13th INTERNATIONAL TELETRAFFIC CONGRESS
COPENHAGEN 1991

DISCUSSION CIRCLES

TELETRAFFIC MODELS IN ADVANCED SWITCHING SYSTEMS SOFTWARE ENGINEERING

B.S. GOLDSTEIN

Leningrad Institute of Telecommunication (LONIIS)
USSR, 196128, Leningrad, Varshavskaya st., 11

The report introduces the teletraffic models for the 5-layers methodology for telecommunication software engineering {R, A, S, P, C}.The methodology is oriented on the CCITT language SDL-88.The class of telecommunication software teletraffic model corresponding to software development phases (layers A and S) is discussed: mixed preemptive and non-preemptive priority model with the applications to the analysis and optimization of real time switching software systems. Two modifications of such complex priority system are analyzed. Optimization algorithm for telephone program priority fixing is introduced.

1. INTRODUCTION

In today's telecommunication software design the implementation of analytical models and methods becomes very important. The 5-layers methodology for switching systems software engineering {R,A,S,P,C} is used at LONIIS. This methodology is based on five design levels: R-technical reqierements, A - architecture, S -specifications, P - programming, C - coding and debugging. It is oriented on the CCITT language SDL-88 [1].

The important part of levels A and S design is the development of teletraffic models for software processes, blocks and subsystems.

These models are the basis for implemented software characteristics (performance and quality) analysis. In this way the representation of the complicate queuing systems with mixed priorities comes to the deadlock due to the increase of the event phase space dimension. So we use direct methods, based on "physical" interpretation of service processes. Particulary, the conservation law must be used. It reflects the most significant properties of ergodic random processes. The main idea of this law is the conservation of work that must be done by service unit and is invariant to a wide class of service disciplines. This law was estimated by L.Kleinrock [2] for priority queue systems without losses $\overline{M}_k/G_k/1$ and was spread for a wider class $\overline{GI_k}/G_k/1$ by Bronshtein and Duchovny [3]. The organization of the rest of the paper is as follows. Section 2 presents a detailed mixed (preemptive and non preemptive) priority queueing model for telecommunication software. Two modifications of such complex priority strategy is proposed in section 3. Section 4 presents the special direct optimization algorithm. Numerical results are given for analysis and optimization.

2. MATHEMATICAL MODEL

The mathematical model has as a basis the following simple geometrical structure, given in fig. l. Two priority indexes (k, m) are introduced, where k - number of preemptive priority, $k = \overline{1, N}$, and m - number of non-preemptive priority inside the k preemptive priority level. The number of non-preemptive priorities inside k-level is M_k and the whole number is

$$\sum_{k=1}^{K} M_k = N .$$

As shown on fig. l the telephone program of priority (k, m) can be interrupted by any request for telephone program of priority (1,1), …, (k-l, M_{k-1}) and can interrupt programs of priorities (k+1,1),..., (k, M_k) .Programs with priorities (k, 1), ...,(k ,m), ..., (k, M_k.) are served according to the non-preemptive priority discipline.

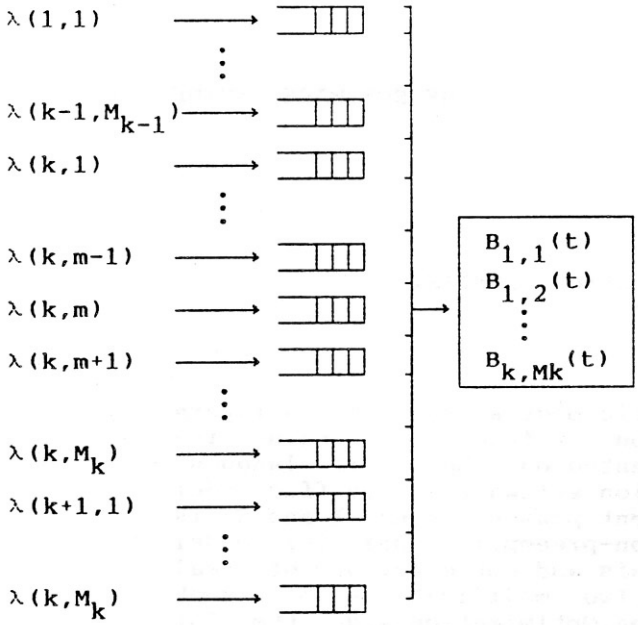

Figure 1

λ(k, m) - average arrival rate of requests for program of priority (k, m). The service - time distribution function at priority (k, m) is denoted by $B_{k,m}(t)$. The mean and second moment of $B_{k,m}(t)$ are b(k, m) and $b^{(2)}(k, m)$ respectively. Let R be the total utilization of the system,

$$R = \sum_{i=1}^{k} \sum_{j=1}^{M_i} \lambda\,(i, j)\; b\,(i, j) = \sum_{i=1}^{k} \sum_{j=1}^{M_i} \rho\,(i, j) < 1$$

