最適チェックポインティング方策における システム障害の影響について*

小田切政徳**・土肥 正**・海生 直人***・尾崎 俊治**

On the Effect of System Failures in Optimal Checkpointing Policies*

Masanori ODAGIRI**, Tadashi DOHI**, Naoto KAIIO*** and Shunji OSAKI**

In a file system, checkpointing is inevitably necessary to recover the system once the system failure occurs. If the checkpointing is frequently carried out, the cost for generating checkpoints increases. On the other hand, if the number of checkpointing is small, the recovery cost via the journal increases when the system failure occurs. Thus, it is important to determine theoretically the optimal checkpoint sequence which minimizes the expected cost. In this paper we consider the situation in which all data files in a secondary storage can be destroyed by the media failure occurring at the checkpoint, and derive the optimal checkpointing policy. The fault tolerant design by taking such a catastrophic media failure into account is of great use to achieve the high reliability from the standpoint of risk avoidance.

1. はじめに

近年、コンピュータシステムは大規模化・複雑化の一途をたどっており、我々が日常生活で利用する機会も多くなっている。ゆえに、コンピュータシステムがひとたびシステムダウンを引き起こすと、それが社会生活に与える影響は非常に大きなものとなる。したがって、コンピュータシステムのシステムダウンによる影響を低減させるために、様々な方法を用いて高信頼性を実現する必要がある。

このような目的に対して、従来からいかに信頼性の高い部品を作るか、いかにバグのないソフトウェアを作るかというフォールトアボイダンス技術が研究されてきた。しかし、障害の発生しないハードウェアを作ることや、バグのないソフトウェアを作るとは極めて困難である。そこで、故障やバグがあったとしても、それがシステム全

体に重大な影響を与えないよう制御・設計を行うフォールトトレランス技術が注目を集めている¹⁾。

一般にファイル系システムにおいて、データが障害発生により失われてしまうことは、システム管理者にとって非常に大きな損失となる。よって、データベースに代表されるファイル系システムにおけるデータの回復技術をフォールトトレランスの観点から取り扱うことは、非常に有効であると考えられる。通常、ファイル系システムにおいて考えられる障害の種類として、主記憶上の内容を損なうシステム障害と、何らかの二次記憶媒体をハード的に損なう媒体障害とに分類される。システム障害の発生により主記憶上のデータが破壊されてしまった際に、障害が発生する直前の状態までデータを回復させる代表的な手法の一つにロールバック・リカバリがある。一般に、データベースシステムなどにおいて、到着するトランザクションの要求によりデータの変更が成されるとき、その処理に必要な情報はすべてジャーナルと呼ばれる記憶媒体に記録される。もしシステム障害の発生により主記憶上のデータが失われたときには、過去にやりとりされたデータの履歴を主記憶上に呼び込み、ジャーナルに記録されたトランザクションの処理に必要な情報を利用することにより、障害発生直前の状態までシステムをリカバリすることが可能となる。

* 原稿受付 1995年3月13日

** 広島大学 工学部 Department of Industrial and Systems Engineering, Hiroshima University; 4-1 Kagamiyama 1-chome, Higashi-Hiroshima city, Hiroshima 739, JAPAN

*** 広島修道大学 商学部 Department of Management Science, Hiroshima Shudo University; 1717 Ohtsuka, Numata-cho, Asaminami ward, Hiroshima city, Hiroshima 731-31, JAPAN

Key Words: checkpointing, fault tolerance, catastrophic failure, file system, stochastic modeling.

しかしながら、ジャーナルだけを用いた回復動作は、手間がかかるだけでなく多くの費用がかかることから決して得策であるとは言い難い。そこで、主記憶上のデータを安定な二次記憶媒体にあらかじめ定められた時点ごとに保存するチェックポイント技法^{1)~9)}との併用により、ジャーナルを使用する場合のリカバリにかかる費用を抑えることができると考えられている。チェックポイントの生成をあまり行わない場合、チェックポイント自体にかかる費用は少なくてすむが、システム障害が発生したときにジャーナルによるリカバリにかかる費用が大きくなる。これに対して、チェックポイントの生成を頻繁に行う場合、システム障害が発生したときにジャーナルによるリカバリにかかる費用は少なくてすむが、チェックポイント自体にかかる費用が多くなることは自明である。そこで、チェックポイントの生成頻度もしくは時機をいかに決定するかが重要な問題となる。

初期の文献^{2),3)}では、システム障害の発生がポアソン過程を形成するという仮定のもとで、システムのアベイラビリティを最大にする最適チェックポイント問題について議論している。福本⁴⁾は、システム障害発生の時間分布に一般分布を仮定し、単位時間当たりの最適チェックポイント回数を表す濃度関数を導入することにより、定常状態における単位時間当たりの近似期待費用を最小にするチェックポイントスケジュールを求めている。また、これらの応用として、データベースシステム内でリカバリにかかる時間がトランザクションの数に比例し、システムの故障率が時間と共に変化するモデル⁵⁾、累積ページ数の観点からトランザクションの到着率とシステムの故障率が時間と共に変化するモデル⁶⁾、あるいはチェックポイントを行う前に点検を行うことによってシステム障害を未然に防ぐモデル⁷⁾などが考えられている。最近、Odagiri et al.⁸⁾は、事後保全の考えを踏まえた上で最適チェックポイント方策を再構成している。

