ON THE LAW OF PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

Zotova, Marina A. – 1st category technologist at Design-Technological Bureau for Information Systems Development (PKTB TSKI) of JSC «Russian Railways», Ph.D. student of the department of computer engineering of Russian Open Transport Academy of Moscow State University of Railway Engineering (MIIT), Moscow, Russia.

ABSTRACT

The article provides a study on probabiity distribution of a random variable at the example of response time to enquiry within the data processing center (hereinafter-DPC), considered as a queuing system. To solve this problem the author examines the hypothesis of an exponential distribution, uses in the study queuing theory and simulation methods. The results obtained provide a well-founded guidance to DPC users and ability to correlate computed values with existing standards and growing needs of information system's users.

ENGLISH SUMMARY

Background. *Data Processing Center (DPC) functions in the environment of random stream of requests for processing and of random time of their execution. Studies conducted earlier, allow considering DPC as a queuing system (hereinafter-QS) of M/M/1/∞ type. This reference designation means a simple input stream, exponential service time distribution, one processor, infinite input stream, when a probability of a new enquiry does not depend on the number of enquiries, already existing in the system. Calculating DPC response time (time elapsed from the moment of the enquiry's input from external environment) is quite consistent with experimental data, that is why search for the answer to the question of compliance of DPC operations with requirements of standard ITIL [1], should be based on a variety of QS models.*

Studies on the period of engineering of the first man-machine systems, in which people solved their problems through dialogue with the machine, show that a user has negative emotions if response time exceeds 2 seconds [2, 3]. DPC standard means that developer should negotiate allowable response time with client of the system [1].

Objective. *The objective of the author is to study application of the law of probability distribution of a random variable at the example of DPC response time using hypothesis of exponential distribution.*

Methods. *Due to the fact that DPC response time is a random variable, it can be evaluated using a certain model only as a mean value. Thus in order to obtain a probability that response time will exceed the predetermined value of maximum allowable time of reaction, provided by standards, it is necessary to use methods of queuing theory and simulation. If model M/M/1 / ∞ is applied, then the probability distribution of a random variable (life time of enquiry within the system, defined as the sum of random queuing time and of service proceeding time itself) can be applied based on the considerations given below.*

Results. *If a request comes in QS and finds it free with probability p_0, then stay period in the system is a random variable having an exponential distribution with µ-parameter of service time since there is processing of the request.*

If a request comes in QS and finds another input service request, already existing in the system, with probability p_1, then stay period in the system is a random variable, which is a sum of two exponential variables with µ-parameter of service time. The first exponential value is a remainder of time necessary to finish the proceeding of the request which existed in the system at the moment of the input of a new request, the second exponential value is linked to processing of a new request. According to probability theory it is known [4] that in case of exponential distribution, a remainder of afterservice time has the same exponential distribution with the same parameter µ. Consequently, time distribution of stay in the QS is Erlang distribution of the 2nd order [5].

*If a request comes in the QS and finds that there are **i** requests in it (one for service, (i-1) in the queue) with probability p_i, then stay period in the system is a random variable, which is the sum (i+1) of exponentially distributed random variables with µ-parameter of service time. The first one is a remainder of afterservice time of a request that was at processing at the time of input of new request in the system, the rest are linked with servicing of requests selected from the queue, and of a new request. Afterservice and service of requests received from the queue are random variables having an exponential distribution with the same parameter µ. Consequently, time distribution of stay in the QS is in this case is Erlang distribution of (i+1) order.*

Erlang distribution with order v=(i+1) and parameter µ has density of probability distribution [5]:

$f(t) = \mu(\mu t)^{v-1}e^{-\mu t}/(v-1)!,\ t\geq 0,\ v>0.$ (1)

And that means that density of probability distribution of a random variable (stay period of a request in QS) $f_{reaction}(t)$ is defined by formula:

$$f_{reaction}(t)=\sum_{i=1}^{\infty}p_{i-1}\mu(\mu t)^{\nu-1}e^{-\mu t}/(\nu-1)!,t\geq 0. \qquad (2)$$

Summation is made from 1 to infinity, because QS M/M/1/∞ receives an infinite input stream, thus theoretically there is a non-zero probability that there is an arbitrarily large number of requests.

Probabilities p_i, i=0,1,2,3,... are calculated by formulas given in queuing theory [5]. When a request comes in a system, we should talk about conditional probabilities $\tilde{p}_i$, i=0,1,2,3,... and the situation that the request will meet there i requests at the moment of its input. In case of closed QS these conditional probabilities differ from the stationary probabilities of finding the system in state i. However, as it is shown in [6], for M/M/1/∞ probabilities $\tilde{p}_i$ coincide with stationary probabilities of finding QS in state i.

