DCEでのファイルアクセス

The File Transfer to and from the Server on DCE Services

日本歯科大学新潟歯学部　山下陽介
日立情報システムズ　土門哲也
新潟大学　高清水直美
高エネルギー物理学研究所　佐々木節

Yousuke YAMASHITA
The Nippon Dental University, School of Dentistry at Niigata,
Hamaura-cho, Niigata 951, Japan

Tetsuya DOMON
Hitachi Engineering Systems, Dougenzaka, Shibuya-ku, Tokyo 150, Japan

Naomi TAKASHIMIZU
Niigata University, Nino-machi, Ikarashi, Niigata 950-21, Japan

Takashi SASAKI
National Laboratory for High Energy Physics, Oho, Tukuba-shi, Ibaraki-ken 305, Japan

(1996年11月29日　受理)

1. はじめに

高エネルギー実験におけるコンピューティングソフトウェアの開発環境は，大規模・複雑化するソフトウェアの開発を余儀なくされてきている。そのため，開発には，世界中に分散する非常に多くの研究者達がたずさわってきている。このような環境においては，ソ

フトウェアの開発や配布バージョンの同期が正しく容易に行えることが不可欠である。NFSなど従来の分散ファイル・システムは，スケーラビリティ，セキュリティ，管理の容易さなどの面で問題を持っており，DCE(Distributed Computing Enviorment)がこれらの要求に答えることができる開発環境を与えると考えられている。

DCEとは，OSF(Open Software Foundation)により統合化・標準化作業が行われている，異機種混在における大規模分散環境を構築するためのソフトウエアパッケージのことであり，特に，分散ファイル・サービス(DFS)は，ネットワーク上に分散しているファイルを一つのまとまったファイル・システムとして扱うサービスを提供するDCEの中核となるアプリケーションである。ここでは，DCEの実際の適用についての基礎的な知見を得るために，ファイル転送について，テストしたデータを示して，DCEの適用性について考察する。

2．ATM(Asynchronous Transfer Mode)回線を使用したDFS性能評価

ATM回線は，スイッチング型の高速回線で，セルと呼ばれる53バイトの固定長パケットでデータ転送が行われる。セルの転送のほとんどの処理がハードウエアで行われるため高速なデータ転送を行うことができる。

2.1　テスト環境

テスト環境を説明する。使用したマシンは，SUN S 1000 1台，SUN SS 10 2台，HP 9000/735 1台である。全てのマシンは，ATM回線にFore社のATM Switchにより接続されており，閉じたネットとしてのATM Intranetを構築した。各ベンダーのマシンのみによるセルを作り，それぞれSUN Cell，HP Cellとした。HPのマシンは，1台の構成であるが，これはATM Networkを使用したIntercellでの性能を評価するためである。また，SUNのマシンの一台は，NTTマルチメディア実験線のATM回線を利用して接続されている新潟大学高エネルギー研究室に設置してあるマシンである。これにより，Wide Area Network(WAN)におけるDFSの性能評価も合せて行なった。DFSのシステムパラメータであるキャッシュサイズは100 MB，チャンクサイズは，32 KB-256 KBまでとった。

表 1. テスト環境で使用したマシンの性能

Host	Model	CPU	Clock (MHz)	Memory (MB)	Network Interface	OS
sun 1	S 1000	SuperSPARC	40 x 2	192	ATM	Solaris 2.4
sun 2	Axil-320	SuperSPAR C	75	64	ATM	Solaris 2.4
sun 3	SS 10	SuperSPARC	50 x 2	64	ATM	Solaris 2.4
hp 1	HP 735/99	PA-7100	99	128	ATM	HP-UX 9.x

表 2. テスト環境における各マシンの DCE の役割

	SEC	CDS	DTS	FLDB	DFS	NTP
SUN Cell						
sun 1	Master Server	Master Server	Local Server	Master Server	Server	-
sun 2	Client	Client	Local Server	Client	Client	-
sun 3	Client	Client	Clinet	Client	Client	-
HP Cell						
hp 1	Master Server	Master Server	Local Server	Master Server	Server	-

2.2 テスト方法

性能測定には，iozone ベンチマークプログラムを使用した。iozone は，システム関数を使用してメモリの内容をファイルに書き込み，その後同じファイルからメモリに読みこむことで sequential I/O の性能を測定するプログラムである。結果は，[bytes/sec] として得られ，このパラメータが大きい程性能が優れている。DFS には，Server から read してきたデータを local disk あるいは memory に蓄えて二度目以降の read アクセスを高速に行うキャッシュ機能が備わっている。よって，DFS 上において iozone を実行すると，read の際に DFS のキャッシュにヒットしてしまい，read 時における DFS 本来の転送性能を正確に評価することができない。そこで，iozone のプログラムを改良し DFS のキャッシュをフラッシュした後に，read するようプログラムを改良した。また，一般的な UNIX OS では，kernel レベルでキャッシュ機能を有しており，この効果も測定に少なか

