On the Quantitative Specification of Jitter Constrained Periodic Streams

Claude-Joachim Hamann
Dresden University of Technology
Department of Computer Science
D-01062 Dresden, Germany
claude.hamann@inf.tu-dresden.de

Abstract

An increasing number of application systems can be characterized by their requirement to process sequences of events in real-time. These sequences are principally of constant rate, but may vary within given limits. Hence, several parameter sets that seem to differ only slightly have been proposed to describe such sequences. This includes the parameter sets used in the Tenet Protocol Suite, in the traffic description of an ATM connection, and in the model of linear bounded arrival processes (LBAP) for transferring continuous media.

The existence of several parameter sets raises the question whether or not the parameter sets differ in principle or in notation only. To answer this question, the paper proposes a generalized model for jitter constrained periodic streams. That model subsumes the parameter sets mentioned above, allows to prove their equivalence and to transform the different sets of parameters each to another.

1. Introduction

Multimedia and other real-time applications can be characterized by their requirements to process sequences of events in real-time. To do that, proper resources must be made available in time by systems sometimes called quality of service architectures. To provide or reserve resources in time, the load for such systems, i.e. the characteristics for these sequences of events, must be described precisely. The sequences are principally of constant rate that may vary within given limits. These sequences occur at various levels of a multimedia system, e.g., at the cell relay level of an ATM hardware adapter or at the interface of a file system.

Hence, processes of the following structure are of interest: events occur at an interface of a distributed system principally at a constant rate or with constant distance, but the occurrence may vary over a given interval. In other words, events may occur late or early but do not exceed a given time span.

Surprisingly (or not surprisingly), it is not a single scheme that has been proposed to describe the simple abstraction of sequences of events but several that differ somehow. Each new quality of service architecture project seems to invent its own, new model and sets of parameters. Hence, at the start of Dresden Real-time Operating System Project, we tried to understand the principle differences of the model to select one of them, or derive our own.

For this purpose we looked at three of such parameter sets of major importance: the traffic description parameters for the Tenet Protocol Suite, the parameters according to the traffic description used in ATM networks, and the LBAP model proposed by Anderson. As will be shown in the paper, these three models turn out to be equivalent in a sense described below. To show this, yet another model is introduced and it is shown how the parameters of the three mentioned models can be transformed into our new model.

The literature on real-time systems and communication contains many investigations and results concerning particular aspects of traffic models, above all buffer space requirements and end-to-end delay guarantees. For instance, the „leaky bucket“ model (that means the Generic Cell Rate Algorithm in ATM) is described in [4], burst streams are considered in [9], both with respect to end-to-end delay founded on an deterministic model. Zhang and Ferrari [13] introduced a probabilistic model, and [7], 10], [12] deal with bounds of buffer space based on Markovian chains (discrete-time, finite state processes). On the other hand a determistic model is used for the same reason in [11].

However, only few papers compare several models or traffic descriptions but with respect to server disciplines [14] or traffic constraint functions [8]; the latter includes the Tenet traffic model. Khan’s approach [6] is similar to ours: he introduced the “(L, M, T) mechanism” to evaluate the leaky bucket behaviour, where L and T are the minimum and mean distance respectively between two

ATM cells and M denotes the maximum burst size, but he also supposed a stochastic input process. However, to the best of our knowledge, there are no quantitative comparisons between parameter sets of traffic descriptions.

The paper is organized as follows. Section 2 describes the new generalized model. Section 3 contains some results, above all the computation of the maximum burst size and a lower bound of buffer size required to avoid loss of data. These results enable to investigate the equivalence of the three parameter sets mentioned above in section 4. Finally, section 5 contains the conclusion and outlines the future work.

The paper summarizes the results only, with exception of a few cases.. The accurate formal definitions and the complete proofs are included in [5].

2. The Generalized Model

This paper deals with sequences of events (e.g., sending or receiving data units, all units are of the same size) called *event streams* in the following sense: Beginning from a starting point t_0 the events occur one after each other with a principally constant distance T (that means with constant rate $R = 1/T$). However they may vary over a given interval: events may occur τ time units too early or τ' time units too late as long as they obey a minimum distance $D < T$ (see Figure 1). Hence, T is the average distance between two events. Without loss of generality we assume $t_0 = 0$.

