한국시뮬레이션학회 논문지
Vol. 20, No. 2, pp. 1-10 (2011. 6)

시뮬레이션을 이용한 대기행렬 네트워크 도착과정의 변동성함수에 관한 연구

김선교[1†]

A Simulation Study on the Variability Function of the Arrival Process in Queueing Networks

Sunkyo Kim

ABSTRACT

In queueing network analysis, arrival processes are usually modeled as renewal processes by matching mean and variance. The renewal approximation simplifies the analysis and provides reasonably good estimate for the performance measures of the queueing systems under moderate conditions. However, high variability in arrival process or in service process requires more sophisticated approximation procedures for the variability parameter of departure/arrival processes. In this paper, we propose an heuristic approach to refine Whitt's variability function with the k-interval squared coefficient of variation also known as the index of dispersion for intervals(IDI). Regression analysis is used to establish an empirical relationships between the IDI of arrival process and the IDI of departure process of a queueing system.

Key words : Queueing networks, Decomposition approximation, Index of dispersions for intervals(IDI), Variability function, Regression analysis

요 약

본 연구에서는 대기행렬네트워크 성과측정 방법 중의 한 가지로서 널리 이용되는 분해법의 구성요소로 제안된 변동성 함수의 이론적 근거를 살펴보고 성과척도 측정의 정확도 제고를 위하여 회귀분석을 통한 변동성 함수의 모수추정 개선방안을 제안하고자 한다. 이를 위하여 변동성이 높은 도착과정과 서비스 과정이 포함된 직렬 대기행렬 네트워크에서의 이탈과정의 자동상관계수 함수를 추정하여 분해법에 사용할 수 있는 방안을 알아본다.

주요어 : 대기행렬 네트워크, 분해법, 변동성 함수

1. Introduction

The decomposition approach is a reasonable alternative to simulation when the analysis of a queueing network is not tractable by analytic method. The simplicity of the method is due to the two-moment approximation of the arrival and departure processes; for details of the decomposition method see Buzacott and Shanthikumar (1993), Kuehn(1979), Shanthikumar and Buzacott(1981), and Whitt(1983). The squared coefficient of variation (SCV) of inter-arrival times is approximated by either the stationary interval(SI) method or the asymptotic (ASY) method. The SI method completely ignores the autocorrelation in inter-arrival times whereas the ASY method accounts for infinitely many lags of autocorrelation; Albin(1984) and Whitt(1981, 1982). Then the SCV of inter-arrival times is used to approximate the

*이 연구는 아주대학교 교내연구비 지원에 의하여 이루어졌음.
접수일(2011년 3월 10일), 심사일(1차 : 2011년 4월 1일), 게재 확정일(2011년 4월 9일)
[1)] 아주대학교 경영대학 경영학부
주 저 자 : 김선교
교신저자 : 김선교
E-mail; sunkyo@ajou.ac.kr

performance measures such as average waiting time and average queue length. However, using SCV as a single measure of the variability of a non-renewal process may result in significant error in the approximation of performance measure. It is shown that the two-moment parametrization that does not account for the autocorrelation is subject to a serious under-approximation of waiting time especially at a bottleneck station in a network with highly variable arrival process or highly variable service process; Fendick, Saksena, and Whitt(1989, 1991), Jageman et. al.(2004), Kim(2004), Livny et. al.(1993), and Suresh and Whitt (1981). Motivated by this phenomenon Whitt (1995) proposed the variability function that parameterizes the variability of a non-renewal process not as a constant but as a function of given traffic intensity. Unlike the SI or the ASY approximation of SCV of a non-renewal process, the variability function takes on different values and results in better approximation of queueing performances. The decomposition approximation of queueing networks has been refined in many different ways to overcome the limitation of two-moment renewal approximation of arrival processes; Bitran and Tirupati(1988), Kim(2000, 2004, 2005), and Whitt(1994, 1995). Recent studies include Markov Modulated Poisson Process(MMPP) approximation of arrival processes to account for the dependence in arrival processes; Heindl(2001).

This paper proposes a method to refine the variability function with an ad hoc approach to approximate the index of dispersion for intervals(IDI) sequence, i.e. k-interval SCV discussed in Fendick et al.(1989, 1991), Sriram and Whitt(1986), and Whitt(1995). Two different cases of tandem networks considered in Suresh and Whitt (1990a, 1990b) are used for the regression analysis of the variability function queueing formula.

The organization of the paper is as follows. In section 2, a brief review of the decomposition approximation and the variability function is given. Model description and simulation results are given in section 3. An approximation formula for the IDI sequence is given in section 4. Section 5 concludes with the discussion of the implication of our result and future direction of research.