So we can describe such priority system by notation

$M = \{M_1, M_2, \ldots, M_k\}$.

For this system average total delay of program performance T(k, m) is equal to the average waiting time W(k, m) & average processing time V(k, m). Average processing time V(k, m) start from the beginning of program of prority (k, m) up to the finish of this program with possible interruption.

Then

$$T(k,m)=W(k,m)+V(k,m) \qquad (1)$$

V(k, m) value doesn't depend on lower priority requests (k, m+l), ... ,(K,M_k), thus it's possible to ignore them.

I proposed a method for evaluation V(k, m) .Telephone programs of priorities (1, 1), … , (k-1, M_{k-1}) have preemptive priority considering the program of priority (k, m). Let's consider the same features for programms of priorities (k,l), (k,2), ... , (k, m-l). For this case there it is aclassical queueing theory result [4] that

$$V'(k,m) = \frac{b(k,m)}{1-\sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m-1)}\rho(i,j)},$$

where

$$\varphi(x,y) = \begin{cases} M_i, when & i < x, \\ Y, & when \quad i = x. \end{cases}$$

In our system V(k, m) can be evaluated by following way. During the time after the last interruption in telephone program of priority (k, m) requests for telephone programs of priorities (k, l), ... , (k, m-l) can be stored in processor unit (fig. 2).These requests can't interrupt the program of priority (k, m) and their total rate is

$$\sum_{j=1}^{m-1}\lambda(k,j)$$

and their average service time

$$\sum_{j=1}^{m-1}\rho(k,j) / \sum_{j=1}^{m-1}\lambda(k,j).$$

The total number of these requests is

$$\theta(k,m) \quad \sum_{j=1}^{m-1}\lambda(k,j).$$

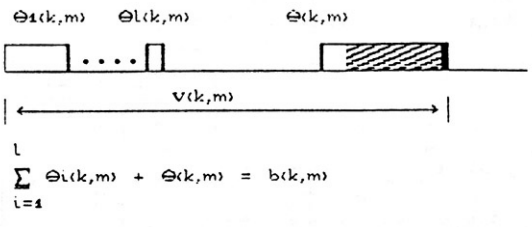

Figure 2

In addition in this modificated system with only preemptive priori ties new requests for telephone programs of priorities (1, 1), ... , (k, m-l) would be served. It's taken into account by factor

$$\frac{1}{1-\sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m-1)}\rho(i,j)}.$$

Then, one can express the average processing time as

$$V(k,m) = \frac{b(k,m) - \theta(k,m)\sum_{j=1}^{m-1}\rho(k,j)}{1 - \sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m-1)}\rho(i,j)} \qquad (2)$$

Let's determine value θ(k, m). Let the time of program of priority (k, m) performance equal t. This performance would not be interrupted with probability

$$\exp\left\{-t * \sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\lambda(i,j)\right\}$$

and in this case θ(k, m) = b (k, m). For priorities (1, 1), ... ,(1, M_i) this probability equal 1 and θ (l, m) = b (k, m).With probability

$$1 - \exp\left\{-x * \sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\lambda(i,j)\right\}$$

the performance of the telephone program of priority (k, m) would be interrupted and time θ(k, m) would be less the value x.

