Probing a M/G/1 Queue with General Input and Service Times

Nelson Antunes
CEMAT/University of Algarve
nantunes@ualg.pt

Gonçalo Jacinto
CIMA-UE/Évora University
gjcj@uevora.pt

António Pacheco
CEMAT/IST-TULisbon
apacheco@math.ist.utl.pt

ABSTRACT

We consider the estimation of the arrival rate and the service time moments of a $M/G/1$ queue with probing, i.e., with special customers (probes) entering the system. The probe inter-arrival times are i.i.d. and probe service times follow a general positive distribution. The only observations used are the arrival times, service times and departure times of probes. We derive the main equations from which the quantities of interest can be estimated. Two particular probe arrivals, deterministic and Poisson, are investigated.

1. INTRODUCTION

The problem addressed is motivated by the estimation of the characteristics of traffic in a bottleneck link/router modeled through a $M/G/1$ queue using probing packets. The assumption that (regular) packet arrivals follow a Poisson process is reasonable if traffic is generated by a large number of sources, each with a low peak rate relative to the link capacity. It is known that the distribution of packet sizes is very far from exponential and shows a bimodal structure on the min. (40B) and max. (around 1500B) sizes. If the buffer size is sufficiently large, infinite capacity can be assumed.

As pointed out in [1] there is a small literature on the study of classical queueing systems when probes are intrusive (i.e., non-zero packet sizes). An early paper [4] estimates the traffic intensity of a $M/G/1$ queue assuming that times between probe arrivals are general distributed. For this queue, [1] estimates the arrival rate and service distribution under Poisson probing. Other quantity, like the residual processing capacity of a $M/G/1$, has been estimated in [2] for a particular probing method (see also [3]).

The main contribution of this paper is to consider an arbitrary arrival distribution between probes for the estimation of the arrival rate and the service time moments of a $M/G/1$ queue. This allows to compare different probing streams. We study the properties of the estimators in the case of deterministic and Poisson probing. We focus on variance of the estimators since that is the best distinguishing factor among unbiased estimators.

2. THE MODEL

We consider a FIFO $M/G/1$ queue where both the arrival rate, λ, and service distribution of (regular) customers are unknown. Let S be the service time of a customer. Additionally, probes (special customers) arrive to the queue

according to a renewal process with rate λ_p and general positive service distribution (different from regular customers). Let A_i, S_i, and D_i, $i \geq 1$, denote the arrival time, service time and departure time of the ith probe entering the system, respectively. These are the only probe related quantities observed.

3. MAIN EQUATIONS AND ESTIMATION

Let $\{N_n\}_{n \geq 1}$ be the sequence of probes that see the arrival of the next consecutive probe before departing from the queue,

$$N_n = \inf\{i : i > N_{n-1}, D_i > A_{i+1}\}, \qquad (1)$$

$n \geq 1$ and $N_0 = 0$. Let

$$T_n = A_{N_n+1} - A_{N_n} \qquad (2)$$

denote the inter-arrival time between the N_nth and $(N_n + 1)$th probes, which are together in the queue at some point in time, and

$$W_n = D_{N_n+1} - D_{N_n} - S_{N_n+1} \qquad (3)$$

the amount of work of (regular) customers that arrive between both probes, during the interval (A_{N_n}, A_{N_n+1}) of length T_n. We let

$$I_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{W_i=0\}} \quad \text{and} \quad L_n(k) = \sum_{i=1}^{n} W_i^k \bigg/ \sum_{i=1}^{n} T_i . \quad (4)$$

