ショートノート

ディスク負荷を評価するための実用的な尺度†

末永 正†† 景川耕宇††† 藤村直美††

大規模な計算機システムにおいて，数多くのディスク装置の中から，入出力（IO）の負荷バランスを欠いた装置をみつけるのは，煩わしく，システム管理者の経験とカンに負うところが大きい．本論文では，待ち行列理論で用いられる平均客数によって，ディスクのIO負荷を評価する方法を提案する．この方法によれば，従来，ディスクのビジー率，IO頻度，IO応答時間などを基に人間が行っていた判断を，大部分機械化でき，しかも評価のための測定オーバヘッドも削減できる．

1. まえがき

近年，汎用計算機を中心とするシステムの規模拡大は著しく，ディスク装置が数百台，あるいは数千台になることもある．こうした大規模な計算機システムを効率よく運転するためには，入出力負荷（以下，単に負荷と書く）に関するチューンナップ作業（IOチューニング）が必須である．通常，このIOチューニングが必要とされるのは，①ビジー率の高いチャネル，② IOの競合が激しいディスク・コントローラ，③応答性の悪いディスクに対してである．

①および②の負荷については，許容ビジー率等の基準値[1]があり，一意に評価できる．それに対して，③の評価については，決まった方法がなく，システム管理者が，ディスクのビジー率，IO頻度，IO応答時間などを参考に，ある程度，経験とカンで行っている．しかし，このような方法では，計算機システムの規模が大きくなり，対象となるディスクの数が多くなると，現実的に対応できなくなる．しかも，ディスク・キャッシュや半導体ディスクなどの新しい装置が従来のディスクに混在して使われるようになると，それらを含めて適切に判断することは極めて難しい．また，このためのエキスパート・システムを構築するにしても，これらはまだ研究段階にあり[2]，実用化には，収集データの充実，および分析技術の確立など，残された課題も多い．

以下では，こうした大規模な計算機システムにおいて，ディスク，ディスク・キャッシュ，半導体ディスクなどを一体として，負荷バランスを計測し評価するための新しい尺度について提案する．

2. 評価法

2.1 尺度

計算機システムの動作分析には，待ち行列モデルをつくり，それを解析する方法が知られている[3]．ここでは，ディスクの負荷分析に関する問題を，待ち行列モデルの隘路分析に置き換えて論じることにする．しかしながら，現実の処理方式，例えば，チャネル，ディスク・コントローラおよびディスクの制御は，逐次的ではないため，これらを待ち行列モデルで忠実に表現し解析することは極めて困難である．したがって，ここでは，チャネル，ディスク・コントローラ，およびディスクの処理を1つのサーバ内の処理として単純化し，個々のサーバに個別の待ちキューが存在するものとする．また，各サーバにおける客のサービス時間は指数分布をなすものと仮定する．さらに，以下では，計算機利用者に入れ代わりがない閉鎖型モデルを仮定するが，開放型モデルを想定しても以下の理論は一般性を失わない．

待ち行列理論によると，各サーバが先着順にサービスし，1つのサーバでは，すべての客が同じ指数分布サービス時間をもつ閉鎖型モデルの場合，隘路になるサーバは，全サーバの相対利用率（relative utilization）λ/μ（λは相対到着率（relative arrival rate），μは平均サービス時間（average service time）を表す）のうち，最大の値をもつ[4]．また，全サーバの相対利用率が近似的に等しければ，隘路は生じない[4]．

このことから，ディスクの負荷バランスを実現するためには，IOサーバ間の相対利用率を均衡化すれば

† Practical Measure for Disk-load Evaluation by TADASHI SUENAGA (Educational Center for Information Processing, Kyushu University), KOU KAGEKAWA (Computer Center, Kyushu University) and NAOMI FUJIMURA (Educational Center for Information Processing, Kyushu University).

