

AN OPTIMAL N -JOB BACKUP POLICY MAXIMIZING AVAILABILITY FOR A HARD COMPUTER DISK

Hiroaki Sandoh

Hajime Kawai

University of Marketing & Distribution Sciences Tottori University

(Received July 18, 1990; Revised December 20, 1990)

Abstract Personal computers and engineering work stations are generally used with hard computer disks which preserve a variety of files for reference. Some or all of these files are sometimes lost because of human errors and failures of the computer system or the hard disk. Making a backup copy of files on floppy disks or magnetic tapes protects a user from such serious losses. The present study discusses an optimal backup policy for a hard computer disk where a backup copy of files in the hard disk is made when N jobs of updating or creating files are finished. The availability of the proposed backup policy is formulated and the existence of the optimal integer N^* which maximizes the availability is examined. It is shown that there always exists an optimal integer N^* equal to or greater than 1. A numerical example is presented to illustrate the theoretical model for determining the optimal backup policy using gamma distributions to describe the setup time, the job processing time and the backup time for a system subject to exponential failures.

1. Introduction

Personal computers and engineering work stations have rapidly come into wide use, since they can, in recent years, be purchased at lower prices and can be handled more easily. Most of systems contain hard computer disks (hard disks) which store many files for reference. While using a system, some or all the files in the hard disk may be lost because of human errors or failures of the hardware devices which comprise the computer system. This is called "a hard disk failure". It is important to make a backup copy of files on floppy disks or magnetic tapes to protect ourselves from such serious losses. Similar problems can be observed in the main internal memory of a main frame computer, where data stored in the main memory are sometimes lost because of a failure of the main frame computer. For these systems, several studies on rollback and recovery strategies have been reported [1-5] which suggest to periodically make a copy of data in the main memory on hard disks.

For protecting files on the hard disks of personal computers and engineering work stations, an optimal economic backup policy has been proposed [6]. It was shown that an optimal "age" T exists at which a backup copy of files should be made, where "age" refers to the elapsed time since the previous operation of a backup copy or the recovery from a hard disk failure. In this formulation, however, users of the hard disk may have to stop their jobs of creating or updating files at age T . A more practical consideration would be to make a backup copy of files at the end of their processing jobs.

The present study considers a modified policy for a hard disk where a backup copy is made when N jobs of creating or updating files are finished. The system availability is used as the objective function in order to determine the backup policy described by the selection of an optimal integer N^* to maximize system availability. It is shown, as a result, that there always exists an optimal integer. An example is shown to illustrate the theoretical underpinnings of the backup policy formulation.

2. N -Job Backup Policy and Assumptions

Consider an N -job backup policy where a backup copy of files are to be made after N ($N = 1, 2, \dots$) jobs are completed. In the case of a hard disk failure, the hard disk can only be partially recovered by copying the contents of the previous backup floppy disks to the hard disk. It is to be noted that, the recovery would be partial since the backup disks preserve only the files created and/or updated at the previous backup operation. Under these conditions it is desired to determine the optimal N after which the files generated or changed since the last backup will be backed up so as not to lose the newly run N jobs. When these newly run N jobs are lost, the processing time of these jobs represents an ineffective usage time and decreases the useful/effective system time or the systems effective availability.

We make the following assumptions:

- (a) At each backup time, a copy of the files updated and/or created since the preceding backup time is added to the backup disks.
- (b) The failure time, X , of the hard disk follows an exponential distribution with mean $1/\lambda$, because these failures occur randomly in time.
- (c) The processing time T for each job of updating files is independently and identically distributed (iid) and the cumulative distribution function (cdf) of a processing time is denoted as $H(t)$.
- (d) The cdf of a setup time S for a backup operation is denoted as $A(s)$ and the time U for making a backup copy of files updated by each job is iid $B(u)$ and U is statistically independent of S .
- (e) The time V for a backup is given by the sum of the setup time and the total backup time of the N jobs.
- (f) The cdf of the time W for recovering a hard disk failure (recovery time) is $G(w)$ with mean g , and any hard disk failure can instantly be detected. A hard disk failure does not occur during the recovery.

