

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 10 日現在

機関番号：15301

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730046

研究課題名(和文)メニーコアCPUにおける高性能なオペレーティングシステム構成法の研究

研究課題名(英文)On the study of high performance operating system on many-core CPU

研究代表者

山内 利宏(Yamauchi, Toshihiro)

岡山大学・自然科学研究科・准教授

研究者番号：80359942

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：モノリシックカーネルOSのTenderにおいて、排他制御を削減する方式を設計し、実現した結果、LinuxやFreeBSDと同等以上の性能であり、理想値に近い性能を出せることを明らかにした。また、コア毎にOSを起動する方式を実現し、評価した。評価の結果、NUMA構成における提案方式の課題を明らかにし、NUMA構成における設計指針を示した。

研究成果の概要(英文)：In Tender operating system that is a monolithic kernel, we designed a method that reduces mutual exclusion. We evaluated the method. The evaluation results show that the performance of the proposed method is equal to or more than that of Linux and FreeBSD. In addition, we designed a method that can boot an operating system on each core. The evaluation results of the proposed method found the problem of the method in a machine with NUMA. Thus, we investigated the performance of the method in a machine with NUMA, then showed the design principle in a machine with NUMA.

研究分野：基盤ソフトウェア

キーワード：オペレーティングシステム 並列・分散処理 マルチコア 分散OS

1. 研究開始当初の背景

現在はコアが数個のマルチコア CPU が主流であるが、今後は、さらなる高性能化のため、メニーコアと呼ばれる 10 個以上、将来的には数百のコアを有する CPU の登場が予想されている。しかし、既存の Windows や Linux などのオペレーティングシステム(OS)では、マルチコア CPU 環境において、コア数が増加すればするほど、コア当たりの処理性能が低下するという問題がある。これは、OS は、マルチコア環境で、異なるコアで走行するプログラムから同時に処理を要求されると、OS の共有データ領域を排他的に利用する必要があり、排他処理が必要な部分では、OS 内部で同時に一つずつしか要求を処理できないためである。

既存の OS は、様々な機能を取り込んで、高機能化や高性能化を実現しているものの、OS 内部の構造に統一的な設計方針がない。このため、マルチコア CPU やメニーコア CPU に対応する場合に、OS の機能ごとに排他が必要な部分を調査し、それぞれ異なる粒度で排他制御を実現している。また、排他を制御する部分は、OS の機能毎に点在しており、排他によるプログラムの修正箇所は、OS 内に分散して多数存在する。したがって、既存の Linux や Windows の構造では、修正箇所が多くの場合に分散し、OS 機能の追加が難しくなる。また、将来的にコア数が増えた時に、OS の処理性能に限界が生じることになる。

2. 研究の目的

本研究では、メニーコア CPU 環境を想定し、モノリシックカーネル内部において、排他制御を制御するプログラム部分を一箇所に集約するメニーコア CPU 向け OS 構成法を研究する。この構成法は、排他制御が必要な部分を局所化し、CPU コア数にスケールするカーネル処理性能を実現する OS の実現を目指す。また、一つの OS で多数のコアを管理し、スケーラブルな性能発揮するには限界があるため、メニーコア環境では、複数の分散 OS を同時に起動し、コアを各分散 OS で排他的に管理し、分散 OS を連携させる機構の実現を目指す。

3. 研究の方法

モノリシックカーネル OS の **Tender** において排他制御箇所を細粒度化する手法を以下の方法で検討する。

- (1) 少ない修正量で排他制御部分を大幅に削減する方式を設計し、実現する。
OS が制御し管理する対象 (以降、資源) 間の処理の呼び出し関係を調査し、排他制御の無駄な箇所とその理由の明確化する。
- (2) 資源間の処理呼び出し部分の対処で、より粒度の小さい排他制御を実現する方式の設計

現在、資源管理表単位で排他制御を行っており、これにより資源種別単位での排他制御になっている。これを解決するために、各資源を操作する 5 種類の個々のプログラム単位を排他制御単位とする方式を検討し、同一資源種別の異なる資源操作プログラムを同時に実行できる方式を実現する。