2. Decomposition Approximation and the Variability Function

In this section, we briefly review the decomposition methodology with a focus on the queueing operation and the variability function. The decomposition approximation represents the queueing network as a system of linear equations in terms of the SCV parameter of the arrival process to each station. The system of equations is obtained by three network operations; queueing, splitting, and superposition. Each of these three operations describes the approximate linear relationship among SCVs of intervals of point processes associated with each station; see Whitt(1994) for details. Notations used in this paper are summarized below. For station j,

λ_j: the effective arrival rate
ρ_j: the traffic intensity
c_{aj}^2: the SCV parameter of inter-arrival times
c_{dj}^2: the SCV parameter of inter-departure times
c_{sj}^2: the SCV of service times

In the decomposition approximation of queueing networks, the relationship between c_d^2 and c_a^2 at a given station is approximated by a queueing operation. The splitting operation describes the relationship between departure process from a station and its split processes. The superposition operation approximates the SCV of aggregated arrival processes to a station. These three network operations are based on simplifying assumption that each point process is a renewal process. The variability of each point process is represented by a single parameter. The dependence in inter-arrival times is either completely ignored by the SI approximation or fully accounted for by the ASY approximation; Albin(1984) and Whitt(1981, 1982). This dependence is sometimes critical in queueing performance analysis; Fendick et al.(1989, 1991), and Sriram and Whitt(1986). Moreover, its effect is dependent on traffic intensity at a station; that is, traffic intensity determines the range of relevant lag of autocorrelations. We are interested in parametrization of the dependence in inter-departure times within the queueing operation.

As an approximation of the SCV of inter-departure time, Whitt(1982) proposed

$$c_{dj}^2 = \rho_j^2 c_{sj}^2 + (1-\rho_j^2) c_{aj}^2 \tag{1}$$

based on the SI approximation and

$$c_{dj}^2 = c_{aj}^2 \tag{2}$$

by the ASY approximation. The SI queueing formula is typically chosen over the ASY queueing formula since the latter is appropriate only under heavy traffic and that the SI formula gets close to (2) as the traffic intensity goes to 0. Since the SI method ignores all the autocorrelation, however, the effect of autocorrelation in inter-arrival times is not reflected in the approximation of waiting time in queue. That is the reason why the IDI is needed in some situations for better estimation of the queueing performance. Refinements in this regard have been proposed with some modifications of the weights given to c_s^2 and c_a^2 in (1). Suresh and Whitt(1990a) used traffic intensities of current and next stations based on observations of two stations in tandem. Whitt(1995) proposed the following the variability function for general open queueing networks:

$$c_{dj}^2(\rho) = \alpha(\rho_{j,}\rho) c_{sj}^2 + (1-\alpha(\rho_{j,}\rho)) c_{aj}^2(\rho)$$

where

$$\alpha(\rho_{j,}\rho) = \rho_j^2 \min\left\{1, \frac{(1-\rho)^2}{(1-\rho_{\mathrm{j}}^2)}\right\}$$

and ρ is the traffic intensity of the station of interest (i.e., the bottleneck station). With this variability function, the system of equations is solved as many times as the number of distinct traffic intensities of interest. In order to approximate $\mathrm{E}(W_j)$, the system of equations is solved for $c_{aj}^2(\rho_j)$ with $\rho = \rho_j$; see Whitt(1983, 1995). Then, $c_{aj}^2(\rho_j)$ is used as the SCV parameter for the arrival process to station j in the following approximation formula:

$$\mathrm{E}(W_j) = \frac{\rho_j^2 (c_{sj}^2 + c_{aj}^2)}{2\lambda_j (1-\rho_j)} \tag{3}$$

In Whitt's variability function, ρ plays the role of choosing the level of autocorrelation among inter-arrival times by controlling the weights given to c_s^2 and $c_{aj}^2(\rho_j)$. In this paper we propose an approach to fill the gap between the SI and the ASY queueing formula. Simple tandem queueing networks considered in Suresh and Whitt(1990b) and Whitt(1995) are used to illustrate two different structures of autocorrelations in inter-departure times. By regression analysis, a variability function for queueing formula is obtained that incorporates these two different cases. Since the proposed formula is given in terms of the IDI sequence, it can be considered as an IDI approximation formula for departure processes in queueing networks. That is, the autocorrelation among successive inter-departure times at each station is approximated and characterized by the IDI sequence, $c^2(k)$. This sequence measures the cumulative autocorrelation among inter-departure times. Let $S_k = X_1 + \cdots + X_k$ be the sum of k consecutive intervals. Then the k-interval SCV is defined, for $k \geq 1$, as follows:

$$\begin{aligned} c^2(k) &\equiv k \frac{\mathrm{Var}(S_k)}{\mathrm{E}^2(S_k)} \\ &= c^2(1)[1 + \frac{2}{k}\sum_{j=1}^{k-1}(k-j)\gamma_j] \end{aligned}$$

where γ_j is the lag-*j* autocorrelation of intervals. As two special cases, $c^2(1)$ and $c^2(\infty)$ correspond to SI and ASY method, respectively. In the next section, we discuss a heuristic approach to refine the variability function.

