Control and Cybernetics

vol. **22** (1993) No. 1/2

Optimal periodic buffer exhaustion strategy

by

Tadeusz Nowicki

Military University of Technology
Warsaw
Poland

A mathematical model of exploitation of communication buffer is described. Periodic buffer exhaustion, very useful for rhytmical work of transmission system, is defined. The problem of choosing an optimal value of exhaustion strategy parameter is formulated. This value minimizes expected unit limit cost of transmission aided by communication buffer. A method to solve the problem is proposed.

1. Introduction

In this paper we assume that a random stream of data comes to a comminication buffer and fills it. Many different strategies can be offered for buffer exhaustion. One of them is periodic strategy. We consider the problem of choosing an optimal value of a parameter connected with buffer exhaustion strategy.

Buffers which are used in computer systems are of hardware or software nature. They are mechanisms supporting data transmission process between different computer devices. Buffers make data transmission easier, and so their role is very important. Proper use of buffers make computer system exploitation more effective. So fixing parameter values of buffer usage is essential for system functioning. The most significant characteristics of buffer usage are: buffer capacity and strategy of its exhaustion. It is very often so that buffer capacity is fixed. The choice of exhaustion parameter values, however, gives designer possibility of data flow control.

In many papers the problem of rhytmical strategy of buffer exhaustion is considered. In particular, in computer networks such kind of buffer usage is smoothing the arhythmical signal sequences Seidler (1979). So, selection of the best period of time between buffer exhaustions is an important problem.

The same problem is considered in case when one processor serves many buffers Arthurs, Stuck (1979), Eisenberg (1972). Processor visits every buffer at one time interval in a cycle. The best moments of processor visits in the cycle could be selected by computer system designer.

In database systems a special work space called differential file is organized in external memory. During database actualization differential file is used as a file for collecting inflowing records in order to deliver them to database files from time to time Kleirock, Levy (1988), Lehman, Severance (1976), Severance (1982). It is often done for higher reliability of database. In this particular case differential file can be thought as a buffer which is exhausted periodically.

Many authors consider the problem of choosing length of waiting time of server during its work Kella (1990), Servi (1986). Others are interested in such system in which waiting time for tasks service is bounded Swensen (1986). Both of the cases above suggest the necessity of non–continuous, and therefore periodic, work of queue (buffer) system.

In many other cases there is no need to wait for buffer being full of data or to make empty too often. Random way of buffer exhausting is not the best at all. Hence, the length T of intervals between successive buffer exhaustions should be fixed. It can happen, however, that a buffer is full of data although time T has not elapsed yet. So, the model of periodic strategy of buffer exhaustion can be desribed in terms of different situations:

a) buffer is exhausted because constant time T has elapsed,
b) buffer must be exhausted after it has been filled with inflowing data, though the constant time T has not elapsed.

After each of the situations above the same strategy a)–b) is continued from the beginning.

Buffer exhaustion is connected with many detailed costs. Most often they have the following sense:

- duration time of different elementary operations,
- number of elementary operations,
- delay of some events, etc.

In this paper we assume that only the time of buffer filling is interesting. The time of buffer exhaustion is omitted and represented only by special detailed costs. Such an assumption makes our results easy to apply not only in computer systems.

Our goal is to choose such a time interval T which minimizes the expected unit limit and total costs of the buffer exhaustion process.

2. Mathematical model

We assume that random stream of records is comming to the buffer. This stream creates the so called renewal process. It means that intervals of time between subsequent records are independent random variables which have identical probability distribution with limited expected value. This value is equal a. Buffer capacity (size) is determined by number of data records which can be contained in the buffer. This capacity is fixed and it is equal k.

Assume

c_1 – cost of one record transmission from the environment to the buffer,

$c_2(l)$ – unit cost of l–sized buffer service,

$c_3(l,n)$ – cost of handling of n records and of transmission from l–sized buffer to its environment when case (a) has occurred,

$c_3^p(l,n)$ – the same cost as above when case (b) has occurred,

c_4 – unit cost of record protection against loss when it arrives during buffer exhaustion.

Functions $c_3(l,n)$ and $c_3^p(l,n)$ are practically concave and monotonicaly increasing for argument n.

We assume that the time of buffer exhaustion can be ommited. Simply, what we consider is not astronomical time but time of buffer filling only. It must be added that the cost of record protection against loss during buffer exhaustion is considered in this model. Let $N(t)$ be the random process which for fixed t is a random variable representing the number of records inflowing to the buffer until time t.

We assume that

$K(t,T)$ is the cost of record handling and transmission until time t while using buffer exhausted periodically,

$K(T)$ is expected, limit and unit cost of record handling and transmission.

