

Mass Storage System を用いたジョブ処理の性能解析†

波 辺 豊 英†† 星 野 聰†††

MSS に作成したデータセットを用いてジョブ処理を行う場合に、ジョブの処理時間がどのようになるかを待ち行列モデルによって解析した。MSS の解析モデルを構成するにあたって、実際に複数のジョブを用いて実験し、MSS の動作特性を調べた。この実験によって、MSS の解析モデルは SDC 系と DRD 系の二つの処理単位から構成された循環待ち行列システムで表現することができる。ジョブ処理の性能解析をステージ・モード、シリンダ不在モード、ページ不在モードという MSS のステージング方法によって分け、各種パラメータを変えることによって種々の場合を想定して検討した。その結果、データセットの一部しか利用しない場合には、ステージ・モードよりもシリンダ、およびページ不在モードの方が処理時間が短い。たとえば、データセットの大きさが平均 2M バイトでは、約 14% 以下の利用であればシリンダ、およびページ不在モードのステージング方法が、また約 71% 以下の利用であればページ不在モードのステージング方法が効率よく処理する。さらに、全データセットが必要とする記憶域の大きさより、ステージング DASD の容量が少ない場合を想定して解析しても、ジョブの処理時間はそのステージング DASD の容量にほとんど影響を受けず、ステージング DASD の容量が十分な場合とはほぼ同じ傾向を示し、各ステージング方法にも特別な結果は現れなかった。

1. はじめに

膨大なデータを磁気テープ並みの価格で記録でき、またオンラインで処理できる記憶媒体 MSS (Mass Storage System)¹⁾ が開発され、利用できるようになった。MSS は利用者にあたかも磁気ディスク・ファイルを扱うかのようなアクセス手順を提供し、磁気テープ・ファイルのようにマウント、デマウント操作を要求しない。MSS が自動的にこれらの操作を行う。したがって、MSS は磁気ディスク・ファイルに比べて機械動作が伴うために、アクセス時間が遅い。しかし、その記憶容量は磁気ディスク・ファイルに比べて $10^2 \sim 10^3$ 倍であり、膨大なデータを経済的に格納できる有用なツールである。膨大なデータも一処理で利用されるのは全体の一部であり、すべてのデータがすぐにアクセス可能な状態である必要はない。とくに、データベースの格納ファイルとして、また業務・管理データの保存ファイルとして利用すれば効果がある。

本稿では、MSS に作成したデータセットを処理するジョブの処理時間を解析・検討する。まず、MSS の解析モデルを待ち行列システムで構成し、続いて解析モデルの評価に必要なパラメータの設定について述

べる。このパラメータの値には、京都大学大型計算機センターに設置されている MSS を用いて実験し^{3),5)}、測定データから求めた。そして、この解析モデルに対して種々のジョブ処理の環境を設定し、MSS の動作特性によるジョブの処理時間を評価する。

2. MSS 動作の概要

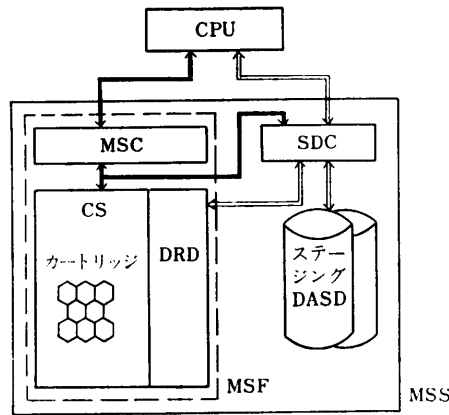
MSS はさまざまな装置・機構から構成され、そのハードウェア構成を図 1 に示す。MSF (Mass Storage Facility) はデータを管理・格納し、MSS 全体を制御する装置で、複数の機構から構成されている。これらは、カートリッジを保管する CS (Cartridge Store) 機構、カートリッジに対してデータを読書きする DRD (Data Recording Device) 機構、MSS を構成する各ハードウェアを管理・制御する MSC (Mass Storage Control) 機構などである。また、CPU が直接データを読書きできるステージング DASD (Direct Access Storage Device) という磁気ディスク装置があり、さらにステージング DASD と CPU 間、DRD とステージング DASD 間のデータ転送などを行う SDC (Staging Disk Control) がある。

MSS は磁気ディスク・ファイルの仮想化を、カートリッジ・ファイルを用いてステージング DASD で実現している。これは、仮想記憶が外部記憶ファイルを主記憶の延長として制御する機構と同様な概念の下に、カートリッジ・ファイルを磁気ディスク・ファイルの延長として制御する機構である。ステージング

† Performance Analysis of Processing Times of Jobs Using Datasets on a Mass Storage System by TOYOHIDE WATANABE (Data Processing Center, Kyoto University) and SATOSHI HOSHINO (University of Library and Information Science).

†† 京都大学大型計算機センター

††† 図書館情報大学



MSF : Mass Storage Facility
 MSC : Mass Storage Control
 DRD : Data Recording Device
 CS : Cartridge Store
 SDC : Staging Disk Control

⇔ : データの流れ
 ← : 制御情報の流れ

図 1 MSS のハードウェア構成図

Fig. 1 Hardware configuration of MSS.

DASD は約 2M バイトのページと呼ばれる記憶域を単位として管理され、1 ページは 1 シリンダ約 248 k バイトが 8 個連続して構成される。

DRD からステージング DASD へのデータ転送をステージング (staging)、反対方向の転送をデステージング (destaging) という。

MSS は表 1 に示す手順に従って CPU の要求を処理する。ステージングの処理はステップ 2 から 6 であり、デステージングの場合には、ステップ 5 でデステージングの実行を指示し、ステップ 6 でステージング DASD から DRD にデータを転送する。表 1 の

表 1 MSS の動作シーケンス
 Table 1 Execution process of MSS.

