Optimal Backup Interval for a Database System with Full and Periodic Incremental Backup

Cunhua Qian and Yingyan Huang

Nanjing University of Technology, College of Management Science and Engineering, Nanjing, China
Email: qch64317@njut.edu.cn

Xufeng Zhao and Toshio Nakagawa

Aichi Institute of Technology, Department of Marketing and Information System, Toyota, Japan
Email: toshi-nakagawa@aitech.ac.jp

Abstract—This paper considers the following backup scheme for a database system: a database is updated at a nonhomogeneous Poisson process and an amount of updated files accumulates additively. To ensure the safety of data, full backup are performed at time NT=L or when the database fails, whichever occurs first, and between them, incremental backups are made at periodic times iT (i=1, 2, ···, N-1) so as to make the backup efficiently. Using the theory of cumulative processes, the expected cost is obtained, and optimal numbers N* of incremental backup which minimizes it are analytically discussed when incremental backup interval T or full backup interval L is given. Finally, it is shown as examples that optimal numbers are numerically computed.

Index Terms—database, incremental backup, cumulative damage model

I. INTRODUCTION

In recent years, database in computer systems have become very important in the highly information-oriented society. The database is frequently updated by adding or deleting data files, and is stored in floppy disks or other secondary media. Even high reliable computers might sometimes break down eventually by several errors due to noises, human errors and hardware faults. It would be possible to replace hardware and software when they fail, but it would be impossible to do a database. One of important things for using computers is to backup data files regularly, *i.e.*, to copy all files in a secondary medium. Fukumoto et al. discussed optimal checkpoint generations for a database recovery mechanism [1].

Cumulative damage models in reliability theory, where a system suffers damage due to shocks and fails when the total amount of damage exceeds a failure level K, generate a cumulative process [2]. Some aspects of damage models from reliability viewpoints were discussed by Esary et al. [3]. It is of great interest that a system is replaced before failure as preventive maintenance. The replacement policies where a system is replaced before failure at time T [4,5], at shock N [6,7], or at damage Z [5,8,9] were considered. Nakagawa and Kijima [10] applied the periodic replacement with minimal repair [11] at failure to a cumulative damage

model and obtained optimal values $T*$, $N*$ and $Z*$ which minimize the expected cost. Satow et al. [12] applied the cumulative damage model to garbage collection policies for a database system. Qian et al. [5,13-17] successfully obtained the optimal Full and Cumulative Backup policies for a database by using cumulative damage models.

In this paper, we apply the cumulative damage model to the backup of files for database media failures, by putting *shock* by *update* and *damage* by *updated files*: a database is updated at a nonhomogeneous Poisson process and an amount of updated files accumulates additively. To lessen the overhead of backup processing, incremental backups with small overhead are adopted between full backups. The expected costs are derived, using the theory of cumulative processes. Further, an optimal number of incremental backup which minimizes the expected cost is analytically derived. Finally, it is shown as examples that optimal numbers are numerically computed when incremental backup interval or full backup interval is given.

II. FULL BACKUP AND INCREMENTAL BACKUP

Backup frequencies of a database would usually depend on the factors such as its size and availability, and sometimes frequency in use and criticality of data. The simplest and most indispensable method to ensure the safety of data would be always to shut down a database, and to make the backup copies of all data, log and control files in other places, and to take them out immediately when some date in the original secondary media are corrupted. This is called the *total backup*. But, such backup has to be made while a database is off-line and unavailable to its users, and would take more hours and costs as data become larger.

To overcome these disadvantages, *export backup* has been developed because only a small percentage changes in most applications between successive backups [14,18]. The total backup is a physical backup scheme which copies all files. On the other hand, export backup is a logical backup scheme which copies the data and the definition of the database where they are stored in the operating system in binary notation. The export backup

makes the copies of only files which have changed or are new since a prior backup. The resources required for such backup are proportional to the transactional activities which have taken place in a database, and not to its size. This can shorten backup times and can decrease the required resources, and would be more useful for larger databases. This approach is generally classified into three schemes: incremental backup, cumulative backup and full backup.

Full backup exports all files, and by this backup, the attributes of archives are updated, that is, a database system returns to an initial state. When full backup copies are made frequently, all images of a database can be secured, but its operation cost is greatly increased. Thus, the scheme of incremental or cumulative backup is usually adopted and is suitably executed between the operations of full backup.
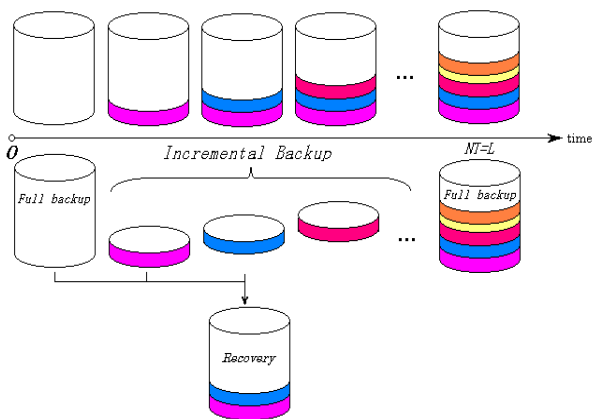

Figure 1. Full backup and incremental backups.

