FURTHER ANALYSIS OF CONFIDENCE INTERVALS FOR LARGE CLIENT/SERVER COMPUTER NETWORKS

Vyacheslav Abramov

School of Mathematical Sciences, Monash University, Building 28, Level 4,
Clayton Campus, Wellington road, Clayton, Victoria 3800, Australia

e-mail: vyacheslav.abramov@sci.monash.edu.au

Abstract

In the recent paper [Abramov, *RTA,* 2 (2007), pp. 34-42], confidence intervals have been derived for symmetric large client/server computer networks with client servers, which are subject to breakdowns. The present paper mainly discusses the case of asymmetric network and provides another representation of confidence intervals.

Key words: Closed networks, Performance analysis, Normalized queue-length process, Confidence intervals.

2000 Mathematical Subject Classification: 60K30, 60K25

1. Introduction

Consider a large closed queuing network containing a server station (infinite-server queuing system) and k single-server client stations. The total number of customers (units) is N, where N is assumed to be a large parameter. The departure process from client stations is assumed to be autonomous. For the definition of queuing systems with autonomous service mechanism in the simplest case of single arrivals and departures see [1], where there are references to other papers related to that subject.

The arrival process from the server to the i-th client station is denoted $A_{i,N}(t)$. The service time of each unit in the server station is exponentially distributed with parameter λ. Therefore, the rate of arrival to client stations depends on the number of units in the server station. If there is N_t units in the server station in time t, then the rate of departure of units from the server in time t is λN_t. There are k client stations in total, and each client station is a subject to breakdown. The lifetime of each client station is a continuous random variable independent of lifetimes of other client stations. The probability distribution of i-th client station is $G_i(x)$. The client stations are not necessarily identical, and a unit transmitted from the server station chooses each one with corresponding probability p_i. This probability p_i depends on configuration of the system in a given time moment, that is on the number of available (not failure) client servers and there indexes. In general such kind of dependence is very complicated. However, in the case of the network with only two client stations this dependence is simple. This simplest case is just discussed in the present paper.

The departure instants from j-th client station (j=1,2,…,k) are $\xi_{j,N,1}$, $\xi_{j,N,1}+\xi_{j,N,2}$, $\xi_{j,N,1}+\xi_{j,N,2}+\xi_{j,N,3}$, … where each the sequence $\xi_{j,N,1}$, $\xi_{j,N,2}$, … forms a strictly stationary and ergodic sequence of random variables (N is the series parameter). The corresponding point process associated with departures from the client station j is denoted

$$S_{j,N}(t)=\sum_{i=1}^{\infty}\mathrm{I}\left\{\sum_{l=1}^{i}\xi_{j,N,l}\leq t\right\},$$

and satisfies the condition

$$\mathrm{P}\left\{\lim_{t\to\infty}\frac{S_{j,N}(t)}{t}=\mu_j N\right\}=1.$$

The relations between parameters λ, p_j, μ_j (j=1,2,…,k) and k are assumed to be

(1.1) $\frac{\lambda p_j}{\mu_j}<1,$ for all j=1,2,…,k,

and

(1.2) $\frac{\lambda}{\mu_j}>1,$ at least for one of j's, j=1,2,…,k.

In the sequel, asymmetric networks will be discussed for k=2, and the relations (1.1) and (1.2) will be assumed for this value of parameter k.

In the case of symmetric network, where $\mu_j=\mu$ for all j, and $p_j=p_j(l)=1/l$, where l is the number of available client stations (in a given time moment), conditions (1.1) and (1.2) correspondingly are as follows:

(1.3) $\frac{\lambda}{k\mu}<1,$

and

(1.4) $\frac{\lambda}{\mu}>1.$

In this case, condition (1.4) means that after one or other breakdown the entire client stations become bottleneck, and there is a value $l_0=\max\left\{l:\frac{\lambda}{l\mu}>1\right\}$.

