A Variational Calculus Approach to Optimal Checkpoint Placement

Yibei Ling, *Member*, *IEEE*, Jie Mi, and Xiaola Lin

Abstract—Checkpointing is an effective fault-tolerant technique for improving system availability and reliability. However, a blind checkpointing placement can result in either performance degradation or expensive recovery cost. By means of the calculus of variations, we derive an explicit formula that links the optimal checkpointing frequency with a general failure rate, with the objective of globally minimizing the total expected cost of checkpointing and recovery. Theoretical result shows that the optimal checkpointing frequency is proportional to the square root of the failure rate and can be uniquely determined by the failure rate (time-varying or constant) if the recovery function is strictly increasing and the failure rate is $\lambda(\infty) > 0$. Bruno and Coffman [2] suggest that optimal checkpointing by its nature is a function of system failure rate, i.e., the time-varying failure rate demands time-varying checkpointing in order to meet the criteria of certain optimality. The results obtained in this paper agree with their viewpoint.

Index Terms—Aperiodic checkpointing, periodic checkpointing, system failure rate.

1 Introduction

Computer and database information systems are vulnerable to system failures. The presence of software bugs and hardware failures makes the computer inherently unreliable. The rollback-recovery technique is a common means of increasing the system reliability against various types of failure. Checkpointing in rollback/recovery schema is an operation that stores the correct state of a process from time to time in a stable storage such that the process can resume its normal computation from the checkpointed state on recovery, avoiding expensive recomputation from scratch in case of a system failure. In addition to fault-tolerant applications, checkpointing is also used as a means of process migration or coarse-grained job-swapping [18] and as a means of reducing the overall expected time of completing a job [7]. The benefit of checkpointing, however, comes at a price; as a result, excessive checkpointing would result in performance degradation, while deficient checkpointing would incur an expensive recovery overhead. Therefore, a trade-off must be made.

A wealth of references in the literature is available to cover a wide range of issues related to checkpointing and recovery. The papers by Chandy [5] and Nicola [17] can serve as an excellent overview of checkpointing and recovery strategies in the literature. The primary objective of using the checkpointing technique is to increase the computational efficiency and enhance the reliability in a faulty environment. The key problem is how to determine strategies for checkpoint placement to meet the system performance objective optimally. Various mathematical models of checkpoint scheduling, with different objectives, are formulated to address problems from different angles. Guided by the principle of simplicity, the research direction primarily focuses on efforts to relax restrictions imposed by the assumptions in order to increase the scope of practical application. In practice, the simplest one among the class of optimal checkpointing strategies is most preferred by system designers and administrators [17].

In this section, we will give a brief overview of existing literature, with emphasis on our closely related work. Many relevant references in the literature can be found in the references section of this paper. In [8], a model with Poisson failure is considered to determine the optimal number of checkpoints which minimizes the overall expected execution time of a program, assuming that the execution time of each program part with equal size is an independently and identically distributed (i.i.d) random variable. Grassi et al. [12] improve the same model and suggest that the optimal number of checkpoints rely on the distribution of the program execution time. In [4], a model is given to optimally schedule checkpoint placement to maximize the system availability, assuming that the recovery cost is proportional to the time interval between the present failure to the most recent checkpoint.

Tantawi and Ruschitzka [23] consider a model with general failure distribution, allowing failures to occur during checkpointing and recovery. As a result of the generality, the evaluation of the model suggests the need for computing an infinite number of embedded integrals, and is basically mathematically intractable. A simplified version, constrained by the restriction that prevents failures from happening during checkpointing and recovery, is proposed. However, its evaluation is still out of reach since the evaluation needs to compute an infinite set of nonlinear equations. To make the problem tractable, Tantawi and Ruschitzka make a useful simplification by imposing additional restrictions. The equidistant strategy assumes

- *Y. Ling is with Applied Research Laboratories, Telcordia Technologies, 445 South St., Morristown, NJ 07960-6438. E-mail: lingy@research.telcordia.com.*
- *J. Mi is with the Department of Statistics, Florida International University, Miami, FL 33199. E-mail: mi@fiu.edu.*
- *X. Lin is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong. E-mail: xlin@eee.hku.hk.*

Manuscript received 13 June 2000; accepted 23 Jan. 2001.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 109270.

that the production time between two successive checkpoints is constant and the equicost strategy assumes that the expected reprocessing time equals the mean checkpointing time. Under these conditions, a computational approach (iteration algorithm) is derived to compute an approximation of the best equicost strategy. L'Ecuyer and Malenfant [9] modify Tantawi and Ruschitzka's model by considering the failure rate with a cyclical fashion and place the focus on the computational approach. Dynamic programming is introduced to improve the computational efficiency. In [24], a stochastic dynamic programming approach is used to determine the optimal checkpoint schedule between tasks of given lengths in a program and the objective is to minimize the total expected program execution time.

For age-dependent failure, Leung and Choo [14] propose a computational procedure for determining the optimal checkpoint placement. Sumita et al. [22] give an analysis of effective service time under age-dependent interruption. Coffman and Gilbert [6] formulate models for minimizing the expected execution time or maximizing the probability of completing a job. They make a useful simplification by considering the checkpointing process as a renewal process, assuming that the system is "as-good-as-new" after each checkpoint. Gelenbe and Hernandez [10] consider a model with the age-dependent Poisson process, assuming that each failure is a renewal point. Any deviation from the assumption about the age-dependent Poisson distribution might compromise the validity of the model in some way [10]. For illustration purposes, the formula for the sensitivity evaluation is derived and the error sensitivity to a Weibull distribution is calculated [10]. Boguslavsky et al. [1] consider the optimization problems associated with the save-and-check scheduling for the system whose failure is not self-evident. The save schedule is for error recovery, while the check schedule is for failure detection. Both schedules are working in tandem according to the different coupling strategies. General save-and-check scheduling strategies with an exponential failure law are discussed and analyzed and the example with the numerical computation of the optimal checkpoint placement in the program is given.

Checkpointing is extended to increase computational fault-tolerance in a multiprocessor environment [2], [3]. Bruno and Coffman [2] derive a stochastic model for determining the number of checkpoints and with the objective to maximize the probability of completing a job before all m processors fail. They give an elegant asymptotic estimate of k (the number of checkpoints made within the job of the given size) for $m = 2$ processors and an exponential failure law for each processor. A computational approach with $m \geq 3$ processors is given and the property of the optimal completion probability is presented [2]. Bruno et al. [3] propose a model for improving computational reliability by using parallel processing and study the policy of processor shadowing so as to maximize the probability of completing a job. The mathematical relation among the optimal time to commence shadowing, the duration of the job, and the number of processors is obtained. In addition to fault-tolerant application, Coffman et al. [7] formulate a stochastic model for determining the optimal checkpoint schedule that yields the minimum expected cost of serving a sequence of requests for an abstract moving-server system.

