A PARAMETRIC REPRESENTATION OF PROGRAM BEHAVIOR IN A VIRTUAL MEMORY SYSTEM

by

M. Z. Ghanem
and
H. Kobayashi

Computer Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York

Abstract

A new model, which characterizes the stochastic and dynamic behavior of the "working set" of a program running in a virtual memory environment, is represented. The model describes the working set size sequence as a random walk with some kind of restoring tendency. Then, applying the diffusion process approximation to the resultant process, we obtain the equilibrium density function of the working set size, the transient behavior of the working set size and its autocorrelation function.

Introduction

One of the major barriers that hampers the progress of computer system performance evaluation methodology is our inability to characterize quantitatively the system workload. In recent years [1], however, an increasing use of stochastic process representation of program behavior has evolved in studies on system performance evaluation. Modeling of program behavior and its parametric representation can play an important role in developing a general treatment of the resource allocation problem in a multiprogrammed computer system. Among other objectives of studying program behavior are (1) to provide important design parameters--e.g., for stroage hierarchy configurations; (2) to provide more realistic inputs to analytic and simulation models.

The present paper proposes a parametric representation of program behavior in a paging environment. Our model is based on a diffusion process approximation of the working set size sequence.

A Stochastic Model of the Working Set

The term "working set" is often used loosely to mean a set of working pages associated with a process (or task) which keeps the process running efficiently. A more formal definition of the working set is given by Denning [2]: Let the behavior of a program in a virtual memory system be represented by a page reference sequence

$$\underline{r} = r_1, r_2, \ldots, r_i, \ldots$$

The working set $W_i(T)$ at time i is the set of distinct pages referenced in the T most recent references. The integer parameter T is called the window size. The working set size $w_i(T)$ is the size of $W_i(T)$. Various properties of the working set are discussed by Denning and Schwartz [3] under the assumption that $\underline{r}$ is stationary. Most of their results seem extendable to a more general case where we require only the weak stationarity of $\{w_i(T)\}$.

Coffman and Ryan [5] conducted a Monte Carlo Simulation in which LRU stack distance sequences [4] were generated randomly following some forms of the first-order distribution. By converting the stack distance sequences thus generated into the corresponding page reference patterns, they observed that the working set size (for values of T) followed the normal distribution.

The present section proposes a new model of program behavior in a paging environment which attempts to capture the dynamic behavior of the working set. In this section we are interested in a characterization of a program executing in one locality. Within a given locality of the program we assume that the working set size sequence is stationary or at least covariance stationary. Furthermore, it is the objective of this study to describe the "steady-state" behavior. A generalization of the following model to a characterization of transitions between different localities and of the initial transient phase in locality is not considered here; it will be discussed in a separate report [10].

The working set size sequence is representable as

$$w_i = w_{i-1} + Z_i \quad , \quad i = 1,2,\ldots \quad , \tag{2.1}$$

where the steps Z_i can only take the values 1, 0, or -1. Here we suppress T, the window size parameter, for brevity of notation.

If we assume that the incremental changes Z_i are independently, identically distributed with

$$\text{Prob}\{Z_i=1\} = r^+ \quad \text{and} \quad \text{Prob}\{Z_i=1\} = r^- \tag{2.2}$$

then the behavior of the discrete-process $\{w_i\}$ can be modeled as a simple one-dimentional random walk [6], in which each state represents the value of w_i. From its definition, it is clear that w_i is bounded by

$$1 \le w_i \le \min\{N,T\} \quad , \tag{2.3}$$

where T is chosen window size and N is the total number of distinct pages in the locality under consideration. The boundary condition (2.3) is met by assuming reflecting barriers at states 1 and $\min\{N,T\}$. We express Eq. (2.3) for notational convenience in later discussions by

