(32)

On the Optimum Checkpoint Interval

EROL GELENBE

Université de Paris-Sud, Orsay, France

ABSTRACT. One of the basic problems related to the efficient and secure operation of a transaction oriented file or database system is the choice of the checkpoint interval. In this paper we show that the optimum checkpoint interval (i.e. the time interval between successive checkpoints which maximizes system availability) is a function of the load of the system. We also prove that the total operating time of the system (and not the total real time) between successive checkpoints should be a deterministic quantity in order to maximize the availability. An explicit expression for this time interval is obtained. These results are a significant departure from previous work where load independent results have been obtained. We also present a rigorous analysis of the queueing process related to the requests for transaction processing arriving at the system, and prove the ergodicity conditions for the system.

KEY WORDS AND PHRASES: database systems, failures, rollback recovery, checkpoints, availability, performance modeling and optimization

CR CATEGORIES: 4.33, 4.35, 4.6, 5.5, 8.1

1. *Introduction*

The model we analyze in this paper arises in the context of reliability theory in general. Our motivation is due to the fact that it appears in the study of reliable database systems and transaction oriented computer systems.

Consider a database system (the server) which executes transactions (the customers) being handled in first-come-first-served order. In such systems it is essential to be able to reconstitute the contents of the primary memory to its correct state after a failure occurs which invalidates memory contents. To this effect each transaction is stored in the *audit trail*; a *checkpoint* is established from time to time, and after a failure all transactions stored in the audit trail since the most recent checkpoint are executed once again. A checkpoint is established by copying the contents of primary memory (which can be affected by failures) into some secondary memory device (disk or tape). After a failure, and before all the transactions on the audit trail can be reexecuted, the contents of the checkpoint have to be copied back into primary memory. In [9] this system has been modeled as a queue whose server can be in one of three states; the server state is controlled by a Markov chain. In [8] the case of multiple types of failures and checkpoints has been considered. Young [15], Kovalenko [10], and Chandy [2, 3] present models of rollback recovery in which the queueing delay and the backlog of transaction requests are not taken in account. They assume a constant checkpoint interval and obtain its optimum value, i.e. the value which maximizes system availability. In [2, 3] the failure rate is assumed to be small. Robin [13]

This work was supported by an IRIA/SESORI (Projet Bases de Données Réparties) Research Contract to Université de Paris-Nord.

Author's address: Laboratoire de Recherche en Informatique, Université de Paris-Sud, Centre d'Orsay, 91405 Orsay, France.

considers a problem of optimum control based upon the model studied in the present paper.

A similar model arises in reliability studies. Consider a server which goes through maintenance at predetermined instants of time, and which is subject to failures. Suppose that during maintenance periods and while failures are being repaired, the server cannot serve customers. We assume that the time necessary to repair a failure is a function of the age of the failure with respect to the most recent maintenance epoch. Also suppose that customers are served in their order of arrival. This is precisely the model which is being considered in this paper.

The main practical contribution of this paper is a new formula for the optimum checkpoint interval which includes the effect of the load of the system, and which is appreciably different from the usual formula [15, 2, 3]. This formula gives the optimal *total operation time* between checkpoints excluding the unpredictable intervals of time when the system is recovering from a failure.

The main theoretical contribution of the paper is to introduce a new queueing model in which the duration of service interruptions (due to recoveries from failures) is a function of the past history of the systen.. The ergodicity conditions for this queueing system are established in Section 3.

In Section 2 the system being analyzed is introduced and system availability, defined as the proportion of time it is able to service transactions, is computed. Section 2.1 contains a computation of the optimum checkpoint interval. Section 2.2 is devoted to the interpretation of this result and to the explanation of its practical significance. In Section 2.3 we compute the total time between checkpoints, including the checkpoint interval during which the system is available for service plus the recovery time from failures.

2. *The Rollback Recovery Process*

The service rendered to customers depends on the "state" of the server X_t, $t \geq 0$. which is

$$X_t = \begin{cases} 2 & \text{if the system is creating a checkpoint,} \\ 1 & \text{if it is recovering from a failure,} \\ 0 & \text{if it is operating normally.} \end{cases} \tag{2.1}$$

Service to requests for transaction processing is rendered only in state 0. $\{X_t, \ t \geq 0\}$ is a stochastic point process with the following properties:

(i) The *total* time spent in state 0 between two consecutive transitions to state 2 is a random variable Y independent of the past history of the process with general distribution function $F(y)$ and density $f(y)$. That is, during this time the server is available for serving customers which are in queue. Let $EY \triangleq \int_0^\infty y dF(y) < \infty$.

The random variable Y is the length of time during which the system is available for service between two successive checkpoints, and *not* the total time between checkpoints (which would include the recovery times from failures which may have occurred). The reason for choosing a random quantity Y in our model is that (i) it provides a more general setting for the choice of Y, which we can specialize to a deterministic choice later, and (ii) in practice human or system related factors can introduce some randomness in the choice of Y. In the sequel (Section 2.1) we shall show that a deterministic value of Y can maximize system availability, and we shall obtain an explicit load dependent formula for its optimum value.

