MECを含む分散クラウドにおける負荷分散を考慮したサービス配置計算手法の基礎評価

城 哲*1,2, 蕨野 貴之*1,2, 鈴木 悠祐*1,2, 大谷 朋広*1,2
*1：KDDI Corporation
*2：KDDI Research, Inc

謝辞：本研究開発は、総務省の「革新的AIネットワーク統合基盤技術の研究開発（JPMI00316）」によって実施した成果を含みます。

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

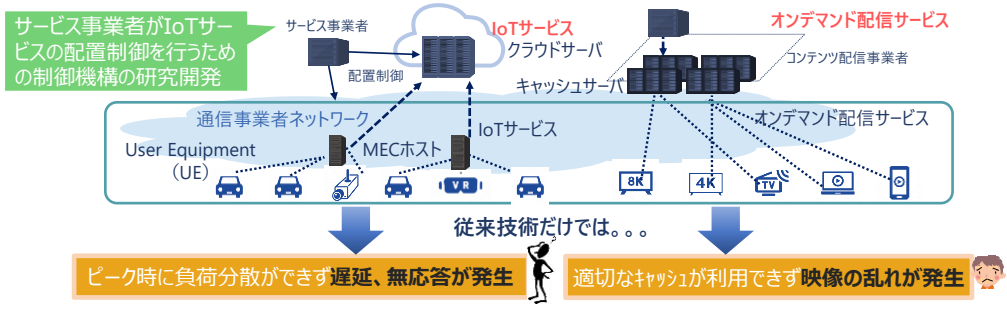
はじめに

総務省様からの受諾案件「革新的AIネットワーク統合基盤技術の 研究開発（JPMI00316）」にて，
サービス事業者が多種多様なサービスを多くのIoT機器に安定的に提供可能な処理配置手法を検討している

背景 5G（第5世代移動通信システム）の導入，IoT機器普及，4K8Kコンテンツの増大等に伴い，通信量の急激な増加や新サービスの創出が見込まれている

課題 設備投資によるネットワーク設備やノードを拡張するだけでは、サービスの品質/信頼性を保持できない

課題 今後のトラヒックの増加の主な要因と想定される、①オンデマンド配信サービス、②IoTサービスに着目
① オンデマンド配信サービス： 4K8Kといった大容量コンテンツの視聴に対しても、映像の乱れを回避
② **IoTサービス：最大需要の瞬間に合わせた負荷分散を行うことで大量のIoTサービスを提供**



Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

アジェンダ

1. MECの概要
2. 標準化動向
3. 先行研究
4. 基礎評価
5. まとめ

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

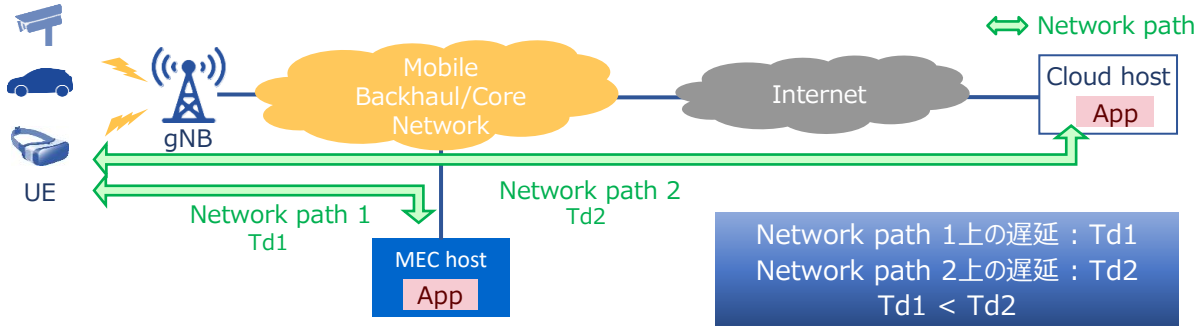
1. MECの概要

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

MECとは？

KDDI Research

- Multi-access Edge Computing
- ネットワークのエッジにサーバ（MECホスト）が分散配置される
- MECホストへのApplication（App）の配置により，ネットワーク区間における遅延が短縮され，QoSが向上する



Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

MECは5G時代に注目されているキーテクノロジー

KDDI Research

- 要求遅延が異なる多種多様なサービスが登場
- E2Eでこれらのサービスの要求を満たせる仕組みがネットワークシステムに必要

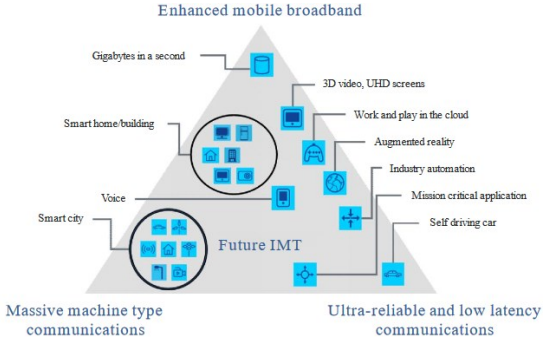


図1-1.IMT 2020における利用シナリオ

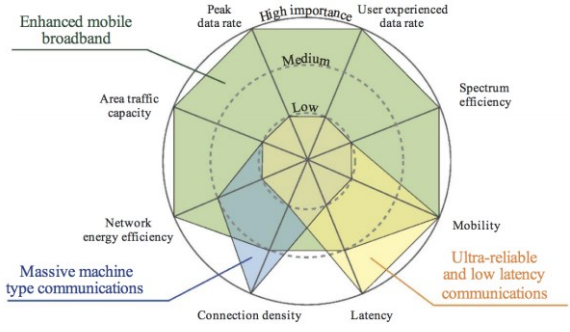


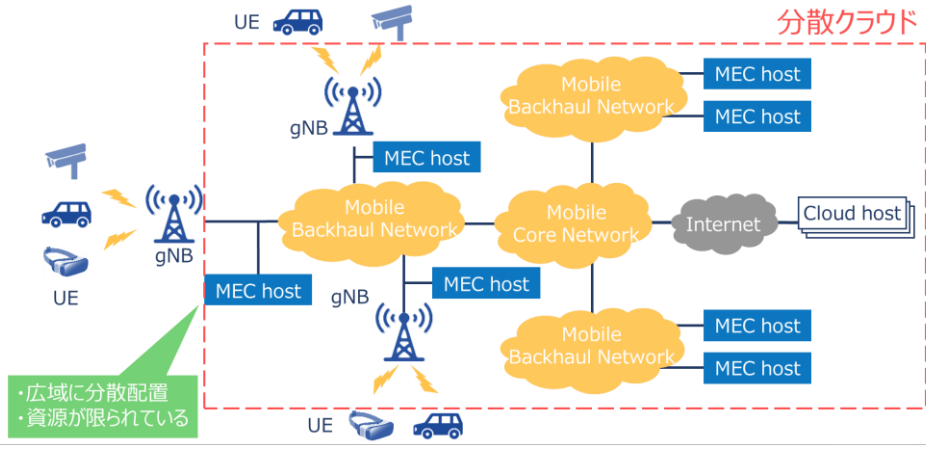
図1-2.利用シナリオにおける指標

https://www.itu.int/rec/R-REC-M.2083

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

複数のホストがAppの配置先の候補

- 実用形態は他のMECホストや中央クラウドを組み合わせた分散クラウド
- Appの配置先ホストの候補は分散クラウド内に複数存在



Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

多量のIoT機器へのサービス提供には負荷分散手法が重要

- E2Eにおける2つの主要な遅延
  - データ配送遅延
    - UEとApp間のデータ転送にかかる遅延
    - UEの際近傍に位置するMECホスト間との値が一番小さい
  - 計算遅延
    - ホスト上におけるAppの処理にかかる遅延
    - ホストの計算力や負荷状況に影響される
- MECホストは資源が限られる
  - 混雑時等の負荷集中による計算遅延の増加は無視できない要素