本論文では、従来のチェックポイントモデルで考慮されていなかった、チェックポイント時にヒューマンエラーや誤動作によって媒体障害が発生するというモデルについて考察する。参考文献9)では、チェックポイントに起因するシステム障害を考慮した保全モデルを取り扱っている。しかしながら、チェックポイントに起因する障害の中で最も危険な状態は、ヒューマンエラーなどによって更新されたファイルの情報だけではなく、二次記憶媒体上のファイル全体が破壊されるという場合であろう。事実、二次記憶媒体へのたび重なるアクセスや人間の誤動作によって、主記憶本体と二次記憶媒体の両方に障害が発生するのは、チェックポイントを行っている最中であると考えられる。このような壊滅的な媒

体障害は頻繁に起こるとは考えにくい、安全側に立脚した耐障害システムの信頼性設計を行うためには無視することが出来ない重要な要因である。そこで本論文の目的は、チェックポイントの際に生じるシステム障害および媒体障害が、(費用関数を最小にするという意味で)経済的最適チェックポイント方策に与える影響を定量的に評価することである。

本論文の構成は次のとおりである。まず2.において、媒体障害を伴うチェックポイントモデルを記述する。次に3.において、リカバリを行う際にジャーナルを用いる場合と用いない場合の2通りについて、チェックポイントとリカバリに要する定常状態における単位時間当たりの期待費用を解析的に導出する。ジャーナルを使う場合は、前述のロールバック・リカバリを実行することに対応している。一方、後者はパーソナルコンピュータ上のキーボードデータ入力操作におけるチェックポイント設定方式を想定している^{10)~12)}。いずれの場合も、ヒューマンエラーなどによるチェックポイント時におけるシステム障害と媒体障害を考慮にいたれたモデルとなっている点に注目すべきである。最後に4.においていくつかの数値例を表す。システム障害発生時間の確率分布関数に指数分布とWeibull分布をそれぞれ仮定した場合について、チェックポイントに起因する2種類の障害発生確率が最適チェックポイント方策に与える影響を数値的に調べる。さらに、各種コストパラメータとジャーナル処理率がチェックポイントにいかに関与するかについての感度分析も合わせて行う。

2. モデルと仮定

参考文献9)に従って、本研究で扱うモデルに対して以下の仮定をおく。

- 仮定1 主記憶と二次記憶から構成されるファイルシステムを取り扱う。システムは時刻0で動作を開始し、計画期間は無量大とする。
- 仮定2 システムはチェックポイント t_i ($i=1,2,3,\dots$) で瞬時にチェックポイントされる ($t_0=0$)。
- 仮定3 システム障害は直ちに発見され、システムのリカバリが瞬時に施される。システムの動作開始からシステムの障害発生後にリカバリが完了するまでの期間(ジャーナルを使う場合は、障害発生時点までのデータをリカバリするまで)を1サイクルとし、このサイクルが繰り返されるものとする。ただし、ジャーナルを使う場合には、リカバリ後すぐに媒体障害の発生を伴わないチェックポイントがなされる。

仮定 4 システムの障害発生時間は、有限な平均 $1/\lambda$ を持つ任意の連続形確率分布 $F(t)$ に従う。

仮定 5 時刻 t_{i-1} ($i = 1, 2, 3, \dots$) において、システムが正常に動作しているとき、時間間隔 $(t_{i-1}, t_i]$ の間にシステム障害が発生する確率を p ($0 < p < 1$)、すなわち、

$$\frac{F(t_i) - F(t_{i-1})}{F(t_{i-1})} = p \quad (1)$$

とする。ここで、一般的に $\bar{\phi} = 1 - \phi$ とする。また、 $F(\cdot)$ の逆関数を $F^{-1}(\cdot)$ とすると $t_i = F^{-1}(1 - \bar{p}^i)$ となる¹³⁾。つまり、 p をいったん設定することによって点列 t_i が決定される。

仮定 6 システム障害は処理期間中だけでなく、チェックポインティングの際にも確率 a ($0 \leq a \leq 1$) で発生する。つまり、チェックポインティングに起因するシステム障害⁹⁾が起こる可能性がある。

仮定 7 チェックポインティング時に確率 b ($0 \leq b \leq 1$) でシステム障害と同時に媒体障害が発生し、二次記憶媒体上のファイルが破壊される。このような媒体障害は、通常の処理期間中にはチェックポイントと独立に発生しないものとする。

