社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ARC－157　(6)
2004－HPC－ 97　(6)
2004／3／1

グリッドデータファームとGNET-1による日米間高速ファイル複製

建部修見[†]　小川宏高[†]　児玉祐悦[†]
工藤知宏[†]　高野了成[†,††]　関口智嗣[†]

グリッド・データファームは，複数組織による高信頼のデータ共有，高速データアクセス，高速データ処理のための Gfarm ファイルシステムを提供する．現在，日米タイの計 9 拠点からなる環太平洋規模のテストベッドを構築しており，Gfarm ファイルシステムの容量は 68.5 TBytes，ディスク I/O 性能は 13 GB/sec である．本テストベッドを利用して，日米間のファイル複製作成に関する性能評価を行った．

ネットワーク装置 GNET-1 を利用したインター・フレーム・ギャップ (IFG) による流量制限により，およそ 10,000 Km 離れた日米間において，理論ピーク性能に近い通信性能を達成した．流量制限によりパケットロスフリーの広帯域ネットワークが日米間に実現し，安定したネットワークフローを達成することができた．このパケットロスフリーのネットワークを利用した日米間ファイル転送では，1,100 Gbytes のデータを米国の 11 ノードから日本の 11 ノードに転送し，3.78 Gbps の安定したデータ転送レートを実現した．これは理論ピーク性能の 3.9 Gbps の 97%に相当する．

Trans-pacific fast file replication using GNET-1 on Grid Datafarm

Osamu Tatebe,[†] Hirotaka Ogawa,[†] Yuetsu Kodama,[†]
Tomohiro Kudoh,[†] Ryousei Takano[†,††] and Satoshi Sekiguchi[†]

The Grid Datafarm provides a Gfarm file system that federates multiple local file systems in a Grid across administrative domains for dependable file sharing, and high-performance file access and processing. We are constructing a Trans-pacific Grid Datafarm testbed, which provides 68.5 TBytes disk capacity, 13 GB/sec disk I/O performance using 5 clusters in Japan, 3 clusters in US, and 1 cluster in Thailand. Here, we present our experiment and evaluation results of Trans-pacific file replication using the testbed.

Thanks to the rate control by adjusting the inter-frame gap (IFG) by a programmable network testbed device GNET-1, we achieved a stable network flow with maximum performance close to the theoretical peak performance between US and Japan (about 6,000 miles), because the rate control realizes a high-speed packet loss free network. Trans-pacific file replication of 1,100Gbytes data via the packet loss free network provides a stable data transfer rate of 3.78 Gbps out of the theoretical peak 3.9 Gbps using 11 cluster node pairs.

1. はじめに

グリッド・データファーム[14),18)] は，ペタバイトスケールのデータ・インテンシブ・コンピューティングのために設計され，複数組織の PC クラスタに分散配置されたデータの共有と高速処理を目指したものである．

グリッド・データファームの大きな特徴の一つとして，ファイル複製によるファイルアクセスの信頼性の向上と，アクセス性能の向上がある．ディスクやネットワークの故障時でも，どれかのファイル複製にアクセスができれば，問題なくファイルを読み出すことが可能である．また，読み込みが集中するファイルは，適宜別のファイル複製から読み込むようにすることで，性能を落すことなくファイルアクセスが可能となる．

また，近年広域ネットワークの広帯域化が進み，10 Gbps 級の高速ネットワークが日米間，米欧間などの長距離でも利用可能となってきた．しかしながら，文献 9) でも指摘されているように，通常の TCP による高遅延の高速ネットワークの効率的な利用は現実的ではない．日米間の往復のネットワーク遅延は 150 msec ～ 300 msec 程であるが，仮に 100 msec とすると，TCP コネクションのバンド幅が 10 Gbps になるのは 5,000,000,000 パケット送る間（約 1.67 時

† 産業技術総合研究所グリッド研究センター
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology (AIST)
E-mail: {o.tatebe,h-ogawa,y-kodama,t.kudoh, s.sekiguchi}@aist.go.jp
†† (株) アックス
Axe, Inc.
E-mail: takano@axe-inc.co.jp