➡ 分散クラウド内の効率的な負荷分散がサービス品質の向上や収容可能なUE数の向上において重要となる

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

2. 標準化動向

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

欧州電気通信標準化機構（ETSI）がMECの国際標準を牽引

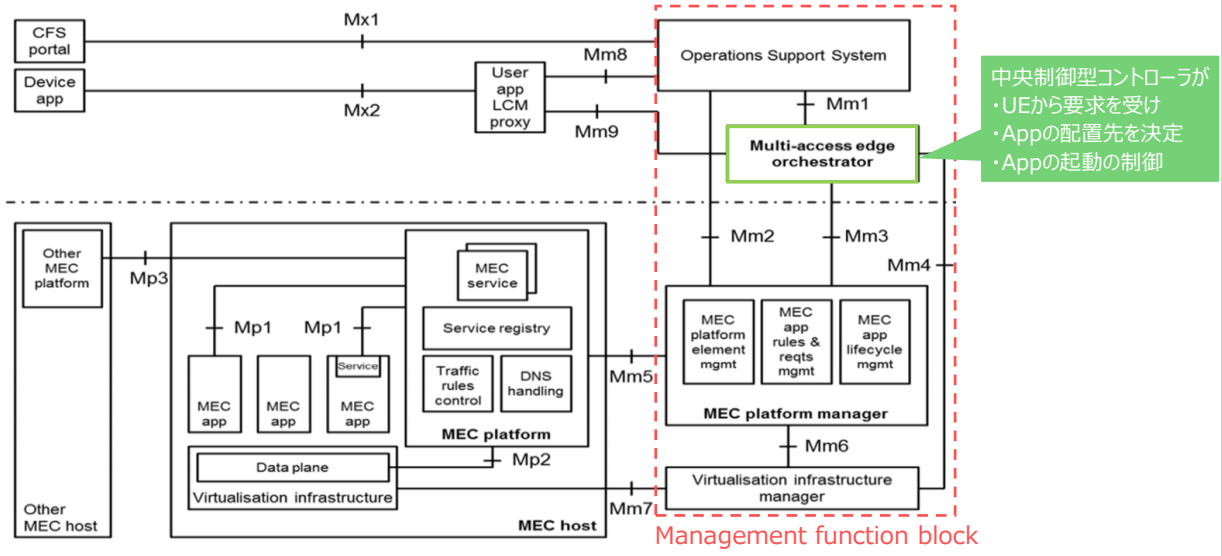
➢2014年から多くのMECに関する国際標準を公開

ETSI deliverable	title
ETSI GS MEC 001 V2.1.1 (2019-01)	Multi-access Edge Computing (MEC); Terminology
ETSI GS MEC 002 V2.1.1 (2018-10)	Multi-access Edge Computing (MEC); Phase 2: Use Cases and Requirements
ETSI GS MEC 003 V2.1.1 (2019-01)	Multi-access Edge Computing (MEC); Framework and Reference Architecture
ETSI GS MEC 005 V2.1.1 (2019-07)	Multi-access Edge Computing (MEC); Proof of Concept Framework
ETSI GS MEC 009 V1.1.1 (2017-07)	Mobile Edge Computing (MEC); General principles for Mobile Edge Service APIs
ETSI GS MEC 009 V2.1.1 (2019-01)	Multi-access Edge Computing (MEC); General principles for MEC Service APIs
ETSI GS MEC 010-1 V1.1.1 (2017-10)	Mobile Edge Computing (MEC); Mobile Edge Management; Part 1: System, host and platform management
ETSI GS MEC 010-2 V2.1.1 (2019-11)	Multi-access Edge Computing (MEC); MEC Management; Part 2: Application lifecycle, rules and requirements management
ETSI GS MEC 011 V2.1.1 (2019-11)	Multi-access Edge Computing (MEC); Edge Platform Application Enablement
ETSI GS MEC 012 V2.1.1 (2019-12)	Multi-access Edge Computing (MEC); Radio Network Information API
ETSI GS MEC 013 V2.1.1 (2019-09)	Multi-access Edge Computing (MEC); Location API
ETSI GS MEC 014 V1.1.1 (2018-02)	Mobile Edge Computing (MEC); UE Identity API
ETSI GS MEC 015 V1.1.1 (2017-10)	Mobile Edge Computing (MEC); Bandwidth Management API
ETSI GS MEC 015 V2.1.1 (2020-06)	Multi-Access Edge Computing (MEC); Traffic Management APIs
ETSI GS MEC 016 V2.2.1 (2020-04)	Multi-access Edge Computing (MEC); Device application interface
ETSI GS MEC 021 V2.1.1 (2020-01)	Multi-access Edge Computing (MEC); Application Mobility Service API
ETSI GS MEC 026 V2.1.1 (2019-01)	Multi-access Edge Computing (MEC); Support for regulatory requirements
ETSI GS MEC 028 V2.1.1 (2020-06)	Multi-access Edge Computing (MEC); WLAN Information API
ETSI GS MEC 029 V2.1.1 (2019-07)	Multi-access Edge Computing (MEC); Fixed Access Information API
ETSI GS MEC 030 V2.1.1 (2020-04)	Multi-access Edge Computing (MEC); V2X Information Service API
ETSI GS MEC-IEG 004 V1.1.1 (2015-11)	Mobile-Edge Computing (MEC); Service Scenarios
ETSI GS MEC-IEG 005 V1.1.1 (2015-08)	Mobile-Edge Computing (MEC); Proof of Concept Framework
ETSI GS MEC-IEG 006 V1.1.1 (2017-01)	Mobile Edge Computing; Market Acceleration; MEC Metrics Best Practice and Guidelines

