

ロボットの挙動制御によるネットワーク性能向上手法の検討

新宮 裕章¹ 本生 崇人¹ 藤橋 卓也¹ 工藤 理一² 高橋 馨子² 村上 友規³ 渡辺 尚¹ 猿渡 俊介¹

概要: 携帯電話網や無線 LAN の進化によっていつでもどこでも無線ネットワークに接続できる環境が整備されたことで, Automatic Guided Vehicle (AGV) による部品や荷物の自動搬送, 無人走行車やドローンの遠隔操作などネットワークを介してロボットを制御する応用が生まれて来ている. しかしながら, 多数のロボットがネットワーク資源を共有するため, お互いのロボットの動作に影響を受けずに無線ネットワークとの接続を安定的に維持することが課題になると考えられる. このような観点から, 本稿では, ロボットの挙動とネットワークの制御の同時最適化を実現する強化学習のフレームワーク「CoRein」を提案する. CoRein は, ロボットの操作とネットワークの設定を行動, ロボットで取得できる情報を状態, ロボットに与えられたタスクとネットワーク性能を報酬関数として, 行動価値関数を Deep Neural Network (DNN) で強化学習する. 計算機シミュレーションを用いて 2 台のロボットが行ったり来たりする環境で評価したところ, 2 台のロボットが移動の速さを変えながら無線ネットワークの性能が高くなるように接続先アクセスポイントとロボットの位置を調整するような動作が確認できた.

1. はじめに

5G をはじめとした遅延を意識したワイヤレスネットワークの登場に相まって, ネットワークを介してロボットを制御することで新たなサービスが生まれつつある. 既に倉庫や工場では, 荷物や部品の運搬に AGV などのロボットがネットワークを介して制御されている. 将来的には, トヨタの e-Pallet などの自動運転車もネットワークを介して制御されるようになると思われる. さらに, ドローンによる運送が実現されれば, 多数のドローンが配達効率が高くなるようにネットワークから制御されるようになることも予想される.

ネットワークを介してロボットを制御することを考えた場合, 様々な要因によってネットワーク性能が低下してロボットの動作に影響が生じることが予想される. ロボットと基地局との間の通信距離, マルチパスによるフェージング, ロボットの移動によるシャドウィング, 他のロボットの通信との通信衝突など, これまでのユーザに向けたネットワークサービスと同様の問題が数多く発生する. 最悪の場合では, ネットワークから切断されたロボットは動作を停止して, ロボット自身が提供するサービスを継続できなくなる可能性がある.

ネットワークを介したロボットの制御の実現に向けて,

本稿ではネットワーク性能の最適化・ロボット挙動の最適化を同時達成できるかどうかを考える. これまで, ネットワーク側の最適化に関しては, 基地局配置設計手法 [1-8], 基地局選択手法 [9-16] や, リソース制御手法 [17-23] などのネットワーク側の最適化を行う研究がなされてきた. 基地局の選択手法に関しては, RSSI に基づいて端末側で基地局を選択する手法も存在するが, 本研究では端末のネットワークインタフェースもネットワークの一部と捉えて基地局選択もネットワーク側であると位置付けている. 関連研究の詳細に関しては 2 節で述べる. 本稿ではこれまでの既存研究の試みからさらに一歩進めて, ロボットの挙動自体を制御してネットワーク性能とロボットに与えられたタスクの遂行効率を同時最適化することを目指す.

このような観点から, 本稿では, 深層強化学習を用いたネットワーク・ロボット挙動同時最適化手法「CoRein」を提案する. CoRein では, ネットワーク性能やロボットのタスク実行性能に応じてネットワークの設定やロボットの挙動を強化学習によって最適化する仕組みを提供する. 具体的には, ロボットの操作とネットワークの設定を行動, ロボットで取得したセンサ情報を状態, ロボットに与えられたタスクやネットワーク性能を報酬関数として, 行動価値関数を ϵ -greedy 法を用いて DNN で近似しながら最適化する強化学習を実現した. DNN は, 全結合層, ReLU, 欠落率 40% のドロップアウト層の組を 2 回繰り返したものをを用いた. 計算機シミュレーションによって提案手法「CoRein」を評価した結果, RSSI に基づいてアクセスポイ

¹ 大阪大学大学院情報科学研究科

² 日本電信電話株式会社 NTT 未来ねっと研究所

³ 日本電信電話株式会社 NTT アクセスサービスシステム研究所

ントを選択する naïve な手法よりも高い性能を発揮することが確認できた。

本稿の構成は以下の通りである。2 節では、関連研究について述べる。3 節では、提案手法「CoRein」について述べる。4 節では、提案手法「CoRein」の有効性を評価するために行った計算機シミュレーションについて述べ、最後に 5 節でまとめとする。

2. 関連研究

本研究は、ネットワーク最適化に関連する。ネットワーク最適化の研究としては、主に基地局配置設計手法、基地局選択手法、リソース制御手法が挙げられる。基地局選択手法としては、RSSI に基づいて基地局を選択するなどの端末側が主導となって基地局を選択する場合もあるが、本研究では、端末の基地局の選択自体もネットワークの最適化であると位置づける点に注意されたい。例えば、ユーザがスマートフォンの通信状況がよい場所を求めて移動するという行動は今までもあった。しかしながら、筆者らの知りうる限り、ネットワークの最適化のためにユーザやロボットの行動を意図的に制御する研究は存在しない。これまでの研究では、ネットワークの利用者は制御ができないユーザを想定していたからだと考えられる。

2.1 基地局配置設計手法

無線基地局の配置を設計する研究としては、平均スループットを向上させるために基地局を配置する研究 [1–4]、平均アクセス遅延を軽減するために基地局を配置する研究 [5]、カバレッジを増加させるために基地局を配置する研究 [6–8] などが挙げられる。文献 [1] では、深層強化学習を用いてデータ収集などのサービスを地上に提供する UAV を都市環境において最適に配置する手法を提案している。文献 [2] では、ユーザの位置が時間的に変化することを考慮して、協調的確率的近似に基づいて AP の位置とカバレッジ半径を共同で最適化する手法を提案している。文献 [5] では、モバイル IoT デバイスとそのトラフィック負荷に動的に適応できるランキングベースの近距離最適配置アルゴリズムを用いた RNOPA を提案している。文献 [7] では、屋内測位精度とサービス範囲の向上を目的とした、多目的進化計算 P-EStra の有効性が検討されている。