さらに以下の諸量を定義する。

c_c : チェックポインティング 1 回当たりの費用。これはチェックポイントでシステム障害が発生した際のチェックポインティングに対しても同様にかかるものとする。

k_t : システム障害発生直前の媒体障害を伴わないチェックポイントからシステム障害発生時点まで、あるいは動作開始からシステム障害発生までの損失期間にかかる単位時間当たりの損失費用。

k_r : ジャーナルを使う場合のリカバリ（以下ジャーナル回復と呼ぶ）にかかる単位時間当たりの処理費用。

c_r : システム障害に対するリカバリ 1 回当たりの費用。

$E_C(p)$: 1 サイクル中の総期待費用。

$E_T(p)$: 1 サイクルの平均時間。

$C(p)$: 定常状態における単位時間当たりの期待費用。

ここで、チェックポインティングが瞬時に行われるという仮定について言及する。いま計画期間は十分大きいものと仮定しているので、一回のチェックポインティングに要するオーバーヘッドは相対的に十分小さいものと考えられる。そこで、各チェックポインティングに要する時間が一定であるとすれば、あたかもチェックポインティングを瞬時的に行っているように考えてもモデル上は問題ないことに注意すべきである。もし各チェックポインティングに

要するオーバーヘッドが τ (> 0) であり、単位オーバーヘッド当たりのチェックポインティング費用を c とすれば、チェックポインティング 1 回当たりの費用が $c_c = c\tau$ となることは自明である。そのとき、本論文で扱っている確率 a と ab で発生する障害は、時間間隔 $[0, \tau]$ の間に発生するシステム障害と媒体障害であると解釈することが可能である。よって、ここでは議論を簡略化するために、参考文献 4), 9) に従って、チェックポインティングは瞬時的に行われるものとする。

3. 解析

3.1 ジャーナルを使わない場合

まず最初に、回復動作にジャーナルを使わない場合の 1 サイクル中の総期待費用 $E_C(p)$ を求める。 $E_C(p)$ は 1 サイクル中の期待チェックポインティング費用、期待損失費用、期待リカバリ費用の和で表される。

1 サイクル中の期待チェックポインティング回数は、処理期間中にシステム障害が発生した場合の期待チェックポインティング数

$$\sum_{i=1}^{\infty} (i-1)(\bar{p}a)^{i-1}p = \frac{\bar{p}ap}{(1-\bar{p}a)^2}$$

と、チェックポイント t_i に起因するシステム障害が発生した場合の期待チェックポインティング数

$$\sum_{i=1}^{\infty} i(\bar{p}a)^{i-1}\bar{p}a = \frac{\bar{p}a}{(1-\bar{p}a)^2}$$

の和になるので、1 サイクル中の期待チェックポインティング費用は

$$c_c \left\{ \frac{\bar{p}ap}{(1-\bar{p}a)^2} + \frac{\bar{p}a}{(1-\bar{p}a)^2} \right\} = \frac{c_c \bar{p}}{1-\bar{p}a} \quad (2)$$

となる。つぎに、1 サイクル中の期待損失時間は、処理期間中にシステム障害が発生した場合の期待損失時間

$$\begin{aligned} & \sum_{i=1}^{\infty} \bar{a} \int_{t_{i-1}}^{t_i} (t - t_{i-1}) dF(t) \\ &= \bar{a} \int_0^{\infty} t dF(t) - \bar{a} \sum_{i=1}^{\infty} t_{i-1} [F(t_i) - F(t_{i-1})] \\ &= \bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \end{aligned}$$

と、チェックポイント t_i に起因する媒体障害を伴わないシステム障害が発生した場合の期待損失時間

$$\begin{aligned} & \sum_{i=1}^{\infty} (t_i - t_{i-1})(\bar{p}a)^{i-1}\bar{p}ab \\ &= \bar{p}ab \sum_{i=1}^{\infty} (t_i - t_{i-1})(\bar{p}a)^{i-1} \end{aligned}$$

および、チェックポイント t_i においてシステム障害と媒体障害が同時に発生した場合の期待損失時間

$$\sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \bar{p}ab = \bar{p}ab \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1}$$

の和となる。したがって、1 サイクル中の期待損失費用は

$$\begin{aligned} k_l \left[\bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} (t_i - t_{i-1}) (\bar{p}a)^{i-1} \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \right] \end{aligned} \quad (3)$$

となる。最後に、1 サイクル中の期待リカバリ回数は、処理期間中にシステム障害が発生した場合の期待リカバリ回数

$$\sum_{i=1}^{\infty} (\bar{p}a)^{i-1} p = \frac{p}{1 - \bar{p}a}$$

と、チェックポイント t_i に起因する媒体障害を伴わないシステム障害が発生した場合の期待リカバリ回数

$$\sum_{i=1}^{\infty} (\bar{p}a)^{i-1} \bar{p}ab = \frac{\bar{p}ab}{1 - \bar{p}a}$$

および、チェックポイント t_i においてシステム障害と媒体障害が同時に発生した場合の期待リカバリ回数

$$\sum_{i=1}^{\infty} (\bar{p}a)^{i-1} \bar{p}ab = \frac{\bar{p}ab}{1 - \bar{p}a}$$