The equally spaced checkpointing is prevalent due to its relatively simpler mathematical treatment. Practically, the checkpoints are very likely to be placed unevenly [1], [7], [17], [23]. For example, for software debugging, checkpoints (breakpoints) are placed unevenly, depending to a large extent on the structure of the program. Boguslavsky et al. [1] suggest that, for certain criteria of optimality and assumptions about the system, the optimal solution may require an aperiodic checkpointing. The example is given in [23] to show that, for a Weibull failure distribution, the equicost checkpointing achieves higher system availability than that for an equidistant strategy. In general, the best equidistant strategy is not optimal [9], [10], [23]. It is suggested in [17] that checkpointing should be performed more frequently toward the end of the program. Bruno and Coffman [2] state that "the checkpointing strategies depend on the distribution of the time-to-failure random variable; however, in many cases, including the exponential failure law, optimal checkpointing is done at interval which are, for the most part, uniformly spaced."

The dependence of the optimal checkpointing upon the failure rate suggested by [2], [4], [8], [9], [10], [17], [23] implies that the optimal checkpointing frequency by its nature is a function of the failure rate, i.e., a time-varying failure rate demands the time-varying optimal checkpointing in order to meet the criteria of certain optimality.

In this paper, using the calculus of variations, we formulate the optimal checkpoint scheduling to globally minimize the overall expected cost and derive a closed form equation that establishes the connection between the optimal checkpointing frequency and a general failure rate. The essential features of our approach are:

1. No particular assumption about the system failure distribution is made except for a technical one $\lambda(\infty) > 0$;
2. No assumption about checkpoint schedule (equally spaced or unequally spaced) is made a priori;
3. The introduction of the continuous checkpointing frequency function forms the basis of our approach, leading to the problem solving by using a new mathematical approach: the calculus of variations (remark: the sequence of discrete checkpoints can be uniquely determined when the checkpointing frequency is given);
4. The global minimum total expected cost is guaranteed. We demonstrate that the optimal checkpointing frequency is proportional to the square root of the system failure rate (time-varying or constant) and that the optimal checkpointing becomes equally spaced under Poisson failure distribution (constant failure rate), which agrees with the results reported in [2], [4], [5], [7], [8], [9], [10].
5. The computational efficiency is ensured due to the closed-form expression for the optimal checkpointing frequency.

The remainder of the paper is organized as follows: Section 2 delves in detail into the mathematical formalism

TABLE 1
Frequently Used Symbols

SYMBOL	MEANING
$n(t)$	Checkpoint frequency function
c_0	Cost for a checkpoint setup
c_1	Recovery Proportionality Constant
c_2	Constant associated with a recovery setup
$L(\cdot)$	Recovery cost function

of our approach; Section 3 gives an example to substantiate theoretical results obtained in Section 2, the numerical gain of the optimal aperiodical checkpointing over the optimal periodic checkpointing is calculated, and the optimal checkpoint placement for failure rate with a cyclical nature is given. Finally, Section 4 concludes the paper and outlines future research.

2 Description of Mathematical Model

In this section, we first present a set of definitions and assumptions relevant to our derivation of the mathematical model. Since some symbols will be used frequently throughout this paper, we summarize them in Table 1 for brevity.

Let's begin with an introduction of the checkpointing frequency function, followed by a discussion of its mathematical constraints and physical implications, as well as the relationship between a continuous checkpointing frequency function and discrete checkpoint instants.

Two natural constraints on the checkpointing frequency function $n(t)$ are introduced in order to exclude unrealistic cases from consideration:

1. The definite integral of $n(t)$ on any finite time interval is finite, that is, $\int_a^b n(\tau)d\tau < \infty$ for any $0 \le a \le b < \infty$;
2. When $t \to \infty$, the limit of $n(t)$ exists and is positive; that is, $n(\infty) = \lim_{t\to\infty} n(t) > 0$. Here, $n(\infty) = \infty$ is permitted.

The family of functions satisfying the above natural constraints is denoted as $\aleph$ throughout this paper. The constraints have a sound physical basis: The first constraint means that a finite number of checkpoints are performed in a finite time interval. The second constraint means that checkpointing is still required against unanticipated system failures when the system has attained its steady state (the time is sufficiently large). The constraints make the system physically realizable.

Given the continuous checkpoint frequency function $n(t)$, the sequence of discrete checkpoint time instants, $0 = t_0 < t_1 < \cdots < t_n$, can be established via the following equation:

$$\int_{t_{i-1}}^{t_i} n(\tau)d\tau = 1, \quad \forall i \ge 1. \tag{1}$$

The following definitions, in conjunction with some examples, are given to distinguish two different types of checkpointing:

Definition 1. *A checkpoint schedule is said to be periodic if* $t_i - t_{i-1} = t_j - t_{j-1}$ $(i \ne j, \forall i, j \ge 1)$.

Definition 2. *A checkpoint schedule is said to be aperiodic if there exists at least one pair of indices i and j* $(i \ne j)$ *such that* $t_i - t_{i-1} \ne t_j - t_{j-1}$.

Example 1. Assume that $n(t) = 1$, then, using (1), we obtain $t_i = i$, $i \ge 0$.

Example 2. Assume that $n(t) = t$, then, using (1), we obtain $t_i = \sqrt{2 \cdot i}$, $i \ge 0$.

It is obvious that Example 1 is periodic checkpointing (equally spaced checkpoints) and that Example 2 is aperiodic checkpointing (unequally spaced checkpoints).

To derive the main theorem, we make the following assumptions, which are widely used in the literature.

1. A system failure is self-evident, and can be instantly detected [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [16], [17], [20], [21], [22], [23], [24], [25].
2. The elapsed times for both checkpoint and recovery are relatively negligible as compared with the average failure time [1], [3], [4], [5], [6], [8], [10], [11], [14], [16], [17], [20], [22].
3. Recovery cost is proportional to the time interval between the last checkpoint and the present failure [2], [3], [7], [9], [10], [12], [13], [14], [17], [24].
4. No failure occurs during checkpointing and recovery processes [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [14], [16], [17], [20], [22], [23].

Remark. Assumption 2 becomes a reality thanks to the rapid advancement of technology. For example, the diskless checkpointing technique [18] removes the stable storage from checkpointing, thereby eliminating the main source of overhead in checkpointing. In addition, the wide adoption of (QMO) quality method operation in system design and implementation (hardware/software) makes the system reliable. All these factors combined make Assumption 2 realistic. A close examination reveals that Assumptions 2 and 4 are somewhat correlated since if Assumption 2 holds, then the probability of a failure occurring during checkpointing and recovery is very small and thus can be reasonably ignored for mathematical tractability.

In Fig. 1, $T_{recovery}$ is the recovery time, and $T_{checkpoint}$ is the elapsed time spent on a checkpoint setup, and Y_i is the elapsed time between two consecutive failures.