2.2 基地局選択手法

基地局の選択は、携帯電話網や無線 LAN においてセルを移動するときのハンドオーバーの際に行われる [24, 25]。ハンドオーバーでは、進行中のセッションの中断を最小限に抑えて通信を継続することが第一の目標となる。ハンドオーバー中の最大の遅延は、切り替え先の無線基地局を見つけるために必要なスキャン段階である [26]。スキャン

を容易にするための技術として、様々なデータリンク層情報を上位相に提供する IEEE 802.11k [9]、FT (Fast Basic Service Set Transition) によって無線基地局切り替え時に高速で認証する仕組みを提供する 802.11r [10] などの規格が提案されている [27]。現状ではほとんどのハンドオーバーは RSSI に依存している [28, 29]。

RSSI だけでなく、消費電力や遅延など様々な要素に基づいて基地局選択をする仕組みも検討されている [11–16]。例えば、ゲーム理論に基づいたネットワーク接続に関する研究として、ゲーム内のプレイヤーがユーザ対ユーザのもの [11] と、ユーザ対ネットワークのもの [12]、ネットワーク対ネットワークのもの [13] などが検討されている。また、文献 [15] では、動的マルチチャネルアクセス問題における未知のダイナミクスと法外な計算量という課題を克服するために、強化学習を利用したアプローチが提案されている。

2.3 リソース制御手法

接続端末に対する適切なリソース割当てに関する研究としては、適応的に周波数を割り当てる研究 [17]、適応的に電力を割り当てる研究 [18, 19]、アンテナを割り当てる研究 [20–23] などが挙げられる。例えば、文献 [17] では、十分に利用されていない周波数帯に動的にアクセスすることを可能にすることを目的として、統計的チャネル割り当てを用いた SCA-MAC プロトコルを提案している。文献 [30] では、セルラーネットワークにおけるエネルギー効率の向上を目的として、新しいマルチホップコグニティブセルラーネットワーク (MC2N) アーキテクチャを提案している。文献 [20] では、電力配分の設計とアップリンクビームフォーミング設計を分離することで、複雑な演算を必要としない深層学習に基づく高速ビームフォーミング設計手法を提案している。また、接続先やトラフィックの集中に起因する問題を解決するためにネットワーク側でのロードバランシングに関する研究 [31–36] も検討されている。

3. 提案手法: CoRein

3.1 全体像

2 節で述べたように、既存の研究はユーザの行動や環境は制御できないという条件の基でユーザや環境の状態に応じてネットワーク側の最適化を行ってきた。それに対して本稿では、強化学習によってネットワーク側とロボットの挙動を同時に最適化する「CoRein」を提案する。

図 1 に提案手法の全体像を示す。CoRein は、ロボット部、ネットワーク部、ロボット・ネットワーク制御部の 3 つの要素から構成される。ロボット部は、ロボット・ネットワーク制御部からコマンドを受け取って物理空間において任意のタスクを実行する。ネットワーク部は、ロボッ

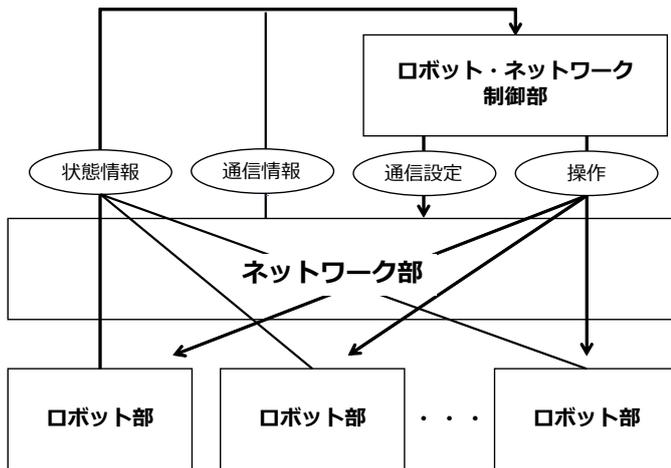


図 1: 提案手法「CoRein」の全体像

ト部とロボット・ネットワーク制御部を接続する。ロボット・ネットワーク制御部は、各ロボット部とネットワーク部から情報を収集して、ロボット部に対して取るべき挙動や、ネットワーク部に対して接続する無線アクセスポイントなどの通信設定の指令を出す。各ロボット部からの情報は、ロボットの位置・向きなどの状態を意味する。ネットワーク部からの情報は、RSSIなどのロボットに具備された無線 LAN インターフェースから取得できるネットワークの情報や、無線アクセスポイントに設定されているチャンネル、スループット、パケットロス率、遅延などの情報を意味する。

3.2 ロボット部

ロボット部は、制御対象となる AGV, ドローン, 作業車両, 自動運転車などのロボット自体を意味している。ロボットに搭載されているネットワークインタフェースの設定に関しては CoRein ではネットワーク部に属することに注意されたい。ロボット部は、入力としてロボットの操作, 出力としてロボットの状態情報を持つ。ロボット i の操作は以下の式で表される。

$$B^i \triangleq \{b^{1,i}, b^{2,i}, \dots, b^{|B^i|,i}\} \quad (1)$$

$b^{n,i}$ はロボットで可能な操作と対応している。例えば, 5 m/s で前進/後退, 10 m/s で前進/後退, 右/左に 5 度回転などが相当する。ロボットの状態は以下の式で表される。

$$U^i \triangleq \{u^{1,i}, u^{2,i}, \dots, u^{|U^i|,i}\} \quad (2)$$

$u^{n,i}$ に割り当てられた数字はロボットの各状態と対応している。より具体的には, ロボット自身が持つセンサで取得可能な情報を意味する。例えば, ロボットの位置情報, 向き, バッテリ残量などが相当する。