https://www.etsi.org/

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

ETSI MEC標準の参照アーキテクチャ



Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

3. 先行研究

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

E2E遅延の推定手法の基本的な考えは[1]を参考とした

- [1] Michal Vondra, et al, “QoS-ensuring Distribution of Computation Load among Cloud-enabled Small Cells,” CloudNet, Oct. 2014.
- [1]の提案では，推定のデータ転送遅延と計算遅延の和が最小となるMECホストがAppの配置先ホストとして選定される
- 推定データ転送遅延：Td
  - Td ≈ 転送データ量 R / ネットワークパス上のスループットの最小値 τ
- 推定計算遅延：Tc
  - Tc ≈ 要求計算量 M / ホスト上の利用可能な計算量 a
- E2E遅延 =Td +Tc

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

4. 基礎評価

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

基礎評価の概要

KDDI Research

- 目的
  - 分散クラウドへの負荷分散手法適応時のApp配置の特性に対する理解を深める
- 評価項目
  - 評価1：複数のUEに対するApp配置の同時計算と逐次計算に関する比較評価
  - 評価2：目的関数と配置傾向の関係性
- 評価指標
  - 評価1：各ホストの負荷，サービスごとの遅延
  - 評価2：各ホストの負荷，サービスごとの遅延，収容可能UE数
- 評価方法
  - 数値シミュレーション

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

システムアーキテクチャとシナリオ

KDDI Research

- 中央型の配置制御機構が
  - 通信特性の異なる3種類のサービスの起動要求をUEから受け
  - 3つの無線エリアと中央クラウドで構成された分散クラウドにAppを配置する

