A Study on the Concurrent Processing System

SETSUO OSUGA*

One motivation for concurrent processing systems stems from ratio of cost to performance. To assure the improved cost/performance and to justify a concurrent processing system, the utilities of central processing unit and peripheral devices as a whole must be made as high as possible. Here, utility is defined as percent of time system or device is busy.

We discuss these utilities when system is composed of one central processing unit and many identical peripheral devices and more particularly we intend to discuss the case of real-time information processing system where peripheral devices undertake main part of processing and one central processing unit controls them.

1. *Utility of System*

Hereinafter, for convenience, simple notations CPU and EPU are used for central processing unit and peripheral device respectively. The model to be analysed is the system composed of one CPU and n EPUs attached to CPU and there are several transactions to be processed waiting in the buffer of CPU. CPU sends one of them to one of idle EPUs. Each EPU, when finished one processing, deliver and inform it through interruption to CPU which, receiving processed transaction from EPU, sends new one again. Meanwhile, a sequence of program must be executed in CPU and this causes queue of EPU waiting for CPU's service. In this model, EPUs are finite number of input sources and CPU is single server. Let us suppose that service time of CPU is constant denoted by α and process time in EPU is of negative exponential distribution with parameter λ. Then, if queue size probability can be obtained, utilities of CPU and EPUs can be derived. Let us denote by p_ν and $\bar{\nu}$ the probability that queue size is ν and mean queue size at steady state respectively and by η_C and η_E the utilities of CPU and EPU. Then definitions of utilities are,

$$\eta_C = 1 - p_0, \qquad \eta_E = 1 - \bar{\nu}/n, \tag{1}$$

2. *Queue Problem*

Let us denote by $p_\nu(x, t)$ the 2 dimensional probability density that queue size is ν at time t and time to end of service is x. Here, "time to end of service is x" means that time to end of service of CPU is within x and $x+h$ where h is small time interval. Then x decreases continuously with time except when abrupt

This paper first appeared in Japanese in Joho Shori (the Journal of the Information Processing Society of Japan), Vol. 5, No. 3 (1964), pp. 142~149.

* Institute of Space and Aeronautical Science, University of Tokyo.

change from 0 to α occurs. Here, $x=0$ means one service end and unless queue size is 0, new service for next EPU begins successively. The queue is treated "first-in, first-out" basis. This transition can be expressed as,

$$p_\nu(h,\ t) \to p_{\nu-1}(\alpha,\ t+h)$$

When $\nu=0$, at $x\to 0$, at abrupt change of x does not occur but x continuously decreases into negative region. This scheme is shown in Fig. 1. The probability that queue size is ν (including what is under processing) and time to end of service is x at time $t+h$ is derived from transition probabilities from t to $t+h$ as following.

First, for the case of $0<x<\alpha$,

$$\begin{aligned}
&p_0(x,t+h)=p_0(x+h,t)(1-n\lambda h)+O(h^2),\\
&p_1(x,t+h)=p_1(x+h,t)[1-(n-1)\lambda h]+O(h^2),\\
&p_\nu(x,t+h)=p_\nu(x+h,t)[1-(n-\nu)\lambda h]+p_{\nu-1}(x+h,t)(n-\nu+1)\lambda h+O(h^2),\\
&p_n(x,t+h)=p_n(x+h,t)+p_{n-1}(x+h,t)\lambda h+O(h^2),
\end{aligned} \tag{2}$$

and for $x=0$ and $x=\alpha$,

$$\begin{aligned}
&p_0(0_-,t+h)=p_1(0_+,t)[1-(n-1)\lambda h]+O(h),\\
&p_1(\alpha,t+h)=p_2(0_+,t)[1-(n-2)\lambda h]+\int_{-\infty}^{0-}p_0(x,t)n\lambda dx+O(h),\\
&p_\nu(\alpha,t+h)=p_{\nu+1}(0_+,t)[1-(n-\nu-1)\lambda h]+O(h),\\
&p_n(\alpha,t+h)=O(h),
\end{aligned} \tag{3}$$

can be obtained.

