Calculating Bounds for Delay in Communication Networks Under Real Time Constraints

Claudine Chaouiya
Departamento de Computação da UFC
Campus do Pici, Bloco 910
60021 Fortaleza - CE BRAZIL

Alain Jean-Marie
INRIA-Sophia Antipolis
06902 Sophia-Antipolis Cedex FRANCE

Abstract

We study communication networks submitted to hard real time constraint. This constraint specifies that each customer must leave its arrival station before some fixed deadline. The study of bounds for the workload allows to deduce bounds for the delay of customers and discuss the "feasibility" of the real-time system.

We first analyse the basic properties of a single server with two levels of priority. Then we analyse a feed-forward network without loops. The case of a network of stations with a ring topology is finally presented. For each of these topologies, we compute bounds for the workload of the stations. This calculus is based on the introduction of evolution equations. We also discuss periodicity and feasibility conditions when the arrival process is a superposition of periodic processes.

Resumo

Estudamos redes de comunicação submetidas a condição estrita de tempo real. Esta condição implica que cada cliente deve deixar sua estação de entrada antes de uma certa data limite (*dead-line*). O estudo de limites para a carga permite deduzir limites para o atraso dos clientes e discutir a "factibilidade" do sistema de tempo real.

Primeiro nós analisamos as propriedades básicas de uma estação isolada com uma prioridade de dois níveis. Pois analisamos uma rede *feed-forward* de estações sem *loops*. Enfim, o caso de uma rede em anel está apresentado. Para cada uma dessas topologias, calculamos limites sobre a carga das estações. O cálculo está baseado em equações de evolução. Discutimos tambem de periodicidade e de condições de factibilidade quando o processo de chegada é uma superposição de processos periódicos.

1 Introduction

In this paper, we consider communication networks under a real time constraint, operating in a packet switched mode. In this context, the emerging high-speed networks address the transmission of data, video and audio streams: it is the so-called service integration. The Broadband Integrated Services Digital Networks (BISDN) are projected to support such diversity of traffic. In these networks, the setting of a new connection is based on a guarantee of service quality to the user. The latter has to specify some parameters of his traffic. This problem is really of current interest since neither the nature of the information provided by the user nor the control access criterias are yet well defined.

Our study also applies to local data networks for real-time distributed applications. The control of such distributed applications is based on the communication between the various elements (captors, processors, etc).

Given a set of customers, described by their arrival dates (phase), their service times and deadlines, the problem is to decide whether or not every customer will meet its deadline when served according to a given policy. Because of the real time constraint, we are mainly interested in the determination of conditions for phase independent feasibility. The determination of bounds for the transmission delay in queuing networks has been studied in recent papers for deterministic models [5, 1] and stochastic models [7, 2]. We develop this bound approach for a single server, a feed-forward network and a ring network. For each of these systems, the special case of periodic arrivals is also discussed. We characterize the exit process of the queue as R. Cruz [5, 6] does. However, our formulation by integral equations allows us to derive more general results. In the case of a network, we believe that this formulation can lead to an improvement to the bounds.

Most of the results presented in this paper can be found also in [4].

2 Single server with two levels priority

2.1 Basic properties

In this section, we study the basic properties of a single queue for which the arrival process satisfies a certain "burstiness" constraint. We assume that the service policy has two levels of priority: the types of customers are grouped into two sets called I_1 and I_2 and customers of type in I_2 have a preemptive priority on the customers of type in I_1. Between customers of type in I_1 (respectively in I_2), we need not to specify the priority. In our applications, we shall use these two sets to model the set of local customers and the set of external customers which have a higher priority.

We discuss the stability of the system and give bounds on the workload of customers of type in I_1 and on the length of its activity periods. The special case of periodical arrivals is discussed and we derive from the bounds previously obtained some feasibility conditions.

For each type $i \in I_1$ and for any $0 \le a \le b$, we define $S_i(a,b)$ as the quantity of work of type i arrived in the interval $[a,b)$. We assume that the arrival process, for each type $i \in I_1$, satisfies the following burstiness constraint:

$$\forall i \in I_1, \quad \exists \rho_i, \Lambda_i, \quad \forall 0 \le a \le b, \ S_i(a,b) \le \rho_i(b-a) + \Lambda_i. \tag{1}$$

Thus for the whole quantity of work of type in I_1 arrived in any interval $[a,b)$, namely $S^1(a,b)$, we have also:

$$\forall 0 \le a \le b, \ S^1(a,b) \le \rho^1(b-a) + \Lambda^1,$$

where $\rho^1 = \sum_{i \in I_1} \rho_i$ and $\Lambda = \sum_{i \in I_1} \Lambda_i$.