3.3 ネットワーク部

ネットワーク部は, ロボット部とロボット・ネットワー

ク制御部を接続する機構である。無線ネットワークや有線ネットワークを想定している。繰り返しになるが, 注意されたいのは, ロボットに具備された無線 LAN インターフェースはネットワーク部に属することである。ネットワーク部は, 入力として通信設定, 出力として通信情報を持つ。ネットワークの通信設定は以下の式で表される。

$$C \triangleq \{c^1, c^2, \dots, c^{|C|}\} \quad (3)$$

c^n はネットワークに設定可能な項目と対応している。例えば, どのロボットがどの AP に接続するか, 各 AP の周波数チャンネル, SDN (Software Defined Networks) を用いた場合には有線ネットワークのトポロジなどが相当する。ネットワークの状態は以下の式で表される。

$$V \triangleq \{v^1, v^2, \dots, v^{|V|}\} \quad (4)$$

v^n はネットワークの各状態と対応している。より具体的にはネットワーク機器でモニタリングして取得可能な情報を意味する。例えば, ネットワーク全体の総スループット, パケットロス率, 最低スループット, 遅延などが相当する。

3.4 ロボット・ネットワーク制御部

ロボット・ネットワーク制御部は, 各ロボット部に対して取るべき挙動や, ネットワーク部に対して接続する無線アクセスポイントなどの通信設定の指令を出す機構である。ロボット・ネットワーク制御部は, 入力として各ロボット部の状態情報 U^i と, ネットワーク部の通信情報 V , 出力として各ロボットへの操作 B^i と, ネットワーク部の通信設定 C を持つ。また, これらの値, ロボットに与えられたタスク, ネットワークの性能から以下の式で報酬が算出される。

$$R \triangleq \{r \in \mathbb{R} : r = g(s, v, a), s \in \mathcal{S}, v \in \mathcal{V}, a \in \mathcal{A}\} \quad (5)$$

ここで行動 \mathcal{A} は, 各ロボットのとるべき挙動 B^i , ロボットの台数 M , ネットワーク部の通信設定 C を用いて以下の式で表される。

$$\mathcal{A} \triangleq \left(\prod_{i=1}^M B^i \right) \times C \quad (6)$$

また, 状態 \mathcal{S} は, 各ロボットの状態 U^i , ロボットの台数 M を用いて以下の式で表される。

$$\mathcal{S} \triangleq \prod_{i=1}^M U^i \quad (7)$$

また, $g(s, v, a)$ はロボットに与えられたタスクやネットワークの状態によってアプリケーション毎に定義する報酬関数である。

ここである状態 s のときに行動 a を取ったときの報酬の推定値を返す行動価値関数 $Q(s, a)$ を考える。CoRein では,

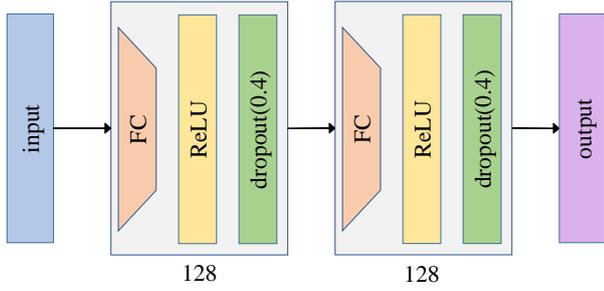


図 2: CoRein で用いる DNN

行動価値関数 $Q(s, a)$ を DNN で近似する. 図 2 に CoRein で用いている DNN を示す. 入力層はロボットの状態情報 S である. 隠れ層は, 各層は 128 ノードの全結合層 (FC: Fully Connected layer), ReLU 活性化層, dropout (0.4) の組を 2 回繰り返したものである. 出力層は行動 A を取った場合の報酬の推定値である.

この DNN のパラメータを θ とする. CoRein では, 各時刻 t において, 以下の Huber 損失関数 $L(\theta_{t-1})$ を最小化するように θ の更新を行う.

$$L(\theta_{t-1}) = \begin{cases} \frac{1}{2}f(t-1)^2 & (|f(t-1)| \leq 1) \\ |f(t-1)| - \frac{1}{2} & (\text{otherwise}) \end{cases} \quad (8)$$

ここで関数 $f(t)$ は以下の式で表される.

$$f(t) = r_t - Q(s_t, a_t | \theta_t) \quad (9)$$

式 (8) の最小化には RMSProp を用いることで以下の式で θ を更新する.

$$v_t = \beta v_{t-1} + (1 - \beta) (\nabla_{\theta_{t-1}} L(\theta_{t-1}))^2 \quad (10)$$

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{v_t} + \gamma} \nabla_{\theta_{t-1}} L(\theta_{t-1}) \quad (11)$$

α は学習率, β は平滑化定数, γ はゼロ除算を防ぐための設定パラメータである.

また, 行動選択の際には ϵ -greedy 法を用いることで行動選択にランダム性を持たせることで局所的な行動に陥らないようにしている. 具体的には, ある時刻 t において, 以下の式で行動 a を選択する.

$$\begin{cases} \arg \max_a Q(s_t, a | \theta_t) & (\text{with probability } 1 - \epsilon) \\ \text{randomly select } a \in \mathcal{A} & (\text{with probability } \epsilon) \end{cases}$$

この時, ϵ は以下の式で算出する.

$$\epsilon = \epsilon_{end} + (\epsilon_{start} - \epsilon_{end}) \exp\left(-\frac{t}{T_{decay}}\right) \quad (12)$$

ϵ_{start} は ϵ の初期値, ϵ_{end} は ϵ の収束値, T_{decay} は ϵ が ϵ_{start} から ϵ_{end} まで変化する速さを表すパラメータである.

4. 評価

CoRein の有効性を確認するために, 計算機シミュレーションによって評価を行った.