3. Model Description and Simulation Results

In this section, we explore the IDI sequence of inter-arrival times in tandem queues. As in Suresh and Whitt (1990b) and Whitt(1995) two extreme cases of 10-station tandem networks are simulated to generate the IDI sequence for each station. The external arrival process is a renewal process with arrival rate 1. At each station, customers are served according to the FIFO discipline by a single server with i.i.d. service times. The queue capacity is assumed to be infinite. Tandem queueing networks are perfect models for the development and test of the queueing formula since no splitting or superposition is involved. In the first case, a highly variable renewal arrival is transformed into a non-renewal process with large variation in the IDI sequence. In the second case, a highly variable service process transforms a Poisson arrival process into a non-renewal process with large variation in the IDI sequence. This large variation in the IDI sequence complicates the problem of identifying the appropriate inter-arrival SCV parameter to be used in the approximation of waiting time, especially at a bottleneck station. The purpose of this experiment is to propose an approach to approximate the IDI sequence. This is a basis for further refinement of queueing formulas and the decomposition method.

In case A, we consider two subcases for external arrival variability: $c_{a1}^2 = 8$ in case A-1 and $c_{a1}^2 = 4$ in case A-2. The i.i.d. inter-arrival times follow hyperexponential distribution with balanced means. That is, a random variable with $c^2 = 8$ is generated as a combination of two exponential random variables each with mean 1/0.12 and 1/1.88, and with weights 0.06 and 0.94 respectively; see Whitt(1982) for details. Service times are exponentially distributed at all stations, i.e., $c_{sj}^2 = 1, j = 1,...,10$. As for the traffic intensities, $\rho_1 = \cdots = \rho_9 = 0.6$ and $\rho_{10} = 0.9$.

In case B, we consider two subcases for service variability at station 1, i.e. $c_{s1}^2 = 8$ in case B-1 and $c_{s1}^2 = 4$ in case B-2. The external arrival process is Poisson with arrival rate of 1. High variability is created at the first station by i.i.d. service times with hyperexponential distribution with balanced means. With the traffic intensity $\rho_1 = 0.9$ the effect of service variability is passed on to the subsequent stations. For stations 2 to 9, the service times are i.i.d. exponential with mean 0.6. The last station is a bottleneck with traffic intensity 0.9 and its service times are i.i.d. exponential with mean 0.9.

In order to develop a regression approximation of the IDI sequence, $c_{aj}^2(k)$'s are obtained from simulation of 10 replications of 500,000 inter-arrival times, of which first 20,000 observations are truncated.

The IDI sequence for case A-1 is summarized in Table 1 and the graph is shown in Figure 1. Simulation results of case A-1 and A-2 can be summarized as follows:

Table 1. $c_{aj}^2(k)$ of interarrival times from simulation (Case A-1)

	k							
j	1	10	50	100	200	300	400	500
1	8	8.02	8.02	8	8.11	7.93	8	8.06
2	6.03	6.53	7.28	7.54	7.86	7.77	7.95	7.93
3	5.05	5.77	6.84	7.27	7.71	7.67	7.89	7.83
4	4.4	5.25	6.52	7.06	7.59	7.6	7.84	7.77
5	3.93	4.84	6.25	6.89	7.47	7.53	7.79	7.71
6	3.56	4.53	6.02	6.73	7.37	7.47	7.73	7.67
7	3.27	4.25	5.82	6.59	7.28	7.4	7.68	7.63
8	3.02	4.02	5.65	6.47	7.2	7.35	7.63	7.59
9	2.82	3.82	5.48	6.35	7.12	7.3	7.58	7.56
10	2.64	3.64	5.33	6.23	7.04	7.25	7.53	7.51

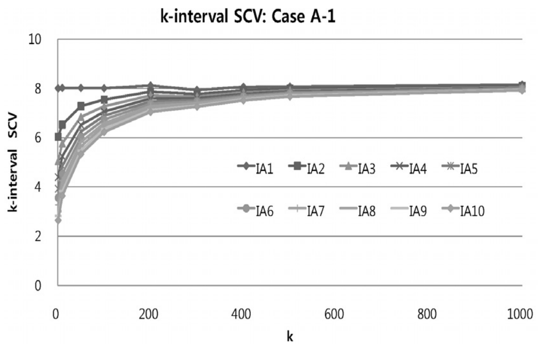

Fig. 1. $c_{aj}^2(k)$ for stations j = 1,...,10(Case A-1)

i) $c_{a1}^2(k) = 8, k \geq 1$.

ii) $c_{aj}^2(1)$ is decreasing with respect to j.

iii) For stations 2 to 10, $c_{aj}^2(k)$ increases to the ASY SCV $c_{a1}^2(\infty)$ = 8 as k goes to ∞.

iv) For station 10, $c_{a10}^2(k)$ ranges from 2.5 to 8 for case A-1 and from 1.5 to 4 for case A-2.

