

# 時間・ジョブ管理型バックアップ政策

河合 一 (鳥取大学工学部)

三道 弘明 (流通科学大学情報学部)

あらまし ハードディスクは、ワークステーションでは言うまでもなく、パーソナルコンピュータにおいても不可欠なシステム構成要素として位置付けられるようになってきた。ハードディスクには、様々なファイルが保存されており、日々その内容が更新されている。

このようなファイルが、人間の誤操作やシステムを構成するハードウェアの故障により、一瞬にして消失してしまうことも少なくない。こうした障害による被害をできる限り小さくする最も単純な方法として、ハードディスクの内容をフロッピーディスクなどに定期的にバックアップするという方法がある。しかし、バックアップ実施時期の決定に関しては、各ユーザーの勘と経験に委ねられているのが現状である。

このようなバックアップ時期決定問題に対し、本研究では、バックアップ完了時点あるいはディスク破損からの復旧終了時点から計測して一定期間  $T$  でバックアップの必要性を警告し、そのとき処理中のジョブがその処理を終了した時点でバックアップを行うという政策を提案する。更に、単位時間あたり期待費用を最小にするという意味で最適バックアップ警告時期を決定する。

## 1 まえがき

ワークステーションやパーソナルコンピュータで用いられるハードディスクは、その記憶容量が年々大きくなるとともに、低価格下が進んでいる。更に、パーソナルコンピュータ用のアプリケーションソフトウェアの中にも、ハードディスクの使用を前提としたものが次々に開発されてきている。このような状況を反映して、ワークステーションは言うまでもなく、パーソナルコンピュータにおいても、ハードディスクが不可欠なシステム構成要素として位置付けられるようになってきた。

一般に、ハードディスクには、種々のファイルが保存されており、多くのファイルの内容が日々のように更新されている。このようなファイルが、ユーザの誤操作あるいはシステムを構成するハードウェアなどの故障により、瞬時に消失してしまうことも少なくない。以下ではこのような障害をディスクの破損と呼ぶこととする。こうした障害に対処するためには、障害が発生しても速やかに正常な状態に復旧できるような方法が必要であり、通常は次のような方法が採られている。それは、ハードディスクの内容を定期的にフロッピーディスクなどの2次記憶装置にバックアップするという最も単純な方法である。しかしながら、バックアップの実施時期は、各ユーザの勘と経験によって決定されているのが現状であり、その妥当性は明らかにされていない。

バックアップの実施時期を決定する問題に対しては、これまで著者らは次のような政策を提案している。

- 1) バックアップ終了後あるいはハードディスク破損からの復旧完了後から計測し、 $T$  時間後にバックアップを行うという政策<sup>(1)</sup>。

- 2) ハードディスクの内容を変更・更新するジョブが  $N$  個終了した時点でバックアップを行うという政策<sup>(2)</sup>.
- 3) フロッピーディスクのように記憶容量の小さい場合には、全記憶容量の  $1/N$  を消費した時点でバックアップを実施するという政策<sup>(3)</sup>.

しかしながら、上の 1), 3) の下では処理中のジョブを中断してまでもバックアップを行うことになる。また、2) の場合には、バックアップ時期が不定期となり、ジョブ処理時間の分布によっては、そのばらつきが大きくなることから管理上不都合が生じる。

本研究では、このような問題を解消することを目的として、次のようなバックアップ政策を提案する。それは、バックアップ終了時点あるいはハードディスク破損からの復旧完了時点から計測して、一定時間  $T$  (以下、これをバックアップ警告時間と呼ぶ) に到達した時点でバックアップの必要性を促す警告を発し、そのとき丁度処理中のジョブ (警告ジョブ) がその処理を終えた時点でバックアップを行うという、ジョブを考慮した時間計画型のバックアップ政策である。

本政策の下では、ジョブ処理時間を確率変数とすると、バックアップ実施時点までに処理されるジョブ数も確率変数となる。しかし、バックアップ警告時刻  $T$  以前にはバックアップを実施しないので、バックアップ時期のばらつきは、前述の 2) の方法のそれよりも小さくなるのが十分に期待できる。また、バックアップ警告時刻  $T$  に到達した時点で処理中の警告ジョブがその処理を終えるまでバックアップを行わないことから、1), 3) の方法が有する前述の問題点も解消される。これらの点で、本バックアップ政策は、これまでに提案されたものと比較してより現実的であると言える。

以下では、上述のような政策の下で、バックアップに固定的に必要な費用と、ディスク破損時にファイルを消滅することによる損失費用を考慮した単位時間あたり期待費用を導出する。更に、これを最小にするという意味での最適バックアップ警告時期の存在について解析を行う。