間）に高々一度輻輳が発生するという頻度である必要があるからである．これに対し，通信プロトコル的な解決を図る研究開発は，HighSpeed TCP[9]，Scalable TCP[7]，FAST TCP[2]，XCP[11]，Tsunami[8]，SABUL/UDT[13]，RBUDP[10]，atou[1]，PSockets[12] など，数多く行われている．これらは，遅延が大きいネットワークに関して，利用可能な最大バンド幅をいかに早く見つけるか，パケットロスが起きたときにいかに早く復帰するか，などに対するプロトコル的，実装的な改良がなされている．

本論文では，ファイル複製作成でも特に重要な長距離高速ネットワークを利用した大規模データの複製作成に注目し議論する．

我々グループは，2002 年の秋，IEEE/ACM SC2002 バンド幅チャレンジ (BWC) に参加し，グリッド・データファームを利用した長距離の高速ファイル複製の実験，性能評価を行った[16]．当時，日米間のネットワーク APAN/TransPAC は，東京–シカゴ間，東京–シアトル間の二本の高速ネットワークで接続されていた．東京–シアトル間は OC-12 POS であり，HighSpeed TCP を利用した日米間のネットワーク転送実験では，1 ノードのペアで HighSpeed TCP ストリームを二本利用し 529 Mbps を達成した．パケットロス後のバンド幅の復帰もスムーズでほぼ問題ないかに思われた．一方，東京–シカゴ間ではパケットロスが多発し，理論ピーク性能が 271 Mbps のところ，ネットワークフローは振動し 10 分平均で 85.9 Mbps であった．恐らく途中経路のルータでパケット破棄が行われたと考えられる．これに対し，100 Mbps に流量を制限した二本の TCP ストリームでトラヒックを流すとほぼネットワークフローは安定し，10 分平均で 166.1 Mbps であった．ネットワークを混雑させ過ぎないことは効率的なネットワーク転送のためには大変重要なことが分かった．

2003 年になり，日米間には 2.4 Gbps × 2 の NII/SuperSINET，および秋には APAN/TransPAC の東京–シアトル間が東京–ロサンジェルス間となり帯域も 2.4 Gbps となった．そこで，現在構築している，日米タイの複数拠点からなる大規模な環太平洋グリッド・データファーム・テストベッドを利用し，SC2003 バンド幅チャレンジに再び参加して，日米間ファイル複製作成の実験，性能評価を行った．

本論文では，まず 2，3 章でグリッド・データファームおよび GNET-1 について簡単に説明する．続いて 4 章でテストベッドの紹介をしたあと，5 章で日米間ネットワーク転送性能，ファイル複製作成性能の評価を行い，6 章でまとめる．

2. グリッド・データファーム — 高性能データ処理をサポートする Gfarm ファイルシステム

グリッド・データファーム[14),18)] は，複数組織による高信頼のデータ共有，高速データアクセス，高速データ処理を目標として設計された．複数組織に分散するデータを，そのデータ格納場所を意識することなく，必要なアクセス制御の基でアクセスを可能とする Gfarm ファイルシステム（広域仮想ファイルシステム）を実現し，高信頼のデータ共有を図る．また，データアクセスの局所性を利用した，データ格納場所における広域並列分散処理のための並列 I/O，プロセス・スケジューリング（ファイル・アフィニティ・スケジューリング）を提供し，高速データアクセス，高速データ処理を図る．

グリッド・データファームの主な特徴を以下に示す．

- グリッド認証技術に基づく安全性
- データサイズ，利用規模に応じたスケーラビリティ
- データ格納場所を意識することのないデータアクセス
- 自動的に適切なデータ複製にアクセスすることによる高信頼性
- 複数のストレージに並列アクセスすることによる高速性