Expanding $p_\nu(x+h,t)$ in the continuous region of x as

$$p_\nu(x+h,t)=p_\nu(x,t)+h(\partial/\partial x)p_\nu(x,t)+O(h^2),$$

and making h tend to 0, we get

$$(\partial/\partial t)p_\nu(x,t)=(\partial/\partial x)p_\nu(x,t)-(n-\nu)\lambda p_\nu(x,t)+(n-\nu+1)\lambda p_{\nu-1}(x,t) \tag{4}$$

For convenience, replacing $\partial/\partial t$ and $\partial/\partial x$ by D_t and D_x respectively, then,

$$\begin{aligned}
&(D_t-D_x)p_0(x,t)=-n\lambda p_0(x,t),\\
&(D_t-D_x)p_1(x,t)=-(n-1)\lambda p_1(x,t),\\
&(D_t-D_x)p_\nu(x,t)=-(n-\nu)\lambda p_\nu(x,t)+(n-\nu+1)\lambda p_{\nu-1}(x,t),\ (1<\nu<n),\\
&(D_t-D_x)p_n(x,t)=\lambda p_{n-1}(x,t),
\end{aligned} \tag{5}$$

can be obtained. Meanwhile, eqs. (3) become

$$\begin{aligned}
&p_0(0_-,t)=p_1(0_+,t)\\
&p_1(\alpha,t)=p_2(0_+,t)+n\lambda\int_{-\infty}^{0-}p_0(x,t)dx,\\
&p_\nu(\alpha,t)=p_{\nu+1}(0_+,t),\ (1<\nu<n)\\
&p_n(\alpha,t)=0.
\end{aligned} \tag{6}$$

Then, $p_0(x,t)$ and $p_\nu(x,t)$ is integrated by x as following,

$$\int_{-\infty}^{0}p_0(x,t)dx=p_0(t),\quad \int_0^\alpha p_\nu(x,t)dx=p_\nu(t). \tag{7}$$

Here, $p_\nu(t)$ is probability that queue size is ν at time t regardless of x.

As $p_0(x,t)$ tend to 0 when x tend to infinity,

$$\int_{0+}^{\alpha} D_x p_\nu(x,t)dx = \Big[p_\nu(x,t)\Big]_{0+}^{\alpha} = p_\nu(\alpha,t) - p_\nu(0_+,t),$$
$$\int_{-\infty}^{0-} D_x p_0(x,t)dx = \Big[p_0(x,t)\Big]_{-\infty}^{0-} = p_0(0_-,t), \tag{8}$$

and so, by integrating eqs. (4) and using eqs. (5),

$$D_t p_0(t) = -n\lambda p_0(t) + p_1(0_+,t),$$
$$D_t p_1(t) = -(n-1)\lambda p_1(t) + n\lambda p_0(t) - p_1(0_+,t) + p_2(0_+,t),$$
$$D_t p_\nu(t) = -(n-\nu)\lambda p_\nu(t) + (n-\nu+1)\lambda p_{\nu-1}(t) - p_\nu(0_+,t) + p_{\nu+1}(0_+,t),\ (1<\lambda<n),$$
$$D_t p_n(t) = \lambda p_{n-1}(t) - p_n(0_+,t), \tag{9}$$

can be obtained.

We can solve these if we know $p_\nu(0_+,t)$ that can be derived from transition probabilities from $t-\alpha$ to t and expressed as sum of following terms,

(1) Probability that queue size is 0 at time $t-\alpha$ and one input (EPU) arrives in time interval h and then $\nu-1$ inputs arrive in succeeding time interval α.

(2) Probabilities that queue size is k $(2 \le k \le \nu+1)$ at time $t-\alpha$ and one service end in time interval h then $\nu-k+1$ inputs arrive in succeeding time interval α.

