

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：10101

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26540040

研究課題名(和文)大規模分散処理環境を革新する「フロントエンドコンピューティング」パラダイムの創出

研究課題名(英文)Frontend computing paradigm toward accelerating large-scale distributed computing platform

研究代表者

本村 真人 (Motomura, Masato)

北海道大学・情報科学研究科・教授

研究者番号：90574286

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)：大規模分散処理環境のフロントエンドコンピューティングコンセプトによる加速処理の一例として、Memcachedのネットワーク・インタフェースによる高速化手法とそのハードウェア化に取り組み、ソフトウェアに比べて最大3.5倍の性能向上を確認した。また、別の一例として、近似最近隣探索アルゴリズムLSHに着目し、その一つであるExact Euclidean Locality Hashingのハードウェア向けアルゴリズムへの改良とそのハードウェア化に取り組み、ソフトウェアに比べて1/9のレイテンシ短縮と約25倍のスループット向上を確認した。

研究成果の概要(英文)：As an example for accelerating large-scale distributed processing platform based on front-end computing concept, we have worked on acceleration of memcached on network interface, and its hardware implementation. We have found 3.5 times speed improvement compared to existing software solution. We have also worked on approximate nearest-neighbor search algorithm LSH, especially Exact Euclidean Locality Hashing. By improving the algorithm and by implementing it on hardware, we have found 1/9 latency and around 25 times throughput improvement over existing software solution.

研究分野：離婚フィギュラブルシステム

キーワード：データマイニング ビッグデータ ストリーム処理

1. 研究開始当初の背景

「ビッグデータ時代への相転移」を可能にするために必要な大量データの記憶・処理は、ハードウェア(HW)技術の発展に頼らず(=なるべく安価で技術的にも熟した HW を利用)、大規模分散処理を指向する多層的なソフトウェア(SW)技術が進展することでもたらされてきた。その代表的な例としては、Google による BigTable、MapReduce、オープンソースで開発が進む Hadoop、memcached 等があげられる。これらの技術では、既存のサーバを大量に並べて高速ネットワーク(NW)で接続する HW 構成を前提とし、その上でサーバ台数に応じて性能が上がる(Scale Out する)SW 的な仕組みを構築することで、大規模分散処理環境を実現していた。一方、計算機アーキテクチャの世界では、そのような大きな流れとあまり交わることなく、CPU システムのアーキテクチャ最適化研究が依然主流を占めてきた(マルチコア化、メモリ階層の最適化、仮想化支援機構、等々)。これらの技術はサーバ内部の高々年率数%程度の性能・電力改善技術でしかなく、ビッグデータ時代に向けた大きな潮流と従来アーキテクチャ技術の成立ちとの間にミスマッチがあった。

2. 研究の目的

しかしながら、「大規模データセンターか所に発電所が一つ必要」と言われるほど大規模分散処理システムが大電力を消費する現実を踏まえると、SW のみの工夫で、サーバ数を増やして処理能力を上げる Scale Out 型アプローチだけでは持続的な解とはならない。サーバ単体ではなく、大規模分散処理システム全体を見通し、データ処理の高速化・低エネルギー化を実現するアーキテクチャ技術が今こそ必要であった。

3. 研究の方法

本研究では、これまでサーバ内 CPU システム(バックエンドと呼ぶ)上の SW が担当して

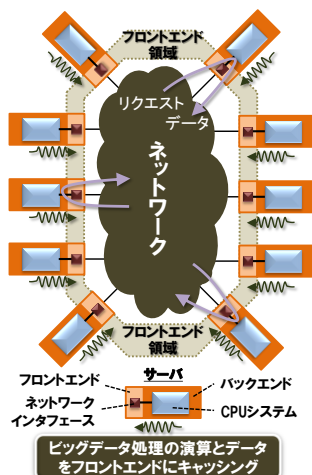


図 1. フロントエンドコンピューティングのコンセプト

きた大規模分散処理の計算機能と記憶機能の一部を、NW インタフェース部(フロントエンド)のリコンフィギュラブル HW に「キャッシング」することで、より高速・低電力なビッグデータ処理の実現を目指す(図 1)。