The last observation is an explanation for the heavy traffic bottleneck phenomenon discussed in Suresh and Whitt(1990). That is, the average waiting time approximation based on the SI method could be significantly lower than the ASY method. With the approximation of the whole

Table 2. $c_{aj}^2(k)$ of interarrival times from simulation (Case B-1)

	k							
j	1	10	50	100	200	300	400	500
1	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
2	6.66	6.6	6.21	5.78	5.18	4.67	4.4	4.11
3	5.02	5.4	5.71	5.56	5.13	4.66	4.4	4.14
4	4.23	4.81	5.43	5.42	5.09	4.65	4.39	4.15
5	3.71	4.4	5.21	5.3	5.04	4.63	4.37	4.15
6	3.33	4.08	5.02	5.19	4.99	4.62	4.35	4.14
7	3.04	3.82	4.86	5.09	4.96	4.6	4.35	4.14
8	2.79	3.6	4.71	5.01	4.92	4.59	4.33	4.15
9	2.59	3.41	4.58	4.93	4.89	4.56	4.32	4.15
10	2.43	3.25	4.45	4.84	4.86	4.54	4.31	4.15

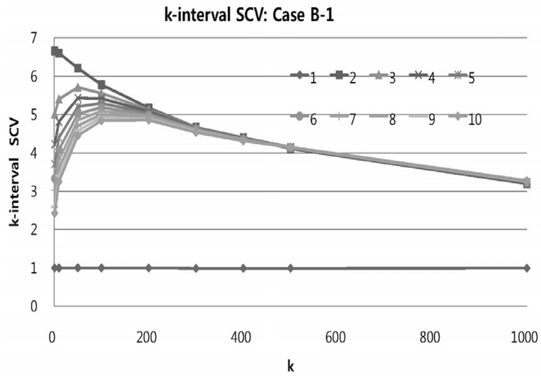

Fig. 2. $c_{aj}^2(k)$ for stations j =1,...,10(Case B-1)

sequence of $c_{a10}^2(k)$, further improvement can be made on the average waiting time approximation.

The IDI sequence for case B-1 is summarized in Table 2. For case B-1, they are also depicted in Figure 2. The following summarizes the simulation for case B:

i) $c_{a1}^2(k)=1, k\geq 1$, since the arrival process is Poisson.

ii) $c_{aj}^2(1)$ is decreasing along the stations from j = 2 to 10.

iii) For station 2, $c_{a2}^2(k)$ is decreasing function of k.

iv) For stations 3 to 10, $c_{aj}^2(k)$'s are increasing and then decreasing.

v) For all stations $c_{aj}^2(k)$ approaches the ASY SCV $c_{a1}^2(\infty)$ = 1 as k goes to ∞. (Simulation shows that $c_{aj}^2(5000)$ is about 1.60 for j =2,...,10.)

The following observations can be made from simulation experiments:

i) For both cases, $c_{aj}^2(1)$, approaches 1 along the stations j =1,..., 10. This is due to exponential service times at all stations in Case A and at all stations except for the first station in Case B.

ii) For both cases, $c_{aj}^2(k)$ approaches the external ASY arrival variability $c_{a1}^2(\infty)$ as k goes to ∞ for all stations.

iii) $c_{a10}^2(k)$ is monotonically increasing with respect to $k\geq 1$ in Case A whereas it is unimodal, increasing and then decreasing, in Case B.

In the next section, we propose an approach to approximate the IDI sequence.

4. Regression Approximation of the Variability Function

4.1 Approximation of the IDI by Regression

In this section, an approximation formula is provided for $c_d^2(k)$. First, it is assumed that the $c_d^2(k)$ is a convex combination of $c_a^2(k)$ and c_s^2 for all k ≥1. By this assumption, if $c_a^2(k)=c_s^2$ for some k, then $c_d^2(k)=c_a^2(k)$. This assumption holds true if there is no dependence between arrival process and service process. Thus, this assumption does not hold under immediate feedback; see Kim(2000) for details.

