ON COMPONENT-BASED RELIABILITY IN COMPUTER NETWORKS

S. Minkevičius[1], G. Kulvietis[2]

[1] *Institute of Mathematics and Informatics and Vilnius Gediminas Technical University*

[2] *Vilnius Gediminas Technical University*

[1,2] *Vilnius, Lithuania*

e-mail: stst@ktl.mii.lt

Performance in terms of reliability of computer networks motivates this paper. Limit theorems on the queue length and virtual waiting time in open queueing networks in heavy traffic are derived and applied to a reliability model for computer networks where we relate the time of failure of a computer network to the system parameters.

Keywords: Performance evaluation, reliability theory, queueing theory, mathematical models of technical systems, open queueing network, heavy traffic, the probability limit theorem, queue length of customers, virtual waiting time of customers.

1. INTRODUCTION

Probabilistic models and queueing networks have long been used to study the performance and reliability of computer systems [1, 2] and to analyse the performance and reliability of computer networks and of distributed information systems [3, 4]. In this paper, we will first briefly review the works related to using the queueing theory of computer system reliability, and then present some new results on the estimation of the time of failure of a computer network.

In one of the first papers of this kind [6], the reliability of execution of programs in a distributed computing system is considered, showing that a program, which runs on multiple processing elements that have to communicate with other processing elements for remote data files, can be executed successfully despite that certain system components may be unreliable. In order to analyse the performance of multimedia service systems which have unreliable resources and to estimate their capacity requirements, a capacity planning model using an open queueing network is presented in [9]. In [5] a novel model for a reliable system composed of N unreliable systems, which can hinder or enhance one another's reliability, is discussed. In [10], the management policy of an $M/G/1$ queue with a single removable and non-reliable server is discussed and analytic results are explored, using an efficient Matlab program to calculate the optimal threshold of the management policy and to evaluate system performance. In [11], the authors consider a single machine subject to breakdown and employ a fluid queue model with repair. In [12], the behaviour of a heterogeneous finite-source system with a single server is considered and applications in the field of telecommunications and reliability theory are treated.

In this paper, first we present probability limit theorems on the queue length and the virtual waiting time of customers in heavy traffic for open queueing networks. We consider open networks with the "first come, first served" service discipline at each station and general distributions of interarrival and service times. The basic components of the queueing network are arrival processes, service processes, and routing processes among the different queues. The queueing network that we study has k single server stations, each of which has an associated infinite capacity waiting room. Every station has an arrival stream from outside the network, and the arrival streams are assumed to be mutually independent renewal processes. Customers are served in the order of arrival and after service they are randomly routed to either another station in the network, or out of the network entirely. Service times and routing decisions form mutually independent sequences of independent identically distributed random variables.

2. THE MATHEMATICAL MODEL

Let us consider the mutually independent sequences of independent identically distributed random variables $\left\{z_n^{(j)}, n \geq 1\right\}$, $\left\{S_n^{(j)}, n \geq 1\right\}$ and $\left\{\Phi_n^{(j)}, n \geq 1\right\}$ for $j = 1, 2, \ldots, k$; defined on the probability space. The random variables $z_n^{(j)}$ and $S_n^{(j)}$ are strictly positive, and $\Phi_n^{(j)}$ have support in $\{0, 1, 2, \ldots, k\}$. We define $\mu_j = \left(M\left[S_n^{(j)}\right]\right)^{-1} > 0$, $\sigma_j = D\left(S_n^{(j)}\right) > 0$ and $\lambda_j = \left(M\left[z_n^{(j)}\right]\right)^{-1} > 0$, $a_j = D\left(z_n^{(j)}\right) > 0$, $j = 1, 2, \ldots, k$; with all of these terms assumed finite. Denote $p_{ij} = P\left(\Phi_n^{(i)} = j\right) > 0$, $j = 1, 2, \ldots, k$.

In the context of the queueing network being considered, the random variables $z_n^{(j)}$ represent an interarrival time from outside the network at the station j, while $S_n^{(j)}$ is the nth service time at the station j, and $\Phi_n^{(j)}$ is a routing indicator for the nth customer served at the station j. If $\Phi_n^{(i)} = j$ (which occurs with probability p_{ij}), then the nth customer served at the station i is routed to the station j. When $\Phi_n^{(i)} = 0$, the associated customer leaves the network. The matrix P is called a routing matrix.

