

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches



笹部 昌弘 (Masahiro Sasabe, Ph. D.)

関西大学 総合情報学部 教授

(Professor, Faculty of Informatics, Kansai University)

電子情報通信学会、IEEE、ACM

研究専門分野：情報ネットワーク、数理最適化、Networking for AI/AI for networking

あらまし

サービスチェイニング(Service Function Chaining: SFC)は、ネットワークサービスを1つ以上の機能列で構成し、それらをサービスの提供者と利用者間で確立したサービスパスと呼ばれる経路上の中継ノードにおいて順次実行する技術である。各機能を汎用サーバ上で動作するソフトウェアとして実現することで、多様なサービスを迅速かつ柔軟に展開できる。サービスチェイニングは、数学的には容量制約付き最短経路ツアー問題(Capacitated Shortest Path Tour Problem: CSPTP)と呼ばれるNP困難な問題として定式化される。本研究では、ラグランジュ緩和法と最短経路ツアーアルゴリズムの融合による軽量な解法、サービス需要やトポロジ変化への適応力を備えた機械学習に基づく手法を提案する。数値評価を通して、サービスパスの最適性、環境変化への適応力、計算量などの観点から、提案手法の有効性・実用性が確認された。

1. 研究の目的

現在の通信インフラは、従来の専用ハードウェアによる固定的な構成から、汎用サーバ上で動作するソフトウェア群への移行の過渡期にある。その中核をなすのがネットワーク機能仮想化(Network Functions Virtualization)*1であり、その上で展開されるサービスチェイニング(SFC)*2は、ファイアウォールやロード

バランサといったネットワーク機能をソフトウェアで実現し、サービスの要求に応じて連結することで、動的にサービスパスを構築する技術である。ここで、ソフトウェア化の技術として、当初想定されていた仮想マシン上で動作する仮想ネットワーク機能(Virtual Network Function: VNF)*3に加え、より軽量なクラウドネイティブネットワーク機能(Containerized Network Function: CNF)*4への展開も進められている。

従来の専用機器ベースの運用と比較し、ソフトウェア化されたNFVネットワークでは、新規サービスの追加や構成変更の即時性が期待されている。しかしながら、動的なネットワーク環境において、サービスごとの要件に応じて、必要となる計算資源(CPUやメモリ)や通信リンクの帯域幅を適切に割り当てたサービスパスを確立するためには、高度な数学的・技術的課題を伴う。

この課題に対し、既存の最適化問題に対する高速な解法や需要・トポロジ変化への適応力を備えたサービスパスの構築手法を確立することで、数理的・実践的アプローチによる持続可能なサービスチェイニングの実現を目指す。

2. 研究の背景

サービスパスの最適化は、グラフ理論における最短経路ツアー問題(Shortest Path Tour Problem: SPTP)の拡張に位置付けられる。SPTP自体は、特定のノード群を所望の順序で巡る最短経路を求める問題となるが、サービスパスの構築においては、各ノードの処理能力や通信リンクの帯域幅に対する容量制約が加わった、容量制約付き最短経路ツアー問題(CSPTP)となる。これまでに、CSPTPは整数線形計画問題(Integer Linear Programming: ILP)*5として定式化できることが示されているが、NP困難な問題となるため、計算量と最適性のバランスのとれた解法が求められている[1]。

また実環境では、サービス需要の変動、リンク故障によるトポロジの変化など、ネットワーク環境が時々刻々と変化する。こうした変動に対し、機械学習や強化学習を用いた適切なサービスパスの予測や構築手法

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches

が検討されている[2,3]。本研究では特に、グラフニューラルネットワーク(Graph Neural Network: GNN)*6を用いた深層強化学習(Deep Reinforcement Learning: DRL)*7によるルーチング[4]を参考に、ルーチングよりも複雑な問題となるサービスチェイニングへの応用を目指す。

3. 研究の方法・結果

まず 3.1 節において、本研究で対象とするサービスチェイニングの概要を示す。その後の 3.2 節と 3.3 節では、サービスチェイニングを効率的に計算するための手法、サービス需要やネットワークなど環境変化への適応力を備えたサービスチェイニングの実現手法について順に紹介する。

