B-015

ディスアグリゲーテッドコンピュータにおける高効率なデータ処理に向けたマッピング手法

A mapping method for highly efficient data processing in Disaggregated Computer

榎原 成則† 右近 祐太† 有川 勇輝† 山崎 晃嗣†
Narunori Ebara Yuta Ukon Yuki Arikawa Koji Yamazaki

1. はじめに

データセンタの高性能化，高効率化に向けて，XPU ,FPGA，CPU，メモリ等のデバイスを分散し，必要に応じて柔軟に活用することでアプリケーションに最適な計算リソースを提供するディスアグリゲーテッドコンピュータ(DC)[1]が注目されている．従来の CPU セントリックな技術では，デバイス間の接続に CPU を介在させる必要があるため，CPU リソースの増加や通信オーバーヘッドが生じる．この問題を解決するためには，高効率かつ柔軟にデバイスを接続する技術が必要となる．そこで，本稿では CPU を極力介在させずに，アプリケーション構成要素(ファンクション)を配置するデバイスを数珠繋ぎに直接接続するファンクションチェイン(FC)技術を検討している[2]．尚，ファンクションはデバイスを機能単位に分割した処理機能部として実装する．

FC 技術を用いたデータ転送において CPU を非介在とするには，アプリケーションの実行前に構成要素であるファンクションの配置および接続方法(マッピング問題)を決定する必要がある．本問題の類似研究として，仮想マシン(VM)のマッピング問題を扱う事例があり，VM の配置を数理計画モデルに帰着して数理最適化を行う方法が提案されている[3,4]．しかし，これらの数理計画モデルでは，単一種類のスイッチ，CPU，メモリから構成される CPU セントリックなコンピュータアーキテクチャしか扱えないため，複数種類のデバイス，スイッチから構成される高度なコンピュータアーキテクチャおよび，データセントリックな処理を扱うことはできない．そこで，本稿では複数種類のデバイス，スイッチの特性，FC 技術を考慮して，DC におけるファンクションの配置，接続方法を新たな数理計画モデルに帰着し，所望の性能指標(目的関数)を最適にするようにファンクションをマッピングする方法を提案する．

2. 提案手法

2.1 節で論じる数理計画モデルを図 1 の手順により実施することで，目的関数の計算結果，ファンクションの配置情報と接続情報を含むマッピング情報等を得る方法を提案する．まず，アプリケーション要件とアプリケーションの構成要素のファンクション情報を入力として，提案技術[5]によりファンクションの性能差を緩和するようにファンクションを並列化してチェイン構築を行い，チェイン情報を得る．尚，本稿ではアプリケーションの構成要素であるファンクションの順序関係を有する組み合わせをチェインと呼ぶ．次に，提案技術[5]で得られたチェイン情報と与えられたシステム構成を数理計画モデルに入力し，一般に広く用いられる最適化ソルバー(GUROBI，CPLEX 等)によって

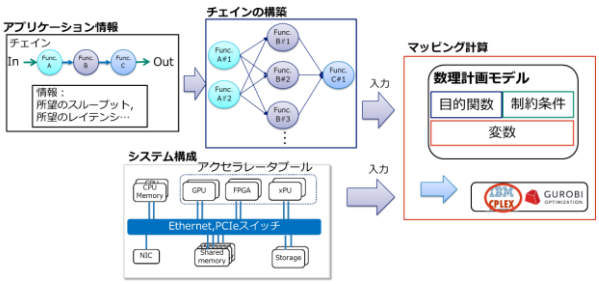

図 1　提案手法概要図

数理最適化を行うことにより，目的関数の計算結果とマッピング情報を得る．

2.1 数理計画モデル

2 種類のスイッチ(PCIe, Ethernet)，内外の伝送経路と複数種類のデバイスで構成される DC のアーキテクチャ[6]におけるファンクションチェインのマッピング問題を解決するために新たな数理計画モデルを提案する．本稿では 複数種類のデバイス，スイッチおよび FC 技術を考慮して，連続変数と整数変数で表す決定変数を用いて制約条件，目的関数を線形関数で記述する混合整数線形計画(MILP)モデルを新たに構築する．

