On the component-based reliability in open multi-server queueing networks

Edvinas Greičius, Saulius Minkevičius

Abstract — This paper is motivated by performance in terms of reliability of multi-server computer networks. Limit theorems on the queue length and virtual waiting time in an open multi-server queueing network in heavy traffic are derived and applied to a reliability model for a multi-server computer network, where the time of failure of a multi-server computer network is related to the parameters of the system.

Keywords — Heavy traffic, performance evaluation, queueing theory, probability limit theorem.

I Introduction

Probabilistic models and queueing networks have long been used to study the performance and reliability of computer systems and to analyse the performance and reliability of computer networks and of distributed information systems. In this paper, we will first review the works related to using the queueing theory of computer systems 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, 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 successfully executed 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], and in [5] a novel model for a reliable system composed of N unreliable systems, which can hinder or enhance each other’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 the system performance. In [11], the authors consider a single machine subject to break down and employ a fluid queue model with repair. In [13], 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 the probability limit theorem on the queue length and virtual waiting time of the customer in heavy traffic for open multi-server queueing networks.

II The network model

Consider a network of j stations, indexed by \( j = 1, 2, \ldots, J \), and the station \( j \) has \( c_j \) servers, indexed by \( \{j, 1\}, \ldots, \{j, c_j\} \). A description of the primitive data and construction of processes of interest are the focus of this section. No probability space will be mentioned in this section, and certainly, one can always think that all the variables and processes are defined on the same probability space.

First, \( \{u_{j}(e), e \geq 1\}, j = 1, 2, \ldots, J \), are \( J \) sequences of exogenous interarrival times, where \( u_{j}(e) \geq 0 \) is the interarrival time between the \( e - 1 \) job and the \( e \)-th job which arrive at the station \( j \) exogenously (from the outside of the network). Define \( U_{j}(0) = 0 \), \( U_{j}(n) = \sum_{e=1}^{n} u_{j}(e), n \geq 1 \) and \( A_{j}(t) = \text{sup}\{n \geq 0 : U_{j}(n) \leq t\} \), where \( A_{j} = \{A_{j}(t), t \geq 0\} \) is called the exogenous arrival process of the station \( j \), i.e., \( A_{j}(t) \) counts the number of jobs that arrived at the station \( j \) from the outside of the network.

Second, \( \{v_{j,k}(e), e \geq 1\}, j = 1, 2, \ldots, J, k_j = 1, 2, \ldots, c_j \), are \( c_1 + \ldots + c_J \) sequences of service times, where \( v_{j,k}(e) \geq 0 \) is the service time for the \( e \)-th customer served by the server \( k_j \) of the station \( j \). Assume that \( V_{j,k}(0) = 0 \), \( V_{j,k}(n) = \sum_{e=1}^{n} v_{j,k}(e), n \geq 1 \) and \( x_{j,k}(t) = \text{sup}\{n \geq 0 : V_{j,k}(n) \leq t\} \), where \( x_{j,k} = \{x_{j,k}(t), t \geq 0\} \) is called the service process for the server \( k_j \) at the station \( j \), i.e., \( x_{j,k}(t) \) counts the number of services completed by server \( k_j \) at the station \( j \) during the server’s busy time. We define \( \mu_{j,k} = (M[v_{j,k}])^{-1} > 0 \), \( \sigma_{j,k} = D(v_{j,k}) > 0 \) and \( \lambda_{j} = (M[u_{j}])^{-1} > 0 \), \( a_j = D(u_{j}) > 0 \), \( j = 1, 2, \ldots, k \), with all of these terms assumed finite. Let \( p_{ij} \) be probability of the job after service at the \( i \)-th station of the network are arrived to the \( j \)-th station of the network, \( i, j = 1, 2, \ldots, J \).

Now we introduce the following process \( Q_{j,k} = \{Q_{j,k}(t), t \geq 0\} \), where \( Q_{j,k}(t) \) indicates the number of customers waiting to be served by server \( k_j \) of the station \( j \) at time \( t \); \( j = 1, 2, \ldots, J \), \( k_j = 1, 2, \ldots, c_j \).
Mathematical Methods in Science and Engineering

III THE MAIN RESULTS

Let the number of servers $k_j$ in $j$-th station of the network divide into parts: $k_j = 1, 2, \ldots, p_j$ (where the probability limit theorem is valid for queue length of customers) and $k_j = 1, 2, \ldots, r_j$ (where the probability limit theorem is valid for the virtual waiting time of customer (workload process) in $k_j$ server of the station $j$ at time $t; j = 1, 2, \ldots, J; k_j = 1, \ldots, c_j$).

The dynamics of the queuing system (to be specified) depends on the service discipline at each service station. To be more precise, “first come, first served” (FCFS) service discipline is assumed for all $J$ stations. When a customer arrives at a station and finds more than one server available, it will join one of the servers with the smallest index. We assume that the service station is work-conserving; namely, not all servers at a station can be idle when there are customers waiting for service at that station. In particular, we assume that a station must serve at its full capacity when the number of jobs waiting is equal to or exceeds the number of servers at that station.