## 2 問題の設定

前にも述べたとおり、本研究では各ジョブの処理時間の累積がバックアップ警告時間  $T$  に達したときに、まさに処理中の警告ジョブがその処理を完了した時点でバックアップを行うという政策を考える。このような政策の下では、バックアップ警告時間  $T$  が決定変数となる。但し、ディスクの破損時には次のような処理を行うこととする。すなわち、前回のバックアップ時点までの内容はバックアップディスクに保存されているので、バックアップディスクの内容をハードディスクにコピーすることで部分的に復旧する。復旧が部分的であるのは、前回のバックアップ時点以降に変更・更新した部分の内容は復旧不能であることによる。

また、以下では、各バックアップ時点において、前回のバックアップ時点以降に変更・更新された部分のみをバックアップすることを考える。

なお、本研究では以下の仮定をし、以下の記号を定義する。

(仮定)

- (i) 各ジョブの処理時間は、互いに独立で同一の分布に従う。さらに処理時間は、密度関数をもつ。
- (ii) ハードディスクの破損は希にしか起こらないと考え、ディスクの破損時間間隔は指数分布に従う。
- (iii) ハードディスクの破損は即座に検出され、復旧時間は無視できる。
- (iv) バックアップに要する時間は無視しうる。

(記号)

$T$  : バックアップ警告時間

$\lambda$  : ハードディスク破損に対する故障率

$a$  : バックアップ1回あたり固定費用

$b$  : 単位時間あたりに変更・更新した情報を消去した場合の損失費用

$H(t)$  : 各ジョブの処理時間分布

$h(t)$  :  $H(t)$  の密度関数 ( $= dH(t)/dt$ )

$H_n(t)$  :  $n$  個のジョブの処理時間分布

$h_n(t)$  :  $H_n(t)$  の密度関数 ( $= dH_n(t)/dt$ )

なお、 $H_n(t)$  は、漸化的に次式で与えられる。

$$H_1(t) = H(t), H_n(t) = \int_0^t H_{n-1}(t-x)h(x)dx \quad (1)$$

### 3 時間平均費用の導出

ここでは、無限期間における単位時間当たり期待費用、すなわち時間平均費用を最小にするという意味での最適な  $T$  を求めることを目的として、時間平均費用を導出する。

#### 3.1 プロセスの振舞

仮定 (iii) より、プロセスを以下の3通りの場合のいずれか1つが生起した時点を再生点とする再生報酬過程<sup>(4)</sup>とみなすことにより、期待費用を導出する事ができる。

- (a) 警告ジョブがその処理を終了し、バックアップも終了した時点。

- (b) バックアップ警告が発せられたものの、警告ジョブが終了する前にディスクが破損し、バックアップディスクによる復旧が完了した時点。
- (c) バックアップ警告が発せられる前にディスクが破損し、バックアップディスクによる復旧が完了した時点。

上述の (a),(b),(c) 3 通りのそれぞれの場合の費用は以下のようなになる。

- (a) の場合：バックアップに必要な固定費用  $a$  のみが生じる。
- (b) の場合：ハードディスクが破損するまでの時間を  $y$  とすると  $by$  という費用が必要となる。
- (c) の場合：(b) と同様、ハードディスクが破損するまでの時間を  $y$  とすると費用  $by$  が発生する。

### 3.2 時間平均費用

時間平均費用を  $C(T)$  で表わし、 $A(T)$ ,  $B(T)$  をそれぞれ、引き続き再生点間における期待時間と期待費用とすると、 $C(T)$  は次式で与えられる<sup>(4)</sup>。

$$C(T) = \frac{B(T)}{A(T)} \quad (2)$$

仮定 (i) ~ (iv) より、 $A(T)$ ,  $B(T)$  は以下のように与えられる。

$$\begin{aligned} A(T) &= \int_T^\infty x e^{-\lambda x} h(x) dx \\ &+ \int_0^T \int_T^\infty y e^{-\lambda y} h(y-x) m(x) dx dy \\ &+ \int_T^\infty y \lambda e^{-\lambda y} \bar{H}(y) dy \\ &+ \int_0^T \int_T^\infty y \lambda e^{-\lambda y} \bar{H}(y-x) m(x) dx dy \\ &+ \int_0^T y \lambda e^{-\lambda y} dy \end{aligned} \quad (3)$$

$$\begin{aligned} B(T) &= a \int_T^\infty e^{-\lambda x} h(x) dx \\ &+ a \int_0^T \int_T^\infty e^{-\lambda y} h(y-x) m(x) dx dy \\ &+ b \int_T^\infty y \lambda e^{-\lambda y} \bar{H}(y) dy \\ &+ b \int_0^T \int_T^\infty y \lambda e^{-\lambda y} \bar{H}(y-x) m(x) dx dy \\ &+ b \int_0^T y \lambda e^{-\lambda y} dy \end{aligned} \quad (4)$$