If $N_n \to \infty$ and $\sum_{i=1}^{n} T_i \to \infty$, as $n \to \infty$, equations in (4) converge almost surely respectively to

$$E[I] = P(W=0) = E[e^{-\lambda T}] \quad \text{and} \quad E[L(k)] = \frac{E[W^k]}{E[T]}, \quad (5)$$

where I, $L(k)$, W and T denote the limiting random variables of the sequences (I_n), $(L_n(k))$, (W_n) and (T_n), resp. Note that $(W|T)$ is a compound Poisson sum with LST

$$E[e^{-sW}|T] = \exp(-\lambda T(1 - E[\exp(-sS)])). \qquad (6)$$

Writing the right equation in (4) for $k = 1, \ldots, 4$ and deriving the moments of W with the LST (6), we have

$$E[S] = E[L(1)]/\lambda, \qquad (7)$$

$$E[S^2] = (E[L(2)]E[T] - E[L(1)]^2 E[T^2])/(\lambda E[T]), \qquad (8)$$

$$E[S^3] = (E[L(3)]E[T] - 3\lambda E[L(1)]E[S^2]E[T^2]$$
$$- E[L(1)]^3 E[T^3])/(\lambda E[T]), \qquad (9)$$

$$E[S^4] = (E[T]E[L(4)] - \lambda E[T^2](3\lambda E[S^2]^2 - 4E[L(1)]E[S^3])$$
$$- E[L(1)]^4 E[T^4] - 6\lambda E[L(1)]^2 E[S^2]E[T^3])/(\lambda E[T]). \quad (10)$$

Equations for other higher moments of the service time of customers can be obtained.

Suppose that, from the observable quantities of probes, $\{(W_k, T_k)\}_{k=1}^n$ is available. The LST of T is not known (and differs from $A_{i+1} - A_i$, $i \geq 1$) in general and the empirical version of the left equation in (5) is implemented as

$$I_n = \frac{1}{n} \sum_{k=1}^n e^{-\widehat{\lambda}_n T_k}. \qquad (11)$$

Solving (11) numerically with respect to $\widehat{\lambda}_n$ provides an estimation for λ. Equations (7) to (10) give the estimators of the service time moments $\widehat{E}_n[S^k]$, $k = 1, \ldots, 4$, where the quantities on the left-hand-sides are replaced by their empirical counterparts (e.g., $E[L(1)]$, $E[T]$, λ are replaced by $L_n(1)$, $\sum_{k=1}^n T_k/n$, $\widehat{\lambda}_n$, resp.). Two particular probe arrivals are considered next.

4. DETERMINISTIC PROBING

We assume that the inter-arrival times between consecutive probes $A_{i+1} - A_i$, $i \geq 1$, are deterministic and equal to $1/\lambda_p$. In this case $T_n = 1/\lambda_p$.

From the left equation in (5), equating I_n with $E[e^{-\lambda T}] = e^{-\lambda/\lambda_p}$, gives an estimator for λ,

$$\widehat{\lambda}_n = -\lambda_p \log I_n. \qquad (12)$$

Using the Delta method, we have $E[\widehat{\lambda}_n] \approx \lambda$ and

$$V[\widehat{\lambda}_n] \approx \frac{\lambda_p^2 V[I_n]}{E[I_n]^2} = \frac{\lambda_p^2}{n} \left(e^{\lambda/\lambda_p} - 1 \right), \qquad (13)$$

where $E[I_n]$ and $V[I_n]$ are given in Appendix. Hence, $\widehat{\lambda}_n$ is a consistent estimator of λ and its variance does not depend on the service times of probes and (regular) customers.

Using the empirical counterparts for $E[L(1)]$ and the estimator $\widehat{\lambda}_n$ of λ in (7), we have the following approximately unbiased estimator of $E[S]$,

$$\widehat{E}_n[S] = -\frac{L_n(1)}{\lambda_p \log I_n}. \qquad (14)$$

Assuming that I_n and $L_n(1)$ provide from two independent samples, the extension of the Delta method to the multivariate case, gives

$$V[\widehat{E}_n[S]] \approx \frac{V[L_n(1)]}{\lambda_p^2 (\log E[I_n])^2} + \frac{E[L_n(1)]^2 V[I_n]}{\lambda_p^2 E[I_n]^2 (\log E[I_n])^4} \qquad (15)$$

$$= \frac{\lambda_p^2 (e^{\lambda/\lambda_p} - 1) E[S]^2}{n\lambda^2} + \frac{\lambda_p E[S^2]}{n\lambda}, \qquad (16)$$

where $E[L_n(1)]$ and $V[L_n(1)]$ are given in the Appendix and $\widehat{E}_n[S]$ is a consistent estimator.

