



メニーコア型クラスタ向けOSカーネル McKernelの開発の現状報告

堀 敦史
理化学研究所

システムソフトウェア技術部会 2015年度活動計画



• システムソフトウェア技術部会

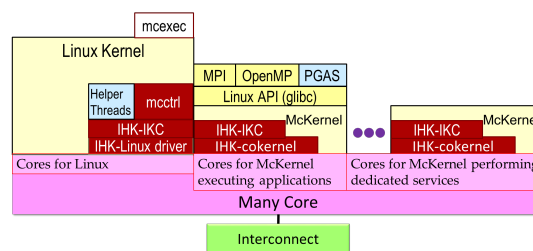
部会長: 堀 敦史 (理化学研究所)
副部会長: 清水 正明 (日立製作所)

◆ 次世代PCクラスタシステムソフトウェア

- ◆ McKernelおよび低レベル通信ライブラリLLCおよびMPI通信ライブラリ(MPICH)の安定稼働および機能拡充の支援として、情報交換のためのWiki page、ソースを管理するためのgit repository、バグ管理のためのbugzillaを管理する
 - ◆ 現在、McKernel, LLC, MPICHの開発は、理研計算科学研究機構、日立、富士通、NECの協力の下進められている
 - ◆ 科学技術振興機構CREST課題「メニーコア混在型並列計算機用基盤ソフトウェア」(代表堀敦史)で提案された新しいプロセス・スレッドモデルPVASのMcKernel移植の支援
- ◆ McKernelリリース&チュートリアル「McKernel, LLC, MPICHを用いた次世代PCクラスタ構築」開催
- ◆ SC15において最新のシステムソフトウェアスタック上でのデモを行うとともにリリース
- ◆ PVAS (Partitioned Virtual Address Space)
- ◆ MPI通信ライブラリ規格書翻訳

軽量マイクロカーネル: McKernel

- 右図はMcKernel動作環境を図示している。McKernel上にglibcが移植されており、McKernelが提供しないLinux APIはLinux側で処理される。アプリケーションプログラムはOSノイズが少ないMcKernel上で稼働する。
- 低レベル通信ライブラリ: LLC
MPI通信ライブラリ、並列プログラミング言語、ファイルI/Oなどで必要とされる低位層通信ライブラリとしてLLCを開発してきている。LLCの参照実装はMellanox Infinibandで行われており、京コンピュータ、FX10にも移植している
- MPI通信ライブラリ: MPICH
MPICHにLLCをポートするとともにアルゴンヌ国立研究所のMPICH開発リポジトリに登録している。今後、アルゴンヌからこのInfiniband版MPICHもリリースされる予定



McKernel開発状況

年月	行数	Linuxのカーネル機能をテストするテストスイート開発プロジェクト Linux Test Project (LTP) が提供するシステムコールテストが1013件中1012件パス
2013年4月	25,900	
2014年6月	34,707	
2015年5月	52,139	

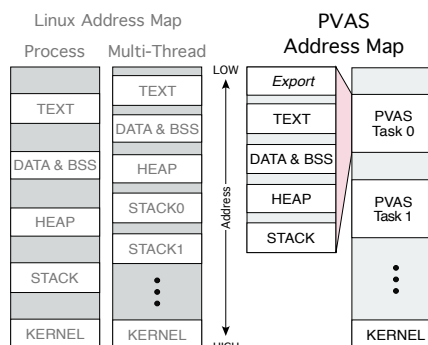
Distribution: 以下のURLから配布予定
<http://www.pcluster.org/ja/download.html>

2015/06/24

PVAS (Partitioned Virtual Address Space)



- ◆ 2015年2月の PCCCワークショップ in 大阪
 - ◆ 従来のプロセスやスレッド代わる新しいタスクモデル
 - ◆ メニーコア時代を意識し、低遅延かつ省メモリなタスク間通信を実現
 - ◆ プロセス: プロセスの境界がプロセス間通信の邪魔となる
 - ◆ スレッド: 変数が共有されてしまうため排他制御が必要
- ◆ PVAS (Paritioned Virtual Address Space)
 - ◆ プロセスモデルとスレッドモデルの「いいとこ取り」
 - ◆ User-Level Process (ULP)
- ◆ 米アルゴンヌ研究所との共同研究
 - ◆ エクサにおいて負荷のアンバランスは不可避
 - ◆ ULP を用いて負荷のアンバランスで生じたコアの空き時間を有効活用
- ◆ プロトタイプ公開 <http://www-sys-aics.riken.jp/releasedsoftware/pvas.html>
 - ◆ 秋に正式版をリリースする予定



2015/06/24

- ◆ MPI version 2.2
 - ◆ MPI の規格を決めている MPI Forum と連携
 - ◆ “Unofficial Version” → 英語版が唯一の “Official Version”
 - ◆ 理研の「有志」6名で内容を確認中
 - ◆ 約 600 ページ
 - ◆ 夏頃に終了する見込み
- ◆ MPI version 3.X (オフィシャル版の最新は 3.1)
 - ◆ 2.2 が終了後, 着手する予定
- ◆ オフィシャル版と同様に “Public Domain” になる予定
 - ◆ コンソーシアムのページで公開 (PDF)
 - ◆ 印刷製本については, 実費で配布することを検討中