Let us denote by $\gamma_\alpha(s,m)$ the probability that s out of m operating EPU end their processes in time interval α, and this is expressed as

$$\gamma_\alpha(s,m) = \binom{m}{s}(1-e^{-\lambda\alpha})^s \bar{e}^{(m-s)\lambda\alpha}.$$

Then,

$$p_1(0_+,t) = n\lambda\gamma_\alpha(0,n-1)p_0(t-\alpha) + \gamma_\alpha(0,n-1)p_2(0_+,t-\alpha)$$
$$p_\nu(0_+,t) = n\lambda\gamma_\alpha(\nu-1,n-1)p_0(t-\alpha) + \sum_{k=1}^{\nu}\gamma_\alpha(\nu-k,n-k)p_{k+1}(0_+,t-\alpha),\ (1<\nu<n)$$
$$p_n(0_+,t) = n\lambda\gamma_\alpha(n-1,n-1)p_0(t-\alpha) + \sum_{k=1}^{n-1}\gamma_\alpha(n-k,n-k)p_{k+1}(0_+,t-\alpha), \tag{10}$$

can be obtained.

In steady state, by putting $D_t p_\nu(t)=0$ in eqs. (9) and denoting $p_\nu(t)$ simply by p_ν,

$$n\lambda p_0 = p_1(0_+,t),$$
$$(n-1)\lambda p_1 = p_2(0_+,t),$$
$$(n-\nu+1)\lambda p_{\nu-1} = p_\nu(0_+,t),$$
$$\lambda p_{n-1} = p_n(0_+,t), \tag{11}$$

can be obtained and substituting them into eqs. (10),

$$np_0 = n\gamma_\alpha(0,n-1)p_0 + (n-1)\gamma_\alpha(0,n-1)p_1,$$
$$(n-1)p_1 = n\gamma_\alpha(1,n-1)p_0 + (n-1)\gamma_\alpha(1,n-1)p_1 + (n-2)\gamma_\alpha(0,n-2)p_2,$$

$$(n-\nu+1)p_{\nu-1}=n\gamma_{\alpha}(\nu-1,n-1)p_0+\sum_{k=1}^{\nu}(n-k)\gamma_{\alpha}(\nu-k,n-k)p_k,\ (1<\nu<n)$$

$$p_{n-1}=n\gamma_{\alpha}(n-1,n-1)p_0+\sum_{k=1}^{n-1}(n-k)\gamma_{\alpha}(n-k,n-k)p_k, \tag{12}$$

can be obtained. From these equations, p_ν for $1\leq\nu\leq n-1$ is expressed by p_0 in the form of

$$p_\nu=\kappa_\nu p_0 \quad (\nu=1,2,\ldots.n-1), \tag{13}$$

where κ_ν depend on ν.

In regard to p_n,

$$p_n=-\sum_{\nu=1}^{n-1}p_\nu+\lambda\alpha\sum_{\nu=0}^{n-1}(n-\nu)p_\nu, \tag{14}$$

can be obtained by solving eqs. (5) including x at steady state and integrating them by x. Substituting them into

$$\sum_{\nu=0}^{n}p_\nu=1, \tag{15}$$

then,

$$p_0=1/[1+\lambda\alpha\sum_{k=0}^{n-1}(n-k)\kappa_k] \tag{16}$$

can be obtained. Thus solving eqs. (12), p_ν and so utilities can be derived.

Examples.

(1) The cases of $n=1$, 2, 3 and 5.

The relationships between $n\lambda\alpha$ and η_c and η_p for the cases of $n=1$, 2, 3 and 5 are shown in Fig. 2. As shown, the more the number of EPUs, the larger the effect of concurrency provided the ratio of mean process rate of CPU to that of n EPUs ($=n\lambda\alpha$) is kept constant.

(2) $n\to\infty$, $\lambda\to0$ in such way $n\lambda$ remains constant.

