

プログラム解析情報可視化システム -PREPS- *

4U-5

草野和寛 妹尾義樹 †渡辺幸光†

NEC C&C 研究所† †NEC 情報システムズ‡

e-mail: kusano@csl.cl.nec.co.jp, seo@csl.cl.nec.co.jp, yukimi@nsis.cl.nec.co.jp

1 はじめに

分散メモリマシンでプログラムを並列化実行するには、処理の流れを各PEにタスク分割することに加えて、データを分割して分散配置することが重要である。このタスク分割とデータの分散配置をうまく調和させないと、実行時に大量のデータ転送が必要になり、性能低下の原因となる。これまでにプログラムの静的な解析情報を表示するツール [1] やプロファイル実行の結果を表示するツールは開発されている。しかし、データの分散配置とタスク分割をうまく調和させるためのツールは存在していない。そこで我々は、タスク分割とデータ分散の情報を同時に表示して、それらを調和させ、効率的な並列化を行なうことを目的としたツールを開発した。本稿では、そのプログラム解析情報可視化システム PREPS の構成と機能について述べる。

2 プログラム解析情報可視化システム PREPS

PREPSは、タスク分割とデータ分散情報をうまく調和させることを目的とした、並列化支援システムである。このために、本システムはデータのアクセスを中心にした表示を行なう。

PREPSは、FORTRANプログラムの解析結果を保持するデータベースファイルからプログラム解析情報を読み込み、その解析情報を加工・変形して、Xウィンドウ上で表示する。本システムは、図1に示すように、データベースサーバ、通信サーバ、可視化エンジン、表示部の4つの部分から成っている。本システムは、プログラ

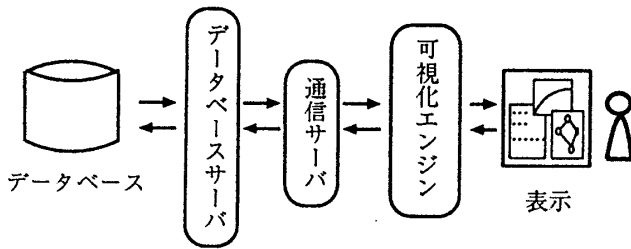


図1: システムの構造

ム解析情報をデータベースファイルから読み込む構造にすることで、プログラムの解析部とは独立に動作できると共に、プログラム全体の解析情報の表示を可能にしている。また、プログラム解析情報は独自フォーマットで

保持しているため、異なるデータベース仕様でも、データベースサーバの変更で対応が可能である。

本システムが表示する解析情報には、プログラムの負荷分布、プログラムのコントロールフロー、変数の依存関係や変数の分割状態、データ転送に関する情報がある。

3 本システムの表示機能

図2を用いて本システムの表示機能を示す。

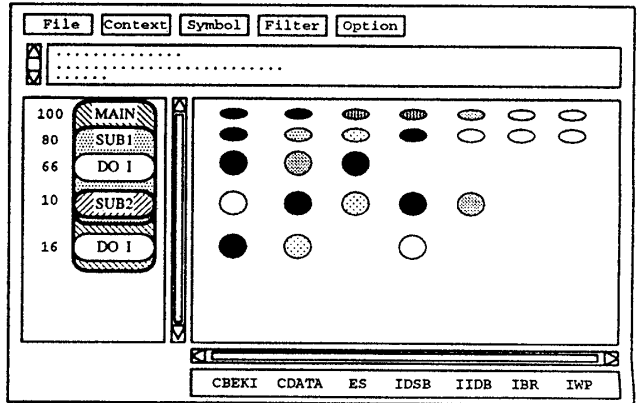


図2: 表示画面イメージ

縦軸にプログラムのコントロールフローを、横軸に変数名をとり、交点にあたる楕円領域で以下の情報を表す。

- アクセス属性 (Read/Write)
- アクセス回数
- データ転送量
- アクセス方向

アクセス回数とデータ転送量の違いは色の濃淡で表し、アクセス属性は色の違いで表示する。アクセス方向の表示は、データ分割と並列化次元のアクセス方向が一致/不一致を各変数で色分けして表す。図2のSUB2とIBRの交点部分のように何も表示されていない場合、SUB2では変数IBRにアクセスしないことを表し、SUB1とCBEKIの交点に示すように濃い色でアクセスが多いことを表す。このアクセス量やデータ転送量の表示は、プログラム構造に対して階層的な表示になっている。

横軸に表示する変数名は、縦軸で示したプログラムの領域で最初に宣言された名前を表示する。例えば、CBEKIはMAINで宣言された名前であり、SUB1では異なる名前であっても、その名前は表示されない。このような別名でのアクセスに関する情報は、後述の変数に関するデータ表示において参照できる。本システムは、これらの変数をユーザ指示で並べ替えるソート機能として、以下の機能を備えている。

