1997年度日本オペレーションズ・リサーチ学会
秋季研究発表会

Throughput Analysis of a Server for Facsimile Communication Networks

*Ken'ichi KAWANISHI *NTT Multimedia Network Service Sector*
Yoshitaka TAKAHASHI *NTT Multimedia Networks Laboratories*

1 Introduction

We present a throughput analysis of a server in the NTT's facsimile communication network, so-called F-NET. The server deals with multi-address calls; see [2, 4]. By a *multi-address call* we mean a facsimile (fax) call that requires multi-address delivery service. From the teletraffic point of view, multi-address calls can make the F-NET heavily congested, and the server leads to a bottleneck. Focusing on the heavy traffic we investigate the departure process of the server via the renewal theory [1]. We obtain the asymptotic result for the departure process. Applying the diffusion approximation [3, 5] together with the asymptotic result enables us to estimate the throughput as well as the performance (queueing) analysis of the server.

2 Operation Mechanisms

The system consists of a single server and a queue with an infinite capacity. Customers arrive at the server and then proceed to the subsequent resource (channels). The customers correspond to multi-address calls which form a batch input. The server's job is to "extract" the individual calls (individual customers) contained in the multi-address call. This server (we call it "extraction server") plays an important role in multi-address delivery service in F-NET. Contrast to the most queueing system, the extraction server does not continuously monitor the queue. The server periodically wakes up and sleeps down and the queue is reviewed at an equally spaced time epoch. Moreover, the server does not necessarily do its job at each wake-up time. The server has a memory to store the number of the calls that have previously extracted. The extraction server sets its counter as many as periods proportional to the value of the memory. Whenever the counter is zero, extractions take place if there are any multi-address calls in queue. The determination how many individual calls are extracted depends on the state of the channels. Basically, the number of extracted calls equals the number of idle channels. However, the server cannot extract individual calls more than a constant l which is a limit of the extraction number at a time. If all the individual calls contained in the multi-address call are not extracted at a time, the multi-address call is queued at its tail of the queue (round-robin queueing mechanism).

3 Throughput Analysis

We denote by G_n the number of extracted calls which is a random variable assigned at time t_n, where t_n is the point when n-th extraction is carried out. We also denote by τ_n the time interval between adjacent extraction points, $\tau_n = t_n - t_{n-1}$. Because the operation mechanism that the extraction server is forced to wait for a period proportional to the number of the previous extracted calls holds at each extraction epoch, τ_n and G_n are related to the following equation

$$\tau_n = T(1 + \beta G_n), \tag{1}$$

where β is a coefficient of waiting factor. The equation (1) indicates that the time intervals are random variables which are related to the previous extracted calls. Thus the extracted calls and the interval are correlated each other. Obviously, the sequence $\{G_n\}$ does not form renewal process. However, it can be said that the departure process is well described by the sequences $\{G_n\}$ and $\{\tau_n\}$. In the following, we analyze this process using an approximated method and derive the mean throughput.

Let us denote by $p_k^{(n)}$ probability that k channels are busy just before the n-th extraction is carried out. Similarly, $q_k^{(n)}$ is probability that k channels are busy just after the n-th extraction has been carried out. Denoting the survivor probability by p, the probabilities $p_k^{(n)}$ and $q_k^{(n)}$ must satisfy the following equations,

$$p_k^{(n)} = \sum_{i=l}^{S} \binom{i}{k} p^k (1-p)^{i-k} q_i^{(n-1)}, \quad (2)$$

$$q_k^{(n)} = \begin{cases} p_{k-l}^{(n)} & l \leq k \leq S-1 \\ \sum_{i=0}^{l} p_{S-l+i}^{(n)} & k = S, \end{cases} \quad (3)$$

where S is the number of the total channels and l the maximum extraction number at a time. We denote by $N(t)$ the number of extraction points until time t, which can be written

$$N(t) = \max\{n : S_n \leq t\}, \quad (4)$$

where S_n is given by $\sum_{i=1}^{n} \tau_i$ and $S_0 \equiv 0$. The cumulative extracted calls $V(t)$, which is equal to the number of individual calls departed from the extraction server until time t, is written

$$V(t) = \sum_{i=1}^{N(t)} G_i. \quad (5)$$

The mean number of individual calls departed from the extraction server until time t is given by

$$\mathrm{E}[V(t)] = \mathrm{E}[N(t)]\mathrm{E}[G]. \quad (6)$$

Then we are naturally led to the definition of the mean throughput α as

$$\alpha = \lim_{t\to\infty} \frac{\mathrm{E}[V(t)]}{t}. \quad (7)$$

In the equilibrium state, we assume that the interval between extractions is allowed to set the mean interval over equilibrium probability of the number of busy channels just before an extraction takes place. Thus we adopt an approximation that the survivor probability p is set to $\exp(-\mu\langle\tau\rangle)$, where $\langle\cdot\rangle$ denotes the average over the equilibrium probability. Passing through the mean interval $\langle\tau\rangle$ between extractions, the expected number of the extracted calls is given by

$$g(\langle\tau\rangle) = \sum_{i=0}^{l} i\pi_{S-i} + l \sum_{i=l+1}^{S} \pi_{S-i}, \quad (8)$$

$$\langle\tau\rangle = T(1+\beta g(\langle\tau\rangle)). \quad (9)$$

where $\pi_k (k = 0, 1, 2, \cdots, S)$ are the equilibrium probabilities which provide the number of busy channels just before the extraction takes place. The asymptotic behavior of $\mathrm{E}[N(t)]$ as t tends to ∞ is evaluated

$$\mathrm{E}[N(t)] \sim t/\langle\tau\rangle. \quad (10)$$

Thus the approximated mean throughput is given by

$$\alpha = g(\langle\tau\rangle)/\langle\tau\rangle. \quad (11)$$

4 Mean Waiting Time

Using the diffusion approximation, we evaluate the mean waiting time. We propose the diffusion parameters in our model as follows [3, 5]

$$a = \lambda(\mathrm{E}[X]^2 C_a^2 + \mathrm{Var}[X]) + \langle\tau\rangle^{-1}(\mathrm{E}[G]^2 C_s^2 + \mathrm{Var}[G]), \quad (12)$$

$$b = \lambda\mathrm{E}[X] - \alpha, \quad (13)$$

where C_a^2 and C_s^2 represent the squared coefficient of variation of the interarrival and departure times. Focusing on a Poisson batch arrival, we finally obtain the mean waiting time W_q for individual calls as

$$W_q = \frac{1}{\alpha}\left(\frac{\frac{\mathrm{E}[X^2]}{\mathrm{E}[X]} + \frac{(\mathrm{E}[G]^2 C_s^2 + \mathrm{Var}[G])}{\mathrm{E}[G]}}{2(1-\rho)} - 1\right), \quad (14)$$

where $\rho \equiv \lambda\mathrm{E}[X]/\alpha$ is the traffic intensity.

References

[1] D. R. Cox: *Renewal Theory*, Methuen, London (1962).

[2] H. Fuju, T. Konishi, Y. Takeyama, and H. Miyaji: A multi-image communication system design, *NTT R&D*, **43**, 841-848 (1994). (in Japanese)

[3] T. Kimura: A unifying diffusion model for state dependent queues, *Optimization*, **18**, 265-283 (1987).

[4] Y. Nakatani and H. Koyano: NTT's facsimile communications network service, *NTT Review*, **6**, 90-94 (1994).

[5] Y. Takahashi: Queueing analysis for mixed loss and delay systems; Exact and diffusion approximation results, *Trans. IEICE*, **E70**, 1195-1202 (1987).