The queue-length in the j-th client station is defined as

$$Q_{j,N}(t)=A_{j,N}(t)-\int_0^t\mathrm{I}\{Q_{j,N}(s-)>0\}dS_{j,N}(s).$$

In the case of N large, the behavior of the queue-length process is as follows. When all of k client stations are available in time t, most of units are concentrated at the server station, and normalized queue-lengths $q_{j,N}(t)=Q_{j,N}(t)/N$ vanish as N increases indefinitely. When after one or another failure the client stations are overloaded in time t, then $q_{j,N}(t)$ converge in probability, as N increases indefinitely, to some positive value. Then the queue-lengths in client stations increase more and more as t increases. Then the system is assumed to be *at risk* if the total number of units in queues in client stations increases the value αN.

Confidence intervals for symmetric large client/server computer networks have been studied in Abramov [1]. The motivation of this problem, review of the related literature and technical details are given in [1]. The present paper mainly discusses confidence intervals for large asymmetric client/server computer networks and provides new representations for confidence intervals.

In the case of symmetric network, a confidence interval is characterized by parameter $\gamma \le \alpha$. More specifically, for given level of probability P, say P=0.95, there is value γ characterizing the guaranteed level of normalized cumulated queue, and a (random) confidence interval is associated with this value of γ. In other words, along with parameter α characterizing the system at risk we have another parameter γ, which is closely associated with α and with probability P. In the particular case k=2, the explicit representation for γ has been established in [1].

In the case of asymmetric network such deterministic parameter cannot longer characterize confidence intervals. To see it, consider a network the parameters of which are: $\lambda = 1$, $p_1 = p_2 = 1/2$, $\mu_1 = 4/3$, $\mu_2 = 2/3$. Then, condition (1.1) is fulfilled, and $\frac{\lambda p_1}{\mu_1} = \frac{3}{8} < 1$, $\frac{\lambda p_2}{\mu_2} = \frac{3}{4} < 1$. Condition (1.2) is fulfilled as well, and $\frac{\lambda}{\mu_1} = \frac{3}{4} < 1$, $\frac{\lambda}{\mu_2} = \frac{3}{2} > 1$. In this case one can expect the situation when the second server breakdowns first, and the cumulative normalized queue-length process will converge to zero as $N \to \infty$ for all t, and there is no observable parameter. Therefore, the network can breakdown unexpectedly without any information on its state. In other example, where the parameters of network are: $\lambda = 1$, $p_1 = p_2 = 1/2$, $\mu_1 = 3/4$, $\mu_2 = 2/3$, we have the following situation. Condition (1.1) is fulfilled with $\frac{\lambda p_1}{\mu_1} = \frac{2}{3} < 1$, $\frac{\lambda p_2}{\mu_2} = \frac{3}{4} < 1$. Condition (1.2) is fulfilled with $\frac{\lambda}{\mu_1} = \frac{4}{3} > 1$ and $\frac{\lambda}{\mu_2} = \frac{3}{2} > 1$. Therefore in the case when the first client station breakdowns first, the limiting cumulative normalized queue-length will increase with the rate, different from that would be in the case when the second client station breakdowns first. Therefore, by following up the limiting cumulative normalized queue-length one cannot uniquely characterize a confidence interval as it has been done in the case of symmetric networks. For this reason we need in another representation for confidence intervals.

The rest of the paper is organized as follows. In Section 2 we recall main equations for limiting (as $N \to \infty$) cumulated normalized queue-length process from earlier paper [1], and derive slightly more general representation than in [1]. We then derive an explicit value for a confidence interval, which are closely related to the result, obtained in [1]. In Section 3, the results are derived for asymmetric networks in the case k=2. We conclude the paper in Section 4.

2. The case of symmetric network

Limiting as $N \to \infty$ cumulated normalized queue-length process is denoted $q(t)$. Let $l_0 = \max\left\{l : \frac{\lambda}{l\mu} > 1\right\}$, let $\tau_1, \tau_2, ..., \tau_k$ be the moments of breakdown of client stations, $0 \le \tau_1 \le \tau_2 \le ... \le \tau_k$. Then $q(t) = 0$ for all $t \le \tau_{k-l_0}$, and in any arbitrary time interval $[\tau_i, \tau_{i+1})$, $i = k - l_0, k - l_0 + 1, ..., k-1$, we have the equation

$$q(t) = q(\tau_i) + [1 - q(\tau_i)]\{([1 - q(\tau_i)]\lambda - \mu(k-i))(t - \tau_i) - [1 - q(\tau_i)]\lambda \int_{\tau_i}^{t} r(s - \tau_i) ds\},$$

where

$$r(t) = \left(1 - \frac{\mu(k-i)}{[1 - q(\tau_i)]\lambda}\right)\left(1 - e^{-[1-q(\tau_i)]\lambda t}\right).$$

