Experimental Evaluation of Regular Events Occurrence in Continuous-time Markov Models

Vaclav Vais
Department of Computer Science and Engineering
University of West Bohemia
Plzen, Czech Republic
vais@fav.zcu.cz

Stanislav Racek
Department of Computer Science and Engineering
University of West Bohemia
Plzen, Czech Republic
stracek@kiv.zcu.cz

Abstract—**Continuous-time Markov process is widely used abstract tool to construct high-level models of complex computer systems in order to evaluate either performance or reliability parameters of the system. Utilization of the continuous-time Markov process is based on an assumption of exponential distribution of the time between random events influencing behavior of the modeled system. Another probability distribution of this time needs an adaptation of the original model. This article uses a representative example to evaluate precision of the modeled system parameters when the exponential distribution of burning time of events is replaced with another probability distribution, including regular distribution of events.**

Keywords; Markov model, influence of probability distribution, experimental evaluation

I. Introduction

A powerful tool for analyzing some probability based systems and or problems from the area of computer science is the probabilistic model based on the mathematical theory of the (stochastic) *Markov processes*. For a thorough review of the basic theory, please consult e.g. [2]. From a computer researcher point of view, Markov process is a kind of finite automata, where a transition between two states is caused by random event, i.e. the time duration of every state is a random variable. When the state is reached, the transition is *fired.* Every transition (an edge in the graph describing the process) has assigned a value of *transition rate*. This value can be interpreted in two ways: (i) it is conditional (transition is fired) frequency of the (subsequent) transitions, and (ii) it is the parameter of the exponential probability distribution of the time interval between the transition firing and transition occuring (denoted here as time interval of the *transition burning*). The model (i.e. its state-transitions graph) can be easily transformed into a set of linear differential equations from which time dependent state probabilities $p_0(t)$, $p_1(t)$, … can be computed using conventional methods.

Markov models are used in two basic categories. First category contains models with one or more *absorbing states*, i.e. states without an output edge. It is apparent that these models have "limited time of life". Time dependent probabilities of model states are computed directly from the corresponding set of differential equations, then the target parameters can be determined, usually as a linear combination of some state probabilities. Markov models from the second category have "infinite life" (i.e. no absorbing states) and here the asymptotic probabilities (i.e time independent limits $p_0 = p_0(\infty)$, $p_1 = p_1(\infty)$, ...) of model states can be computed from a set of linear algebraic equations. Subsequently significant parameters can be determined using known values of the model states asymptotic probabilities. The analyzed case used in this article falls into the second mentioned category. Description of Markov models utilization in the area of computer science can be found e.g. in [1], [3], [4].

Utilization of Markov processes is limited by the assumption of exponential probability distribution of the duration of any transition burning. Exponential distribution is quite "irregular", i.e. its standard deviation has the same value as its mean value (both are $1/\lambda$, where λ is (the single) parameter of the distribution). In some applications (see e.g. [7]), the Markov model is constructed and utilized even when the modeled system time behavior is influenced by more regular events, e.g. by built-in tests with a quite regular period. This article has the aim to use a representative example to evaluate numerically a deviation which occurs when we use Markov model and the assumption of exponential distribution of events burning time is not quite valid.

II. SYSTEM TO BE EVALUATED

For the analysis we have chosen classical model of cooperating parallel processes. The assumed computational environment can be e.g. symmetrical multiprocessor system consisting of n processors and a shared memory. We have n computational processes, every process has its own processor. The processes cooperates using one *critical section* containg all the shared data of processes. All processes have the same program describing their cyclic behavior: local computation without any interaction with another process, then computation within the critical section, etc. The computation inside critical section needs to be locked, i.e. only one process in one time can be in this part of (the shared) program. All processes have the same time behavior with the following parameters:

- λ … mean conditional frequency of a process local computation, i.e. reversed value of the mean time of the local computation,
- μ … mean conditional frequency of any computation inside critical section, i.e. reversed value of the mean time of this computation.