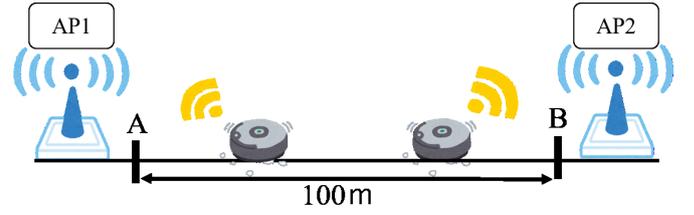


図 3: 本研究で想定した環境

4.1 評価環境

図 3 に本稿で想定する環境を示す. 102 m 離れた 2 つの無線 AP 1, 無線 AP 2 間の直線上にある地点 A, 地点 B 間 100 m を 2 台のロボットが自律的に往復するものとした. 2 台のロボットは地点 A, 地点 B 間のランダムな点から移動を開始するものとした. 各ロボットは 1 秒ごとに接続する無線 AP を選択するものとした. このとき, 各無線 AP が使用する無線チャンネルは周波数が十分に離れているため, 互いに干渉しないものとした. また, 各ロボットは利用可能な 2 つの無線 AP のうち, いずれかの無線 AP を選択して接続し, 自身が保持するデータを無線チャンネルを介してアップロードするものとした. このとき, 各ロボットから接続先無線 AP に対する上り方向の通信スループットは同一の無線 AP に接続するロボット数にしたがって定まるものとした. ここで, ロボット i が無線 AP m に対して接続する場合, 受信信号強度 $S_{i,m}$ [W] はフリスの伝達公式にしたがって次式で定めた.

$$S_{i,m} = P_t \left(\frac{\lambda}{4\pi d_{i,m}} \right)^2 \quad (13)$$

P_t はロボット i の送信電力 [W], λ は送信電波の波長 [m], $d_{i,m}$ はロボット i と無線 AP m 間の通信距離 [m] である. このとき, ロボット i から接続先無線 AP m に対する上り方向の通信スループット $C'_{i,m}$ [bps] をシャノン容量にしたがって次の通り定めた.

$$C'_{i,m} = B \log_2 \left(1 + \frac{S_{i,m}}{N} \right) \quad (14)$$

B は帯域幅 [Hz], N はノイズフロア [W] である. しかしながら, 同一の無線 AP に接続するロボット数が複数の場合, 通信に利用できる帯域は等分配されると仮定した. 他のロボットの通信も考慮した, ロボット i から接続先無線 AP m に対する上り方向の通信スループット $C_{i,m}$ [bps] を次の通り定めた.

$$C_{i,m} = \frac{B}{L_m} \log_2 \left(1 + \frac{S_{i,m}}{N} \right) \quad (15)$$

ここで, L_m は無線 AP m に接続されるロボットの台数である.

表 2 に提案手法「CoRein」の実装に使用したパラメータを示す. 提案手法「CoRein」の実装には深層学習フレームワーク PyTorch を用いた. 式 (1) で定義しているロボット

表 1: 行動の対応表

Action	Robot 1		Robot 2	
	速さ	接続先	速さ	接続先
a^1	5 m/s	AP 1	5 m/s	AP 1
a^2	5 m/s	AP 1	5 m/s	AP 2
a^3	5 m/s	AP 1	10 m/s	AP 1
a^4	5 m/s	AP 1	10 m/s	AP 2
a^5	5 m/s	AP 2	5 m/s	AP 1
a^6	5 m/s	AP 2	5 m/s	AP 2
a^7	5 m/s	AP 2	10 m/s	AP 1
a^8	5 m/s	AP 2	10 m/s	AP 2
a^9	10 m/s	AP 1	5 m/s	AP 1
a^{10}	10 m/s	AP 1	5 m/s	AP 2
a^{11}	10 m/s	AP 1	10 m/s	AP 1
a^{12}	10 m/s	AP 1	10 m/s	AP 2
a^{13}	10 m/s	AP 2	5 m/s	AP 1
a^{14}	10 m/s	AP 2	5 m/s	AP 2
a^{15}	10 m/s	AP 2	10 m/s	AP 1
a^{16}	10 m/s	AP 2	10 m/s	AP 2

i の操作 B^i を構成する要素 $b^{n,i}$ は, $b^{1,i}$ を 5 m/s で前進/後退, $b^{2,i}$ を 10 m/s で前進/後退とした. ここで, 各ロボット部は, 図 3 の地点 A, 地点 B にて自律的に前進と後退を切り替えるものとした. 式 (2) で定義しているロボット i の状態 U^i を構成する要素 $u^{n,i}$ は, $u^{1,i}$ をロボット i の X 座標, $u^{2,i}$ をロボット i の Y 座標, $u^{3,i}$ をロボット i の向きとした. 式 (3) で定義しているネットワークの通信設定 C を構成する要素 c^n は, c^1 を各ロボットの接続先 AP の設定とした. 式 (4) で定義しているネットワークの状態 \mathcal{V} を構成する要素 v^n は, v^1 をネットワーク全体の総スループットとした. 式 (5)(6)(7) で定義しているロボットの台数 M を 2 台, $g(s, v, a)$ を式 (15) の $C_{i,m}$ を用いて次の通りに定めた.

$$g(s, v, a) = \sum_{i=1}^M C_{i,m} \quad (16)$$

表 1 は式 (6) で定義した行動 \mathcal{A} の要素 a^n の対応を示したものである. 例えば, Action a^1 は, ロボット 1 が 5 m/s で前進/後退しながら無線 AP1 に接続し, ロボット 2 が 5 m/s で前進/後退しながら無線 AP1 に接続する行動を示している. 式 (10), (11) の RMSprop の設定パラメータは, PyTorch のデフォルトパラメータにしたがい α を 0.01, β を 0.99, γ を $1.0 \cdot 10^{-8}$ と定めた. 式 (12) で定義しているランダムに行動を選択する割合 ϵ の初期値 ϵ_{start} を 1.0, 収束値 ϵ_{end} を 0.0, T_{decay} を 4.2 節で示す評価にしたがって 40 と定めた. 図 4 に, 式 (12) で示した各学習時間に対するランダムに行動を選択する確率 ϵ の変化の様子を示す. 縦軸の epsilon は ϵ を意味している. ϵ は, 学習時間が約 200 秒の時点で 0 に収束する. 式 (13) で用いるロボット i の送信電力 P_i を一般的な無線 LAN モジュールの送信電力にしたがって 0.001W, 送信電波の波長 λ を 2.4GHz の電磁波

の波長 0.124m と定めた. 式 (14), (15) で用いる帯域幅 B を周波数 2.4GHz の WiFi の帯域幅である $22 \cdot 10^6$ Hz, ノイズフロア N を $1.0 \cdot 10^{-12}$ W と定めた.