We propose a formula for the whole sequence of $c_d^2(k)$ given $c_a^2(k)$, c_s^2, and ρ. We consider the IDI sequence of the following form which can be justified by the simulation result:

$$c_d^2(k)=\rho^{g(k)}c_s^2+(1-\rho^{g(k)})c_a^2(k)$$

which can be written as

$$g(k)=\ln[(c_d^2(k)-c_a^2(k))/(c_s^2-c_a^2(k))]/\ln\rho$$

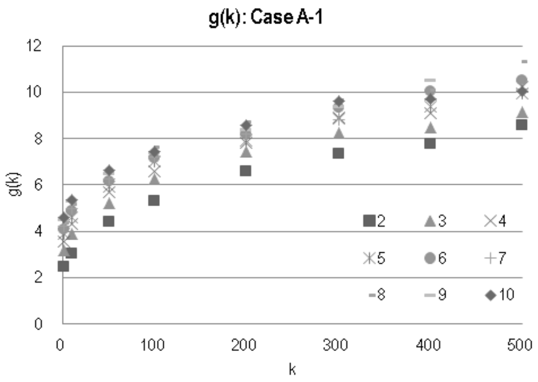

Fig. 3. $g(k)$ of inter-arrival times(Case A-1)

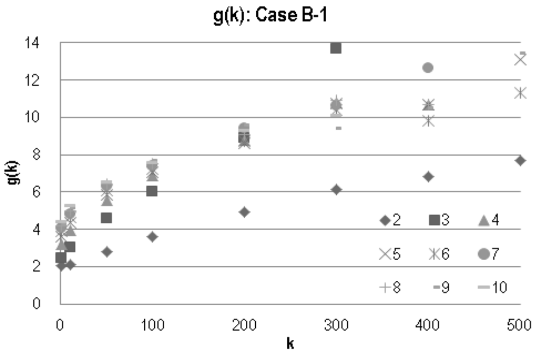

Fig. 4. $g(k)$ of inter-arrival times(Case B-1)

Regression data collected from simulation not satisfying the above assumption has been omitted in regression analysis. That is, there are few observations such that $c_d^2(k) > \max(c_a^2(k), c_s^2)$. Based on the values of $g(k)$ collected from simulations of both cases A and B, a logarithmic function $g(k) = a + b \ln k$ is chosen for the simple regression analysis; see Figures. 3 and 4. Two regression models are considered below, with and without the restriction on the constant a. In the first model, we set $a = 2$ to make the model consistent with the SI queueing formula (1), that is, $c_{dj}^2 = \rho_j^2 c_{sj}^2 + (1-\rho_j^2)c_{aj}^2$ for $k = 1$, where $g(1) = 2$. The second model has no restriction on a.

Based on the IDI sequences collected from the simulation of 10-station tandem networks of cases A and B, estimates of $g(k)$'s are given in table 3.

For the first model with the constraint $a = 2$, both cases show regression coefficient b ranging from 1.18 to 1.36. The second model shows constant a between 1.7 and 3 and coefficient b between 1 and 1.4. So, we may choose $g(k) = 2 + 1.2 \ln k$ for use with the variability function.

Table 3. Regression approximation of g(k)

	Model 1	Model 2
A-1	$g(k) = 2 + 1.182 \ln k$	$g(k) = 2.96 + 0.999 \ln k$
	($R^2 = 0.965$)*	($R^2 = 0.808$)
A-2	$g(k) = 2 + 1.234 \ln k$	$g(k) = 2.516 + 1.135 \ln k$
	($R^2 = 0.975$)	($R^2 = 0.863$)
B-1	$g(k) = 2 + 1.275 \ln k$	$g(k) = 2.288 + 1.217 \ln k$
	($R^2 = 0.915$)	($R^2 = 0.660$)
B-2	$g(k) = 2 + 1.363 \ln k$	$g(k) = 1.732 + 1.416 \ln k$
	($R^2 = 0.923$)	($R^2 = 0.727$)

*R^2 is the coefficient of determination.

For validation of our formula, a two-station tandem queue is used with the same parameter values as in Heindl(2001). Two different arrival processes are chosen for the test queues. One is the interrupted Poisson process (IPP) as a special case of MMPP with parameters $r_0 = 0.9$, $r_1 = 0.1$, $\lambda_0 = 5$, and $\lambda_1 = 0$ following the standard notations used in the literature; see Heindl(2001) for details. The mean and the SCV of the inter-arrival times is 2 and 10 respectively. The other arrival process is the balanced-mean hyperexponential(H2) inter-arrival times with the same first two moments as the IPP, i.e. combination of two exponential distributions one with mean of 1.05 with probability 1/1.05 and the other with mean of 20.95 with probability 1/20.95. In fact, it can be shown that the above IPP is equivalent to a hyperexponential inter-arrival times with non-balanced mean, i.e. one with mean of 11.832 with probability 0.157 and the other with mean of 0.169 with probability 0.843. The service time is uniform(0, 2) distribution at the first station and deterministic with mean 1 at the second station. In order to test the performance of our formula, the IDI sequences estimated from simulation are given in Table 4 along with the average waiting time.

