

Phase サーバの現状と展望

西田晃*、長谷川秀彦†、関口智嗣‡、小柳義夫*

概要

Phase (Parallel and High-performance Application Software Exchange) はソフトウェアライブラリの蓄積、公開を通じて理論研究から計算機ソフトウェア開発に渡る研究者間の情報流通を促進し、より高品質なソフトウェアを育成していくことを目的に発足したプロジェクトであり、既に高性能計算に関する国内最大のライブラリとなっている。本稿では Phase サーバの概念、提供内容、及び現在の利用状況を報告し、あわせて今後の課題について述べる。

1. はじめに

Phase (Parallel and High-performance Application Software Exchange) は高性能計算に関するソフトウェアライブラリの蓄積、公開を通じて理論研究から計算機ソフトウェア開発に渡る研究者間の情報流通を促進し、より高品質なソフトウェアを育成していくことを目的に 1995 年 4 月より発足したプロジェクトであり、現在以下のメンバーによりボランティアベースで運営されている。

東京大学大学院理学系研究科 小柳義夫
東京大学大型計算機センター 村尾裕一
お茶の水女子大学理学部 長嶋雲平
新情報処理開発機構 佐藤三久
電子技術総合研究所 関口智嗣
電子技術総合研究所 中田秀基
図書館情報大学 長谷川秀彦
東京大学大学院理学系研究科 西田晃

Phase では

- 1) 文書: 論文, テクニカルレポート, 本など
- 2) ソフトウェア: 開発者から提供された (主としてソース形式の) プログラムとマニュアル, 使用方法など
- 3) データ: 問題, 測定結果, コメントなどのデータ
- 4) 案内: 学会, 研究機関の連絡先, ハードウェア・ソフトウェアのベンダ, シンポジウムのスケジュール, メーリングリストの案内, 関連する情報提供サービスなど

の提供を行っており、既にハイパフォーマンスコ

ンピューティングに関する国内最大のライブラリとなっている。

また現在までに Netlib¹, Matrix Market² 両プロジェクトの協力を得て公式ミラーサイトの運営を開始しており、今後さらにその対象を広げていく予定である。

2. 背景

このような数学ソフトウェア配布システムとして、海外では Tennessee 大学 / Oak Ridge 国立研究所の Dongarra らを中心として Netlib をはじめとするサービスが知られている。Netlib では 1985 年に電子メール経由での配布が開始されて以来、FTP, HTTP などの通信手段の普及とともに急速に利用規模が拡大し、1996 年には年間利用件数が 600 万件を超えている (図 1³ 参照)。現在 HTTP ベースでの運営は University of Tennessee (USA), AT&T Bell Laboratories (USA), University of Bergen (Norway), The Higher Education National Software Archives (UK), Konrad-Zuse-Zentrum für Informationstechnik Berlin (Germany), 電子技術総合研究所, 及び Centre National de la Recherche Scientifique (France) の 7 箇所で行なわれている。

さらに近年では Netlib を含む HTTP ベースの情報提供サービス (NHSE, National HPC Software Exchange)⁴ が米国科学財団によって運営されており、NHSE Software Catalog, Netlib, HPC-Netlib (high performance branch of Netlib), PTLIB (Parallel Tools Library) などのライブ

¹<http://www.netlib.org>, <http://netlib.bell-labs.com>

²<http://math.nist.gov/MatrixMarket>

³<http://www.netlib.org/utk/misc/count.html>

⁴<http://www.nhse.org>

*東京大学 {nishida,oyanagi}@is.s.u-tokyo.ac.jp

†図書館情報大学 hasegawa@ulib.ac.jp

‡電子技術総合研究所 sekiguti@etl.go.jp

Requests Made to the Netlib Repositories
at the Univ. of Tennessee & ORNL

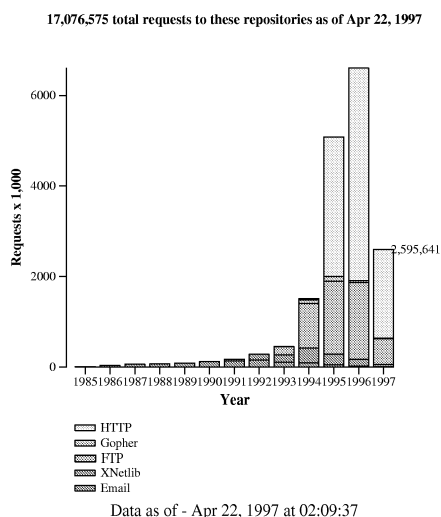


図 1. Netlib (Tennessee 大学 / Oak Ridge 国立研究所) へのアクセス状況 (Netlib 公表資料による)

ラリの公開、運営を行っている (図 2⁵, 表 1⁶ 参照). 現段階では 1 日の平均利用件数は 670 件程度であり、依然として Netlib の利用の方が圧倒的に多いが、NHSE に関しても利用件数は毎年増加しており、国内においても同様な公的基盤整備を行なっていく必要がある。