Incremental backup exports only files which have changed or are new since the last incremental or full backup, and updates the attributes of archives (see Figure 1). On the other hand, *cumulative backup* exports only modified files since the last full backup, however, does not update the attributes of archives. For most failures, a database can recover from these points by log files and restore a consistent state by importing files of all incremental backups and the full backup for the incremental backup scheme, and by importing files of the last cumulative backups and the full backup for the cumulative backup scheme. *i.e.*, the incremental backup or cumulative backup cannot take the place of full backup; however, it can reduce the frequency of full backup which is required. From the above point of view, we can reduce the frequency of full backup by incremental backups or cumulative backups.

An important problem in actual backup schemes is when to create the full backup. The full backup with large overhead is done at long intervals and the incremental or cumulative backup with small overhead is done at short intervals. We want to lessen the number of full backups with large overhead. However, the overhead of cumulative backup is increasing with the number of newly updated trucks, because the list of modified files is growing up each day until the next full backup which will

clear all attributes of archives. For the incremental backup scheme, the amount of data transfer is constant approximately. However, the overhead of recovery which imports files of all incremental and the full backups is remarkably increased with the number of incremental backups, when some errors have occurred in storage media. That is, both overheads of cumulative backup and recovery of incremental backup increase adaptively with the amount of newly updated trucks. From this point of view, we should decide the full backup interval, by comparing two overheads of backup and recovery.

Chandy, et al. and Reutr considered some recovery techniques for database failures [19,20]. The optimal checkpoint intervals of such models which minimize the total overhead were studied by many authors: Young [21], Gelenbe [22] and Dohi, et al. [23]. Further, Sandoh, et al. [24] discussed optimal backup policies for hard disks.

In this paper, we apply the cumulative damage model to backup policy for a database system with periodic incremental backup scheme and derive an optimal full backup interval and incremental backup interval. We have to pay only attention to the matter what are essential laws of governing systems, and take a grasp of updated processes and try to formulate it simply, avoiding small points. In other words, it would be necessary to form mathematical models of backup schemes which outline the observational and theoretical features of complex phenomena.

Suppose that a database in secondary media fails according to a general distribution $F(t)$ and full backup is performed when the database fails. The incremental backups are performed at periodic times iT ($i = 1, 2, \cdots, N-1$), and export only updated files which have changed or are new since the last full backup or incremental backup. In order to enhance reliability of the series system of incremental backups, full backup is performed at time $NT = L$.

Taking the above considerations into account, we formulate the following stochastic model of the backup policy for a database system: suppose that a database is updated at a nonhomogeneous Poisson process with an intensity function $\lambda(t)$ and a mean-value function $R(t)$, *i.e.*,

$$R(t) \equiv \int_0^t \lambda(u)\,\mathrm{d}u . \qquad (1)$$

Then, the probability that the j-th update occurs ($j = 0, 1, 2, \cdots$) exactly during $(u, v]$ is

$$H_j(u, v) \equiv \frac{\left[R(v) - R(u)\right]^j}{j!} e^{-[R(v) - R(u)]} \qquad (2)$$

where $R(0) \equiv 0$ and $R(\infty) \equiv \infty$.

Further, let Y_j denote an amount of files, which changes or is new at the j-th update. It is assumed that each Y_j has an identical probability distribution $G(x) \equiv \Pr\{Y_j \le x\}$ ($j = 1, 2, \cdots$). Then, the total amount

of updated files $Z_j \equiv \sum_{i=1}^{j} Y_i$ up to the j-th update where $Z_0 \equiv 0$ has a distribution [25,26]

$$\Pr\{Z_j \le x\} = G^{(j)}(x) \quad (j = 0,1,2,\cdots), \quad (3)$$

which is the j-fold Stieltjes convolution of the distribution $G(x)$ of itself, i.e.,

$$G^{(j)}(x) \equiv \int_0^x G^{(j-1)}(x-u)\,\mathrm{d}\,G(u) \quad (4)$$

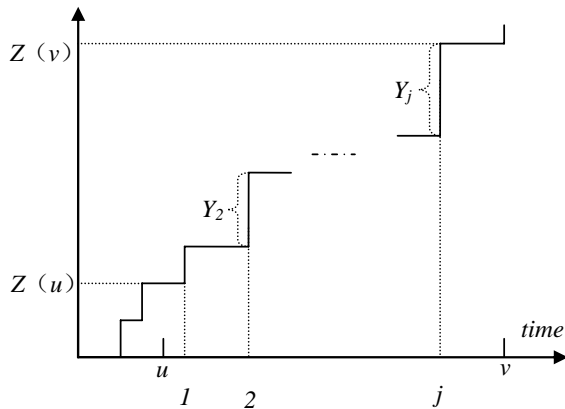with and $G^{(0)}(x) \equiv 1$ for $x \ge 0$, and $G^{(1)}(x) = G(x)$.