In the case of deterministic probing, the empirical counterparts of (8) gives

$$\widehat{E}_n[S^2] = \frac{L_n^2(1) - \lambda_p L_n(2)}{\lambda_p^2 \log I_n}. \qquad (17)$$

Proceeding in a similar way as above, the estimator $\widehat{E}_n[S^2]$ is approximately unbiased and consistent with variance

$$V[\widehat{E}[S^2]] \approx \frac{4E[L_n(1)]^2 V[L_n(1)]}{\lambda_p^4 (\log E[I_n])^2} + \frac{V[L_n(2)]}{\lambda_p^2 (\log E[I_n])^2}$$
$$+ \frac{(E[L_n(1)]^2 - \lambda_p E[L_n(2)])^2 V(I_n)}{\lambda_p^4 (\log E[I_n])^4 E[I_n]^2}. \qquad (18)$$

There is no simple closed form of (18) and the quantities are given in Appendix. We note that $V[L_n(2)]$ involves $E[S^3]$ and $E[S^4]$ which are estimated from the empirical version of (9) and (10). Also note that variances (15) and (18) do not depend on the service times of probes.

The variances of the estimators $\widehat{\lambda}_n$, $\widehat{E}_n[S]$ and $\widehat{E}_n[S^2]$ are strictly convex upward in λ_p. Dividing (13) by λ^2, the variance has a minimum at $\lambda/\lambda_p \approx 1.6$. Since $E[I_n] = e^{-\lambda/\lambda_p}$, the optimal probing to estimate λ is attained when $I_n \approx e^{-1.6}$. If we divide (16) by $E[S]^2$ and use that $E[S^2]/E[S]^2 > 1$, it can be shown that $V[\widehat{E}_n[S]]$ has a minimum at λ/λ_p, with $\lambda/\lambda_p < 2$. However, for the variance in (18) a similar result seems outreach. In practice, to have just one probing rate, a strategy consist in adjust the sending of probes such that $I_n \approx e^{-1.6}$.

5. POISSON PROBING IN HEAVY TRAFFIC

In order to simplify the exposition and due to the limit space, we assume that the inter-arrival times between probes $A_{i+1} - A_i$, $i \geq 1$, follow an exponential distribution with rate λ_p. (The analysis can be done assuming the more general gamma distribution.) We consider that the queue is near saturation, i.e., the load is close to but less than unity. In this case, since the workload in the system is large, each probe tends to see the next probe in the queue before depart. Hence, T_n's are approximately i.i.d. exponential variables with parameter λ_p.

Using the empirical counterpart of $E[I]$ and $E[e^{-\lambda T}] = \lambda_p/(\lambda + \lambda_p)$ in the left equation in (5), leads to

$$\widehat{\lambda}_n = \lambda_p \left(I_n^{-1} - 1 \right). \qquad (19)$$

By the Delta Method, $E[\widehat{\lambda}_n] \approx \lambda$ and

$$V[\widehat{\lambda}_n] \approx \frac{\lambda(\lambda_p + \lambda)^2}{n\lambda_p}, \qquad (20)$$

which goes to zero has $n \to \infty$. From the empirical counterparts of (7), an approximately unbiased estimator of first moment of the service time, is given by

$$\widehat{E}_n[S] = \frac{L_n(1)}{\lambda_p \left(I_n^{-1} - 1 \right)}. \qquad (21)$$