大規模分散処理システム全体としてこれを捉えるならば、NW の海に隣接する海岸線たる「フロントエンド領域」で実行時間の大部分を占める処理をカバーする系を実現する。具体的には、基本ライブラリ memcached 等をビークルとして、研究期間内にコンセプト具体化、システム試作・評価、高性能化・低エネルギー化の効果実証を進め、より大きな「フロントエンドコンピューティング」パラダイムの創出につなげる。

4. 研究成果

Memcached の NIC Cache による高速化手法の改良とハードウェア実装を中心に研究に取り組んだ。Memcached は大規模データセンターなどでウェブサーバの応答を高速化するために用いられるデータベース技術であり、Memcached サーバのデータと処理の一部を同じサーバの Network Interface Card (NIC) においてキャッシュすることで Memcached を高速化するシステムの有効性をソフトウェアシミュレーションにより検証した(図 2)。

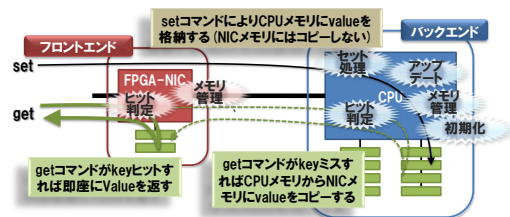


図 2 Memcached のキャッシング

課題は、①NIC 上のキャッシュ容量が小さい問題を解決することと、②ハードウェア実装を行うことの 2 点であった。①の問題は、FPGA 上のハッシュテーブルを実装するための Block RAM の容量が少ないことが原因であった。そこでハッシュテーブルに保持するタグの幅を減らすことでハッシュテーブルのエントリ数を増やし、同容量の Block RAM でキャッシュ容量を増やす方法を考案した。また②の問題に関しては、提案した設計仕様を元に FPGA 搭載 NIC 上に実装を進め、レイテンシを通常のソフトウェアによる Memcached と比較して最大 3.5 倍の改善を達成した。FPGA 搭載 NIC は 20Gbps のネットワークインタフェース、FPGA (Virtex-5)、1GB の DRAM メモリ、そして 20Gbps の PCI Express インタフェースを備えている。

また、近年研究されている近似最近傍探索アルゴリズム Locality-Sensitive Hashing (LSH) に着目し、その一つである Exact Euclidean Locality-Sensitive

Hashing (E2LSH) を元に(図 3)、ハードウェア指向 LSH アルゴリズムへと改良した。

本研究で提案するアルゴリズムの特徴として、二重ハッシングとセットアソシアティブ方式を取り入れた「限定動的テーブル」と

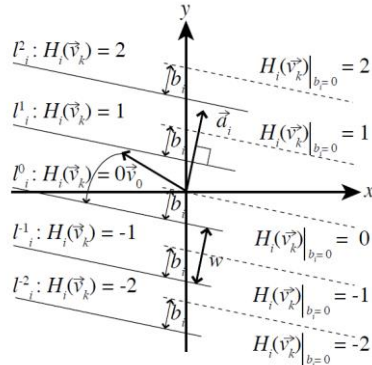


図 2. ユークリッド距離による LSH,

いうデータ構造を用いている。これにより限られたハードウェア資源をうまく使うことができ、実装が容易になった。アルゴリズムの精度として Recall 率を計測したところ、E2LSH と同様のテストにより、約 82%を達成した。また本論文では、そのハードウェアアーキテクチャを提案している。提案するアーキテクチャでは、Liner Feedback Shift Register (LFSR) と、中心極限定理を用いて、回路内部で乱数を生成する方法を取った(図 4、図 5)。これによりメモリ使用量を抑えつつ複数の乱数列を生成することが出来た。ま