The arrival process of types in I_2 satisfies a similar "burstiness" constraint:

$$\exists \rho^2, \Lambda^2, \quad \forall 0 \le a \le b, \quad S^2(a,b) \le \rho^2(b-a) + \Lambda^2. \tag{2}$$

Let $W^1(t)$ be the workload of types in I_1 in queue at time t. Recall that a service policy is said to be non-idling (or work-conserving) if it is such that the server works as long as the queue is not empty.

The following lemmas establish the evolution equation of the workload processes and the stability condition for the server.

Lemma 2.1 *For all non-idling service policy, W^1 is characterized by the functional equation:*

$$\forall 0 \le a \le b, \qquad W^1(b) = W^1(a) + S^1(a,b) - \int_a^b 1_{\{W^1(u)>0\}} 1_{\{W^2(u)=0\}} du.$$

The workload process W^2 satisfies:

$$\forall 0 \le a \le b, \qquad W^2(b) = W^2(a) + S^2(a,b) - \int_a^b 1_{\{W^2(u)>0\}} du.$$

Finally, the total workload of the server, given by $W = W^1 + W^2$, satisfies:

$$\forall 0 \le a \le b, \qquad W(b) = W(a) + S^1(a,b) + S^2(a,b) - \int_a^b 1_{\{W(u)>0\}} du.$$

Proof The workload of types in I_1 at an instant b equals the load at an instant a ($a \leq b$), increased by the quantity of work of types in I_1 arrived in the interval $[a, b)$, minus the work done in the same interval. Since the policy is non-idling, the server works on types I_1 if and only if the workload of types in I_1 is positive and there is no customer of type in I_2 in queue. The evolution equation for W^2 is justified by the fact that types in I_2 are strictly preemptive on types in I_1.

Conversely, these equations can be shown to have a single solution.

□

Lemma 2.2 *If $\rho^1 + \rho^2 < 1$ then the system is stable, that is:*

$$\forall t \geq 0,\ \exists t' > t \text{ such that } W(t') = 0.$$

Proof Assume that for some t, $\forall t' > t$, $W(t') > 0$. Then, by lemma 2.1:

$$W(t') = W(t) + S^1(t, t') + S^2(t, t') - (t' - t).$$

Using the burstiness constraints (1) and (2),

$$W(t') \leq W(t) + \Lambda^1 + \Lambda^2 + (\rho^1 + \rho^2 - 1)(t' - t).$$

This last inequality leads to a contradiction: when t' is large enough, the quantity in the right-hand side becomes negative. □

In the following, we derive an upper bound for the quantity of work done by the server on types in I_1 and for the workload W^1. These results are completed by a study of the activity periods. One can note that these results are valid for any single server operating under any policy, by considering an empty set I_2. We assume that the queue is stable, *i.e.* $\rho^1 + \rho^2 < 1$, and that properties (1) and (2) hold. For simplicity, we also assume that at $t = 0$ the queue is empty ($W(0) = 0$).

In addition to the previous notation, let, for any set of types I, $\Lambda_I = \sum_{i \in I} \Lambda_i$ and $\rho_I = \sum_{i \in I} \rho_i$.

An *activity period* of the system is defined as a time interval $[a, b]$ such that $\int_a^b \mathbf{1}_{\{W(u)=0\}} du = 0$ and such that there is no interval including $[a, b]$ and satisfying this property. When there is no preemption, this corresponds to the classical notion of a *busy period* during which the queue is not empty, with the small difference that we allow $W(t) = 0$ in such a busy period if there is an arrival at t. In others words, idle periods have a strictly positive duration.

Define the function f by: $f(t) = 0$ if the server is not working, and $f(t) = i$ if the server is working on a customer of type i. The key result for obtaining bounds on the workload W^1 is the following bound on the quantity of work done by the server on customers of types in $I \subseteq I_1$.

Lemma 2.3 *For any subset $I \subseteq I_1$ and any $0 \leq a < b$,*

$$\int_a^b \mathbf{1}_{\{f(u)\in I\}}du \leq \Lambda_I + \rho_I \frac{\Lambda^1 + \Lambda^2 - \Lambda_I}{1 - \rho^2 - (\rho^1 - \rho_I)} + \rho_I(b-a).$$