Gfarm はグリッド・データファーム・アーキテクチャの参照実装であり，WWW ページ[5)] によりソースコードが公開されている．

2.1 Gfarm ファイルシステム

Gfarm ファイルシステムは，複数組織からなるグリッドに分散配置されたファイル・データを管理する仮想的な広域ファイルシステムである．利用者やアプリケーションは，POSIX ファイル I/O API，あるいは拡張された機能のための Gfarm ファイル I/O API により，Gfarm ファイルシステムを利用する．また，Linux では Gfarm ファイルシステムをマウントして利用することもできる．

Gfarm ファイルシステムは，仮想ファイルシステムを構成するため，

- 仮想ディレクトリ階層管理，ファイル複製管理を行うファイルシステム・メタデータ・サーバ
- 遠隔ファイルアクセスのための I/O サーバ
- それらサーバを利用して仮想ファイルシステムに対する I/O を実現するファイル I/O ライブラリ

を提供する．

仮想ディレクトリ階層管理では，階層的なファイルのパス名からのファイルのルックアップ，アクセス制御を行う．ファイル複製管理では，ルックアップされたファイルの実体がどこに複製され，存在するかを管

理する．ファイル I/O ライブラリは，それらの情報を元に実際にファイルを保持する I/O サーバを利用してアクセスを実現する．

特に指定されなければ，ファイル I/O ライブラリはファイル格納位置，I/O サーバの CPU 負荷などを元に適切なファイル複製を選ぶ．そのとき利用可能なファイル複製を選ぶことで耐故障性を向上させるだけではなく，ファイルへのアクセスを分散させることにより，アクセス集中を回避し性能低下を防ぐことができる．

2.2 高速データアクセス，高速データ処理のサポート

高性能データ処理に向けた世界規模の並列分散処理を支援するために，ファイル集合の管理，ファイル・アフィニティ・スケジューリングと呼んでいるファイル格納位置に基づくプロセス・スケジューリング手法，および SPMD 並列データアクセスのための並列ファイルアクセスの方式を提案した．

複数のクラスタノードに分散配置されたファイルの集合は，一つのレギュラ・ファイルとして扱うことができ，Gfarm ファイルあるいはスーパ・ファイルと呼ばれる．ファイル・アフィニティ・スケジューリングは，指定された Gfarm ファイルの構成ファイルに対し，そのファイル複製の所在を元に並列プロセスを割り当てる．ローカル・ファイル・ビューと呼ばれる新しいファイル・ビューは，Gfarm ファイルの構成ファイルのそれぞれに対する並列アクセスを可能とする．

ファイル・アフィニティ・スケジューリングとローカル・ファイル・ビューは，単一システムイメージで，“owner computes” 戦略，あるいは「計算をデータに移動させる」アプローチによる並列分散データ処理を可能とする．

2.3 ファイル複製作成

Gfarm ファイルのファイル複製は，その Gfarm ファイルを構成するファイル集合の複製を意味する．それぞれの構成ファイルの複製作成には依存性がないため，並列に独立してファイル複製を作成することができる．Gfarm ファイルの構成ファイルが異るクラスタノードに分散配置されている場合，ファイル複製作成は複数のクラスタノードから異る複数のクラスタノードへの第三者並列ファイル転送となる．これにより，ファイル複製作成において，スケーラブルなディスク I/O 性能を達成することができる．

3. GNET-1 — プログラマブル・ネットワーク装置

GNET-1[4),19)] は，再構成可能なゲートアレイ (FPGA) と 4 つの高速なメモリバンク，4 つの Gigabit Ethernet ポートをもつネットワーク実験装置である．FPGA の設定により，広域ネットワークのエミュレーション，GPS を利用したマイクロ秒単位の片道遅延の測定，ミリ秒以下の精密な通信バンド幅の測定，インターフレームギャップ (IFG) を調整することによるネットワーク流量制限，スイッチ機能の試験などに利用することができる．