ステップ	時刻	動作内容
1		CPU は MSC にステージング要求を行う。
2	T 1	MSC はこの要求に対してミニヘッダ (MSS が要求を処理するための管理情報) を作成する。
3	T 2	MSC はこのミニヘッダを処理できる DRD を選択する。
4	T 3	アクセス機構は該当のカートリッジを選択して DRD に設定する。そして、DRD はカートリッジ・データを読み、種々の制御データを照合する。
5	T 4	MSC は SDC にステージングの実行を指示する。
6	T 5	DRD は SDC を介してカートリッジ・データをステージング DASD に転送する。
7		CPU は SDC を介してステージング DASD のデータを参照する。

注) 時刻 T_i は各ステップの処理が終了した時刻である。ステップ 2 からステップ 6 が MSS のステージング動作を表している。

T_i は各ステップが終了した時刻を表す。もちろん、すでにステージング DASD に必要なデータがある場合には、CPU の要求はステップ 7 の処理だけで終了する。データセットのマウント時の処理もほぼ同じ手順で行われ、1 シリンダの制御データがステージング DASD に転送される。ページ単位に管理されているステージング DASD では、1 シリンダの記憶域が占有され、残りは空き状態になる。他のページに属するデータは、この空き領域に割り付けられない。

DRD からステージング DASD にデータを転送するには種々のモードがあり、代表的なステージング方法は、

- ステージ・モード (s) : データセットのオープン時にすべてのデータをステージング DASD に転送
- シリンダ不在モード (c) : ステージング要求のつど、1 シリンダのデータをステージング DASD に転送
- ページ不在モード (p) : ステージング要求のつど、1 ページのデータをステージング DASD に転送

である。これらの処理モードを以下、s, c, p と表し、本文ではこの記号を用いて述べる。

3. MSS の解析モデル

MSS に格納したデータセットを読書きするジョブの処理時間を求めるために、MSS をモデル化して評価する。ここで、ジョブの処理時間とは、ジョブがデータ処理を行うのに必要なすべての読書き動作を終了するまでの時間とする。

3.1 解析モデルの構成

実際に MSS の動作特性を調べて MSS の解析モデルを構成することは、モデルを単純で正確な構造にする。MSS を用いて*複数の FORTRAN ジョブを連続して実行させ、表 1 の各ステップがどのように処理されるかを調べた³⁾。図 2 にジョブごとの処理経過を示した。横軸に 28 個のジョブを開始時刻が早いものから並べ、縦軸に経過時間を表した。

図 2 には、ステージング動作の場合 (case 1, case 2) とデステージング動作の場合 (case 3, case 4) を示した。各ジョブが時間を置いて、ほぼ等間隔に到着している case 1 では、各ジョブの処理時間にほとんど

* 実験時の MSS の構成は、実装可能なカートリッジの容量 102G バイト、DRD 4 台、SDC 2 台であり、またステージング DASD の容量は 200M バイトである。

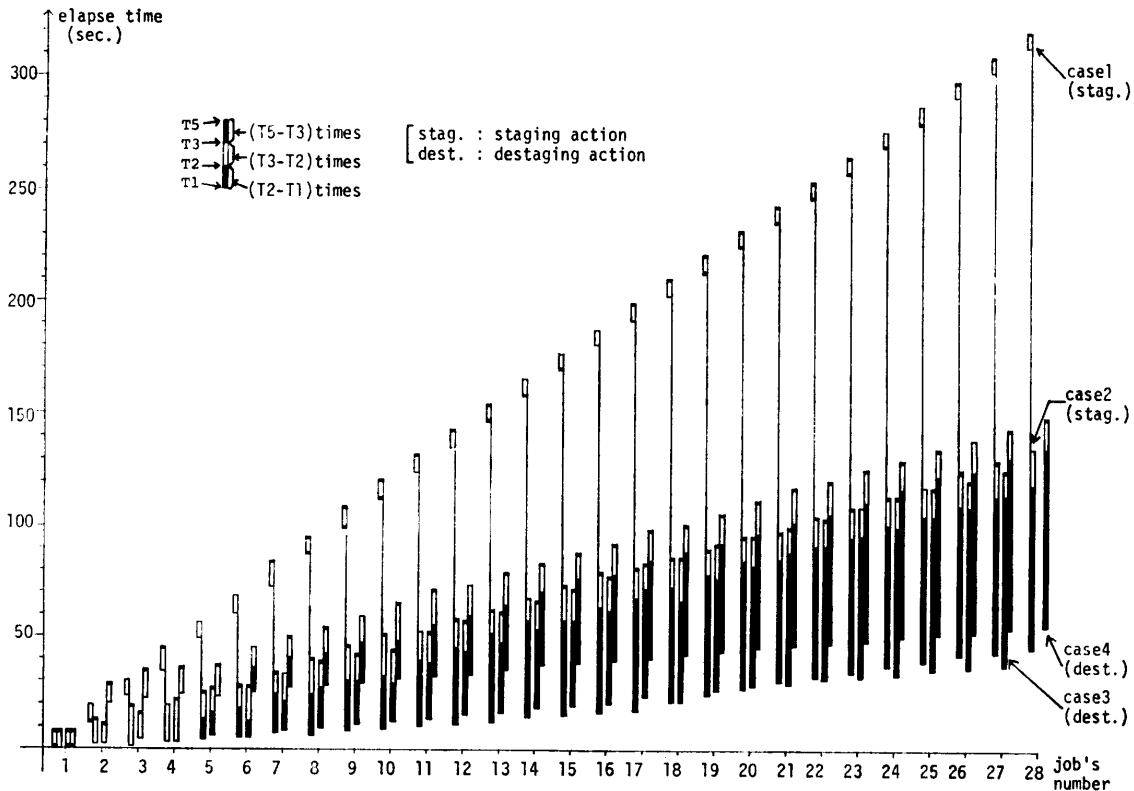


図2 ステージング, デステージング時の MSS の動作 (1 シリンダの転送)
 Fig. 2 MSS actions of staging or destaging.

ど差はないが, すべてのジョブがほぼ同時に到着している case 2 から case 4 では, case 1 に比べて ($T2-T1$) 時間はジョブの番号が大きくなるほど, 長くなっている. また, ($T3-T2$) 時間も若干長くなっている. すなわち, case 2 から case 4 では, DRD の負荷が非常に大きく, 各ジョブは DRD のサービスを受けるために待ち行列を構成している. 一方, ($T5-T3$) 時間は case 1 から case 4 ではほとんど差がなく, SDC のサービスと考えられる.

したがって, MSS の解析モデルは, ($T5-T3$) 時間の SDC のデータ転送のサービス部と ($T3-T1$) 時間の DRD のカートリッジの選択・参照, およびデータの照合などのサービス部から構成できる. もちろん, 時刻 $T3$ から $T5$ でも DRD が関与しているが, この時点では時刻 $T2$ ですでに割り当てられた DRD が利用され, SDC の割当てとサービスだけが問題となる. case 2 から case 4 において, ステップ 3 の処理時間が最初の 4 個を除いて三つのジョブにしか重なっていない. すなわち, あるジョブの時刻 $T2$ は四つの前のジョブの時刻 $T3$ が終了した後である. これは, DRD があるジョブのサービスを終了しても使

用したカートリッジを元の格納場所に戻す時間が必要であり, DRD はこれら一連の処理が完了してから次のサービスを開始できるからである. このようにカートリッジを返却することによって次の要求が待たされ

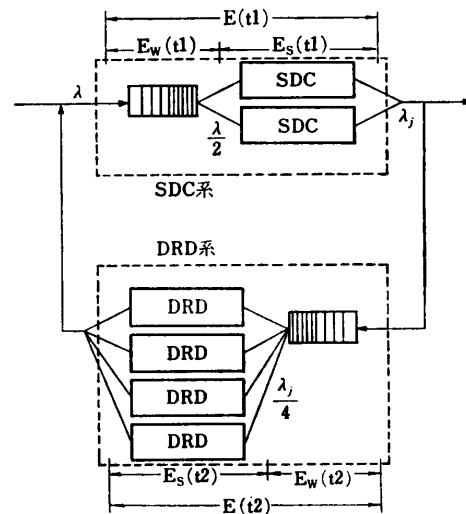


図3 MSS の解析モデル
 Fig. 3 Analysis model of MSS.