Proof Consider first the case where $W(a) > 0$. Let $c = sup\{t \leq a | W(t) = 0\}$. Such an instant exists, since the system is assumed to be stable and the initial workload is zero. This instant is the beginning instant of the *activity period* containing a. We have:

$$\begin{aligned}\int_a^b \mathbf{1}_{\{f(u)\in I\}}du &= \int_c^b \mathbf{1}_{\{f(u)\in I\}}du - \int_c^a \mathbf{1}_{\{f(u)\in I\}}du \\ &= \int_c^b \mathbf{1}_{\{f(u)\in I\}}du - (a-c) + \int_c^a \mathbf{1}_{\{f(u)\notin I\}}du \\ &= \int_c^b \mathbf{1}_{\{f(u)\in I\}}du - (a-c) + \int_c^a \mathbf{1}_{\{f(u)\in \bar{I}\}}du + \int_c^a \mathbf{1}_{\{W^2(u)>0\}}du,\end{aligned}$$

where $\bar{I}$ is the complementary of I in I_1. This follows from the fact that $[c, a]$ is included in an activity period, so that if the server is not working on type I, then it is either working on a type in I_2 or working on a type in $\bar{I}$. Observe that the sum of the two last integrals is less than $a - c$.

Given that $W(c) = 0$, we have $\int_c^b \mathbf{1}_{\{f(u)\in I\}}du \leq S_I(c, b)$. The same holds for $\bar{I}$. Using properties (1) and (2), we obtain:

$$\begin{aligned}&\int_a^b \mathbf{1}_{\{f(u)\in I\}}du \\ &\leq \Lambda_I + \rho_I(b-a+a-c) - (a-c) + \min\{a-c, \Lambda_{\bar{I}} + \rho_{\bar{I}}(a-c) + \Lambda^2 + \rho^2(a-c)\} \\ &= \Lambda_I + \rho_I(b-a) + \min\{\rho_I(a-c), \Lambda^1 - \Lambda_I + \Lambda^2 + (\rho^1 - \rho_I + \rho^2 - 1)(a-c)\}\ .\end{aligned}$$

It is easily checked that under the stability condition, $\max_{x\geq 0} \min\{\rho_I x, \Lambda^1 - \Lambda_I + \Lambda^2 + (\rho^1 + \rho^2 - 1)x\} = \rho_I(\Lambda^1 + \Lambda^2 - \Lambda_I)/(1 - \rho^1 - \rho^2 + \rho_I)$. The lemma follows in this case.

If now $W(a) = 0$, the proof can be easily adapted. □

A similar approach provides a bound on the workload of types in I_2 of the server (see [3, 4] for more details):

Theorem 2.4 *If the system is stable and $W(0) = 0$, then*

$$\forall t \geq 0, \qquad W^1(t) \leq \Lambda^1 + \rho^1 \frac{\Lambda^2}{1-\rho^2}.$$

Remark that if the set I_2 is empty, the bound reduces to $W^1(t) \leq \Lambda^1$. One can consult [8] for further discussion on this latter bound.

Finally, we give a bound on the length of the activity periods of the system (the proof should be found in [4]).

Theorem 2.5 *If $\rho^1 + \rho^2 < 1$ and $W(0) = 0$, then the duration δ of any activity period satisfies:*

$$\delta \leq \frac{\Lambda^1 + \Lambda^2}{1 - \rho^1 - \rho^2}.$$

2.2 Periodicity and feasibility

Let us now assume that the arrival process of the queue is periodic. For each type of customer $i \in I_1 \cup I_2$, we define:

- L_i the (constant) service time of a customer,
- R_i the period of the arrival process, which is assumed to be an integer,
- Φ_i the instant of the first arrival (initial phase),
- for types in I_1, σ_i the constant *deadline*, such that $L_i \leq \sigma_i \leq R_i$.

The quantities L_i, R_i and Φ_i will be expressed in the same unit. Let T^1 be the common period of types in I_1, that is the least common multiple of the R_is, $i \in I_1$, and let T^2 be the period of the arrival process of types in I_2. We define also T the common period (least common multiple of T^1 and T^2).

One can prove that under the above hypothesis, the workload W^1 of types in I_1 becomes periodic after a transient period depending on the initial workload of types in I_1 [3].