The described example can serve as a model of parallel computation based on utilization of data parallelism – a process works on separate (local) piece of a large set of data, then updates (global) result of computation. This activity is performed periodically until all the set of data is exhausted. Time intervals of the processes behavior needs then to be taken as random variables due to different values of data processed within different cycles of activity. When

$$(n-1) / \mu < 1/ \lambda$$

then the ideal (linear) speedup of parallel computation $s_{max} = n$ can be reached assuming deterministic time behavior. When the processes have a random time behavior, their conflicts (i.e necessary synchronization at the input of critical section) decreases the reachable value of the speedup. Keeping the above stated condition, maximum frequency (i.e. frequency without conflicts) of every process computation is as follows:

$$f_{max} = 1/ (1/\lambda + 1/\mu)$$

Due to the conflicts, the real frequency of computation f is decreased compared to f_{max}, so we can define a speedup degradation coefficient d as the ratio:

$$d = f_{max} / f$$

Corrected value of the speedup can be then expressed as

$$s = s_{max} /d = n/d .$$

When the time intervals of local computation and the time intervals within critical section have the exponential distribution (i.e. they are quite irregular), then the analytic solution for the degradation coefficient d can be found (see the next section). Variables λ and μ are then taken as the parameters of the corresponding exponential probability distribution and the reversed values $1/\lambda$ and $1/\mu$ serve as the mean values of the corresponding distribution. But an assumption of exponential distribution (i.e. total irregularity) of the processing time intervals can be questionable, especially in the case of time spent inside the critical section. In the analyzed example (model of parallel computing with data parallelism utilization) this time corresponds a global result update, which operation can be quite regular. That is why in section IV an influence of increased regularity of the time spent inside the critical section will be taken into account, still using (modified) Markov model. In section V an influence of combined regularity of the both processing times will be computed by means of discrete-time simulation model.

The basic Markov model (see below, Sec.III) is general enough and it can be used as an abstract model of many systems or problems from the area of the applied computer science. For example we can use it for a closed queuing network with one server (here μ is the serving rate) and n clients which are non-stop generating (with the rate λ) their requests to be processed at the server. Here the degradation coefficient d reflects time lost with clients unproductive waiting for the server to start their request. Another example is from the area of fault-tolerant systems. Highly available information system uses n (identical) servers. The fault rate of a server is then λ . Rate of repairs is μ, these are performed in a sequence (one repairman assumption). In both given examples, the assumption that λ represents parameter of exponential probability distribution is acceptable (irregular corresponding times), but similar assumption for μ is questionable (more regular time of single services or more regular time of single repairs).

III. ANALYTICAL SOLUTION

The Markov model of an example described above can be represented by a simple state diagram.

Figure 1: The basic Markov model

This system can be represented by matrix equation

$$\mathbf{A}\,\mathbf{p} = \mathbf{0}$$

where $\mathbf{p} = (p_0 ,\ p_1, \ldots\ldots, p_{n-1} ,\ p_n)^T$ is a column vector of asymptotic probabilities and $\mathbf{A}$ is a matrix coefficient of the system

$$\mathbf{A} = \begin{bmatrix} n.\lambda & -\mu & 0 & . \ . & 0 & 0 \\ -n.\lambda & (n-1).\lambda+\mu & -\mu & . \ . & 0 & 0 \\ 0 & -(n-1).\lambda & (n-2).\lambda+\mu & . \ . & 0 & 0 \\ . & . & . & . \ . & . & . \\ 0 & 0 & . & . \ . & -\mu & 0 \\ 0 & 0 & 0 & . \ . & \lambda+\mu & -\mu \\ 0 & 0 & 0 & . \ . & -\lambda & \mu \end{bmatrix}$$

Generally any system of linear equations representing Markov process by this way is linearly dependent. The rank of matrix

A is *n* (number of states minus 1). This degradation will be eliminated by replacing any equation by equation:

$$p_0 + p_1 + \ldots\ldots + p_{n-1} + p_n = 1$$

(system will always be in some state with probability 1).

The form of matrix **A** clear the way for deriving of analytical solution of vector **p**. Suppose that a value of p_0 is known. Due to the first row of matrix **A** p_1 can be expressed straightly in terms of p_0, due to the second row p_2 can be expressed straightly in terms of p_0 and p_1, etc. After a sequence of algebraic transformations all the probabilities p_i can be expressed in terms of p_0 (in this special case, not in general):

$$p_i = p_0 \cdot \left[\prod_{j=0}^{i-1} (n-j) \right] \left(\frac{\lambda}{\mu} \right)^n$$