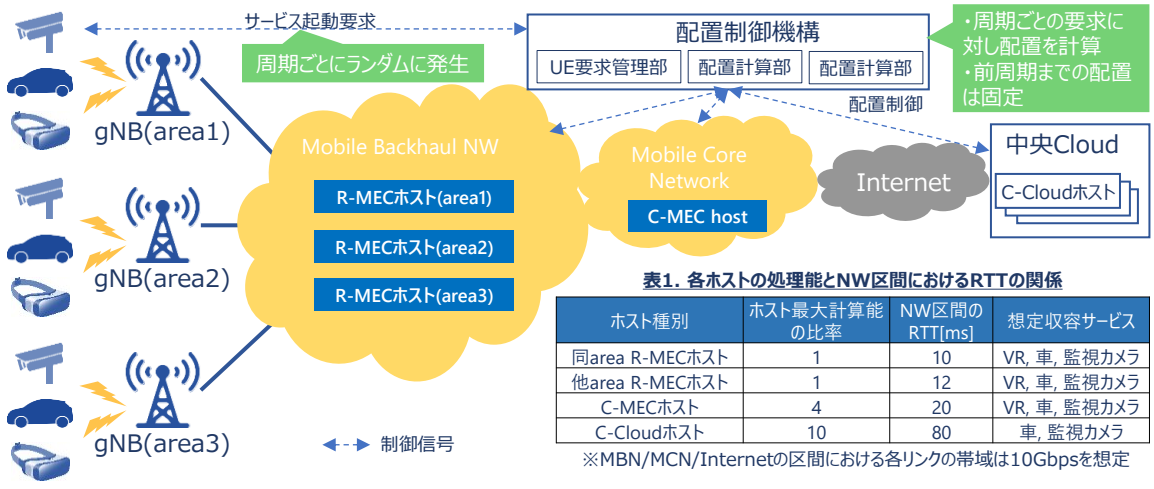


表1. 各ホストの処理能とNW区間におけるRTTの関係

ホスト種別	ホスト最大計算能の比率	NW区間のRTT[ms]	想定収容サービス
同area R-MECホスト	1	10	VR, 車, 監視カメラ
他area R-MECホスト	1	12	VR, 車, 監視カメラ
C-MECホスト	4	20	VR, 車, 監視カメラ
C-Cloudホスト	10	80	車, 監視カメラ

※MBN/MCN/Internetの区間における各リンクの帯域は10Gbpsを想定

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

想定サービス概要

KDDI Research

➢要求の遅延/通信量/計算量が異なる3種類のサービスを想定

サービス種別	サービス概要	通信モデル	要求遅延	秒間要求通信量	秒間要求計算量	処理周期
VR	Appから各UEにVR映像を提供する 1. UEがAppにセンサ情報を送付 2. Appがセンサ情報をもとにVR映像を生成 3. Appが作成したVR映像をUEに送付	①UE -> App ②UE <- App	70ms	①1Mbps ②40Mbps	3.7Gcycle/s	30ms
車	Appが各UEに対し交差点の優先制御を行う 1. UEがAppにセンサ情報を送付 2. Appが各UEからの情報をもとに優先制御に関する計算を実施 3. AppがUEに制御の計算結果を通知	①UE -> App ②UE <- App	100ms	①0.03Mbps ②0.3Mbps	0.11Gcycle/s	100ms
監視カメラ	AppがUEから配信される映像の異常を判定 1. UEがカメラ映像をAppに送付 2. AppがUEからのカメラ映像を分析 3. Appが分析結果を中央クラウド上の監視センタに通知	①UE -> App ②App -> 監視センタ	2s	①40Mbps ②1Mbps	4.2Gcycle/s	30ms

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

モデル

KDDI Research

- ➢推定遅延率 [%]
  - ➢推定遅延率 = E2E推定遅延 t / 要求遅延 T-req x 100
  - ➢E2E推定遅延 t = 推定NW遅延 t-nw + 推定計算遅延 t-c
    - ➢t-nw = NWパスの固定遅延t-nw_f + NWパスにおけるデータ転送遅延 t-nw_d
      - ➢t-nw_f = UEの在圏エリアとホストの組み合わせで一意に定まる固定値
      - ➢t-nw_d = Σ (リンクにおける各Appの処理周期単位の要求通信量) [bit] / リンクの回線容量 [bps]
    - ➢t-c = ホストにおけるAppの処理周期単位の要求計算量 [cycle] / ホストの処理能 p [cycle/s]
      - ➢p = ホストの最大処理力[cycle/s] / ホスト上のApp数
- ➢推定負荷率 [%]
  - ➢ホストの推定計算負荷率 l-h
    - ➢l-h = Σ (ホストにおけるAppの秒間要求計算量) [cycle/s] / ホストの最大処理能 [cycle/s] x 100
  - ➢リンクの推定回線負荷率 l-l
    - ➢l-l = Σ (リンクにおける各Appの秒間要求通信量) [bps] / リンクの回線容量 x 100