Observe that conditions (1) and (2) hold with:

$$\Lambda_i = L_i, \forall i \quad , \Lambda^1 = \sum_{i \in I_1} L_i, \quad \text{and} \quad \Lambda^2 = \sum_{i \in I_2} L_i,$$

$$\rho_i = L_i / R_i, \forall i, \quad \rho^1 = S^1(t, t+T)/T \quad \text{and} \quad \rho^2 = S^2(t, t+T)/T.$$

When $\rho^1 + \rho^2 < 1$, the system empties in every period. We have proved in [4] that another bound for the length of the activity periods holds when the arrival process and the preemption process are periodic:

$$\delta \leq \min\{\frac{\Lambda^1 + \Lambda^2}{1 - \rho^1 - \rho^2}, T(\rho^1 + \rho^2)\}. \tag{3}$$

We shall say that a task set is *feasible* if, starting from an empty system ($W(0) = 0$), the delay of every customer is less than its deadline, for any initial phase. One may envision several approaches to feasibility:

- *tests with trajectory* which reduce the problem to the study of certain trajectories obtained with particular phases. They consist in verifying the feasibility of task sets on these trajectories, which may involve the simulation of the system. For the case of a single server, first studies of this type of feasibility tests were done by C.L. Liu and J.W. Layland [10]. S. Lefebvre-Barbaroux has established such a test in the case of nonpreemptive policies [8]. In [3] we have shown that this approach fails in general for networks.
- *test on parameters* of the system, such as message lengths and periods, or load factors. These conditions are much easier to compute but are often only necessary or only sufficients. In the case of a single server with periodical arrivals operating under the preemptive policy "Rate Monotonic", a sufficient condition for feasibility has been proven [10, 11]: if the load factor (our ρ) satisfies the stability condition then the set is feasible. The bound approach leads to this kind of tests for feasibility.

We are interessed here in feasibility test for customers of type in I_1. Inequality (3) provides a sufficient condition for the feasibility of the task set I_1 under any work conserving policy, be it preemptive or nonpreemptive:

$$\min\{\frac{\Lambda^1 + \Lambda^2}{1 - \rho^1 - \rho^2}, T(\rho^1 + \rho^2)\} \leq \min_{i \in I_1} \sigma_i. \tag{4}$$

Indeed, the delay of any customer is clearly bounded by the length of the activity period to which it belongs. Because of its general nature, this condition is very weak, and better conditions are known for some classes of service disciplines [4]. Using the information of the discipline for types in I_1, one can improve the above condition.

2.2.1 Preemptive priorities

In this paragraph, we assume that each type of customer in I_1 has a fixed priority and can preempt customers with lower priority. Assume that customer types are ordered according to decreasing priority. The delay of a customer of type $i \in I_1$ is not influenced by customers of types $j > i$ (but is preempted by any customer of type in I_2). This allows us to refine condition (4).

Theorem 2.6 *A sufficient condition for the task set I_1 to be feasible under the fixed preemptive priority is:*

$$\forall i \in I_1, \qquad \frac{\sum_{j=1}^{i} L_i + \Lambda^2}{1 - \sum_{j=1}^{i} L_i/R_i - \rho^2} \leq \sigma_i \; .$$

2.2.2 FIFO service discipline

In this paragraph, we assume that the server operates under the FIFO (First In First Out) service policy for types in I_1. In this case, the delay of a customer in I_1 depends only on the workload found upon its arrival, and on the process of types in I_2 (and not on future arrivals in I_1). This allows to derive a feasibility condition which is better than (4)[3]:

Theorem 2.7 *The task set I_1 with $\rho^1 < 1$ is feasible by FIFO if:*

$$\frac{\Lambda^1 + \Lambda^2}{1 - \rho^2} \leq \min_i \sigma_i. \tag{5}$$

3 A feed-forward communication network

3.1 Basic properties

The general problem of the stability of a network with loops and multiple classes is complicated [1]. We first study the case of a feed-forward network. In [3, 4] we have considered the simpler case where stations are connected in series with preemptive bus. We now generalize the results to any topology such that the network does not contain loops (a tree topology for instance).

Consider a feed-forward network with N stations. Each type of customer has its route specified. In this type of feed-forward network without loops, one can define a depth level. The stations are numbered in order of their depth (see figure 1). Since there are no loops in the network, we can ignore communication delays.

Let us give some additional notations:

- $I(e)$ the set of the messages arriving directly at station e (local messages),
- $M(e)$ the set of the messages arriving at station e from stations of lower level (external messages),