Then

$$K(T) = E\left\{\lim_{t\to\infty} \frac{K(t,T)}{t}\right\} \tag{1}$$

and

$$K(t,T) = \sum_{r=1}^{\left[\frac{t}{T}\right]} K_r(T) + \bar{K}(t,T), \tag{2}$$

where

$K_r(T)$ is the cost of record handling and transmission borne during the r–th period of buffer exploitation $[(r-1)T, rT]$,

$\bar{K}(t,T)$ is the same cost of the interval of time $\left(\left[\frac{t}{T}\right]T, t\right)$,

$[b]$ is the entire part of number b.

It is easy to show that

$$\lim_{t\to\infty} \frac{\bar{K}(t,T)}{t} = 0 \tag{3}$$

with probability equal to 1. So we have

$$K(T) = E\left\{\lim_{t\to\infty} \frac{\sum_{r=1}^{\left[\frac{t}{T}\right]} K_r(T)}{t}\right\}. \tag{4}$$

We notice that for every elementary event the system of following inequalities is true

$$\frac{\sum_{r=1}^{n} K_r(T)}{n\frac{n+1}{n}T} \leq \frac{K(t,T)}{t} \leq \frac{\sum_{r=1}^{n+1} K_r(T)}{(n+1)\frac{n}{(n+1)}T}, \tag{5}$$

where $n = \left[\frac{t}{T}\right]$ for $t \geq T$. Inequalities (5) give us the following result

$$\lim_{n\to\infty} \frac{\sum_{r=1}^{n} K_r(T)}{n\frac{n+1}{n}T} \leq \lim_{n\to\infty} \frac{K(t,T)}{t} \leq \lim_{n\to\infty} \frac{\sum_{r=1}^{n+1} K_r(T)}{(n+1)\frac{n}{(n+1)}T}, \tag{6}$$

because $t \to \infty$ follows from $n \to \infty$.

Variables $K_r(T)$, $r = 1, 2, 3 \ldots$, are dependent random variables because time to the first inflowing record in each period can be the remaining time from the moment of arrival of the previous record. In practice we can assume that correlation of random variables $K_i(T)$ and $K_{i+j}(T)$ converges to zero with $j \to \infty$. Additionally, these random variables have limited variance.

For such assumptions the Berstein weak law of large numbers takes the following form

$$P - \lim_{n\to\infty} \left(\frac{\sum_{r=1}^{n} K_r(T)}{n} - \frac{\sum_{r=1}^{n} E\{K_r(T)\}}{n} \right) = 0, \tag{7}$$

where $P - \lim$ means convergence in probability.

It is known from the literature that if the following limit exists

$$\eta(T) = \lim_{n\to\infty} \frac{\sum_{r=1}^{n} E\{K_r(T)\}}{n} \tag{8}$$

then inequalities given below are true

$$P - \lim_{n\to\infty} \frac{\sum_{r=1}^{n} K_r(T)}{n\frac{n+1}{n}T} \leq \lim_{n\to\infty} \frac{K(t,T)}{t} \leq P - \lim_{n\to\infty} \frac{\sum_{r=1}^{n+1} K_r(T)}{(n+1)\frac{n+1}{(n+1)}T}. \tag{9}$$

The following lemma is proved in Knopp (1956).

LEMMA *If there exists a sequence of numbers* $\{x_n\}$ *for which*

$$\lim_{n\to\infty} x_n = \bar{x}_n \quad \textit{and} \quad x'_n = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \textit{then} \quad \lim_{n\to\infty} x'_n = \bar{x}_n.$$

So, using this result, we see that

$$\begin{aligned} \eta(T) &= \lim_{n\to\infty} \frac{\sum_{r=1}^{n} E\{K_r(T)\}}{n} = \lim_{n\to\infty} \frac{\sum_{r=1}^{n+1} E\{K_r(T)\}}{n+1} = \\ &= \lim_{r\to\infty} E\{K_r(T)\}. \end{aligned} \tag{10}$$

This implies more important equalities

$$K(T) = E\left\{\lim_{t\to\infty} \frac{K(t,T)}{t}\right\} = \frac{\eta(T)}{T} = \frac{1}{T}\lim_{r\to\infty} E\left\{K_r(T)\right\} \tag{11}$$

We can show that:

$$\begin{aligned} E\left\{K_r(T)\right\} &= \\ &= E\left\{c_1 N_r(t) + c_2(k)\right\} + P\left\{N_r(T) \le k\right\} E\{c_3^p(k, N_r(T))/N_r(T) \le k\} + \\ &+ P\left\{N_r(T) > k\right\} E\left\{c_3(k,k) + c_5 \int_0^T [N_r(t) - k] I_{\{N_r(t)>k\}} dt\right\}, \end{aligned} \tag{12}$$

where $I_{\{A\}}$ is the so called indicator of the event A,

$$I_{\{A\}}(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases}$$

$N_r(t)$ is the random process representing the number of records incoming to the buffer during r–th period till time t $[(r-1)\cdot T, (r-1)\cdot T + t]$.