$$-b \le w_i \le a \tag{2.4}$$

where

$$b = -1 \tag{2.5a}$$

$$a = \min\{N,T\} \tag{2.5b}$$

Then the limiting equilibrium distribution of the state occupation probabilities is given by the following truncated geometric distribution:

$$\text{Prob}\{w=\ell\} = \frac{1-r^+/r^-}{1-(r^+/r^-)^a} (r^+/r^-)^{\ell-1} \quad ,$$

$$\ell = 1,2,\ldots,a \quad . \tag{2.6}$$

The simple model described above assumes that values of r^+ and r^- are independent of the current state; this may be acceptable when T is very small, but is clearly unrealistic for a reasonably large value of T (say $T\geq 20$). The probability r^+ that Z_i takes on 1 should be dependent on the state w_{i-1}. A reasonable assumption we can make is that the larger w_{i-1} the smaller r^+. For the same reason we assume the probability r^- should be a monotone increasing function of w. Figure 2.1 depicts the case where $(r^+(w)$ and $r^-(w)$ are both linear functions of w with slopes $-\frac{\beta}{2}$ and $\frac{\beta}{2}$ respectively and take on the common value $\frac{\gamma}{2}$ at $w=w_c$:

$$r^+(w) = \frac{\gamma}{2} - \frac{\beta}{2}(w-w_c) \quad , \tag{2.7}$$

$$r^-(w) = \frac{\gamma}{2} - \frac{\beta}{2}(w-w_c) \quad . \tag{2.8}$$

This model describes the process $\{w_i\}$ as a random walk with some kind of restoring tendency, since the probability of moving one unit toward the center w_c is greater than the probability of moving one unit away from the center by an amount proportional to the distance from the center, i.e., $\beta|w-w_c|$. This formulation was originally suggested by Coffman and Ryan [5], although they have not carried out a detailed analysis.

Let $p(w,i|w_0)$ be the conditional probability that the working set size is equal to w at time i given that it starts from w_0 at time $i=0$. Because of the stationarity assumption made on the process $\{w_i\}$, this conditional probability is equal to the probability that w_{i+i_0} takes on value w given that $w_{i_0}=w_0$, for any i_0. In other words, $i=0$ simply defines the time origin relative to other points. Then we can obtain the following equation:

$$p(w,i+1|w_0) = p(w-1,i|w_0)\, r^+(w-1) + p(w+1,i|w_0)\, r^-(w+1) + p(w,i|w)\,\{1-r^+(w)-r^-(w)\} \quad , \quad -\infty<i<\infty \; . \tag{2.9}$$

This difference equation is rather tedious to solve. Therefore, we will resort to one of the conventional techniques: i.e., we replace the discrete time series w_i by a diffusion process, viz., a continuous-path Markov process $x(t)$, and we find the partial differential equation which defines the process $x(t)$. We now write $p(x,t|x_0)$ instead of $p(w,i|w_0)$ and therefore Eq. (2.9) becomes

$$\begin{aligned} p(x,t+\delta t|x_0) = {} & p(x-1,t|x_0) r^+(x-1)\delta t \\ & + p(x+1,t|x_0) r^-(x+1)\delta t \\ & + p(x,t|x_0)\,\{1-r^+(x)-r^-(x)\}\delta t \end{aligned} \tag{2.10}$$

By taking the limit as $\delta t\to 0$, Eq. (2.10) can be written as:

$$\frac{\partial p(x,t|x_0)}{\partial t} = p(x-1,t|x_0)r^+(x-1) + p(x+1,t|x_0)r^-(x+1) - p(x,t|x_0)[r^+(x)+r^-(x)] \tag{2.11}$$

By applying Taylor series expansion around x with t being fixed, we have the following approximation:

$$p(x-1),t|x_0) \approx p(x,t|x_0) - \frac{\partial}{\partial x} p(x,t|x_0) + \frac{1}{2}\frac{\partial^2 p(x,t|x_0)}{\partial x^2} \tag{2.12}$$

$$p(x+1,t|x_0) \approx p(x,t|x_0) + \frac{\partial}{\partial x} p(x,t|x_0) + \frac{1}{2}\frac{\partial^2 p(x,t|x_0)}{\partial x^2} \tag{2.13}$$