(ii) When the process enters state 2 it remains there for a random epoch, independent of the past history, of general distribution function $C(y)$. At the end of this time it returns to state 0. This is the time it takes to create a checkpoint. We suppose that $EC \triangleq \int_0^\infty y dC(y) < \infty$.

(iii) During epochs when the system is in state 0, instants of transition to state 1 are defined by a Poisson process of parameter γ. γ is the failure rate of the server.

(iv) When a transition into state 1 occurs, the time spent by the server in that state until

its return to state 0 is defined as follows. Let $h: R^+ \to R^+$ be a measurable function. For a failure (transition from state 0 to 1) taking place at time t, let

$$t' = \sup\{\sigma : \sigma < t \text{ and } X_\sigma = 2\}.$$

Let the random variable Y_t be defined by

$$Y_t = \begin{cases} \displaystyle\int_{t'}^{t} 1_0(\tau)\,d\tau & \text{if } X_t = 0 \text{ or } 1, \text{ where } 1_0(\tau) = 1 \text{ if } X_\tau = 0 \text{ and } 1_0(\tau) = 0 \text{ otherwise,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

That is, Y_t is the total time spent by the server in state 0 in the interval $[t', t]$, where t' is the most recent time before t at which the server was in state 2. Then the server remains in state 1 (after a failure which occurred at time t) for a time of duration $h(Y_t)$.

Thus the recovery time of the server is a function of the *age* Y_t of the server with respect to the most recent checkpoint.

Returning now more specifically to the stochastic process $(X) = \{X_t,\ t \geq 0\}$ defined by (2.1), (i) to (iv), and (2.2), we are interested in the stationary (equilibrium) probabilities associated with it:

$$\Pi_j \triangleq \lim_{t\to\infty} \Pr[X_t = j], \quad j = 0, 1, 2. \tag{2.3}$$

This limit is easily obtained by noting that the process (X) is regenerative [4]. Recall that a stochastic process is regenerative if, roughly speaking, there exist random times forming a renewal process such that after each of these times the stochastic process replicates itself in the probabilistic sense. The instants at which (X) enters state 2 are these random times.

By application of a well-known theorem [4, Ch. 9, Th. 2.25], we have

$$\Pi_i = (1/m)A_i, \quad i = 0, 1, 2,$$

where m is the average time between successive visits of the process (X) into state 2, and A_i is the average time it spends in state i between successive visits to state 2.

Consider an interval between two successive checkpoints such that the total time spent in state $X_t = 0$ is y. Since failures occur according to a Poisson process, conditioned on n occurrences of failures, the instants of occurrence of the n failures are independent and each is uniformly distributed in the interval; this follows from well-known properties of the Poisson process (cf. [6, pp. 27–28]). Thus the expected value of the length of this interval is

$$EC + y + \sum_{n=0}^{\infty} n\frac{(\gamma y)^n}{n!}e^{-\gamma y}\int_0^y (h(x)/y)\,dx = y\left[1 + \gamma\int_0^y (h(x)/y)\,dx\right] + EC$$

for given y, and its expectation over all values of y is

$$EY + \gamma\int_0^\infty dF(y)\int_0^y h(x)\,dx + EC, \tag{2.4}$$

and therefore

$$\Pi_1 = \frac{\gamma\displaystyle\int_0^\infty dF(y)\int_0^y h(x)\,dx}{EC + EY + \gamma\displaystyle\int_0^\infty dF(y)\int_0^y h(x)\,dx}, \tag{2.5}$$

$$\Pi_0 = \frac{EY}{EC + EY + \gamma \int_0^\infty dF(y) \int_0^y h(x)dx}, \tag{2.6}$$

$$\Pi_2 = 1 - \Pi_1 - \Pi_2. \tag{2.7}$$

Since Π_0 is the stationary probability that the server is available for service, we shall call it the *availability* of the server. In the following section, it will reappear in the ergodicity conditions for the queue length process. We write it in the more convenient form:

$$\Pi_0 = \left[1 + EC/EY + (\gamma/EY) \int_0^\infty h(y)(1 - F(y))dy \right]^{-1}. \tag{2.8}$$

2.1 THE OPTIMUM CHECKPOINT INTERVAL. Several authors (Young [15], Chandy [2, 3], Gelenbe [8, 9], and Kovalenko [10]) have examined the problem of the choice of the optimum checkpoint interval. Before briefly reviewing previous work let us introduce some definitions.

The *checkpoint interval* (CI) is the random variable Y defined in (i). It is the total time between two successive checkpoints during which the system is available for processing transactions which have not been processed before, and $F(y)$ is its distribution function.

The *total real time between checkpoints* (TRBC) is the total real time which elapses between two successive checkpoints; it is made up of the CI plus all the intervals of failure recovery. We shall denote it by the random variable ξ, and its distribution function by $\hat{F}(y)$.

Young [15] and Chandy [2, 3] assume that the failure rate is very small: $\gamma E\xi \ll 1$. Thus they compute the value of $E\xi$ which maximizes the availability. Chandy [2, 3] assumes also that ξ is constant (i.e. deterministic). Gelenbe [9] assumes that Y is exponentially distributed and computes the value of EY which maximizes the availability. Kovalenko [10] assumes that Y is constant.