Figure 2. Process for database updated.

Let $Z(t)$ be the total amount of updated files at time t (see Figure 2). Then, the distribution of $Z(t)$ [3] is

$$\Pr\{Z(t) \le x\} = \sum_{j=0}^{\infty} H_j(t)G^{(j)}(x) \quad (5)$$

where $H_j(t) \equiv H_j(0,t)$, and

$$\Pr\{Z(v) - Z(u) \le x\} = \sum_{j=0}^{\infty} H_j(u,v)G^{(j)}(x). \quad (6)$$

III. OPTIMAL FULL BACKUP INTERVAL

A. Expected Cost

We discuss optimal full backup interval $L^*(T) = N^*T$ for fixed incremental backup periodic time T, e.g., daily or weekly(see Figure 3).

Let us introduce the following costs: cost c_F is suffered for the full backup; cost $c_D + c_0(x)$ is suffered for the incremental backup when the amount of export files at the backup time is x, and $c_0(x)$ is increasing with x; cost $c_{FF} + jc_{FD} + c_0(x)$ is suffered for recovery if database fails when the total amount of import files at the latest incremental backup time is x, where jc_{FD} denotes recovery cost of incremental backups when the number of incremental backups is j, c_{FF} denotes recovery cost of the last full backup.

Then, the expected cost of incremental backup, when it is performed at time iT $(i = 1, 2, \cdots, N-1)$, is

$$C_{ED}(i,T) \equiv \int_0^{\infty} [c_D + c_0(x)]\,\mathrm{d}\Pr\{Z(iT) - Z((i-1)T) \le x\},$$

and hence,

$$C_{ED}(i,T) \equiv \sum_{j=0}^{\infty} H_j((i-1)T,iT)$$
$$\times \int_0^{\infty} [c_D + c_0(x)]\,\mathrm{d}\,G^{(j)}(x). \quad (7)$$

Further, if the database fails during $[iT, (i+1)T]$, then the expected recovery cost till time iT is (see Figure 3)

$$C_{EF}(i,T) \equiv \sum_{j=0}^{\infty} H_j(iT)$$
$$\times \int_0^{\infty} [c_{FF} + ic_{FD} + c_0(x)]\,\mathrm{d}\,G^{(j)}(x). \quad (8)$$

We define time between two full backups is one cycle, then mean time to full backup is

$$E(N) \equiv NT\overline{F}(NT) + \int_0^{NT} t\,\mathrm{d}\,F(t)$$
$$= \int_0^{NT} \overline{F}(t)\,\mathrm{d}\,t. \quad (9)$$

From (7), the total expected backup cost of one cycle is

$$C_B(N) \equiv c_F + \overline{F}((N\text{-}1)T)\sum_{m=1}^{N-1} C_{ED}(m,T)$$
$$+ \sum_{i=2}^{N-1} [F(iT) - F((i-1)T)]\sum_{m=1}^{i-1} C_{ED}(m,T)$$
$$= c_F + \sum_{i=1}^{N-1} C_{ED}(i,T)\overline{F}(iT) \quad (10)$$

and from (8), the total expected recovery cost of one cycle is

$$C_R(N) \equiv \sum_{i=1}^{N} [F(iT) - F((i-1)T)]C_{EF}(i\text{-}1,T). \quad (11)$$

Therefore, the expected cost per unit time in the steady-state is

$$C(N) \equiv \frac{C_B(N) + C_R(N)}{E(N)}. \quad (12)$$

$\square$: updated point $\bigcirc$: incremental backup point
$\blacksquare$: failure point $Z(2T)$–$Z(T)$:amount of backup files
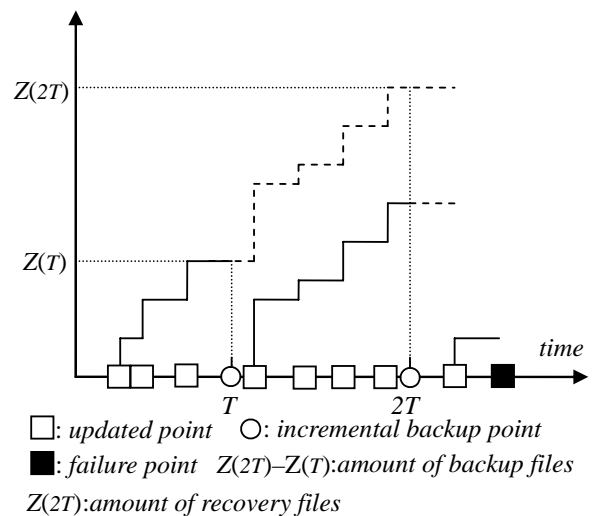$Z(2T)$:amount of recovery files

Figure 3. Periodic incremental backup model.

B. Optimal Policy

We discuss optimal value N* which minimize the expected cost $C(N)$ in (12).