Substituting Eqs. (2.7), (2.8), (2.12) and (2.13) into Eq. (2.11), we obtain the diffusion equation

$$\frac{\partial p}{\partial t} = \left(\frac{\gamma}{2}+\frac{\beta}{2}\right)\frac{\partial^2 p}{\partial x^2} + \beta\frac{\partial}{\partial x}[(x-\bar{x})p] \quad . \tag{2.14}$$

Note that for practical purposes we can assume that $\beta \ll \gamma$, otherwise r^+ and r^- would take negative values for x not far from $\bar{x}$.

$$\frac{\partial p}{\partial t} = \frac{\gamma}{2}\frac{\partial^2 p}{\partial x^2} + \beta\frac{\partial}{\partial x}[(x-\bar{x})p] \quad , \tag{2.15}$$

So the probability density function $p(x,t|x_0)$ of the working set size is governed by the forward Kolmogorov equation. If it were not for boundary conditions, the diffusion process which satisfies Eq. (2.15) would be the so-called Ornstein-Uhlenbeck process [9]. In our problem, however, the working set size is not an unrestricted process like the O-U process; it follows from Eq. (2.4) that

$$-b \leq x(t) \leq a \quad , \quad \text{for all } t \geq 0 \quad , \tag{2.16}$$

The boundary condition (2.16) is equivalent to the following equation [7]

$$\frac{\gamma}{2}\frac{\partial p(x,t|x_0)}{\partial x} + \beta(x-\bar{x})\,p(x,t|x_0) = 0 \quad , \quad \text{at } x=a \text{ and } x=-b \text{ for all } t \; . \tag{2.17}$$

Furthermore, the following initial condition must be met

$$p(x,0|x_0) = \delta(x-x_0) \quad , \tag{2.18}$$

where $\delta(x)$ is Dirac's delta function. The solution of the diffusion equation (2.15) with conditions (2.17) and (2.18) is given by Sweet and Hardin [8]:

$$\begin{aligned} p(Z,t|Z_0) = {} & e^{-\frac{Z^2}{2}}\left[\int_{-Z_b}^{Z_a} e^{-\frac{Z^2}{2}}\,dZ\right]^{-1} \\ & + \sum_{n=1}^{\infty} e^{-\beta\lambda_n t}\left\{e^{-\frac{1}{4}(Z^2-Z_0^2)}\right\} Y(\lambda_n,Z)\,Y(\lambda_n,Z_0) \\ & \left[\int_{-Z_b}^{Z_a} Y^2(\lambda_n,Z)\,dZ\right]^{-1} \quad , \end{aligned} \tag{2.19}$$

where the argument of the probability density is the normalized variable defined by

$$Z = (x-\bar{x})\left(\frac{2\beta}{\gamma}\right)^{\frac{1}{2}} \quad . \tag{2.20}$$

The interval $-b\leq x\leq a$ is transformed accordingly into $-Z_b \leq Z \leq Z_a$, where

$$Z_a = (a-\bar{x}) \left(\frac{2\beta}{\gamma}\right)^{\frac{1}{2}} \quad . \qquad (2.21)$$

and

$$Z_b = (b+\bar{x}) \left(\frac{2\beta}{\gamma}\right)^{\frac{1}{2}} \quad . \qquad (2.22)$$