Here we shall examine (see (2.6)) the choice of $F(y)$ which maximizes the availability Π_0: This is the general problem of the optimum checkpoint interval. Previous work has been carried out under restrictive assumptions because either γ has been assumed to be very small, or a special form for $\hat{F}(y)$ has been assumed, or both. We shall see that Π_0 is optimized by letting Y (and *not* ξ) be deterministic.

Let $\mathcal{H}(y) \triangleq \int_0^y h(x)dx$, and denote by $\mathcal{F}_a$ the set of all probability distribution functions of fixed and finite expectation $a \geq 0$. Let $E\mathcal{H}$ be the expectation of $\mathcal{H}(Y)$ if Y is distributed according to some $F \in \mathcal{F}_a$.

Remark 1. If $h(x)$ is such that $E\mathcal{H} \geq \mathcal{H}(a)$ for all $F \in \mathcal{F}_a$ then the element of $\mathcal{F}_a$ which maximizes Π_0 (for fixed $a \geq 0$) is

$$U_a(y) = \begin{cases} 1 & \text{if } y \geq a, \\ 0 & \text{if } y < a. \end{cases}$$

Note that $E\mathcal{H} \geq \mathcal{H}(a)$ if $\mathcal{H}(y)$ is convex (i.e. $h(y)$ is nondecreasing).

PROOF. Consider, for fixed a,

$$\Pi_0 = \left[1 + EC/a + (\gamma/a) \int_0^\infty dF_a(y)\mathcal{H}(y) \right]^{-1}$$
$$= [1 + EC/a + (\gamma/a)E\mathcal{H}]^{-1}.$$

Clearly $\int_0^\infty dU_a(y)\mathcal{H}(y) = \mathcal{H}(a)$, completing the proof.

So far we have treated the recovery time as an abstract quantity $h(y)$. As in previous work, we shall take

$$h(y) = \alpha y + \beta, \tag{2.9}$$

where $\alpha, \beta > 0$ are constants. This form is intuitive in the context of database operation: β is a fixed time associated with reloading the information stored at the checkpoint back into primary memory, and αy corresponds to the time necessary to reexecute transactions which were processes in time y. We will return to the choice of α later. Note now that

$$\mathcal{H}(y) = (\alpha/2)y^2 + \beta y \tag{2.10}$$

and

$$E\mathcal{H} = (\alpha/2)EY^2 + \beta EY \geq \mathcal{H}(EY) = (\alpha/2)(EY)^2 + \beta EY.$$

Suppose that the probability distribution function of the optimum checkpoint interval $F^*(y)$ is in $\mathcal{F}_a$ (i.e. $a = \int_0^\infty y dF^*(y)$) for some $a \geq 0$. Then by the above remark $F^*(y) = U_a(y)$, and

$$\Pi_0(F^*(y)) = [1 + EC/a + \gamma\beta + \alpha\gamma a/2]^{-1}. \tag{2.11}$$

Let us now examine the problem of the choice of the parameters α and β in (2.9).

Define $p^*(0, 0)$ as the probability that the system is idle, given that it is in state 0 (normal operation). Then, in a long interval of normal operation of length y the system will be busy during a time of average value $y(1 - p^*(0, 0))$. Suppose that transactions arrive to the system at a rate λ, and let μ be the service rate. Then during the interval of length y, the number of transactions processed will be on the average $\mu y(1 - p^*(0, 0))$. Let k be the proportion of transactions which have to be reprocessed after a failure; if each transaction reprocessing is as long as the transaction processing it is reasonable to suppose that the reprocessing time for a failure which occurs after y time units of normal operation will take, on the average, a time $\mu^{-1}\mu ky(1 - p^*(0, 0))$. Therefore we shall take

$$\alpha \approx k(1 - p^*(0, 0))$$

and

$$h(y) = ky(1 - p^*(0, 0)) + \beta. \tag{2.12}$$

In Section 3 we shall show rigorously that the stationary probability that the server is busy given that it is in state 0 is

$$p^*(0, 0) = 1 - \lambda/\mu\Pi_0.$$

Therefore we take

$$\alpha = k\lambda/\mu\Pi_0. \tag{2.13}$$

Combining (2.11) and (2.13) we have

$$\Pi_0[1 + (EC/a) + \gamma\beta] = 1 - k\gamma a\lambda/2\mu. \tag{2.14}$$

We are now ready to obtain the value $\hat{a}$ of the optimum checkpoint interval from (2.14). This is our main practical contribution.

THEOREM 1. *The value of a which maximizes* Π_0 *in* (2.14) *is*

$$\hat{a} = \frac{EC}{1 + \beta\gamma}\left[\left(1 + \frac{2(1 + \beta\gamma)}{\rho\gamma kEC}\right)^{\frac{1}{2}} - 1\right], \tag{2.15}$$

where $\rho = \lambda/\mu$.

The expression (2.15) is our new formula for the optimum checkpoint interval: Clearly it is dependent on ρ, the load factor of the transaction processing model. The value $\hat{a}$ is quite different from the formula $(2EC/\alpha\gamma)^{\frac{1}{2}}$ for the optimum checkpoint interval derived in [15] and [3] which is not applicable if (2.13) is used for the choice of α.