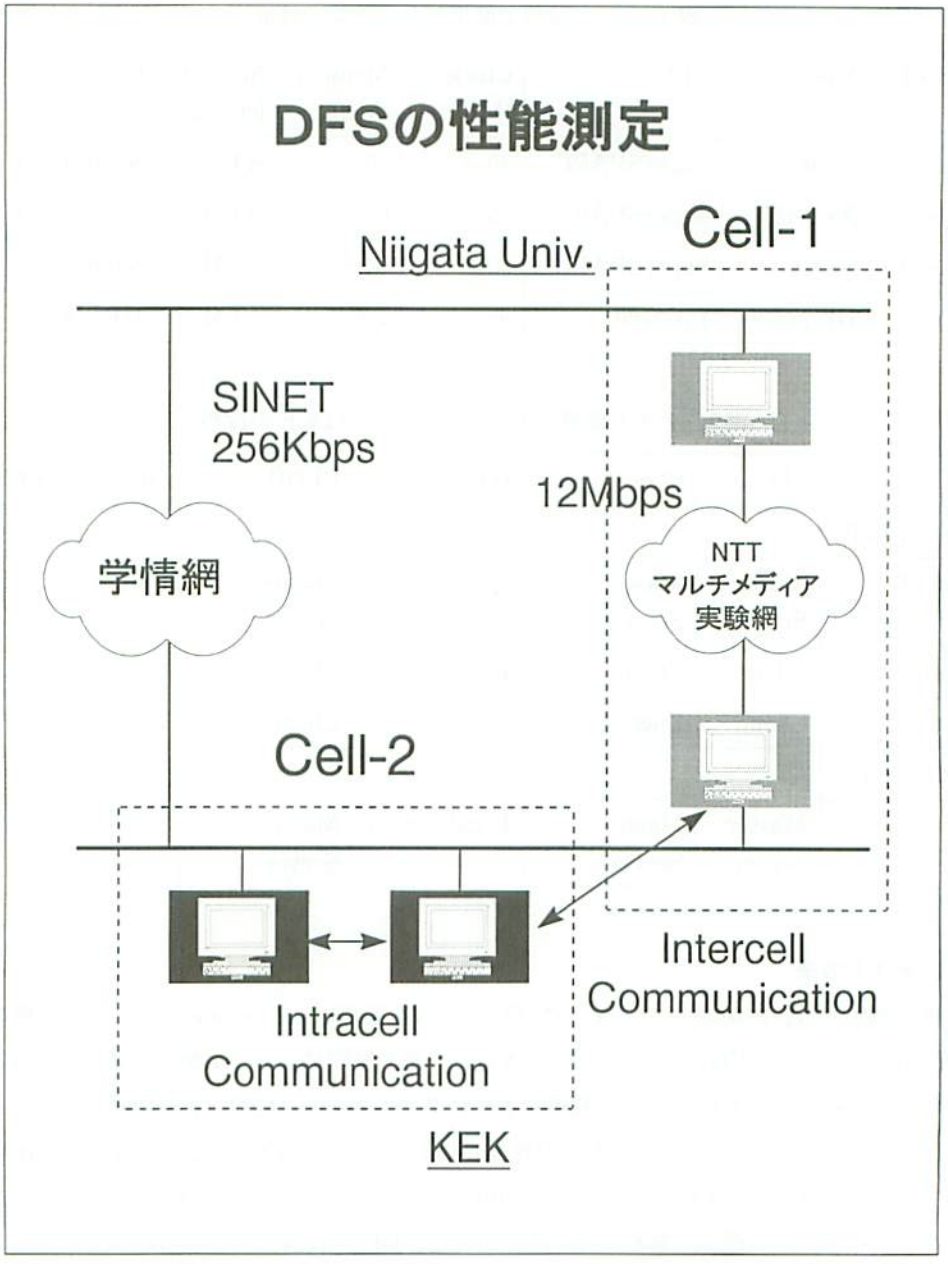

図1. DFSの性能測定における Network

らず影響を与えることが予想されたため，iozoneでのファイルサイズを50 MBとある程度大きなサイズにすることとした。また，DFSのシステムパラメータとしてチャンクサイズと呼ばれるものがある。これは，DFSがサーバ・クライアント間でのデータ転送を行う際のデータサイズである。このチャンクサイズは，デフォルト値64KBで行うこととした。