*Program Information Visualization System -PREPS-

†Kazuhiro KUSANO, Yoshiki SEO, Yukimitsu WATANABE

‡C&C Research Laboratories, NEC Corporation

§NEC Informatec Systems

- アルファベット順
- アクセス回数順
- データ転送量順

また、表示変数がある条件により制限するフィルター機能として以下の機能を持っている。

- アクセス方向とデータ分散が一致していない変数
- 書き込みがある変数
- データ転送が必要な変数
- 大域変数

これらソート機能とフィルター機能を組み合わせて利用することができる。例えば、データ転送が必要な変数を転送量の順で表示することができる。これらの機能がないと、変数が大量に表示されているところからデータ転送が必要な変数を探していくという、非常に困難な作業が必要になる。図2では変数はMAINでのアクセス回数順に並んでいる。ソート/フィルター機能で手続きやループを指定することで、そのループ(手続き)においてアクセス回数やデータ転送の多い変数を見つけることも可能である。例えば、図でSUB2において書き込みがある変数だけを表示することで、手続き呼びだしで変更される変数一覧を見ることができる。

縦軸に示すコントロールフロー情報は、ループを表示の最小単位として、ループや手続き呼びだしを階層的に表示する。図2のプログラム構造は、MAINルーチンに手続き呼びだしSUB1とDOループがあり、さらにSUB1の中でもループと手続き呼びだしSUB2があることを示している。この表示状態からSUB2を選択すれば、SUB2内部の構造を階層的に表示する。また、表示領域の変更が可能であり、プログラムの注目しているループのみを表示することができる。例えば、図2の手続きSUB1以下のみを表示することができる。図2で、プログラム構造の左横に示した数字は、プログラムをプロファイル実行させた結果に基づいた、ループまたは手続き呼びだしがプログラム全体の実行時間に対して占める割合である。

3.1 変数情報の表示機能

変数に関する情報は、専用のウィンドウを用いて表示する(図3)。表示情報には、データの宣言に関する情報、Equivalence(手続き引数によるものを含む)情報がある。配列変数の場合、ユーザが指定した2次元について配列イメージで表示し、この表示イメージ上で配列のデータ分割やアクセス方向の表示をする。この表示を用いることで、大量のデータ転送が発生する、データ分割とそのアクセスがうまくあっていない場合を容易に見つけることができる。

3.2 並列化作業での本システムの利用

以上述べた本システムの表示機能をプログラム並列化作業で利用することを考えてみる。まず、プログラムで負荷が集中している手続き、さらに負荷が集中しているループを探していくのに、前述の負荷分布の表示が利用

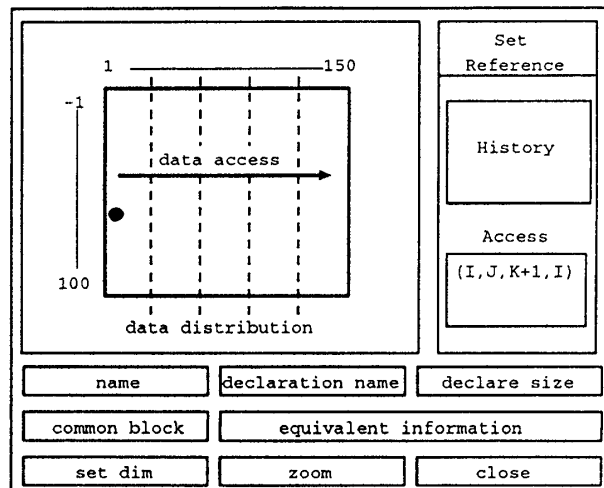


図3: 変数情報の表示

できる。図2では、MAINルーチン中のSUB1に処理が集中しており、さらにSUB1中ではDOループに処理が集中していることが判る。次に、並列化判定ではデータ依存関係情報の表示を利用する。並列化できないと判定された場合には、その判定理由を表示できる。この時に、依存関係に関するユーザ指示を行なうインタフェースを備えており、ユーザ指示の並列化が可能である。

並列化対象ループでは、データのアクセス回数やデータ分割との一致、データ転送量などを見て、データ分割の決定などを行なう。この時、データ分割とアクセス方向の一致に関する情報をプログラム全体で見ること、プログラム全体でデータ転送を減らすデータ分割を容易に選ぶことができる。また、データ分割方法の変更に伴うデータ転送の変化を見ることができる。

4 まとめ

本稿では、プログラム解析助方を表示して並列化を支援する、プログラム解析情報可視化システムPREPSの概要とその表示機能について述べた。本システムにより、プログラムの並列化でタスク分割とデータの分散配置を効果的に決定することができる。本システムは、開発を本年度末までに終了させ、プログラムの並列化における有用性を実証する予定である。

謝辞 最後に、本研究の機会を下さいましたNEC C&C研究所の山本所長、中崎部長、中田課長、また、有益な助言を下さいました皆様方に感謝致します。なお、本研究は科学技術振興調整費科学技術総合研究「数値シミュレーションプログラム開発のための能動的支援方策の研究」の一環として行なった。

参考文献

- [1] M.W.Hall, et al., Experiences Using the ParaScope Editor: an Interactive Parallel Programming Tool, ACM Sigplan Symp. on PPOPP, pp.33-43, 1993.