る時間を DRD の返却時間と呼ぶ。実験結果より、これを DRD のサービスと考える。

以上の内容から、MSS の解析モデルを図 3 のように構成できる⁴⁾。すなわち、SDC のサービスに関する SDC 系の待ち行列システムと DRD のサービスに関する DRD 系の待ち行列システムに分かれ、これらが連結した循環待ち行列システムになる。このモデルは Misra の単純モデルとほぼ同じである¹⁾。彼のモデルが MSS の各装置・機構の動作時間を解析するモデルであるのに対して、われわれのモデルはジョブの処理時間を解析するモデルである。すなわち、彼のモデルはステージング、デステージングの処理における MSS の動作特性を問題としているが、われわれのモデルはジョブが MSS のデータセットを讀書きするための処理時間を問題にしている。

3.2 解析モデルの定式化

次に、図 3 の解析モデルを定式化する⁶⁾。この定式化に用いた記号を表 2 に整理した。各要求の到着はポアソン分布に、また各サービス時間は指数分布に従うとする²⁾。なぜなら、SDC 系のサービスには MSC との制御情報の転送、ステージング DASD 上のデータ

表 2 待ち行列の解析に用いられる記号
Table 2 Parameters used in the analysis of MSS queuing model.

記号	意味
1)	λ_1 ステージング要求の到着率 λ_2 デステージング要求の到着率 λ_3 マウント要求の到着率 λ_4 CPU のデータ読書き要求の到着率 λ_5 MSC と SDC とのデータ転送要求の到着率 λ $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5$ λ_j $\lambda_1 + \lambda_2 + \lambda_3$
2)	$E_s(t1)$ SDC 系の平均サービス時間 $E_d(t2)$ DRD 系の平均サービス時間 $E_{s1}(t1)$ SDC 系の DRD とステージング DASD 間のデータ転送の平均サービス時間 $E_{s2}(t1)$ SDC 系の CPU とステージング DASD 間のデータ転送の平均サービス時間 $E_{s3}(t1)$ SDC 系の MSC と SDC 間のメッセージ、制御情報の転送の平均サービス時間 $E_{s4}(t1)$ SDC 系のデータ・チェックなどの平均サービス時間
3)	$E_w(t1)$ SDC 系の平均待ち時間 $E_w(t2)$ DRD 系の平均待ち時間
4)	$E(t1)$ SDC 系の平均応答時間 $E(t2)$ DRD 系の平均応答時間 $E_1(t1)$ SDC 系の DRD とステージング DASD 間のデータ転送の平均応答時間 $E_2(t1)$ SDC 系の CPU とステージング DASD 間のデータ転送の平均応答時間 $E_3(t1)$ SDC 系の MSC と SDC 間のメッセージ、制御情報の転送の平均応答時間 $E_4(t1)$ SDC 系のデータ・チェックなどの平均応答時間

管理、CPU とのデータ転送、DRD とのデータ転送があり、サービス時間はそれぞれ異なっている。一方、DRD 系のサービスも、カートリッジの格納場所によってアクセス時間が異なり、また DRD の返却時間の影響が要求の到着間隔によって異なるからである。

3.2.1 SDC 系について

この系の平均サービス時間は、

$$E_s(t1) = (1/\lambda) \{ \lambda_1 E_{s1}(t1) + \lambda_2 E_{s2}(t1) + \lambda_3 E_{s3}(t1) + (\lambda_4 + \lambda_5) E_{s4}(t1) \} \quad (1)$$

である。SDC は 2 台の窓口を有し、ピンチン・ポラック式より M/M/2 型になる。平均待ち時間は、

$$E_w(t1) = P_1(B) E_s(t1) / (2 - \lambda E_s(t1)) \quad (2)$$

となる。P₁(B) は 2 台の窓口の稼働率である。

$$P_1(B)$$

$$= \left\{ 1 - \frac{\sum_{n=0}^1 [\lambda E_s(t1)]^n / n!}{\sum_{n=0}^2 [\lambda E_s(t1)]^n / n!} \right\} / \left\{ 1 - (\lambda E_s(t1) / 2) \frac{\sum_{n=0}^1 [\lambda E_s(t1)]^n / n!}{\sum_{n=0}^2 [\lambda E_s(t1)]^n / n!} \right\}$$

$$= \lambda^2 E_s^2(t1) / (2 + \lambda E_s(t1)) \quad (3)$$

したがって、式(2)は

$$E_w(t1) = \{ \lambda^2 E_s^2(t1) / (4 - \lambda^2 E_s^2(t1)) \} E_s(t1) \quad (4)$$

となる。SDC 系における個々の要求の平均応答時間は、

$$E_1(t1) = E_w(t1) + E_{s1}(t1) \quad (5)$$

$$E_2(t1) = E_w(t1) + E_{s2}(t1) \quad (6)$$

$$E_3(t1) = E_w(t1) + E_{s3}(t1) \quad (7)$$

$$E_4(t1) = E_w(t1) + E_{s4}(t1) \quad (8)$$

である。

3.2.2 DRD 系について

この系には λ_1 から λ_5 に対する要求が到着し、M/M/4 型の待ち行列になる。この平均待ち時間は、

$$E_w(t2) = P_2(B) E_d(t2) / (4 - \lambda_j E_d(t2)) \quad (9)$$

である。ここで、P₂(B) は P₁(B) と同様に 4 台の窓口を有した DRD の稼働率であり、次式で示される。

$$P_2(B) = \lambda_j^4 E_d^4(t2) / \{ 24 + 18 \lambda_j E_d(t2) + 6 \lambda_j^2 E_d^2(t2) + \lambda_j^3 E_d^3(t2) \} \quad (10)$$

これにより、式(9)は、

$$E_w(t2) = \lambda_j^4 E_d^5(t2) / \{ (24 + 18 \lambda_j E_d(t2) + 6 \lambda_j^2 E_d^2(t2) + \lambda_j^3 E_d^3(t2)) (4 - \lambda_j E_d(t2)) \} \quad (11)$$