It is easy to see that it is quite difficult to use cumbersome formula (2) for engineering calculations of a probability that a random variable exceeds a predetermined value, since it is necessary to perform numerical integration. Therefore we shall use the methods of simulation (statistical modelling) [7]. Namely, we take a general-purpose simulation system of discrete complex systems GPSS [8].

QS model in GPSS language looks relatively simple and it is shown in Pic. 1.

Average query time in DPC mainframe is 0,0189 s for an option with performance of 5939 Mips. The model takes 0,1 ms as a time unit, so the first parameter of the block GENERATE, simulating simple request stream in the QS, is equal to 1890 model time units, and the first parameter of the block ADVANCE, simulating service time, is equal to 189 time units. With these parameters, load factor of a processor QS1 theoretically should be equal to 0,1.

```
*   MODEL  OF   OPEN- END  Queuing    system    type    M/M/1/  infinity
* 1     MODEL  TIME    UNIT    = 0.1       MILLISECOND OF REAL TIME
SIMULATE
  EXPON FUNCTION   RN1,C24                EXPONENTIAL DISTRIBUTION           ( AVERAGE =1)
0,0/.1,.104/.2,.222/.3,.355/.4,.509/.5,.69/
0.6,.915       /.7,1.2/.75,1.38/.8,1.6/.84,1.83/.88,2.12/
0.9,2.3/.92,2.52/.94,2.81/.95,2.99/.96,3.2/.97,3.5/
0.98,3.9/.99,4.6/.995,5.3/.998,6.2/.999,7/.9997,8.0
  XTIME TABLE     M1,0,1000,50
INPUTGENERATE          1890,   FN$EXPON   POISSON INPUT STREM
QUEUE              SYSTEM                          MEASURING    TIME    IN THE SYSTEM
*   SIMULATE SERVICE IN QS
              QUEUE     PHASE1               MEASURING TIME IN PHASE           1
              QUEUE     LINE1                MEASURING TIME IN QUEUE           1
              SEIZE     CMO1                 OCCUPANCY   OF   ONE   SERVICE      CHANNEL IN QS1
              DEPART    LINE1                END   OF   TIME    MEASURING IN QUEUE 1
  OBSLU ADVANCE    189,FN$EXPON                  SERVICE SIMULATION IN PHASE 1
              RELEASE   CMO1           RELEASE     OF   ONE   SERVICE      CHANNEL IN QS1
              DEPART    PHASE1               END   OF   TIME    MEASURING IN PHASE 1
              DEPART    SYSTEM               END   OF   TIME    MEASURING IN THE SYSTEM
TABULATE        XTIME        HISTOGRAM OF TOTAL TIME
              TERMINATE      1
START     10000,NP                  MODEL OVERCLOCKING
              RESET                                RESET     OF   ACCUMULATED STATISTICS
              START     10000                    BASIC SIMULATION
              CLEAR
  OBSLU ADVA  NCE    900,FN$EXPON                SERVICE SIMULATION IN PHASE 1
              START     10000,NP                   MODEL OVERCLOCKING
              RESET                                RESET     OF   ACCUMULATED STATISTICS
              START     10000                    BASIC SIMULATION
              END
```

Pic. 1. GPSS-model to receive histogram of DPC response time.

The simulation results show that the experimentally obtained (by simulation) load of a device FACILITY/UTIL is 0,099, that is almost equal to the theoretical one.

Histogram of a random variable (stay period of request in the system) is described by lines:

```
TABLE                  MEAN     STD.DEV.        RANGE            RETRY FREQUENCY CUM.%
 XTIME               209.869   206.684                              0
                                         0.000   -      100.000              3808    38.08
                                       100.000   -      200.000              2271    60.79
                                       200.000   -      300.000              1512    75.91
                                       300.000   -      400.000               898    84.89
                                       400.000   -      500.000               585    90.74
                                       500.000   -      600.000               381    94.55
                                       600.000   -      700.000               207    96.62
                                       700.000   -      800.000               129    97.91
                                       800.000   -      900.000                84    98.75
                                       900.000   -     1000.000                46    99.21
                                      1000.000   -     1100.000                34    99.55
                                      1100.000   -     1200.000                20    99.75
                                      1200.000   -     1300.000                10    99.85
                                      1300.000   -     1400.000                 6    99.91
                                      1400.000   -     1500.000                 3    99.94
                                      1500.000   -     1600.000                 3    99.97
                                      1600.000   -     1700.000                 1    99.98
                                      1700.000   -     1800.000                 0    99.98
                                      1800.000   -     1900.000                 0    99.98
                                      1900.000   -     2000.000                 1    99.99
                                      2000.000   -     2100.000                 0    99.99
                                      2100.000   -     2200.000                 0    99.99
                                      2200.000   -     2300.000                 1   100.00
```

The average stay period in the system is 209,869, root-mean-square deviation (square root of dispersion) of this variable is 206,684. Thus, the coefficient of variation of a random variable is almost equal to 1, that allows (taking into account histogram frequency) making an assumption that the law of distribution of a random variable (stay period in QS) is exponential.