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

目的関数・制約条件・その他

KDDI Research

- 目的関数
  - 目的1：min (推定遅延率の最大値)
  - 目的2：min (推定負荷率の最大値)
    - 推定負荷率の最大値：ホストの推定負荷率の最大値 ＋ リンクの推定負荷率の最大値
  - 目的3：min (推定遅延率の最大値 ＋ 推定負荷率の最大値)
- 制約条件
  - 制約1：推定遅延率 < 100
  - 制約2：ホストの推定負荷率 l-h < 100
  - 制約3：リンクの推定負荷率 l-l < 100
  - 制約4：各Appの配置数 = 1
  - 制約5：過去周期において，ホストに配置されたAppの配置箇所は変更しない
- その他
  - 問題：Mixed Integer Nonlinear Programming（MINLP）
  - ソルバ：Solving Constraint Integer Programs（SCIP）

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

評価1：同時計算 vs 逐次計算

KDDI Research

- 評価条件（下表）

総UE数	同時計算UE数 (step数)	1周期当たりの サービス起動要求数	目的関数	試行数	App比率 (VR:車:監視カメラ)
180	1, 10, 20台の3通り	総UE数 / step数の3通り	目的2:min (推定負荷率の最大値)	10	1:1:1

- 評価結果(負荷率観点)

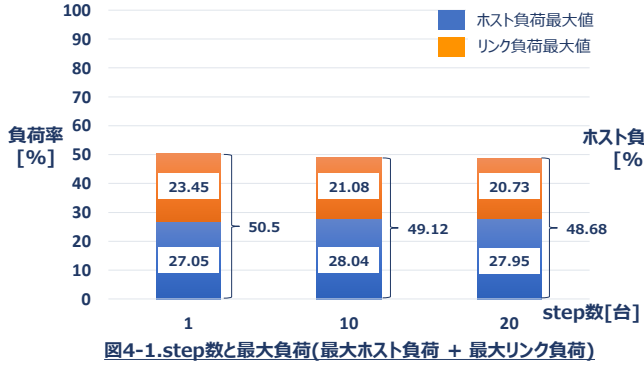


図4-1.step数と最大負荷(最大ホスト負荷 ＋ 最大リンク負荷)

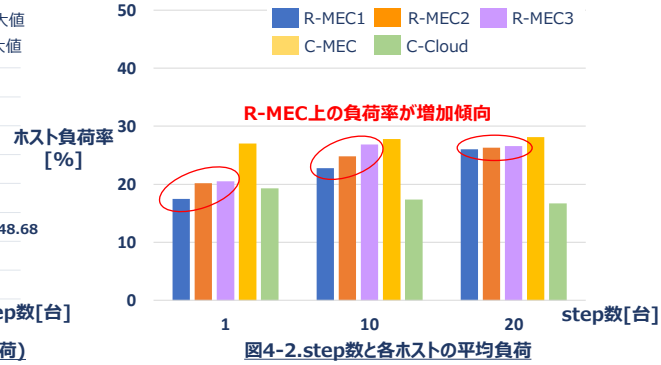


図4-2.step数と各ホストの平均負荷

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

評価1：同時計算 vs 逐次計算

KDDI Research

➢評価結果（サービス観点）

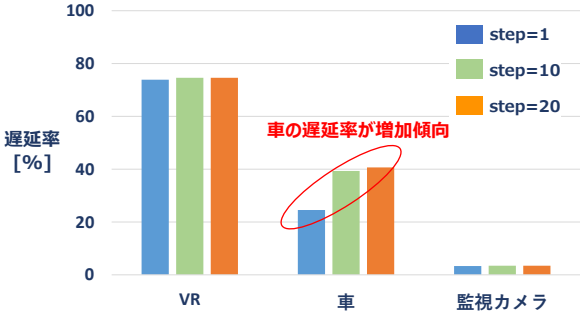


図4-3.step数と各サービスの平均遅延

表2. 評価1におけるstep数とホスト種別ごとのApp配置比率

Step数	R-MECホスト [%]	C-MECホスト [%]	Cloudホスト [%]
1	40.00	22.44	37.56
10	36.67	24.00	39.33
20	35.39	24.50	40.11