Consider the case when $p^*(0, 0) \approx 0$ (the system is heavily loaded). Then

$$2(1 + \beta\gamma)/\rho\gamma kEC \gg 1 \quad \text{and} \quad \hat{a} \approx \left(\frac{2EC}{\rho\gamma k(1 + \beta\gamma)}\right)^{\frac{1}{2}},$$

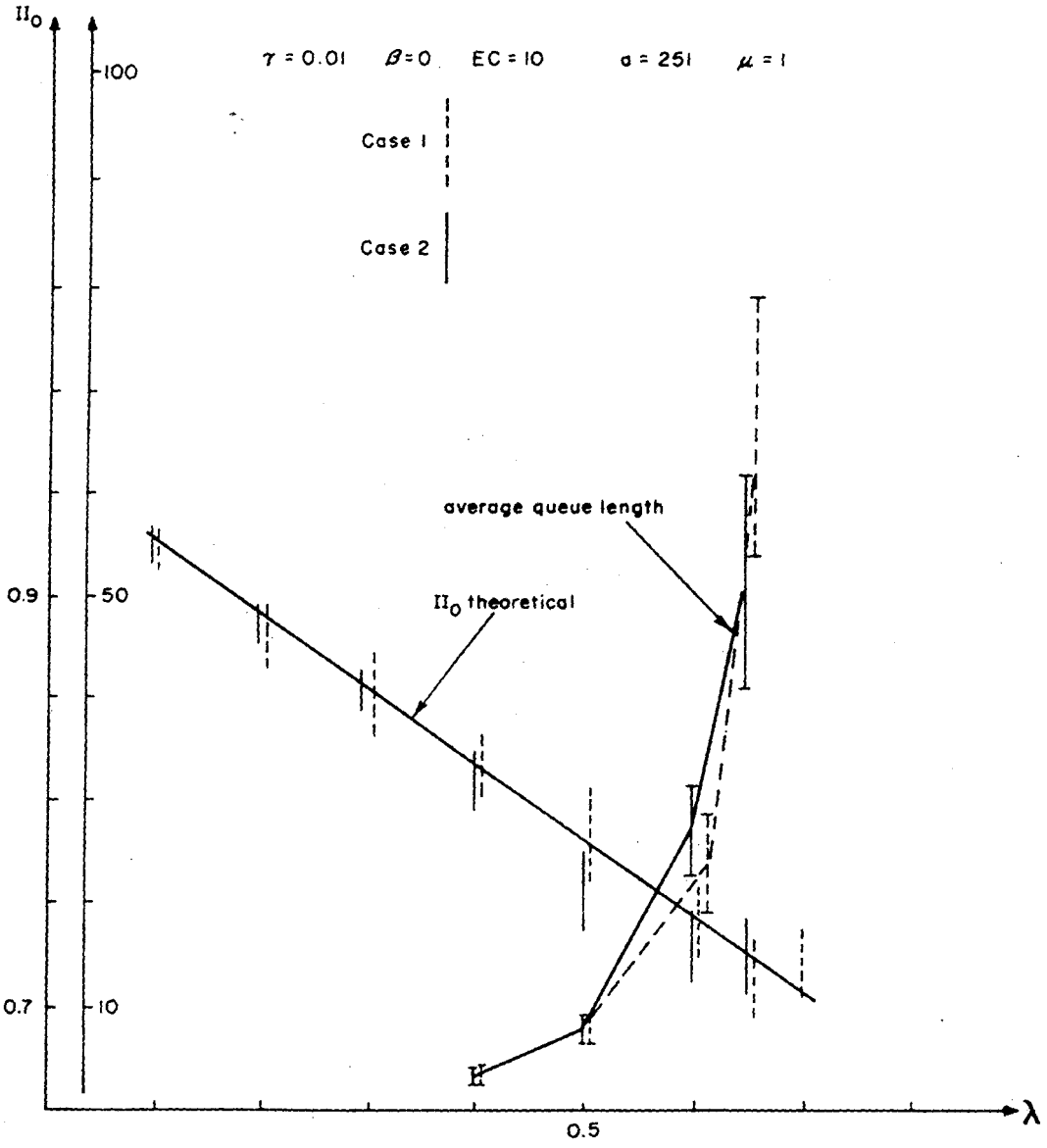

FIG. 1

which resembles somewhat the formula provided by Young [15] and Chandy [3]. Another case of interest arises when

$$2(1 + \beta\gamma) \ll \rho\gamma kEC.$$

We then have

$$\hat{a} \approx 1/\rho\gamma k.$$

The use of (2.12) for the recovery time and the use of (2.13) have been validated in a series of simulations reported in [7]. In Figure 1 we show a sample result. Two simulations are conducted. In the first case (case 1) the actual number of transactions processed (the contents of the audit trail) is recorded; when a failure occurs the same number of transactions is reprocessed. In case 2, $p^*(0, 0)$ is estimated by measuring the proportion of time up to a failure when the server was idle given that it was in state 0; this is then used in (2.12). The simulations show very good agreement between case 1 and case 2 for Π_0, and an acceptable agreement (the confidence intervals cover each other) for the average

queue length of transaction requests. In both cases we show the 95-percent confidence intervals. The theoretical result (using (2.14)) for Π_0 is also shown.