となる。したがって、平均応答時間は、

$$E(t2) = E_w(t2) + E_d(t2) \quad (12)$$

である。

以上のように応答時間を決定でき、 λ_1 から λ_5 に対

する要求の平均応答時間は、

(a) ステージング要求時:

$$E_{s1}(t) = E_1(t1) + E_4(t1) + E(t2) \quad (13)$$

(b) デステージング要求時:

$$E_{s2}(t) = E_1(t1) + E_4(t1) + E(t2) \quad (14)$$

(c) マウント時:

$$E_{s3}(t) = E_1(t1) + E_4(t1) + E(t2) \quad (15)$$

(d) CPU のデータ読書き時:

$$E_{s4}(t) = E_2(t1) + E_4(t1) \quad (16)$$

(e) SDC と MSC 間のデータ転送時:

$$E_{s5}(t) = E_3(t1) \quad (17)$$

である。これらがそれぞれ λ_1 から λ_5 で到着するために、1 要求当りの平均応答時間は、

$$E(t) = (1/\lambda) \{ \lambda_1 E_1(t1) + \lambda_4 E_2(t1) + \lambda_5 E_3(t1) + (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) E_4(t1) + \lambda_j E(t2) \} \quad (18)$$

となる。

4. 解析モデルのパラメータ

式(18)を用いてジョブの処理時間を求めるには、各パラメータの値を定める必要がある。適当な処理状況を想定し、実験の測定データに基づいてパラメータ値を決める。ここでは、CPU がデータを讀込む処理だけを考え、デステージング動作は発生しないとする。デステージング動作は、データセットがクローズ処理されたときか、またステージング DASD に空きページがなく、更新されたデータがあったときに発生し、このような状況を想定しても解析を困難にし、また冗長な結果しか得られない。

まず、サービス時間を考える。 $E_{s3}(t1)$, $E_{s4}(t1)$ は、 $E_{s1}(t1)$, $E_{s2}(t1)$ に比べて非常に小さく、無視できる。なぜなら、MSS への要求を管理する情報 (ミニヘッド) の大きさは $(32 + \alpha)$ バイトであり、 α は要求の内容にも依存するが 256 バイト前後である。一方、制御情報の大きさも $(38 + \beta)$ バイトであり、 β は 128 バイト以下である。これらを 1 シリンダ約 248 k バイトと比べると、SDC と MSC 間のデータ転送のサービス $E_{s3}(t1)$ はほとんど問題にならない。また、SDC のデータ管理のサービス $E_{s4}(t1)$ もデータ転送に伴うジョブの処理時間を検討する場合に演算速度は速く、無視できる。したがって、 $E_{s3}(t1)$, $E_{s4}(t1)$ を値 0 とすれば、式(18)は、

$$E(t) = (1/\lambda) \{ \lambda_1 E_1(t1) + \lambda_4 E_2(t1) + \lambda_j E(t2) \} \quad (19)$$

になる。測定データより、 $E_{s1}(t1)$, $E_{s2}(t1)$, $E_s(t2)$ を次のように定めることができる。

$E_{s1}(t1)$ は 1 シリンダの転送に 0.655 秒を要し、 $E_{s2}(t1)$ も 2 k バイトの転送に 0.005 秒を要する。また、 $E_s(t2)$ は DRD の稼働率が非常に低く、DRD の返却時間が影響しなければ 7.414 秒である。DRD の返却時間は要求の到着率の関数になり、この時間を d とすれば一般に $(7.414 + d)$ 秒となる。ただし、 $0 \leq d < 5.000$ である。付録に d の決定方法を示した。

次に、要求の到着率を考える。 $E_{s3}(t1)$ を値 0 としたために λ_5 は考えない。 λ_1 から λ_4 はデータセットの大きさや個数、およびステージング方法によって異なる。いま、データの讀込み処理を考えているので、あるジョブが何回の入力要求を発生すればデータを読み終えてしまうかを問題にする。したがって、データセットの大きさによって必要な讀込みの要求数を計算でき、これらがある時間 T の間に到着するとして、 λ_1 から λ_4 を定める。たとえば、データセットの大きさが 4 M バイト、CPU の 1 回のデータ讀込み量が 8 k バイトであれば、1 ジョブの処理に必要な CPU の要求数は 500 回になる。このとき、ステージング方法が ■ であれば、ステージング要求とマウント要求はそれぞれ 1 回になる。したがって、1 ジョブしか考えなければ、 $\lambda_1 = 1/T$, $\lambda_2 = 0$, $\lambda_3 = 1/T$, $\lambda_4 = 500/T$ と決まる。複数のジョブであれば、要求数はジョブの数に比例して多くなる。時間 T は要求の到着率を変えるパラメータで、MSS への要求を密に到着させるか否かなどの状況を設定できる。

5. 解析モデルによる性能評価

以上のようなパラメータの設定に従い、さまざまな状況に応じたジョブの処理時間を求める。式(19)を用いて 1 要求の平均応答時間を求め、1 ジョブの処理に必要な要求数を積算すれば 1 ジョブの平均処理時間を求めることができる。

5.1 ステージング方法の比較

30 個のジョブを想定し、平均 4 M バイトのデータセットを 8 k バイト単位に CPU が読み込む状況を考える*。1 ジョブがデータセット全体を読み込むには 500 回の CPU 要求が必要になり、マウント要求の数はデータセットの個数と同じく、30 回になる。一方、ステージング要求の数はステージング方法によって異なる。ただし、ステージング DASD の容量は十分大きいとする。この結果を図 4 に示す。縦軸は平均処理