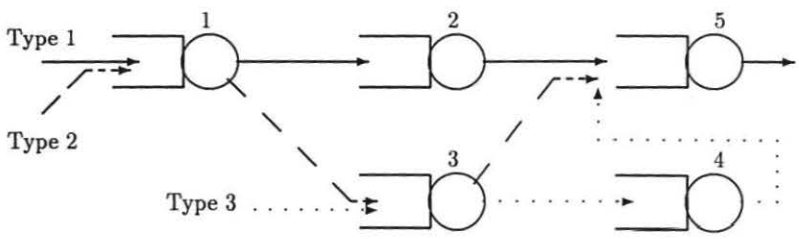

Figure 1: Feed-forward network

- let N_k be the set of stations at the k^{th} level. Stations of the lowest level (N_1) are never preempted by external messages,
- $\Lambda^e = \sum_{i\in I(e)} \Lambda_i$ the iargest quantity of local work which can arrive simultaneously in the station e,
- $\rho^e = \sum_{i\in I(e)\cup M(e)} \rho_i$ the load factor of station e,
- $\rho_I^e = \sum_{i\in I} \rho_i$ the load factor of messages of type belonging to some set $I \subset (I(e)\cup M(e))$ at station e,
- W^e the workload process in station e,
- $\forall 0 \leq a \leq b,\ \ S^e(a,b)$ the quantity of local work arrived at station e during the interval $[a,b]$,
- f^e the *trajectory function* of station e defined by: $f^e(t) = 0$ is station e is not emitting at instant t (because $W^e(t) = 0$), and $f^e(t) = i$ if the station is emitting a message of type i.

We assume that the local arrival process at each station e satisfies the burstiness property (1). By lemma 2.3 and induction we shall prove that the quantity of work from other stations also satisfies such property.

The following lemma states the functional equation for the workload W^e of any station of the network.

Lemma 3.1 *For all station $e \in N_k$, the workload process W^e satisfies the following evolution equation:* $\forall b \geq a \geq 0$,

$$W^e(b) = W^e(a)+S^e(a,b)+\sum_{p\in\cup_{i<k}N_i}\int_a^b 1_{\{f^p(u)\in M(e)\}}du-(b-a)+\int_a^b 1_{\{W^e(u)=0\}}du. \quad (6)$$

We shall assume that the condition of stability is satisfied for each station:

$$\rho^p < 1 \quad \forall p. \tag{7}$$

The following theorem establishes a bound for the workload for any station of the network:

Theorem 3.2 *Define by recurrence the sequences:*

$$B^e = \sum_{p \in \cup_{i \leq k} N_i} B^p_{M(e)} + \Lambda^e, \tag{8}$$

$$B^e_I = \Lambda_I + \rho_I \frac{B^e - \Lambda_I}{1 - \rho^e + \rho^e_I}, \tag{9}$$

for $e \in N_{k+1}$, $k = 1, \ldots, K-1$ (where K is the number of levels in the network) and $I \subseteq I(e) \cup M(e)$, with $B^p = \Lambda^p$ for all stations $p \in N_1$ at the lowest level. Then, assuming the system is stable, we have:

$$\forall e, \forall t, \qquad W^e(t) \leq B^e_{M(e)}. \tag{10}$$

Proof We actually prove by induction that for all e,

i/

$$\forall t, \qquad W^e(t) \leq B^e_{M(e)},$$

and

ii/

$$\forall I \subseteq I(e),\ \forall a < b,\ \int_a^b 1_{\{f^e(u) \in I\}} du \leq B^e_I + \rho^e_I (b-a).$$

For $e \in N_1$, claim i/ reduces to $W^e(t) \leq B^e_{M(e)} = \Lambda^e$, which is true by theorem 2.4. Likewise, claim ii/ is true by lemma 2.3.

Assume the claims have been proved for all $p \in \cup_{i<k} N_i$. Any station $e \in N_k$ can be seen as a server in isolation with its local arrivals and the arrivals of messages emitted by some stations $p \in \cup_{i<k} N_i$. Remind that the local arrival process satisfies:

$$\forall i \in I(e),\ \forall 0 \leq a \leq b,\ S_i(a,b) \leq \Lambda_i + \rho_i (b-a),$$

In addition, from the induction hypothesis, we have for all $p \in \cup_{i<k} N_i$:

$$\int_a^b 1_{\{f^p(u) \in M(e)\}} du = \int_a^b 1_{\{f^p(u) \in M(e) \cap I(p)\}} du \leq B^p_{M(e)} + \rho^p_{M(e)} (b-a),$$

so that finally:

$$\int_a^b \sum_{p \in \cup_{i<k} N_i} \mathbf{1}_{\{J^p(u) \in M(e)\}} du \leq \sum_{p \in \cup_{i<k} N_i} [B^p_{M(e)} + \rho^p_{M(e)}(b-a)]$$
$$= \mathcal{B}^e - \Lambda^e + (\rho^e - \rho^e_{M(e)})(b-a).$$

Let now $I \subseteq I(e) \cup M(e)$. Applying a reasoning similar to the proof of lemma 2.3 yields claim ii/. Claim i/ follows from theorem 2.4. □

We now give a bound on the length δ of the activity periods of any station. Based on equation (6), an argument similar to that of theorem 2.5 yields:

Theorem 3.3 *The length δ of any activity period of station e satisfies the bound:*

$$\delta \leq \frac{\mathcal{B}^e}{1-\rho^e}.$$

3.2 Periodicity and feasibility

Let us now assume, as in section 2.2, that the arrival process is periodic. In the case where the service policy is FIFO or a static priority (for the local customers), we can prove that the systems couples in finite time with a periodic regime [3]. The period of station $e \in N_{k+1}$ is T^e the least common multiple of local messages and of all messages emitted by stations $p \in \cup_{i \leq k} N_i$.