From assumption (c), the total processing time T of N jobs has N -fold convolution of $H(t)$ with itself, i.e.,

$$H^{(N)}(t) = \int_0^t H^{(N-1)}(t-y) dH(y), \quad (2.1)$$

$$H^{(1)}(t) = H(t). \quad (2.2)$$

From assumptions (d) and (e), the operation time V of a backup follows;

$$B_N(v) = \int_0^\infty B^{(N)}(v-y) dA(y), \quad (2.3)$$

where $B^{(N)}(v)$ is the N -fold convolution of $B(v)$ with itself.

3. Formulation of Availability

Under the proposed N -job backup policy and the assumptions, a renewal reward process [7] is generated, where the time at which one of the following three events occurs corresponds to a renewal point;

- (a) A backup operation has successfully been executed after N jobs had been completed.
- (b) A hard disk failure occurred during a backup operation immediately after N jobs had been completed, and the recovery using the backup disks has just been finished.

In this case, the state of the hard disk is identical to that at the previous backup time.

- (c) A hard disk failure occurred before N jobs were completed, and the recovery has just been finished. The state of the hard disk is, also in this case, identical to that at the previous backup time.

Figure 1 shows these three cases with effective and ineffective times in each case. In Fig. 1, E_0 , E_1 and E_2 express a renewal point, a start point of a backup operation and a point of a hard disk failure, respectively.

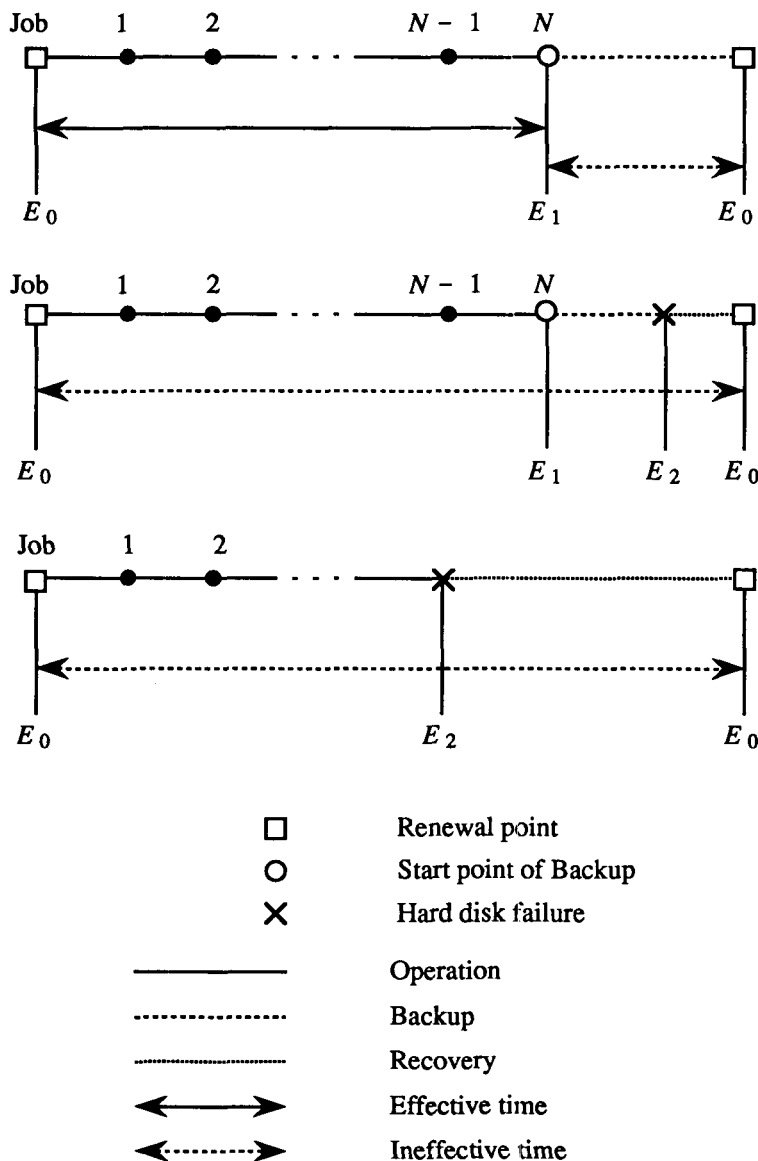


Fig. 1 Cases with effective and ineffective times.