Then it follows:

$$p_0 = \frac{1}{1 + \sum_{i=1}^{n} \left\{ \left[\prod_{j=0}^{i-1} (n-j) \right] \left(\frac{\lambda}{\mu} \right)^n \right\}}$$

The real frequency of computation (i.e. frequency with conflicts) is :

$$f = \frac{\mu.(1-p_0)}{n}$$

The table I. shows numerical values of speedup degradation coefficient *d* based on the analytical model presented above. These results were computed for a representative set of parameters. The meaning of parameters is explained in the previous text. For example the ratio $\lambda / \mu = 0.1$ means ten times longer local computation (in average) compared to the average time of the critical section utilization. Value $d = 1.0424$ for $n = 5$ processes means about 4% longer computation due to the influence of conflicts when accessing the critical section.

TABLE I.

λ/μ	n	2	3	5	10
0,01		1,0001	1,0002	1,0004	1,0010
0,1		1,0083	1,0179	1,0424	1,1575
0,2		1,0278	1,0631	1,1653	1,6979
0,5		1,1111	1,2667	1,7302	3,3335

IV. NUMERICAL SOLUTION

In this section the time intervals inside the critical section will be observed as a serial connection of *k* stages, each of them with the exponential distribution. The mean conditional frequency of any stage is $k.\mu$. Therefore the aggregate time within the critical section has Erlang-*k* distribution in this case. The method of stages is discussed in [5]. In general, the state with non-exponential distribution can be split into some serial-parallel cluster of two or more exponentially distributed stages.

The following state diagram represents Markov model where computation inside critical section is divided into *k* stages with identical mean times of all stages of computation.

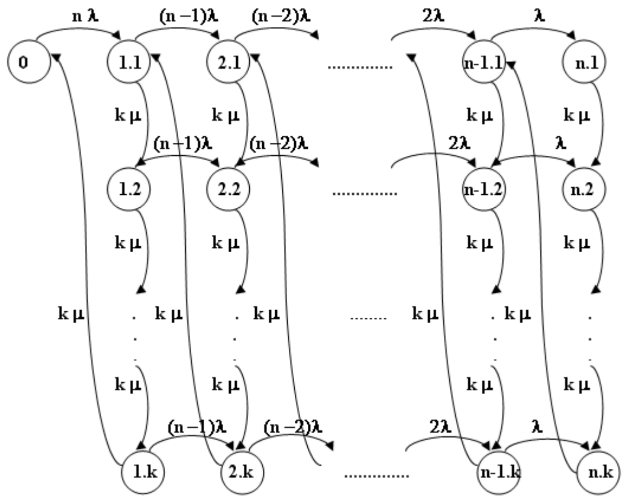

Figure 2: The extended Markov model

Asymptotic probabilities of model states can be computed from a system of $n.k + 1$ linear algebraic equations. We will illustrate it using case for $n = 3$ and $k = 2$.

$$\begin{bmatrix} 3\lambda & -2\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\lambda+2\mu & -2\mu & 0 & 0 & 0 & 0 \\ -3\lambda & 0 & 2\lambda+2\mu & -2\mu & 0 & 0 & 0 \\ 0 & -2\lambda & 0 & \lambda+2\mu & -2\mu & 0 & 0 \\ 0 & 0 & -2\lambda & 0 & \lambda+2\mu & -2\mu & 0 \\ 0 & 0 & 0 & -\lambda & 0 & 2\mu & -2\mu \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} . \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

This matrix equation is not suitable for deriving of the analytical solution. In general, probabilities p_i cannot be simply expressed in terms of p_0. Therefore numerical solution of this system seems to be the best way to obtain vector of asymptotic probabilities ***p***. We used the standard MS Excel software to automatically construct the matrix coefficient **A**, to solve the system of linear equations, and to compute the real frequency of computation *f*, and a speedup degradation coefficient *d.*

The next table shows the numerical values of speedup degradation coefficient *d* based on the numerical model presented above. This results were computed for the same set of parameters as in section III. The lower row presents the relative deviation between the exact analytical solution (i.e. for exponential distribution of the intervals inside the critical section) and the numerical solution for $k = 10$ stages with exponential distribution.

TABLE II.