2.2 THE TRBC WITH THE OPTIMUM CI. It is of interest to obtain $\hat{F}(y)$, the distribution of the total real-time ξ between checkpoints (TRBC) when the optimum CI is used. From (2.4) we have

$$E\xi = (2EC/\alpha\gamma)^{\frac{1}{2}}(1 + \gamma\beta) + EC.$$

We will derive $\hat{f}^*(s) \triangleq \int_0^\infty e^{-sy} d\hat{F}(y)$. Note that for constant $\hat{a}$, and given n events in the interval $[0, \hat{a}]$ according to a Poisson process, the n instants of occurrence are independent and uniformly distributed in $[0, \hat{a}]$. Thus the n recovery times are independent; the density function of any one is

$$f_h(y) = \begin{cases} (\alpha\hat{a})^{-1} & \text{if } y \in [\beta, \beta + \hat{a}\alpha], \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if $f_h^*(s) \triangleq \int_0^\infty f_h(y)e^{-sy}dy$,

$$\begin{aligned} \hat{f}^*(s) &= [e^{-s\hat{a}}]\left[\sum_{n=0}^{\infty} \frac{(\gamma\hat{a})^n}{n!} e^{-\gamma\hat{a}}(f_h^*(s))^n\right] \\ &= \exp[-s\hat{a} + \gamma\hat{a}(f_h^*(s) - 1)], \end{aligned}$$

where

$$f_h^*(s) = (\hat{a}\alpha s)^{-1}e^{-\beta s}[1 - e^{-\alpha\hat{a}s}].$$

The variance of ξ is obtained as

$$E[(\xi - E\xi)^2] = \gamma\hat{a}[\beta^2 + \hat{a}\alpha\beta + (\alpha\hat{a})^2/3]. \tag{2.16}$$

We see that, even for small values of γ, the variance can be quite large at the optimum value of the CI.

3. *Analysis of the Queue*

In this section we analyze the queueing process representing the number of transactions awaiting processing and the state of the server. Our purpose is to determine measures of performance such as the *utilization* of the server (i.e. the steady state probability that the queue length is not zero), the *saturation condition* (i.e. the value of the arrival rate λ of transaction requests beyond which the stationary queue length will be infinite), and the distribution of the number of customers in queue. This queueing analysis is new. Of interest to us is the stochastic process $(N, X, Y) \triangleq \{N_t, X_t, Y_t, t \geq 0\}$ where N_t and X_t are the number of customers in queue and the state of the server, respectively, at time t; Y_t is defined in (2.2).

Intuitively speaking, the random variable Y_t allows us to compute the amount of time $h(Y_t)$ which is necessary for recovery from a failure which occurs at time t when the server is in state 0.

The analysis of the queueing system is undertaken under the following assumptions. Requests for transactions arrive according to a Poisson process of parameter λ, and are served in first-come-first-served order when the server is in state 0. No service is rendered in states 1 and 2 of the server. The service time distribution for a transaction is exponentially distributed of parameter μ.

Our study begins with the conditional process

$$(N, Y | X = 0) \triangleq \{N_t, Y_t | X_t = 0, t \geq 0\}.$$

3.1 THE CONDITIONAL PROCESS $(N, Y | X = 0)$. Let $p(n, y, t) = P[N_t = n, Y_t = y | X_t = 0]$ (see Footnote 1), and let us introduce the following notation:

[1] We mean that $p(n, y, t)dy = P[N_t = n, y \leq Y_t < y + dy | X_t = 0]$ at $t \geq 0$.

$$\eta(y) = f(y)/(1 - F(y)),$$
$$r_y(j) = \frac{(\lambda h(y))^j}{j!} \exp(-\lambda h(y)), \quad j \geq 0. \tag{3.1}$$

Clearly $\eta(y)dy$ is the probability that a checkpoint is established in the interval $[t, t + dy[$ given that $X_t = 0$ and $Y_t = y$. $r_y(j)$ is the probability that j arrivals occur during a recovery from a failure which took place at time t when $X_t = 0$ and $Y_t = y$.

The following equations are obtained in the usual manner. For $y \geq 0$, $n > 0$:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial y}\right)p(n, y, t) = -(\lambda + \mu + \eta(y) + \gamma)p(n, y, t) + \lambda p(n - 1, y, t) + \mu p(n + 1, y, t) + \gamma \sum_{j=0}^{n} r_y(j)p(n - j, y, t), \tag{3.2}$$

and for $y \geq 0$, $n = 0$:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial y}\right)p(0, y, t) = -(\lambda + \eta(y) + \gamma)p(0, y, t) + \mu p(1, y, t) + \gamma r_y(0)p(0, y, t). \tag{3.3}$$

For $y = 0$ we have for $n \geq 0$:

$$p(n, 0, t) = \int_0^\infty \sum_{j=0}^{n} c(j)p(n - j, y, t)\eta(y)dy, \tag{3.4}$$

where

$$c(j) = \int_0^\infty \frac{(\lambda z)^j}{j!} e^{-\lambda z} dC(z), \quad j \geq 0 \tag{3.5}$$

is the probability of j arrivals during the establishment of a checkpoint (see (ii) of Section 2).