2.1.1 制約条件

提案する数理計画モデルでは，ファンクションの配置制約とファンクション間の接続制約，アプリケーション要件を満たすためのチェイン制約の大きく 3 カテゴリの制約条件を設ける．各制約について以下説明する．

配置制約として下記 3 つの制約を設ける．

- ファンクションを必ず 1 デバイスに割り当てる制約
- 1 つのデバイスに許容容量以下でファンクションを複数割り当て可能である制約
- デバイスが搭載できるファンクション数，種類，速度を限定する制約

上記制約により，デバイスの容量だけでなく，最大処理速度，配置できるファンクション数、ファンクションの種類を考慮でき，アクセラレータを含む複数種類のデバイスにファンクションを割り当て可能となる．

接続制約として下記 3 つの制約を設ける．

- ファンクションを配置したデバイスや使用するスイッチに伝送経路が接続される制約
- 伝送経路の整合を担保する制約
- 伝送経路の帯域幅を超えない制約

上記制約により，DC のスイッチ，伝送経路を考慮し，伝送経路の流量や伝送経路の整合性を担保する．

FC 技術では，CPU セントリックな処理とは異なり，柔軟にファンクション間を接続するために接続方法が複雑化する．そのため，従来技術にはない要素として，本数理計画モデルでは接続制約設けている．

チェイン制約として下記 2 つの制約を設ける．

† 日本電信電話株式会社デバイスイノベーションセンタ
Device Innovation Center, NTT Corporation

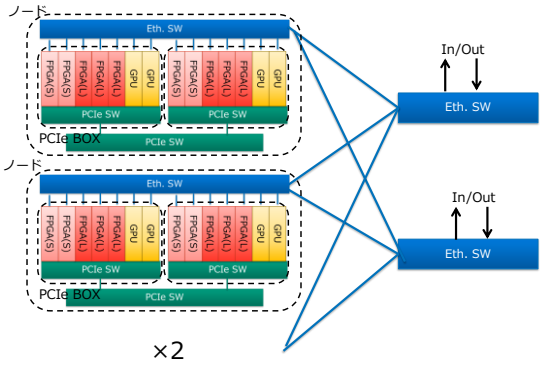

図 2　DC 概要図

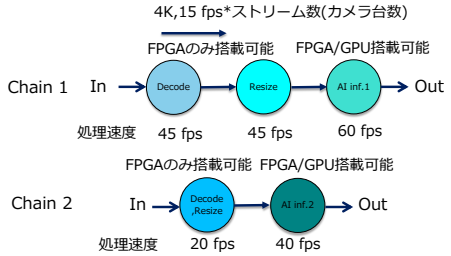

図 3　映像 AI 推論処理アプリケーション

- チェイン要求の伝送時間，出力間隔を満たす制約
- 同一チェインで伝送経路を決定する制約

上記制約により，ユーザからのアプリケーション要件を満たすようにチェインを構築することができる．

2.1.2 目的関数

提案する数理計画モデルでは，消費電力，ネットワーク帯域，レイテンシなどの様々なパラメータを目的関数として扱うことができる．本稿では，目的関数を実機の電力に見立てた電力スコアとし，電力スコアの最小化問題を解く．電力スコアは，デバイスの静的電力$W_{\text{node}}^{\text{total}}$，ファンクションの動的電力$W_{\text{function}}^{\text{total}}$，2 種類のスイッチの電力$W_{\text{PCIe}}^{\text{total}}$，$W_{\text{Ethernet}}^{\text{total}}$の総和とする(式 1).

$$Object = W_{\text{node}}^{\text{total}} + W_{\text{function}}^{\text{total}} + W_{\text{PCIe}}^{\text{total}} + W_{\text{PCIe}}^{\text{total}} \quad [1]$$