The values of the parameters λ_n are the zeros of the function

$$y_o(\lambda-1,Z_a) y_e(\lambda-1,Z_b) + y_e(\lambda-1,Z_a) y_o(\lambda-1,Z_b) \qquad (2.23)$$

where

$$y_e(\lambda,x) = e^{-\frac{x^2}{2}} [1+(-\lambda)\frac{x^2}{2!} + (-\lambda)(-\lambda+2)\frac{x^4}{4!} + (-\lambda)(-\lambda+2)(-\lambda+4)\frac{x^6}{6!} + \ldots] \quad , \qquad (2.24)$$

$$y_o(\lambda,x) = e^{-\frac{x^2}{2}} [1+(-\lambda+1)\frac{x^3}{x!} + (-\lambda+1)(-\lambda+3)\frac{x^5}{5!} + (-\lambda+1)(-\lambda+3)(-\lambda+5)\frac{x^7}{7!} + \ldots] \qquad (2.25)$$

and

$$Y(\lambda_n,Z) = y_e(\lambda_n,Z) + \lambda_n \frac{y_o(\lambda_n-1,Z_a)}{y_e(\lambda_n-1,Z_b)} y_o(\lambda_n,Z) \quad . \qquad (2.27)$$

Note that the solution given by (2.19) consists of two parts: the steady state part which is a truncated Gaussian distribution with mean

$$\bar{x} + \frac{\gamma}{2\beta} \left[e^{-\frac{Z_a^2}{2}} - e^{-\frac{Z_b^2}{2}} \right] \left[\int_{-Z_b}^{Z_a} e^{-\frac{Z^2}{2}} dZ \right]^{-1} \quad , \qquad (2.27)$$

and the transient part which is a complicated function of time and x.

Now we wish to compute $\mu(t|x_0)$, the average value of $x(t)$ given in the initial value x_0.

$$\mu(t|x_0) = E[x(t)|x(t_0)=x_0] \qquad (2.28)$$

Multiplying both sides of equation (2.15) by $(x-\bar{x})$ and integrating them from $-b$ to a, we obtain

$$\frac{d}{dt}\mu(t|x_0) = -\beta[\mu(t|x_0)-\bar{x}] - \frac{\gamma}{2}[p(a,t|x_0)-p(-b,t|x_0)] \qquad (2.29)$$

Now by using the transformation

$$\mu(t|x_0) - \bar{x} = \eta(t|x_0)\, e^{-\beta t} \quad , \qquad (2.30)$$

we obtain

$$\frac{d}{dt}\mu(t|x_0) = -\frac{\gamma}{2} e^{\beta t}[p(a,t|x_0)-p(-b,t|x_0)] \quad . \qquad (2.31)$$

But from equation (2.19) $p(a,t|x_0)-p(-b,t|x_0)$ can be written as

$$A + \sum_{n=1}^{\infty} c_n(x_0)\, e^{-\beta\lambda_n t} \quad , \qquad (2.32)$$

where

$$A = \left[\int_{-Z_b}^{Z_a} e^{-\frac{Z^2}{2}} dZ \right]^{-1} \qquad (2.33)$$

and

$$c_n(x_0) = \left\{ e^{-\frac{1}{4}(Z^2-Z_0^2)} Y(\lambda_n,Z_n) Y(\lambda_n,Z_0) - e^{-\frac{1}{4}(Z^2-Z_0^2)} Y(\lambda_n,Z_b) Y(\lambda_n,Z_0) \right\} \cdot \left[\int_{-Z_b}^{Z_a} Y^2(\lambda_n,Z)\, dz \right]^{-1} \quad , \quad n = 1,2,\ldots \quad . \qquad (2.34)$$