の和により 1 となるので、1 サイクル中の期待リカバリ費用は c_r である。以上の結果より、1 サイクル中の総期待費用 $E_C(p)$ は、

$$\begin{aligned} E_C(p) = \frac{c_c \bar{p}}{1 - \bar{p}a} + k_l \left[\bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} (t_i - t_{i-1}) (\bar{p}a)^{i-1} \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \right] + c_r \end{aligned} \quad (4)$$

のように求めることができる。

つぎに、1 サイクルの平均時間 $E_T(p)$ を導出する。処理期間中にシステム障害が発生した場合の平均時間は

$$\sum_{i=1}^{\infty} \bar{a} \int_{t_{i-1}}^{t_i} t dF(t) = \frac{\bar{a}}{\lambda}$$

となる。また、チェックポイント t_i に起因するシステム障害が発生した場合の平均時間は

$$\sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \bar{p}a = \bar{p}a \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1}$$

となるので、1 サイクルの平均時間 $E_T(p)$ は

$$E_T(p) = \frac{\bar{a}}{\lambda} + \bar{p}a \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \quad (5)$$

のように求めることができる。これより、単位時間当たりの期待費用 $C(p)$ は、

$$C(p) = \frac{E_C(p)}{E_T(p)} \quad (6)$$

となる。

いま、特別な場合として、システムの障害発生時間が指数分布と Weibull 分布に従う場合を考える。システム障害の発生時間がパラメータ λ の指数分布に従うとき、つまり、

$$F(t) = 1 - \exp(-\lambda t), \quad \lambda > 0 \quad (7)$$

であれば、チェックポイント t_i ($i=1, 2, 3, \dots$) は簡単に次のように表現される¹³⁾。

$$t_i = -\frac{i}{\lambda} \ln \bar{p} \quad (8)$$

これより、1 サイクル中の総期待費用 $E_C(p)$ は

$$\begin{aligned} E_C(p) = \frac{c_c \bar{p}}{1 - \bar{p}a} \\ + k_l \left[\bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} \left(-\frac{i-1}{\lambda} \ln \bar{p} \right) \bar{p}^{i-1} \right) \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} \left(-\frac{i}{\lambda} + \frac{i-1}{\lambda} \right) \ln \bar{p} (\bar{p}a)^{i-1} \right. \\ \left. + \bar{p}ab \sum_{i=1}^{\infty} \left(-\frac{i}{\lambda} \ln \bar{p} \right) (\bar{p}a)^{i-1} \right] + c_r \\ = \frac{c_c \bar{p}}{1 - \bar{p}a} + \frac{k_l}{\lambda} \left[\bar{a} \left(1 + \frac{\bar{p}}{p} \ln \bar{p} \right) \right. \\ \left. - \frac{(1 - \bar{p}ab) \bar{p}a}{(1 - \bar{p}a)^2} \ln \bar{p} \right] + c_r \end{aligned} \quad (9)$$

となり、1 サイクルの平均時間 $E_T(p)$ は

$$\begin{aligned} E_T(p) = \frac{\bar{a}}{\lambda} + \bar{p}a \sum_{i=1}^{\infty} \left(-\frac{i}{\lambda} \ln \bar{p} \right) (\bar{p}a)^{i-1} \\ = \frac{1}{\lambda} \left(\bar{a} - \frac{\bar{p}a}{(1 - \bar{p}a)^2} \ln \bar{p} \right) \end{aligned} \quad (10)$$

となる。

つぎに、システム障害の発生時間が次のような 2 パラメータ Weibull 分布に従うものとする。

$$F(t) = 1 - \exp \left\{ - \left(\frac{t}{\eta} \right)^m \right\}, \quad \eta > 0, \quad m > 0 \quad (11)$$

このとき、チェックポイント t_i ($i=1,2,3,\dots$) は

$$t_i = \eta \left(i \ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \quad (12)$$

となるので¹⁴⁾、1 サイクル中の総期待費用 $E_C(p)$ は

$$\begin{aligned} E_C(p) = & \frac{c_c \bar{p}}{1 - \bar{p}a} + k_l \eta \left[\bar{a} \left(\Gamma \left(1 + \frac{1}{m} \right) \right. \right. \\ & - p \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} (i-1)^{\frac{1}{m}} \bar{p}^{i-1} \Big) \\ & + \bar{p} a \bar{b} \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} \left(i^{\frac{1}{m}} - (i-1)^{\frac{1}{m}} \right) (\bar{p}a)^{i-1} \\ & \left. + \bar{p} a b \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} i^{\frac{1}{m}} (\bar{p}a)^{i-1} \right] + c_r \quad (13) \end{aligned}$$

となり、1 サイクルの平均時間 $E_T(p)$ は

$$\begin{aligned} E_T(p) = & \eta \left(\bar{a} \Gamma \left(1 + \frac{1}{m} \right) \right. \\ & \left. + \bar{p} a \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} i^{\frac{1}{m}} (\bar{p}a)^{i-1} \right) \quad (14) \end{aligned}$$

