



Title	Optical Micro Disaggregated Data Centers Capable of Executing Many Tasks Simultaneously
Author(s)	生駒, 昭繁
Citation	大阪大学, 2025, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/101762
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏 名 (生駒 昭繁)	
論文題名	Optical Micro Disaggregated Data Centers Capable of Executing Many Tasks Simultaneously (多くのタスクを同時実行可能な光リソース分離型マイクロデータセンター)
論文内容の要旨	
<p>近年、多くのサービスがクラウドコンピューティングによって提供されているが、クラウドデータセンターとの高い通信遅延や通信トラフィックの増大が問題となっている。そこで、ユーザの近くでデータを処理することができるエッジコンピューティングが注目されている。エッジコンピューティングでは、小規模なデータセンター（以下、マイクロデータセンター）をユーザの近くに配置し、サービスに対する処理を行う。ただし、マイクロデータセンターは、大規模なクラウドデータセンターに比べて保有資源（CPU や GPU 等）が少ないという問題があり、多くのサービスタスクを実行するためには、限られた資源の効率的な利用が必要となる。本論文では、効率的な資源利用に有用なアーキテクチャである、resource disaggregation に着目し、resource disaggregation を適用したマイクロデータセンター（以下、リソース分離型マイクロデータセンター（μDDC）と呼ぶ）によって、多数のサービスタスクを同時に実行できるマイクロデータセンターを構成する。</p> <p>μDDC においてタスクを実行するとき、タスクの実行資源間でネットワーク通信が発生する。このような通信の頻度は高いため、資源間の通信遅延はサービス性能に大きな影響を与える。そこで、低遅延かつ高帯域で安定通信ができる光インターコネクトが有用である。しかし、資源間の通信遅延はインターコネクトにのみ依存せず、資源割当やネットワークトポロジにも依存する。そこで、本論文では、多くのタスクを実行できる μDDC を構成するための資源割当手法とネットワークトポロジを提案する。</p> <p>まず、多数のタスクを同時に実行するために、性能要求を満たしながら将来のタスクに必要な資源をできるだけ残す資源割当手法 RA-CNP を提案し、その有効性を評価する。その結果、従来の手法においてタスクの棄却が発生する環境においても、RA-CNP は棄却を 0 に抑えることができることを示した。</p> <p>次に、柔軟な資源割当を可能とするために、性能要件を満たす通信が可能な資源ペアを多く持つ光ネットワークトポロジ設計を提案する。RA-CNP を用いた資源割当評価によって、提案したネットワークトポロジ設計は、従来のネットワークトポロジと比較してタスクの資源割当の棄却を 50% 以上削減できることを示した。</p> <p>最後に、各タスクの割当資源数を最適化するために、RA-CNP を拡張する。本論文では、深層学習モデルの分割による並列実行に着目し、分割を最適化することで各タスクに割り当てられる資源数を最適化する。評価の結果、提案手法は同時実行できるタスクの数において、従来手法よりも最大30%改善できることを示した。</p>	

論文審査の結果の要旨及び担当者

氏 名 (生 駒 昭 繁)			
	(職)	氏 名	
論文審査担当者	主 査	教授	村田 正幸
	副 査	教授	渡辺 尚
	副 査	教授	山口 弘純
	副 査	教授	下西 英之

論文審査の結果の要旨

近年、多くのサービスがクラウドコンピューティングによって提供されているが、クラウドデータセンターとの高い通信遅延や通信トラフィックの増大が問題となっている。そこで、ユーザの近くでデータを処理することができるエッジコンピューティングが注目されている。エッジコンピューティングでは、小規模なデータセンター（以下、マイクロデータセンター）をユーザの近くに配置し、サービスに対する処理を行う。ただし、マイクロデータセンターは、大規模なクラウドデータセンターに比べて保有資源（CPU や GPU 等）が少ないという問題があり、多くのサービスタスクを実行するためには、限られた資源の効率的な利用が必要となる。本論文では、効率的な資源利用に有用なアーキテクチャである、resource disaggregation に着目し、resource disaggregation を適用したマイクロデータセンター（リソース分離型マイクロデータセンター（ μ DDC））によって、多数のサービスタスクを同時に実行できるマイクロデータセンターを構成する。

μ DDC においてタスクを実行するとき、タスクの実行資源間でネットワーク通信が発生する。このような通信の頻度は高いため、資源間の通信遅延はサービス性能に大きな影響を与える。そこで、低遅延かつ高帯域で安定通信ができる光インターコネク트가有用である。しかし、資源間の通信遅延はインターコネク트에のみ依存せず、資源割当やネットワークトポロジにも依存する。そこで、本論文では、多くのタスクを実行できる μ DDC を構成するための資源割当手法とネットワークトポロジを提案する。

まず、多数のタスクを同時に実行するために、性能要求を満たしながら将来のタスクに必要な資源をできるだけ残す資源割当手法 RA-CNP を提案し、その有効性を評価する。その結果、従来の手法においてタスクの棄却が発生する環境においても、RA-CNP は棄却を 0 に抑えることができることを示した。

次に、柔軟な資源割当を可能とするために、性能要件を満たす通信が可能な資源ペアを多く持つ光ネットワークトポロジ設計を提案する。RA-CNP を用いた資源割当評価によって、提案したネットワークトポロジ設計は、従来のネットワークトポロジと比較してタスクの資源割当の棄却を 50% 以上削減できることを示した。

最後に、各タスクの割当資源数を最適化するために、RA-CNP を拡張する。本論文では、深層学習モデルの分割による並列実行に着目し、分割を最適化することで各タスクに割り当てられる資源数を最適化する。評価の結果、提案手法は同時実行できるタスクの数において、従来手法よりも最大30%改善できることを示した。

以上のように本論文は、多くのタスクの同時実行可能性という観点から、リソース分離型のマイクロデータセンター・アーキテクチャの実現に対する有用な研究成果をあげている。よって、博士（情報科学）の学位論文として価値のあるものと認める。