Under these assumptions, we consider the sequence of $\{Y_i, i \ge 1\}$ as a renewal process [1], [6], [9], [10], [11], [14], [15], [16], [18], [21]. Because of $\frac{T_{recovery}}{Y_i} << 1$, by Assumption 2, the ith cycle denotes the time interval $[\sum_0^{i-1} Y_k, \sum_0^i Y_k]$ with $Y_0 \equiv 0$ and the length of the ith cycle equals the ith system lifetime $Y_i, (i \ge 1)$. Our primary goal is to determine the

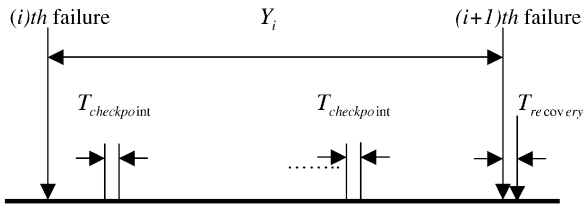

Fig. 1. Diagram for timing of failure, recovery, and checkpoint.

optimal sequence of checkpoint instants by minimizing the total expected cost associated with the renewal process $\{Y_i, i \geq 1\}$. Notice that the variable t used in the checkpoint frequency function $n(t)$ represents the elapsed time measured from the beginning of each cycle.

The recovery time can be expressed as $T_{recovery} \propto y$, based on Assumption 3, where y is illustrated in Fig. 2. With these assumptions in mind, we can claim that, the recovery time, $T_{recovery}$, can be approximately expressed as a function of the reciprocal of $n(Y_i)$. The justification is given as follows:

Suppose that $0 \leq t_1 < t_2 < \cdots$ are checkpoint time instants starting from the beginning of the ith cycle and a system failure will occur in the time interval $[t_k, t_{k+1}]$, that is, $t_k < Y_i \leq t_{k+1}$ for a $k \geq 0$. Then, $y = Y_i - t_k$ is the elapsed time starting from the last checkpoint t_k and has conditional probability distribution function:

$$P(Y_i - t_k < y \mid t_k < Y_i \leq t_{k+1}) = \frac{P(Y_i \leq y + t_k, t_k < Y_i \leq t_{k+1})}{P(t_k < Y_i \leq t_{k+1})}$$
$$= \begin{cases} 0, & y \leq 0 \\ \frac{P(t_k < Y_i \leq y + t_k)}{P(t_k < Y_i \leq t_{k+1})}, & 0 < y \leq t_{k+1} - t_k, \\ 1, & t_{k+1} - t_k < y. \end{cases}$$

That is, for $0 < y \leq t_{k+1} - t_k$,

$$P(Y_i - t_k > y \mid t_k < Y_i \leq t_{k+1}) = \frac{F(y + t_k) - F(t_k)}{F(t_{k+1}) - F(t_k)} = \frac{\overline{F}(y + t_k) - \overline{F}(t_{k+1})}{\overline{F}(t_k) - \overline{F}(t_{k+1})},$$

where $F(x)$ and $\overline{F}(x) = 1 - F(x)$ are the distribution and survival functions of Y_i, respectively. The first order of Taylor expansion could be expressed as:

$$\overline{F}(y + t_k) - \overline{F}(t_{k+1}) \approx f(t_{k+1})(t_{k+1} - t_k - y)$$

and

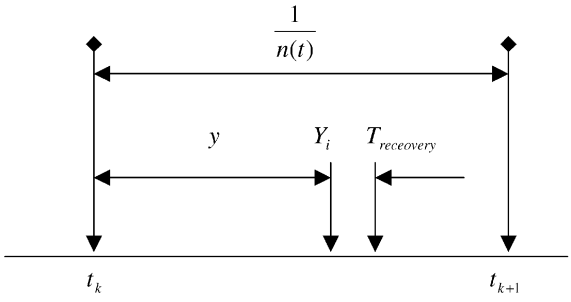

Fig. 2. Timing diagram for checkpointing.

$$\overline{F}(t_k) - \overline{F}(t_{k+1}) \approx f(t_{k+1})(t_{k+1} - t_k),$$

where $f(y)$ is the probability density function of y. The expected time from the last checkpoint to the end of the ith cycle is given as:

$$E(Y_i - t_k \mid t_k < Y_i < t_{k+1}) = \int_0^{t_{k+1} - t_k} P(Y_i - t_k > y \mid t_k < Y_i < t_{k+1}) dy \approx \int_o^{t_{k+1} - t_k} \frac{f(t_{k+1})(t_{k+1} - t_k - y)}{f(t_{k+1})(t_{k+1} - t_k)} dy = \frac{1}{2}(t_{k+1} - t_k). \quad (2)$$

On the other hand, by the Mean-Value Theorem, from (1), we have:

$$\int_{t_k}^{t_{k+1}} n(\tau) d\tau = 1 \approx (t_{k+1} - t_k) \cdot n(Y_i),$$

where $t_k \leq Y_i < t_{k+1}$. Namely,

$$t_{k+1} - t_k \approx 1/n(Y_i). \quad (3)$$

Combining (3) and (2) we have:

$$E(Y_i - t_k \mid t_k < Y_i \leq t_{k+1}) \approx \frac{1}{2 \cdot n(Y_i)},$$

and, thus, the expected recovery cost during a cycle $\propto \frac{1}{n(Y_i)}$, which can be approximately expressed by a linear function of $\frac{1}{n(Y_i)}$. If we use the second order of Taylor expansion

$$\overline{F}(y + t_k) - \overline{F}(t_{k+1}) \approx f(t_{k+1})(t_{k+1} - t_k - y) + f'(t_{k+1}) \frac{(t_{k+1} - t_k - y)^2}{2},$$

where $0 < y < t_{k+1} - t_k$. In a similar approach, it can be seen that the mean elapsed time from the last checkpoint to the end of the ith cycle can be approximately expressed as:

$$\frac{1}{2 \cdot n(Y_i)} + \frac{f'(t_{k+1})}{6 \cdot f(t_{k+1})} \cdot \frac{1}{n^2(Y_i)}.$$

From the above, we can see that mean recovery cost associated with the ith cycle can be approximately expressed by $L(1/n(Y_i)), i \leq 1$, where $L(z)$ is a recovery function defined on $(0, \infty)$.