測定は，ゆらぎ等の測定誤差の影響を少なくするため測定を計5回実施し，上下2つを除いた3つのデータの平均をとって性能評価を行った。

2．3 結果

DFS上で，NFSと同等以上のI/O性能が達成された。NTTマルチメディア実験のATM線上でのDFSの性能は，十分に高い性能を示した。

3．DCE/RPCにおけるファイル転送プログラム

従来のTCP/IPを利用したネットワークプログラミングでは，ホスト名の指定，サー

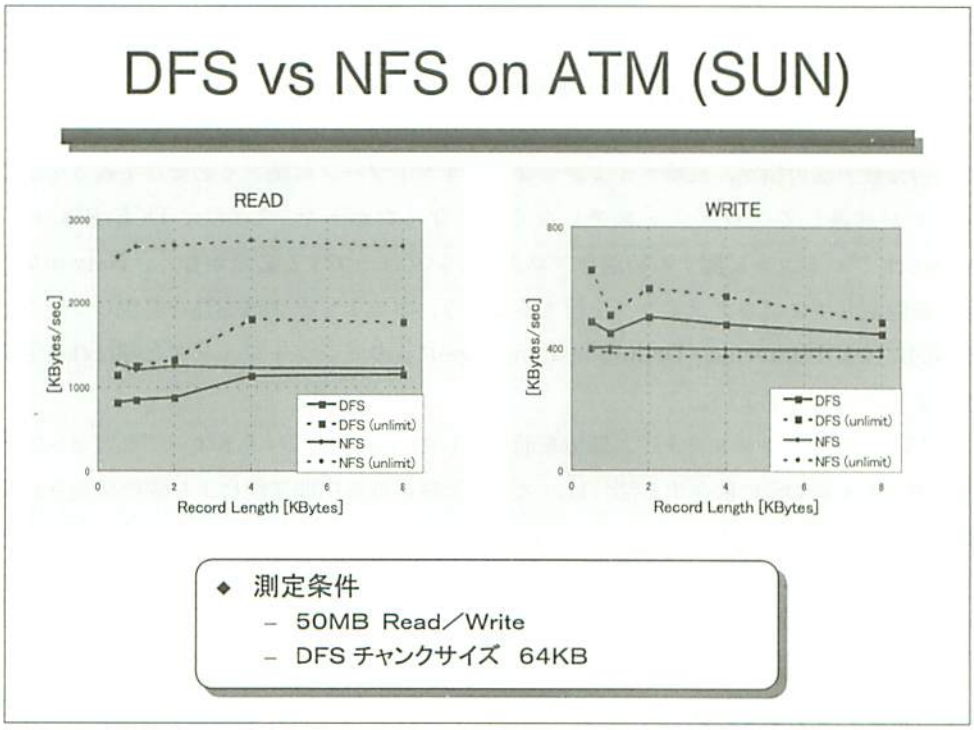

図2．ATM上でのDFSとNFSの性能比較

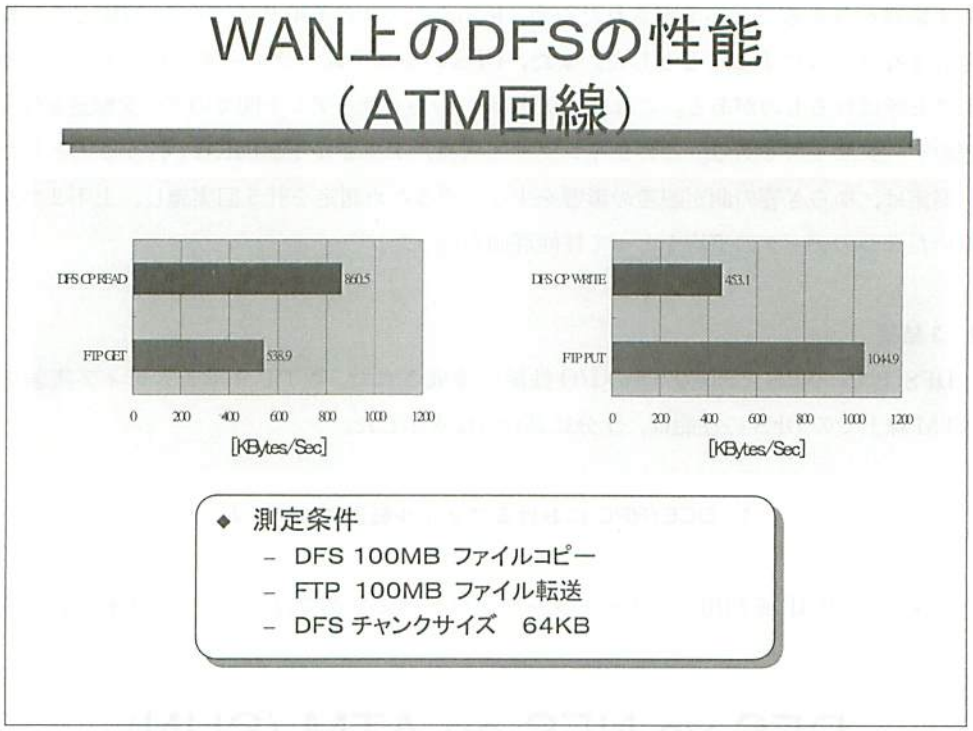

図3．WAN(ATM)上での DFS の性能

バへの接続方法の指定，接続タイミングなど，ネットワークに関する必要な手続きを全てユーザが意識してプログラミングしなくてはならなかった。しかし，DCE/RPC は，Network プロトコルに関する特別なプログラミングは一切する必要がなく，Network の存在を意識せずにプログラミングを行えるという，ネットワーク透過性を実現している。基本的には IDL(Interface Definition Language)によりインターフェースと呼ばれる手続きを定義するだけでよい。

また，マルチプラットフォーム環境を前提としているため，バイトオーダなどマシンアーキテクチャや OS に依存する部分は，スタブと呼ばれる中間言語により変換処理される。よって，OS ごとにプログラム内容を変更するなど，プラットフォームを意識したプログラミングをする必要が全くないのも特徴である。しかし，IDL によるインターフェースの定義など，RPC 独特のコーディングを行わなくてはならない。

3.1 RPC プログラミング

RPC では，IDL によりインターフェースを定義する。インターフェースとは，サーバ

が実行すべき関数(プロシージャ)の集合である。クライアントは，IDLにより定義された関数をコールすると，その関数を提供するサーバにアクセスし，サーバにおいて関数が実行される。RPCプログラミングは，サーバ部とクライアント部の二つに分けてコーディングする。サーバ部は，RPCによる通信を行うための初期設定部と，クライアントからコールされるマネージャルーチンと呼ばれる関数本体からなる。クライアント部は，基本的にはRPC独自の処理は必要なくプログラム中において関数をコールするだけであるが，コーディングの方法や実装する機能に関連した手続きが必要となる。