但し

$$m(t) \equiv \sum_{n=1}^{\infty} h_n(t) \quad (5)$$

であり、 $m(t)dt$  は時刻  $t$  と  $t+dt$  の間で、何番目のジョブかを問わず丁度ジョブの処理が終了する確率を表す。ここで式 (3), (4) において、右辺の第 1,2 項は節 3.1 の (a) に、第 3,4 項は (b) に、第 5 項は (c) にそれぞれ対応している。なお式 (3), (4) を整理した結果は次のようになる。

$$A(T) = p[1 + \int_0^T e^{-\lambda t} m(t) dt] \quad (6)$$

$$\begin{aligned} B(T) &= a(1 - \lambda p) + b\lambda Q \\ &+ \lambda \int_0^T [bpt + bQ - ap] e^{-\lambda t} m(t) dt \end{aligned} \quad (7)$$

ここで

$$p = \int_0^{\infty} e^{-\lambda x} \bar{H}(x) dx \quad (> 0) \quad (8)$$

$$Q = \int_0^{\infty} x e^{-\lambda x} \bar{H}(x) dx \quad (> 0) \quad (9)$$

である。

式 (2), (6), (7) より、 $C(T)$  は次式で与えられる。

$$\begin{aligned} C(T) &= \frac{a(1 - \lambda p) + b\lambda Q + \lambda \int_0^T [bpt + bQ - ap] e^{-\lambda t} m(t) dt}{p[1 + \int_0^T e^{-\lambda t} m(t) dt]} \end{aligned} \quad (10)$$

## 4 最適バックアップ政策

本節では、前節で導出した式 (10) の時間平均費用  $C(T)$  を最小にする  $T$  の存在について分析する。

まず式 (2) を  $T$  に関して微分すると

$$C'(T) = \left[ \frac{B(T)}{A(T)} \right]' = \frac{B'(T)A(T) - A'(T)B(T)}{A(T)^2} \quad (11)$$

が得られる。式 (11) において、 $A(T)^2 > 0$  であるから、分子の  $B'(T)A(T) - A'(T)B(T)$  の符号を調べれば  $C'(T)$  の符号がわかる。さらに、

$$\begin{aligned} &B'(T)A(T) - A'(T)B(T) \\ &= A'(T) \left[ \frac{B'(T)}{A'(T)} A(T) - B(T) \right] \end{aligned} \quad (12)$$

と書くことができる。ここで  $A'(T)$  を直接計算すると

$$A'(T) = pe^{-\lambda T} m(T) > 0 \quad (13)$$

となる。そこで式 (12), (13) に着目して,  $D(T)$  を次のように定義する。

$$D(T) = \frac{B'(T)}{A'(T)} A(T) - B(T) \quad (14)$$

$C'(T)$  の符号を調べるには  $D(T)$  の符号について調べればよい。  $D(T)$  を  $T$  に関して微分すると

$$D'(T) = \left[ \frac{B'(T)}{A'(T)} \right]' A(T) \quad (15)$$

となる。さらに,  $A(T) > 0$  であり

$$\left[ \frac{B'(T)}{A'(T)} \right]' = \lambda b > 0$$

が成立することから,  $D'(T)$  は正の値をとることがわかる。

一方,

$$\lim_{T \rightarrow \infty} D(T) = +\infty \quad (16)$$

$$\lim_{T \rightarrow 0} D(T) = -a < 0 \quad (17)$$

が成立する。以上より,  $D(T)$  の符号は唯 1 回変化し, さらに負から正へ変化する。このことは  $C(T)$  は下に単峰であることを意味し, したがって,  $C(T)$  を最小にする  $T$  が唯一存在することがわかる。

## 参考文献

- [1] 三道弘明, 海生直人, 河合 一: “フロッピーディスク・ハードディスクの最適バックアップ政策” 信学論 (D-I), **J73-D-I**, 3, pp. 336-341 (1990-03).
- [2] 三道弘明・河合 一, “ハードディスクに対するジョブを考慮した最適バックアップ政策”: 信学論 (A), **J73-A**, 9, pp. 1496-1500 (1990-09).
- [3] 三道弘明・河合一, “ワードプロセッサ用文書フロッピーディスクに対する最適  $1/N$  バックアップ政策”: 信学論 (D-I), **J73-D-I**, 11, pp. 900-904 (1990-11).
- [4] S. M. Ross: “Applied Probability Models with Optimization Applications”, Holden-Day, San-Francisco, California (1970).