様々なロボットの初期位置に対する提案手法「CoRein」の評価を行うため 3 種類のテストデータを用意した. 1 つ目のテストデータでは, 2 台のロボットの初期位置の間隔を 100 m に設定した. 具体的には, 1 台目のロボットが地点 A, 2 台目のロボットが地点 B から移動を開始するものとした. このとき, 1 台目のロボットは最初に地点 B に向かって進み, 地点 A, 地点 B 間を 30 秒間往復する一方で, 2 台目のロボットは最初に地点 A に向かって進み, 地点 A, 地点 B 間を 30 秒間往復するものとした. 2 つ目のテストデータでは, 2 台のロボットの初期位置の間隔を 50 m に設定した. 具体的には, 1 台目のロボットが地点 A, 2 台目のロボットが中点から移動を開始するものとした. このとき, 2 台のロボットは最初に地点 B に向かって進み, 地点 A, 地点 B 間を 30 秒間往復するものとした. 3 つ目のテストデータでは, 2 台のロボットの初期位置の間隔を 0 m に設定した. 具体的には, 2 台のロボットが地点 A から移動を開始するものとした. このとき, 2 台のロボットは最初に地点 B に向かって進み, 地点 A, 地点 B 間を 30 秒間往復するものとした.

表 2: パラメータの対応表

Hyperparam	CoRein
B^i	{5 m/s で前進/後退, 10 m/s で前進/後退}
U^i	{X 座標, Y 座標, 向き}
C	{各ロボットの接続先 AP}
\mathcal{V}	{ネットワーク全体の総スループット}
M	2
$g(s, v, a)$	$\sum_{i=1}^M C_{i,m}$
α	0.01
β	0.99
γ	$1.0 \cdot 10^{-8}$
ϵ_{start}	1.0
ϵ_{end}	0.0
T_{decay}	40
P_i	0.001
λ	0.124
B	$22 \cdot 10^6$
N	$1.0 \cdot 10^{-12}$

4.2 パラメータ T_{decay} による影響

図 5 に $T_{decay} = 100, 300, 500, 700$ における学習時間に対する平均合計スループットを示す. $T_{decay} = 100$ としてラベルを付与した実線のグラフは, T_{decay} を 100 に設定した場合の CoRein の平均合計スループットを示している. 平均合計スループットは, 10 回の試行平均に基づいた 100 秒単位での平均値とした. 図 5 の評価結果から, T_{decay} を

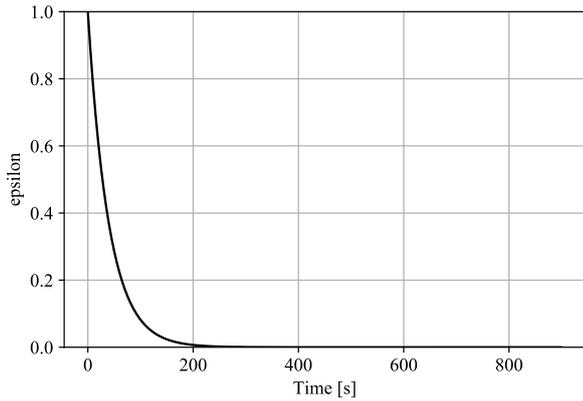


図 4: $T_{\text{decay}} = 40$ における ϵ の変化

300 に設定した場合, T_{decay} を 500 に設定した場合, T_{decay} を 700 に設定した場合と比較して, T_{decay} を 100 に設定した場合は, 学習結果の収束にかかる学習時間が短く, 平均合計スループットが大きい時間が多いことが分かった.

図 6 に各 T_{decay} における学習収束時の平均合計スループットを示す. 図 5 の評価結果に基づいて, $T_{\text{decay}} = 100$ 付近の値である $T_{\text{decay}} = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180$ における学習収束時の平均合計スループットを計測した. 学習収束時の平均合計スループットとは, 学習結果が収束しはじめる学習時間から, 3000 秒までの合計スループットの平均値を意味する. 図 5 の評価結果に基づいて, 学習結果が収束しはじめる学習時間を 300 秒として評価を行った. 図 6 の評価結果から, $T_{\text{decay}} = 40, 140, 120, 20, 130, 60, 90, 110, 30, 100, 50, 170, 80, 180, 150, 160, 70, 10$ の順で学習収束時の平均合計スループットが大きいことが分かった. 図 7 に, 図 6 において学習収束時の平均合計スループットが大きい順に T_{decay} を 4 つ選択した, $T_{\text{decay}} = 20, 40, 120, 140$ における学習時間に対する平均合計スループットを示す. 図 7 の評価結果から, T_{decay} を 20 に設定した場合や, T_{decay} を 120 に設定した場合, T_{decay} を 140 に設定した場合と比較して, T_{decay} を 40 に設定した場合の学習結果が収束するために必要な学習時間は短いことが分かった.

図 6, 図 7 の評価結果から, 評価を行った範囲では, CoRein のパラメータ T_{decay} は 40 が最適であることが分かった.