Let us here consider the expected effective time per unit time over an infinite time span expressed by

$$W(N) = \lim_{t \rightarrow +\infty} \frac{E[\text{total effective time during}(0, t]]}{t}, \quad (3.1)$$

which can be regarded as the limiting availability of the proposed backup policy. An optimal backup policy can be obtained by maximizing $W(N)$ in Eq. (3.1) with respect to N . It is well known in renewal theory [7] that $W(N)$ can be rewritten as

$$W(N) = \frac{D(N)}{C(N)}, \quad (3.2)$$

where $C(N)$ and $D(N)$ respectively signify the expected time and the expected effective time over the time between two successive renewal points. Many models similar to the above are seen in studies on replacement policies [8]. From assumptions (b)-(f), $C(N)$ is expressed by

$$C(N) = \int_0^\infty (t + m_1)e^{-\lambda t} dH^{(N)}(t) + \int_0^\infty (t + m_2)\overline{H}^{(N)}(t)\lambda e^{-\lambda t} dt, \quad (3.3)$$

where

$$\overline{H}^{(N)}(t) = 1 - H^{(N)}(t), \quad (3.4)$$

and m_1 and m_2 denote the mean times to E_0 from E_1 and E_2 respectively.

It can be noticed in Fig. 1 that m_1 is expressed by

$$m_1 = \int_0^\infty ve^{-\lambda v} dB_N(v) + \int_0^\infty (v + m_2)\lambda e^{-\lambda v} \overline{B}_N(v) dv. \quad (3.5)$$

Figure 1 also implies that m_2 is given by

$$m_2 = \int_0^\infty \overline{G}(w) dw = g. \quad (3.6)$$

From Eqs. (3.5) and (3.6), we have

$$m_1 = (g + 1/\lambda)(1 - ab^N), \quad (3.7)$$

$$a \equiv A^*(\lambda), \quad (3.8)$$

$$b \equiv B^*(\lambda), \quad (3.9)$$

where

$$A^*(\lambda) \equiv \int_0^\infty e^{-\lambda s} dA(s), \quad (3.10)$$

$$B^*(\lambda) \equiv \int_0^\infty e^{-\lambda u} dB(u). \quad (3.11)$$

Hence, we obtain, from Eqs. (3.3), (3.6) and (3.7), the expected time $C(N)$ over the time between two successive renewal points is

$$C(N) = (g + 1/\lambda)(1 - ab^N h^N), \quad (3.12)$$

$$h \equiv H^*(\lambda), \quad (3.13)$$

where

$$H^*(\lambda) \equiv \int_0^\infty e^{-\lambda t} dH(t). \quad (3.14)$$

On the other hand, the expected effective time is given by

$$\begin{aligned} D(N) &= \int_0^\infty e^{-\lambda t} \int_0^t x dH^{(N)}(x) dB_N(t-x) \\ &= \int_0^\infty t e^{-\lambda t} dH^{(N)}(t) \int_0^\infty e^{-\lambda t} dB_N(t) \\ &= p N h^{N-1} a b^N, \end{aligned} \quad (3.15)$$

where

$$p \equiv \int_0^\infty t e^{-\lambda t} dH(t). \quad (3.16)$$

From Eqs. (3.2), (3.12) and (3.15), we have

$$W(N) = \frac{ap}{(g+1/\lambda)h} \frac{N b^N h^N}{1 - a b^N h^N}. \quad (3.17)$$

It should be noted in Eq. (3.17) that the optimal value of N minimizing $W(N)$ is independent of g , the mean recovery time.