本論文では，上記の GNET-1 のさまざまな機能のうち，IFG による流量制限（ペーシング）を主に利用する．IFG はイーサ・フレーム間のギャップのことで，このギャップを変化させることにより利用可能バンド幅の制限を任意に行うことができる．例えば，ジャンボフレームを利用して，MTU (Maximum Transfer Unit) が 9000 Bytes の場合，IFG を x とすると，Gigabit Ethernet のネットワーク流量は $8948/(9026+x) \times 1000$Mbps に制限される．

ペーシングではフレーム毎の IFG を調整するため，精密に通信バンド幅を測定したとしても，制限された流量以上には流れることがない．そのため，ピークのない安定したネットワークフローを実現することができる[15)]．また，GNET-1 では，それぞれのネットワーク・インターフェース毎に 16 MBytes のバッファをもっており，パケットロスを大幅に防ぐことができる．

4. 環太平洋グリッド・データファーム・テストベッド

Gfarm によるグリッド・データファームの実証実験環境として，国内外の複数の組織からなる環太平洋規模のテストベッドを構築している．テストベッドは，日本では産総研，東工大，筑波大，高エネ機構，APAN 東京 XP の 5 拠点，米国はインディアナ大学，サンディエゴ・スーパ・コンピュータ・センタ (SDSC)，および国際会議 SC2003 の期間中は会議の開催地のフェニックス国際会議場 (SC2003 会場) の 3 拠点，タイの Kasetsart 大学の計 9 拠点の PC クラスタからなる．

4.1 ネットワーク構成

図 1 にテストベッドのネットワーク構成を示す．日米間には APAN/TransPAC と SuperSINET の高速ネットワークがある．APAN/TransPAC は東京とロサンジェルス (LA) を接続する LA 線と，東京とシカゴを接続するシカゴ線がある．LA 線は OC-48 POS であり，ネットワーク帯域は 2.4 Gbps である．シカゴ線は OC-12 ATM である．シカゴ線は ATM であるため，ヘッダのオーバヘッドが大きく，TCP により利用可能なバンド幅は 500 Mbps 程度となる．SuperSINET は東京とニューヨーク (NY) を二本の OC-48 で接続し，帯域は 4.8 Gbps である．ただし，今回の実験ではそのうち 1 Gbps を利用している．

日本国内の 5 拠点は，つくば WAN，SuperSINET，Maffin 線などによりそれぞれ 1 Gbps で接続されている．米国内では，SC2003 会場，SDSC，インディアナ大学は Abilene にそれぞれ 10 Gbps，1 Gbps，1

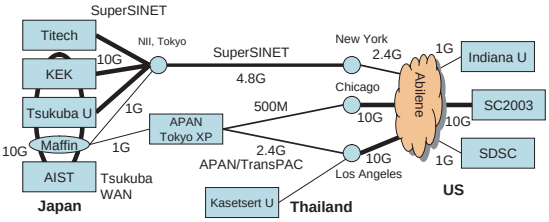

図 1 環太平洋グリッド・データファーム・テストベッドのネットワーク構成

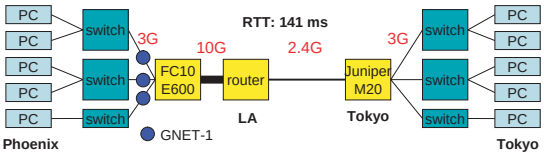

図 2 APAN/TransPAC の LA 線を利用した通信性能の評価環境.

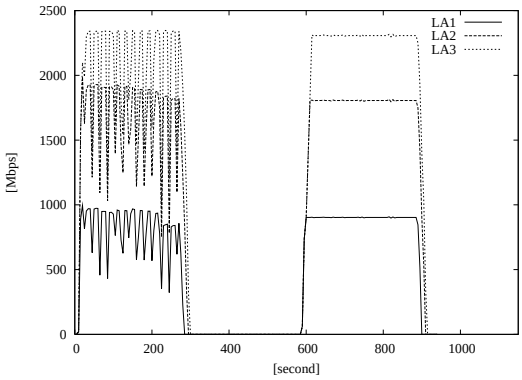

図 3 APAN/TransPAC の LA 線を利用した日米間ネットワーク転送性能. 左 (300 秒まで) は IFG による流量制限をしていない場合. 右 (600 秒から 900 秒まで) はそれぞれ 900 Mbps, 900 Mbps, 500 Mbps で流量制限した場合.