Suppose that the database is updated at a homogeneous Poisson process with an intensity function $\lambda(t) = p\lambda$, and it fails according to distribution $F(t) = 1 - e^{-q\lambda t}$, where $p + q = 1$, further suppose that $E\{Y_j\} = 1/\mu$ and $c_0(x) = c_0 x$, then the expected cost in (12) is

$$\frac{C(N)}{q\lambda} = (c_{FF} - c_{FD} - \frac{c_0 p\lambda T}{\mu})$$

$$+ \frac{c_D + c_{FD} + \frac{2c_0 p\lambda T}{\mu}}{1 - e^{-q\lambda T}} + \tilde{C}(N) . \quad (13)$$

where

$$\tilde{C}(N) \equiv \frac{c_F - c_D - \frac{c_0 p\lambda T}{\mu}}{1 - e^{-q\lambda NT}}$$

$$- \frac{(c_{FD} + \frac{c_0 p\lambda T}{\mu})N e^{-q\lambda NT}}{1 - e^{-q\lambda NT}} . \quad (14)$$

From (13), we know that optimal $N*$ which minimizes the expected cost $C(N)$ is equality to optimal $N*$ which minimizes $\tilde{C}(N)$.

From (14), we have

$$\tilde{C}(I) = \frac{c_F - c_D - \frac{c_0 p\lambda T}{\mu} - (c_{FD} + \frac{c_0 p\lambda T}{\mu})e^{-q\lambda T}}{1 - e^{-q\lambda T}} , \quad (15)$$

$$\tilde{C}(\infty) = c_F - c_D - \frac{c_0 p\lambda T}{\mu} , \quad (16)$$

then exists a positive $N*$ ($I \leq N* \leq \infty$) which minimizes $\tilde{C}(N)$ in (14).

From the inequality $\tilde{C}(N+1) - \tilde{C}(N) \geq 0$, we have

$$Q(N+1) \geq \frac{c_F - c_D - \frac{c_0 p\lambda T}{\mu}}{c_{FD} + \frac{c_0 p\lambda T}{\mu}} \quad (17)$$

where

$$Q(N) \equiv N - \frac{1 - e^{-q\lambda NT}}{1 - e^{-q\lambda T}} . \quad (18)$$

Thus $Q(1) = 0$, $Q(N+1) - Q(N) = 1 - e^{-q\lambda NT} > 0$ and $\lim_{N \to \infty} Q(N) = \infty$, i.e., $Q(N)$ is strictly increasing with N.

Therefore, we have the following optimal policy:

1. If $c_F - c_D > (c_{FD} + \frac{c_0 p\lambda T}{\mu})(1 - e^{-q\lambda T}) + \frac{c_0 p\lambda T}{\mu}$, then there exists a finite and unique minimum $N*$ ($I < N* < \infty$) satisfying (17), minimizes $\tilde{C}(N)$, and

$$(c_{FD} + \frac{c_0 p\lambda T}{\mu})\frac{Q(N*) - N* e^{-q\lambda N*T}}{1 - e^{-q\lambda N*T}} < \tilde{C}(N*)$$

$$\leq (c_{FD} + \frac{c_0 p\lambda T}{\mu})\frac{Q(N*+1) - N* e^{-q\lambda N*T}}{1 - e^{-q\lambda N*T}} . \quad (19)$$

2. If $c_F - c_D \leq (c_{FD} + \frac{c_0 p\lambda T}{\mu})(1 - e^{-q\lambda T}) + \frac{c_0 p\lambda T}{\mu}$, then $N* = 1$, i.e., only full backup needs to be done, and the resulting cost is

$$\frac{C(1)}{q\lambda} = c_{FF} + \frac{c_F}{1 - e^{-q\lambda T}} . \quad (20)$$

C. Numerical Example

Suppose that $c_F/(c_0/\mu) = 2000$, $c_D/(c_0/\mu) = 40$, $c_{FF}/(c_0/\mu) = 2400$, and $c_{FD}/(c_0/\mu) = 50$. Then an $N*$ is given by a finite and unique minimum such that

$$N - \frac{1 - e^{-q\lambda T(N+1)}}{1 - e^{-q\lambda T}} \geq \frac{1910}{50 + p\lambda T} . \quad (21)$$

Table I gives the optimal number $N*$ and the resulting cost $C(N*)/(\lambda c_0/\mu)$ for $q = 10^{-2}$, 10^{-3}, 10^{-4} and $\lambda T = 200, 400, 800, 1000$. Note that all costs are relative to cost c_0/μ and all times are relative to $1/\lambda$. In this case, it is evident that $(50 + p\lambda T)(1 - e^{-q\lambda T}) + p\lambda T < 1960$ except for $q = 10^{-2}$ and $\lambda T = 1000$. It is shown from Table I that optimum number $N*$ decreases with both λT and q, and $C(N*)/q\lambda$ decreases with λT, and conversely, increases with q.