In table 5, the approximated IDI sequences are given based on $g(k) = 2 + 1.2 \ln k$ for both IPP an H2 interarrival times. The result shows good performance

Table 4. Estimated IDI and E(W) for tandem queues

	Hyperexponential with balanced mean			IPP		
k	$c_{a1}^2(k)$	$c_{d1}^2(k)$	$c_{d2}^2(k)$	$c_{a1}^2(k)$	$c_{d1}^2(k)$	$c_{d2}^2(k)$
1	10.18*	8.85	8.38	10.01	5.68	5.13
10	10.09	9.25	8.96	9.93	6.78	6.44
50	9.97	9.64	9.53	9.83	8.42	8.24
100	9.68	9.48	9.42	9.74	9.07	8.96
200	9.91	9.80	9.80	10.44	10.20	10.13
300	10.22	10.08	10.04	10.31	9.93	9.88
400	9.84	9.83	9.85	9.60	9.49	9.45
500	8.69	8.64	8.64	10.39	10.04	9.96
1000	9.60	9.45	9.44	9.51	9.37	9.34
E(W)	2.90	0.98	-	8.92	1.39	-

*Based on 120,000 observations of inter-arrival times

Table 5. Approximation of $c_{dj}^2(k)$ with $g(k)$ = 2 + 1.2 ln k

	Hyperexponential with balanced mean		IPP	
k	$c_{d1}^2(k)$	$c_{d2}^2(k)$	$c_{d1}^2(k)$	$c_{d2}^2(k)$
1	7.72	5.79	7.59	5.69
	(-0.13)*	(-0.31)	(0.34)	(0.11)
10	9.73	9.37	9.58	9.22
	(0.05)	(0.05)	(0.41)	(0.43)
50	9.88	9.78	9.74	9.64
	(0.02)	(0.03)	(0.16)	(0.17)
100	9.63	9.58	9.69	9.64
	(0.02)	(0.02)	(0.07)	(0.08)
200	9.88	9.85	10.41	10.38
	(0.01)	(0.01)	(0.02)	(0.02)
300	10.20	10.18	10.29	10.27
	(0.01)	(0.01)	(0.04)	(0.04)
400	9.82	9.81	9.58	9.57
	(0.00)	(0.00)	(0.01)	(0.01)
500	8.68	8.67	10.38	10.36
	(0.00)	(0.00)	(0.03)	(0.04)
1000	9.59	9.58	9.50	9.50
	(0.02)	(0.02)	(0.01)	(0.02)

*Number in ( ) is relative error.

for H2 inter-arrival times except for k = 1. On the other hand the result is not as good in IPP arrival process.

It is worthwhile to note that the average waiting time under IPP is significantly greater than the case of H2 inter-arrival times even though the two arrival processes have the mean and variance of inter-arrival times, that is, λ = 5 and SCV =10. One key difference is that the non-centered third moment of the IPP is 1,560 whereas that of H2 is 2,640. This result is consistent with the claim in Wolff(2003) that average waiting time decreases with the increase of the third moment of inter-arrival times. The result partially validates our formula but also presents the limitation of two-moment decomposition method. Another interesting result can be found in the second queue where the average waiting time is larger for the IPP arrival case even with smaller second moment. In fact, the $c_{a2}^2(k)$ under IPP is shown to be less than $c_{a2}^2(k)$ under H2 for almost all k.

4.2 The Correspondence between $c_a^2(k)$ and traffic intensity

This section briefly discusses the issue of using the IDI sequence for the approximation of average waiting time in queue especially with the bottleneck station in cases A-1 and B-1 described above. In order to use the approximate sequence $c_{a10}^2(k)$ for the approximation of $\mathrm{E}(W_{10})$, k needs to be chosen. Whitt (1981) proposed

$$k(\rho) = \frac{\rho c_a^2(\infty)}{2(1-\rho)^2} \tag{4}$$

where k depends on $c_a^2(\infty)$ and ρ but not on c_s^2. Our result shows that the correspondence between $c_{aj}^2(k)$ and $\mathrm{E}(W_j)$ is not quite consistent with (4). By simulation estimate of $\mathrm{E}(W_j)$ together with (3) the imputed SCV, $\hat{c}_{aj}^2$, can be determined as follows:

$$\hat{c}_{aj}^2(\rho_j) = \left(\frac{2\lambda_j(1-\rho_j)}{\rho_j^2}\mathrm{E}(W_j) - c_{sj}^2\right)^+$$