3. 1 サービスチェイニング

図 1 にサービスチェイニングの概要を示す。サービスチェイニングでは、ユーザからのサービスチェーン要求(Service Chain Request: SCR)に対し、物理ネットワーク上で適切なノード、機能、通信リンクを割り当てたサービスパスを構築する。図では 3 つの機能列で構成されたサービスを想定しており、SCR には、始点ノード、終点ノード、必要な機能の列、各機能が要する CPU 資源量、通信帯域が含まれる(図 1 上部)。物理ネットワークは、始点・終点を含む複数の物理ノードとそれらの間の通信リンクで構成される。物理ノードと通信リンクの資源は有限であり、以降では CPU と帯域を容量制約の対象とする。また、SCR で指定された機能毎に架空ノードを用意し、その機能を実行可能な物理ノードとの間に仮想リンクを追加することで、拡張ネットワークを形成する(図 1 中部)。最終的に得られるサービスパスは始点から終点へと、所望の順番に中継ノード上で機能を実行可能な通信経路として描かれる(図 1 下部)。この例では機能数が 3 であるため、4 本の部分パスの連結によりサービスパスが構成されている。

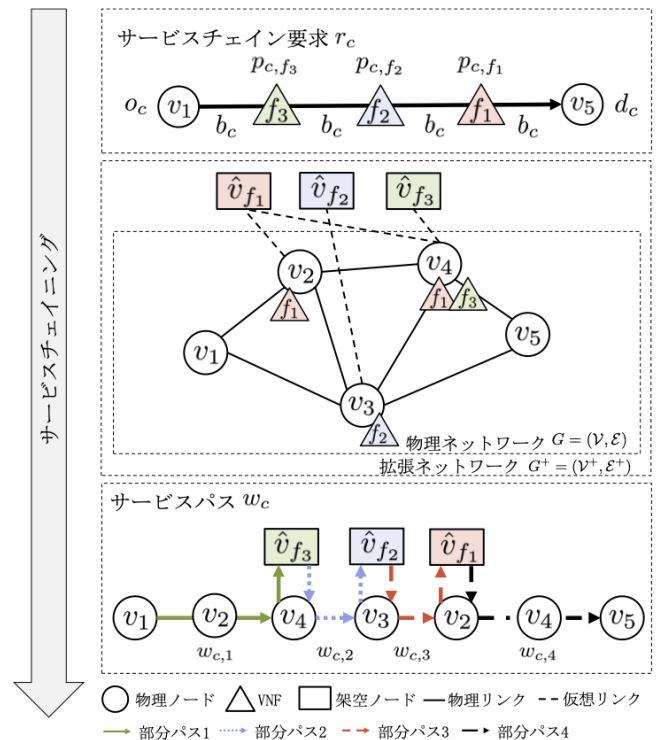


図 1 サービスチェイニング

本研究では、サービスパスの遅延を各リンクにおける伝播遅延、中継ノード上での転送遅延、機能の実行時間の総和として定義し、これが最小となるようなサービスパスの構築を目指す。この問題は CSPTP-based ILP として定式化される[1]。

3. 2 迅速性と最適性を考慮したサービスチェイニング

サービスチェイニングは CSPTP-based ILP となるため、既存の線形ソルバ*8が解法の 1 つとして挙げられる。ただし、ネットワークの規模増大にともない、最適解の導出に長時間を要する場合や、解の導出自体が困難となる点が実用上の課題となる。一方、実用上は、厳密な最適解の導出が難しい場合であっても、準最適解を高速に導出できれば有用である場面が多い。

そこで、ユーザからの SCR 毎にサービスパスを確立するオンライン型 SFC において、ILP に基づく最適解からの性能低下を抑えつつ、計算時間を大幅に短縮するために、ラグランジュ緩和法*9、深さ優先ツアー探索(Depth First Tour Search: DFTS)*10、劣勾配法*11

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches

を組み合わせた解法を提案した[6]。

具体的には以下の(1)から(3)の手順をサービスパスの遅延に対する改善効果が見られなくなるまで繰り返す。(1) ラグランジュ緩和法により、CSPTP-based ILPにおける複雑な容量制約を目的関数のペナルティ項として置き換えることで、より簡単なSPTPに変換する。(2) 得られたSPTPをDFTSにより多項式時間で解く。(3) 容量制約を目的関数のペナルティ項として追加する際のラグランジュ定数を劣勾配法により調整する。