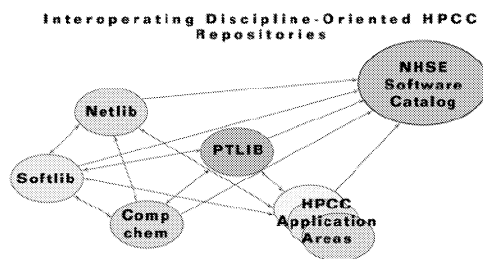


図 2. National HPC Software Exchange の概念 (NHSE 公表資料による)

⁵<http://www.nhse.org>

⁶http://www.nhse.org/stats/comprehensive_stats.html

表 1. <http://www.nhse.org> のドメイン別利用状況 (1994/2/11-1997/4/26, NHSE 公表資料による)

#reqs	%reqs	domain
145310	18.90	edu (USA Educational)
137920	17.93	utk.edu (University of Tennessee)
123129	16.01	com (Commercial, mainly USA)
85638	11.14	[unresolved numerical addresses]
38898	5.06	net (Network)
37755	4.91	gov (USA Government)
20531	2.67	de (Germany)
20339	2.64	ca (Canada)
17050	2.22	fr (France)
16963	2.21	uk (United Kingdom)
16740	2.18	jp (Japan)

3. 利用方法

現在 Phase サーバは電子技術総合研究所内に置かれており、主に HTTP, FTP ベースのサービスを行なっている。主なサービスのアドレスはそれぞれ以下の通りとなっている。

HTTP index <http://phase.etl.go.jp>
Contrib /contrib/
Netlib /netlib/
Matrix Market /MatrixMarket/

FTP index <ftp://phase.etl.go.jp>
Netlib /pub/netlib/
Matrix Market /pub/MatrixMarket2/
MPI Standard /pub/mpi/

Netlib については email でのサービスも行なっており、index は netlib@phase.etl.go.jp 宛に "send index" を本体に持つメールを送付することにより入手する仕組みとなっている。

4. 現在までの利用状況

現在までの利用状況を表 2-5, 及び図 3 に示す。94 年 12 月から 97 年 4 月下旬までの利用件数、及び転送されたバイト数の全体に占める割合を集計したものである。現在までの総利用件数は 16 万件に上り、1 日の平均利用件数は 300 件を越えているが、表 3 から国内での利用が海外、特に米国からの利用に比べて少ないことが分かる。海外からのアクセスに関しては、民間の検索サービス運営サイトによるミラーリングが大きな割合を占めているが、国内における認知度が低いことにも問題があり、今後より安定的で充実したサービスを行なっていく必要がある。

表 2. <http://phase.etl.go.jp> 利用状況
(1994/11/30-1997/4/22)

Total completed requests	160,453
Average completed requests per day	316
Number of distinct files requested	21,549
Number of distinct hosts served	9,328
Total data transferred	2,868 Mbytes
Average data transferred per day	5,781 kbytes

表 3. ドメイン別利用状況

#reqs	%bytes	domain
58170	36.76%	jp (Japan)
48562	22.64%	com (Commercial, mainly USA)
20484	9.87%	[unresolved numerical addresses]
13955	19.85%	org (Non-Profit Making Organizations)
5534	1.08%	[unknown]
4046	3.33%	edu (USA Educational)
2187	0.90%	net (Network)
849	1.05%	de (Germany)
829	0.44%	fr (France)
679	0.21%	uk (United Kingdom)
554	0.28%	ca (Canada)
524	0.22%	au (Australia)
519	0.70%	gov (USA Government)

5. ライブラリの登録方法

Phase では

- Free: 料金を徴収しない
- As is: 提供される情報に手を加えない
- Volunteer: 非営利目的に限定する
- No requirement: ユーザへの負担を求めない
- Clear Source: 作成者及び入手元を明示する
- Self-responsibility: 利用者の自己責任を原則とする

のガイドラインのもとに良質なサービスを提供することを目的としており、ライブラリの公開やリンクに関する情報の提供は email: phase@etl.go.jp で随時受け付けている。各ライブラリの管理は原則

表 4. ディレクトリ別利用状況

#reqs	%bytes	directory
94192	73.60%	/netlib/ (Netlib)
8492	1.21%	/ninf/ (Ninf)
8100	1.41%	[root directory]
6615	0.32%	/sighpc/ (IPSJ SIGHPC)
5508	15.98%	/contrib/ (Contributed Software Repository from Nation Wide)
2435	0.14%	/phase/ (Phase-related Files)
1758	0.90%	/jsiam/ (JSIAM)

表 5. Phase/Netlib 利用状況

#reqs	%bytes	page
2295	0.17%	/netlib/
1224	1.98%	/netlib/master/
574	0.17%	/netlib/benchmark/
404	0.01%	/netlib/netlib/search/search.html
239	0.01%	/netlib/f2c/
235	0.04%	/netlib/benchmark/top500.html
202	0.13%	/netlib/lapack/

Each unit (≡) represents 250 requests, or part thereof.