We also define

$$\tilde{\beta}_{jkj} = \sum_{i=1}^{j} \hat{\beta}_{ikj} = \frac{\lambda_j}{c_j \cdot \mu_{jkj} \cdot p_j} > 1 > 0,$$

$$\sigma_j = \frac{\lambda_j}{\mu_{jkj}} \cdot \frac{1}{c_j \cdot p_j^2} + 1 > 0, j = 1, 2, \ldots, J; k_j = 1, 2, \ldots, c_j; t \geq 0.$$

We also assume that the following “overload conditions” are fulfilled

$$\sum_{i=1}^{c_j} \mu_{ikj} \cdot p_j + \lambda_j > 0, j = 1, 2, \ldots, J. \quad (1)$$

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

$$\lim_{n \to \infty} \mathbb{P} \left( \frac{\tilde{Q}_{ikj}(nt) - \tilde{\beta}_{jkj} \cdot n \cdot t}{\sigma_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp \left( -\frac{x^2}{2} \right) dx,$$

$$0 \leq t \leq 1, k_j = 1, 2, \ldots, r_j; j = 1, 2, \ldots, J.$$

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

IV THE RELIABILITY OF A MULTI-SERVER COMPUTER NETWORK

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

If $t \geq \max \left( \frac{\max_{1 \leq j \leq 1} m_j}{\beta_{jkj}}, \frac{\max_{1 \leq \gamma_j \leq r_j} \gamma_j}{\beta_{jkj}} \right)$ and conditions (1) are fulfilled, the computer network becomes unreliable (all computers fail).

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

$$\lim_{n \to \infty} \mathbb{P} \left( \frac{\tilde{Q}_{ikj}(nt) - \tilde{\beta}_{jkj} \cdot n \cdot t}{\sigma_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp \left( -\frac{x^2}{2} \right) dx,$$

$$k_j = 1, 2, \ldots, p_j \quad (2)$$

and

$$\lim_{n \to \infty} \mathbb{P} \left( \frac{\tilde{V}_{jkj}(nt) - \tilde{\beta}_{jkj} \cdot n \cdot t}{\sigma_j \cdot \sqrt{n}} < x \right) = \int_{-\infty}^{x} \exp \left( -\frac{x^2}{2} \right) dx,$$

$$k_j = 1, 2, \ldots, r_j, j = 1, 2, \ldots, J. \quad (3)$$

Let us investigate a computer network which consists of the elements (computers) $\alpha_i$ that are indicators of stations $X_j$, $j = 1, 2, \ldots, p_j$ and the elements (computers) $\gamma_i$ that are indicators of stations $Y_i$, $i = 1, 2, \ldots, r_j$

Denote

$$X_j = \begin{cases} 1, & \text{if the element } \alpha_j \text{ is reliable} \\ 0, & \text{if the element } \alpha_j \text{ is not reliable,} \\ j = 1, 2, \ldots, p_j \end{cases}$$

and

$$Y_i = \begin{cases} 1, & \text{if the element } \beta_i \text{ is reliable} \\ 0, & \text{if the element } \beta_i \text{ is not reliable,} \\ i = 1, 2, \ldots, r_j \end{cases}$$

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

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

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

$$P(y = 1) + P(X_1 = 1) =$$

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

$$\sum_{i=1}^{m} \sum_{i=1}^{n} P(Q_{ik}(nt) \leq m_{jk}) +$$

$$\sum_{j=m+1}^{j} \sum_{i=1}^{n} P(V_{ik}(nt) \leq \gamma_{jk}).$$

Assuming that $k_j = p_j + r_j$

$$0 \leq h(X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_r, t) \leq$$

$$\sum_{i=1}^{m} \sum_{i=1}^{n} P(Q_{ik}(nt) \leq m_{jk}) +$$

$$\sum_{j=m+1}^{j} \sum_{i=1}^{n} P(V_{ik}(nt) \leq \gamma_{jk}). \quad (4)$$

Applying Theorem 1, we obtain that for $m_{jk} < \infty$

$$\lim_{n \to \infty} P(Q_{jk}(nt) < m_{jk}) =$$

$$\lim_{n \to \infty} P\left(\frac{Q_{jk}(nt) - \beta_j \cdot n^{-1}}{\sigma_j \cdot n^{-1}} < \frac{m_{jk} - \beta_j \cdot n^{-1}}{\sigma_j \cdot n^{-1}}\right) =$$

$$\int_{-\infty}^{\gamma_{jk}} \exp\left(-\frac{y^2}{2}\right) dy = 0, \quad (5)$$

where $k_j = 1, 2, \ldots, p_j$ and $j = 1, 2, \ldots, J$.

It follows from (5), that, for $m_{jk} < \infty$,

$$\lim_{n \to \infty} P(Q_{jk}(nt) < m_{jk}) = 0, \quad (6)$$

where $k_j = 1, 2, \ldots, p_j$ and $j = 1, 2, \ldots, J$.

Similarly as in (5) - (6), we prove that for $\gamma_{jk} < \infty$,

$$\lim_{n \to \infty} P(V_{jk}(nt) < \gamma_{jk}) = 0, \quad (7)$$

where $k_j = 1, 2, \ldots, r_j$ and $j = 1, 2, \ldots, J$.

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.

V. CONCLUDING REMARKS AND FUTURE RESEARCH

1. Conditions (1) are fundamental, - the behaviour of the whole network and its evolution is not clear, if conditions (1) are not satisfied. Therefore, this fact is the object of further research and discussion.

2. Note that a computer with Windows operating system functions steadily if the number of jobs does not exceed 5 (therefore, $m_{jk} = 5$). In other cases, the computer fails (see paragraph 1).

REFERENCES