3.2 プログラム設計・実装

DFSでの大容量ファイルのシーケンシャルアクセスは，キャッシュあふれをおこし，DFSのパフォーマンスに影響をもたらす。そこで，DFSへ影響を与えずに，大容量ファイルのシーケンシャルアクセスを可能にするプログラムを作成することを試みた。今回プログラムを作成するにあたり，以下の点を重視してプログラムを作成した。

○ 高性能データ転送

データ転送の一連の処理において，サーバ，クライアント間でのデータ転送が最もボトルネックになる。そこで，readにおいては，先読み，writeにおいては，バッファードI/Oの機能を組み込んだ。先読みは，一度目のread要求で，あらかじめ確保したバッファサイズの分だけデータを先読みし，二度目以降のデータの読み込みは，バッファから行うというものである。バッファードI/Oは，あらかじめ確保したバッファに一時的にデータを蓄積し，一杯になった時点でサーバにデータを転送するという機能である。シーケンシャルI/Oを行うことが前提ではあるが，これによりデータ転送の回数を減らし，Network使用時でのデータ転送におけるボトルネックを最小限に抑えることができる。

また，RPCには，パイプ転送と呼ばれる大容量向けデータ転送機能が装備されており，この機能も利用した。

○ 標準システム関数との互換性

今までの解析プログラムとの互換性を重視し，なるべくスムーズに移行できるよう，標準システム関数との互換性を保つように設計した。ユーザは，標準システム関数として定義されている，open，read，write等の関数をRPCによって作成した関数dce-open，dce-read，dce-write等と置きかえればよい。各々の関数の引数は，標準システム関数と同じ仕様にした。また，リモートホストの指定方法であるが，ファイルオープンの際に，指定するファイル名の先頭に"ホスト名+コロン(:)"を付けて指定するようにし，ホスト名を省略すれば，自動的にローカルホストが指定されるようにした。