It was shown in Nowicki (1992), that

$$\begin{aligned} K(T) &= \frac{1}{T}\lim_{r\to\infty} E\{K_r(T)\} = \frac{1}{T}\left[c_1 \bar{H}(T) + c_2(k) + \right. \\ &+ [1 - F_{S_k}(T)]\, E\left\{c_3^p(k, \bar{N}(T))/\bar{N}(T) \le k\right\} + \\ &\left. + F_{S_k}(T) \cdot \left(c_3(k,k) + c_5 \int_0^T [\bar{H}_k(t) - k] F_{S_{k+1}}(t) dt\right)\right], \end{aligned} \tag{13}$$

where $\bar{N}(t)$ – is the same renewal process as $N(T)$ in which the time elapsing until the first record flows in is the so called limit rest time to renewal,

$\bar{H}(t) = E\{\bar{N}(t)\}$,

$\bar{H}_k(t) = E\{\bar{N}_k(t)\}$,

$$\bar{N}_k(t) = \begin{cases} \bar{N}(t) & \text{if } \bar{N}(t) \ge k, \\ k & \text{if } \bar{N}(t) < k. \end{cases}$$

S_k – is the time in which k records arrive, calculated from the beginning of each period of buffer exploitation,

$F_{S_k}(t)$ – distribution function of the random variable S_k.

Using Jensen inequality Klimov (1986) form concave functions we have

$$\begin{aligned} E\left\{c_3^p(k, \bar{N}(T))/\bar{N}(T) \le k\right\} &\le c_3^p\left(k, E\left\{\bar{N}(T)/\bar{N}(T) \le k\right\}\right) \\ &= c_3^p\left(k, \bar{H}^k(T)\right), \end{aligned} \tag{14}$$

where $\bar{H}^k(t) = E\{\bar{N}^k(t)\}$,

$$\bar{N}^k(t) = \begin{cases} \bar{N}(t) & \text{if } \bar{N}(t) \le k, \\ k & \text{if } \bar{N}(t) > k. \end{cases}$$

Finally, from (13) and (14) we have the upper approximation of function (13) in the following form

$$\begin{aligned}\bar{K}(T) &= \frac{1}{T}\left(c_1\bar{H}(T) + c_2(k) + [1 - F_{S_k}(T)] \cdot c_3^p(k, \bar{H}^k(T)) + \right.\\ &\quad \left. + F_{S_k}(T) \cdot \left[c_3(k,k) + c_5 \int_0^T [\bar{H}_k(t) - k] F_{S_{k+1}}(t) dt\right]\right), \qquad (15)\end{aligned}$$

3. Optimization problem

We want to find value T^* for which

$$\bar{K}(T^*) = \min_{T \in R^*} \bar{K}(T) \qquad (16)$$

Because of fact that functions: $\bar{H}$, $\bar{H}_k$, $\bar{H}^k$, c_3, c_3^p are monotonically increasing, problem (16) can be solved using one of the available numerical methods. In practice, the variable T is always limited from above.

4. Concluding remarks

The problem considered in the paper was complicated in view of necessity of determining the upper approximation $\bar{K}(T)$ of the criterion function $K(T)$. For typical distributions of random process $N(t)$ the function $\bar{K}(T)$ has only one solution. It is easy to find this solution using numerical methods. This is the main reason for not giving more information on the optimization problem. Construction of the formula of criterion function for problem (16) was the main idea of the present paper.

Reference

ARTHURS E., STUCK B.W. (1979) Analysis of a Single–Buffer Loss-Delay System. Operations Research No.1.

EISENBERG M. (1972) Queues with Periodic Services and Changeover Times, *Operations Research*, **20**.

KELLA O. (1990) Optimal Control of the Vacation Scheme in an M/G/1 Queue. Operations Research No.4

KLEIROCK L., LEVY H. (1988) The Analysis of Random Polling Systems. Operations Research No.5

KLIMOV G. (1986) Probability Theory and Mathematical Statistics. Mir Publishers, Moscow.

KNOPP K. (1956) Szeregi nieskończone. PWN, Warszawa.

LEHMAN, G., SEVERANCE D. (1976) Differerntial Files: Their Application to the maintenance of Large Database, ACM Transactions on Database Systems, 2.

NOWICKI T. (1992) Periodyczna strategia opróżniania buforów komunikacyjnych. *Bulletin of Military University of Technology*, 9.

SEIDLER J. (1979) Analiza i synteza sieci łączności dla systemów teleinformatycznych. PWN, Warszawa.

SERVI L.D. (1986) D/G/1 Queues with Vacations. *Operations Research*, 4.

SEVERANCE D. (1982) A Practical Guide to the Design of Differerntial Files for Recovery of On-Line Database. ACM Transactions on Database Systems, 4.

SWENSEN A.R. On a GI/M/c Queue with Bounded Waiting Times, Operations Research No.6