We observe that this system is quite general, encompassing tandem systems, acyclic networks of $GI/G/1$ queues, networks of $GI/G/1$ queues with feedback and open queueing networks.

Let us define $Q_j(t)$ as the queue length of customers at the jth station of the queueing network at time t, $\beta_j = \lambda_j + \sum_{i=1}^{k} \mu_i \cdot p_{ij} - \mu_j > 0$, $\hat{\sigma}_j^2 = (\lambda_j)^3 \cdot Dz_n^{(j)} + \sum_{i=1}^{k} (\mu_i)^3 \cdot DS_n^{(i)} \cdot (p_{ij})^2 + (\mu_j)^3 \cdot DS_n^{(j)} > 0$, $j = 1, 2, \ldots, k$ and $t > 0$. Also, let us define $W_j(t)$ as the virtual waiting time of customers at the j-th station of a queueing network at time t (one must wait until a customer arrives at the j-th station of the queueing network to be served at time t), $\bar{\beta}_j = \frac{\lambda_j + \sum_{i=1}^{k} \mu_i \cdot p_{ij}}{\mu_j} - 1 > 0$, $\bar{\sigma}_j^2 = \sum_{i=1}^{k} p_{ij}^2 \cdot \mu_i \cdot \left(\sigma_j + \left(\frac{\mu_i}{\mu_j}\right)^2 \cdot \sigma_i\right) + \lambda_j \cdot \left(\sigma_j + \left(\frac{\lambda_j}{\mu_j}\right)^2 \cdot a_j\right) > 0$, $j = 1, 2, \ldots, k$ and $t > 0$.

We will also assume that the following "overload conditions" are fulfilled:

$$\lambda_j + \sum_{i=1}^{k} \mu_i \cdot p_{ij} > \mu_j, \quad j = 1, 2, \ldots, k. \tag{1}$$

Note that these conditions guarantee that the length of all the queues will grow indefinitely with probability one. The results of the present paper are based on the following theorems:

Theorem 1. *If conditions (1) are fulfilled, then*

$$\lim_{n\to\infty} \left(\frac{Q_j(nt) - \beta_j \cdot n \cdot t}{\hat{\sigma}_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2t}\right) dy,$$

$0 \le t \le 1$ *and* $j = 1, 2, \ldots, p$.

and

Theorem 2. *If conditions (1) are fulfilled, then*

$$\lim_{n\to\infty} P\left(\frac{V_i(nt) - \tilde{\beta}_i \cdot n \cdot t}{\hat{\sigma}_i \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2t}\right) dy,$$

$0 \le t \le 1$ *and* $i = 1, 2, \ldots, r$.

These theorems are proved in [7], and the proof is therefore omitted here so as not to lengthen this short paper.

3. THE RELIABILITY OF A COMPUTER NETWORK

Now we present a technical example from the computer network practice. Assume that queues arrive at a computer v_j at the rate λ_j per hour during business hours, $j = 1, 2, \ldots$,p+r. These queues are served at the rate μ_j per hour in the computer v_j, $j = 1, 2, \ldots$,p+r. After service in the computer v_j, with probability p_j (usually $p_j \ge 0.9$), they leave the network and with probability p_{ji}, $i \ne j$, $1 \le i \le$p+r (usually $0 < p_{ji} \le 0.1$) arrive at the computer v_i, $i = 1, 2, \ldots$,p+r. Also, we assume the computer v_j fails when the queue length of queues is more than k_j, $j = 1, 2, \ldots, p$, and the computer v_i fails when the virtual waiting time of queues is more than γ_i, $i = 1, 2, \ldots, r$.

In this section, we prove the following theorem on the probability that a computer network fails due to overload.

