仮想マシン環境における MPI プログラム高速実行手法

2010 年 4 月 2 日

本多、近藤研究室
0953016
西川　優

概要

近年、仮想化環境においてMPIプログラムを実行する需要が拡大していく可能性について指摘されている

長所：稼働中の仮想マシンをユーザにとって簡単に別ハードウェアに移動することが可能
短所：パフォーマンスの低下

仮想化マシン環境においてMPIプログラムを高速実行できるような仮想マシンの配置手法を研究していく

関連研究 [1]

各ジョブ Ji にはメモリニーズ、CPU ニーズがあらかじめ付与されている [2]

$$\text{Ji の CPU ニーズ} = \frac{\text{Ji の CPU 使用時間}}{\text{Ji の処理時間}} \qquad 0 < \text{CPU ニーズ} < 1$$

$$\text{Ji のメモリニーズ} = \frac{\text{Ji の使用するメモリ容量}}{\text{物理ホストのメモリ容量}} \qquad 0 < \text{メモリニーズ} < 1$$

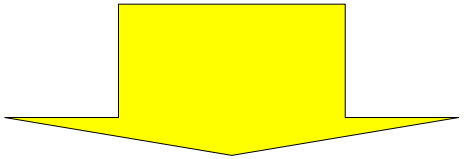
各物理マシンの CPU やメモリの使用状況を考慮しながらジョブを振り分けていく

[1]Mark Stillwell1 David Schanzenbach Frederic Vivien Henri Casanova:
Resource Allocation using Virtual Clusters, *Proceedings of the 2009 9th CCGrid*,Pages260267,2009

[2]William Leinberger,George Karypis,Vipin Kumar:
”Multi-Capacity Bin Packing Algorithms with Applications to Job Scheduling under Multiple Constraints”
Proceedings of the 1999 International Conference *on Parallel Processing*,Page: 404

本研究の目的

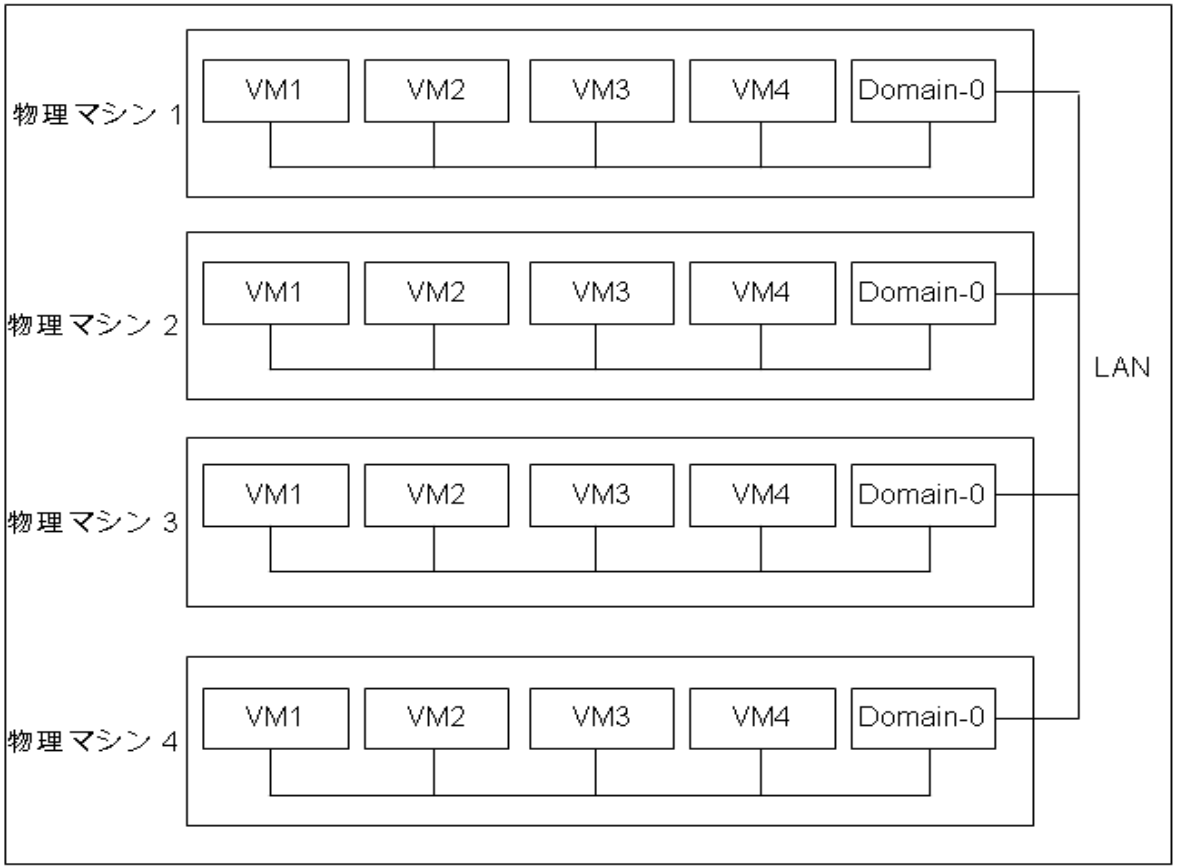
本研究ではCPUとメモリのニーズだけでなく、
通信の頻度もパラメータとして取り入れる