TABLE I.
OPTIMAL NUMBER $N*$ AND THE RESULTING COST
$C(N*)/(\lambda c_0/\mu)$ FOR λT AND q WHEN
$c_F/(c_0/\mu) = 2000$, $c_D/(c_0/\mu) = 40$,
$c_{FF}/(c_0/\mu) = 2400$, AND $c_{FD}/(c_0/\mu) = 50$

| λT | q | | | | | |
| | 10^{-2} | | 10^{-3} | | 10^{-4} | |
	$N*$	$\frac{C(N*)}{\lambda c_0/\mu}$	$N*$	$\frac{C(N*)}{\lambda c_0/\mu}$	$N*$	$\frac{C(N*)}{\lambda c_0/\mu}$
200	8	44.7607	12	6.4879	29	2.1670
400	3	44.1646	6	6.0945	14	1.9892
800	2	44.0056	3	5.6392	6	1.8295
1000	1	44.0009	2	5.4372	5	1.7658

IV. OPTIMAL INCREMENTAL BACKUP INTERVAL

A. Expected Cost

We discuss an optimal incremental backup interval $T*(L) = L/N*$ for fixed full backup periodic time L.

It is assumed that a database in secondary media fails according to a general distribution $F(t)$, and full backup is performed at time L or when the database fails, whichever occurs first. A database returns to an initial state by full backups. Incremental backups are performed at periodic times iT $(i = 1, 2, \cdots, N-1)$, where $NT = L$. In order to ensure point-in-time recovery after media failure, transaction log backups are made since the last full backup, its size depends on the amount of updated data in database.

Let us introduce the following costs: cost c_F is suffered for the full backup; cost c_D is suffered for one incremental backup or cost suffered for importing by one incremental backup because of exporting and importing data are the reverse process ; cost c_L is suffered for every transaction log backup; cost $c_R(x)$ is suffered for reconstructing data by transaction logs when the amount of updated data in database is x, and $c_{RO} + ic_{CD}$ is suffered for recovery if database fails, where ic_{CD} denotes importing cost of incremental backups when the number of incremental backups is i, c_{RO} denotes importing cost of the last full backup.

Then, the expected cost of incremental backups is

$$C_1 \equiv \sum_{i=1}^{N} [F(iT) - F((i-1)T)](i-1)c_D$$
$$+ \overline{F}(L)(N-1)c_D \qquad (22)$$

and the expected cost of transaction log backups is

$$C_2 \equiv \sum_{j=0}^{\infty} \int_0^L jH_j(t) \, c_L \, \mathrm{d}F(t)$$
$$+ \overline{F}(L) \sum_{j=0}^{\infty} jH_j(L)c_L . \qquad (23)$$

Further, if the database fails during $[(i-1)T, iT]$, then the expected recovery cost till time $(i-1)T$ is

$$C_3 \equiv \sum_{i=1}^{N} [F(iT) - F((i-1)T)][c_{RO} + (i-1)c_D] \qquad (24)$$

and the expected cost of reconstructing data by transaction logs till failure time is

$$C_4 \equiv \sum_{i=1}^{N} \sum_{j=0}^{\infty} \int_{(i-1)T}^{iT} H_j[(i-1)T, t] \mathrm{d}F(t)$$
$$\times \int_0^{\infty} c_R(x) \mathrm{d}G^{(j)}(x) \qquad (25)$$

Then the total expected backup and recovery cost during the interval L is

$$E(C) \equiv c_F + \sum_{i=1}^{4} C_i . \qquad (26)$$

B. Optimal Policy

To discuss optimal values $T^*(L) = L/N^*$ which minimizes expected cost $E(C)$ in (26) analytically , we suppose that the database is updated at a homogeneous Poisson process with an intensity function $\lambda(t) = p\lambda$, and it fails according to distribution $F(t) = 1 - e^{-q\lambda t}$,

where $p + q = 1$, further suppose that $E\{Y_j\} = 1/\mu$ and $c_R(x) = c_R x$, then the expected cost in (26) is

$$E(C) \equiv (c_F - c_D) + (\frac{p}{q}c_L + c_{RO} - c_D)(1 - e^{-q\lambda T})$$
$$+ p\lambda L c_L + C(N) \qquad (27)$$

where

$$C(N) \equiv \frac{2c_D(1 - e^{-q\lambda L})}{1 - e^{-\frac{q\lambda L}{N}}} - Nc_D e^{-q\lambda L}$$

$$+ \frac{pNc_R(1 - e^{-\frac{q\lambda L}{N}})}{q\mu} - \frac{c_R p\lambda L}{\mu} e^{-\frac{q\lambda L}{N}} \qquad (28)$$

From (27), we know that optimal N^* which minimizes the expected cost $E(C)$ is equality to optimal N^* which minimizes $C(N)$.

From (28), we know that
$$C(\infty) \equiv \lim_{N \to \infty} C(N) = \infty , \qquad (29)$$

$$C(1) = (c_D + \frac{pc_R}{q\mu} + \frac{p\lambda L c_R}{\mu})(1 - e^{-q\lambda L})$$

$$+ c_D - \frac{p\lambda L c_R}{\mu} , \qquad (30)$$

then there exists a finite number N^* $(1 \leq N^* < \infty)$ which minimizes $C(N)$.