Theorem 3. *If* $t \ge \max\left(\max\limits_{1\le j\le p} \frac{k_j}{\beta_j}, \max\limits_{1\le i\le r} \frac{\gamma_i}{\tilde{\beta}_i} \right)$ *and conditions (1) are fulfilled, the computer network becomes unreliable (all computers fail).*

At first, using Theorem 1 and Theorem 2, we get that for $x > 0$

$$\lim_{n\to\infty} P\left(\frac{Q_j(nt) - \beta_j \cdot n \cdot t}{\hat{\sigma}_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2t}\right) dy, \; j = 1, 2, \ldots, p \tag{2}$$

and

$$\lim_{n\to\infty} P\left(\frac{V_i(nt) - \tilde{\beta}_i \cdot n \cdot t}{\tilde{\sigma}_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2t}\right) dy, \; i = 1, 2, \ldots, r. \tag{3}$$

Let us investigate a computer network which consists of the elements (computers) α_j that are indicators of stations X_j, $j = 1, 2, \ldots, p$ and elements (computers) ν_i that are indicators of stations Y_i, $i = 1, 2, \ldots, r$

Denote

$$X_j = \begin{cases} 1, & \text{if the element } \alpha_j \text{ is reliable} \\ 0, & \text{if the element } \alpha_j \text{ is not reliable,} \end{cases}$$

$j = 1, 2, \ldots, p$ and

$$Y_i = \begin{cases} 1, & \text{if the element } \nu_i \text{ is reliable} \\ 0, & \text{if the element } \nu_i \text{ is not reliable,} \end{cases}$$

$i = 1, 2, \ldots, r$.

Note that $\{X_j = 1\} = \{Q_j(nt) < k_j\}$, $j = 1, 2, \ldots, p$ and $\{Y_i = 1\} = \{V_i(nt) < \gamma_i\}$, $i = 1, 2, \ldots, r$. Denote the structural function of the system of elements, connected by scheme 1 from $p + r$ (see, for example, [8]), as follows:

$$\phi(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) = \begin{cases} 1, & \sum_{j=1}^{p} X_j + \sum_{i=1}^{r} Y_i \geq 1 \\ 0, & \sum_{j=1}^{p} X_j + \sum_{i=1}^{r} Y_i < 1. \end{cases}$$

Assume $y = \sum_{j=2}^{p} X_j + \sum_{i=1}^{r} Y_i$. Estimate the reliability function of the system (computer network) using the formula of conditional probability

$$h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) = E\phi(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) =$$

$$P(\phi(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) = 1) = P(\sum_{j=1}^{p} X_j + \sum_{i=1}^{r} Y_i \geq 1) =$$

$$P(X_1 + y \geq 1) = P(X_1 + y \geq 1 | y = 1) \cdot P(y = 1) + P(X_1 + y \geq 1 | y = 0) \cdot P(y = 0) =$$

$$P(X_1 \geq 0) \cdot P(y = 1) + P(X_1 \geq 1) \cdot P(y = 0) \leq P(y = 1) + P(X_1 \geq 1) =$$

$$P(y = 1) + P(X_1 = 1) \leq P(y \geq 1) + P(X_1 = 1) =$$

$$P(\sum_{j=2}^{p} X_j + \sum_{i=1}^{r} X_i \geq 1) + P(X_1 = 1) \leq \cdots \leq \sum_{j=1}^{p} P(X_j = 1) + \sum_{i=1}^{r} P(Y_i \geq)$$

$$\leq \cdots \leq \sum_{j=1}^{p} P(X_j = 1) + \sum_{i=1}^{r} P(Y_i = 1) = \sum_{j=1}^{p} P(Q_j(nt) \leq k_j) + \sum_{i=1}^{r} P(V_i(nt) \leq \gamma_i).$$

Thus,

$$0 \leq h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) \leq \sum_{j=1}^{p} P(Q_j(nt) \leq k_j) + \sum_{i=1}^{r} P(V_i(nt) \leq \gamma_i). \quad (4)$$

Applying Theorem 1, we obtain that

$$0 \le \lim_{n\to\infty} P(Q_j(nt) < k_j) =$$

$$\lim_{n\to\infty} P\left(\frac{Q_j(nt) - \beta_j \cdot n \cdot t}{\hat{\sigma}_j \cdot \sqrt{n}} < \frac{k_j - \beta_j \cdot n \cdot t}{\hat{\sigma}_j \cdot \sqrt{n}}\right) = \int_{-\infty}^{-\infty} \exp\left(-\frac{y^2}{2t}\right) dy = 0, \qquad (5)$$

$j = 1, 2, \ldots, p$.