Using the same assumptions as in (15), we have

$$V[\widehat{E}_n[S]] \approx \frac{(\lambda_p + \lambda)^2 E[S]^2}{n\lambda\lambda_p} + \frac{\lambda_p E[S^2]}{\lambda(n-1)}, \qquad (22)$$

with $n > 1$ and hence $\widehat{E}_n[S]$ is a consistent estimator of $E[S]$. Finally, based on the empirical counterpart of (8), an approximately unbiased second moment estimator of S is

$$\widehat{E}_n[S^2] = \frac{\lambda_p L_n(2) - 2L_n^2(1)}{\lambda_p^2 (I_n^{-1} - 1)} \qquad (23)$$

and

$$V[\widehat{E}_n[S^2]] \approx \frac{V[L_n(2)]}{\lambda_p^2 (E[I_n]^{-1} - 1)^2} + \frac{16E[L_n(1)]^2 V[L_n(1)]}{\lambda_p^4 (E[I_n]^{-1} - 1)^2}$$
$$+ \frac{(\lambda_p E[L_n(2)] - 2E[L_n(1)]^2)^2 V(I_n)}{(\lambda_p E[I_n](E[I_n]^{-1} - 1))^4}. \qquad (24)$$

It can be shown that $\widehat{E}_n[S^2]$ is a consistent estimator of $E[S^2]$ by replacing in (24) the quantities in Appendix. Both

λ/load/total load	Deterministic			Poisson	
(w/probes)	Mean	Std. dev.	Std. dev. (approx.)	Mean	Std. dev.
714/0.4/0.45	(713.6, 700.9, 971.6)	(20.0, 26.5, 84.3)	(19.8, 26.1, 85.3)	(714.8, 697.7, 960.5)	(30.1, 42.1, 124.2)
1071/0.6/0.68	(1070.6, 699.8, 967.8)	(29.7, 25.2, 86.7)	(29.8, 26.1, 85.3)	(1071.3, 699.7, 960.2)	(41.3, 36.6, 125.4)
1429/0.8/0.90	(1427.7, 701.3, 972.7)	(40.1, 24.7, 81.9)	(39.7, 26.1, 85.3)	(1433.9, 698.5, 949.8)	(48.9, 30.8, 135.6)

Table 1: Estimation of packet arrival rate and 1st and 2nd moments of packet size $(\lambda, 700, 964.8(\times 10^3))$ with $\lambda/\lambda_p = 1.6$ and mean probe packet size 150B.
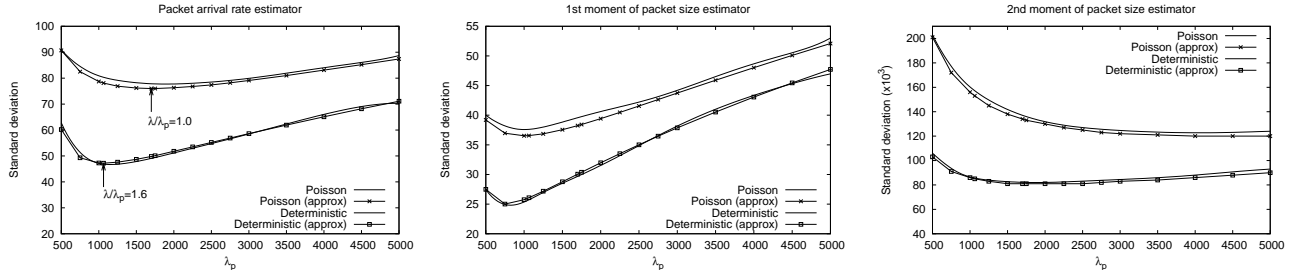

Figure 1: Standard dev. of the estimates in heavy traffic (total load 0.99) for different probing rates with $\lambda = 1700$. The mean estimates for deterministic ($\lambda/\lambda_p = 1.6$) and Poisson ($\lambda/\lambda_p = 1.0$) probing are $(1700.5, 700.5, 967.9(\times 10^3))$ and $(1698.1, 702.0, 968.1(\times 10^3))$, resp.

variances (22) and (24) do not depend on the service times of probes.

The variances of estimators $\widehat{\lambda}_n$, $\widehat{E}_n[S]$ and $\widehat{E}_n[S^2]$ are strictly convex upward in λ_p. Equation (20) has a minimum at $\lambda/\lambda_p = 1$. The $V[\widehat{E}_n[S]]$ has minimum at λ/λ_p, with $\lambda/\lambda_p < 1.4$. The complexity of the variance in (24) does not allow to obtain a similar result. In order to have just one probing rate, a strategy consist in adjusting the sending of probes to the optimal estimation of λ, i.e., when $I_n \approx \lambda_p/(\lambda + \lambda_p) = 0.5$ is observed.