In the last endpoint τ_k we set $q(t) = 1$. In the case of k=2 the confidence time interval is the sum of two intervals. The first interval is [0, τ_1). The second one is [τ_1, θ), where the endpoint θ is defined as follows. In [τ_1, τ_2) for $q(t)$ we have the equation:

$$q(t) = (\lambda - \mu)(t - \tau_1) - \lambda \int_{\tau_1}^{t} r(s - \tau_1) ds,$$

where

$$r(s) = \left(1 - \frac{\mu}{\lambda}\right)\left(1 - e^{-\lambda t}\right).$$

To find θ we have the following equations:

$$\mathrm{P}\{q(t) = 0\} = [1 - G(t)]^2,$$

$$\mathrm{P}\{q(t) \le \gamma < 1\} = [1 - G(t)][1 - G(t - t_\gamma)],$$

where $G(x)$ denotes lifetime distribution of each of identical client stations and t_γ is such the value of t under which

$$(\lambda - \mu)t - \lambda \int_0^t r(s) ds = \gamma.$$

The value of t_γ can be found from the relation

$$\frac{\int_0^\infty [1-G(t)][1-G(t-t_\gamma)]dt}{\int_0^\infty [1-G(t-t_\gamma)]^2\,dt} = P.$$

If the corresponding value of γ is not greater than α, then the value θ of the interval $[\tau_1, \theta)$ should be taken $\theta = \tau_1 + t_\gamma$. Otherwise, if $\gamma > \alpha$, the value θ should be taken $\theta = \tau_1 + t_\alpha$.

The above result has been obtained in [1]. Let us now extend this result for a more general situation of an arbitrary number of client stations $k \geq 2$ under the special setting assuming that $l_0 = 1$. This means that $q(t) = 0$ in the random interval $[0, \tau_{k-1})$, and $q(t) > 0$ in the interval (τ_{k-1}, τ_k). In this case we have the following relationships:

$$\mathrm{P}\{q(t) = 0\} = \sum_{i=2}^{k} \binom{k}{i} [1-G(t)]^i [G(t)]^{k-i},$$

$$\mathrm{P}\{q(t) \leq \gamma < 1\} = [1-G(t)] \sum_{i=1}^{k-1} \binom{k-1}{i} [1-G(t-t_\gamma)]^i [G(t-t_\gamma)]^{k-i},$$

where the value of t_γ can be found from the relation

$$\frac{\int_0^\infty [1-G(t)] \sum_{i=1}^{k-1} \binom{k-1}{i} [1-G(t-t_\gamma)]^i [G(t-t_\gamma)]^{k-i}\,dt}{\int_0^\infty \sum_{i=2}^{k} \binom{k}{i} [1-G(t-t_\gamma)]^i [G(t-t_\gamma)]^{k-i}\,dt} = P.$$

Again, if the corresponding value of γ is not greater than α, then the value θ of the interval $[\tau_{k-1}, \theta)$ should be taken $\theta = \tau_{k-1} + t_\gamma$. Otherwise, if $\gamma > \alpha$, the value θ should be taken $\theta = \tau_{k-1} + t_\alpha$.

The above construction gives us a random confidence interval [0, θ) corresponding to the level of probability not smaller than *P*. We now find a *deterministic* confidence interval corresponding to the level of probability not smaller than *P*. That deterministic confidence interval will be a guaranteed interval, and the probability that the system will be available is not smaller than *P*.

We have

$$\mathrm{P}\{\theta > t\} = \mathrm{P}\{\tau_1 + t_\gamma > t\} = \mathrm{P}\{\tau_1 > t - t_\gamma\} = [1-G(t-t_\gamma)]^2 = P.$$

Therefore, the desired deterministic interval is [0, $z + t_\gamma$], where the value *z* is given from the condition $[1-G(z)]^2 = P$.

Then the construction of deterministic interval in the case *k*=2 is as follows.