In the sequel we study the stationary probability distribution for the process $(N, Y | X = 0)$.

THEOREM 2. *The stationary solution of* (3.2), (3.3), *and* (3.4) *exists if and only if* $(\lambda/\mu) < [1 + EC/EY + (\gamma/EY) \int_0^\infty h(y)(1 - F(y))dy]^{-1}$.

PROOF. "Only if" part (necessary condition). Set $(\partial/\partial t) = 0$ in (3.2) and (3.3). Let

$$p^*(n, s) = \int_0^\infty e^{-sy} p(n, y)dy,$$

where we have dropped the dependence on t. Denote

$$G_y(x) = \sum_0^\infty x^n p(n, y), \qquad G^*(s, x) = \int_0^\infty G_y(x)e^{-sy}dy.$$

We obtain for $n > 0$,

$$sp^*(n, s) - p(n, 0) = -(\lambda + \mu + \gamma)p^*(n, s) + \mu p^*(n + 1, s) + \lambda p^*(n - 1, s) - \int_0^\infty e^{-sy}\eta(y)p(n, y)dy + \sum_{j=0}^{n} \gamma \int_0^\infty r_y(j)p(n - j, y)e^{-sy}dy \tag{3.6}$$

and

$$sp^*(0, s) - p(0, 0) = -(\lambda + \gamma)p^*(0, s) + \mu p^*(1, s) - \int_0^\infty \eta(y)e^{-sy}p(0, y)dy + \gamma \int_0^\infty r_y(0)p(0, y)e^{-sy}dy,$$

yielding

$$G^*(s, x)[s + \lambda(1 - x) + \mu(1 - 1/x) + \gamma] - G_0(x)$$
$$= p^*(0, s)\mu(1 - 1/x) + \gamma H^*(s, x) - \int_0^\infty \eta(y)G_y(x)e^{-sy}dy, \quad (3.7)$$

where

$$H^*(s, x) = \int_0^\infty e^{-sy}e^{-\lambda(1-x)h(y)}G_y(x)dy.$$

We use (3.4) to obtain

$$G_0(x) = C^*(\lambda(1 - x)) \int_0^\infty \eta(y)G_y(x)dy, \quad (3.8)$$

where

$$C^*(\lambda(1 - x)) = \int_0^\infty e^{-\lambda(1-x)z}dC(z).$$

From (3.7) we remain with

$$G^*(s, x) = \frac{p^*(0, s)\mu(1 - 1/x) - \int_0^\infty \eta(y)G_y(x)e^{-sy}dy + G_0(x)}{s + \lambda(1 - x) + \mu(1 - 1/x)}$$
$$+ \frac{\gamma[H^*(s, x) - G^*(s, x)]}{s + \lambda(1 - x) + \mu(1 - 1/x)}, \quad (3.9)$$

where $G_0(x)$ is given by (3.8). It is clear that if the stationary distribution $\{p(n, y), n \geq 0, y \geq 0\}$ exists we must have

$$\lim_{x\to 1} \lim_{s\to 0} G^*(s, x) = 1.$$

Taking the limits in (3.9), we have an indeterminate form in both terms on the right-hand side. Therefore take first

$$\lim_{s\to 0} G^*(s, x) = \frac{p^*(0, 0)\mu(1 - 1/x) + [C^*(\lambda(x - 1)) - 1] \int_0^\infty \eta(y)G_y(x)dy}{\lambda(1 - x) + \mu(1 - 1/x)}$$
$$+ \frac{\gamma[H^*(0, x) - G^*(0, x)]}{\lambda(1 - x) + \mu(1 - 1/x)}; \quad (3.10)$$

now apply l'Hôpital's rule using

$$\lim_{x\to 1} \frac{d}{dx} C^*(\lambda(1 - x)) = \lambda EC,$$

$$\lim_{x\to 1} C^*(\lambda(1 - x)) = 1,$$

$$\lim_{x\to 1} \frac{d}{dx} H^*(0, x) = \lim_{x\to 1} \frac{d}{dx} G^*(0, x) + \int_0^\infty \lambda h(y)G_y(1)dy,$$

so that

$$\lim_{x\to 1}\lim_{s\to 0} G^*(s, x) = \frac{\mu p^*(0, 0) + \lambda EC\int_0^\infty \eta(y)G_y(1)dy + \lambda\gamma\int_0^\infty h(y)G_y(1)dy}{-\lambda + \mu}. \quad (3.11)$$

To complete the proof, we shall call upon the following lemma.