†† 九州大学情報処理教育センター
††† 九州大学大型計算機センター

よいことになる．しかし，この場合の IO サーバは，チャネル，ディスク・コントローラ，およびディスクにおける処理が複雑にからんでおり，その平均サービス時間を計測することは，不可能に近い．したがって，この相対利用率によって直接判断することはできない．なお，IO サービスの相対利用率に類似したものとして，ディスクのビジー率が考えられる．しかし，このビジー率については，

- ディスク単独の利用率であるため，入出力経路（パス）の競合によって，利用率が低下したディスクの負荷を正しく評価できない．
- 上限が 100% であり，負荷が高くなるにつれて，得られる値の変化が緩慢になる．そのため，ビジー率の高いディスク間の比較が難しい．

といった問題がある．

閉鎖型モデルにおいて，あるサーバが隘路になるということは，システム内の客数が増加した場合，そのサーバのキューがそれに比例して増加し，他のサーバのキューがほとんど増加しないことを意味する[4]．したがって，そのサーバの平均客数（system size）は，各サーバの平均客数の中で最も大きな値になり，相対利用率と同様の性質をもつ．また，各サーバの平均客数に著しい差がなければ，そのシステムの負荷バランスに問題がないことが推察できる．

以上のことから，ここでは，各 IO サーバにおける平均客数を，相対利用率の代替尺度として用いる方法を提案する．待ち行列における平均客数は，測定ツールによって直接計測することもできるが，リトルの公式[5]（平均客数＝到着率×平均待ち時間）から，次式で計算することもできる．

$$g = F \cdot R \qquad (1)$$

ここで，g は平均客数，F は IO 頻度（回/秒），R は IO 平均応答時間である．

ところで，開放型モデルの場合には，閉鎖型モデルでの相対利用率，相対到着率はそれぞれ利用率，到着率になり，次のことがいえる．開放型モデルにおける隘路は，各サーバに対する到着率を一定の割合で増加させたとき，利用率が最初に 1 に達し，その平均客数が無限大になるサーバということになる．この場合，客のサービス時間を指数分布と仮定すると，平均客数の一番大きなサーバが隘路になりうるサーバとなり，閉鎖型モデルと同様な議論ができる．

2.2 特　　徴

平均客数（g 値）を用いた評価は，主に，次のような特徴をもつ．

i) システムの隘路になるにつれて，g 値が急激に増加するため，負荷の高いディスクに対する評価性能がよい．

ii) g 値を算出するためのデータ（IO 応答時間および IO 頻度）は，IO の発行時，およびその完了時を契機として収集されるため，サンプリング方式のソフトウェア測定ツールによって得られるビジー率等に比べて，精度がよく，しかも測定自身のオーバヘッドを少なくできる．

iii) g 値の計測は，その測定方式から，比較的短い時間（5 分程度）でも十分である．しかも測定時のオーバヘッドが少ないことから，システムの動作を常時観測でき，動的な負荷の変化を把握できる．

iv) 測定のための特別なデータ収集プログラムを開発する必要がなく，既存の測定ツール（例えば，計算機メーカ提供の効率測定ツール）を利用できる．

3. 適 用 結 果

図 1 に，九州大学大型計算機センターの FACOM M 382 システムで測定したディスクの負荷状況を示す．測定時のディスク台数は，ディスク 112 台（1 台の装置が複数の論理装置に分れる場合，それぞれを 1 台と数えた）であった．図 1 は，各ディスクの IO 頻度および IO 応答時間を対数スケールでプロットしたものである（ただし，IO 頻度が 1 秒あたり 1 回未満のものは除いた）．また，**表 1** は，このときの各ディスク負荷を(1)式（g 値）によって評価し，上位 6 つ（全ディスクの 5%）を列挙したものである．なお，参考のために，従来の効率評価ツールによって得られる値（IO 頻度，IO 応答時間，ビジー率，待ちキュー（サービス中のものを含まないキュー）の平均長）とその値による評価順位も表示した．

g 値による評価は，図 1 において，直線 a に平行な，各点を通る直線の原点からの距離によって評価することに等しい．得られた結果に対しては，

- 1 位のディスクが OS の一機能（ファイルの階層管理）の効率を低下させていること．
- 3，4，6 位のディスクが動的なファイル割当ての集中（割当て可能なディスク台数の不足）のためであること．