Then, the imputed SCV and the IDI sequence $c_{aj}^2(k)$ obtained from simulation can be compared with each other to determine the relevant lag $\hat{k}$ given c_s^2 and ρ_j. Table 6 shows the value of $k(\rho)$ for bottleneck station

Table 6. Correspondence between IDI and ρ_{10}

Case	ρ_{10}	0.7	0.8	0.9	0.95
A-1	E(W_{10})	3.05	8.35	28.3	69.01
		(0.06)*	(0.23)	(1.66)	(6.23)
	$\hat{c}^2_{a10}$	2.74	4.22	5.99	6.65
	$k(\rho_{10})$	31	80	360	1520
	$\hat{k}$	5	20	100	200
B-1	E(W_{10})	2.83	6.85	19.71	40.83
		$\hat{c}^2_{a10}$	(0.21)	(0.88)	(2.05)
	$\hat{c}^2_{a10}$	2.47	3.28	3.87	3.52
	$k(\rho_{10})$	4	10	45	190
	$\hat{k}$	1	10	20	700

*The number in ( ) is 90% confidence half width.

under ρ_{10} = 0.7, 0.8, 0.9 and 0.95 for cases A-1 and B-1. Recall that $c^2_{a10}(\infty)=8$ for case A-1 and that $c^2_{a10}(\infty)=1$ for case B-1. For case A-1, the imputed SCV increases monotonically with respect to ρ_{10}. Since $c^2_{a10}(k)$, $k \geq 1$, is also monotonically increasing, one-to-one mapping is possible between ρ_{10} and k; see Figure 1. For each ρ_{10} = 0.7, 0.8, 0.9 and 0.95, $k(\rho_{10})$ is significantly larger than $\hat{k}$. This implies that waiting time can be overapproximated by (4). In case B-1, on the other hand, the imputed SCV increases and then decreases with respect ρ_{10}. This is consistent with $c^2_{a10}(k)$, $k \geq 1$, obtained from the simulation; see Figure 2. This unimodality complicates the correspondence between the relevant value of k and the traffic intensity. As for ρ_{10} = 0.95, $\hat{k}$ could possibly be as large as 700 whereas Whitt's formula gives $k(\rho_{10})$ = 190.

As pointed out in previous section, the third moment can also make a significant difference in average waiting time even under the same first two moments. Therefore, trying to establish the correspondence between autocorrelation and traffic intensity would be meaningless under such situations as both in the first and second queue of our test model.

5. Discussions and conclusions

The IDI sequence, $c^2(k)$, is a measure of cumulative autocorrelation among inter-arrival times. It has been proposed as a variability measure for arrival processes to be used in the approximation of waiting time in queue. However, simulation is the only way to obtain the IDI sequence of inter-departure times for a queueing network with non-renewal arrival processes and general service times. In this paper we propose an approximation method for the IDI sequence of inter-departure times in queueing networks. Our approach is based on regression analysis of the IDI data obtained from simulation of 10-station tandem networks. Two extreme cases, one with highly variable arrival process and the other with highly variable service times, are considered and are shown to have similar regression estimates of the parameters of the variability function that can be used in queueing operation of the decomposition approximation of general open queueing networks. By comparing the IDI sequence and the average waiting time under various traffic intensities at the bottleneck station, the range of relevant lag of autocorrelations of inter-arrival times can be determined as a function of the traffic intensity and other parameters of the system. A test model shows the validity of our proposed variability function but also presents a potential limitation of two-moment approximation with an example of queue where the third moment determines the average waiting time.

In order to apply our result to general open queueing networks, however, other factors should be taken into consideration such as lag-k correlation between split processes. Also Bitran and Tirupati's(1988) and Whitt's refinement(1994) of decomposition method for deterministic routing can be combined with our result as in Kim(2005). Another direction of future research is the validation of (4) proposed by Whitt(1995). It is well known that the SI SCV $c^2_a(1)$ is appropriate for estimating E(W) under moderate traffic intensity and that the ASY SCV $c^2_a(\infty)$ should be used under heavy traffic. We leave the problem of identifying this correspondence for future research. Moreover, the characterization of the autocorrelation structure of departure process from a queue with specific service times and Markovian arrival processes can be used for further refinement of decomposition method; Shioda(2003) and Yeh, and Chang(2000). Further improvement also requires taking into account of the third moment of inter-arrival times which is shown to be a critical factor under some situations when the arrival is highly variable.