From the inequality $C(N+1) - C(N) \geq 0$, we have

$$2c_D(1 - e^{-q\lambda L}) \frac{e^{-\frac{q\lambda L}{N+1}} - e^{-\frac{q\lambda L}{N}}}{(1 - e^{-\frac{q\lambda L}{N}})(1 - e^{-\frac{q\lambda L}{N+1}})}$$

$$- \frac{pc_R}{q\mu} e^{-\frac{q\lambda L}{N+1}} - \frac{pc_R}{q\mu}(N + q\lambda L)(e^{-\frac{q\lambda L}{N+1}} - -e^{-\frac{q\lambda L}{N}})$$

$$\geq c_D e^{-q\lambda L} - \frac{pc_R}{q\mu} . \qquad (31)$$

It is very difficulty to discuss optimal analytic values N^* accurately. So we seek analytic values $\tilde{N}^*$ approximately. Optimal accurate numerical values will be given in numerical example and be compared with the approximate numerical values; error analysis will be given at last.

Using two approximations that [27]
$$e^{-q\lambda L} \approx 1 - q\lambda L , \qquad (32)$$
and

$$e^{-\frac{q\lambda L}{N+1}} - e^{-\frac{q\lambda L}{N}} \approx \frac{q\lambda L}{N} - \frac{q\lambda L}{N+1} . \qquad (33)$$

thus (28) can be written

$$\tilde{C}(N) \equiv Nc_D(1 + q\lambda L) + \frac{c_R pq(\lambda L)^2}{N\mu} \qquad (34)$$

and the inequality (31) is simplified as

$$Q(\tilde{N}) \geq \frac{pq\lambda^2 L^2 c_R}{\mu(1 + q\lambda L)c_D} \qquad (35)$$

where

$$Q(\tilde{N}) \equiv \tilde{N}(1 + \tilde{N}) . \qquad (36)$$

Then $Q(1) = 2$, $\lim_{\tilde{N}\to\infty} Q(\tilde{N}) = \infty$ and $Q(\tilde{N})$ is strictly increasing with $\tilde{N}$.

Therefore, we have the following optimal policy:

1. If $\dfrac{pq\lambda^2 L^2 c_R}{\mu(1+q\lambda L)c_D} > 2$, there exists a unique optimal $\tilde{N}^*$ ($1 < \tilde{N}^* < \infty$) satisfies (35) which minimizes $\tilde{C}(\tilde{N})$ in (34), and

$$2N - 1 < \frac{\tilde{C}(\tilde{N}^*)}{c_D(1+q\lambda L)} \le 2N + 1 . \qquad (37)$$

2. If $\dfrac{pq\lambda^2 L^2 c_R}{\mu(1+q\lambda L)c_D} \le 2$, then $\tilde{N}^* = 1$, i.e., only backup and log back need to be done, and the resulting cost is

$$\tilde{C}(\tilde{N}^*) = \tilde{C}(1) . \qquad (38)$$

C. Numerical Example

Suppose that $1/\mu = 1$. Table II gives the optimal approximate numerical values $\tilde{N}^*$, optimal accurate numerical values N^* and the resulting costs $C(N^*)$, $C(\tilde{N}^*)$, $\tilde{C}(N^*)$ and $\tilde{C}(\tilde{N}^*)$ for $\lambda L = 1000$, 2000, 3000, 4000 when $c_D = 100$, $c_R = 10$ and $q = 10^{-4}$.

TABLE II.
OPTIMAL NUMBER N^* , $\tilde{N}^*$ AND THE RESULTING COSTS $C(N^*)$, $C(\tilde{N}^*)$, $\tilde{C}(N^*)$ AND $\tilde{C}(\tilde{N}^*)$ FOR λL WHEN $c_D = 100$, $c_R = 10$ AND $q = 10^{-4}$

λL	1000	2000	3000	4000
N^*	2	4	7	9
$\tilde{N}^*$	3	6	8	11
$C(N^*)$	451.075	899.456	1341.765	1776.383
$C(\tilde{N}^*)$	472.086	940.590	1364.307	1818.864
$\tilde{C}(N^*)$	719.950	1479.900	2195.586	3037.600
$\tilde{C}(\tilde{N}^*)$	663.300	1386.600	2164.888	2994.400

Similarly, Table III gives optimal approximate numerical values $\tilde{N}^*$, optimal accurate numerical values N^* and the resulting cost $C(N^*)$ for $q = 10^{-4}$, 2×10^{-4} , 2.5×10^{-4} , 3×10^{-4} , $c_R = 5$, 10, 15 and $c_D = 100$, 200, 300 when $\lambda L = 4000$.

These indicate that both optimal numbers N^* and the expected costs $C(N^*)$ are increasing with λL , c_R and q . Although optimal numbers N^* are decreasing when c_D is increasing, the expected costs $C(N^*)$ are increasing.