4.3 各ロボットの接続先とする無線 AP の選択

CoRein が, 各ロボットの接続先とする無線 AP の選択にどのような影響を与えているか評価した. 図 8, 図 9, 図 10 にそれぞれのテストデータにおける, 各学習時間に対する 2 台の無線 AP が通信に利用されていた時間の割合を示す. 1 台目のロボットが無線 AP1 に接続している一方

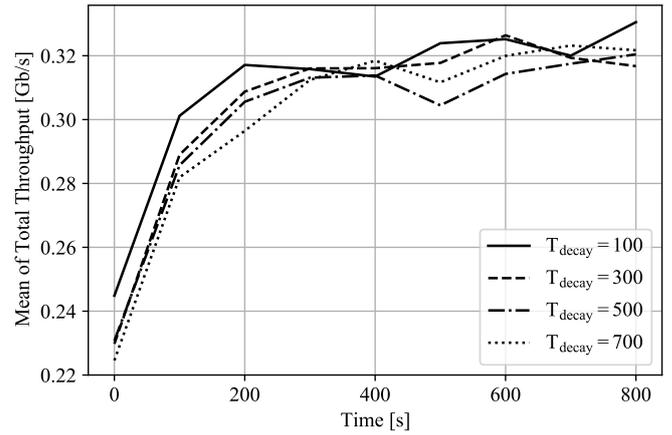


図 5: $T_{\text{decay}} = 100, 300, 500, 700$ における学習時間に対する平均合計スループット

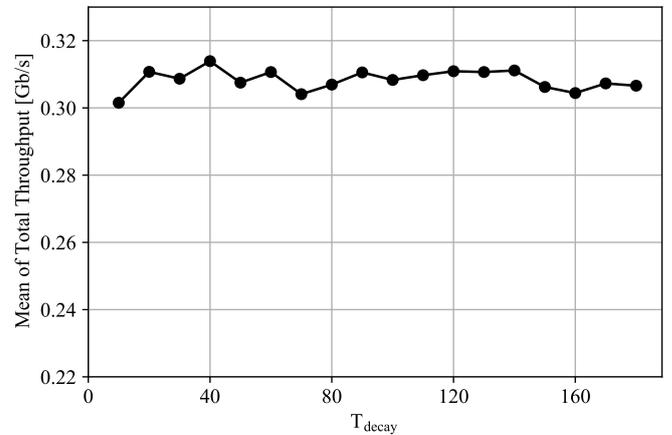


図 6: 各 T_{decay} における学習収束時の平均合計スループット

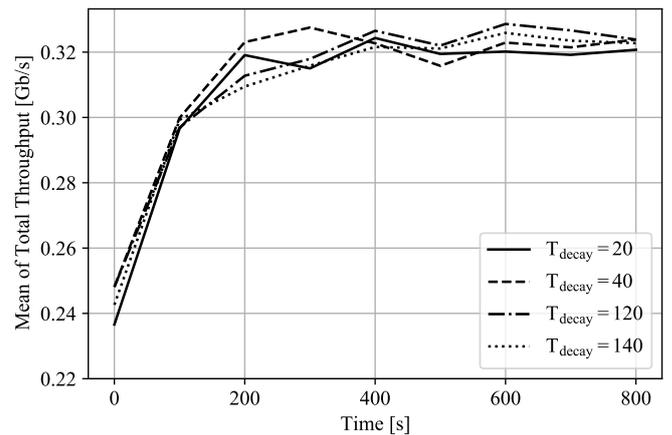


図 7: $T_{\text{decay}} = 20, 40, 120, 140$ 時における学習時間に対する平均合計スループット

で, 2 台目のロボットが無線 AP2 に接続している時間, もしくは 1 台目のロボットが無線 AP2 に接続している一方で, 2 台目のロボットが無線 AP1 に接続している時間が, 2 台の無線 AP が通信に利用された時間に相当する. 縦軸

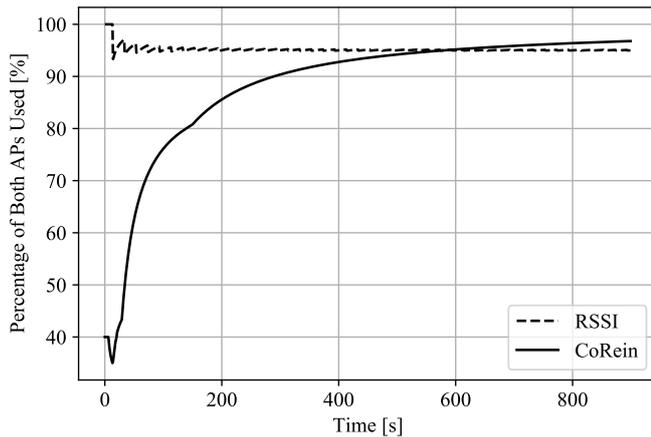


図 8: 2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでの各学習時間における 2 台の無線 AP が通信に利用された割合

は、各学習時間に占める 2 台の無線 AP が通信に利用された時間の割合を示している。ここで、2 台の無線 AP が通信に利用された時間の割合は、10 回の試行平均とした。繰り返しになるが、式 (15) の通り、同一の無線 AP に接続するロボット数が複数の場合、通信に利用できる帯域は等配分されるため、各ロボットは同一の無線 AP に接続し通信を行うと、上り方向のスループットは減少する。したがって、2 台のロボットは別々の無線 AP を利用することが望ましい。

図 8 の評価結果からは、2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでは、約 600 秒の時点で提案手法が、RSSI に基づいて接続先の無線 AP を選択する手法と同等かそれ以上に、2 台の無線 AP を利用している時間が多いことが分かった。図 9 の評価結果からは、2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでは、約 50 秒の時点で提案手法が、RSSI に基づいて接続先の無線 AP を選択する手法と比較して、2 台の無線 AP を利用している時間が多いことが分かった。図 10 の評価結果からは、2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでは、提案手法が、RSSI に基づいて接続先の無線 AP を選択する手法と比較して、2 台の無線 AP を利用している時間が多いことが分かった。

3 種類のテストデータの評価から、提案手法「CoRein」は十分な学習時間を得ることで、さまざまなロボットの初期位置に対して、2 台の無線 AP を利用するように接続先とする無線 AP を選択することが分かった。

4.4 2 台のロボットの位置関係

CoRein が、2 台のロボットの位置関係にどのような影響を与えているか評価した。2 台のロボットがそれぞれ最も電波強度の大きい無線 AP に接続しながら、通信に利用できる帯域が等配分されないようにするためには、2 台のロ