LEMMA 1. *$G_y(1) = \lim_{t\to\infty}\sum_{n=0}^\infty p(n, y, t)$ is given by*

$$G_y(1) = (1 - F(y))/EY.$$

PROOF. This is again a consequence of the well-known result [14, pp. 10, 11] concerning renewal processes. Write

$$K(y, t) = \int_0^y \sum_{n=0}^\infty p(n, z, t)dz,$$

so that

$$\lim_{t\to\infty} K(y, t) = \int_0^y G_z(1)dz.$$

Then

$$K(y, t) = \sum_{k=1}^\infty P[t - y < \sigma_k \le t < \sigma_{k+1}], \quad (3.12)$$

where (since we are dealing with the conditional process $(N, Y|X = 0)$) any instant of time t is relative to epochs during which the state of the server is 0; that is, t corresponds to the "real" time t' (in process time of (N, Y, X)) where

$$t = \int_0^{t'} i_0(\tau)d\tau.$$

In (3.12) $\sigma_1 < \cdots < \sigma_k < \sigma_{k+1} < \cdots$ are defined by $Y_{\sigma_i} = 0$. Clearly the $(\sigma_{k+1} - \sigma_k)$ are independent and identically distributed of probability distribution function $F(y)$. Thus

$$\lim_{t\to\infty} K(y, t) = \int_0^y [1 - F(x)]dx/EY, \quad (3.13)$$

and the lemma is established.

Now returning to (3.11) and the proof of Theorem 2, we see that

$$\int_0^\infty \eta(y)G_y(1)dy = \int_0^\infty [dF(y)/(1 - F(y))][1 - F(y)]/EY = 1/EY.$$

Therefore

$$p^*(0, 0) = 1 - \left(\frac{\lambda}{\mu}\right)\left[1 + EC/EY + \gamma\int_0^\infty dyh(y)(1 - F(y))/EY\right] = 1 - \frac{\lambda}{\mu\Pi_0} \quad (3.14)$$

if the stationary solution to (3.2), (3.3), and (3.4) exists. But we must have $p^*(0, 0) = \int_0^\infty p(0, y)dy > 0$, completing the first part of the proof.

"If" part (sufficient condition). To prove that the condition is sufficient consider the instants σ_i^+ just *after* the instants σ_i when $Y_{\sigma_i} = 0$ in the process time of $(N, Y|X = 0)$. The process

$$(\hat{N}|X = 0) \equiv \left\{N_{\sigma_i^+}, i \ge 1 | X_t = 0, t \ge 0\right\}$$

is a discrete time Markov chain, and it is easy to see that it is aperiodic and irreducible. We first prove that it is ergodic under the condition of Theorem 2 by applying a lemma of Pakes [12], which we recall below.

LEMMA. *Let $\{B_i, i \geq 1\}$ be an aperiodic irreducible Markov chain. It is ergodic if the following conditions are satisfied:*

$$|E[B_{i+1} - B_i | B_i = j]| < \infty \text{ for all } j, \limsup_{j\to\infty} E[B_{i+1} - B_i | B_i = j] < 0.$$

Let $\alpha = \sigma_{i+1} - \sigma_i$; in the sequel all events are conditioned on $N_{\sigma_i^+} = j$. Let T be the total time in the interval $[\sigma_i^+, \sigma_{i+1}^+[$ during which the queue is nonempty. Clearly $T \leq \alpha$. Let $D(\alpha)$ be the number of departures from the queue in the interval $[\sigma_i^+, \sigma_{i+1}^+[$.

$$1 \geq P[T = \alpha] \geq P[D(\alpha) < j].$$

But

$$P[D(\alpha) < j] \geq \int_0^\infty dF(\tau) \sum_0^{j-1} \frac{(\mu\tau)^l}{l!} e^{-\mu\tau},$$

since the event $[D(\alpha) < j]$ implies the event [(the number of departures in a busy period of length α) $< j$]. Therefore

$$\lim_{j\to\infty} P[T = \alpha] = 1. \tag{3.15}$$

Now consider

$$E\left[N_{\sigma_{i+1}^+} - N_{\sigma_i^+} | N_{\sigma_i^+} = j\right] = E[A(\alpha) - D(\alpha)],$$

where $A(\alpha)$ is the number of arrivals in $[\sigma_i^+, \sigma_{i+1}^+[$. Clearly

$$E[A(\alpha)] = \lambda\left[EY + EC + \gamma\int_0^\infty h(y)(1 - F(y))dy\right].$$

Also $E[D(\alpha)] = \mu E[T]$; however,

$$\lim_{j\to\infty} E[T] = EY.$$

Therefore, if the condition of Theorem 2 is satisfied then both conditions of Pakes' lemma are satisfied by $(\tilde{N} | X = 0)$ so that this Markov chain is ergodic. It is easily seen that $p(n, y)$ also exists.

As an immediate consequence we have the following theorem.

THEOREM 3. *Under the conditions of Theorem 2, the stationary probability distribution for the process $(N | X = 0)$ exists and its generating function $G(x) \triangleq \lim_{s\to 0} G^*(s, x)$ satisfies*

$$G(x) = \frac{\gamma H^*(0, x) + p^*(0, 0)\mu(1 - 1/x) + (C^*(\lambda(x-1)) - 1)\int_0^\infty \eta(y)G_y(x)dy}{\gamma + \lambda(1 - x) + \mu(1 - 1/x)}.$$

3.2. THE PROCESS (N). Let the (unconditional) queue length process be denoted by $(N) \triangleq \{N_t, t \geq 0\}$.

In this section we extend the results of Sections 2 and 3.1 to (N). The remaining work is largely technical.