3. シミュレーション条件

一定間隔，一定流量で流れる 4K,15 fps 映像ストリームに対して 2 種類の映像 AI 推論処理アプリケーションを図 2 に示す DC にマッピングするシミュレーションを行った．2 種類のアプリケーションは，それぞれ図 3 に示すチェイン 1(Decode, Resize, AI inf.1)とチェイン 2(Decode and Resize, AI inf.2)とする．本シミュレーションで扱う DC は， 2 階層で構成されるリーフスパイン型のアーキテクチャを想定しており，内部伝送経路は PCIe でノード内を接続し，外部伝送経路は Ethernet でノード内外を接続する．なお，全てのデバイスは PCIe，Ethernet の両方と接続する．デバイスは 2 種の性能の異なる FPGA と 1 種の AI 推論処理に特化した GPU の 3 種類であり，最大数は 56 個である．尚，本実験における GPU は，大規模な AI 推論処理に関して FPGA より電力性能比が良いものとして定義している．スイッチは Ethernet スイッチと PCIe BOX を含む PCIe ファブリックスイッチの 2 種類とし，それぞれ最大数は 6 個，12 個である．

本シミュレーション結果は，Python3.9 で構築した MILP モデルを CPLEX22.1.1 最適化ソルバーによって双対ギャップ 1%以下として数理最適化を行って得た結果である．

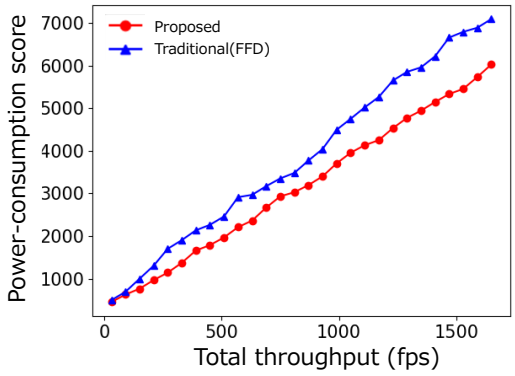

図 4　所望のスループットにおける電力スコア

3.1 シミュレーション結果

図 4 に提案手法と従来手法 FFD(First-Fit Decreasing)における電力スコアを示す．図 4 の x 軸は 15 fps*ストリーム数で表すスループット(fps)の合計値，y 軸は電力スコアを示す．図 4 の赤丸は，提案手法により得られた解を示す．青三角は，従来のマッピング問題で一般的に用いられる FFD によりファンクションを配置し，得られた解である．FFD は，詰められるデバイスの中で最も添字番号の小さいデバイスから順にファンクションを配置する手法である．尚，ファンクション間の接続は従来手法では考慮されていないため，提案する数理計画モデルを用いて決定した．

本結果より，本制約条件下において提案手法を用いてファンクションチェインをマッピングすることで従来手法に比べて，電力スコアを平均約 20%削減できるということが分かった．これによって，提案手法によりファンクションチェインをマッピングすることで，電力性能比を向上することができる．

4. おわりに

本稿では，従来手法では扱えない複数種類のデバイス，スイッチおよび FC を考慮して数理計画モデルを構築し，FC を用いる DC のマッピング方法を提案した．リーフスパイン型の DC に 2 種類の映像 AI 推論処理アプリケーションをマッピングするシミュレーションにより，提案手法を用いて、ファンクションチェインをマッピングすることで，従来手法に比べて電力スコアを削減し，電力性能比を向上できることを示した．今後は，数理計画モデルの改良や提案する数理計画モデルの求解手法の探索を進める．

参考文献

[1] Kiyo Ishii, *et al.*, "Disaggregated optical-layer switching for optically composable disaggregated computing [Invited],"JOCN, vol. 15, no. 1, pp. A11-A25, (2023).
[2] 樽林 亮介ら，”IOWN 時代のデータ処理を支えるデータセントリック基盤とそのコンセプト実証”, NTT ジャーナル, (2023).
[3] Pages, Albert, *et al.*, "On the benefits of resource disaggregation for virtual data centre provisioning in optical data centres." Computer Communications, vol. 107, pp. 60-74, (2017).
[4] L. Ferreira, *et al.*, "Optimizing Resource Availability in Composable Data Center Infrastructures," LADC, Natal, Brazil, pp. 1-10, (2019).
[5] 榎原 成則ら,”光ディスアグリゲーテッドコンピュータの実現に向けたアクセラレータ間通信の高効率化に関する検討”,電子情報通信学会ソサイエティ大会, C-12-19, (2023).
[6] “Data-Centric Infrastructure Functional Architecture”, pp. 14, (2023). https://iowngf.org/wp-content/uploads/2023/04/IOWN-GF-RD-DCI_Functional_Architecture-2.0.pdf