We now proceed to derive the main result. The total expected cost within a cycle consists of 1) the cost of checkpoint setup accumulated during the cycle, 2) a recovery cost. Hence, the random cost associated with the first cycle is given as:

$$R_1 = c_0 \int_0^{Y_1} n(\tau) d\tau + L\left(\frac{1}{n(Y_1)}\right). \quad (4)$$

Similarly, the random cost associated with the ith cycle is given by:

$$R_i = c_0 \int_0^{Y_1} n(\tau) d\tau + L\left(\frac{1}{n(Y_i)}\right).$$

Obviously R_i, $i \geq 1$, are independently and identically distributed (i.i.d) since $Y_i, i \geq 1$ are (i.i.d). For any given $t \geq 0$, let $N(t)$ be the number of system failures in the

interval [0, t]. If we consider the mean cost in this interval, that is,

$$\frac{E\left(\sum_{i=1}^{N(t)} R_i\right)}{t},$$

then, by the renewal reward process [18], we have:

$$\lim_{t\to\infty} \frac{E\left(\sum_{i=1}^{N(t)} R_i\right)}{t} = \frac{E(R_1)}{E(Y)}. \tag{5}$$

Equation (5) means that a cycle is completed every time the system is recovered from a failure and that the long-run expected cost is just the expected cost incurred during a cycle divided by the expected time of a cycle (the expected interfailure time). It is easy to see that, to minimize the expected long-run average cost, it suffices to minimize $E(R_1)$ since the expected $E(Y)$ is fixed. Let the probability density function for a system failure time be $f(t)$, then the probability for system failure within $[t, t+dt]$ is $f(t)\cdot dt$. By conditioning on the system failure time, the expected cost $E(R_1)$ during a cycle is written by the integral:

$$C(n(\cdot)) \equiv E(R_1) = \int_0^\infty \left[\int_0^t c_0 \cdot n(\tau)d\tau + L(1/n(t))\right] \cdot f(t)dt, \tag{6}$$

where $C(n(\cdot))$ is a functional defined on the family $\aleph$. The following theorem illustrates how to determine an optimal function $n^*(t)$ that minimizes $C(n(\cdot))$ in the family of functions $\aleph$, i.e.,

$$C(n*(\cdot)) \le C(n(\cdot)) \qquad \forall n(\cdot) \in \aleph.$$

It is well-known that system failure rate is defined as $\lambda(t) = \frac{f(t)}{1-F(t)}$ [19], where $F(t)$ is the probability distribution of system failure time.

Definition 3. *A system is said to be physically realizable if and only if its failure rate satisfies $\lambda(\infty) > 0$. A system is said to be physically unrealizable if and only if $\lambda(\infty) = 0$.*

Remark. Such a system being free of failure when time is sufficiently large contradicts our observation.

Example 3. A system having Poisson failure distribution (exponential failure law) is physically realizable.

Example 4. A system having Weibull distribution with shape parameter $\alpha = 0.5$ implies that it is eventually free of failure ($\lim_{t\to\infty}\lambda(t) = 0$). By Definition 3, the system is physically unrealizable.

Theorem 1. *Suppose that $f(\infty) = \lim_{t\to\infty} f(t)$ exists and $\lambda(\infty) > 0$. Then, the optimal function $n^*(t)$ that minimizes the overall expected cost $C(n(\cdot))$ involved in the checkpointing process is the unique solution of the following equation:*

$$n^*(t) = \sqrt{\frac{L'\left(\frac{1}{n^*(t)}\right)\cdot f(t)}{c_0\cdot(1-F(t))}} = \sqrt{\frac{L'\left(\frac{1}{n^*(t)}\right)}{c_0}}\cdot\sqrt{\lambda(t)}.$$

To derive the above main result, we need the following auxiliary result, which guarantees the existence and uniqueness of the optimal solution $n^*(t)$.

Lemma 1. *Suppose $0 < L(z)$ is a strictly increasing convex function on $(0,\infty)$ with $L'(0) > 0$. Then, for any positive number u, the equation*

$$\frac{y^2}{L'\left(\frac{1}{y}\right)} = u, \qquad 0 < y < \infty \tag{7}$$

has a unique solution. If we denote the solution by $y = y(u)$, then y is an increasing function of u.

Proof of Lemma 1. Let

$$\varphi(y) = \frac{y^2}{L'\left(\frac{1}{y}\right)}.$$

Since $L(z)$ is convex and strictly increasing, the first derivative $L'(z) > 0$ is increasing. Hence, $L'(\infty) > 0$. This gives $\varphi(0+) = 0$. The condition $L'(0) > 0$ implies $\lim_{y\to\infty}\varphi(y) = \infty$. The function $\varphi(y)$ is obviously continuous on $(0,\infty)$. From $\varphi(0+) = 0$, we conclude that the range of $\varphi(y)$ is $(0,\infty)$. Thus, for any given u $(0 < u < \infty)$, (7) has at least one solution. It is easy to see that $L'(\frac{1}{y})$ decreases in $y > 0$ since L is convex and, thus, L' is increasing. Since $L' > 0$, it follows that $\varphi(y)$ is increasing in $y > 0$. The monotonicity of $y(u)$ immediately follows from the fact that $\varphi(y)$ is increasing in $y > 0$. Therefore, for any positive u, the equation has a unique solution, say, $y = y(u)$. □

Now, we are in a position to derive our main result.

Proof of Theorem 1. Let $x(t) = \int_0^t n(\tau)d\tau$. Then, (6) becomes

$$C(n(\cdot)) = \int_0^\infty \left[c_o\cdot x(t) + L\left(\frac{1}{x'(t)}\right)\right]\cdot f(t)dt.$$

Let $\Psi(t,x,x') = [c_0\cdot x(t) + L(\frac{1}{x'(t)})]\cdot f(t)$. The extreme value can be obtained by Euler's theorem:

$$\frac{\partial\Psi}{\partial x} - \frac{d}{dt}\frac{\partial\Psi}{\partial x'} = 0.$$

Taking the partial derivative of Ψ with respect to x and x', respectively, we obtain

$$\frac{\partial\Psi}{\partial x} = c_o\cdot f(t)$$

and

$$\frac{\partial\Psi}{\partial x'} = L'\left(\frac{1}{x'(t)}\right)\cdot\left(\frac{-1}{(x'(t))^2}\right)\cdot f(t).$$

Substituting the above equations into (8), we have:

$$c_0\cdot f(t) + \frac{d}{dt}\cdot\frac{L'\left(\frac{1}{x'(t)}\right)\cdot f(t)}{(x'(t))^2} = 0. \tag{9}$$

Integrating (9) on both sides, we have:

$$\begin{aligned} c_0 \cdot F(t) + \frac{L'\left(\frac{1}{x'(t)}\right) \cdot f(t)}{(x'(t))^2} &= K, \\ c_0 F(t) + \frac{L'(1/n(t)) \cdot f(t)}{(n(t))^2} &= K, \end{aligned} \tag{10}$$

for a constant K. Note that $n(\cdot) \in \aleph$ implies $n(\infty) > 0$. Since the recovery function $L(\cdot)$ is convex, L' must be bounded on any finite interval. The condition $x(0) = 0$ is automatically satisfied since $x(t) = \int_0^t n(\tau)d\tau$. The natural boundary of Euler's equation $\lim_{t\to\infty} \frac{\partial \Psi}{\partial x'} = 0$ is met (notice that the probability density $f(\infty) = 0$ follows directly the existence of $\lim_{t\to\infty} f(t)$ [19]) since

$$\lim_{t\to\infty} \frac{\partial \Psi}{\partial x'} = \lim_{t\to\infty} \frac{L'\left(\frac{1}{x'(t)}\right) \cdot f(t)}{(x'(t))^2} = 0. \tag{11}$$