となる。ここで、 $\Gamma(\cdot)$ は標準ガンマ関数である。

3.2 ジャーナルを使う場合

1 サイクル当たりの期待費用 $E_C(p)$ は、回復動作にジャーナルを使わない場合と同様に求めることができる。まず、処理期間中にシステム障害が発生した場合の期待チェックポインティング回数

$$\sum_{i=1}^{\infty} \{ (i-1)(\bar{p}a)^{i-1} p + (\bar{p}a)^{i-1} p \} = \frac{p}{(1 - \bar{p}a)^2}$$

と、チェックポインティング時にシステム障害が発生した場合の期待チェックポインティング回数

$$\sum_{i=1}^{\infty} \{ i(\bar{p}a)^{i-1} \bar{p}a + (\bar{p}a)^{i-1} \bar{p}a \} = \frac{\bar{p}a}{(1 - \bar{p}a)^2} + \frac{\bar{p}a}{1 - \bar{p}a}$$

を用いることによって、1 サイクル中の期待チェックポインティング費用は

$$\begin{aligned} c_c \left(\frac{p}{(1 - \bar{p}a)^2} + \frac{\bar{p}a}{(1 - \bar{p}a)^2} + \frac{\bar{p}a}{1 - \bar{p}a} \right) \\ = \frac{c_c \bar{p}}{1 - \bar{p}a} + c_c \quad (15) \end{aligned}$$

となる。1 サイクル中の期待損失時間は、ジャーナル回復（ジャーナルを使う場合のリカバリ）にかかる時間が損失時間に対応するので、つぎの三つの場合についての損失時間を求めればよい。

- (i) 処理期間中にシステム障害が発生した場合のジャーナル回復にかかる時間を以下のように求める。

$$\begin{aligned} \sum_{i=1}^{\infty} \int_{t_{i-1}}^{t_i} \bar{a} \frac{\kappa}{\mu} (t - t_{i-1}) dF(t) \\ = \bar{a} \frac{\kappa}{\mu} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \end{aligned}$$

ここで、 κ と μ はそれぞれ処理するデータの到着率とサービス率である。これより κ/μ はジャーナルの処理率を表す。

- (ii) チェックポイントに起因する媒体障害を伴わないシステム障害が発生した場合のジャーナル回復にかかる時間を以下のように求める。

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{\kappa}{\mu} (t_i - t_{i-1}) (\bar{p}a)^{i-1} \bar{p}a \bar{b} \\ = \frac{\kappa}{\mu} \bar{p}a \bar{b} \sum_{i=1}^{\infty} (t_i - t_{i-1}) (\bar{p}a)^{i-1} \end{aligned}$$

- (iii) チェックポイント時にシステム障害と媒体障害が同時に発生した場合のジャーナル回復にかかる時間を以下のように求める。

$$\sum_{i=1}^{\infty} \frac{\kappa}{\mu} t_i (\bar{p}a)^{i-1} \bar{p}a b = \frac{\kappa}{\mu} \bar{p}a b \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1}$$

以上より、1 サイクル中のジャーナル回復にかかる費用は

$$\begin{aligned} k_r \frac{\kappa}{\mu} \left[\bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \right. \\ \left. + \bar{p}a \bar{b} \sum_{i=1}^{\infty} (t_i - t_{i-1}) (\bar{p}a)^{i-1} + \bar{p}a b \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \right] \quad (16) \end{aligned}$$

となる。また、1 サイクル中の期待リカバリ費用は、回復動作にジャーナルを使わない場合と同様に c_r となる。これより、1 サイクル中の総期待費用 $E_C(p)$ は

$$\begin{aligned} E_C(p) = & \frac{c_c \bar{p}}{1 - \bar{p}a} + c_c + c_r \\ & + k_r \frac{\kappa}{\mu} \left[\bar{a} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \right. \\ & + \bar{p}a \bar{b} \sum_{i=1}^{\infty} (t_i - t_{i-1}) (\bar{p}a)^{i-1} \\ & \left. + \bar{p}a b \sum_{i=1}^{\infty} t_i (\bar{p}a)^{i-1} \right] \quad (17) \end{aligned}$$

のように求めることができる。

一方、処理期間中にシステム障害が発生した場合の平均時間は

$$\sum_{i=1}^{\infty} \int_{t_{i-1}}^{t_i} \bar{a} \left(t + \frac{\kappa}{\mu} (t - t_{i-1}) \right) dF(t) \\ = \frac{\bar{a}}{\lambda} + \bar{a} \frac{\kappa}{\mu} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right)$$

であり、チェックポイントに起因する媒体障害を伴わないシステム障害が発生した場合の平均時間は

$$\sum_{i=1}^{\infty} \left(t_i + \frac{\kappa}{\mu} (t_i - t_{i-1}) \right) (\bar{p}\bar{a})^{i-1} \bar{p}ab \\ = \bar{p}ab \sum_{i=1}^{\infty} \left(t_i + \frac{\kappa}{\mu} (t_i - t_{i-1}) \right) (\bar{p}\bar{a})^{i-1}$$