4. Optimal Backup Policy

This section examines the conditions under which an optimal value of N exists. Let $V(N)$ be defined as

$$V(N) = \frac{1 - a q^N}{N q^N}, \quad (4.1)$$

where,

$$q \equiv b h (< 1). \quad (4.2)$$

Since maximization of $W(N)$ with respect to N is equivalent to minimization of $V(N)$, we minimize $V(N)$ in the following.

We here have

$$\text{sgn}[V(N+1) - V(N)] = \text{sgn}[a q^{N+1} - (N+1)q + N], \quad (4.3)$$

where $\text{sgn}[\cdot]$ signifies the sign of \cdot . Letting $U(N)$ be defined as

$$U(N) = a q^{N+1} - (N+1)q + N, \quad (4.4)$$

we examine the sign of $U(N)$ in the following. We have

$$U(1) = a q^2 - 2q + 1, \quad (4.5)$$

$$\lim_{N \rightarrow \infty} U(N) = +\infty. \quad (4.6)$$

Furthermore, since $a \leq 1$ and $q < 1$, we obtain

$$U(N+1) - U(N) = (1-q)(1 - a q^{N+1}) > 0, \quad (4.7)$$

which indicates that $U(N)$ is increasing in N .

From Eqs. (4.5), (4.6) and (4.7), the existence of the optimal solution is discussed for the following two cases:

(1) $a < 1$

This case can furthermore be classified into two subcases:

$$(i) \quad bh \leq (1 - \sqrt{1-a})/a$$

In this subcase, $U(1) \geq 0$, and thus we have $U(N) \geq 0$ for $N = 1, 2, \dots$. Hence, $V(N)$ is increasing in N and consequently, the optimal value N^* is 1, i.e., the optimal policy is to make a backup copy after each job is finished.

$$(ii) \quad bh > (1 - \sqrt{1-a})/a$$

In this subcase, we have $U(1) < 0$ from Eq. (4.5), and it follows that the sign of $U(N)$ varies from negative to positive. This indicates that $V(N)$ first decreases and then increases with increasing N and that there exists a unique finite solution N^* minimizing $V(N)$.

$$(2) \quad a = 1$$

When $a = 1$, the setup time is zero. In this case, we have $U(1) > 0$ from Eq. (4.5) and thus $U(N) > 0$ for $N = 1, 2, \dots$. This indicates that $V(N)$ is increasing in N , and consequently the optimal value N^* of N is 1.

5. Numerical Example

This section presents a numerical example of the proposed backup policy. It is assumed that the setup time S , the backup time V for each job and the processing time U of each job follows gamma distributions $Ga(\alpha, \beta)$, $Ga(\gamma, \delta)$ and $Ga(\zeta, \eta)$, respectively, where the pdf of $Ga(\alpha, \beta)$ is defined as

$$p(z; \alpha, \beta) = \frac{\alpha^\beta z^{\beta-1}}{\Gamma(\beta)} e^{-\alpha z}, \quad \alpha, \beta > 0. \quad (5.1)$$

Then a, b, h and p defined in Section 4 respectively become

$$a = \left(\frac{\alpha}{\lambda + \alpha}\right)^\beta, \quad (5.2)$$

$$b = \left(\frac{\gamma}{\lambda + \gamma}\right)^\delta, \quad (5.3)$$

$$h = \left(\frac{\zeta}{\lambda + \zeta}\right)^\eta, \quad (5.4)$$

and

$$p = \frac{\eta \zeta^\eta}{(\lambda + \zeta)^{\eta+1}}. \quad (5.5)$$

Table 1 show values of N^* and their corresponding availability in the case of $\alpha = 2.0, \beta = 0.1, \gamma = 5.0, \delta = 0.5, \zeta = 2.0, \eta = 2.0$ and $g = 3.0$. In this case, the mean setup time, the mean backup time for each job, the mean processing time of each job and the mean recovery time are 0.05, 0.1, 1.0 and 3.0, respectively. It is observed in Table 1 that the optimal value of N decreases with increasing λ which is intuitively anticipated. It is also seen in Table 1 that $W(N^*)$ tends to be decreasing in λ .