Applying (11) to (10), we obtain $K = c_0$. Hence, (10) becomes

$$n^*(t) = \sqrt{\frac{L'\left(\frac{1}{n^*(t)}\right) \cdot f(t)}{c_0 \cdot (1 - F(t))}} = \sqrt{\frac{L'\left(\frac{1}{n^*(t)}\right)}{c_0}} \cdot \sqrt{\lambda(t)}. \tag{12}$$

Equation (12) is equivalent to:

$$\frac{n^2(t)}{L'\left(\frac{1}{n(t)}\right)} = \frac{\lambda(t)}{c_0}$$

for each fixed $t \geq 0$. Let $u = \frac{\lambda(t)}{c_0}$. By Lemma 1, the checkpointing frequency function $n(t)$ can be uniquely determined by $\frac{\lambda(t)}{c_0}$ and, as a matter of fact, $n(t)$ can be expressed as $n(t) = g(\lambda(t))$ for certain increasing function $g(\cdot)$. The optimal checkpointing frequency function $n^*(t)$, which is determined by (12), belongs to the family of functions $\aleph$ because $\lambda(\infty) > 0$. The proof is thus completed. □

We can see from Theorem 1 that 1) the optimal checkpointing frequency $n^*(\cdot)$ is directly proportional to the square root of the failure rate $\lambda(t)$; 2) the optimal checkpointing frequency $n^*(\cdot)$ is proportional to the derivative of recovery function $L(\cdot)$; and 3) a decrease in the checkpoint setup results in an increase in the optimal checkpointing frequency.

The qualitative description of Theorem 1 is in line with our intuition and observation. Moreover, the theorem provides us with a quantitative account of intrinsic mathematical association among the optimal checkpointing frequency, recovery cost, and failure rate, illustrating that the optimal checkpointing frequency by its nature is a function of failure rate. The following corollary, which shows the mathematical equivalence of the Poisson failure rate and constant checkpointing, is a special case of Theorem 1. Its proof is straightforward and thus is omitted.

Corollary 1. *Let the recovery cost function L be linear with the proportional constant c_1 and let the cost for a checkpoint setup c_0 be a constant. The optimal checkpointing is equally spaced if and only if the system failure time is exponentially distributed (failure rate λ). In mathematical terms this means:*

$$n^* = \sqrt{\frac{c_1}{c_0}} \cdot \sqrt{\lambda}.$$

It is worth mentioning that the conclusion stated in the corollary agrees with the results reported in [3], [4], [9], [10], [22], [25], although different criteria of optimality are used.

3 Suboptimal Checkpointing vs. Optimal Checkpointing

In this section, we use an example to illustrate that, with the objective of minimizing the expected overall cost, periodic checkpointing fails to yield the optimal solution when the system follows a non-Poisson failure distribution. The example is described as follows.

Example 3.1. Suppose system failure follows a Weibull distribution with shape parameter $\alpha = 1.5$ and a mean of 60 hours. The corresponding failure rate is calculated as $\lambda(t) = 1.5 \cdot \left(\frac{\Gamma(5/3)}{60}\right)^{1.5} \cdot t^{1/2}$ with $\lambda(\infty) = \infty$, where $\Gamma(\cdot)$ is the Gamma function. The checkpoint setup is assumed to be constant as 1 minute, i.e., $c_0 = \frac{1}{60}$ hour. The recovery function $L(\cdot)$ is of the form: $L(z) = 0.5 \cdot z + 0.1$, i.e., $c_1 = 0.5$. Thus, the probability density function is written as $f(t) = 1.5 \cdot (\frac{\Gamma(5/3)}{60})^{1.5} t^{0.5} \cdot \exp(-(\frac{\Gamma(5/3)}{60} \cdot t)^{1.5})$ by the well-known mathematical relationship [18] between failure rate $\lambda(t)$ and probability density $f(t)$.

$$f(t) = \exp\left(-\int_0^t \lambda(\tau)d\tau\right) \cdot \lambda(t).$$

Based on Theorem 1, the optimal checkpointing frequency $n^*(t)$ such that $C(n^*(\cdot)) = \min_{n(\cdot)\in\aleph} C(n(\cdot))$ is given by:

$$n^*(t) = \sqrt{\frac{c_1}{c_0} \cdot \lambda(t)} = \sqrt{45} \cdot \left(\frac{\Gamma(5/3)}{60}\right)^{\frac{3}{4}} t^{\frac{1}{4}}. \tag{13}$$

Then, the sequence of checkpoint instants is determined by (1) as:

$$t_i = (i)^{4/5} \cdot \left(\frac{5}{4}\right)^{4/5} \cdot \frac{1}{(45)^{2/5}} \cdot \left(\frac{60}{\Gamma(5/3)}\right)^{3/5}, \quad i = 1, 2, \cdots,$$

with $t_0 \equiv 0$. It is clear that it is an aperiodic checkpointing by Definition 2. The total expected cost (elapse time) of the optimal aperiodic checkpointing is:

$$\begin{aligned} C(n^*(\cdot)) &= \int_0^\infty \left(\int_0^t \frac{n^*(\tau)}{60} d\tau + \frac{0.5}{n^*(t)} + 0.1\right) \cdot f(t)dt \\ &= \frac{2 \cdot \Gamma(5/6)}{\sqrt{3 \cdot \Gamma(5/3)}} + 0.1. \end{aligned}$$

We will compare the total expected cost of the optimal aperiodic checkpointing with that of the best periodic checkpointing with the same mean failure rate. We define $\aleph_c = \{n_\alpha(t) = \alpha, \ \alpha > 0, \ t > 0\}$ to represent the family of all constant checkpoint frequency functions.

