6G-4

広域環境での使用に対応した 分散ファイルシステム Gfarm の性能評価

竹川知孝[†1]　佐藤未来子[†2]　並木美太郎[†3]

東京農工大学大学院工学府電子情報工学専攻[†1]　東京農工大学大学院工学府[†2]

東京農工大学大学院工学研究院先端情報科学部門[†3]

1. はじめに

近年，新たなコンピューティングの形態としてクラウドコンピューティングが普及し始めている．広域環境での帯域幅が増すことで，ローカルのコンピュータで処理をするのではなく，ネットワークの先にあるサーバで処理をするクラウドコンピューティングを使用する環境が整ってきた．しかし，広域環境でコンピュータを使用する事が容易になった一方で，ネットワーク上に分散したファイルの管理や，増加するデータサイズへの対応など新たな課題が生まれている．既存のネットワークファイルシステムはファイルを単体のサーバで管理，保存を行う．しかし，広域環境での使用を想定した場合，一つのサーバで管理するとネットワーク遅延，サーバへの負荷が問題になる．そこで本研究では広域環境での使用に対応した分散ファイルシステム Gfarm の性能測定を行い，ローカル環境と広域環境のそれぞれの環境で Gfarm を利用する利点を示す．Gfarm は筑波大学の建部氏を中心にオープンソースで開発されているネットワークファイルシステムであり，広域環境でスケーラブルなアクセス性能を実現するために複製作成や，RTT を元にした最短経路でのファイルの取得といった特徴を持っている．

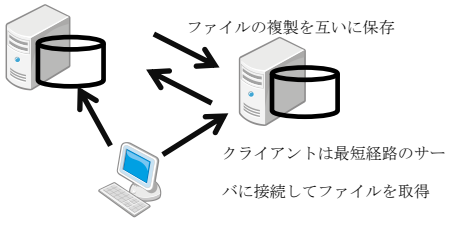

図１　Gfarm の特徴

Evaluation of the Gfarm Filesystem Corresponded to Use in the Wide Area Network

†1 Tomotaka TAKEKAWA
†1 Computer and Information Sciences, The Graduate School of Engineering at Tokyo University of Agriculture and Technology
†2 Mikiko SATO
†2 The Graduate School of Engineering at Tokyo University of Agriculture and Technology
†3 Mitaro NAMIKI
†3 Institute of Symbiotic Science and Technology at Tokyo University of Agriculture and Technology

2. 実験

Gfarm は大きく分けて，メタデータサーバ，ファイルシステムノード，クライアントの３つのパートで構成される．実験ではメタデータサーバ，ファイルシステムノードを BUFFALO 社の TeraStation で動作させ，PC をクライアントとしてサーバに接続する事で Gfarm を評価する．ローカル環境，広域環境それぞれの環境における Gfarm の性能を測定するために二つの実験を行った．

2.1 ローカル環境での実験

ローカル環境において TeraStation を 1 台，PC を 1 台配置し，これらを Gigabit Ether で接続する．TeraStation をメタデータサーバとファイルシステムノードとし，PC をクライアントとして Gfarm を動作させた場合と NFS を動作させた場合の転送速度を計測する．転送するファイルは 1MB，10MB，100MB，1GB のファイルであり，これらを以下のコマンドで転送し，転送速度を比較する．

```
time dd if=/dev/zero of=/mnt/gfarm/zero.dat bs=1M count=1
```

2.2 広域環境での実験

広域環境における実験では東京農工大学と筑波大学を結ぶ VPN を構築して Gfarm の性能を評価する．本実験で使用する VPN は KURO-SHEEVA において PacketiX VPN を動作させる事で構築し，東京農工大学と筑波大学の RTT は 16msec，東京農工大学内ネットワークの帯域は 1Gbps である．東京農工大学ではクライアント，ファイルシステムノード A を用意し，筑波大学にはメタデータサーバとファイルシステムノード B を用意する．筑波大学と東京農工大学のそれぞれのファイルシステムノードに 1MB，10MB，100MB，1GB のファイルを準備し，クライアントの読み込み時の転送速度と書き込み時の転送速度，さらに筑波大学にあるファイルの複製を東京農工大学のファイルシステムノードに作成した場合の転送速度を以下のコマンドで計測する．

```
time dd if=/dev/zero of=/mnt/gfarm/zero.dat bs=1M count=1
time dd if=/mnt/gfarm/zero.dat of=/home/takekawa
```