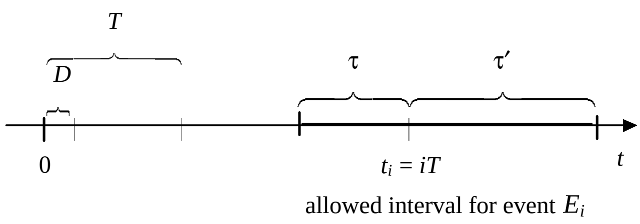

Figure 1. Jitter constrained stream

Definition 1. Let

$$D, T, \tau, \tau' \in \mathbb{R} \quad \text{with} \quad T > D > 0, \quad \tau, \tau' \geq 0, \qquad (1)$$

and $i \in \mathbb{N}$. A sequence of events $(E_i)_{i=0,1,...}$ is a *(τ,τ')-constrained periodic stream with a constant period T and minimal distance D* (shortly: *jitter constrained stream*) iff the following holds: event E_i occurs at

$a_i \in [t_i - \tau, t_i + \tau'] \subseteq \mathbb{R}$ real event time

with

$t_i = iT$ periodical (or expected) event time

and the distances obeys the condition

$$a_{i+1} - a_i \geq D \qquad \forall i \in \mathbb{N}.$$

Let us now consider bursts, i.e. streams where events may follow each other at shortest possible intervals. Furthermore, let us identify the events E_i with the moment a_i of their occurrence. So an "event sequence of length k+1" is a sequence $a_i a_{i+1} ... a_{i+k}$ $(i, k \in \mathbb{N})$, where the elements of the sequence obey definition 1. An event sequence obeying definition 1 is called a *burst of length l* (shortly *l-burst* or *burst*), $l \in \mathbb{N}$, $l > 0$, if the events follow each other at distance D, but the distance to an earlier and to a later event time is greater than D. (A l-Burst is assumed to have length l only, before and after that we expect a break.)

Obviously, a burst B is definitely given by its starting point $a(B) = a_i$ and its length $l(B) = l$. Furthermore, in our model the length of a burst is bounded by a natural number L. Thus, a burst is called a *maximum burst* if it has maximum length L.

Next we investigate two "worst cases":

- *burst streams*, i.e. streams consisting of maximum bursts only (data are to be transferred in packets as large as possible; the breaks between the bursts can vary respecting definition 1);
- *tightly packed streams*, i.e. transferring the data units through packets with varying size, but at earliest possible time.

It is the objective of the next section to derive important performance measures:

– maximum burst size L
– earliest b_e and latest b_l starting point of a maximum burst
– earliest starting point b^l of an l-burst
– intervals between bursts (interburstiness), especially the smallest and the largest distance I_s, I_l between maximum bursts
– lower bound P of buffer size to avoid loss of data
– number of events, $N(t)$, occurring before time t
– calculating the parameters τ, τ' from L.

3. Some Results

3.1. Burst streams

A jitter constrained stream is called a *(maximum) burst stream* in our model if it is a sequence $(B_i)_{i=0,1,...}$ of bursts B_i where

$$l(B_i) = L \qquad \forall i \in \mathbb{N}.$$

Obviously, a burst with maximum length L is generated if the burst initial time $a(B)$ is as late as possible, which means $a(B) = b_s$, $b_s = \tau'$. Thus, it follows (see Figure 2)