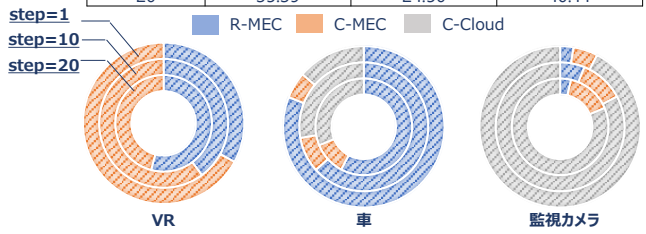


図4-4.step数と各サービスの配置傾向

➢結果分析

- ➢同時計算による負荷削減効果は逐次計算と比較して限定的
- ➢同時計算時は，車サービス(処理/通信量：少)がCloudホストに，監視カメラサービス(処理/通信量：多)がMECホストに配置される頻度が増加
- ➢結果，資源が限られ，負荷が上昇しやすいR-MECホストの負荷が増加。また，要求遅延が監視カメラより短い車サービスの応答遅延が増加

結論：同時計算による負荷の削減効果が限定的であり，かつ，特定サービスの遅延が上昇傾向となることから，サービス品質の観点から逐次計算が効果的

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

評価2：目的関数と配置傾向の関係性

KDDI Research

➢評価2-1条件（下表）

総UE数	同時計算UE数(step数)	目的関数	試行数	App比率
180	1	目的1：min (推定遅延率の最大値) 目的2：min (推定負荷率の最大値) 目的3：min (推定遅延率の最大値 +推定負荷率の最大値)	10	1:1:1

➢評価結果

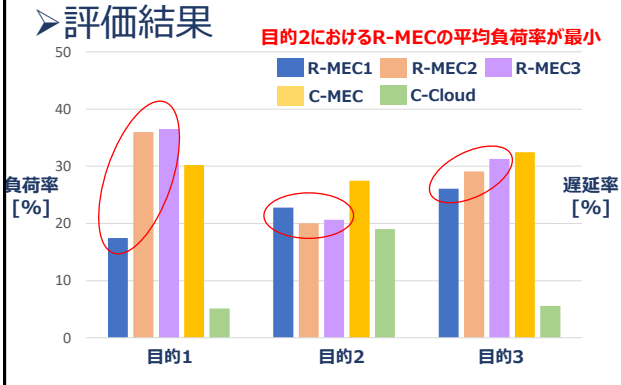


図4-5.評価2-1における目的関数と各ホストの平均負荷

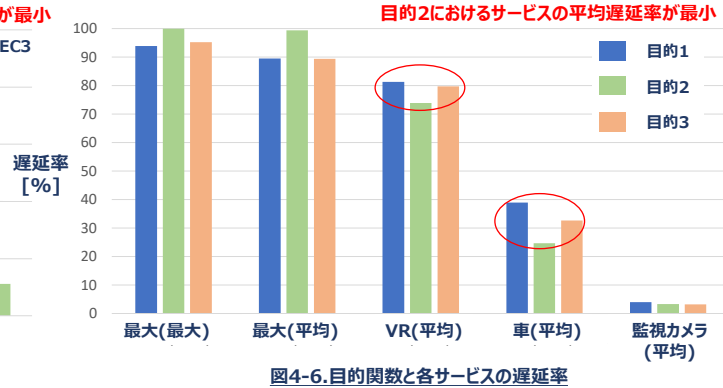


図4-6.目的関数と各サービスの遅延率

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

評価2：目的関数と配置傾向の関係性

KDDI Research

➢評価2-2条件（下表）

総UE数	同時計算数(step数)	目的関数	試行数	App比率
配置不可発生まで	1	目的2：min (推定負荷率の最大値) 目的3：min (推定遅延率の最大値 +推定負荷率の最大値)	10	1. 3:1:1（VR3倍） 2. 1:3:1（車3倍） 3. 1:1:3（監視カメラ3倍）