提案手法の有効性を Intel Core i9-9900K (8 コア/16 スレッド)、64 GB メモリを搭載した計算機を用いて数値計算とシミュレーションにより評価する。提案手法(DFTS-based Lagrangian heuristics)を C++言語と Boost グラフライブラリを用いて実装し、線形ソルバには CPLEX 12.8 を用いる。物理ノード数 200 の物理ネットワークを想定し、各物理ノードに 10 コアの CPU 資源を割り当てる。物理ノード間にはランダムに物理リンクを構築し、それぞれの容量を 10 Gbps、伝播遅延時間を平均 10ms とする。各ノードにおける転送時間を平均 10 ms、各機能の実行時間を平均 50 ms とする。想定する 4 種類のサービスの機能構成、需要、要求帯域を表 1 に示す。

また 5 種類の機能(NAT: Network Address Translator, FW: Firewall, TM: Traffic Monitor, WOC: WAN Optimization Controller, IDPS: Intrusion Detection Prevention System, and VOC: Video Optimization Controller)が要する CPU 資源量と需要を表 2 に示す。

表 1 サービスの機能構成、需要、要求帯域

Service	Sequence of functions	Demand	b_c
Web service	NAT-FW-TM-WOC-IDPS	18.2%	500 kbps
VoIP	NAT-FW-TM-FW-NAT	11.8%	320 kbps
Video streaming	NAT-FW-TM-VOC-IDPS	69.9%	20 Mbps
Online gaming	NAT-FW-VOC-WOC-IDPS	0.1%	20 Mbps

表 2 機能に対する要求 CPU 資源量と需要

Function type	$p_{c,j,c,k}^{func}$	Demand
NAT	0.0046	22.6%
FW	0.0045	22.6%
TM	0.0665	20.2%
IDPS	0.0535	16.6%
VOC	0.0270	14.2%
WOC	0.0270	3.7%

図 2 に SCR の到着率に対するサービスパス遅延の変化を示す。図より、手法によらず到着率の増加に応じてサービスパス遅延は増加の傾向を示している。これは、資源制約により新規のサービスパスが選択可能なノードや通信リンクに制限が生じているためである。ただし、提案手法は到着率によらず最適解(CSPTP-based ILP)と同程度のサービスパス遅延を実現できている。また 図 3 に、SCR の到着率に対する SCR あたりのサービスパスの計算に要する平均実行時間を示す。図より、提案手法は CSPTP-based ILP に比べて 20%程度の実行時間に抑えられている。以上より、提案手法は最適解と同等の解を迅速に導出できることが示された。