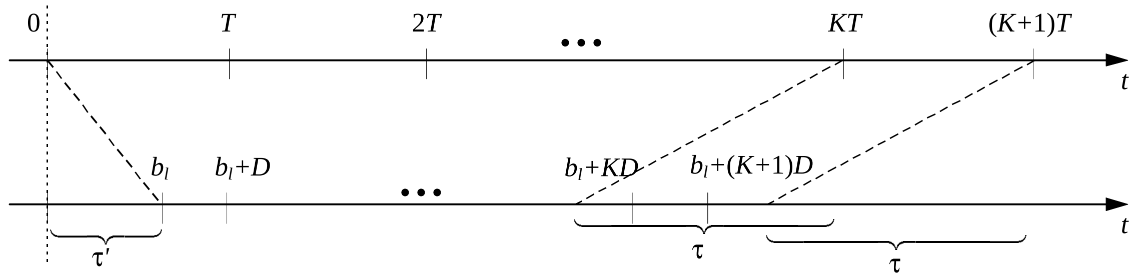

Figure 2. Maximum burst length

$$K = \max_{b_s + kD \geq kT - \tau} (k \in \mathbb{N}) =$$

$$\max_{k \leq \frac{\tau+\tau'}{T-D}} (k \in \mathbb{N}) = \left\lfloor \frac{\tau + \tau'}{T - D} \right\rfloor$$

where $K = L - 1$ and therefor

$$L = 1 + \left\lfloor \frac{\tau + \tau'}{T - D} \right\rfloor . \qquad (2)$$

Similarly:

$$b_l = \tau' \qquad (3.1)$$

$$b_e = K(T-D) - \tau = (T-D)\left\lfloor \frac{\tau + \tau'}{T - D} \right\rfloor - \tau . \qquad (3.2)$$

$$I_s = 2K(T-D) + T - (\tau + \tau') . \qquad (3.3)$$

$$I_l = T + \tau + \tau' . \qquad (3.4)$$

Furthermore, a burst stream is periodically with the period LT. That is, it holds for a stream $(B_i)_{i=0,1,...}$ of maximum bursts:

$$a(B_i) \in [b_e + iLT,\ b_l + iLT] \quad \forall\, i \in \mathbb{N}. \qquad (4)$$

Often it is not of interest to compute the maximum burst length L from the parameters given in (1). Rather, the maximum burst size is given and we need to determine the parameters τ, τ′. It follows from (2):

$$\tau + \tau' \in \left[(L-1)(T-D), L(T-D) \right) . \qquad (5)$$

Now let us derive a lower bound P of buffer size that loss of data is impossible. We suppose that "event" means the arrival of a data unit (e.g., an ATM cell) for a server process; all data units are of the same size. The process handles or consumes the data units with the same constant period T as the arriving units, but without variation or jitter. To avoid loss of data it is necessary to store data units arriving too early in a buffer. For the server process, we consider two policies:

(P1) It is guaranteed that the server process has always work to do (may be the process must wait at the beginning).

(P2) The policy is work conserving, i.e., the server process begins its work when a data unit arrives.

Furthermore, we assume: if the arrival time of a data unit coincides with the time when the server process takes the next data unit from the buffer then the size of used buffer does not change. Then it follows:

Theorem 1. Given a jitter constrained stream of data units of equal size exclusively, consisting of maximum bursts, and a server process. Loss of data units is not allowed; then the lower bound of buffer size P is

$$P = \left\lceil \frac{\tau + \tau'}{T} \right\rceil \qquad (6)$$

if the server begins to take the data units out of the buffer not later than at the latest starting point b_s of a maximum burst. That is true for both policies, (P1) and (P2).

3.2. Tightly packed streams

Now we investigate event streams consisting of bursts B_i with different length. All the starting points $a(B_i)$ are supposed to be as early as possible obeying definition 1.

Definition 2. A jitter constrained stream is a *tightly packed stream* iff it is a sequence $(B_i)_{i\ =\ 0,1,...}$ of bursts B_i where

$$l_i := l(B_i) \leq L \qquad L \text{ according to (2)},$$

$$a_i := a(B_i) = t_i\, T + b^{l_i}$$

with

$$t_i = \sum_{j=1}^{i} l_j\ , \quad b^{l_i} = (l_i - 1)(T-D) - \tau$$

for all $i \in \mathbb{N}$ (l_i are the free parameters of the sequence).

