蓄積転送サービスにおけるディスク容量管理方法の考察

5U－6

関野公彦　小西隆介
ＮＴＴ情報通信研究所

1．背景

電子ソフト流通に代表される情報流通サービスの電子化要求の高まりに伴い、データをファイルとして、ある期間ネットワークホスト上のディスク内に保存し、要求に応じて転送する蓄積転送型サービスの重要性が増してきている。蓄積転送サービスに用いられるファイルは、以下のような特徴を持つ。

(1) ネットワーク通信における一時的な蓄積場所として機能するため、一定の寿命を持ち、また、ファイルの格納場所は、蓄積ホスト上でファイル転送処理を行うプロセス（以降サーバと呼ぶ）が意識するものであり、システム運用者が意識するものではない。

(2) 単純な文書などのサイズの小さなものから、画像や音声などの大きなサイズのものまで含む。

従来のオペレーティングシステム（OS）は、これらの特徴を意識して設計されていないため、必ずしも蓄積転送サービスに対して有用な機能を提供しているとは言えない。蓄積転送型サービスに向けたOS機能の検討課題としては、ディスク容量の管理、リアルタイム性の保証、過負荷時の対処、ディスク使用効率の向上、フラグメンテーションの回避、生成削除の性能向上などが挙げられるが、本稿では、ディスク容量の管理方式に着目する。

2．従来技術

木構造のファイルシステムを持つOSは、従来より複数ディスクを一つのファイルシステムとして見せることにより、ディスクの管理を行ってきた。また、ファイルシステムとそれの格納媒体を分離することにより、ディスクが過負荷状態になったときのディスク増設を容易にする研究もされている［1］。また、UNIXでは、ディスクの過負荷状態の管理として、利用者の使用可能な領域をquotaとして設定し、過負荷状態を利用者に通知する機能を提供している［2］。しかし、これらには、以下のような問題点がある。

(1) 負荷の偏り

一つのディレクトリ配下のファイルに対する書き込みが集中したときなど、使用ディスクの偏りが起こる可能性がある。

(2) 過負過時の通知単位

quotaの単位はディスク単位であるため、使用できる複数のディスクが全体として過負荷なのか、ディスクの使い方が偏っているため過負荷状態なのかが判断できない。使用できるディスクの中には、空き容量があるディスクがあるにも関わらず、負荷が一つのディスクに集中することにより過負荷通知されることがある。

(3) 転送失敗時のオーバヘッド

同一のディスクに複数のサーバが同時に書き込みを行っている場合、個々のサーバは他のサーバの要求量を知ることができない。従って、書き込み開始時に空き容量が十分でも、書き込み中に容量不足が生じる可能性がある。この場合、書き込み中データの再送／再書き込みが必要となる。データのサイズが大きい場合、再送のオーバヘッドは無視できない。

3．ディスク容量管理機能の実装

以上のような問題点を回避するためには、ホストに収容されている全てのディスクを集中的に管理し、複数のサーバに対して、容量に空きのあるディスクを割り当てていく機能をOSとしてサポートする必要がある。本稿で提案するディスク容量管理機能では、蓄積転送サービスで使用する一時蓄積ファイルの格納場所は運用者が直接意識するものではないことに着目し、空きのあるディスクを、サーバに対して割り当てる機能を提供する。また、割り当て対象となるディスク群の定義／変更機能、過負荷状態の管理機能を持つ。

ディスク容量管理機能は、以下のシーケンスに従って、サーバから使用される。

(1) サーバは、要求容量を入力パラメータとして、ディスク割り当て要求を行う。

(2) ディスク容量管理は、要求容量が空いているディスクを、あらかじめ指定された割り当て戦略に従って検索し、そのディスクをサーバに割り当てる。

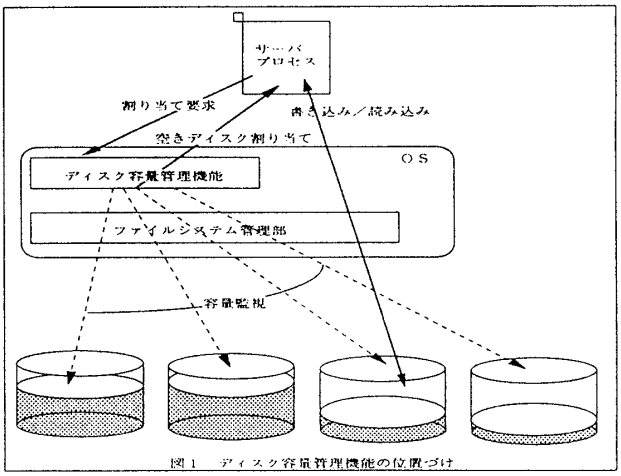

図1　ディスク容量管理機能の位置づけ

Disk Storage Management System for Store-and-Forward Service
Kimihiko Sekino, Ryusuke Konishi
NTT Information and Communication Systems Labs.