- According to the aforementioned relations we find the value of interval t_γ. If the corresponding value of γ is not greater than α, then we accept this interval and set $\mathrm{T} := t_\gamma$.
- Otherwise, we set $\mathrm{T} := t_\alpha$, where the value t_α is determined from the relation $(\lambda-\mu)t - \lambda\int_0^t r(s)ds = \alpha$, and $r(s) = \left(1-\frac{\mu}{\lambda}\right)\left(1-e^{-\lambda t}\right)$.
- We find the value *t* from the relation $[1-G(t)]^2 = P$.
- The confidence interval is then taken as [0, *t*+T].

3. The case of asymmetric network

The case of asymmetric network is similar to that of symmetric network. It is based on the formula for the total probability. Specifically, in the case *k*=2 we are to study the cases as (1) the first client station breakdowns first and (2) the second client station breakdowns first. Using the notation, $G_i(x) = \mathrm{P}\{\chi_i \le x\}$, we have $\mathrm{P}\{\chi_1 \le \chi_2\} = \int_0^\infty [1-G_2(x)]dG_1(x)$. Next, we have the following two values γ_1 and γ_2, such that

$$\mathrm{P}\{q(t) \le \gamma_1 < 1 \mid \chi_1 \le \chi_2\} = [1-G_2(t)][1-G_1(t-t_{\gamma_1})],$$

$$\mathrm{P}\{q(t) \le \gamma_2 < 1 \mid \chi_2 \le \chi_1\} = [1-G_1(t)][1-G_2(t-t_{\gamma_2})],$$

and the corresponding values of t_{γ_1} and t_{γ_2} are found from the relationships

$$\frac{\int_0^\infty [1-G_2(t)][1-G_1(t-t_{\gamma_1})]dt}{\int_0^\infty [1-G_1(t-t_{\gamma_1})][1-G_2(t-t_{\gamma_1})]dt} = P,$$

$$\frac{\int_0^\infty [1-G_1(t)][1-G_2(t-t_{\gamma_2})]dt}{\int_0^\infty [1-G_1(t-t_{\gamma_2})][1-G_2(t-t_{\gamma_2})]dt} = P.$$

where in each case, if γ_1 or γ_2 is greater than α, then the corresponding value is replaced by α, and then the corresponding value of t_{γ_1} or t_{γ_2} is to be replaced by t_α as well.

Similarly to the case of symmetric network in this case we have the following.

- We find the value of intervals t_{γ_1} and t_{γ_2}. If the corresponding value of γ_1 or γ_2 is not greater than α, then we accept this interval and set $\mathrm{T}_1 := t_{\gamma_1}$ or $\mathrm{T}_2 := t_{\gamma_2}$
- Otherwise, we set $\mathrm{T}_1 := t_\alpha$ or $\mathrm{T}_2 := t_\alpha$, where the value t_α is determined from the relation $(\lambda-\mu)t - \lambda\int_0^t r(s)ds = \alpha$, and $r(s) = \left(1-\frac{\mu}{\lambda}\right)\left(1-e^{-\lambda t}\right)$.

- Using the formula for the total expectation we find $\mathrm{T} = \mathrm{T}_1 \int_0^\infty [1 - G_2(x)] dG_1(x) + \mathrm{T}_2 \int_0^\infty G_2(x) dG_1(x).$
- We find value t from the relation $[1 - G_1(t)][1 - G_2(t)] = P.$
- The confidence interval is then taken as [0, t+T].

4. Concluding remark

In the present paper we established confidence intervals for large closed client/server computer networks with two client stations. Unlike the earlier result established in [1] for symmetric network, the confidence intervals are deterministic. The advantage of the result of [1] is that one can judge about the quality of system from the information on the system state. However, approach of [1] is not longer available for asymmetric systems. The advantage of the results of the present paper is that they provide confidence intervals for both symmetric and asymmetric networks that give us the entire lifetime of data system in the network with probability not smaller than P.

Acknowledgement

The author thanks the support of the Australian Research Office (grant # DP0771338).

Reference

[1] **V.M.Abramov,** 2007. Confidence intervals associated with performance analysis of symmetric large closed client/server computer networks. *Reliability: Theory and Applications,* 2, Issue 2, 34-42.