λ/μ	n	2	3	5	10
0,01	results	1,0001	1,0001	1,0002	1,0005
	deviation	0,00%	0,01%	0,02%	0,04%
0,1	results	1,0048	1,0105	1,0258	1,1134
	deviation	0,34%	0,72%	1,59%	3,81%
0,2	results	1,0170	1,0400	1,1172	1,6725
	deviation	1,05%	2,17%	4,13%	1,50%
0,5	results	1,0759	1,2045	1,6844	3,3333
	deviation	3,17%	4,91%	2,64%	0,00%

V. SIMULATION BASED SOLUTION

The last part of our analysis is aimed to evaluate combined influence of increased regularity of both probability distributions – distribution of the process local activity computation time as well as distribution of the time spent inside the critical section. We can still use the method of stages explained in the previous section. But the resultant Markov model is complex enough and its complexity growth exponentially both with number of processes and number of assumed stages of both activities. So for this case we decided to use a simulation model. The used model is discrete-time and Monte Carlo based, i.e. it uses random numbers to determine single values of duration of modeled processes activities. As the implementation tool we used C-Sim library [6]. The simulation model can be easily verified when we use it for the cases described above in sections III and IV and when we compare the results. If we let to run the simulation program for 10^6 cycles of modeled processes then the relative error of the computed d is about 10^{-3}.

TABLE III.

λ/μ	n	2	3	5	10
0,01	Results	0,9996	0,9998	1,0000	0,9998
	Deviation	0,05%	0,04%	0,04%	0,12%
0,1	Results	1,0041	1,0095	1,0233	1,0935
	Deviation	0,41%	0,82%	1,83%	5,53%
0,2	Results	1,0146	1,0337	1,0932	1,6655
	Deviation	1,28%	2,76%	6,19%	1,91%
0,5	Results	1,0539	1,1525	1,6675	3,3306
	Deviation	5,15%	9,01%	3,62%	0,09%

The table III was computed for the same set of parameters as previous two tables. It shows results obtained when both time intervals were divided into $k = 10$ stages, i.e. the modeled probability distribution was the Erlang`s distribution of the k-th degree. For this case the coefficient of variance C (as a measure of regularity) for both distributions is $C = 1/\text{sqrt}(k) = 0.32$. It approximately corresponds to the Gaussian distribution with the standard deviation about one third of its mean value.

When comparing the results with the previous table(s), we can see the expected influence of the increased regularity – the computed d has better (i.e. decreased) values. When we use simulation model for quite regular values for both time intervals of activity, we obtain the expected result (processes are fully synchronized) $d = 1.0$ with a sufficient precision.

VI. CONCLUSION AND FUTURE WORK

This article uses a representative example to evaluate precision of the system parameters computed using Markov model, when the assumed exponential distribution of burning time of events is replaced with another probability distribution. The chosen example is from the area of parallel processing, but the results can be generalized into another parts of computer science, e.g. queuing networks or fault-tolerant systems. The results show, that for an integral parameter like the evaluated degradation coefficient, the deviation of result created when we replace exponential distribution of events burning time with another distribution is not too large. Within our analysis this deviation did not exceed 20%. In fact, the evaluated degradation coefficient d is combined from probabilities of many states of the used Markov model where deviations in evaluated probabilities of single states can eliminate each other. It is possible to expect, that probability values (or time functions) of some (chosen) states of the model can be influenced by the change of the events burning time probability distribution much more, but it is the matter of our future work.

REFERENCES

[1] J.Hlavička et al.: Číslicové systémy odolné proti poruchám, In: Vydavatelství ČVUT, Praha, Czech rep. (1997), p.330.

[2] P. Mandel: Pravděpodobnostní dynamické metody, In: Academia, Praha (1995), p.181.

[3] D.P.. Siewiorek – R.S. Swartz: The Theory and Practice of Reliable System Design, In: Digital Press, Bedford, USA-MA (1996), p.772.

[4] K.S. Trivedi: Probability and Statistics with Reliability, Queuing ad Computer Science Applications, In: Prentice Hall, USA, (1998), p. 623.

[5] R. Billinton – R.N. Allan: Reliability Evaluation of Engineering Systems, In: Plenum Press, New York, USA (1983), p.349.

[6] http://www.c-sim.zcu.cz

[7] R.Dobias – J.Konarski – H.Kubatova: Dependability Evaluation of Real Railway Interlocking Device, In: Proceedings of IEEE CS, 11th Euromicro Conference on Digital Systems Design, (2008), pp. 228 – 233.