We assume here that external messages have a strictly preemptive priority upon the local ones. In this context we shall speak on the feasibility of the local set of messages (as the deadline refers to the date of the end of emission at the entry station). A general feasibility test similar to condition (4) is derived from theorem 3.3:

$$\forall e, \qquad \frac{\mathcal{B}^e}{1-\rho^e} \leq \min_{i \in I(e) \cup M(e)} \sigma_i,$$

which is valid for any work conserving discipline. Once again, this condition can be improved is the service discipline is a fixed preemptive priority. If messages are emitted in the FIFO order, we have an improved condition. The following condition is a consequence of theorem 2.7.

Corollary 3.4 *If all stations are FIFO, a sufficient condition for feasibility is*

$$\forall e, \qquad \frac{\mathcal{B}^e}{1-\rho^e+\rho^e_{M(e)}} \leq \min_{i \in I(e)} \sigma_i.$$

4 Ring network

In this section we consider the case of n stations which communicate around a unidirectional ring network (2). The model is similar to the one presented in section 3 but we shall assume now that a message emitted will preempt the stations it will go through. Thus, external messages never accumulate in a station, but only preempt it. It is now necessary to take transmission delays into account by introducing d_{ep}, the transmission delay between the station e and the station p.

Again, the arrival process at each station is supposed to satisfy the burstiness constraint. We first establish a sufficient condition for stability. Then we discuss the case of periodical arrivals. Finally we present some bound for the quantity of work done by any station during a certain time interval, and therefore a bound for the workload of any station.

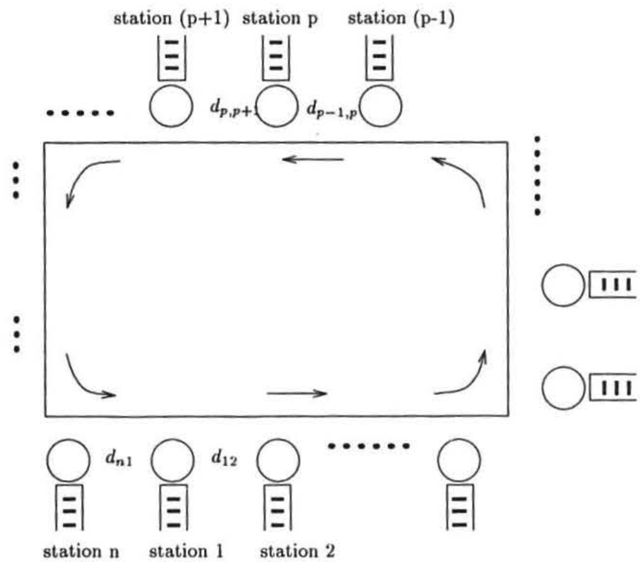

Figure 2: Ring network with n stations

Lemma 4.1 *For any station e of the ring network, the workload process W^e is characterized by the functional equation:* $\forall t \geq 0$,

$$W^e(t) = W^e(0)+S^e(0,t)-t+\int_0^t \mathbf{1}_{\{W^e(u)=0\}}du+\int_0^t \sum_{p\neq e}\mathbf{1}_{\{f^p(u-d_{pe})\in M(e)\}}\mathbf{1}_{\{W^e(u)>0\}}du.$$

Proof One can remark that for any station e and for any instant t,

$$1 = \mathbf{1}_{\{W^e(t)=0\}} + \mathbf{1}_{\{W^e(t)>0\}}[\sum_{p\neq e}\mathbf{1}_{\{f^p(t-d_{pe})\in M(e)\}}] + \mathbf{1}_{\{W^e(t)>0\}}[\sum_{p\neq e}\mathbf{1}_{\{f^p(t-d_{pe})\notin M(e)\}}]. \quad (11)$$

It simply signifies that we have three disjoint events corresponding respectively: to an empty server (in this case, the channel state is not relevant), to a server not empty but preempted, and to a server occupied and not preempted.