Gbps で接続している.

4.2 Gfarm ファイルシステムと PC クラスタ構成

環太平洋グリッド・データファーム・テストベッドでは，容量 68.5 TBytes，ディスク書き込み性能 11.0 GBytes/sec，読み込み性能 13.3 GBytes/sec の Gfarm ファイルシステムを構成している．それぞれの拠点における，構成している PC クラスタのノード数，ファイルシステム容量，およびディスク I/O 性能を表 1 に示す.

Gfarm ファイルシステムは，ローカル・ファイル・ビューによる並列 I/O と，並列ファイル複製作成において，それぞれのクラスタノードのローカルファイルシステムのディスク I/O 性能を合わせた，スケーラブルなディスク I/O 性能を出すことができる．また，POSIX I/O を利用したプログラムからも，ファイル集合を構成するそれぞれのファイルへの操作と見做すことにより，アプリケーションのソースコードの修正なしで，スケーラブルなディスク I/O 性能を出すことができる.

それぞれの拠点に関するクラスタ構成の詳細は 15) を参照されたい.

5. 日米間ファイル複製作成の性能評価

環太平洋グリッド・データファーム・テストベッドを利用して，日米間の高速ネットワークによるファイル複製作成の性能評価を行う．まず，GNET-1 を利用した IFG による流量制限の効果を調べるため，APAN/TransPAC の LA 線を利用してネットワーク転送実験を行った．その後，APAN/TransPAC の LA 線，シカゴ線および SuperSINET の NY 線を利用して，日米間で Gfarm によるファイル複製作成の評価を行う.

5.1 ネットワーク転送性能の評価

APAN/TransPAC の LA 線を利用し，IFG による流量制限の効果を調べた．評価環境を図 2 に示す．米国フェニックス SC2003 会場と APAN 東京 XP に設置した PC クラスタのうちそれぞれ 5 ノードを利用した．5 ノードは，2 ノード，2 ノード，1 ノードの 3 組に分けられ，Gigabit スイッチ，GNET-1 を介し，3 本の Gigabit Ethernet で 10G スイッチ (E600) に接続されている．以下では，その 3 本の Gigabit Ethernet にそれぞれ LA1，LA2，LA3 とラベルをつける．APAN 東京 XP もほぼ同様であるが，こちらは GNET-1 は介していない.

ボトルネックリンクのバンド幅は APAN/TransPAC の 2.4 Gbps，往復の遅延は 141 msec である．評価では Iperf を利用し，HighSpeed TCP の転送バンド幅を 5 分間計測した．それぞれのノード・ペアに対し単一の TCP コネクションを張り，5 並列の TCP コネクションでデータを転送した.

図 3 に，ネットワーク転送性能を示す．バンド幅の計測は，E600 に対し 5 秒間隔で SNMP リクエストを発行することにより行っている[17]．0 秒から 300 秒までは IFG による流量制限はしておらず，600 秒から 900 秒は LA1，LA2，LA3 に設置された GNET-1 により IFG を調節し，それぞれ 900 Mbps，900 Mbps，500 Mbps に流量を制限した.

流量制限をしない場合は，3 本の Gigabit Ethernet により，ボトルネックの 2.4 Gbps を上回るトラヒックを発生させてしまい，途中経路のルータのバッファを溢れさせ，一斉にパケットロスを引き起こしている．HighSpeed TCP により，パケットロス後の輻輳ウインドウのサイズの回復は通常の TCP に比べ早いが，再びパケットロスを引き起こし，結果として通信バンド幅は不安定に振動している.