Let us introduce the following notation for $t \geq 0$:

$$a(n, u, t) \triangleq P[N_t = n, U_t = u, X_t = 2],$$

where for $X_t = 2$:

$$t - U_t \triangleq \sup\{\tau = \tau < t \wedge X_\tau = 0\},$$

$$b(n, y, v, t) = \begin{cases} 0 & \text{if } v > h(y), \\ P[N_t = n,\ Y_t = y,\ V_t = v,\ X_t = 1] & \text{otherwise,} \end{cases}$$

where for $X_t = 1$:

$$t - V_t \triangleq \sup\{\tau : \tau < t \wedge X_\tau = 0\}.$$

Let $a(n, u)$, $b(n, y, v)$ be the corresponding quantities obtained by taking $\lim t \to \infty$, if they exist.

Remark 1. $a(n, u) = \Pi_0.(1 - C(u)) \sum_{j=0}^{n} ((\lambda u)^j/j!)\, e^{-\lambda u} \int_0^\infty p(n - j, y).\eta(y)dy.$

Remark 2. $b(n, y, v) = \Pi_0 \gamma \sum_{j=0}^{n} ((\lambda v)^j/j!)\, e^{-\lambda v} p(n - j, y).$

THEOREM 4. *$a(n, u)$, $b(n, y, v)$ exist for all $n \geq 0$, $u \geq 0$, and $y \geq 0$ such that $0 \leq v \leq h(y)$, if and only if $p(n, y)$ exists for all $n \geq 0$, $y \geq 0$.*

4. *Conclusions*

In this paper we give a detailed theoretical treatment of the problem of determining the optimum checkpoint interval, i.e. the total time between successive checkpoints during which the database system is *not* recovering from failures, which maximizes system availability.

We first obtain, under some general assumptions, the expression for the availability. We then show that the optimum checkpoint interval must be *deterministic* and that *it is a function of the system load.* An explicit expression for its value is given. These are the main practical contributions of the paper.

These results of practical interest are obtained by means of a theoretical analysis of a queueing system representing system behavior. This queue has failures and repair times which are a function of the age of the failure with respect to the most recent maintenance (checkpoint) epoch. The queueing model appears to be novel and may have applications to other reliability studies.

REFERENCES

(Note. References [1, 5, 11] are not cited in the text.)

1. BOROVKOV, A.A. *Stochastic Processes in Queueing Theory.* Springer-Verlag, New York, 1976.
2. CHANDY, K.M. A survey of analytic models of roll-back and recovery strategies. *Computer 8*, 5 (May 1975), 40–47.
3. CHANDY, K.M., BROWNE, J.C., DISSLY, C.W., AND UHRIG, W.R. Analytic models for rollback and recovery strategies in data base systems. *IEEE Trans. Software Eng. SE-1*, 1 (March 1975), 100–110.
4. ÇINLAR, E. *Introduction to Stochastic Processes.* Prentice-Hall, Englewood Cliffs, N.J., 1975.
5. COX, D.R. *Renewal Theory.* Methuen, London, 1962.
6. COX, D.R., AND LEWIS, P.A.W. *The Statistical Analysis of Series of Events.* Methuen, London, 1966.
7. FLAMAND, J., AND GELENBE, E. Simulation of roll-back recovery in a data-base system. To appear.
8. GELENBE, E. A model of roll-back recovery with multiple checkpoints. Proc. 2nd Int. Symp. on Software Eng., Oct. 1976, pp. 251–255 (available from ACM, New York).
9. GELENBE, E., AND DEROCHETTE, D. On the stochastic behaviour of a computer system under intermittent failures. In *Modelling and Performance Evaluation of Computer Systems,* H. Beilner and E. Gelenbe, Eds., North-Holland Pub. Co., Amsterdam, 1976.
10. KOVALENKO, I.N., AND STOIKOVA, L.S. On the productivity of a system and the problem solving time in the presence of random failures and aperiodic memorization of the results. Transl. from *Kibernetika*, issue no. 5 (Sept.–Oct. 1974), 73–75. (*Cybernetics 10*, 5 (Feb. 1976), 820–823, Plenum Pub. Corp., New York).
11. LOYNES, R.M. The stability of a queue with non-independent interarrival and service times. *Proc. Cambridge Philos. Soc. 58* (1962), 497–520.
12. PAKES, A.G. Some conditions for ergodicity and recurrence of Markov chains. *Oper. Res. 17* (1969), 1058–1061.
13. ROBIN, M. Quelques problèmes de contrôle stochastique. Th., U. Paris-Dauphine, 1978.
14. TAKACS, L. *Introduction to the Theory of Queues.* Oxford U. Press, New York, 1962.
15. YOUNG, J.W. A first-order approximation to the optimum checkpoint interval. *Comm. ACM 17*, 9 (Sept. 1974), 530–531.

RECEIVED JANUARY 1977; REVISED JANUARY 1978