となる。さらに、チェックポイント時にシステム障害と媒体障害が同時に発生した場合の平均時間は

$$\sum_{i=1}^{\infty} \left(t_i + \frac{\kappa}{\mu} t_i \right) (\bar{p}\bar{a})^{i-1} \bar{p}ab \\ = \bar{p}ab \left(1 + \frac{\kappa}{\mu} \right) \sum_{i=1}^{\infty} t_i (\bar{p}\bar{a})^{i-1}$$

となる。以上より、1サイクルの平均時間 $E_T(p)$ は

$$E_T(p) = \bar{a} \left[\frac{1}{\lambda} + \frac{\kappa}{\mu} \left(\frac{1}{\lambda} - p \sum_{i=1}^{\infty} t_{i-1} \bar{p}^{i-1} \right) \right] \\ + \bar{p}ab \sum_{i=1}^{\infty} \left(t_i + \frac{\kappa}{\mu} (t_i - t_{i-1}) \right) (\bar{p}\bar{a})^{i-1} \\ + \bar{p}ab \left(1 + \frac{\kappa}{\mu} \right) \sum_{i=1}^{\infty} t_i (\bar{p}\bar{a})^{i-1} \quad (18)$$

のように表すことができる。

先の議論と同様に、システム障害の発生時間が指数分布と Weibull 分布に従う場合について考察する。 $F(t) = 1 - \exp(-\lambda t)$ のとき、1サイクル中の総期待費用 $E_C(p)$ ならびに1サイクルの平均時間 $E_T(p)$ は

$$E_C(p) = \frac{c_c \bar{p}}{1 - \bar{p}\bar{a}} + c_c + \frac{\kappa k_r}{\mu \lambda} \left[\bar{a} \left(1 + \frac{\bar{p}}{p} \ln \bar{p} \right) - \frac{(1 - \bar{p}\bar{a}) \bar{p}a}{(1 - \bar{p}\bar{a})^2} \ln \bar{p} \right] + c_r \quad (19)$$

$$E_T(p) = \frac{1}{\lambda} \left[\bar{a} \left(1 + \frac{\kappa}{\mu} \left(1 + \frac{\bar{p}}{p} \ln \bar{p} \right) \right) - \frac{\bar{p}ab}{1 - \bar{p}\bar{a}} \left(\frac{1}{1 - \bar{p}\bar{a}} + \frac{\kappa}{\mu} \right) \ln \bar{p} - \frac{\bar{p}ab}{(1 - \bar{p}\bar{a})^2} \left(1 + \frac{\kappa}{\mu} \right) \ln \bar{p} \right] \quad (20)$$

のようになる。また、システム障害の発生時間が Weibull 分布 $F(t) = 1 - \exp\{-(t/\eta)^m\}$ に従う場合、1サイクル中の期待費用 $E_C(p)$ ならびに1サイクルの平均時間 $E_T(p)$ は、それぞれ

$$E_C(p) = \frac{c_c \bar{p}}{1 - \bar{p}\bar{a}} + c_c + c_r + k_r \eta \frac{\kappa}{\mu} \left[\bar{a} \left(\Gamma \left(1 + \frac{1}{m} \right) - p \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} (i-1)^{\frac{1}{m}} \bar{p}^{i-1} \right) + \bar{p}ab \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} \left(i^{\frac{1}{m}} - (i-1)^{\frac{1}{m}} \right) (\bar{p}\bar{a})^{i-1} + \bar{p}ab \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} i^{\frac{1}{m}} (\bar{p}\bar{a})^{i-1} \right] \quad (21)$$

$$E_T(p) = \eta \left[\bar{a} \left\{ \Gamma \left(1 + \frac{1}{m} \right) + \frac{\kappa}{\mu} \left(\Gamma \left(1 + \frac{1}{m} \right) - p \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} (i-1)^{\frac{1}{m}} \bar{p}^{i-1} \right) \right\} + \bar{p}ab \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} \left\{ i^{\frac{1}{m}} + \frac{\kappa}{\mu} \left(i^{\frac{1}{m}} - (i-1)^{\frac{1}{m}} \right) \right\} (\bar{p}\bar{a})^{i-1} + \bar{p}ab \left(1 + \frac{\kappa}{\mu} \right) \left(\ln \frac{1}{\bar{p}} \right)^{\frac{1}{m}} \sum_{i=1}^{\infty} i^{\frac{1}{m}} (\bar{p}\bar{a})^{i-1} \right] \quad (22)$$