複数のアプリケーションを同時実行させる際、各
アプリケーションの実行時間の合計が最短になる
ようなスケジューリングアルゴリズムを考案する

■各アプリケーションの特徴を抽出するために予備実験を行った

予備実験環境



物理マシンの CPU	Core i7 (Hyper threading off)
物理マシン 1 台あたりのメモリ	6GB
物理マシン数	4
仮想マシンのメモリ	256MB
仮想マシン数	16

Nas Parallel Benchmark

Kernel	
EP	乗算合同法による一様乱数、正規乱数の生成
MG	簡略化されたマルチグリッド法のカーネル
CG	正値対称な大規模疎行列の最小固有値を求めるための共役勾配法
FT	FFTを用いた3次元偏微分方程式の解法
IS	大規模整数ソート
Simulated CFD Application Benchmarks	
LU	Synmetric SOR iterationによるCFDアプリケーション
SP	Scalar ADI iterationによるCFDアプリケーション
BT	5x5 block size ADI iterationによるCFDアプリケーション

EP(class B) の実験結果と考察

仮想マシン数	仮想マシン の配置	実行時間 [秒]
1	●○○○ ○○○○ ○○○○ ○○○○	142.23
2	●●○○ ○○○○ ○○○○ ○○○○	77.19
	●○○○ ●○○○ ○○○○ ○○○○	76.97
4	●●●● ○○○○ ○○○○ ○○○○	38.63
	●●○○ ●●○○ ○○○○ ○○○○	38.35
	●○○○ ●○○○ ●○○○ ●○○○	38.51
8	●●●● ●●●● ○○○○ ○○○○	19.51
	●●○○ ●●○○ ●●○○ ●●○○	17.79
16	●●●● ●●●● ●●●● ●●●●	9.59

表中の□は物理マシン、○は物理コア、●は物理コア上の仮想マシン

- 実行に使用する仮想マシン数を増やしていくと実行時間は短縮した
- 実行に使用する仮想マシン数が同じ場合には通信量が少ないため仮想マシンの配置は実行時間に影響しなかった

CG(class B) の実験結果と考察

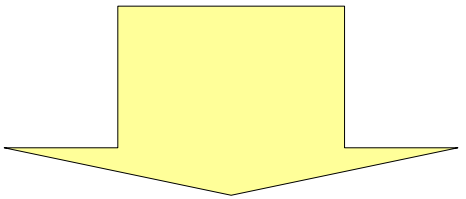
仮想マシン数	仮想マシンの配置	実行時間[秒]
1		83.88
2		46.77
		48.58
4		40.26
		64.60
		52.37
8		115.34
		87.97
16		173.49

表中の□は物理マシン、○は物理コア、●は物理コア上の仮想マシン

- 実行に使用する仮想マシン数が 4 以下の時は 1 つの物理マシン上の仮想マシンのみ使用すると実行時間が最短になることがわかった
- 物理マシン間を隔てた通信を行う場合に同じ物理マシン上で CG の実行に使用する仮想マシン数を増やしていくと、実行時間が長くなる傾向にあることがわかった

まとめ

今回の予備実験から、 Domain-0 が通信においてボトルネックになっていること仮説を立てることができた



仮想マシン環境に合わせたスケジューリングアルゴリズムを考案する必要性があることが明らかになった

今後の予定

- まだ実験していないアプリケーションや問題サイズで実験
- 同時並列的に複数のアプリケーションを実行させて実験
- 各アプリケーションに適切な仮想マシンを選択するスケジューリングアルゴリズムを考案