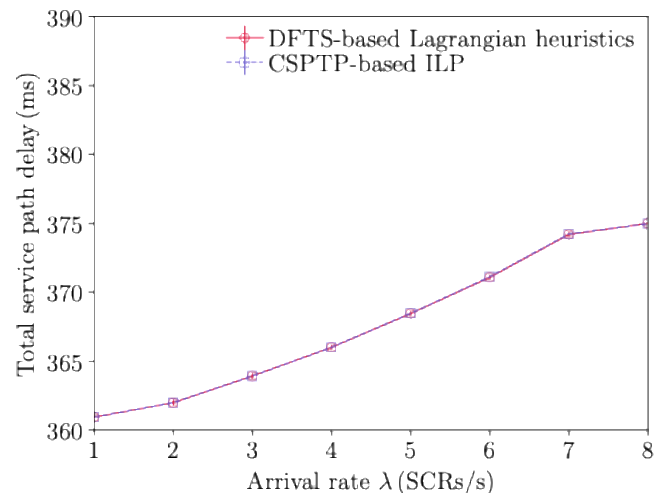


図 2 SCR の到着率とサービスパス遅延

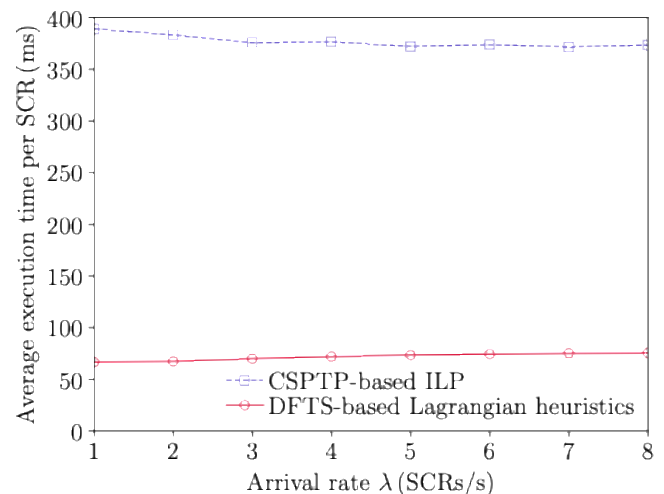


図 3 SCR の到着率と SCR あたりの計算時間

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches

3. 3 サービス需要とネットワークトポロジの変化に 適応可能なサービスチェイニング

実環境では、サービスに対するユーザからの需要が時々刻々と変化する、また、機器の障害等によりネットワークトポロジが変化する可能性がある。こうした環境変動を学習し、適切なサービスチェイニングを実現するために、GNN を利用した DRL フレームワークを提案する[7]。具体的には、サービスチェイニングを実行するオーケストレータをエージェントとみなし、できる限り多くの SCR を受理可能な方策を、GNN を利用した DRL で実現する。ネットワークの現在の利用状況や新たなサービスパス候補による利用状況の変化を GNN でモデル化し、DRL により適切なサービスパス候補を選択する。類似のアプローチとして、ルーティング問題への応用[4]が挙げられるが、サービスチェイニングでは経路全体で見た場合にループ構造が生じる可能性があり、より複雑な問題となる。

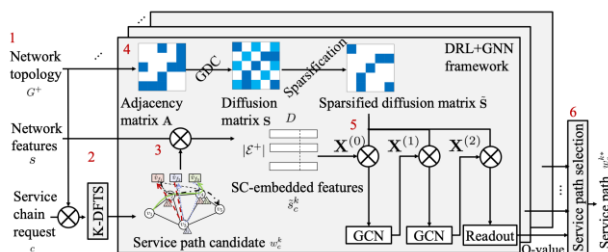


図 4 GNN+DRL に基づくサービスチェイニング

図 4 に提案手法の概要を示す。サービスチェイニングを実現する SDN コントローラ (エージェント) が、ネットワーク (環境) を監視し、ネットワークの状態と新規の SCR を環境から取得する (手順 1)。ここで、ネットワークの状態は前述の拡張ネットワークにおける各リンクの特徴として表現される。次に、エージェントは K-DFTS アルゴリズムにより K 本のサービスパスの候補を計算する (手順 2)。各サービスパス候補に対し、エージェントは現在のネットワークの状態を元に、各物理・仮想リンクの特徴を、媒介中心性、残余容量、サービスパス内での利用回数、1 回あたりの必要資源量、候補利用後の利用率からなる 5 次元ベクトルで表す (手順 3)。また、GNN の学習性能を高めるために、隣接行列を Graph Diffusion Convolution

(GDC)*¹²により拡散行列に変換し、さらにスパース化する (手順 4)。エージェントはこれらの情報を元にそのサービスパス候補の価値を示す Q 値を計算する (手順 5)。その後、候補間で Q 値に基づき確率的にサービスパスを選択する (手順 6)。

提案手法の有効性を Intel Xeon Gold 6226R (16 コア)、196 GB メモリ、NVIDIA GeForce RTX 3090 GPU を搭載した計算機を用いて数値計算とシミュレーションにより評価する。評価対象のネットワークには、米国の研究機関を結ぶバックボーンネットワークである、ノード数 14、リンク数 21 の NSFNET を用いる。表 1 の 4 種類のサービスに対する需要分布を Base とし、需要分布を変化させた 3 種類のシナリオを表 3 に示す。なお θ はコサイン類似度であり、1 に近いほど Base の需要分布に近いことを意味する。

表 3 サービス需要の変化シナリオ

Scenario	Web service	VoIP	Video streaming	Online gaming
Base	18.2%	11.8%	69.9%	0.1%
$\theta = 0.96$	24.2%	17.8%	51.9%	6.1%
$\theta = 0.82$	30.2%	23.8%	33.9%	12.1%
$\theta = 0.54$	36.2%	29.8%	15.9%	18.1%