Then:

$$\theta(k,m) = \begin{cases} b(k,m), \text{ when } k = 1 \\ \dfrac{1}{\sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\lambda(i,j)} * \int_0^{\infty}\left[1 - \exp\left\{-t * \sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\lambda(i,j)\right\}\right] d\ B_{k,m}(t), \end{cases} \qquad (3)$$

$$\text{when} \quad k > / < 1.$$

The expressions for average waiting time have been found by the analogous way:

$$W(k,m) = \frac{\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{M_i}\lambda(i,j)\, b^{(2)}(i,j)}{\left[1 - \sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m-1)}\rho(i,j,)\right]} + \frac{\left[1 - \sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\rho(i,j,)\right]\ \sum_{j=m+1}^{M_k}\rho(k,j)\,\upsilon(k,j)}{\left[1 - \sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m)}\rho(i,j,)\right]} \qquad (4)$$

where

$$\upsilon(k,j) = \begin{cases} \dfrac{b^{(2)}(k,j)}{2\,b(k,j)}, & \text{when } k = 1 \\ \dfrac{b(k,j) - \theta(k,j)}{b(k,j)\sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\lambda(i,j)}, & \text{when } k = 1 \end{cases} \qquad (5)$$

Numerical results for priority system M = {3, 2, 5, 6, 4} and b(k, m) and λ(k, m)=λ for all k, m are shown on fig.3.

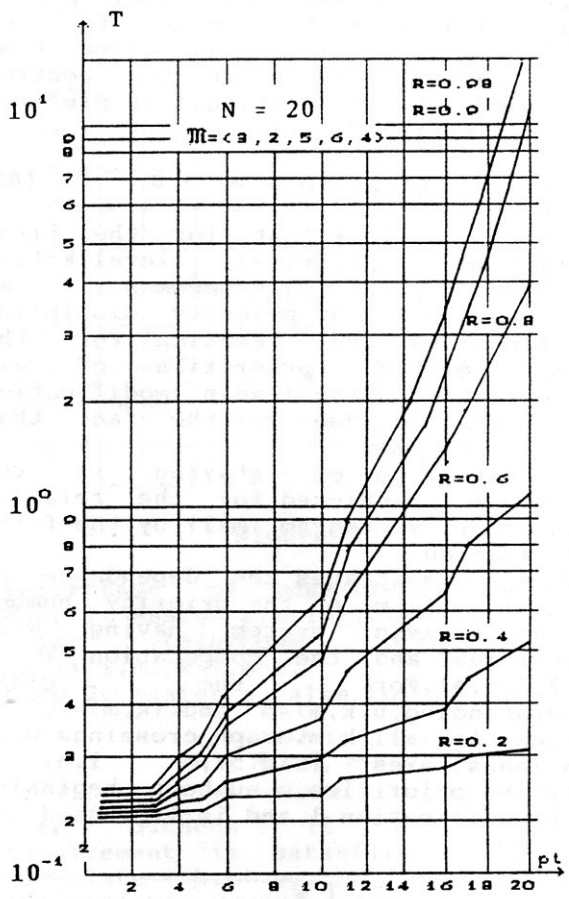

Figure 3

3. TWO MODIFICATIONS OF MIXED PRIORITY SYSTEM.

Enhancement of interrupt request priority (k, m) and its posthandling with relative most significant priority (k, m) within specified absolute level, k (modification 2), is an alternative to the discussed (modification 1) discipline posthandled at its own level. Difference between the two modifications of mixed priority systems shown in fig. 4.

For modification b average total delay $T_b(k, m)$ of program performance of a random priority (k, m) is a determinedby the forms that may be derived either directly from (1) - (5) or applicating the general results [2] to the system described (fig. 4):

$$V_z(k,m) = \frac{b(k,m)}{1-\sum_{i=1}^{k-1}\sum_{j=1}^{M_i}\rho(i,j)} \tag{6}$$

$$W_z(k,m) = \frac{\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{M_i}\lambda(i,j)\,b^{(2)}(i,j)}{\left[1-\sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m-1)}\rho(i,j,)\right]\left[1-\sum_{i=1}^{k}\sum_{j=1}^{\varphi(k,m)}\rho(i,j,)\right]} \tag{7}$$

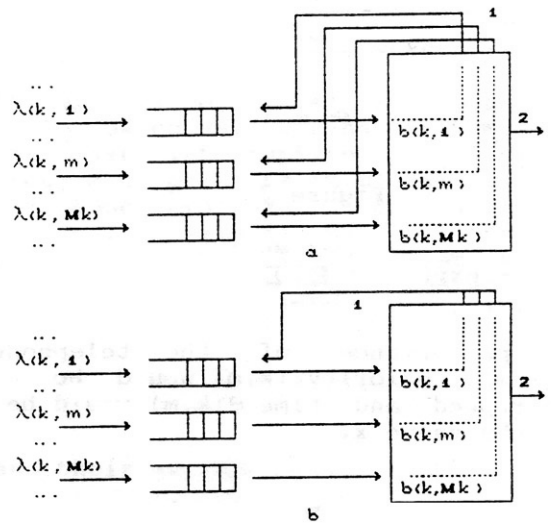

Figure 4

In the telecommnunication software with mixed priorities the effectiveness of modification depends on the control processor structure and capabilities, technical considerations, the nature and hierarchy of the executed tasks. A priory it is reasonable to compare the two modifications by minimum averaging time of the request staying in the control processor, i.e. modification 1 is prefered to modification 2 if:

$$T(k, m)=T_2(k, m) - T_1(k, m) > 0 \qquad (8)$$

If should be clear that for the first preemptive priority level, k=l, in equality (8) becomes an identity, i.e., both priority discipline modifications are identical. For the least relative priorities of all preemptive priority levels modification 2 is preferred due to the fact that $b(k, m) \geq \theta(k, m)$.

The less time of staying in the processor is achieved for the relative most priority program (m=l) by the first: modification.

Figure 5 demonstrates the dependence of .T(k, m)/b(k, m) on the priority number of the given system having N=20 priorities and the computation M = {3,2,5,6,4}. For the clear understanding, b(k, m)=b and λ(k, m)=λ, is assumed for all k, m. Graph crossings with abscissas axes (fig.5), i.e. , relative priorities, m, numbers beginning with modification 1 and is expressed as follows:

$$\overline{m} = \left]\frac{1}{2\rho} - (\rho + 2d - \sqrt{\rho^2 + 4d(d - \rho M_k)})\right[, \qquad (9)$$

where signed], [denotes an overflow adjustment of the nearest integer.

$$d = 1 - \beta\sum_{i=1}^{k-1} M_i, \quad k = \overline{2K}.$$

In physical sense the dependence shown in fig. 5 means that modification 2 equals to some extent the time during which requests with different relative priorities within each absolute level may stay in the processor. In some cases this limit is compensated by preemptive priority levels available providing required time hierarchy for different requests staying in the processor.

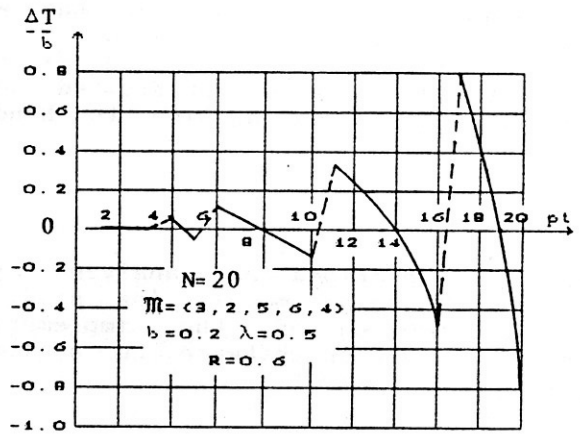

Figure 5

4. OPTIMIZATION

When a telephone program is interrupted, we need some processor resources (processing time and RAM) to store the appropriate attributes of this program. So we can define the optimum priority system: with a specified set of program modules and there characteristics the priority strategy that provides all the time constraints of the following type

$$\{T(i,j) \le t_{\max}(i,j)\}, \quad i = \overline{1,K}\ ,\ j = \overline{1,M_k} \tag{10}$$

is optimum at the/least number of priority levels K.

Consider the order of specifying permissible arrival time values t_{max}(k, m). These values are determined by a given operation quality of the switching node and its software structure (the order of telephone algorithms decomposition to program modules and their quantity).

Denote the available program module set by P = {p_1, p_2, ... , p_N}. A rearrangement of this set is to be defined, P_i = {p_{i1}, p_{i2}, ... , p_{iN}}, that specifies the increase of permissible processing times and their durations needed to start those programs.

A precedence ratio may be introduced when there are multiple programs P. The expression $p_i \infty p_j$ equals the statement that the p_j - program is controlled or receives information from p - program. This ratio is a transitive one, i.e.,

$\nabla p_i, p_j, p_k \quad p_i \infty p_j, p_j \infty p_k \rightarrow p_i \infty p_k$.

This ratio is neither symmetrical nor reflexive. The two programs are independent at $p_i \infty p_j$ and $p_j \infty p_i$.