一方，IFG による流量制限をした場合は，それぞれ

表 1　Gfarm ファイルシステムの容量と並列ディスク I/O 性能のそれぞれの拠点における内訳

	AIST	U. Tsukuba	Titech	KEK	APAN	(SC2003)	SDSC	Indiana	Kasetsart U.
#nodes	24	10	147	7	16	(32)	8	13	1
capacity [GB]	13,454	934	15,659	3,768	11,946	(23,892)	254	48	38
write [MB/s]	2,300	295	3,461	180	1,534	(3,068)	170	138	11.8
read [MB/s]	2,847	319	4,195	233	1,898	(3,796)	170	128	43.4

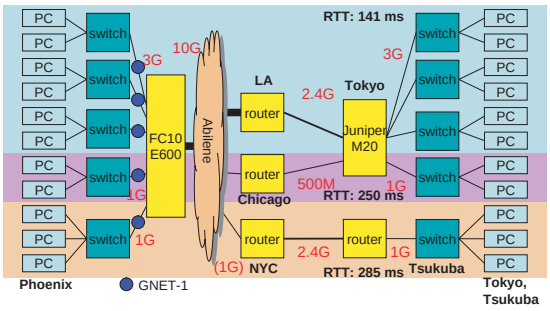

図 4　日米間ファイル複製作成におけるクラスタ，ネットワーク環境

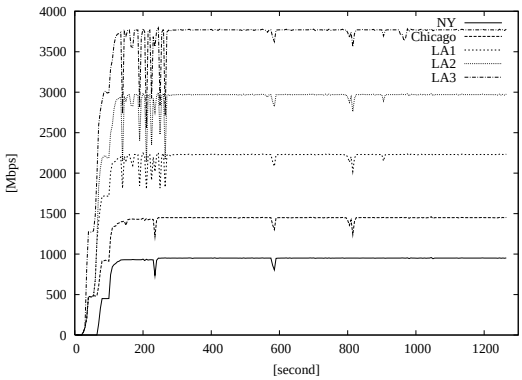

図 5　日米間ファイル複製作成におけるデータ転送性能．シカゴ線，SuperSINET の NY 線は安定している．LA 線は，300 秒頃に LA1 の流量制限を 800 Mbps から 780 Mbps に変更して以降，安定した．

の TCP コネクションに関して流量を制限したバンド幅で安定している．これにより，2.3 Gbps のパケットロスフリーのネットワークが日米間に実現されていることが分かる．さらに，900 Mbps に流量制限された通信路は TCP コネクションが 2 本張られているが，それぞれのバンド幅は 450 Mbps 程度でバランスし，安定したネットワークフローとなっていた．安定した高効率のネットワーク転送性能を得るためには，パケットロスフリーのネットワークを実現することが重要であり，逆に，GNET-1 による IFG の制御で実現可能であることが分かった．

5.2　ファイル転送性能の評価

APAN/TransPAC の LA 線，シカゴ線および SuperSINET の NY 線を利用して，日米間で Gfarm によるファイル複製作成の評価を行う．利用した PC クラスタ，ネットワーク環境を図 4 に示す．Phoenix の SC2003 会場の PC クラスタ 11 ノード，および APAN 東京 XP の 8 ノード，つくば産総研の 3 ノードを利用した．SC2003 会場の 11 ノードを 6 ノード，2 ノード，3 ノードと 3 組に分け，それぞれ APAN 東京 XP の 6 ノード，2 ノードおよびつくば産総研の 3 ノードと，APAN/TransPAC の LA 線，シカゴ線，および SuperSINET の NY 線を利用しファイル複製の作成を行った．LA 線は，通信性能評価と同様に Gigabit Ethernet を 3 本利用し，それぞれ LA1，LA2，LA3 とラベルをつける．流量制限のための GNET-1 は SC2003 会場の PC クラスタノードと E600 の間に 5 台設置した．