Testing the hypothesis of an exponential distribution with the help of χ^2 (a square of criterion) showed that the hypothesis can be accepted with confidence probabiity of 0,62269368. This result was expected, since the main contribution to the random variable (DPS response time in our case) is made by the first summand to the formula (2), while average number of requests in the system (simulation results) is 0,1, and average time in the system is 209,869 that is slightly different from pure service time.

Results are presented graphically in Pic. 2. Rectangles represent experimental histogram obtained by simulation. Curve of probability distribution density is shown dotted. Model time units are indicated on x-axis.

DPC simulation as of a QS at load level of 0,1 made it possible to receive probabilities that response time will be less than a certain value T (see Table 1).

The results of optimization of DPC mode showed that the DPC operates at a load factor of 0,225537 (Table 2). Simulation at 838 model time units gave average time in the system of 248,047; root-mean-square deviation of this value is 247,601. Thus, we

Table 1

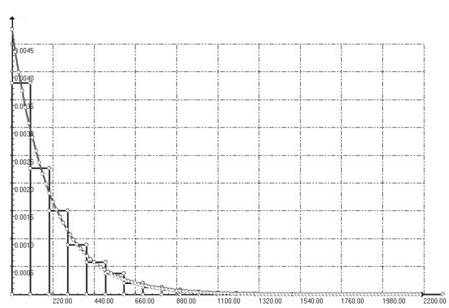
Probabilities that DPC response time will not be more than T, with load factor 0,1

Load factor	Border, **T**	Probability
0,1	< 0,0100 s.	0,3443
	< 0,0200 s.	0,5647
	< 0,0300 s.	0,7145
	< 0,0400 s.	0,8114
	< 0,0500 s.	0,8770
	< 0,0600 s.	0,9212
	< 0,0700 s.	0,9462
	< 0,0800 s.	0,9647
	< 0,0900 s.	0,9763
	< 0,1000 s.	0,9834
	< 0,1100 s.	0,9897
	< 0,1200 s.	0,9930
	< 0,1300 s.	0,9959
	< 0,1400 s.	0,9975
	< 0,1500 s.	0,9986
	< 0,1700 s.	0,9994

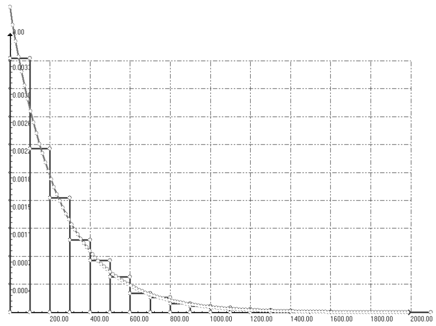
Table 2

Probabilities that DPC response time will not be more than T at load factor 0,225537

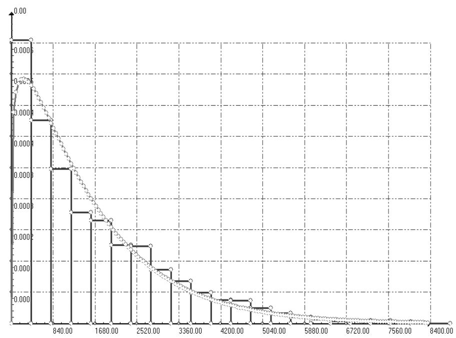
Load factor	Border**, T**	Probability
0,225537	< 0,0100 s.	0,3355
	< 0,0200 s.	0,5516
	< 0,0300 s.	0,7030
	< 0,0400 s.	0,7990
	< 0,0500 s.	0,8675
	< 0,0600 s.	0,9148
	< 0,0700 s.	0,9402
	< 0,0800 s.	0,9600
	< 0,0900 s.	0,9716
	< 0,1000 s.	0,9802
	< 0,1100 s.	0,9866
	< 0,1200 s.	0,9911
	< 0,1300 s.	0,9944
	< 0,1400 s.	0,9968
	< 0,1500 s.	0,9981
	< 0,1600 s.	0,9993



Pic. 2. Load is 0,1. Histogram and exponential distribution.