➢評価結果

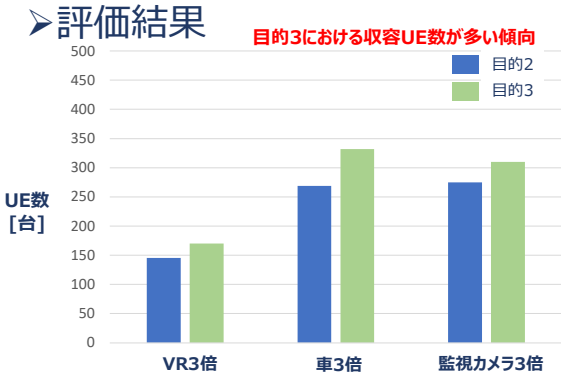


図4-7.目的関数と各サービスの収容数

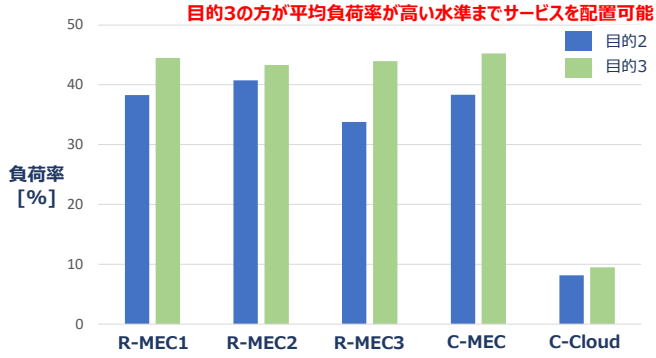


図4-8.評価2-2における目的関数と各サーバの平均負荷

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

評価2：目的関数と配置傾向の関係性

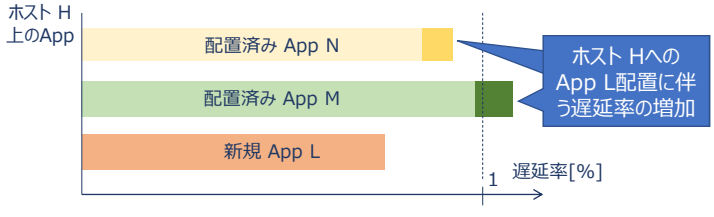
KDDI Research

➢結果分析

- ➢目的2（最大負荷の最小化）は各サーバの負荷の平均値，各サービスの遅延の平均値の観点で優れている
- ➢一方，目的3（最大負荷 ＋ 最大遅延の最小化）は各ホストの資源をより多く利用できるため、目的2よりもAppの配置不可が発生しにくい

➢目的2の方がAppの配置不可が発生しやすい理由

- ➢ホスト上への新規App配置に伴い，ホスト上の配置済Appの遅延率が増加
- ➢配置済Appの遅延率が大きい(≒1)場合、遅延率の制約条件により、ホストへの新規App配置が不可となる
- ➢サービスの最大遅延率が最も大きい傾向にある目的2では、ホストがこのような状態に遷移しやすい



結論：
多量のIoT機器へのサービス提供の観点においては、目的3（最大負荷 ＋ 最大遅延の最小化）が最も効果的

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

5. まとめ

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

まとめ

- MECは
  - 5G時代の多種多様なユースシナリオの提供における重要な技術
  - ETSIによって国際標準化されている
  - 各ホストの資源量が限られており，中央クラウドと組み合わさった分散クラウドでの実用化が考えられる
- 分散クラウド内での効率的な負荷分散がサービス品質や収容UE数の向上において重要
- 基礎評価から以下の知見を得た
  - 計算の所用時間に加え，配置特性からも，UEの配置計算を逐次的に行う手法が優位な場合がある
  - 負荷分散を通じた多量のIoT機器へのサービス提供には、最大遅延と最大負荷の考慮が可能な目的関数の設定が効果的
- 今後
  - サービス種別追加時の配置特性の評価
  - NWトポロジと目的関数の組合せによる配置特性の評価
  - 上記評価結果に沿った，目的関数の組合せや重みの設計（手法）の検討

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.

KDDI Research

ご清聴ありがとうございました

Copyright(C) 2020 KDDI Research, Inc. All Rights Reserved.