* これらの値は実験時の状況に近い数値であり、ここで特別な意味をもたない。

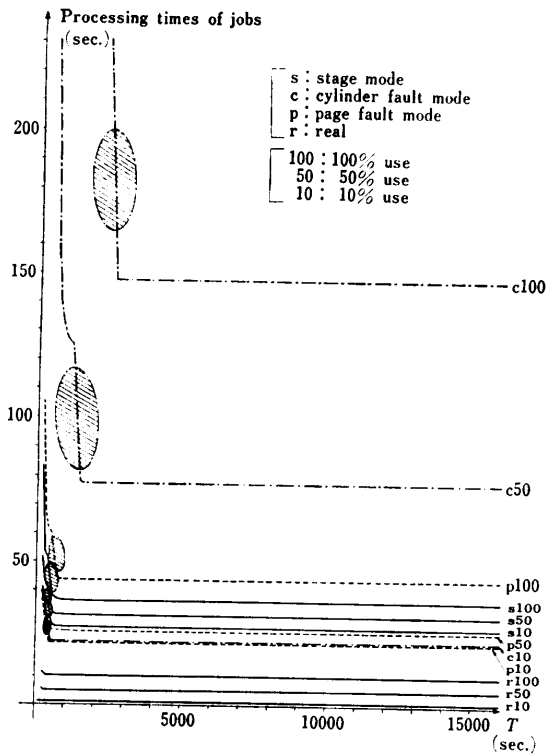
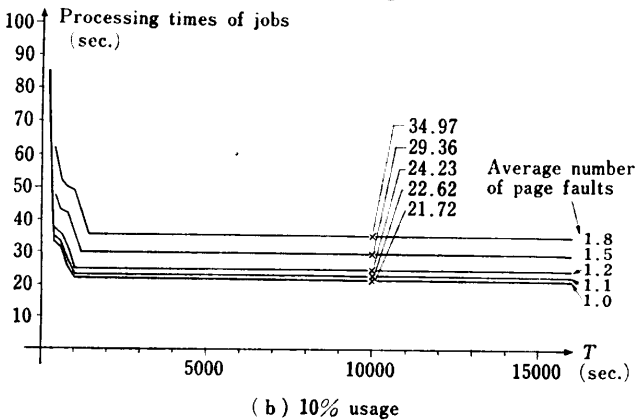
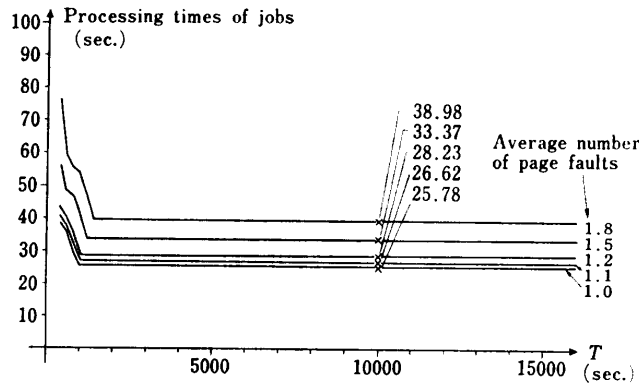


図4 ステージング方法による平均処理時間
Fig. 4 Average processing times by staging methods.



時間であり、横軸は時間 T の値で、要求の到着率に対応している。各グラフで急激に変化している部分（斜線部）は DRD の返却時間の変化が激しいことを表している。これより左では DRD の返却時間が最大限に影響している。一方、DRD の返却時間が影響しなければ、到着率にあまり関係なく、処理されている。

図4には、ステージング方法が **s, c, p** および磁気ディスク装置だけの場合を想定し、またデータセットを 100, 50, 10% 利用したときの平均処理時間を示した。さらに、図5にはステージング方法を **p** として、ページ不在要求の発生回数が異なる場合の値を図示した。ページ不在要求の発生回数によって平均処理時間は大きく変化する。図4はページ不在要求の発生回数が最も少ない場合である。これらの計算に用いたパラメータの設定方法を付表1に示した。

100% 利用の場合には **s** で用いたほうが効率がよい。**c** では **s** の処理時間の4倍にも達し、また **p** では約1.2倍となっている。しかし、50% 利用になると、**c** 対 **s** の比は 2.52, **p** 対 **s** の比は 0.81 となり、**p** がよい。さらに、10% 利用では **c** 対 **s** の比は 0.81, **p** 対 **s** の比は 0.80 となり、**c** と **p** が **s** よりもよくなる。

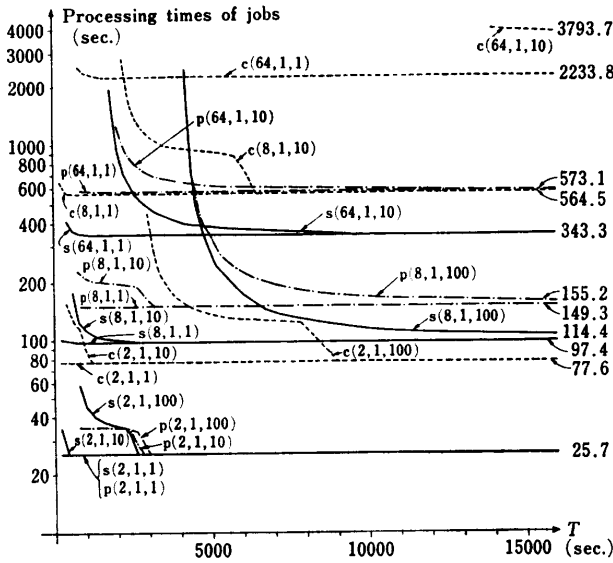
c と **s** では約 14% 利用時に、**p** と **s** では約 71% 利用時に同じ処理時間になる*。データセットの一部しか利用しないときには **c** と **p** が **s** よりもよい。ただし、**c** と **p** のグラフは参照されるデータがデータセット上に最も局所化され、不在要求がいちばん少ない場合を想定して示されている。データが一様に分布している場合にはこの結果よりも悪くなる。たとえば、図5に示した **p** の場合に、不在要求の発生回数が 1.5 になれば、1.0 のときに比べて 50% 利用では処理時間は約 1.3 倍になる。

5.2 データセットの大きさと CPU の 1 回のデータ読み量の比較

次に、データセットの大きさ、CPU の 1 回

* データセットが 64 M バイトの場合には **c** と **s** では 8%, **p** と **s** では 44% の利用で、また 16 M バイトの場合には **c** と **s** では 9%, **p** と **s** では 48% の利用で同じ処理時間になる。

図5 ページ不在モードにおけるページ不在要求の発生回数による平均処理時間
Fig. 5 Average processing times in the case of altering page fault rate under page fault mode.



s(x,y,z): stage mode x: capacity of data sets (MB)
 p(x,y,z): page fault mode y: capacity of data read by CPU (KB)
 c(x,y,z): cylinder fault mode z: number of jobs

図6 データセットの大きさ、CPUの読み込み量、ジョブ数を変えた場合の平均処理時間

Fig. 6 Average processing times in the case of altering the capacity of data sets and data read by CPU, and the number of jobs.

のデータ読み込み量、および同時処理のジョブ数を変えたときの処理時間を検討する。この場合にもステージング DASD の容量は十分であるとする。パラメータの設定方法を付表2に示した。図6がおもな結果である。ただし、CPUの1回のデータ読み込み量を変えても結果にほとんど差がなく、この量を変化させたグラフを示していない。

各モードにおいて、DRDの返却時間が影響しなければジョブ数を変えても、またCPUの1回のデータ読み込み量を変えても、図4と比べて各グラフにそれほど著しい現象は観測されない。データセットの大きさが非常に大きな因子で、cの64Mバイトでは2,000秒以上も要している。しかし、データセットの大きさと処理時間は係数1の傾きで比例せず、データセットの大きさの増加に比べて処理時間は相対的に短い。そ

表3 データセットの大きさに対する平均処理時間の比率
 Table 3 Rate of average processing times concerned with the capacity of data sets.