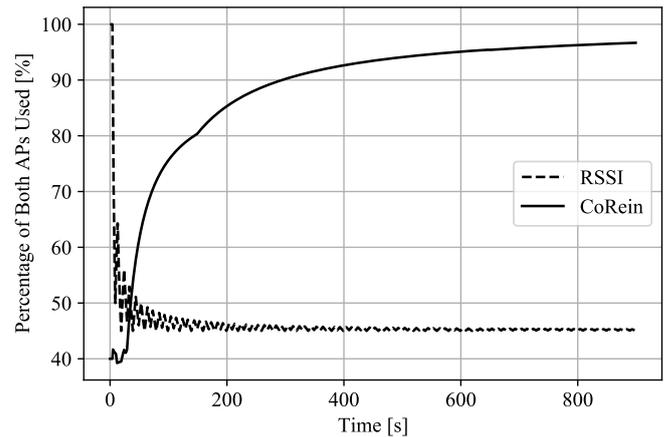


図 9: 2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでの各学習時間における 2 台の無線 AP が通信に利用された割合

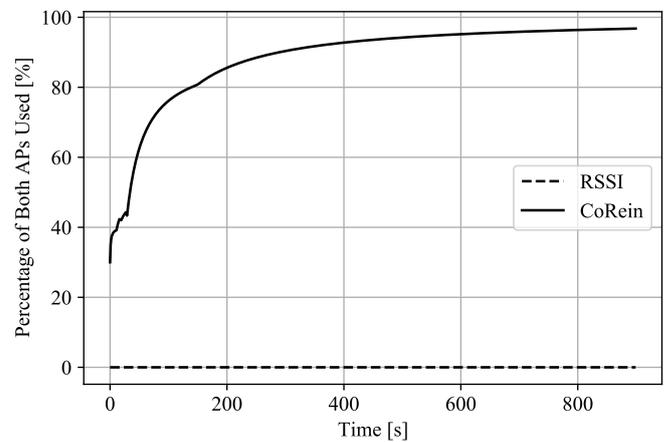


図 10: 2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでの各学習時間における 2 台の無線 AP が通信に利用された割合

ボットの位置は、どちらか片方の無線 AP 周辺に偏らないことが望ましい。

図 11、図 12、図 13 に各テストデータにおける、各学習時間に対する 2 台のロボットが左右の領域に分配された時間の割合を示す。ロボットが左右の領域に分配された状態とは、2 台のロボットが地点 A、地点 B 間の中間地点を挟むような位置関係の状態を意味する。例えば、1 台目のロボットが地点 A から 5 m の地点に位置している一方で、2 台目のロボットが地点 B から 10 m の地点に位置している状態は、ロボットが左右の領域に分配された状態に該当する。ここで、2 台のロボットが左右の領域に分配された時間の割合は、10 回の試行平均とした。

図 11 の評価結果からは、2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでは、提案手法は速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と比較すると、

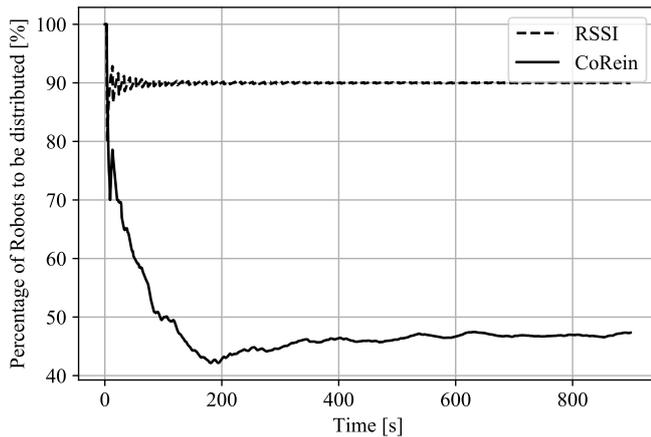


図 11: 2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでの各学習時間に対するロボットが左右の領域に分配された時間の割合

2 台の各学習時間に対するロボットが左右の領域に分配された状態の時間が少ないことが分かった。十分な学習時間が経過した場合、ロボットが左右の領域に分配された状態の割合は約 48 % に収束することが分かった。図 12 の評価結果からは、2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでは、速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と比較すると、提案手法が優れた性能を達成できていることが分かった。十分な学習時間が経過した場合、ロボットが左右の領域に分配された状態の割合は約 50 % に収束することが分かった。図 13 の評価結果からは、2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでは、速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と比較すると、提案手法が優れた性能が達成できていることが分かった。十分な学習時間が経過した場合、ロボットが左右の領域に分配された状態の割合は約 43 % に収束することが分かった。

3 種類のテストデータの評価から、提案手法「CoRein」は十分な学習時間を得ることで、さまざまなロボットの初期位置に対して、ロボットが左右の領域に分配される割合は 40 % 以上 50 % 以下の範囲に収束することが分かった。CoRein の報酬関数は瞬間的な報酬に基づいているため、各ロボットは瞬時的な位置関係の調整を行っていた。長期的な報酬に基づいた報酬関数を設定することで、各ロボットは長期的な観点で位置関係を調整することが可能となるため、左右の領域に分配される割合は増加すると考えられる。

4.5 収束するまでの時間

図 14, 図 15, 図 16 にそれぞれのテストデータにおける学習済エピソード数に対する上り方向のスループットの合計の平均値を示す。ここで、合計スループットの平均値は

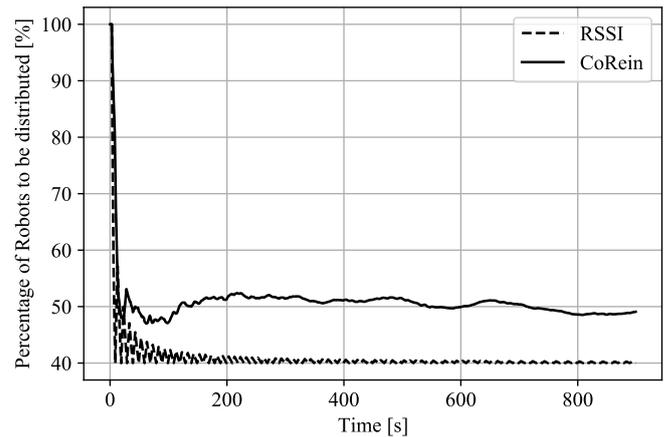


図 12: 2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでの各学習時間に対するロボットが左右の領域に分配された時間の割合