表１　ローカル環境での Gfarm と NFS

		1MB	10MB	100MB	1GB
Gfarm	平均転送時間(sec)	0.0918	0.4330	4.562	50.01
	平均転送速度(MB/s)	10.9	23.1	21.9	20.0
NFS	平均転送時間(sec)	0.1734	0.5140	4.657	47.31
	平均転送速度(MB/s)	5.77	19.5	21.5	21.1

表２　広域環境でファイルシステムノード B のファイルを読み込む

	1MB	10MB	100MB	1GB
平均転送時間(sec)	1.431	10.87	93.80	1091
平均転送速度(MB/s)	0.699	0.920	1.07	0.916

表３　広域環境でファイルシステムノード B からファイルシステムノード A へ複製を作成

	1MB	10MB	100MB	1GB
平均転送時間(sec)	2.538	7.764	80.18	802.1
平均転送速度(MB/s)	0.394	1.29	1.25	1.25

3. 結果

ローカル環境での実験結果である表１から Gfarm の転送速度は NFS と比較して大きな差がない事が確認できる．

また，広域環境での実験結果を示した表２から 10MB 以上のファイルを読み込んだ場合，値が約 1MB/s に収束している事がわかる．広域環境でファイルを転送するとローカル環境での結果と比較して RTT が長いため転送速度が低下する．

表３より転送経路はクライアントがファイルシステム B のファイルを読み込んだ場合と同様であるにも関わらず，10MB 以上のサイズのファイルを複製した場合においてファイル複製の転送速度が 1.25MB/s に収束している事がわかる．

表２と表４を比較する事で，複製作成後の転送速度は作成前に比べて 1GB のファイル読み込みにおいて約 20 倍向上している．

ローカル環境における Gfarm の転送速度の結果はファイルシステムノード A への書き込みと捉える事ができるので表１と表５の比較より，1GB のファイル書き込みにおいて広域環境での書き込みの転送速度はローカル環境での書き込みに対して約 4%に低下している事がわかる．この結果から読み込みだけでなく，書き込みにおいても広域環境での実験では転送速度の低下が確認できる．

4. 考察

ローカル環境での動作では既存のネットワークファイルシステムとして NFS を比較対象にした転送速度を計測した．結果，転送速度に関して Gfarm と NFS で大きな差がない事がわかった．Gfarm の複製作成を用いる事でローカル環境でも障害に強いネットワークファイルシステムを構築できる．また実際に広域環境で実験して転送速度を計測した．クライアントが RTT を考慮して最短経路でファイルシステムノードと接続するため，1GB のファイル転送において読み込みで約 20 倍，書き込みにおいても約 24 倍の転送速度の向上が確認できた．これらの結果から Gfarm はローカル環境と広域環境のどちらにおいても使用する利点がある．

表４　広域環境での実験においてファイルシステムノード A から複製を読み込む

	1MB	10MB	100MB	1GB
平均転送時間(sec)	0.200	0.585	5.330	54.96
平均転送速度(MB/s)	5.00	17.1	18.8	18.2

表５　広域環境でファイルシステムノード B へファイルを書き込む

	1MB	10MB	100MB	1GB
平均転送時間(sec)	1.406	11.96	121.0	1180
平均転送速度(MB/s)	0.711	0.836	0.826	0.847

5. おわりに

本研究では広域環境での使用に対応した分散ファイルシステム Gfarm の評価を行った．ローカル環境と，広域環境でのそれぞれの評価から，耐故障性，広域環境での位置透過性をもったファイルシステムとして Gfarm が有効である事がわかった．今後は Gfarm の特性を生かして広域環境での動作に対応したシンクライアントシステムを構築したい．

謝辞

本研究を進めるにあたり，筑波大学　建部修見准教授，株式会社バッファロー　稲垣達夫氏には大変お世話になりました．

参考文献

1) 建部 修見，曽田 哲之：広域分散ファイルシステム Gfarm v2 の実装と評価,情報処理学会研究報告，2007-HPC-113，pp.7-12 (2007).
2) Gfarm ファイルシステム. http://datafarm.apgrid.org/index.ja.html.