REFERENCES

1. Albin, S.L. (1984), "Approximating a point process by a renewal process, II: superposition arrival processes to queues", *Operations Research,* vol. 32, pp. 1133-1162.
2. Bitran, G.R. and Tirupati, D. (1988), "Multiproduct queueing networks with deterministic routing: decomposition approach and the notion of interference", *Management Science*, vol. 34, pp. 75-100.
3. Buzacott, J.A. and Shanthikumar, J.G. (1993), "Stochastic models of manufacturing systems". Prentice Hall.
4. Fendick, K.W., Saksena, V.R. and Whitt, W. (1989), "Dependence in packet queues", *IEEE Transactions on Communications,* vol. 37, pp. 1173-1183.
5. Fendick, K.W., Saksena, V.R. and Whitt, W. (1991), "Investigating dependence in packet queues with the index of dispersion for work", *IEEE Transactions on Communications,* vol. 39, pp. 1231-1244.
6. Jagerman, D., Baris, B., Altiok, T., and Melamed, B. (2004), "Mean waiting time approximations in the G/G/1 queue", *Queueing Systems*, vol. 46, pp. 481-506.
7. Heindl, A. (2001), "Decomposition of general tandem queueing networks with MMPP input", *Performance Evaluation*, vol. 44, pp. 5-23.
8. Kim, S. (2000), *Refined parametric decomposition approximation of queueing networks*, Ph.D thesis, Purdue University.
9. Kim, S. (2004), "The heavy traffic bottleneck phenomenon under splitting and superposition", *European Journal of Operational Research*, vol. 157, pp. 736-745.
10. Kim, S., Muralidharan, R. and O'Cinneide, C. (2005), "Taking account of correlations between streams in queueing network approximations", *Queueing Systems*, vol. 49, pp. 261-281.
11. Kim, S. (2005), "Approximation of multi-class queueing networks with highly variable arrivals under deterministic routing", *Naval Research Logistics*, vol. 52, pp.399-408.
12. Kuehn, P. (1979), "Approximate analysis of general queueing networks by decomposition", *IEEE Transactions on Communications*, vol. 27, pp. 113-126.
13. Livny, M., Melamed, B., Tsiolis, A. (1993), "The Impact of autocorrelation on queueing systems", *Management Science*, vol. 39(3), pp. 322-339.
14. Shanthikumar, J.G. and Buzacott, J.A. (1981), "Open queueing network models of dynamic job shops", *International Journal of Production and Research*, vol. 19, pp. 255-266.
15. Shioda, S. (2003), "Departure process of the MAP/SM/1 queue", *Queueing Systems*, vol. 44, pp. 31-50.
16. Sriram, K. and Whitt, W. (1986), "Characterizing superposition arrival processes in packet multiplexers for voice and data", *IEEE Journal on Selected Areas in Communications, SAC*, vol. 4, pp. 833-846.
17. Suresh, S. and Whitt, W. (1990a), "Arranging queues in series: a simulation experiment", *Management Science*, vol. 36, pp. 1080-1091.
18. Suresh, S. and Whitt, W. (1990b), "The heavy traffic bottleneck phenomenon in open queueing networks", *Operations Research Letter*, vol. 9, pp. 355-362.
19. Whitt, W. (1981), "Approximating a point process by a renewal process: The view through a queue, an Indirect Approach", *Management Science*, vol. 27, pp. 619-636.
20. Whitt, W. (1982), "Approximating a point process by a renewal process, I: Two basic methods", *Operations Research*, vol. 30, pp. 124-147.
21. Whitt, W. (1983), "The Queueing Network Analyzer", *The Bell System Technical Journal*, vol. 62, pp. 2779-2815.
22. Whitt, W. (1994), "Towards better multi-class parametric-decomposition approximations for open queueing networks", *Annals of Operations Research*, vol. 48, pp. 221-248.
23. Whitt, W. (1995), "Variability functions for parametric-decomposition approximations of queueing networks", *Management Science*, vol. 41, pp. 1704-1715.
24. Wolff, R. W. and Wang, C.-L. (2003), "Idle period approximations and bounds for the GI/G/1 queue", *Advances in Applied Probability*, vol. 35, pp. 773-792.
25. Yeh, P.-C. and Chang, J.-F. (2000), "Characterizing the departure process of a single server queue from the embedded markov renewal process at departures", *Queueing Systems*, vol. 35, pp. 381-395.

김 선 교 (sunkyo@ajou.ac.kr)

1987 서울대학교 경영학과 학사
1989 서울대학교 경영학과 석사
1994 University of Illinois, Urbana-Champaign응용수학 석사
1996 University of California, Berkeley, 산업 공학 석사
2000 Purdue University, 산업 공학 박사
2001～현재 아주대학교 경영대학 대우조교수, 조교수, 부교수

관심분야 : 대기행렬 네트워크 근사이론, 시뮬레이션