From (4) it follows that for $k_j < \infty$,

$$\lim_{n\to\infty} P\left(Q_j(nt) < k_j\right) = 0, \; j = 1, 2, \ldots, p. \qquad (6)$$

Similarly as in (5) - (6) we prove that for $\gamma_i < \infty$

$$\lim_{n\to\infty} P\left(V_i(nt) < \gamma_i\right) = 0, \; i = 1, 2, \ldots, r. \qquad (7)$$

Consequently, $\lim_{n\to\infty} h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) = 0$ (see (4), (6) and (7)), which completes the proof.

Finally, we provide an expression for $h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t)$, $t > 0$ by proving the following theorem.

Theorem 4. *$h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t)$ is equal to* $\exp(-\sum_{j=1}^{p} P(Q_j(nt) < k_j) - \sum_{i=1}^{r} P(V_i(nt) < \gamma_i))$.

Let λ_j, $j = 1, 2, \ldots, p$ and $q_i, i = 1, 2, \ldots, r$ be the traffic intensities related to each server. Then the probability of stopping this system is equal to $e^{-\sum_{j=1}^{p} \lambda_j - \sum_{i=1}^{r} q_i}$ (see, for example, [13]). But

$$\lambda_j = MX_j = P(X_j = 1) = P(Q_j(nt) < k_j), \; j = 1, 2, \ldots, p \qquad (8)$$

and

$$q_i = MY_i = P(Y_i = 1) = P(V_i(nt) < \gamma_i), \; i = 1, 2, \ldots, r. \qquad (9)$$

Applying (8) and (9), we obtain that $h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t)$ is equal to

$$e^{-\sum_{j=1}^{p} \lambda_j - \sum_{i=1}^{r} q_i} = e^{-\sum_{j=1}^{p} P(Q_j(nt) < k_j) - \sum_{i=1}^{r} P(V_i(nt) < \gamma_i)}$$

and the proof is complete.

We thus see that Theorem 4 provides an estimate for the probability that the computer network fails due to overload at time t.

REFERENCES

1. *Gelenbe, E.* Unified approach to the evaluation of a class of replacement algorithms // IEEE Transaction on Computers. 1973.– V. 2.– № 6.– P. 611-618.
2. *Gelenbe, E.* Probabilistic models of computer systems, Part II // Acta Informatica. 1979.– V. 12.– № 1. – P. 285-303.
3. *Gelenbe, E., Finkel, D., Tripathi, S.K.* Availability of a distributed computer system with failures // Acta Informatica. 1986.– V. 23.– № 6. – P. 643-655.

4. *Gelenbe, E, Wang, X. W., Onvural, R.* Diffusion based statistical call admission control in ATM // Performance Evaluation. 1996.– V. 27.– № 8. – P. 411-436.
5. *Gelenbe E., Fourneau J. M.* G-networks with resets // Acta Informatica. 2002.– V. 49.– № 1-4.– P. 179-191.
6. *Lin M. S., Chen D. J.* The computational complexity of reliability problem on distributed systems // Information Processing Letters. 1997.– V. 64.– № 3.– P. 143-147.
7. *Minkevičius S.* Extreme values in open queueing networks. I, II // to appear, 2009.
8. *Morder J.J., Elmaghraby S. E. (eds.)* Handbook of operational research models and applications, Van Nostrand Reinhold, New York, 1978.
9. *Park K., Kim S.* A capacity planning model of unreliable multimedia service // Journal of Systems and Software. 2002.– V. 63.– № 1.– P. 69-76.
10. *Pearn W. L, Ke J. C., Chang Y. C.* Sensitivity analysis of the optimal management policy for a queueing system with a removable and non-reliable server // Computers and Industrial Engineering. 2004.– V. 46.– № 1.– P. 87-99.
11. *Perry D., Posner M. J. M.* A correlated M/G/1-type queue with randomized server repair and maintenance modes // Operations Research Letters. 2000.– V. 26.– № 3.– P. 137-148.
12. *Sztrik J., Kim C. S.* Markov-modulated finite-source queueing models in evaluation of computer and communication systems // Mathematical and Computer Modelling. 2003.– V. 38.– № 7-9.– P. 961-968.
13. *Zubov V. (ed.)* Mathematical theory of reliability of queueing systems, Radio i Sviaz, Moscow, 1964.