- (3) 資源を操作する単位を個々の資源にする方式の実現
さらに排他制御の単位を細分化するため、個々の資源ごとに排他制御を実現する。

次に、マルチコア計算機上で、複数の OS (**Tender**) を起動し、コアを各 OS が分割して管理し、排他制御を大幅に減らす OS 構成法の実現方式を以下の方法で検討する。

- (1) マルチコア CPU 上で、複数の **Tender** を起動させる方式の実現
個々の OS を個別に起動し、ハードウェア資源を排他的に利用する方法、および一つの OS を起動し、その後、OS の資源管理部を複数起動し、ハードウェア資源を各資源管理部で排他的に管理する方法がある。どちらの方式が適切な実現方式かを検討し、実現する。
- (2) マルチコア CPU 上に存在する **Tender** 間で、高速な資源呼び出しを実現する手法の実現

OS 間で高速な通信をするための方法を検討する。分散 OS である **Tender** には、OS 間の RPC が実現されており、リモートの OS への処理依頼方法が実現されている。これをマルチコア向けに再設計し、コア間通信 (IPI: Inter-Processor Interrupt) を利用した高速な RPC の設計を行う。

4. 研究成果

モノリシックカーネル OS の **Tender** において排他制御箇所を細粒度化する手法を検討し、以下のことを明らかにした。

- (1) 資源種別単位での排他制御では、資源の種類ごとに排他制御することで、粗粒度ロックによるマルチコア対応を実現した。
- (2) 資源識別子単位での排他制御では、個々の資源に与えられた識別子ごとに排他制御することで、細粒度ロックによるマルチコア対応を実現した。
- (3) これらの方式の有用性を評価した。評価結果から、細粒度ロックを実現した **Tender** は、細粒度ロックを実現した FreeBSD よりも行数における排他制御量の割合を約 0.12%、Linux よりも約 0.2% だけ少なくしてマルチコア対応を実現していることを示した。
- (4) また、性能向上比の評価では、粗粒度ロックを実現した **Tender** は、操作対象とする資源の種類数以上のコア数では、性能を向上させることができず、コア数の増加に対して有用な方式でないことが

分かった。一方、細粒度ロックを実現した **Tender** は、メモリ処理、プロセス生成処理、およびプロセス間通信処理において4コアの場合、性能向上率が約4.0倍となり、細粒度ロックを実現したLinuxやFreeBSDと同程度以上の性能であり、理想値に近い性能であることを示した。

次に、マルチコア計算機上で、複数のOS (**Tender**) を起動し、コアを各OSが分割して管理し、排他制御を大幅に減らすOS構成法について以下のことを明らかにした。

- (1) 1コアを管理するOSをコア毎に起動する構成を設計した。具体的には、共有データを削減するために、資源をコアごとに用意し、個別に管理するOS構成を提案した。これにより、アプリケーションプログラム(以降、AP)による資源操作における排他制御箇所を削減できる。
- (2) (1)の設計に基づき、本構成の実装を行い、各コアで1コアを管理するOSを起動することを実現した。具体的には、最初に、一つのOSを起動してハードウェア全体を初期化し、利用できる状態にし、その後、OSの資源管理部を複数起動し、ハードウェア資源を各資源管理部で排他的に管理する方法を実現した。Xeonプロセッサ(8コア)を2つ搭載した計算機上でこの機構を評価した。評価の結果、提案手法の構成法では、コア数が8個までは十分な性能が得られたが、コア数が9個以上で十分な性能が出ないことがわかった。この結果から、NUMA構成において性能を十分に出すために、新たな検討が必要であることがわかった。
- (3) さらに、NUMA構成において、プロセッサとメモリの関係について評価を行い、提案手法をNUMA構成の計算機においてどのように実現すべきかの設計指針を示した。
- (4) また、(1)の方式において、各OSは、他OSが管理する資源を利用するために、遠隔手続き呼び出しで利用できる機構を設計した。これにより、提案方式では、APは、走行するコアの資源のみを利用することになり、計算機全体の資源を効率的に利用できない問題に対処する。具体的には、資源操作をコア間で連携させ、遠隔手続き呼出制御により処理を連携できる機能を設計した。
- (5) プロセス間データ用のメモリ空間を利用して、OS内のプロセス間で高速なプロセス間通信を実現する機構を設計した。実装と評価により、提案方式を実現した**Tender**オペレーティングシステムの既存のプロセス間通信より、高速であることを示した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に