At any instant $t \geq 0$, the workload of any server e is equal to the quantity of work arrived (*i.e.* $W^e(0) + S^e(0,t)$) minus the quantity of work done. This latter is t minus the time during which the server was not working (either because it was empty, or because is was preempted). This quantity is directly deduced from (11). □

Theorem 4.2 *Any station e of the ring network is stable if:*

$$\rho^e = \sum_{i\in M(e)} \rho_i < 1.$$

The proof is similar to the proof of lemma 2.2, see [3] for details.

In the case of periodicals arrivals, we have no succeed in determining the periodicity of the workload process of any station. Even in the case for which the periodicity could be established, its value is not the least commun multiple of the periods of all the customers [3].

We are now interessed in the obtention of bounds. The arrival process does not need to be periodic. We have to define additional notations:

- for all station $p = 1,\ldots,n$, d_p is the delay associated to the transmission through the section between station $p-1$ and p (d_1 is the delay of transmission between station n and station 1),
- for any couple (p,k), where p is a station number and $k \in \mathbb{N}^*$, we define the time interval: $I_k^p = [0, \sum_{i=1}^k d_{p-i}]$,
- for all station p, all subset of types $J \subseteq I(p)$ and all $k \in \mathbb{N}^*$:

$$B_J^{p,k} = \Lambda_J^p + \rho_J^p \frac{\Lambda^p - \Lambda_J^p + \sum_{i=1}^{min(k,n-2)} B_{M(p)}^{p-i,k-i}}{1 - (\sum_{i=0}^{min(k-1,n-2)} \rho_{M(p)}^{p-i}) + \rho_J^p}, \quad (12)$$

with $B_J^{p,0} = 0$ and $\rho^p_{M(p)} = \rho^p$, by convention.

Theorem 4.3 *For all $k \in \mathbb{N}^*$, the quantity of work of type J done by a station p during the interval $[a,b]$ included in I_k^p verifies:*

$$\int_a^b \mathbf{1}_{\{f^p(u)\in J\}} du \leq B_J^{p,k} + \rho_J^p(b-a) \tag{13}$$

Proof For $k=1$, any station p observed during the interval $I_1^p = [0, d_{p-1}]$ behaves as if it was in isolation. Actually, messages from station $(p-1)$ have not time to arrive at p. We can apply lemma 2.3 which gives a bound for the quantity of work done by a single server:

$$\forall [a,b] \subseteq I_1^p,\ \int_a^b \mathbf{1}_{\{f(u)\in J\}} du \leq \Lambda_J^p + \rho_J^p \frac{\Lambda^p - \Lambda_J^p}{1-\rho^p+\rho_J^p} + \rho_J^p(b-a) = B_J^{p,1} + \rho_J^p(b-a).$$

Now, suppose the assertion verified for $k-1$, and for any station of the network.

During the interval I_k^p, the station p may be preempted by stations $p-i$ with $i = 1,\dots,\min(k, n-2)$. Note also that if $u \in I_p^k$ then $(u - \sum_{i=1}^{l} d_{p-i}) \in I_{k-l}^{p-l}$.

Let p be a station, $J \subseteq I(p)$ and $[a,b]$ included into I_k^p. We denote $a_p = \sup\{t \leq a, W^p(t) = 0\}$ the beginning of the occupancy period of station p which contains a (if $W^p(a) = 0$, the proof is easily adapted, considering the next instant less than b for which the workload is positive). Then,

$$\begin{aligned}
\int_a^b \mathbf{1}_{\{f^p(u)\in J\}} du &= \int_{a_p}^b \mathbf{1}_{\{f^p(u)\in J\}} du - \int_{a_p}^a \mathbf{1}_{\{f^p(u)\in J\}} du \\
&= \int_{a_p}^b \mathbf{1}_{\{f^p(u)\in J\}} du - (a-a_p) + \int_{a_p}^a \mathbf{1}_{\{f^p(u)\notin J\}} du \\
&\leq S_J(a_p, b) - (a-a_p) + \int_{a_p}^a \mathbf{1}_{\{f^p(u)\in I(p)\setminus J\}} du + \int_{a_p}^a \sum_{p\neq e} \mathbf{1}_{\{f^p(u-d_{pe})\in M(e)\}} du.
\end{aligned}$$