6. EXPERIMENTS

The link is modeled as a single-server FIFO queue with infinite capacity and processing speed $C = 10$Mbps. The packet size P follows a distribution with values 40, 800, 1500B and weights 0.5, 0.1, 0.4, resp. Probe packet sizes are Poisson distributed. We construct a simulation program to obtain the probing observations. Experiments have $n = 2000$ and are repeated 500 times.

Table 1 depicts the estimates of the traffic characteristics for different values of λ (packets/s). (The moments of the service time $S = P \times 8/C$ are converted to packet size). The means and standard deviations of the estimates of the 500 runs are presented. For Poisson probing the estimates are obtained from the estimators in Sect. 3, since the queue is not in heavy traffic. In the case of deterministic probing, the standard deviations are compared with the analytical approximation formulas (13), (16) and (18).

Figure 1 compares the standard deviations of the estimates in heavy traffic for both types of probing varying the probing rate. Using the estimators in Sects. 4 and 5, we compute the estimates in each run (500) and plot their standard deviations against the analytical approximation formulas. We point out the good accuracy of the approximations, the convexity of the standard deviations in λ_p, and the minimum of the standard deviation of $\widehat{\lambda}_n$ as a function of λ/λ_p. As in Table 1, the deterministic probing outperforms the Poisson probing.

7. DISCUSSION AND FUTURE WORK

We considered the estimation of the traffic characteristics in a motivated $M/G/1$ queue assuming a general dis-

tribution between probe arrivals. This allows to compare different probing streams. Results showed that deterministic probing leads to estimators with smaller variances than Poisson probing. Future work includes the comparison of other probing streams and the study of the distribution of time between two consecutive probes which are together in the queue at some point in time (non-deterministic probing when the queue is not in heavy-traffic).

8. APPENDIX

If T_n, $n \geq 1$, are i.i.d then since the arrival of regular customers is Poisson, W_n are also i.i.d. In this case $E[I_n] = E[I]$ and $V[I_n] = E[I](1 - E[I])/n$. Additionally, using the conditional mean and variance formulas, we have $E[L_n(1)] = \lambda E(S)$, $V[L_n(1)] = \lambda E[S^2]E[1/\overline{T}_n(1)]$,

$$E[L_n(2)] = \lambda E[S^2] + \lambda^2 E[S]^2 E\left[\frac{\overline{T}_n(2)}{\overline{T}_n(1)}\right], \qquad (25)$$

and $V[L_n(2)]$ is equal to

$$E[S^4]\lambda E\left[\frac{1}{\overline{T}_n(1)}\right] + \lambda^2(2E[S^2]^2 + 4E[S]E[S^3])E\left[\frac{\overline{T}_n(2)}{\overline{T}_n^2(1)}\right]$$

$$+ 4E[S]^2 E[S^2]\lambda^3 E\left[\frac{\overline{T}_n(3)}{\overline{T}_n^2(1)}\right] + \lambda^4 E[S^2]^4 V\left[\frac{\overline{T}_n(2)}{\overline{T}_n^2(1)}\right], \quad (26)$$

where $\overline{T}_n(k) = \sum_{i=1}^n T_i^k$.

9. REFERENCES

[1] F. Baccelli, B. Kauffmann, and D. Veitch. Inverse problems in queueing theory and internet probing. *Queueing Syst.*, 3(63):59–107, 2009.

[2] S. Y. Nam, S. Kim, and D. K. Sung. Estimation of available bandwidth for an $M/G/1$ queueing system. *Appl. Math. Model.*, 33(8):3299–3308, 2009.

[3] A. Novak, P. G. Taylor, and D. Veitch. The distribution of the number of arrivals in a subinterval of a busy period in a single server queue. *Queueing Syst.*, 53(3):105–114, 2006.

[4] V. Sharma and R. Mazumdar. Estimating traffic parameters in queueing systems with local information. *Perform. Eval.*, 3(32):217–230, 1998.