Mode	Stage	Cylinder fault	Page fault
8MB: 2MB	0.47	0.91	0.73
64MB: 2MB	0.42	0.90	0.70
64MB: 8MB	0.88	0.99	0.96

の比率を表3に示した。すなわち、データを分割して複数のデータセットに格納するよりも、一つのデータセットに格納するほうが、すべてのデータを参照する場合には効率が良い。また、cとs、pとsの処理時間の比は2Mバイトで3.02と1.00、8Mバイトで5.80と1.52、64Mバイトで6.51と1.67になり、sが相対的に優れ、それはデータセットが大きくなるほど顕著である。ただし、MSSの一つのカートリッジの容量は50Mバイトであり、64Mバイトのデータを格納するには二つのカートリッジが必要になる。したがって、実際にはsの場合に50Mバイトを越えると、もう1回のカートリッジのアクセス動作が発生し、この結果よりも処理時間が若干長くなる。

5.3 ステージング DASD の容量が不十分な場合

30個のジョブが4Mバイトのデータセットを処理する場合に、必要となる容量90ページ分がステージング DASD の容量として提供されておらず、ステージング DASD のデータが置き換えられる状況を想定する*。パラメータの設定方法を付表3に示す。

ステージング DASD 上のデータがどのように置き換えられるかを予想することは困難である。ここでは、データの置換回数が必要となる容量とステージング DASD の容量の比に依存するとして、この値をさらに重み付けして置換回数を設定する。すなわち、 y はステージング DASD の容量で $0 < y \leq 90$ を満たし、 $90/y$ に依存すると考えるが、 y が90の場合にちょうど必要な容量となり、実際には置換は発生しない。このために、 $(90/y-1)$ と修正してこれに重み係数を乗じた値を置換回数とする。

y を9, 30, 45, 60ページの場合に分け、実データの重み係数 a が0.5, 1, 2、マウント時の制御データの重み係数 b が0.1, 0.5, 1の場合について算定した。 y が9の場合の結果を図7に示したが、他の場合もおおよそ同じ傾向である。付表4には y を変えたときのおもな値を示した。

各処理時間はステージング DASD の容量にそれほど依存せず、ステージング DASD が十分な場合と同

* 1ジョブの実データは2ページを要し、マウント時の制御データ1シリンダに対して1ページを必要とする。したがって、1ジョブの処理には3ページが必要になり、総計で90ページとなる。

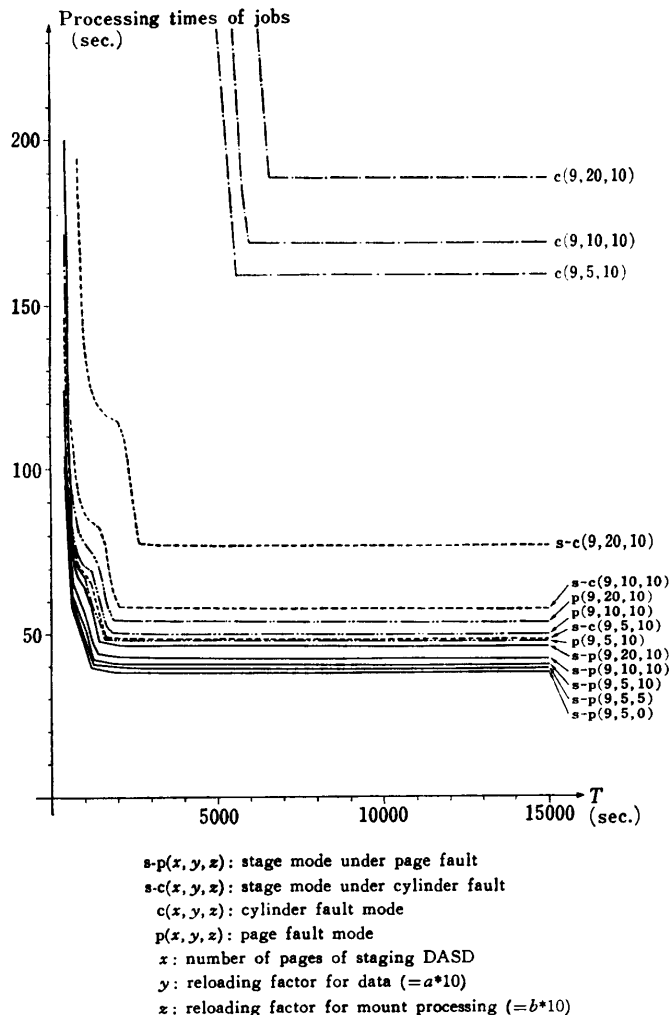


図7 ステージング DASD の容量が不十分な場合の平均処理時間
Fig. 7 Average processing times in the case that the capacity of staging DASD is insufficient.

じ傾向を示す。ただし、ステージング要求が置換によって増えた分だけ、処理時間の伸びが各モードに現れている。また、ステージング方法が同じ **s** であっても、不在要求時の転送モードが **p** (図7の **s-p**) の場合のほうが、**c** (図7の **s-c**) よりもよい。ステージング DASD の容量が十分な場合でも、また不十分な場合でも **p** のステージング方法は有効であり、とくにデータセットが一部しか利用されないときは **p** がよい。

6. おわりに

われわれは MSS の利用がジョブの処理時間にどのように影響するかを調べるために、MSS の解析モデルを構成して種々の処理状況を解析した。その結果、CPU の 1 回のデータ読み込み量にあまり関係なく、扱

うデータセットの大きさに依存すること、データセットの 7~9% 利用時にはシリンダ、およびページ不在モードが、38~45% 利用時にはページ不在モードがステージ・モードのステージング方法よりも有効であることなどがわかった。また、ステージング DASD の容量が不十分な場合にもその容量にそれほど大きく依存しないことがわかった。もちろん、ここで検討したのは限られた状況である。データベースのように一つのデータセットを複数の利用者が扱うという状況を想定していない⁴⁾。

MSS の解析モデルは MSS の動作実験を踏まえて構成され、またパラメータの設定も測定データから検討された。各サービス時間はシステムの構成・環境によって多少異なるかもしれないが、本稿で述べた MSS の解析モデルではパラメータ値を変えるだけでそのまま適用する。このモデルでは SDC を中心に考えて、SDC 系と DRD 系を構成し、ステージング DASD を構成する磁気ディスク装置の台数を考慮していない。われわれが行った実験では台数 (容量ではない) の差がジョブの処理時間に影響することが確認できた³⁾。それは磁気ディスク装置の読書き用のヘッドの機械的なサービスに関係するからである。より一般的には、台数も考慮する必要があるが、モデルが複雑になるだけで、本稿で述べた結果に比べて非常に有効な結論が得られるわけではない。われわれのモデルで

は、 $E_{s1}(t1)$ を装置台数で補正すれば、このような状況にも適用可能である。当然 $E_{s2}(t1)$ も同様に考える必要がある。したがって、われわれがここで述べた結果・方法は MSS をジョブの処理時間から検討した場合に、十分 MSS の動作特性を反映している。