Pic. 3. Load is 0,225537. Histogram and exponential distribution.



Pic. 4. Load is 0, 9. Histogram and gamma distribution. The order of distribution is 1, 18370930, parameter 1/μ = 1333, 70355714 model time units.

obtain a distribution close to exponential. The results are presented in graphical form in Pic. 3, probabilities that response time will be less than a certain value T are shown in Table 2.

At optimal DPC mode service time for almost all requests will not exceed 0,12 s.

High (0,9) load of DPC was simulated. Average stay period in the system was 1578, 7173 model time units, root-mean-square deviation – 1451,0482 model time units. The results of computing a probability that response time does not exceed a certain border are presented in Table 3.

Graph of probability distribution density of DPC response time at load 0, 9 is shown in Pic. 4.

The hypothesis was tested that a random variable (stay period of a request in the QS) has gamma distribution with density:

$$f(t) = \mu (\mu t)^{v-1} e^{-\mu t} / \Gamma(v-1),\ t \geq 0, \qquad (4)$$

where Γ (v) – gamma-function:

$$\Gamma(v) = \int_0^{\infty} e^{-t} t^{v-1} dt. \qquad (5)$$

For integer values Γ (v) there is a simple factorial (v – 1).

Conclusion. *Studies have shown that in case of mainframe with performance 5939 Mips DPC response time will not exceed 0,64 seconds, even at a very high load (0,9).*

Table 3

Probabilities that DPC response time will not be more than T at load factor 0,9

Load factor	Border, **T**	Probability
0,9	< 0,0400 s.	0,2295
	< 0,0800 s.	0,3942
	< 0,1200 s.	0,5194
	< 0,1600 s.	0,6093
	< 0,2000 s.	0,6929
	< 0,2400 s.	0,7566
	< 0,2800 s.	0,8193
	< 0,3200 s.	0,8633
	< 0,3600 s.	0,8977
	< 0,4000 s.	0,9230
	< 0,4400 s.	0,9423
	< 0,4800 s.	0,9610
	< 0,5200 s.	0,9739
	< 0,5600 s.	0,9824
	< 0,6000 s.	0,9877
	< 0,6400 s.	0,9913
	< 0,6800 s.	0,9932
	< 0,7200 s.	0,9953
	< 0,7600 s.	0,9978
	< 0,8000 s.	0,9992

Keywords: *queuing theory, probability theory, data processing center, probability distribution of random variables, response time, simulation.*

REFERENCES

1. Jan van Bon, et al. IT Service Management: An Introduction [*Vvedenie v IT Servis-menedzhment*]. Russian edition. Translated into Russian under the ed. of Pototsky, M.Yu. IT InfrustractureLibrary, 2003, 228 p. Electronic version: http://www.yessoft.ru/ITIL/ITIL–Introduction%20to%20itSMF (rus) –2003.pdf. Last accessed 20.06.2014.

2. Sherr, A. An Analysis of Time-Shared Computer Systems [*Analiz vychislitel'nyh sistem s razdeleniem vremeni*]. Moscow, Mir publ., 1970, 135 p.

3. Sakman, H. Man-computer problem solving: experimental evaluation of time-sharing and batch processing [*Reshenie zadach v sisteme chelovek–EVM*]. Moscow, Mir publ., 1973, 351 p.

4. Venttsel, E. S., Ovcharov, L. A. Probability theory and its engineering applications [*Teoriya veroyatnostey i ee inzhenernye prilozheniya*]. Moscow, Vysshaya shkola publ., 2000, 480 p.

5. Venttsel, E. S. Probability theory [*Teoriya veroyatnostey*]. Moscow, Nauka, Fizmatgiz publ., 1969. 576 p.

6. Jaiswal, N. K. Priority Queues [*Ocheredi s prioritetami*]. Moscow, Mir publ., 1973, 280 p.

7. Venttsel, E. S. Operations' research [*Issledovanie operatsiy*]. Moscow, Sovetskoe radio publ., 1972, 552 p.

8. Kudryavtsev, E.M. GPSS World. Fundamentals of simulation of various systems [*GPSS World. Osnovy imitatsionnogo modelirovaniya razlichnyh system*]. Moscow, DMK Press publ., 2004, 320 p.

Координаты автора (contact information): Зотова М. А. (Zotova M. A.) – mzotova@zmail.ru.

Статья поступила в редакцию / article received 27.03.2014
Принята к публикации / article accepted 22.05.2014