Considerations similarly to section 3.1 yield:

Theorem 2. Given a jitter constrained stream of data units consisting of bursts with different length and a server process. The bursts arrive as early as possible and loss of data units is not allowed. Then the minimum buffer size P is

$$P = \left\lceil \frac{(L-1)(T-D)}{T} \right\rceil \quad (7)$$

if the server begins to take the data units out of the buffer not later than at time

$$b^L = (L-1)(T-D) - \tau. \quad (8)$$

Taking into account the desired comparison of different models, we compute an upper bound $N(t)$ of number of events occuring until time t inclusively. Let us assume $\tau' = 0$ with respect to chapter 4, then $b \le 0$ follows where $b = a(B_0)$ denotes the first starting point of a tightly packed stream. Next we transform the time axis such that the time begins with the first event. Then we can prove:

Lemma. Let $N(t)$ the number of events occuring in a tightly packed stream until time t; time is measured from the first burst event. Then it holds for all $t \in \mathbb{R}$:

$$N(t) \le \overline{N}(t)$$

where

$$\overline{N}(t) = \begin{cases} 0 & \text{if} \quad t < 0 \\ 1 + \left\lfloor \frac{t}{D} \right\rfloor & \text{if} \quad 0 \le t < LT \\ 1 + \left\lfloor \frac{t+\tau}{T} \right\rfloor & \text{if} \quad t \ge LT \end{cases} \quad (9)$$

On closer inspection of the lemma's assumption, we expect that it is not necessary to introduce *two* parameters τ, τ' in definition 1. More precisely:

Definition 3. Two models (D, T, τ_1, τ'_1), (D, T, τ_2, τ'_2) according to definition 1 are *equivalent* if they both determine the same maximum burst length L.

It follows that all further stream characteristics are equal respectively, except for the absolute time measures. In particular, the minimum buffer size according to (6) and (7) and the interburstiness are equal in two equivalent models. If we transform the time as described above (time "begins" with the first event) then the streams are identical. Hence, it follows from theorem 2:

Corollary. The models (D, T, τ, τ'), $(D, T, \tau + \tau', 0)$, and $(D, T, 0, \tau + \tau')$ are equivalent.

Before we compare our model with some other models and parameter sets we consider the borderline case $D = 0$ that was not allowed so far in definition 1. A precise analysis of section 2 and 3 shows that all essential notions can be extended and results are true in that case too (however, we cannot identify an event E_i with its time $a(E_i)$).

4. Equivalence of Parameter Sets

4.1. ATM

The comparison between event streams according definition 1 and sequences of events in ATM connections is based on [1]; refer to chapter 4.4: Traffic Contract Parameters and Related Algorithms.

The following traffic parameters describe the traffic characteristics of an ATM connection:

PCR	Peak Cell Rate
CDVT	Cell Delay Variation Tolerance
SCR	Sustainable Cell Rate (optional)
BT	Burst Tolerance (optional).

The *Generic Cell Rate Algorithm* GCRA(I,L) controls the traffic of an ATM connection. A stream of cells is in accordance to that algorithm if ATM-cells may arrive earlier than a theoretical arrival time TAT but not too early. The ability to send cells prematurly can be used to build bursts. This algorithm has two parameters I (increment) and L (limit). It generates a jitter constrained stream with two specific features:

(a) The delay is not bounded. If an event occurs at time t_a after its theoretical arrival time then TAT is updated to the current time t_a.

(b) TAT is initialized to the current arrival time of the first cell event.

Furthermore, in [1] are defined:

– $R_p = \frac{1}{T_p}$ Peak Cell Rate PCR,

T_p: Peak Emission Interval ; minimum inter-arrival time between two basic events in the PCR reference model;

– τ^*: Cell Delay Variation Tolerance; limit parameter L of the GCRA describing the "cell clumping phenomenon";

– δ: time required to send an ATM cell.

Hence, the "worst case" processes generated by the GCRA in the PCR reference model are $(\tau^*, 0)$-jitter constrained with period T_p and minimal distance δ; $\tau' = 0$ follows from (a). The formula for the maximum number N of conforming back-to-back cells (i.e. the maximum length of bursts if $D = \delta$)

$$N = \left\lfloor 1 + \frac{\tau}{T_p - \delta} \right\rfloor$$

given in [1] is obviously the same as (2) because $L = N$ (note $\tau' = 0$).

Analogously, in [1] are defined:

– $R_s = \frac{1}{T_s}$ Sustainable Cell Rate SCR

T_s: minimum inter-arrival time between two basic events in the SCR reference model;

– τ_s: limit parameter L of the GCRA (called Burst Tolerance BT).