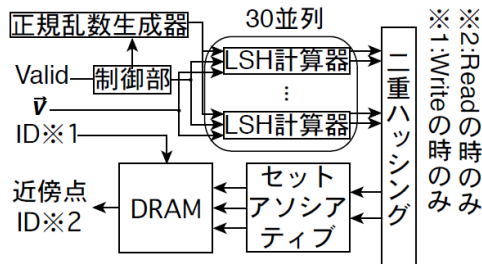


図 4. アーキテクチャ全体図

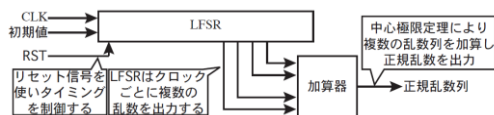


図 5 LFSR による正規乱数発生器

たパイプライン化により、データストリームを止めることなく処理することが可能になっており、高速なアーキテクチャといえる。結果として、CPU に比べ、レイテンシは 1/9 に短縮、スループットは 24.4 倍に増加させることができた。

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 2 件)

1. Yamamoto K., Ikebe M., Asai T., and Motomura M., "FPGA-based stream processing for frequent itemset mining with incremental multiple hashes," Circuits and Systems, vol. 7, (2016), in press. (査読有)
2. Fukuda E. S., Inoue H., Takenaka T., Kim D., Sadahisa T., Asai T., and Motomura M., "Enhancing memcached by caching its data and functionalities at network interface," IPSJ Journal, vol. 56, no. 3, pp. 143-152 (2015). (査読有)

[学会発表] (計 6 件)

1. Yamamoto K., Asai T., and Motomura M., "Hardware architecture for online frequent items mining with memory-efficient data structure," COOL Chips XIX, Yokohama Media & Communications Center, Yokohama, Japan (Apr. 20-22, 2016). (査読有)
2. Yamamoto K., Fukuda E. S., Asai T., and Motomura M., "An accelerator for frequent Itemset mining from data stream with parallel item tree," The 19th Workshop on Synthesis And System Integration of Mixed Information Technologies, Evergreen Resort Hotel, Yilan, Taiwan (Mar. 16-17, 2015). (査読有)
3. Fukuda E. S., Inoue H., Takenaka T., Kim D., Sadahisa T., Asai T., and Motomura M., "Achieving higher performance of memcached by caching at network interface," The 2014 International Conference on Field Programmable Technology, Parkyard Hotel, Shanghai, China (Dec. 10-12, 2014). (査読有)
4. Kim D., Hida I., Fukuda E. S., Asai T., and Motomura M., "A study of transparent on-chip instruction cache for NV microcontrollers," The 7th International Conference on Advances in Circuits, Electronics and Micro-electronics, Mercure Lisboa, Lisbon, Portugal (Nov. 16-20, 2014). (査読有)
5. Kim D., Fukuda E. S., Sadahisa T., Asai T., and Motomura M., "Hardware architecture for accelerating key-value retrieval implemented on FPGA," The 3rd Japan-Korea Joint Workshop on Complex Communication Sciences, Paradise Hotel, Busan, Korea (Oct. 27-28, 2014). (査読有)
6. Fukuda E. S., Inoue H., Takenaka T., Kim D., Sadahisa T., Asai T., and Motomura M.

M., "Caching memcached at reconfigurable network interface," The 24th International Conference on Field Programmable Logic and Applications, Technische Universität München, Munich, Germany (Sep. 2-4, 2014). (査読有)

〔図書〕 (計 1 件)

天野英晴編著、「FPGA の原理と構成」、オーム社、ISBN978-4-274-21864-4、8 章、2016

〔産業財産権〕

○出願状況 (計 0 件)

○取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://lalsie.ist.hokudai.ac.jp/jp/>

6. 研究組織

(1) 研究代表者

本村 真人 (Motomura, Masato)

北海道大学・情報科学研究科・教授

研究者番号：90574286

(2) 研究分担者

浅井 哲也 (Asai, Tetsuya)

北海道大学・情報科学研究科・準教授

研究者番号：00312380