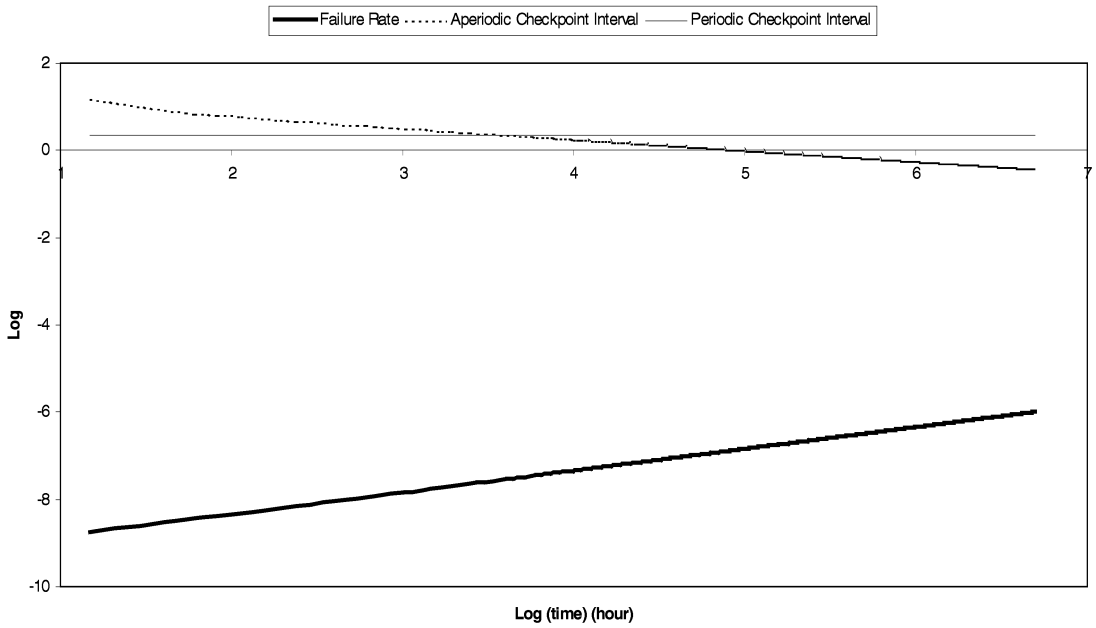

Fig. 3. Failure rate vs. checkpoint interval for aperiodic and periodic checkpointing.

It is clear that $\aleph_c$ is a subfamily of $\aleph$. Then, the total expected cost of the best periodic checkpointing is given by:

$$C(n_\alpha(\cdot)) = \int_0^\infty \left(\int_0^t \frac{n_\alpha(\tau)}{60} d\tau + \frac{0.5}{n_\alpha(t)} + 0.1 \right) f(t) dt = \frac{\alpha}{60} \cdot \mu + \frac{0.5}{\alpha} + 0.1,$$

where μ is the mean failure time, which is defined as:

$$\mu = \int_0^\infty \tau \cdot f(\tau) d\tau.$$

Within the family $\aleph_c$, let's choose α^* such that $C(n_{\alpha^*}(\cdot)) = \min_{n_\alpha(\cdot) \in \aleph_c} C(n_\alpha(\cdot))$. For this purpose, we take the derivative of $C(n_\alpha(\cdot))$ with respect to α and obtain:

$$\frac{d}{d\alpha} C(n_\alpha(\cdot)) = \mu \cdot c_0 - \frac{c_1}{\alpha^2} = \frac{\mu}{60} - \frac{0.5}{\alpha^2}$$

and

$$\frac{d^2}{d\alpha^2} C(n_\alpha(t)) = \frac{1}{\alpha^3} > 0.$$

Hence, $\min_{n_\alpha(\cdot) \in \aleph_c} C(n_\alpha(\cdot))$ exists and is obtained at α^*, satisfying

$$\alpha^* = \sqrt{\frac{30}{\mu}} = \sqrt{\frac{1}{2}}.$$

Using (1), we obtain the sequence of checkpoint instants as $t_i = i \cdot \sqrt{2}, i = 1, 2, \cdots$, with $t_0 \equiv 0$. That is, the best periodic checkpoint interval is $\sqrt{2} \cdot (i - (i-1)) = 1.414$ hour. The overall expected cost for the best periodic checkpointing is calculated as:

$$C(n_{\alpha^*}(\cdot)) = \frac{\alpha^* \cdot \mu}{60} + \frac{0.5}{\alpha^*} + 0.1 = 2 \cdot \sqrt{0.5} + 0.1.$$

Hence, the cost difference between the best periodic checkpointing and the optimal aperiodic checkpointing, $C(n_{\alpha^*}(\cdot)) - C(n^*(\cdot))$, is calculated as:

$$C(n_{\alpha^*}(\cdot)) - C(n^*(\cdot)) = 2 \cdot \sqrt{0.5} + 0.1 - \frac{2 \cdot \Gamma(5/6)}{\sqrt{3} \cdot \Gamma(5/3)} - 0.1 = 0.0429801623 \text{ hour}.$$

The calculation shows that the optimal aperiodic checkpointing is better than the best periodic checkpointing in terms of the total expected cost because the best periodic checkpointing requires 2.5788 minutes (0.04298*60) more than necessary in expected cost.

Fig. 3 is plotted by using the log/log representation of the measurements (y axis) and time (x axis) for better visualization, representing the relationship among the time-varying failure rate, and the checkpoint intervals generated by the both optimal aperiodic and best periodic checkpointing.

Fig. 3 illustrates the dependence of the optimal aperiodic checkpointing on the instantaneous failure rate: Its checkpoint interval (the reciprocal of checkpointing frequency) decreases as the failure rate increases, whereas the checkpoint interval for the periodic checkpointing is constant and, hence, is insensitive to the instant failure rate, initially suffering from an excessive checkpointing and ultimately performing a deficient checkpointing.

The superiority of the optimal aperiodic checkpointing over the best periodic checkpointing is ensured and justified by its mathematical basis: The use of calculus of variation leads to a global minimum $n^*(\cdot)$ in $\aleph$, as distinct from a local minimum $n_{\alpha^*}(\cdot)$ obtained in $\aleph_c$. Since $\aleph_c$ is a subfamily of $\aleph$, it can be inferred that the optimal solution obtained in $\aleph$ is superior to one obtained in its subfamily $\aleph_c$ because the periodic checkpointing is a special case of the aperiodic checkpointing.

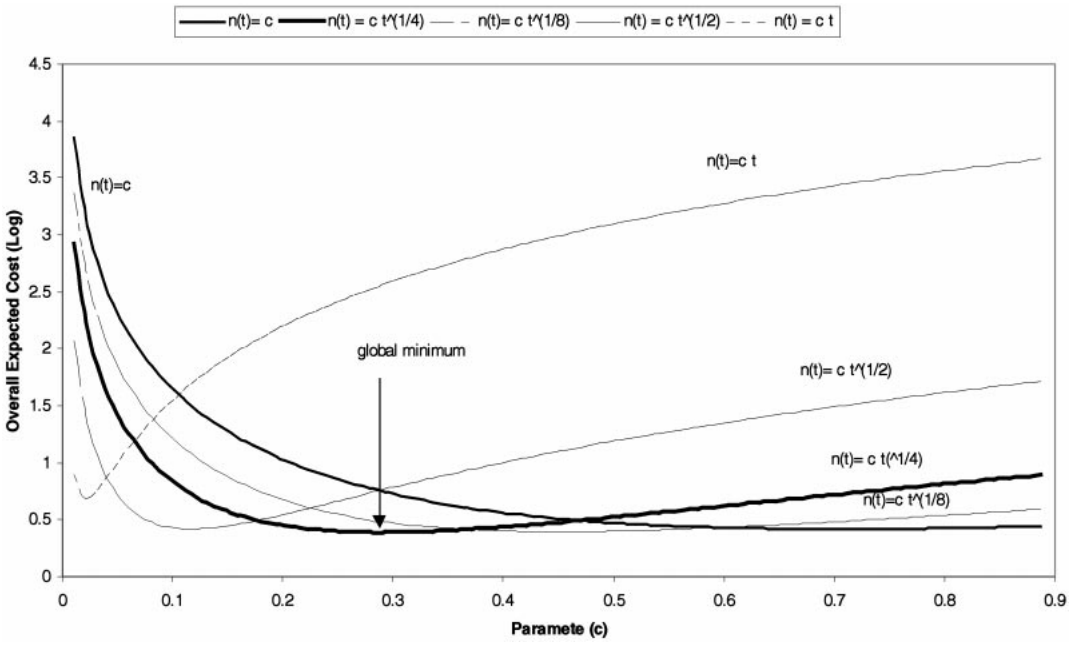

Fig. 4. Diagram of total expected cost vs. parameter c.