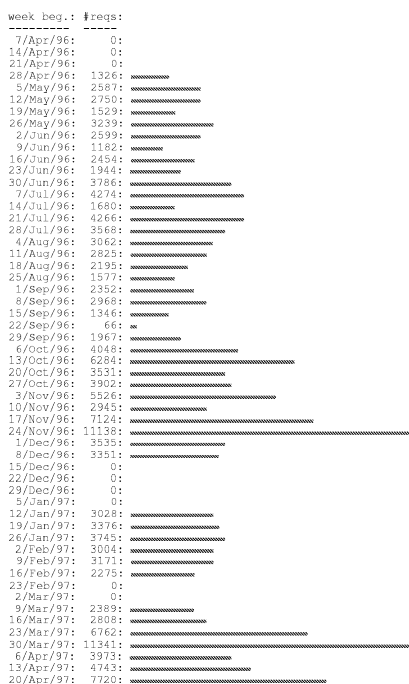


図 3. <http://phase.etl.go.jp> 週間利用状況
(1996/4/1-1997/4/20)

として Phase で行なっているが、この他に Phase による定期的なミラーリングを用いた登録も可能となっている。現時点ではソフトウェアライブラリ、数値データ、及び LAPACK, MPI, PVM 等ハイパフォーマンスコンピューティングに関する和訳資料などが登録されており、既に国内外で広く利用されている (表 6 参照)。

6. 今後の課題と展望

Phase のシステムは現在 SS20/514 256MB, Solaris 2.4 上で運営されている。httpd, ftpd として CERN-HTTPD-3.0, wu-2.4(2) を利用しているが、これらはそれぞれ Apache 等への変更も含めて随時更新される予定である。記憶容量については約 6 GB が確保されているが、これについても 16 GB

表 6. ソフトウェアライブラリ利用状況

#reqs	page
1569	/contrib/
647	/contrib/MPI-j/
236	/contrib/LAPACK-j/
228	/contrib/master/
219	/contrib/PHASE/
192	/contrib/wavelet/
110	/contrib/moriplot/
101	/contrib/MathFunc/
91	/contrib/PVM-j/
69	/contrib/psc/psc95/

程度に拡張される計画となっている。

ミラーリングに関しては、現在 HTTP ベースで Matrix Market, また FTP ベースではこれに加えて Netlib, MPI Standard⁷ についてサービスを行なっている。Netlib では FTP プロトコルを用いたシステムを採用しており、安全性や通信量の削減に重点を置いているが、近年では他のソフトウェアに比べて従来ほどの利点はなくなってきている。

Matrix Market は行列アルゴリズムの比較のためのテストデータを提供するライブラリであり、NIST (National Institute of Standards and Technology) 及び DARPA (Defense Advanced Research Projects Agency) のもとで運営されている。行列生成機能なども整備されつつあり、大量のデータを転送する性質上ミラーリングを行なうことが望ましい。

これらの運営は、基本的にフリーウェアのスク립トを利用しており、FTP については mirror-2.8, HTTP については w3mir-alpha⁷ を採用している。FTP ベースでのミラーリングは既に一般的であるが、Phase ではこれらについても今後更に内容を充実させていく予定である。

HTTP ベースのサービスについては動作に不安定な面もあり、proxy caching server 等への移行も含めた対策を検討している。また検索機能は最も利用頻度の高いサービスの一つであるが、現段階では検索リクエストはマスターサイトに転送されるよう設定しており、参照すべき対象が特定しにくい場合にはミラーリングの利点が発揮できない問題があるため、今後はミラー部分を含む独自の検索サービスを提供する予定である。その他 archie, search engine 等の一般的な検索手段についても、引続き積極的に活用していく方針である。

⁷<ftp://info.mcs.anl.gov/pub/mpi>

謝辞

日頃より本プロジェクトの運営に御協力頂いているメンバーの皆様に感謝致します。またミラーリングの実現に当り、御賛同と御支援を頂いた J. Dongarra, E. Grosse, R. Boisvert の諸氏をはじめとする Netlib, Matrix Market 両プロジェクトの運営グループの皆様に感謝致します。

参考文献

- [1] R. F. BOISVERT, R. POZO, K. REMINGTON, R. F. BARRETT, AND J. DONGARRA, *Matrix Market: a web resource for test matrix collections*, Chapman and Hall, London, 1997, pp. 125–137.
- [2] S. BROWNE, J. DONGARRA, S. GREEN, K. MOORE, T. ROWAN, R. WADE, G. FOX, K. HAWICK, K. KENNEDY, J. POOL, R. STEVENS, R. OLSEN, AND T. DISZ, *The national hpc software exchange*, IEEE Computational Science and Engineering, 2 (1995), pp. 62–69.
- [3] J. DONGARRA AND E. GROSSE, *Distribution of mathematical software via electronic mail*, Commun. ACM, 30 (1987), pp. 403–407.
- [4] E. GROSSE, *Repository mirroring*, ACM Trans. Math. Softw., 21 (1995), pp. 89–97.
- [5] H. HASEGAWA, *High Performance Computing at the University of Tennessee*, IPSJ SIG Notes, 95 (1995), pp. 43–50.