まず、サービス需要の変化への適応力を評価するために、図 5 に各シナリオにおける、提案手法(DRL+GNN scheme)で受理された SCR の数 (SCR 受理数) を示す。なお、比較対象として前述の CSPTP-based ILP とラグランジュ緩和法の結果を合わせて示す。ただし、提案手法とラグランジュ緩和法では、Base シナリオで学習・調整したモデルとパラメータを用いる。Base シナリオでは、事前に想定されたサービス需要に対し、サービスパスを確立するため、方式間で性能差は殆ど見られない。一方、Base シナリオからの需要の変化が大きくなる (θ が小さくなる) と、ラグランジュ緩和法では他の手法に比べて性能が低下する傾向が見られるが、提案手法は CSPTP-based ILP と同等の性能を示している。なお、 θ の減少に応じて方式を問わず、SCR 受理数が増加しているのは、広帯域を要求する動画ストリーミングの需要が減るためである。

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches

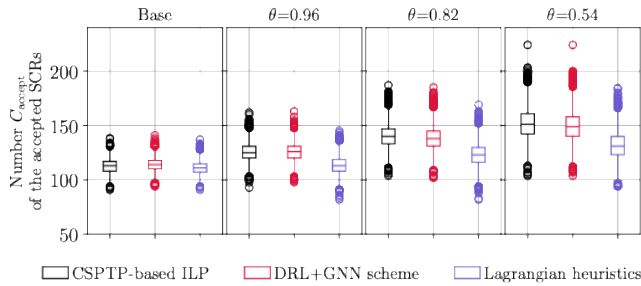


図5 サービス需要のシナリオとSCRの受理数

次に、障害により一部の物理リンクが利用できなくなった場合に、各方式が実現可能なSCRの受理数を評価する。先程のBaseシナリオにおける、リンクの障害数とSCRの受理数の関係を図6示す。なお、提案手法とラグランジュ緩和法では、障害発生前のBaseシナリオ上で学習・調整したモデルとパラメタを用いる。図より、障害リンク数の増加に対して、提案手法はラグランジュ緩和法よりも高い性能を示しており、最適解と同程度のSCR受理数を実現している。

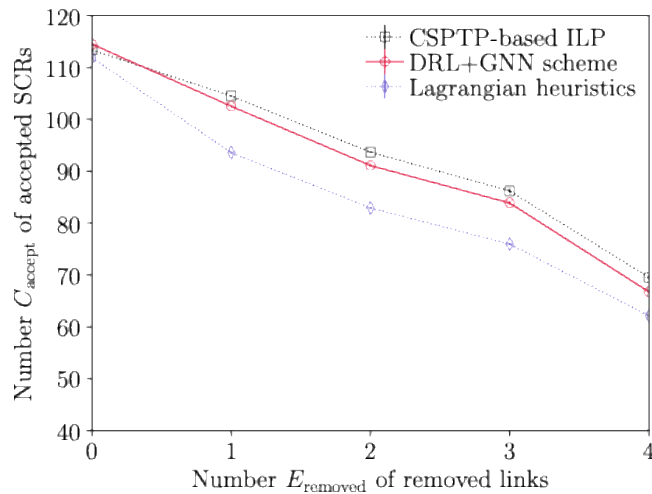


図6 障害リンク数とSCRの受理数

おわりに

本稿では、多様化するサービス需要やネットワーク環境の変化に対して、数理的・実践的アプローチに基づく持続可能なサービスチェイニングの実現方法を提案した。サービスチェイニングは数理的にはNP困難となるCSPTP-based ILPとしてモデル化される。まず、この問題を効率的に解くためにラグランジュ緩和法に基づく手法を提案した。また、サービス需要やネット

ワーク環境の変動への適応力を実現するために、DRLとGNNの組み合わせた手法を提案した。

数値評価を通し、サービスパスの最適性、環境変化への適応力、計算量などの観点から、提案手法の有効性・実用性を示した。今後の展望として、サービスチェイニングと機械学習・生成AIの融合が挙げられる。近年、機械学習・生成AIの社会への浸透はますます加速している一方、精度向上を目指したモデルの肥大化による実行環境の制約や入力データに対するデータプライバシーが課題となっている。モデルを分割し、各部分モデルをサービスチェイニングにおける機能とみなすことで、資源制約を緩和できるとともに中継ノードや通信リンクなどネットワーク環境に対して柔軟に適応可能なサービスの実現が期待できる。