6. Conclusions

This study proposed an optimal backup policy for a hard computer disk which makes a backup copy of files in the hard computer disk when N jobs of creating or updating files are finished. The system availability was formulated as the objective function to determine the optimal value of N . As a result, it was shown that there always exists an optimal integer N^* maximizing the system availability. A numerical example was also shown to illustrate

Table 1 Optimal Backup Policies

 $(\alpha = 2.0, \beta = 0.1, \gamma = 5.0, \delta = 0.5, \zeta = 2.0, \eta = 2.0, g = 3.0)$

λ	N^*	$W(N^*)$	λ	N^*	$W(N^*)$
0.00001	88	0.9060	0.001	9	0.8971
0.00002	65	0.9055	0.002	6	0.8905
0.00003	52	0.9085	0.003	5	0.8847
0.00004	46	0.9074	0.004	5	0.8795
0.00005	41	0.9066	0.005	4	0.8746
0.00006	37	0.9061	0.006	4	0.8699
0.00007	34	0.9070	0.007	3	0.8653
0.00008	32	0.9065	0.008	3	0.8611
0.00009	30	0.9060	0.009	3	0.8569
0.0001	29	0.9056	0.01	3	0.8527
0.0002	20	0.9043	0.02	2	0.8157
0.0003	17	0.9033	0.03	2	0.7823
0.0004	14	0.9021	0.04	1	0.7510
0.0005	13	0.9012	0.05	1	0.7254
0.0006	12	0.9003	0.06	1	0.7012
0.0007	11	0.8995	0.07	1	0.6782
0.0008	10	0.8986	0.08	1	0.6564
0.0009	10	0.8978	0.09	1	0.6357
			0.1	1	0.6161

the proposed method. The proposed method is effective in case it is difficult for the user of a hard disk to suspend his/her job in order to conduct a backup operation.

Acknowledgement

The authors are grateful to Professor S.J. Deutsch of Northeastern University. His suggestions have made the paper more readable. This research was in part supported by UMDS (University of Marketing & Distribution) Research Fund (1989–1990).

References

- [1] K.M. Chandy and C.V. Ramamoorthy: Rollback and recovery strategies for computer programs, *IEEE Trans. Computer*, Vol. C21 (1972), 546–556.
- [2] K.M. Chandy, J.C. Browne, C.W. Dissly and W.R. Uhrig: Analytic models for rollback and recovery strategies in data base system, *IEEE Trans. Software Eng.*, Vol. SE-1 (1975), 100–110.
- [3] K.M. Chandy: A survey of analytic models of rollback and recovery strategies, *Computer*, Vol. 8 (1975), 40–47.
- [4] S. Toueg and Ö Babaoglu: On the optimum checkpoint selection problems, *SIAM J. Computer*, Vol. 13 (1984), 630–649.
- [5] N. Kaio and S. Osaki: A note on optimum checkpointing policies, *Microelectronics & Reliability*, Vol. 25 (1985), 451–453.
- [6] H. Sandoh, N. Kaio and H. Kawai: An Optimal Back-up Policy of Floppy and Hard Disks

- (in Japanese), *Transactions of Institute of Electronics, Information & Communication Engineers*, Vol. J73-D1 (1990), 336–341.
- [7] S.M. Ross: *Applied probability models with optimization applications*, Holden-Day, San Francisco, CA (1970).
- [8] R.E. Barlow, F. Proschan: *Mathematical theory of reliability*, Wiley, NY (1965).

Hiroaki Sandoh
Dept. of Business Administration
Faculty of Commerce
University of Marketing &
Distribution Sciences
Gakuen-nishi-machi, Nishi-ku,
Kobe, 651-21
Japan