But,

$$\begin{aligned}
&\int_{a_p}^a \sum_{p\neq e} \mathbf{1}_{\{f^p(u-d_{pe})\in M(e)\}} du \\
&= \sum_{i=1}^{\min(k,n-2)} \int_{a_p}^a \mathbf{1}_{\{f^{p-i}(u-\sum_{i=1}^{\min(k,n-2)} d_{p-i})\in M(p)\}} du \\
&\leq \sum_{i=1}^{\min(k,n-2)} [B_{M(p)}^{p-i,k-i} + \rho_{M(p)}^{p-i}(a-a_p)],
\end{aligned}$$

the induction hypothesis justifies the last inequality. Finally, with $(a - a_p) = \beta$

$$\begin{aligned}\int_a^b 1_{\{J^p(u)\in J\}}du &\leq \Lambda_J^p + \rho_J^p(b-a+\beta) - \beta \\ &\quad + \min\left\{\beta, S_{(I(p)\setminus J)}(a_p, a_p+\beta) + \sum_{i=1}^{\min(k,n-2)} [B_{M(p)}^{p-i,k-i} + \rho_{M(p)}^{p-i}\beta]\right\} \\ &\leq \Lambda_J^p + \rho_J^p(b-a) \\ &\quad + \min\left\{\rho_J^p\beta, \beta(\sum_{i=1}^{\min(k,n-2)} \rho_{M(p)}^{p-i} - 1) + \Lambda_J^p + \sum_{i=1}^{\min(k,n-2)} B_{M(p)}^{p-i,k-i}\right\}.\end{aligned}$$

The maximum of this minimum is reached for:

$$\beta = \frac{\Lambda^p - \Lambda_J^p + \sum_{i=1}^{\min(k,n-2)} B_{M(p)}^{p-i,k-i}}{1 - (\sum_{i=1}^{\min(k,\,n-2)} \rho_{M(p)}^{p-i}) + \rho_J^p},$$

hence the expression for $B_J^{p,k}$.

□

We have obtained a linear recurrence for terms $B_J^{p,k}$. This recurrence converges under certain conditions which do not always correspond to the stability condition. It seems that the sole stability condition does not insure the existence of bounds. For a more detailed discussion of that point, see [3, 6]. We have assumed that external messages have a strictly preemptive priority upon local messages. We could write bounds for more general case but these bounds would probably diverge sooner.

In networks with loops, the analysis is difficult, essentially because messages emitted by some station have an influence on the traffic of other stations, which influence in return this particular station. It is hope that our approach based on evolution equations can be improved to derive better bounds. The question of the stability of networks is currently under study.

References

[1] C.S. Chang, Stability, Queue Length and Delay, Part 1: Deterministic Queueing Networks, *IBM Research Report 17708*, IBM Research Division, 1992. In *Proc. IEEE Conf. on Decision and Control*, vol. 1,1992.

[2] C.S. Chang, Stability, Queue Length and Delay, Part 2: Stochastic queuing networks, *IBM Research Report 17709*, IBM Research Division, 1992. In *Proc. IEEE Conf. on Decision and Control*, vol. 1,1992.

[3] C. Chaouiya, *Outils pour la validation de contraintes de synchronisation dans des systèmes distribués*, PhD Thesis, University of Nice-Sophia Antipolis, October 1992.

[4] C. Chaouiya, S. Lefebvre-Barbaroux and A. Jean-Marie, Real-Time Scheduling of Periodic Tasks, To appear in *Scheduling Theory and its Applications*, P. Chretienne, E. G. Coffman, J. K. Lenstra, Z. Liu, (Eds.) J. Wiley, 1993.

[5] R.L. Cruz, A Calculus for Network Delay, Part I: Network Elements in Isolation, *IEEE Trans. Information Theory*, **37**, No 1, pp. 114–131, January 1991.

[6] R.L. Cruz, A Calculus for Network Delay, Part II: Network Analysis *IEEE Trans. Information Theory*, **37**, No 1, pp. 132–141, January 1991.

[7] J. Kurose, On computing per-session performance bounds in high-speed multihop computer networks, In *Performance Evaluation Review*, vol.20, Newport, 1992. ACM SIGMETRICS and Performance'92.

[8] S. Lefebvre-Barbaroux, *Files d'attente avec arrivées atypiques: environnement aléatoire et superposition de flux périodiques*, PhD Thesis, University of ParisXI, Orsay, February 1992.

[9] S. Lefebvre-Barbaroux, A. Jean-Marie and C. Chaouiya, Problèmes d'ordonnancement d'une superposition de flux périodiques sous contrainte temps-réel, INRIA Research Report 1576, March 1992.

[10] C.L. Liu and J.W. Layland, Scheduling Algorithms for Multiprogramming in Hard-Real-Time Environment, *Journal of the Association for Computing Machinery*, **20**, No 1, January 1973, pp. 46-61.

[11] O. Serlin, Scheduling of Time Critical Processes, *AFIPS Conference Proceedings SJCC*,**40**, Atlantic City, 1972.