謝辞 日頃から、種々の研究の機会にご便宜いただいている丹羽義次センター長、萩原宏研究開発部長に深謝します。また、北川一、島崎真昭助教授を初めとして実験等に協力していただいた柴山守、堀池博巳、小澤義明技官など研究開発部員の方々、および富士通、京大 CE・SE の方々に感謝します。さらに、適切な指摘をいただいた査読者に感謝します。

参考文献

- 1) Misra, P. N.: Capacity Analysis of the Mass Storage System, *IBM Syst. J.*, Vol. 20, No. 3, pp. 346-360 (1981).
- 2) Tsurubo, S., Matsuda, K., Ohminami, M. and Ito, Y.: Mass Storage Systems Performance Analysis Using a Queuing Model, the 3rd USA-JAPAN Computer Conference, pp. 320-324 (1978).
- 3) 堀池, 柴山, 小澤, 渡辺: データベースにおける MSS 利用に関する実験と考察, 全国共同利用大型計算機センター研究開発論文集, No. 2, pp. 129-140 (1980).
- 4) 小澤, 堀池, 渡辺, 星野: 大容量記憶システム(MSS)を用いた情報検索の性能について, 情報処理学会第22回(昭和56年前期)全国大会論文集, 4L-11 (1981).

- 5) 金沢, 飯田, 柴山, 萩原: 大容量記憶システムの利用特性と動作解析, 情報処理学会: 計算機システムの解析と制御研究会 13-2, p. 9 (1981).
- 6) 国友義久: オンラインネットワークの構造的設計-最適化ネットワークの理論と手法, p. 236, 近代科学社, 東京 (1978).
- 7) FACOM OS IV/F 4 MSS 解説書, MSS 運用手引書, 富士通 (1980).

付録

DRD の返却時間 (d) は次のように決定できる。この場合, いま計算しようとする T の値を T_i とする。そして, それより前の T を T_{i-1} とする。また, T_i についての $\lambda_i, d, E_i(t1), E_i(t2)$ の値をそれぞれ $\lambda'_i, d_i, E'_i(t1), E'_i(t2)$ とする。さらに, n をステージング要求, デステージング要求, マウント要求の回数の和とする。

- 1) $0 \leq \lambda'_i \leq 4/T_i$ ($0 \leq n \leq 4$) のとき
 $d_i = 0.0$
- 2) $\lambda'_i \geq 5/T_i$ ($n \geq 5$) のとき
 $d_0 = 5.000 \times (n-4)/n$: d の初期値
 i) $1/\lambda'_i \geq \{E_i^{-1}(t1) + E_i^{-1}(t2)\}$ ならば,
 $d_i = 0.0$
 ii) $1/\lambda'_i < \{E_i^{-1}(t1) + E_i^{-1}(t2)\}$ ならば,
 (イ) $\{E_i^{-1}(t1) + E_i^{-1}(t2)\} - 1/\lambda'_i > d_{i-1}$ であれば,
 $d_i = d_{i-1}$
 (ロ) $\{E_i^{-1}(t1) + E_i^{-1}(t2)\} - 1/\lambda'_i \leq d_{i-1}$ であれば,

付表 1 ステージング方法の比較時のパラメータ
Table A.1 Parameters on comparison of staging methods.

		ステージ・モード (s)	シリンダ不在モード (c)	ページ不在モード (p)	磁気ディスク装置 (r)
100% 利用	$E_{s1}(t1)$	0.655*17/2	0.655	0.655*17/3	0.0
	$E_{s2}(t1)$	0.005*4	0.005*4	0.005*4	0.005*4
	λ_1	30.0/T	30.0*16/T	30.0*2/T	0.0
	λ_2	0.0	0.0	0.0	0.0
	λ_3	30.0/T	30.0/T	30.0/T	0.0
	λ_4	15000/T	15000/T	15000/T	15000/T
	$E_s(t2)$	7.414+d	7.414+d	7.414+d	0.0
	$E_{total}(t)$	502*E(t)	517*E(t)	503*E(t)	500*E(t)
50% 利用	$E_{s1}(t1)$	0.655*17/2	0.655	0.655*9/2	0.0
	$E_{s2}(t1)$	0.005*4	0.005*4	0.005*4	0.005*4
	λ_1	30.0/T	30.0*8/T	30.0/T	0.0
	λ_2	0.0	0.0	0.0	0.0
	λ_3	30.0/T	30.0/T	30.0/T	0.0
	λ_4	7500/T	7500/T	7500/T	7500/T
	$E_s(t2)$	7.414+d	7.414+d	7.414+d	0.0
	$E_{total}(t)$	252*E(t)	259*E(t)	252*E(t)	250*E(t)
10% 利用	$E_{s1}(t1)$	0.655*17/2	0.655	0.655*9/2	0.0
	$E_{s2}(t1)$	0.005*4	0.005*4	0.005*4	0.005*4
	λ_1	30.0/T	30.0*1.6/T	30.0/T	0.0
	λ_2	0.0	0.0	0.0	0.0
	λ_3	30.0/T	30.0/T	30.0/T	0.0
	λ_4	1500/T	1500/T	1500/T	1500/T
	$E_s(t2)$	7.414+d	7.414+d	7.414+d	0.0
	$E_{total}(t)$	52*E(t)	52.6*E(t)	52*E(t)	50*E(t)

注1) $E_{total}(t)$ は1要求の平均応答時間 $E(t)$ にすべてのデータを読むのに必要な CPU 読み込み回数 (それぞれ, 500, 250, 50), ステージング要求回数, マウント要求回数(1)を積算した値である。たとえば, シリンダ不在モード 100% 利用では, CPU 読み込み回数が 500, ステージング要求回数が 16, マウント要求回数が 1 で合計 517 の積算が必要である。

注2) ページ不在モードにおいては, ページ不在の回数を 50, 10% 利用では 1 回としているが $x(1 \leq x \leq 2)$ 回であれば, 50% 利用で $E_{s1}(t1) = 0.655 * (8x+1)/(x+1)$, $\lambda_1 = 30.0 * x/t$, $E_{total}(t) = (251+x) * E(t)$ となる。10% 利用でも同様に考える。また, シリンダ不在モードでも同様に考える。

付表 2 データセットの大きさ, CPU の読み込み量の比較時のパラメータ

Table A.2 Parameters in the case that the capacity of data sets and data read by CPU change.