Just as above the processes generated by the GCRA in the SCR reference model are (τ_s, 0)-jitter constrained with period T_s and minimum distance T_p. The formulas for the maximum burst size MBS and the computation of τ_s from MBS

$$\mathrm{MBS} = \left\lfloor 1 + \frac{\tau_s}{T_s - T_p} \right\rfloor$$

$$\tau_s \in \left[(\mathrm{MBS} - 1)(T_s - T_p),\ \mathrm{MBS}(T_s - T_p) \right)$$

are identical to (2) and (5) because L = MBS.

In the same way the formula in [1] for the maximum number of cells, $N(t)$, that can be emitted with spacing no less T_p

$$N(t) \le \begin{cases} 1 + \left\lfloor \dfrac{t}{T_p} \right\rfloor & \text{if } t < \mathrm{MBS} \cdot T_s \\ 1 + \left\lfloor \dfrac{t + \tau_s}{T_s} \right\rfloor & \text{if } t \ge \mathrm{MBS} \cdot T_s \end{cases}$$

corresponds to (9).

4.2. Linear Bounded Arrival Processes

Anderson introduced a continuous media resource model [2] that decomposes a distributed system into a set of resources such as CPU, networks, and file systems. Several data streams with different quality are to be processed, stored, and transferred between the system components, while maintaining guaranteed end-to-end performance, e.g., throughput and delay. The workload (or the traffic) occurring at the interface of a resource is described in terms of discrete messages with three parameters:

– M maximum message size (it is not considered in this paper)
– R maximum message rate
– W workahead limit.

The workahead results from the ability of processes or devices to generate bursts that are transmitted within a short interval with a rate greater than R. Thus, sequences of messages are considered where bursts with a maximum size can occur, but they are not allowed to follow each other too tightly packed, and they must obey an average rate. A message arrival process is called a *linear bounded arrival process* (LBAP) if

$$N_I(t_0,t_1) \le W + R(t_1 - t_0) \qquad \forall\ t_1 > t_0. \qquad (10)$$

where $N_I(t_0,t_1)$ denotes the number of messages arriving at interface I in the time interval $[t_0, t_1]$.

Considering the remark at the end of section 3.2 it does not surprise: A linear bounded arrival process with the parameters R and W is a $\left(\frac{W-1}{R}, 0\right)$-jitter constrained stream with period $T = R^{-1}$ and minimum distance $D = 0$.

Since for such a stream yields with (2)

$$L = 1 + \left\lfloor \frac{W-1}{RT} \right\rfloor = W\,,$$

and it follows from (9) (withour transformation of the time axis) for the number of events occuring until time t, $N(t)$:

$$N(t) \le 1 + \left\lfloor \frac{t}{T} + \frac{W-1}{RT} \right\rfloor = 1 + \left\lfloor \frac{t}{T} \right\rfloor + W - 1.$$

Since we can omit $\lfloor\ \rfloor$ because $N(t)$ is a natural number a priori it follows

$$N(t) \le W + Rt.$$

This is the same like (10) except the notations and the assumption $t_0 = 0$.

4.3. The Tenet Protocol Suite

The Tenet Real-Time Protocol Suite [3] was developed to meet the real-time demands and guaranteed performance requirements of applications communicating in an internetwork, such as bounds on throughput, delay, or reliability. This protocol suite is a set of communication protocols that can transfer real-time streams with guaranteed quality in packet-switching internetworks. For the Tenet suite, clients must specify their worst-case description of the traffic it will transmit over a so called real-time channel (e.g. network devices, or processors). That description consists of four traffic parameters:

$$\left.\begin{array}{ll} X_{\min} & \text{minimum inter - message time} \\ X_{\text{ave}} & \text{minimum average inter - message time} \\ I & \text{averaging interval} \end{array}\right\} (11)$$

$$S_{\max} \quad \text{maximum message size.}$$

As in section 4.2 we ignore $S_{\max}$ in this paper. Although it is not recognizable at first glance, the traffic has the same properties: it consists of events with a constant rate in principle but varies over an interval. It is obvious that $D = X_{\min}$, $T = X_{\text{ave}}$ with respect to definition 1. Now $N(I)$ denotes the number of messages arriving or sent within an interval with length I. Thus, including the bounds of the interval it holds per definition