The reason would be explained that the optimal incremental backup numbers N^* during time interval L are related to both costs suffered for incremental backups

TABLE III.
OPTIMAL NUMBER N^* , $\tilde{N}^*$ AND THE RESULTING COST $C(N^*)$, FOR c_R , c_D AND q WHEN $\lambda L = 4000$

q	c_R	c_D	N^*	$\tilde{N}^*$	$C(N^*)$
10^{-4}	5	100	6	8	1257.883
		200	4	5	1785.173
		300	3	4	2201.787
	10	100	9	11	1776.383
		200	6	8	2515.766
		300	5	6	3084.363
	15	100	11	13	2173.769
		200	8	9	3082.193
		300	6	8	3773.649
2×10^{-4}	5	100	9	9	1728.258
		200	6	7	2445.691
		300	5	5	2999.491
	10	100	13	13	2442.371
		200	9	9	3456.516
		300	7	8	4234.162
	15	100	16	16	2989.919
		200	11	12	4230.083
		300	9	9	5184.774
2.5×10^{-4}	5	100	10	10	1896.161
		200	7	7	2683.551
		300	5	6	3292.367
	10	100	14	14	2680.863
		200	10	10	3792.322
		300	8	8	4645.394
	15	100	18	17	3283.052
		200	12	12	4644.213
		300	10	10	5688.484
3×10^{-4}	5	100	11	10	2035.400
		200	8	7	2882.528
		300	6	6	3522.677
	10	100	19	15	2878.965
		200	11	10	4070.801
		300	9	9	4956.040
	15	100	20	18	3526.619
		200	14	13	4987.885
		300	11	10	6106.201

and costs suffered for reconstructing data by transaction logs:

1. Incremental backup costs are increasing with $N*$ and Log Backup costs are decreasing with $N*$, So if we want to minimize expected costs $C(N)$, incremental backup numbers $N*$ should be decreasing with incremental backup cost c_D;

2. on the one hand, costs suffered for reconstructing data by transaction logs are related to c_R and q, when c_R and q are increasing, reconstructing data costs are increasing if $N*$ is constant, so if we want to minimize expected costs $C(N)$, incremental backup numbers $N*$ should be increasing;

3. λL denotes the frequency of updating of database, if λL increases, the probability of database failure increases, so reconstructing data costs will increase, optimal incremental backup numbers $N*$ should be decreased to minimize the excepted costs.

Table II and Table III also give approximate numerical values $\tilde{N}*$ using (35). Compared with optimal accurate numerical values $N*$, it is evident that $\tilde{N}*$ and $N*$ are not identical. It is interesting that $\tilde{N}*$ are almost the same with $N*$ when $q = 2.5 \times 10^{-4}$; when $q < 2.5 \times 10^{-4}$, $\tilde{N}*$ are greater than or equal to $N*$, *i.e.*, values of $N*$ may be $\tilde{N}*$, $\tilde{N}*-1$ or $\tilde{N}*-2$; when $q > 2.5 \times 10^{-4}$, $\tilde{N}*$ are less than or equal to $N*$, *i.e.*, values of $N*$ may be $\tilde{N}*$, $\tilde{N}*+1$ or $\tilde{N}*+2$. So the approximate numbers $N*$ would be useful for seeking accurate numbers $N*$.

V. CONCLUSIONS

We have considered two schemes of full and incremental backup for a database system, and have analytically discussed optimal backup policies which minimize the expected cost, using theory of cumulative processes. It would be of interest that there must be optimal number of incremental backups minimizing the expected cost rate. This result would be applied to the backup of a database, by estimating the costs of two backups, recovery, the amount of updated file from actual data and the probability of database failure. However, backup schemes become very important and much complicated, as database systems have been largely used in most computer systems and information technologies have been greatly developed. These formulations and techniques used in this paper would be useful and helpful for analyzing such backup policies.

ACKNOWLEDGMENT

REFERENCES

[1] S. Fukumoto, N. Kaio and S. Osaki, "A study of checkpoint generations for a database recovery mechanism," *Comput Math. Appl.*, vol.24, pp. 63-68, 1992.

[2] D. R. Cox, *Renewal Theory*, London: Methuen, 1962.

[3] J. D. Esary, A. W. Marshall and F. Proschan, "Shock models and wear processes," *Annals of Prob.*, vol.1, pp. 627-649, 1973.

[4] H. M. Taylor, "Optimal replacement under additive damage and other failure models," *Naval Res. Logist. Quart.*, vol.22, pp. 1-18, 1975.

[5] C. Qian, K. Ito and T. Nakagawa, "Optimal preventive maintenance policies for a shock model with given damage level," *Journal of Quality in Maintenance Engineering*，vol.11, pp. 216-227, 2005.

[6] T. Nakagawa, "A summary of discrete replacement policies," *Euro. J. of Oper. Res.*, vol.17, pp. 382-392, 1984.

[7] C. Qian, S. Nakamura and T. Nakagawa, "Replacement and Minimal Repair Policies for a Cumulative Damage Model with Maintenance," *Computers and Mathematics with Applications*, vol.10, pp. 1111-1118, 2003.

[8] R. M. Feldman, "Optimal replacement with semi-Markov shock models," *J. Applied Prob.*, vol.13, pp.108-117 , 1976.