は下線)

[雑誌論文] (計 3件)

- [1] 山本貴大, 山内利宏, 谷口秀夫, 排他制御を局所化するマルチコア向け **Tender** の実現, 情報処理学会 コンピュータシステム・シンポジウム(ComSys2013)論文集, vol.2013, pp.14-23, (12, 2013).
- [2] 山本貴大, 山内利宏, 谷口秀夫, 資源の独立化機構により排他制御を局所化するマルチコア向け **Tender** の実現, 情報処理学会論文誌 コンピューティングシステム(ACS), Vol. 7, No. 3, pp. 25-36, (8, 2014).
- [3] Toshihiro Yamauchi, Masahiro Tsuruya, Hideo Taniguchi, Fast Control Method of Software-Managed TLB for Reducing Zero-Copy Communication Overhead, IEICE Transactions on Information and Systems, Vol. E98-D, No. 12, pp. 2187-2191, (12, 2015). DOI: 10.1587/transinf.2015PAL0003

[学会発表] (計 6件)

- [1] 山本 貴大, 山内 利宏, 谷口 秀夫, マルチコア **Tender** における排他制御の細粒度化による並列性向上手法, 並列/分散/協調処理に関する『北九州』サマー・ワークショップ(SWoPP 北九州 2013), 情報処理学会研究報告, vol. 2013-OS-126, no. 16, 電子媒体, (8, 2013).
- [2] 山本貴大, 山内利宏, 谷口秀夫, マルチコア **Tender** における性能向上率の評価, 組込み技術とネットワークに関するワークショップ(ETNET2014), 電子情報通信学会技術研究報告, vol. 113, no. 497, pp. 277-282, (3, 2014).
- [3] 川野 直樹, 山内 利宏, 谷口 秀夫, **Tender** におけるプロセス間通信データ域に特化したプロセス間通信の設計, 並列/分散/協調処理に関する『新潟』サマー・ワークショップ(SWoPP 新潟 2014), 情報処理学会研究報告, vol. 2014-OS-130, no. 22, pp. 1-8, (7, 2014).
- [4] 堀井 基史, 山内 利宏, 谷口 秀夫, **Tender** におけるコアごとに資源を用意し個別に管理するOS構成の設計, 情報処理学会第131回システムソフトウェアとオペレーティング・システム研究会, 情報処理学会研究報告, vol. 2014-OS-131, no. 7, pp. 1-8, (11, 2014).
- [5] 堀井 基史, 山内 利宏, 谷口 秀夫, OS資源をコア毎に管理する **Tender** の性能評価, 情報処理学会第135回システムソフトウェアとオペレーティング・システム研究会, 情報処理学会研究報告, vol. 2015-OS-135, no. 15, pp. 1-6, (11, 2015).
- [6] 堀井 基史, 山内 利宏, 谷口 秀夫, **Tender**

におけるコアメモリ間の関係が性能に与える影響, 情報処理学会第136回システムソフトウェアとオペレーティング・システム研究会, 情報処理学会研究報告, vol. 2016-0S-136, no. 14, pp. 1-7, (3, 2016).

〔図書〕(計 0件)

〔産業財産権〕

○出願状況(計 0件)

○取得状況(計 0件)

〔その他〕

なし.

6. 研究組織

(1) 研究代表者

山内 利宏 (YAMAUCHI TOSHIHIRO)

岡山大学・大学院自然科学研究科・准教授

研究者番号: 80359942