$$N(I) \le 1 + \frac{I}{X_{ave}}$$

or with another notation and respecting $N(I) \in \mathbb{N}$:

$$N(t) \le 1 + \left\lfloor \frac{t}{T} \right\rfloor .$$

Furthermore, $1 + \left\lfloor \frac{I}{X_{ave}} \right\rfloor$ is the maximum burst length (note that bursts do not result from earlier but from later arrival of messages). Thus, summarized it follows:
A stream described by the Tenet parameters X_{min}, X_{ave}, I according to (11) is a $\left(0, \frac{I}{X_{\text{ave}}}(X_{\text{ave}} - X_{\text{min}})\right)$-jitter constrained stream with period X_{ave} and minimum distance $X_{\text{min.}}$ according to definition 1.

5. Conclusions and Future Work

A generalized set of parameters to describe jitter constrained periodic streams has been introduced and used to prove some important characteristics and properties for such systems:

- maximum burst size
- lower bound of buffer size to avoid loss of data
- interburstiness
- number of events occurring before time *t*.

Then, it has been shown, that our model subsumes well-known parameter sets to describe traffic in distributed systems.

Future work to be done includes the probabilistic extension of the model. And it includes to apply the model what it was designed for, i.e., to build system components that obey the parameters of the model.

References

[1] ATM. *Traffic Management Specification, Version 4.0.* ATM Forum Technical Committee, atmforum/95-1434R12, 1996.

[2] D. J Anderson. *Metascheduling for Continuous Media*. ATM Transactions on Computer Systems, 11(3): 226-252, August 1993.

[3] A. Banerjea, D. Ferrari, B. A. Mah, M. Moran, D. C. Verma, H. Zhang. *The Tenet Real-Time Protocol Suite: Design, Implementation, and Experiences*. University of California at Berkeley and ICSI, TR-94-059, November 1994.

[4] P. Goayl, S. S. Lam, H. M. Vin. *Determing End-to-End Delay Bounds in Heterogeneous Networks*. Proceedings NOSSDAV ‘95, Lecture Notes in Computer Science, Vol. 1018.

[5] Cl.-J. Hamann. *Jitter constrained periodic streams.* Dresden University of Technology, Dep. of Computer Science, Technical Report SFB 358 - TP G2 - 02/1996.

[6] J. Khan, V. O. K. Li. *Traffic control in ATM networks*. Computer Networks and ISDN Systems, 27(1): 85 - 100, 1994.

[7] K. P. Kontovassilis, J.T. Tsiligaridis; G.I. Stassinopoulos. *Buffer dimensioning for delay/loss-sensitive traffic*. Computer Communications, 18(5): 315-328, May 1995.

[8] E. W. Knightly, D. E. Wrege; J. Liebeherr; H. Zhang. *Fundamental Limits and Tradeoffs of Providing Deterministic Guarantees to VBR Video Traffic*. Proceedings of ACM SIGMETRICS ‘95.

[9] S. S. Lam, G. G. Xie. *Burst Scheduling Networks: Flow Specification and Performance Guarantees*. Proceedings NOSSDAV ‘95, Lecture Notes in Computer Science, Vol. 1018.

[10] J. F. Meyer, S. Montagna; R. Paglino. *Dimensioning of an ATM switch with shared buffer and threshold priority*. Computer Networks and ISDN Systems, 26(1): 95-108, September 1993.

[11] J. Williamson, L. C. Wolf. *Reducing Buffer Space Requirements for Multimedia Data Streams*. IBM European Network Center, Technical Report 43.9504, 1995.

[12] G.-L. Wu, J. W. Mark. *Discrete time analysis of leaky-bucket congestion control*. Computer Networks and ISDN Systems, 26(1): 79 - 94, September 1993.

[13] H. Zhang, D. Ferrari. *Rate Controlled Service Disciplines.* Journal of High Speed Networking, to appear.

[14] H. Zhang, S. Keshav. *Comparison of Rate-Based Service Disciplines*. Proceedings NOSSDAV ‘95, Lecture Notes in Computer Science, Vol. 1018.