To visualize the bearing of the different checkpointing frequency functions on the total expected cost, we use the five checkpointing frequency functions:

1. $n(t) = c$;
2. $n(t) = c \cdot t^{1/4}$;
3. $n(t) = c \cdot t^{1/8}$;
4. $n(t) = c \cdot t^{1/2}$; and
5. $n(t) = c \cdot t$.

Each curve in Fig. 4 is plotted by changing the respective parameter c in the five given checkpointing frequency functions from 0.01 to 0.9. For a given parameter c and the form of function, the total expected cost is calculated by using (6). The numerical calculation demonstrates that, among the five checkpointing frequency functions of interest, the function $n(t) = c \cdot t^{1/4}$ reaches the lowest point (a global minimum) with the parameter $c = \sqrt{45} \cdot (\frac{\Gamma(5/3)}{60})^{3/4} = 0.287924$. With the help of calculus of variations, we are able to identify the optimal checkpointing function $n^*(t)$ that makes its total expected cost globally minimal.

In practical applications, the system failure rate is believed to vary with the working load of a computer system. For example, software failures or system crashes are more frequent when the system load is high [9], [10], [21]. A system load is cyclic, having a slack period at night and a period of heavy load in daytime. The observation suggests that the system failure can be described as a nonhomogeneous Poisson process [9], [10] in which the arrival rate varies over time. The following example illustrates how the derived formula is used to tackle cyclical system failure.

Example 3.2. Assume that the system failure rate is a periodic step function described below, while the cost for establishing a checkpoint is a constant c_0 and the recovery cost function is $L(z) = c_1 \cdot z$.

$$\lambda(t) = \begin{cases} \lambda_1 & 0 = \tau_0 \le t < \tau_1 \\ \lambda_2 & \tau_1 \le t < \tau_2 \\ \lambda_3 & \tau_2 \le t < \tau_3 \\ \lambda_4 & \tau_3 \le t < \tau_4 \end{cases}$$

and $\lambda(t) = \lambda(t - \tau_4)$ if $t\tau_4$. The optimal checkpoint rate is determined by using Theorem 1:

$$n^*(t) = \begin{cases} \sqrt{\frac{c_1 \cdot \lambda_1}{c_0}} & 0 = \tau_0 \le t < \tau_1 \\ \sqrt{\frac{c_1 \cdot \lambda_2}{c_0}} & \tau_1 \le t < \tau_1 \\ \sqrt{\frac{c_1 \cdot \lambda_3}{c_0}} & \tau_2 \le t < \tau_3 \\ \sqrt{\frac{c_1 \cdot \lambda_4}{c_0}} & \tau_3 \le t < \tau_4 \end{cases}$$

and $n^*(t) = n^*(t - \tau_4)$ if $t \ge \tau_4$.

Once the checkpointing frequency function $n^*(t)$ is obtained, the sequence of discrete checkpoint instants $\{t_i, i \ge 1\}$ can be determined by (1).

4 Conclusion

In this paper, we derive a closed-form formula that establishes the connection between the optimum checkpointing frequency and a general failure rate, with the objective of minimizing the total expected cost of checkpointing and recovery. Our theoretical approach is based on the calculus of variations, differing in fundamental ways from the stationary analysis in the literature. The derived formula is applicable to the family of failure rate functions with the eventual property $\lambda(\infty) > 0$, i.e., physically realizable systems, ensuring that the optimum solution makes the total expected cost globally minimal. The present study differs from the work in the literature in its departure

from the stationary analysis, and its adoption of the calculus of variations as a mathematical means of problem solving.

The gain of optimal aperiodic checkpointing over the best periodic one is analyzed in Section 2 and is fully substantiated by the examples given in Section 3. Periodic checkpointing is a special case of aperiodic checkpointing. The theoretical results obtained indicate that the optimal checkpointing is equally spaced if and only if the system follows a Poisson failure, which is consistent with the conclusion [4], [5], [9], [10], [23], [25] that periodic checkpointing (equidistant strategy) is optimal under Poisson failures, assuming that no failure would occur during checkpointing and recovery. Furthermore, the derived result also indicates that non-Poisson failure demands that the optimal checkpointing be aperiodic.

The complexity of the model usually varies significantly with the assumptions used and the criteria of optimality sought. With the variety of assumptions and criteria of optimality, it is hard to make sweeping generalizations; however, there are many useful and realistic extensions that need to be addressed:

1. The overhead of the time-varying checkpoint setup should be considered. In this paper, we only consider the constant overhead of checkpoint setup. This restriction, though widely used in the literature, is somewhat unrealistic in some circumstances since the overhead of checkpointing in fact varies with the load of the system.
2. The different criterion of optimality such as maximizing system availability should also be addressed.
3. System failure is not self-evident. Hence, a system failure is detected by the self-inspection process. The latency between the occurrence of a failure and failure detection should be taken into consideration.

The research direction, guided by the principle of simplicity, primarily focuses on formulating a more realistic model with less restrictive assumptions, with different criteria of optimality. Relaxing restrictions imposed by the assumptions would add a layer of complications to the model, but it is rewarded by the increasing scope of application. Seeking tractable and widely applicable solutions, demanded by the practical needs of efficient implementation and easy performance tuning, would be a challenging task.

ACKNOWLEDGMENTS

The authors would like to thank Scott Knoke, Tracy Mullen, and Gretchen Katzenberger for the constructive comments and valuable discussion in the preparation of the manuscript, which goes much beyond the usual inspection, amounting to redrafting. The authors wish to express their deep gratitude to the anonymous reviewers for their valuable comments. Yibei Ling would like to thank his mother, Meiqin Zhou, for her encouragement and long-time support.

REFERENCES

[1] L.B. Boguslavsky, E.G. Coffman, E.N. Gilbert, and A.Y. Kreinin, "Scheduling Checks and Saves," *OOSA J. Computing,* vol. 4, no. 1, pp. 60-69, 1992.

[2] J.L. Bruno and E.G. Coffman, "Optimal Fault-Tolerant Computing on Multiprocessor Systems," *Acta Informatica,* vol. 34, pp. 881-904, 1997.