が分り，これを改善することができた．ただし，第

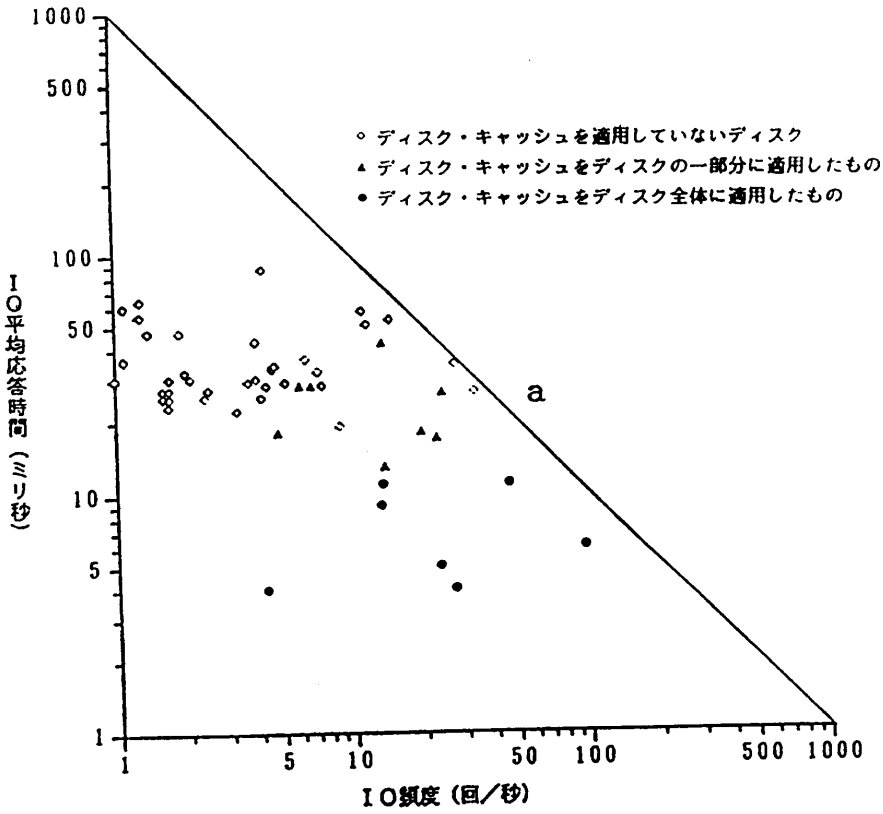


図 1 ディスクの IO 負荷
Fig. 1 IO loads of disks.

表 1 g 値によるディスク負荷の評価
Table 1 Disk load evaluation based on value g.

順位	評価値 g	IO 頻度	平均応答時間（ミリ秒）	デバイス・ビジー率 (%)	待ちキュー平均長
1	0.94	27.8(4)	34(23)	69.4(2)	0.26(6)
2	0.88	33.7(3)	26(45)	88.7(1)	0.02(28)
3	0.76	14.8(10)	52(7)	48.8(3)	0.32(2)
4	0.64	11.3(16)	57(4)	37.2(4)	0.31(4)
5	0.64	24.5(6)	26(44)	34.0(6)	0.32(2)
6	0.59	11.8(15)	50(8)	33.8(7)	0.31(4)

括弧内は，各項目で評価した時の順位を示す．

2位のディスクについては，一過性の負荷であったため，確認できなかった．表1によれば，このディスクは，待ちキューの平均長が 0.02 と異常に小さく，単一のジョブによって使用されていたと推測される．

これに対して，IO 頻度や IO 応答速度による評価は，図1の点をそれぞれの軸に写像して行うことになる．表1に示された評価順位が g 値による順位と異なっていることに注意されたい．実際に，それぞれの上位 1〜2 のディスクに対して，IO チューニングの可能性を調査したところ，

- IO 頻度の高いディスクでは，既に，ファイルの分散化が図られており，しかもディスク・キャッシュの対象にもなっていることから，応答時間が短い．
- IO 応答時間の長いディスクでは，ページング・ファイルへのアクセスしかなく，その回数も非常に少ない．