Multiple switching node programs have a number of given time intervals t_{max} that limits execution of some definite program sequences, connected with node equipment state changes,the accepted signaling system exchanging control and interaction signals with adjacent nodes and stations, subscriber services of a definite node. Permitted intervals include the time period from taking of the receiver to the ready -to-receive signal, the time between dialled digits of a telephone number (e.g.,500ms for connection with step-by-step exchange), the time between

separate control pulses (e.g.,45+5ms with crossbar exchange), intervals between charge accounting pulses, time periods between dialing pulses, etc..

Some tolerable time intervals t_{max} are specified minding the following consideration. Possibility of obtaining answerers to operating stuff requests should be provided. For that purpose the switching node is equipped with an operator console provided with special working programs. The work of these programs may be appreciated as satisfactory, if the time required to answer the operator console request is not too large comparing to the monopoly time of the processor serving only this operator console request. The quantity criteriummay be expressed as follows: the answering time to an operator console request at peak loading must not exceed monopoly time multiplied by 20 or5 seconds.

Specified tolerable time intervals t_{max} are likely to be pair programs related by precedence ratio $p_i \infty p_j$ . .Generally, these programs are not related by direct precedence ratio, i.e., $\exists p_e$: $p_i \infty p_e \infty p_j$. Then, specification of permissible holding time reduces to finding maximum t values satisfying all the in equalities of the type

$$\dots \sum_{i=1}^{N} C_{k,l}^{i} \quad t \le t_{per} \quad (k,l),$$

where

$$C_{k,L}^{i} = \begin{cases} I, \textit{ if both ratios, } p_k \infty p_i \textit{ and } p_i \infty p_l \textit{ are feasible} \\ 0, \textit{ if at least one of these ratios boes not exist.} \end{cases}$$

Since the request flow number, N programs is a finite one the priority optimization problem may become a simple listing of best versions possible and choicing the best decomposition version M = {M_1, M_2, … ,M_k} providing that (10) would be executed at least for a single version. It is not difficult however to show that total number of possible combinations is 2^{N-1} and the listing method is not feasible with larger N - values.

Directed optimization algorithm is offered allowing to achieve a result considering minimum persentage of possible combinations.

Step 1. Assign a priority (1,1) to the first telephone program from (1,1), i.e., k=l, m=l.

Step 2. Assign a value to

$$M_k = N - \sum_{i=1}^{k-1} M_i$$

Step 3. Check if the $M_k \ge 0$ requirement is satisfied if it does calculate T(k, m) by the forms (1) - (5). Otherwise produce a message on constraint system inconsistency (10) and complete the algorithm.

Step 4. Check the requirements of (10) for T(k, m). If (10) for T(k, m) is not satisfired, i.e., time limits mismatch this priority combinations, assign a value to $M_k=M_k$-l and go back to the step 3. If (7) for T(k, m) is satisfired, go on to the step 5.

Step 5. Assign the priority (k, m+l), i.e., m=m+l to the next telephone program.

Step 6. Check the requirements of m+l=M_k. If this equality is not satisfired, go back to step l. The first telephone program from 1, 2, ... , N to the step 3. Otherwise go on to the step 7.

Step 7. Check if the equality

$$\sum_{i=1}^{k} M_i = N$$

is satisfied. If it does, assume K=k and print out the $\{M_1, M_2, \ldots, M_k\}$ value and complete the algorithm. Otherwise, assign the priority (k+1,1), i.e., k=k+l, m=l to the next telephone program from 1, 2, ... , N and go back to the step 2.

The offered algorithm may be given a clear physical interpretation: processor handles the telephone program invoke requests in the way providing all time ratios with minimum margins, herein using minimum part of RAM to store interrupted task and minimum processing time to handle single.

CONCLUSIONS

We have described and analyzed two priority queueing models for studying the performance of the implemented telecommunication software. The analysis showed that the use of optimal mixed priority strategy with preemptive and non-preemptive priorities is an ideal solution for telecommunication software engineering.

REFERENCES

[1] CCITT Recomendation Zl00 (COM X-R15-E), GENEVA, 12-23 January, 1987

[2] L.Kleinrock. Queueing Systems, Vol.l:Theory, New York, Wiley, 1975

[3] O.Bronstein, I.Duchovny. Priority service models in computing systems; Moscow, Nayka, 1976 (in Russian).

[4] U.Herzog. Optimal sheduling strategies for real-time computers IBM Journal of Research and Development, 1975, p.494-504.