[3] J.L. Bruno, E.G. Coffman, J.C. Lagarias, T.J. Richardson, and P.W. Shor, "Processor Shadowing: Maximizing Expected Throughput in Fault-Tolerant Systems," *Math. Operations Research,* vol. 24, no. 2, pp. 362-382, May 1999.

[4] K.M. Chandy, J.C. Browne, C.W. Dissly, and W.R. Uhrig, "Analytic Models for Rollback and Recovery Strategies in Database Systems," *IEEE Trans. Software Eng.,* vol. 1, no. 1, pp. 100-110, Mar. 1975.

[5] K.M. Chandy, "A Survey of Analytic Models for Rollback and Recovery Strategies," *Computer,* vol. 8, no. 5, pp. 40-47, 1975.

[6] E.G. Coffman and E.N. Gilbert, "Optimal Strategies for Scheduling Checkpoints and Preventive Maintenance," *IEEE Trans. Reliability,* vol. 39, no. 1, pp. 9-18, Apr. 1990.

[7] E.G. Coffman, L. Flatto, and P.E. Wright, "A Stochastic Checkpoint Optimization Problem," *SIAM J. Computing,* vol. 22, no. 3, pp. 650-659, June 1993.

[8] A. Duda, "The Effects of Checkpointing on Program Execution Time," *Information Processing Letters,* vol. 16, no. 5, pp. 221-229, June 1983.

[9] P. L'Ecuyer and J. Malenfant, "Computing Optimal Checkpointing Strategies for Rollback and Recovery Systems," *IEEE Trans. Computers,* vol. 37, no. 4, pp. 491-496, Apr. 1988.

[10] E. Gelenbe and M. Hernandez, "Optimum Checkpoints with Age Dependent Failures," *Acta Informatica,* vol. 27, pp. 519-531, 1990.

[11] A. Goyal, V.F. Nicola, A. Tantawi, and K. Trivedi, "Reliability of System with Limited Repairs," *IEEE Trans. Reliability,* vol. 36, no. 2, pp. 202-207, 1987.

[12] V. Grassi, L. Donatiello, and S. Tucci, "On the Optimal Checkpointing of Critical Tasks and Transaction-Oriented Systems," *IEEE Trans. Software Eng.,* vol. 18, no. 1, pp. 72-77, Jan. 1992.

[13] C.M. Krishna, K.G. Shin, and Y.H. Lee, "Optimization Criteria for Checkpoint Placements," *Comm. ACM,* vol. 27, no. 10, pp. 1008-1012, Oct. 1984.

[14] C.H.C. Leung and Q.H. Choo, "On the Execution of Large Batch Programs in Unreliable Computing Systems," *IEEE Trans. Software Eng.,* vol. 10, no. 4, pp. 444-450, July 1984.

[15] J. Mi, "Interval Estimation of Availability of a Series System," *IEEE Trans. Reliability,* vol. 40, pp. 541-546, 1991.

[16] V.F. Nicola and J.M. van Spanje, "Comparative Analysis of Different Models of Checkpointing and Recovery," *IEEE Trans. Software Eng.,* vol. 16, no. 8, pp. 807-821, Aug. 1990.

[17] V.F. Nicola, "Checkpointing and the Modeling of Program Execution Time," *Software Fault Tolerance,* M.R. Lyu, ed., pp. 167-188, John Wiley & Sons, 1995.

[18] J.S. Plank, K. Li, and M.A. Puening, "Diskless Checkpointing," *IEEE Trans. Parallel and Distributed Systems,* vol. 9, no. 10, pp. 972-986, Oct. 1998.

[19] S.M. Ross, *Stochastic Processes.* New York: Wesley, 1996.

[20] K.G. Shin, T.H. Lin, and Y.H. Lee, "Optimal Checkpointing of Real-time Tasks," *IEEE Trans. Computers,* vol. 36, no. 11, pp. 1328-1341, Nov. 1987.

[21] E. de Souza e Silva and H.R. Gail, "Calculating Cumulative Operational Time Distribution of Repairable Computer Systems," *IEEE Trans. Computers,* vol. 35, no. 4, pp. 322-332, Apr. 1986.

[22] U. Sumita, N. Kaio, and P.B. Goes, "Analysis of Effective Service Time with Age Dependent Interruptions and Its Application to Optimal Rollback Policy for Database Management," *Queuing Systems: Theory and Applications,* vol. 4, pp.193-212, 1989.

[23] A. Tantawi and M. Ruschitzka, "Performance Analysis of Checkpointing Strategies," *ACM Trans. Computer Systems,* vol. 2, no. 2, pp. 123-144, May 1984.

[24] S. Toueg and O. Babaoglu, "On the Optimum Checkpoint Selection Problem," *SIAM J. Computing,* vol. 13, no. 3, pp. 630-649, Aug. 1984.

[25] J.W. Young, "A First Order Approximation to the Optimum Checkpoint Interval," *Comm. ACM,* vol. 17, no. 9, pp. 530-531, 1974.

Yibei Ling received the BS degree in electrical engineering from Zhejiang University in 1982, the MS degree in biostatistics from Shanghai Medical University, in 1985, and the PhD degree in computer science from Florida State University at Miami in 1995. Currently, he is a senior research scientist at Applied Research Laboratories, Telcorida Technologies (formerly Bellcore), where he is developing wireless middleware systems and wireless messaging systems. His research interests include distributed systems, system performance evaluation, query optimization in database management systems, and biological modeling. He has published several papers in the *IEEE transactions on Knowledge and Data Engineering*, *IEEE Transactions on Biomedical Engineering*, *SIGMOD*, *Data Engineering*, *Information Systems*, and *Operating System Review*. He is a member of the IEEE.

Jie Mi received his MSc degree in applied mathematics from Shanghai Jiaotong University, People's Republic of China, his MA degree in mathematics and his PhD degree in statistics from the University of Pittsburgh. His papers have appeared in journals such as the *Journal Shanghai Jiaotong University*, *Journal of Applied Sciences*, *Journal of Engineering Mathematics*, *Communications in Statistics—Theory & Methods*, *Journal of Applied Probability*, *Advances in Applied Probability*, *Journal of Statistical Planning & Inference*, *Statistica*, *Naval Research Logistics*, *Journal of Nonparametric Statistics*, *Operations Research*, *Statistics & Probability Letters*, *Probability in the Engineering & Information Sciences*, *IEEE Transactions on Systems, Man, and Cybernetics*, and *IEEE Transactions on Reliability*. He is an associate professor and the chairperson of the Department of Statistics of Florida International University.

Xiaola Lin received the BS and MS degrees in computer science from Peking University, Beijing, China, and the PhD degree in computer science from Michigan State University, East Lansing, in 1992. He is currently with the Department of Electric and Electronic Engineering, the University of Hong Kong. His research interests include parallel and distributed computing, design and analysis of algorithms, and computer network.