であり，他のディスクに比べて，改善の余地を見出せなかった．

一方，ディスクのビジー率による評価では，測定したシステムの負荷が軽く，チャネル競合等の影響が少なかったため，g 値による評価と類似の結果を得た．しかしながら，ビジー率や待ちキューの測定にあたっては，その精度を上げるために，サンプリング周期を短くする必要があった．表1は，サンプリング周期を 10 ミリ秒にし，10 分間の計測によって得たものであるが，その測定に 58 秒もの CPU 時間を消費した．これに対して，g 値による評価では，1.5 秒以下の CPU 時間で必要な情報を得ることができた．

4. むすび

本論文では，待ち行列理論で用いられる平均客数によって，ディスクの負荷バランスを判定する方法を示した．評価に際しては，各サーバ（ディスク）における客が指数分布サービス時間をもつものと仮定した．これを一般的な分布に拡張した場合，ある時点で最大の平均客数をもつサーバが，負荷が増大したときの隘路サーバになるとは限らないため，評価順位の意味は減少する．しかしながら，評価と改善を何度か繰返せば，隘路になるサーバを改善し，負荷を平均化できる．

ところで，負荷の高いディスクをみつけた場合，さらに詳細な分析が必要となる．すなわち，ディスク中のファイル別アクセス形態（頻度，リード／ライトの区別，アクセス位置）などが分らなければ，適切なIO チューニングはできない．しかし，分析すべきディスクを特定できれば，その調査は比較的容易である．筆者らは，本評価法によって全ディスクの中から負荷の高いディスクを選び，そのアクセス内容まで調べる自動負荷分析システム（Unbalanced Disk-load Analyzer：UDA）を開発したが，これについては別稿で報告したい．

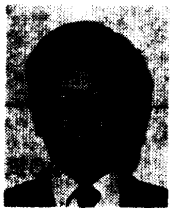
謝辞 本論文作成にあたり，ご教示いただいた九州大学工学部牛島和夫教授に深く感謝いたします．

参考文献

1) Schardt, R. M.: An MVS Tuning Approach, *IBM Syst. J.*, Vol. 19, No. 1, pp. 102-119 (1980).
2) Artis, H. P.: Using Expert Systems for Analyzing RMF Data, *CMS '85. International Conference on the Management and Performance Education of Computer Systems. Conference Proceeding*, pp. 653-657 (1985).
3) 橋田，川島：待ち行列ネットワークモデルによる計算機システムの性能評価，情報処理，Vol. 21, No. 7, pp. 743-750 (1980).
4) Kleinrock, L.: *Queueing Systems Volume II: Computer Applications*, Wiley, N. Y. (1976).
5) Little, J. D. C.: A Proof of the Queuing Formula $L=\lambda W$, *Oper. Res.*, Vol. 9, pp. 383-387 (1961).

（昭和62年11月 2 日受付）
（昭和63年 4 月 14 日採録）

末永 正（正会員）
昭和24年生．昭和47年九州大学工学部電子工学科卒業．九州大学大型計算機センターを経て，現在，九州大学情報処理教育センター助手．オペレーティング・システムの効率，およびマンマシン・インタフェースに関心がある．人工知能学会会員．



景川 耕宇（正会員）
昭和15年生．昭和37年東京大学工学部応用物理学科（数理工学専攻）卒業．昭和39年同大学院修士課程修了．（株）三永通信，（株）新日本パイプを経て昭和42年より九州大学勤務．現在九州大学大型計算機センター講師．主な研究分野：計算機システムの性能評価．

藤村 直美（正会員）
昭和25年生．昭和48年九州大学工学部電子工学科卒業．昭和53年同大学院工学研究科博士課程単位取得退学．同年九州大学工学部助手．昭和56年九州大学情報処理教育センター助教授．工学博士．計算機システムの性能の計測・評価・試験，ソフトウェアの品質・生産性に関心がある．電子情報通信学会，ソフトウェア科学会，ACM，IEEE 各会員．