Therefore, $\frac{d}{dt}\eta(t|x_0)$ can be written as

$$\frac{d\eta(t|x_0)}{dt} = -\frac{\gamma}{2} e^{\beta t} A + \sum_{n=1}^{\infty} c_n(x_0) e^{-\beta\lambda_n t} \quad , \qquad (2.35)$$

which implies that $\mu(t|x_0)$ can be written as:

$$\mu(t|x_0) = \bar{x} - \frac{\gamma A}{2\beta} + c(x_0) e^{-\beta t} + \sum_{n=1}^{\infty} \frac{c_n(x_0)}{\beta(1-\lambda_n)} e^{-\beta\lambda_n t} \quad , \qquad (2.36)$$

where $c(x_0)$ is an integration constant which is uniquely determined at $t=0$:

$$\mu(0|x_0) = \bar{x} - \frac{\gamma A}{2\beta} + c(x_0) + \sum_{n=1}^{\infty} \frac{c_n(x_0)}{\beta(1-\lambda_n)} \qquad (2.37)$$

and this value, by its definition, must be equal to x_0; and as $t\to\infty$, $\mu(t|x_0) \to \bar{x} - \frac{\gamma A}{2\beta}$ which is the mean of the limiting equilibrium distribution.

Now let us examine the correlation function of the diffusion process $x(t)$:

$$R_x(t_1,t_2) = E[x(t_1)x(t_2)] = E\{x(t_1)E[x(t_2)|x(t_1)]\}, \quad t_2 > t_1$$

$$= E[x(t_1)\mu(t_2-t_1|x(t_1))] \quad . \qquad (2.38)$$

Using the result of equation (2.36), we obtain

$$E[x(t_2)x(t_1)] = E\{x(t_1)[\bar{x} - \frac{\gamma A}{2\beta} + x(x(t_1)) e^{-\beta(t_2-t_1)} + \sum_{n=1}^{\infty} \frac{c_n(x(t_1))}{\beta(1-\lambda_n)} e^{-\beta\lambda_n(t_2-t_1)}]\}$$

$$= (\bar{x} - \frac{\gamma A}{2\beta}) E[x(t_1)] + E[x(t_1)c(x(t_1))] e^{-\beta(t_2-t_1)} + \sum_{n=1}^{\infty} \frac{E[x(t_1)c_n(x(t_1))]}{\beta(1-\lambda_n)} e^{-\beta\lambda_n(t_2-t_1)} \qquad (2.39)$$

For almost all the cases of practical interest, the proability that the working set size will be equal to either of the boundary values is very small. So we can assume here that $\{p(a,t,x(t_1))-p(-b,t,x(t_1))\}$ is very small and therefore $c_n(x(t_1))=0$ for $n=1,2,\ldots$, and $A=0$, which implies that $c(x_0)=x(t_1)-\bar{x}$. Consequently $\mu(t|x_0)$ can be expressed in the form

$$\mu(t_2-t_1|x(t_1))=\bar{x}+(x(t_1)-\bar{x})e^{-\beta(t_2-t_1)} \quad . \tag{2.40}$$

Therefore the autocorrelation function is given by

$$R_x(t_1,t_2)=E[x(t_1)x(t_2)]=\bar{x}E(x(t_1))$$
$$+E[x(t_1)[x(t_1)-\bar{x}]]e^{-\beta(t_2-t_1)}$$
$$=\bar{x}(1-e^{-\beta(t_2-t_1)})E(x(t_1))$$
$$+E(x^2(t_1))e^{-\beta(t_2-t_1)} \quad , \quad t_2\geq t_1\geq 0 \tag{2.41}$$

When both t_1 and t_2 are sufficiently large, the initial condition $x(t_0)=x_0$ is negligible and therefore

$$E[x(t_1)] = \bar{x} \tag{2.42}$$

and

$$E[x^2(t_1)] = \bar{x}^2 + Var[x]$$
$$= \bar{x}^2 + \frac{\gamma}{2\beta} Var[Z]$$
$$= \bar{x}^2 + \frac{\gamma}{2\beta} \quad , \tag{2.43}$$

where the last expression was obtained since the steady state distribution of (2.19) is the unit normal distribution under the assumptions we have made. Therefore, the autocorrelation function is reduced to the following form

$$R_x(t_1,t_2) = R_x(t_1-t_2) = \bar{x}^2 + \frac{\gamma}{2\beta} e^{-\beta|t_1-t_2|} \quad ,$$
$$t_1,t_2>>0 \quad . \tag{2.44}$$

This is a property of a Gauss-Markov process.