APAN/TransPAC の LA 線，シカゴ線，SuperSINET の NY 線のボトルネックリンクのバンド幅はそれぞれ 2.4 Gbps，500 Mbps，2.4 Gbps であり，往復遅延はそれぞれ 141 msec，250 msec，285 msec である．ファイル複製作成の評価では，1,100 GBytes の Gfarm ファイルを利用する．評価に利用した Gfarm ファイルは物理的にはそれぞれ 100 GBytes ずつ SC2003 会場の 11 ノードに分散して格納されている．ファイル複製は，Gfarm のファイル複製作成コマンド `gfrep` を利用した．

図 5 に日米間ファイル複製作成におけるデータ転送性能を示す．MTU は 6,000 Bytes としている．300 秒頃までは LA1，LA2，LA3，シカゴ線，NY 線に対しそれぞれ，800 Mbps，750 Mbps，800 Mbps，500 Mbps，930 Mbps の流量制限をしていたが，LA 線でパケットロスが発生し，ネットワークフローが不安定であったため，LA1 で流量制限を 800 Mbps から 780 Mbps に変え，逆に安定していた NY 線を 930 Mbps から 950 Mbps にあげた．以降は，ほぼパケットロスフリーで通信が行われているのが分かる．複数の日米間高速ネットワークを利用して，理論最大性能の 97%に相当する 3.78 Gbps の転送バンド幅を安定したネットワークフローで達成することができた．

6. ま と め

グリッド・データファームは，複数組織による高信頼のデータ共有，高速データアクセス，高速データ処理のための Gfarm ファイルシステムを提供する．現在，日本に 5 拠点，米国に 3 拠点，タイに 1 拠点の計 9 拠点からなる環太平洋規模のグリッド・データファーム・テストベッドを構築しており，Gfarm ファイルシステムの容量は 68.5 TBytes，ディスク I/O 性能は

13 GB/sec である．本テストベッドを利用して，日米間のファイル複製作成に関する性能評価を行った．

ネットワーク・テストベッド装置 GNET-1 を利用した IFG による流量制限により，およそ 10,000 Km 離れた日米間において，理論ピーク性能に近い通信性能を達成した．流量制限によりパケットロスフリーの広帯域ネットワークが日米間に実現し，HighSpeed TCP を利用したネットワーク転送でも，安定したネットワークフローを達成することができた．

このパケットロスフリーのネットワークを利用した日米間ファイル転送では，1,100 Gbytes のデータを米国の 11 ノードから日本の 11 ノードに転送し，6,000 Bytes の MTU を利用して，3.78 Gbps の安定したデータ転送レートを実現した．これは理論ピーク性能の 3.9 Gbps の 97%に相当する．

今回のファイル転送実験では，それぞれのクラスタノードで Linux 2.4 を利用したが，バッファキャッシュをディスクにフラッシュするときに，ネットワーク転送が数分間止ってしまうというバグが存在した．このバグに対し，バッファキャッシュを 0.1 ミリ秒間隔でフラッシュしながらデータを受信するという消極的な対策を講じた．この対策は，ノード間のファイル転送レートを 400 Mbps 程度まで下げる事になったが，数十分間のファイル転送でネットワーク転送が止る現象は収まった．今後，このバグを解決して，より高性能のファイル転送を目指したい．

また，我々グループは，他機関と協力してグローバル・グリッド・フォーラム[3](GGF) にグリッド・ファイルシステム WG[6] (GFS-WG) の設立を提案し，2004 年の 1 月にその提案が認められた．今後，Gfarm ファイルシステムによる知見を元に，グリッド・ファイルシステムの標準化を進めていく予定である．

謝　　辞

環太平洋グリッド・データファーム・テストベッドの構築にあたり，PRAGMA/ApGrid のメンバ諸氏に感謝致します．大規模ファイル転送実験にあたり，貴重なご助言をいただき，ネットワーク・ルータの設定変更などもしていただいた，つくば WAN NOC チーム，APAN NOC チーム，NII/SuperSINET NOC チーム，Abilene NOC チームに感謝致します．SC2003 会場において 10G スイッチをご提供いただいた Force 10 ネットワークスに感謝いたします．また，本研究を遂行するにあたり貴重なご助言，ご討論をいただいた産業技術総合研究所，高エネルギー加速器研究開発機構，東京工業大学，筑波大学のグリッド・データファーム・プロジェクトメンバ諸氏，グリッド研究センターのメンバ諸氏に感謝いたします．