3.3 テスト環境

計算機は，SUN W/S 7台，HP W/S 1台の計8台構成である。また導入したDCEのバージョンは，Transarc DCE 1.1(OSF/DCE 1.1)とHP-DCE 1.4(OSF/DCE 1.1)である。今回測定に用いたマシンは，そのうちのSUN SS-5 2台である。なるべく同じ条件のマシンを使用し，ハードウエア的に測定結果に影響するような要因を少くするようにした。また，2台のマシンは，FDDIにより接続されており，同じネットワークセグメント上に接続されている。DFSのキャッシュサイズは，100 MBディスクキャッシュ，チャンクサイズは，64 KBと固定にした。マシンの詳細な性能とDCEの役割を表に示す。

表2．測定に使用したマシンの性能

host	Model	CPU	Clock (MHz)	Memory (MB)	Network Interface	OS
sun 1	SS-5	MicroSPARC II	85	96	CDDI	Solaris 2.5
sun 2	SS-5	MicroSPARC II	110	96	FDDI	Solaris 2.5

表3．測定に使用したマシンのDCEでの役割

	SEC	CDS	DTS	FLDB	DFS	NTP
SUN Cell						
sun 1	Master Server	Client	Local Server	Client	Server	-
sun 2	Client	Client	Clinet	Client	Client	-

3.4 テスト方法

測定には，DFS性能評価と同様にiozoneベンチマークプログラムを使用した。データサイズを200 MB，レコード長を512 B，64 KB，1 MBと可変させ，1レコード長あたり計5回測定した。測定結果は，上下2つを除いた3つの平均により算出した。

3.5 結果

RPCプログラムによって，DFSに比べて高速なデータ転送が実現できた。またRPCプログラムによるデータ転送の性能は，レコード長に依存せずほぼ一定になっている。

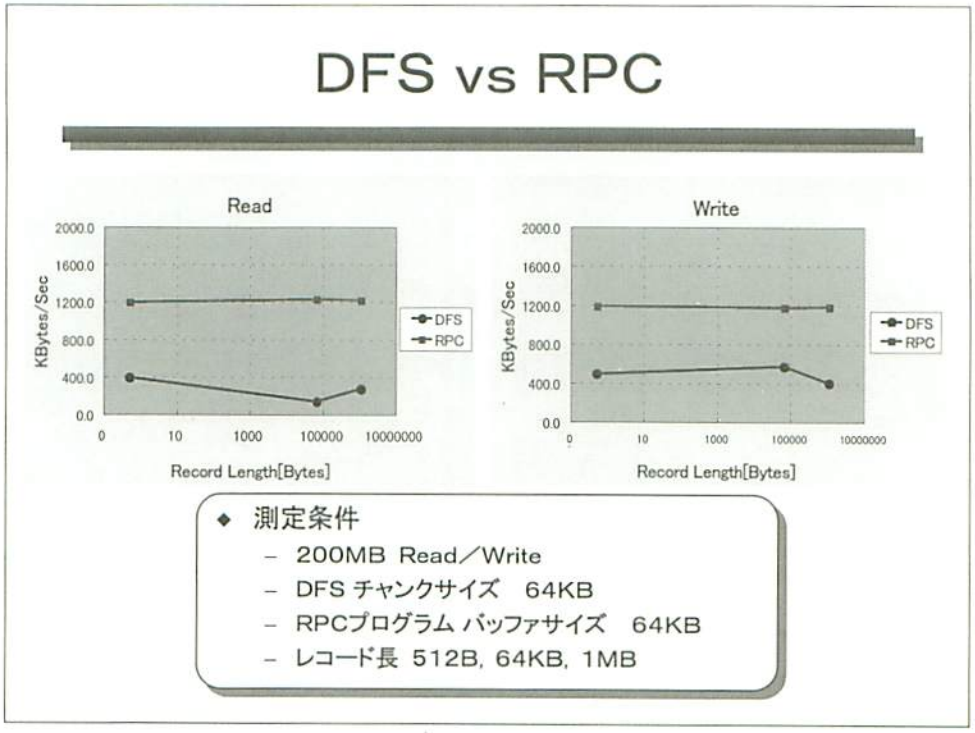

図 4．DFS と RPC を使用したプログラムとの比較

4．まとめ

DCE のシステムに対する適用の基礎データについては，今回のテストを通じて良好な結果が得られたので，実際のシステムに対する運用を通じてより有効な活用について，調べて行く必要がある。

参考文献

1）瓶家 喜代志：OSF DCE 技術解説，SRC,1992.
2）Open Software Foundation：The OSF DCE SERIES, Prentice Hall
3）http://www.osf.org/dce/
4）http://www.opengroup.or.jp/DCE/DCE.html