Concluding Remarks

In this paper, we have presented a method to obtain a parametric representation of program behavior in a virtual memory system. The working set size sequence is assumed to follow a random walk with some kind of restoring tendency. The equilibrium density function of the working set size can be approximated, using the diffusion approximation, as a truncated normal distribution. The second-order property of the process is also discussed in some detail. We need to verify the validity of this model based on real data of program traces. An empirical study, along with generalization of the mathematical model discussed in this paper, will be reported in a separate report [10].

Acknowledgements

The authors would like to thank Dr. W. D. Frazer and Mr. G. S. Graham for their valuable comments and for their careful reading of the manuscript.

References

[1] H. Kobayashi, "Some Recent Progress in Analytic Studies of System Performance", Proc. of First USA-Japan Computer Conference, October 3-5, 1972, Tokyo, Japan, pp. 130-138.

[2] P.J. Denning, "The Working Set Model for Program Behavior", CACM, Vol. 11, No. 5, May 1968, pp. 323-333.

[3] P.J. Denning and S.C. Schwartz, "Properties of the Working Set Model", CACM, Vol. 15, No. 3, March 1972, pp. 191-198.

[4] R.L. Mattson, J. Gecsei, D.R. Slutz, and I.L. Traiger, "Evaluation Techniques for Storage Hierarchies", IBM Syst. J. 9, 2, pp. 78-117.

[5] E.G. Coffman, Jr. and T.A. Ryan, Jr., "A Study of Storage Partition Using a Mathematical Model of Locality", CACM, Vol. 15, No. 3, March 1972, pp. 185-190.

[6] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley and Sons, Inc., N.Y., 1966.

[7] D.R.Cox and M.D. Miller, The Theory of Stochastic Processes, John Wiley and Sons, Inc., N.Y., 1965.

[8] A.L. Sweet and J.C. Hardin, "Solution for Some Diffusion Processes with Two Barriers", J. Appl. Prob., pp. 423-431, 1970.

[9] G.E. Uhlenbeck and L.S. Ornstein, "On the Theory of the Brownian Motion", Physical Review, Vol. 36, No. 3, pp. 823-841 (1930). Also in N. Wax (ed.), Selected Papers on Noise and Stochastic Processes, Dover Publications (1954).

[10] M.Z. Ghanem, "Measurement of the Working Set and Missing Rate of a Program", (in preparation).

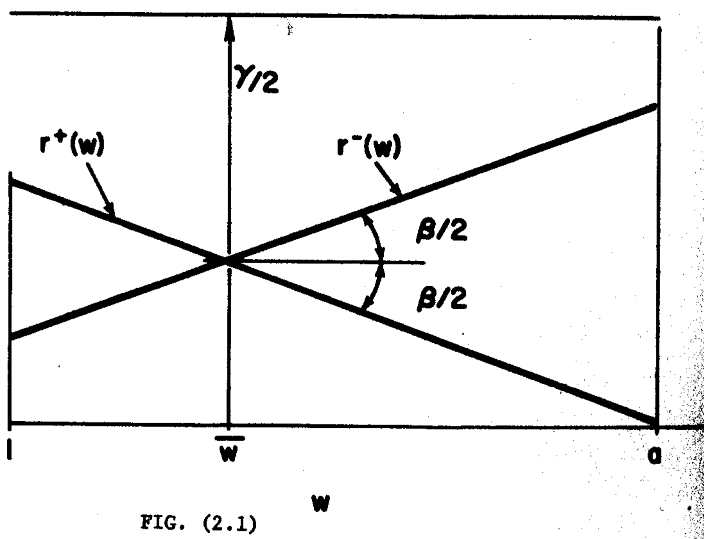

FIG. (2.1)

The Shape of the Restoring Force