本研究の一部は，新エネルギー・産業技術総合開発機構基盤技術研究促進事業（民間基盤技術研究支援制度）の一環として受託を受け実施している「大規模・高信頼サーバの研究」，および文部科学省「経済活性化のための重点技術開発プロジェクト」の一環として実施されている超高速コンピュータ網形成プロジェクト (NAREGI: National Research Grid Initiative) による．

参 考 文 献

1) *Almost TCP over UDP (atou).* http://www.csm.ornl.gov/~dunigan/netperf/atou.html.
2) *FAST TCP.* http://netlab.caltech.edu/FAST/.
3) *Global Grid Forum.* http://www.ggf.org/.
4) *GNET-1: Gigabit network testbed.* http://www.gtrc.aist.go.jp/gnet/.
5) *Grid Datafarm.* http://datafarm.apgrid.org/.
6) *Grid File System WG.* https://forge.gridforum.org/projects/gfs-wg/.
7) *Scalable TCP.* http://www-lce.eng.cam.ac.uk/~ctk21/scalable/.
8) *Tsunami.* http://www.anml.iu.edu/anmlresearch.html.
9) Floyd, S.: *HighSpeed TCP for Large Congestion Windows* (2003). RFC 3649.
10) He, E., Leigh, J., Yu, O. and DeFanti, T. A.: Reliable Blast UDP: Predictable High Performance Bulk Data Transfer, *Proc. IEEE Cluster Computing* (2002).
11) Katabi, D., Handley, M. and Rohrs, C.: Congestion Control for High Bandwidth-Delay Product Networks, *Proc. ACM SIGCOMM 2002* (2002).
12) Sivakumar, H., Bailey, S. and Grossman, R. L.: PSockets: The Case for Application-level Network Striping for Data Intensive Applications using High Speed Wide Area Networks, *Proc. SC2000* (2000).
13) Sivakumar, H., Grossman, R. L., Mazzucco, M., Pan, Y. and Zhang, Q.: Simple Available Bandwidth Utilization Library for High-Speed Wide Area Networks, *Journal of Supercomputing* (2003).
14) Tatebe, O., Morita, Y., Matsuoka, S., Soda, N. and Sekiguchi, S.: Grid Datafarm Architecture for Petascale Data Intensive Computing, *Proc. CCGrid 2002*, pp. 102–110 (2002).
15) Tatebe, O., Ogawa, H., Kodama, Y., Kudoh, T., Sekiguchi, S., Matsuoka, S., Aida, K., Boku, T., Sato, M., Morita, Y., Kitatsuji, Y., Williams, J. and Hicks, J.: The Second Trans-Pacific Grid Datafarm Testbed and Experiments for SC2003, *Proc. SAINT 2004* (2004).
16) Tatebe, O., Sekiguchi, S., Morita, Y., Matsuoka, S. and Soda, N.: Worldwide Fast File Replication on Grid Datafarm, *Proc. CHEP03* (2003).
17) 近藤 秀樹, 石橋 拓也, 建部 修見: SNMP によるクラスタ性能計測手法の検討と評価, 情報処理学会研究報告 2003-HPC-93, pp. 19–24 (2003).
18) 建部 修見, 森田 洋平, 松岡 聡, 関口 智嗣, 曽田 哲之: ペタバイトスケールデータインテンシブコンピューティングのための Grid Datafarm アーキテクチャ, 情報処理学会論文誌：ハイパフォーマンスコンピューティングシステム, Vol. 43, No. SIG 6 (HPS 5), pp. 184–195 (2002).
19) 児玉 祐悦, 工藤 知宏, 佐藤 博之, 関口 智嗣: ハードウェアネットワークエミュレータを用いた TCP/IP 通信の評価, 情報処理学会研究報告 2003-HPC-95, pp. 47–52 (2003).