(3)割り当ての結果として、使用可能なディスク群全体の空き容量が全体のx%(利用者指定値)以上になった場合には、過負荷メッセージを出力し、寿命を過ぎたファイルの削除を行う。
(4)ディスクを検索する際、要求容量を満たすディスクがなかった場合は、割り当て不可として、利用者に通知する。
ディスク容量管理機能のOS内での位置づけを図１に示す。以下、実装上の検討点について述べる。

３．１　ディスク群の導入

利用者が必要とするディスク容量は、サービス中に変化することがある。また、利用者によって、必要なディスク容量、過負荷の定義は異なる。従って、利用者がこれらの値を定義できるようにすることが望ましい。また、必要なディスク容量、過負荷状態の定義は動的に変更されることもある。そこで、割り当て対象となるディスクの集合を、サービス中に変更可能な「ディスク群」として定義できるようにした。利用者は群定義ファイルを書き換えることにより、ディスクの動的追加や、ディスク群の新設などが可能である。

３．２　リザーブ領域の導入

ディスクの空き容量を書き込み時に監視すると、書き込み性能が悪化するため望ましくない。従って、ディスク割り当て要求があった時点で、動的にディスクの空き容量を取得する方法を採用した。さらに、複数サーバの同期書き込みに伴う前述の問題点に対処するため、各ディスクにリザーブ領域を設け、ディスクが（全領域−リザーブ領域）まで使用された時点でそのディスクの新たな割り当てを抑制することにした。リザーブ領域の最適サイズについては、今後検討が必要だが、実際問題として、リザーブのためにあまり大きな領域をとることはできないものと考えられる。従って、複数の要求が輻輳した場合、転送中のディスク容量不足による転送失敗が生じる可能性がある。

４．割り当て戦略の評価

空き領域のあるディスクの割り当て戦略として、(1)ラウンドロビン、(2)最大空き容量優先、(3)ランダムなどが考えられる。本章では、割り当て戦略の評価を行うため、シミュレーションによって実験した結果を示す。

４．１　実施条件

割り当て戦略としては、上記の三戦略を比較した。評価モデルは、既存のファイル転送システムの値をもとに、以下のように定めた。
ホスト数：1台、ディスク台数：6台（各2GB）、ファイル寿命：2日、リザーブ領域：200MB

評価は、蓄積転送サービスにおける主要なオーバヘッド要因として考えられる、転送中のディスク容量不足による転送失敗数により行うこととした。ファイルサイズは、実際のシステムでは様々な組み合わせとなると考えられるが、今回の実験では、0MBからxMBまでの一様分布を行うものと仮定し、xを変化させて転送失敗数を測定した。xの上限、及び、ファイル到着頻度は、群に属する全てのディスクの容量不足によるディスク割り当ての失敗がないように定めた。

４．２　結果

実験結果を図2に示す。ファイルサイズの変動幅が小さいときは、割り当て戦略による差異はほとんどないが、変動幅が増加すると、最大空き容量優先戦略は、他の戦略と比較して転送失敗数の増加が小さくなる。

４．３　考察

転送失敗は、特に、扱うファイルサイズとして、大きいファイルが存在する場合に生じるものと想定されたが、本実験は上記予想を裏付けている。ラウンドロビンは処理の構築が簡単であり、ディスク割り当てにかかるオーバヘッドは少ないため、ファイルサイズの変動が少ない場合は、ラウンドロビンアルゴリズムの採用が望ましいと考えられる。しかし、画像や音声などサイズの変動幅が大きいファイルの場合は、転送失敗によるデータ再送オーバヘッドを避けるため、最大空き容量のあるディスクを割り当てる制御が有効であると考えることができる。

５．まとめ

本稿では、ディスク容量管理機能について概観し、ディスク割り当て戦略について、ファイルサイズの変動幅が大きい場合に、最大空き容量を持つディスクを割り当てる戦略が有効であることを示した。複数ホスト間のディスク容量管理機能間で通信を行うことにより、ネットワーク上に分散したディスクの割り当てを行う機能の実現は今後の検討課題である。

［参考文献］

［１］John Wilkes, "DataMesh, house-building, and distributed systems technology," Position paper for 5th ACM SIGOPS European Workshop.

［２］S. J. Leffler, M. K. Mckusick, M. J. Karels, J. S. Quarterman, "The Design and Implementation of the 4.3BSD UNIX Operating System," Addison-Wesley, Publishing Company, Inc., 1989.

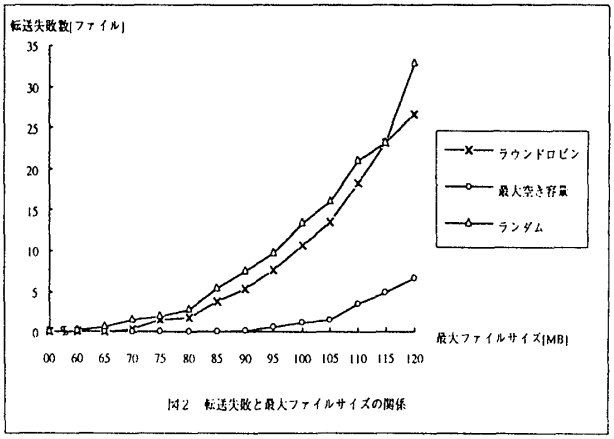

図2　転送失敗と最大ファイルサイズの関係