	ステージ・モード (s)	シリンダ不在モード (c)	ページ不在モード (p)
$E_{s1}(t1)$	0.055*(X+1)/2	0.655	0.655*(X+1)/(X/8+1)
$E_{s2}(t1)$	0.005*Y	0.005*Y	0.005*Y
λ_1	Z/T	XZ/T	(X/8)Z/T
λ_2	0.0	0.0	0.0
λ_3	Z/T	Z/T	Z/T
λ_4	(125XZ/Y)/T	(125XZ/Y)/T	(125XZ/Y)/T
$E_s(t2)$	7.414+d	7.414+d	7.414+d
$E_{total}(T)$	(2+125X/Y)*E(t)	(X+1+125X/Y)*E(t)	(1+125X/Y+X/8)*E(t)
X シリンダ	8,64,256	8,64,256	8,64,256
Y * 2 kB	1,2,8,16	1,2,8,16	1,2,8,16
Z 個	1,10,100	1,10,100	1,10,100

注) X は 1 ジョブ当りの転送シリンダ数, Y は CPU のデータ読み込み量 (2k バイトの倍数), Z は同時処理の MSS 利用のジョブ数を表す。したがって, ステージ・モードでは X=64 のときにはデータセットの大きさが 16M バイト, Y=4 のときには CPU の 1 回のデータ読み込み量が 8k バイトである。

付表 3 ステージング DASD の容量が不十分な場合のパラメータ
Table A.3 Parameters in the case that the capacity of staging DASD is insufficient.

	ステージ・モード (s-p) (転送単位がページ)	ステージ・モード (s-e) (転送単位がシリンダ)	シリンダ不在モード (e)	ページ不在モード (p)
$E_{s1}(t1)$	$\frac{0.655 * (510 + (8a+b)(90/y-1))}{60 + (a+b)(90/y-1)}$	$\frac{0.655 * (510 + (8a+b)(90/y-1))}{60 + (8a+b)(90/y-1)}$	0.655	$\frac{0.655 * (510 + (8a+b)(90/y-1))}{90 + (a+b)(90/y-1)}$
$E_{s2}(t1)$	$0.005 * 4$	$0.005 * 4$	$0.005 * 4$	$0.005 * 4$
λ_1	$(30 + a(90/y-1))/T$	$(30 + 8a(90/y-1))/T$	$(30 * 16 + 8a(90/y-1))/T$	$(30 * 2 + a(90/y-1))/T$
λ_2	0.0	0.0	0.0	0.0
λ_3	$(30 + b(90/y-1))/T$	$(30 + b(90/y-1))/T$	$(30 + b(90/y-1))/T$	$(30 + b(90/y-1))/T$
λ_4	15,000/T	15,000/T	15,000/T	15,000/T
$E_s(t2)$	$7.414 + d$	$7.414 + d$	$7.414 + d$	$7.414 + d$
$E_{total}(t)$	$(502 + (a+b)(90/y-1)/30) * E(t)$	$(502 + (8a+b)(90/y-1)/30) * E(t)$	$(517 + (8a+b)(90/y-1)/30) * E(t)$	$(503 + (a+b)(90/y-1)/30) * E(t)$
y	9, 30, 45, 60	9, 30, 45, 60	9, 30, 45, 60	9, 30, 45, 60
a	0.5, 1.0, 2.0	0.5, 1.0, 2.0	0.5, 1.0, 2.0	0.5, 1.0, 2.0
b	0, 0.1, 0.5, 1.0	0, 0.1, 0.5, 1.0	0, 0.1, 0.5, 1.0	0, 0.1, 0.5, 1.0

注) y はステージング DASD をページ数で, a は再転送されるデータに対する重み, b は再転送されるマウント・データに対する重みである。

付表 4 ステージング DASD の容量が不十分な場合の平均処理時間
Table A.4 Detailed results in the case that the capacity of staging DASD is insufficient. (単位: 秒)

T	y	a	b	Stage mode: page (s-p)			Stage mode: cylinder (s-e)			Cylinder fault mode (e)			Page fault mode (p)		
				1,000	2,000	10,000	1,000	2,000	10,000	1,000	2,000	10,000	1,000	2,000	10,000
9	0.5	0.0	0.0	45.24	38.47	37.88	64.11	46.26	45.67	—	494.23	156.91	63.46	45.89	45.30
			1.0	51.78	40.91	40.31	68.36	48.70	48.09	—	537.86	159.33	67.69	48.33	47.72
	1.0	0.0	50.16	40.44	39.79	81.77	56.19	55.36	—	740.23	166.61	66.53	47.86	47.20	
		1.0	56.77	42.88	42.21	86.56	62.64	57.78	—	851.02	169.03	70.82	50.30	49.62	
30	2.0	0.0	0.0	59.78	44.39	43.59	129.99	109.47	74.73	—	—	186.01	72.84	51.81	51.00
			1.0	63.92	46.83	46.01	138.84	114.40	77.15	—	—	188.44	77.24	54.25	53.42
	0.5	0.0	0.0	41.44	36.94	36.41	46.13	38.67	38.14	—	396.54	149.37	61.09	44.36	43.82
			1.0	42.52	37.38	36.83	51.96	40.84	40.29	—	420.00	151.52	61.76	44.80	44.24
45	2.0	0.0	0.0	44.69	38.25	37.67	62.25	45.17	44.59	—	477.19	155.83	63.12	45.67	45.09
			1.0	40.90	36.73	36.20	43.23	37.59	37.06	—	385.82	148.29	60.76	44.14	43.61
	0.5	0.0	0.0	41.44	36.94	36.41	46.13	38.67	38.14	—	396.54	149.37	61.09	44.36	43.82
			1.0	42.52	37.38	36.83	51.96	40.84	40.29	—	420.00	151.52	61.76	44.80	44.24
60	0.5	0.0	0.0	40.63	36.62	36.09	41.79	37.05	36.52	—	380.69	147.76	60.59	44.03	43.50
			1.0	40.90	36.73	36.20	43.23	37.59	37.06	—	385.82	148.29	60.76	44.14	43.61
	2.0	0.0	0.0	41.44	36.94	36.41	46.13	38.67	38.14	—	396.54	149.37	61.09	44.36	43.82
			1.0	42.52	37.38	36.83	51.96	40.84	40.29	—	420.00	151.52	61.76	44.80	44.24

注) 「—」はこの時間内では処理できないことを表す。

$$d_i^0 = \{E_i^{-1}(t1) + E_i^{-1}(t2)\} - 1/\lambda_i^j;$$

$$k=0;$$

do until ($d_i^k = d_i^{k-1}$);

$$k=k+1;$$

$$d_i^k = \{E_i^{-1}(t1) + E_i^{-1}(t2)\}_{k-1} - 1/\lambda_i^j;$$

end;

$$d_i = d_i^k;$$

ここで, $\{E_i^{-1}(t1) + E_i^{-1}(t2)\}_k$ は d_i^k について

「 $E_i^{-1}(t1) + E_i^{-1}(t2)$ 」を計算した値である。

(昭和 57 年 4 月 5 日受付)

(昭和 58 年 4 月 19 日採録)