[9] T. Nakagawa, "On a replacement problem of a cumulative damage model," *Oper. Res. Quart,* vol.27, pp.895-900, 1976.

[10] T. Nakagawa and M. Kijima, "Replacement policies for a cumulative damage model with minimal repair at failure," *IEEE Trans. Reliability*, vol. 13, pp. 581-584, 1989.

[11] R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, New York : John Wiley & Sons, 1965.

[12] T. Satow, K. Yasui and T. Nakagawa, "Optimal garbage collection policies for a database in a computer system," *RAIRO-Oper. Res.*, vol.30, pp. 359-372, 1996.

[13] C. Qian, S. Nakamura and T. Nakagawa, "Cumulative Damage Model with Two Kinds of Shocks and Its Application to the Backup Policy," *J. Oper. Res. Soc. Japan*, vol. 42, pp.501-511, 1999.

[14] C. Qian, S. Nakamura and T. Nakagawa, "Optimal Backup Policies for a Database System with Incremental Backup," *Electronics and Communications in Japan*, Part 3, Vol. 85, pp.1-9, 2002.

[15] C. Qian, Y. Pan, T. Nakagawa, "Optimal policies for a database system with two backup schemes," *RAIRO Oper. Res.*, vol.36, pp.227-235, 2002.

[16] S. Nakamura, C. Qian, S. Fukumoto and T. Nakagawa, "Optimal backup policy for a database system with incremental and full backup," *Mathematical and computer modelling*, vol.11, pp. 1373-1379, 2003.

[17] T. Nakagawa, *Shock and Damage Models in Reliability Theory*, London: Springer, 2007.

[18] K. Suzuki and K. Nakajima, "Storage management software," *Fujitsu*, vol.46, pp.389-397, 1995.

[19] K. M. Chandy, J. C. Browne, C. W. Dissly and W. R. Uhrig, "Analytic models for rollback and recovery strategies in data base systems," *IEEE Trans. Software Engineering*, SE-1, pp. 100-110, 1975.

[20] A. Reuter, "Performance analysis of recovery techniques," *ACM Trans. Data. Sys.*, vol.4, pp.526-559, 1984.

[21] J. W. Young, "A first order approximation to the optimum checkpoint interval," *Communications of ACM*, vol. 17, pp. 530-531, 1974.

[22] E. Gelenbe, "On the optimum checkpoint interval," *J. Association on Computing Machinery*, vol. 26, pp. 259-270, 1979.

[23] T. Dohi, T. Aoki, N. Kaio and S. Osaki, "Computational aspects of optimal checkpoint strategy in fault-tolerant database management," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E80-A, pp. 2006-2015, 1997.

[24] H. Sandoh, N. Kaio and H. Kawai, "On backup policies for computer disks," *Reliability Engineering & System Safety*, vol. 37, pp. 29-32, 1992.

[25] T. Nakagawa, *Maintenance Theory of Reliability*, London: Springer, 2005.

[26] T. Nakagawa, *Advanced Reliability Models and Maintenance Policies*, London: Springer, 2008.

[27] T. Nakagawa, K. Yasui and H. Sandoh, "Note on optimal partition problems in reliability models," *Journal of Quality in Maintenance Engineering*, vol.10, pp.282-287, 2004.

Cunhua Qian is a Professor at College of Management Science and Engineering, Nanjing University of Technology, China. He received his BS degree in Mathematics from Yangzhou University, Yangzhou, China, in 1984, MS degree in Industrial Engineering in 1998 and Ph.D. degree in reliability in 2001, both from Aichi Institute of Technology, Toyota, Japan. He is currently interested in stochastic processes and reliability. His papers have been published in *Journal of the Operations Research Society of Japan*, *RAIRO Operations Research*, *Computers and Mathematics with Applications*, *Journal of Quality in Maintenance Engineering* and other journals. Prof. Qian is a member of ORSJ.

Yingyan Huang is a graduate student at College of Management Science and Engineering, Nanjing University of Technology, China. She received her BS degree in Industrial Engineering from Nanjing University of Technology in 2008.

Xufeng Zhao is a graduate student at Department of Marketing and Information System, Aichi Institute of Technology, Toyota, Japan. He received his BS degree in Information Engineering in 2006 and MS degree in System Engineering in 2009 both from Nanjing University of Technology, China.

Toshio Nakagawa is a Professor at Department of Marketing and Information System, Aichi Institute of Technology, Toyota, Japan. He received his B.S.E. and M.S. degrees from Nagoya Institute of Technology in 1965 and 1967, and a Ph.D. degree from Kyoto University in 1977. His research interests are optimization, failure-time distributions, and computer systems in reliability theory. His papers have been published in *IEEE Trans. Reliability*, *European Journal of Operational Research*, *Naval Research Logistics Quart.,* and other journals. His books *Maintenance Theory of Reliability* (2005), *Shock and Damage Models in Reliability Theory*(2007) and *Advanced Reliability Models and Maintenance Policies*(2008) have been published by London: Springer. Prof. Nakagawa is a member of ORSJ, IEICE, IPSJ, and others.