用語解説

- *1 ネットワーク機能仮想化 (Network Functions Virtualization: NFV)
ハードウェアで実現されていたネットワーク機能を汎用サーバ上の仮想マシンやコンテナとしてソフトウェア化するアーキテクチャ。
- *2 サービスチェイニング (Service Function Chaining: SFC)
サービスを一連のネットワーク機能の列として表現し、経路を流れるトラフィックに対し、中継ノード上で所望の順にそれらの機能を適用する技術。
- *3 仮想ネットワーク機能 (Virtualized Network Function: VNF)
仮想マシン上でソフトウェアとして実行可能なネットワーク機能 (ファイアウォール、ロードバランサなど)。
- *4 クラウドネイティブ・ネットワーク機能 (Cloud-native Network Function: CNF)
コンテナ技術を前提とし、マイクロサービス構造や継続的インテグレーション/デリバリー (CI/CD) の利点を最大限に活かせるように設計されたネットワーク機能。
- *5 整数線形計画問題 (Integer Linear Programming: ILP)

数理的・実践的アプローチによる持続可能なサービスチェイニング

Sustainable Service Function Chaining through Mathematical and Practical Approaches

目的関数が線形、制約条件が線形の等式または不等式となる最適化問題は線形計画問題 (Linear Programming: LP) と呼ばれ、その中で変数が実数もしくは整数で表される問題。

*6 グラフニューラルネットワーク (Graph Neural Network: GNN)

グラフ構造を持つデータに対して、ノードやエッジの関係性を学習するディープラーニングモデル。

*7 深層強化学習 (Deep Reinforcement Learning: DRL)

強化学習の行動価値関数や方策関数の表現にニューラルネットワークを用いた手法。

*8 ソルバ (Solver)

LP や ILP などの数理モデルに対し、単体法や内点法、分枝限定法などの汎用的なアルゴリズムを用いて最適解や近似解を求めるプログラム。

*9 ラグランジュ緩和法 (Lagrangian Relaxation)

複雑な制約条件を持つ最適化問題において、それらを目的関数に組み込むことで問題を簡略化し、元の問題の最適値に対する下界 (最小化問題の場合) を求める手法。

*10 深さ優先ツアー探索 (Depth First Tour Search: DFTS)

グラフ上の経路探索手法の一つ。あるノードから深さを優先して探索し、決められたノード群を順序どおりに通過するツアーと呼ばれる経路を探索する手法。

*11 劣勾配法 (Subgradient Method)

微分不可能な凸関数を最小化するための反復アルゴリズム。

*12 Graph Diffusion Convolution (GDC)

グラフ上の拡散プロセスを利用してグラフ構造を平滑化し、GNN の学習性能を高める手法。

参考文献

[1] M. Sasabe and T. Hara, "Capacitated Shortest Path Tour Problem Based Integer Linear Programming for Service Chaining and Function Placement in NFV Networks," *IEEE Transactions on Network and Service*

Management, vol. 18, no. 1, pp. 104-117, Mar. 2021.

[2] A. Rafiq, T. A. Khan, M. Afaq, and W.-C. Song, "Service Function Chaining and Traffic Steering in SDN using Graph Neural Network," in *Proc. of International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2020, pp. 500-505.

[3] Z. Ning, N. Wang, and R. Tafazolli, "Deep Reinforcement Learning for NFV-based Service Function Chaining in Multi-Service Networks," in *Proc. of IEEE 21st International Conference on High Performance Switching and Routing (HPSR)*, May 2020, pp. 1-6.

[4] P. Almasan, J. Suárez-Varela, A. Badia-Sampera, K. Rusek, P. Barlet-Ros, and A. Cabellos-Aparicio, "Deep Reinforcement Learning meets Graph Neural Networks: exploring a routing optimization use case," *arXiv:1910.07421 [cs]*, Feb. 2020.

[5] A. Bittar, Z. Wang, and C. Huang, "Service Function Chaining Implementation using VNFs and CNFs," in *Proc. of IEEE International Conference on Cloud Engineering (IC2E)*, Sep. 2023, pp. 193-201.

[6] T. Hara and M. Sasabe, "Speedy and Efficient Service Chaining and Function Placement Based on Lagrangian Heuristics for Capacitated Shortest Path Tour Problem," *Journal of Network and Systems Management*, vol. 31, no. 24, pp. 1-34, Dec. 2022.

[7] T. Hara and M. Sasabe, "Capacitated Shortest Path Tour Based Service Chaining Adaptive to Changes of Service Demand and Network Topology," *IEEE Transactions on Network and Service Management*, vol. 21, no. 2, pp. 1344-1358, Jan. 2024.

この研究は、令和3年度SCAT研究助成の対象として採用され、令和4～6年度に実施されたものです。