As

$$\lim_{\substack{n\to\infty\\ \nu\to0}}\gamma_{\alpha}(\nu,n-k)=\lim_{\substack{n\to\infty\\ \nu\to0}}\binom{n-k}{\nu}(1-e^{-\lambda\alpha})^{\nu}e^{-(n-k-\nu)\lambda\alpha}=e^{-\lambda'\alpha}(\lambda'\alpha)^{\nu}/\nu!,\ (\lambda'=n\lambda)$$

can be obtained, eqs. (12) reduce to

$$p_0=e^{-\lambda'\alpha}(p_0+p_1),$$
$$p_1=\lambda'\alpha e^{-\lambda'\alpha}(p_0+p_1)+e^{-\lambda'\alpha}p_2,$$
$$p_{\nu-1}=(\lambda'\alpha)^{\lambda-1}e^{-\lambda'\alpha}(p_0+p_1)/(\nu-1)!+\sum_{k=2}^{\nu}(\lambda'\alpha)^{\nu-k}e^{-\lambda'\alpha}p_k/(\nu-k)!,\ (2<\nu).$$

The solutions of them are

$$p_\nu=p_0[(-1)^{\nu-1}e^{-\lambda'\alpha}/(\nu-1)!+\sum_{k=0}^{\nu-2}e^{(\nu-1-k)\lambda'\alpha}(-1)^k(\lambda'\alpha)^k\{(\nu-k)^ke^{\lambda'\alpha}-(\nu-k-1)^k\}/k!],$$
$$p_0=1-\lambda'\alpha.$$

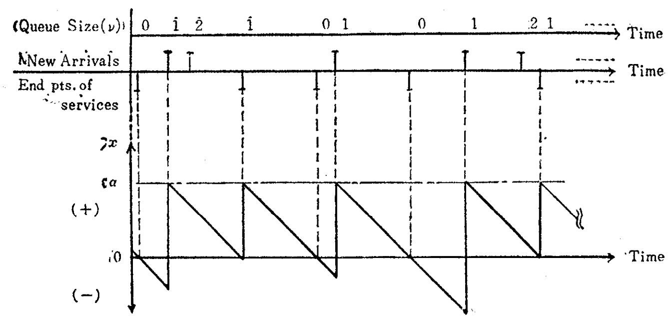

Fig. 1.

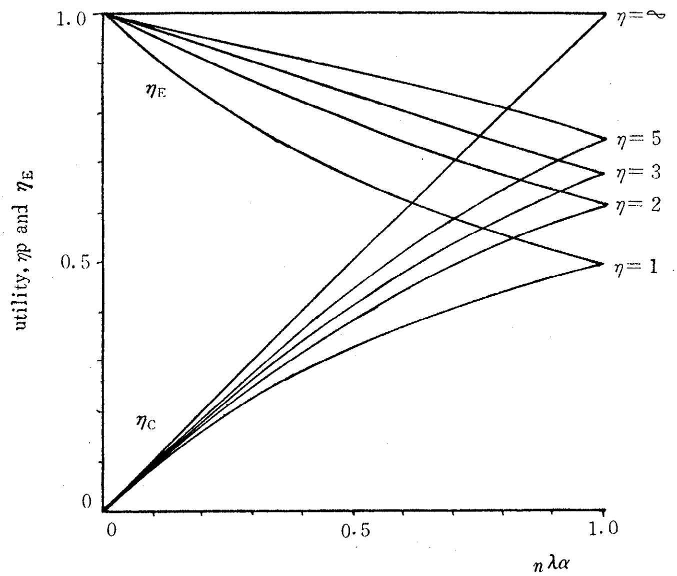

Fig. 2.

References

[1] Feller, W., *An Introduction to Probability Theory and Its Applications, 1.* John Wiley & Sons, 1957.

[2] Khintchine, A., Математические Методы Теории Массового Обслживания, Труды Математического Института Имени В. А. Стеклова, *49*, 1955 (Translated into Japanese).

[3] Takac, L., *Introduction to the Theory of Queues.* Oxford University Press, 1962.