となる。

4. 数値例

3.1 と 3.2 で導出された解析的結果を用いることによって、障害発生時間分布が指数分布と Weibull 分布に従う場合について定常状態における単位時間当たりの期待費用 $C(p)$ を最小にする p を Newton 法により求める。Fig. 1 から明らかなように、システム障害が発生する時間が指数分布に従う場合、最適チェックポイント間隔は等間隔となるが、故障時間分布の年令特性が狭義に IFR (Increasing Failure Rare) と DFR (Decreasing Failure Rare) を示す場合、最適チェックポイント時刻列はそれぞれ減少列ならびに増加列となる (Figs. 2, 3 を参照)。これは従来からよく知られていた知見と一致する結果である^{4), 13), 14)}。Tables 1~4 は、チェックポイントに起因する2種類の障害発生確率の値に対する最適チェックポイント間隔と最小期待費用の従属性について調べたものである。表の結果から、壊滅的な媒体障害の発生確率 b が大きくなるにつれて、最適なチェックポイントの値は増大する傾向にあることが分かる。これはある程度確率 b が大きいならば、壊滅的な媒体障害を避けるためにチェックポイントインテングを頻繁に行わないことが得策であることを示している。また、チェックポイントに起因するシステム障害が発生する確率 a が大きければ、最適なチェックポイントの値は低めに見積もられるべきである。この結果は参考文献9) のものと全く同一である。とりわけ、チェックポイントインテング時に発生する障害の程度に応じて、チェック

Fig. 1 Optimal checkpoint sequence (the case of exponential failure)

Fig. 2 Optimal checkpoint sequence (the case of Weibull failure ($m=2$))

Fig. 3 Optimal checkpoint sequence (the case of Weibull failure ($m=0.5$))

ポイント間隔に関する単調性が異なる点は興味深い。つまり、ここで論じられたモデルは、直前のチェックポイントからの情報だけが失われる場合とすべての情報が失われる場合という極端な事象のみがファイルシステムの障害発生メカニズムに組み込まれているにもかかわらず、障害の規模がチェックポイント方策に強く影響を及ぼしていることを示唆している。

つぎに、他のモデルパラメータに関する感度分析の結果を整理する。Figs. 4～6はチェックポインティング費用 c_c 、ジャーナル回復費用 k_r 、ジャーナル処理率 κ/μ の変化と期待費用関数の振る舞いを図示したものである。Fig. 4では、回復動作にジャーナルを使う場合において、チェッ

Table 1 Dependence of the optimal checkpointing policy on the catastrophic media failure probability; the case of the exponential failure and the recovery without journal

$$\lambda = 0.2, c_c = 1, k_l = 2, c_r = 3$$

a	b	p	$C(p)$	t_1
0.05	0	0.318	1.329	1.912
	0.05	0.321	1.335	1.939
	0.10	0.325	1.341	1.965
0.10	0	0.266	1.304	1.549
	0.05	0.272	1.316	1.585
	0.10	0.277	1.327	1.621
0.20	0	0.217	1.314	1.225
	0.05	0.220	1.330	1.245
	0.10	0.224	1.346	1.267

Table 2 Dependence of the optimal checkpointing policy on the catastrophic media failure probability; the case of the exponential failure and the recovery with journal

$$\lambda = 0.2, c_c = 1, k_r = 2, c_r = 2, \kappa = 0.5, \mu = 1$$

a	b	p	$C(p)$	t_1
0.05	0	0.579	0.780	4.327
	0.05	0.582	0.781	4.356
	0.10	0.584	0.781	4.387
0.10	0	0.552	0.790	4.016
	0.05	0.558	0.790	4.077
	0.10	0.563	0.791	4.140
0.20	0	0.514	0.816	3.607
	0.05	0.525	0.818	3.717
	0.10	0.536	0.819	3.836

Table 3 Dependence of the optimal checkpointing policy on the catastrophic media failure probability; the case of the Weibull failure and the recovery without journal

$$\eta = 2, m = 2, c_c = 1, k_l = 3, c_r = 2$$

a	b	p	$C(p)$	t_1	t_2	t_3
0.05	0	0.404	2.732	1.438	2.034	2.491
	0.05	0.406	2.738	1.444	2.043	2.501
	0.10	0.409	2.745	1.450	2.051	2.512
0.10	0	0.344	2.680	1.300	1.838	2.251
	0.05	0.350	2.694	1.312	1.855	2.272
	0.10	0.355	2.708	1.324	1.873	2.293
0.20	0	0.252	2.618	1.078	1.524	1.867
	0.05	0.259	2.644	1.094	1.549	1.896
	0.10	0.266	2.670	1.112	1.572	1.925

Table 4 Dependence of the optimal checkpointing policy on the catastrophic media failure probability; the case of the Weibull failure and the recovery with journal

$$\eta = 2, m = 2, c_c = 1, k_r = 2, c_r = 2, \kappa = 0.5, \mu = 1.0$$

a	b	p	$C(p)$	t_1	t_2	t_3
0.05	0	0.664	1.971	2.088	2.952	3.616
	0.05	0.665	1.972	2.090	2.956	3.621
	0.1	0.666	1.972	2.093	2.960	3.625
0.1	0	0.615	1.989	1.953	2.762	3.382
	0.05	0.617	1.990	1.959	2.771	3.393
	0.1	0.619	1.991	1.965	2.779	3.404
0.2	0	0.501	2.016	1.667	2.357	2.887
	0.05	0.506	2.020	1.680	2.376	2.910
	0.1	0.511	2.023	1.693	2.394	2.932

