91, Campbell 94] with different planes and functionality. We believe that our framework is more general and best suited to the congestion definition and metric proposed in the first section of this paper.

3. CONCLUSION

This paper presented an architectural framework for the problem of congestion control in communication systems that provide services with a variety of needs in terms of quality of service.

A characterization of the congestion control problem was made, that presented a congestion definition and proposed a metric for its quantitative evaluation. According to the presented definition, congestion is evaluated in terms of its effects on communication services — or applications — that is, in terms of the final objective of all communication systems. The congestion definition is supported by a metric for its quantification and continuous control, based on the deviation of the traffic parameters in relation to the service contract.

This paper also analyzed the congestion variation along the communication path, introduced by the various communication system modules. A set of expressions that relate the end-to-end QoS parameter values with the corresponding QoS parameters of each module was presented. This decomposition enables the control of the QoS parameters on a module-by-module basis, avoiding the need to implement distributed mechanisms with end-to-end scope, with the guarantee of the end-to-end QoS values specified in the service contract.

An architectural framework for congestion control functions was proposed. The complexity of the congestion control problem led to the adoption of a multi-dimensional model in which each dimension represents a particular vision — or plane — of the problem. The framework proposed three congestion control planes: the communication services plane, the communication resources plane, and the communication protocols plane.

The service plane covers all the aspects of congestion control related to the communication services. The resource plane includes the mechanisms related to planning and resource reservation, as well as the mechanisms related to resource monitoring. The protocol plane includes all the congestion control aspects that relate to the operation of the communication system protocol architecture. For each plane, the main functions and mechanisms were identified.

In addition to provide the means for the quantification of the congestion degree of communication services and systems, and to the identification of mechanisms, functions and respective classes, the proposed framework is flexible enough to cover aspects of quality of service other than congestion control, like, for instance, service priority and service security. The large scale use of new communication services and systems requires such a framework, along with the existent protocol and management frameworks of communication systems.

Bibliography

[Campbell 94] Andrew Campbell, Geoff Coulson and David Hutchison, A Quality of Service Architecture, *ACM Computer Communication Review,* Volume 24 (2), pp. 6-27, April 1994.

[ISO 89] International Organization for Standardization, *Information Processing Systems — Reference Model of Open Systems Interconnection — Part 4: Management Framework,* International Standard 7498-4, ISO/IEC, November 1989.

[Jacobson 88] Van Jacobson, Congestion Avoidance and Control, In *Proceedings of the SigComm'88 Conference,* ACM, pp. 314-329, Stanford, EUA, August 1988.

[Jain 84] Raj Jain, D. Chiu and W. Hawe, *A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems*, DEC-TR-301, Digital Equipment Corporation, Littleton, USA, September 1984.

[Jain 89] Raj Jain, A Delay-Based Approach for Congestion Avoidance in Interconnected Heterogeneous Computer Networks, *ACM Computer Communication Review,* Volume 19 (5), pp. 56-71, October 1989.

[Keshav 91] Srinivasan Keshav, *Congestion Control in Computer Networks*, PhD Thesis , Technical Report UCB:CSD-91-649, University of California, Berkeley, USA, December 1991.

[Ramakrishnan 88] K. Ramakrishnan and Raj Jain, A Binary Feedback Scheme for Congestion Avoidance in Computer Networks with a Connectionless Network Layer, In *Proceedings of the SigComm'88 Conference,* ACM, pp. 303-313, Stanford, EUA, August 1988.

[Ramamurthy 91] Gopalakrishnan Ramamurthy and Rajiv Dighe, A Multidimensional Framework for Congestion Control in B-ISDN, *IEEE Journal on Selected Areas in Communications,* Volume 9 (9), pp. 1440-1451, December 1991.

[Woodruff 88] G. Woodruff, R. Rogers and P. Richards, A Congestion Control Framework for High-speed Integrated Packetized Transport, In *Proceedings of the GLOBECOM'88 Conference,* IEEE, 1988.