クポインティング費用が増加すれば対応する最適な p の値も増加している。逆に、ジャーナルによるリカバリにかかる費用が増加すれば、最適な p は減少することを Fig. 5 は示している。いずれの場合も、期待費用の低減効果により p が単調に変化しているものと考えられる。さらに Fig. 6 において、トランザクションの処理率が増加すると最適な p は減少傾向を示すことがわかる。処理率が向上するということは、システムの主記憶上で取り扱う実データ量が増加することを意味するので、ジャーナル回復にかかる費用の増加を抑制するためにチェックポイント間隔を小さめに設定すべきであると解釈することが出来る。上記以外のパラメータについて、著しく最適方策に

Fig. 4 Behavior of the long-run average cost for various checkpointing costs (the case of exponential failure)

Fig. 5 Behavior of the long-run average cost for various recovery costs of the journal (the case of exponential failure)

Fig. 6 Behavior of the long-run average cost for various crossing rates of the journal (the case of exponential failure)

影響を及ぼすという結果は数値実験において全く観察されなかった。

5. おわりに

本論文では、チェックポインティング時に媒体障害が発生する場合のチェックポインティングモデルについて、障害発生時間分布に任意の確率分布関数を仮定し、定常状態における単位時間当たりの期待費用を解析的に導出した。実際のファイル系システムの運用において、オペレータの操作ミスもしくはたび重なる二次記憶媒体へのアクセスによるヘッドクラッシュといったハード故障は極めて深刻な問題である。壊滅的な媒体障害が頻繁に起こるとは考えられないが、この要因を考慮に入れることにより、リスク回避の見地からの耐障害システムの信頼性設計を行うことができる。

つぎに数値例において、障害発生時間の累積分布関数に指数分布と Weibull 分布を仮定した場合について考察を行った。その結果、2種類の障害発生確率の最適チェックポインティング方策に与える影響が異なるという興味深い結果を得た。また、各コストパラメータとトランザクションの処理率が最適方策に与える影響についても言及した。

本論文で議論した処理モデルは、主記憶と二次記憶装置から構成される単純化されたファイルシステムであった。しかしながら、壊滅的な媒体障害による影響を実装例と対比させながら定量的に評価するためには、現実の汎用機系 OLTP システムに忠実に、大規模かつ複雑なモデル化を行う必要がある。よって今後の課題として、本論文で取り上げた壊滅的な障害の概念を用いることによって、実際の大規模主記憶データベースシステムなどの評価を行うことが望まれる。

参考文献

- 1) 当麻：コンピュータシステムの高信頼化技術入門，日本規格協会 (1988)
- 2) J. W. Young: A first order approximation to the optimum checkpoint interval; *Communications of the ACM*, Vol. 7, No. 9, pp. 530-531 (1974)
- 3) E. Gelenbe: On the optimum checkpoint interval; *Journal of the ACM*, Vol. 26, No. 2, pp. 100-110 (1975)
- 4) 福本，海生，尾崎：チェックポイント濃度を使用した最適チェックポインティング方策；情報処理学会論文誌，Vol. 31, No. 6, pp. 887-893 (1990)
- 5) S. Fukumoto, N. Kaio and S. Osaki: Evaluation for a database recovery action with periodical checkpoint generations; *IEICE Transactions*, Vol. E-74, No. 7, pp. 2076-2082 (1991)
- 6) S. Fukumoto, N. Kaio and S. Osaki: A study of checkpoint generations for a database recovery mechanism; *Computers & Mathematics*, Vol. 24, No. 1/2, pp. 63-70 (1992)
- 7) L. B. Boguslavsky, E. G. Coffman, Jr., E. N. Gilbert and A. Y. Kreinin: Scheduling checks and saves; *ORSA Journal of Computing*, Vol. 4, No. 1, pp. 60-69 (1992)
- 8) M. Odagiri, N. Kaio and S. Osaki: A note on optimal checkpoint sequence taking account of preventive maintenance; *IEICE Trans. Fundamentals*, Vol. E77-A, No. 1, pp. 244-246 (1994)
- 9) 柳原，小田切，海生，尾崎：チェックポインティングに起因するシステム障害を考慮した最適チェックポインティング方策；電子情報通信学会論文誌，Vol. J78-A, No. 1, pp. 54-63 (1995)
- 10) 三道，河合：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；電子情報通信学会論文誌，Vol. J73-D-I, No. 11, pp. 900-904 (1990)
- 11) 三道，河合：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；電子情報通信学会論文誌，Vol. 73-A, No. 9 pp. 1496-1500 (1990)
- 12) 福本，石井，前沢，海生，尾崎：キーボードデータ入力操作におけるチェックポイント設定方式とその有効性に関する考察；電子情報通信学会論文誌，Vol. J76-D-I, No. 7, pp. 390-404 (1993)
- 13) A. G. Munford and A. K. Shahani: A nearly optimal inspection policy; *Operational Research Quarterly*, Vol. 23, No. 3, pp. 373-379 (1972)
- 14) A. G. Munford and A. K. Shahani: An inspection policy for the weibull case; *Operational Research Quarterly*, Vol. 24, No. 3, pp. 453-458 (1973)