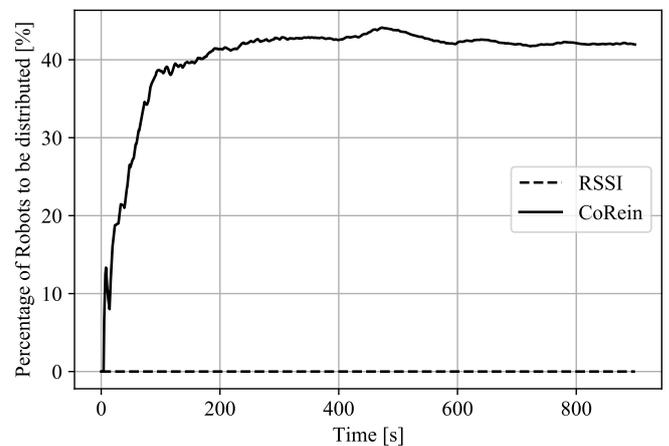


図 13: 2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでの各学習時間に対するロボットが左右の領域に分配された時間の割合

10 回の試行平均に基づいた 10 秒単位での平均値とした。各ロボットが速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法を RSSI, 提案手法を CoRein とした。

図 14 の評価結果からは、2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでは、約 200 秒の学習時間で提案手法が平均的には、速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と同等の性能を達成できていることが分かった。図 15 の評価結果からは、2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでは、約 50 秒の学習時間で提案手法が平均的には、速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と同等の性能を達成できていることが分かった。図 16 の評価結果からは、2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでは、

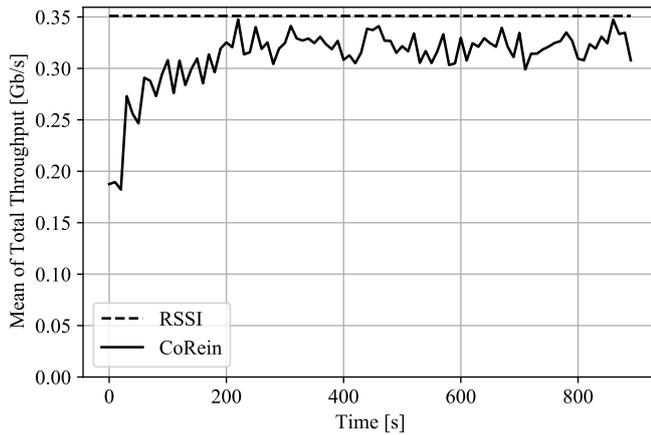


図 14: 2 台のロボットの初期位置の間隔を 100 m に設定したテストデータでの各学習時間における合計スループット

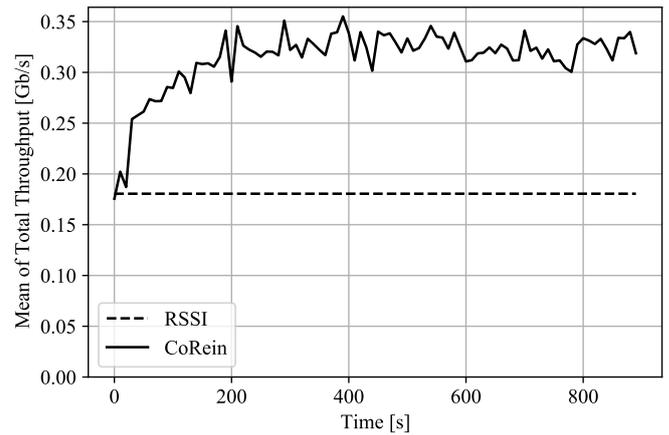


図 16: 2 台のロボットの初期位置の間隔を 0 m に設定したテストデータでの各学習時間における合計スループット

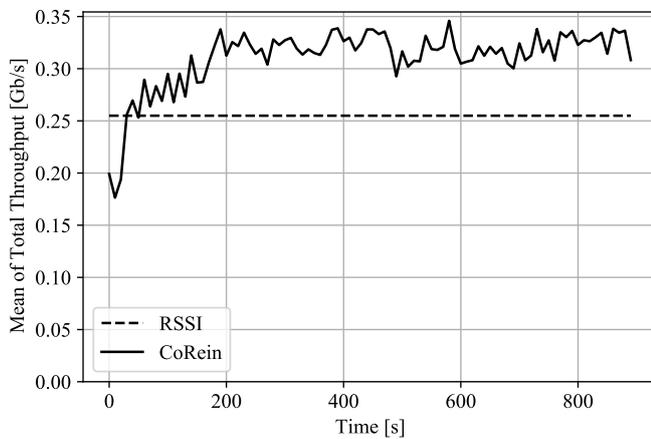


図 15: 2 台のロボットの初期位置の間隔を 50 m に設定したテストデータでの各学習時間における合計スループット

提案手法が速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法よりも優れた性能を達成できていることが分かった。

3 種類のテストデータの評価から、提案手法「CoRein」は十分な学習時間を得ることで、さまざまなロボットの初期位置に対して、各ロボットが速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と同等か、もしくは優れた性能を達成できていることが分かった。

5. おわりに

本稿では、ネットワークを介したロボットの安定的な制御の実現に向けて、深層強化学習を用いたネットワーク・ロボット挙動同時最適化手法「CoRein」を提案した。具体的には、ロボットの操作とネットワークの設定を行動、ロボットで取得したセンサ情報を状態、ロボットに与えられたタスクやネットワーク性能を報酬関数とすることでロボットの挙動制御に深層強化学習を適用する。シミュレー

タ上で深層強化学習を行い、3 つのテストデータを用いて評価を行った。合計スループット最大化の観点で、CoRein はさまざまなロボットの初期位置に対して、各ロボットが速さ 10 m/s で移動しながら無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法と同等か、もしくは優れた性能を達成できることを示した。

謝辞

本研究は JSPS 科研費 (JP19H01101, JP20K19783), NTT 未来ねっと研究所の支援の下で行った。

参考文献

- [1] Krijestorac.Enes, Hanna.Samer, and Cabric.Danijela, "Uav access point placement for connectivity to a user with unknown location using deep rl," 2019 IEEE Globecom Workshops (GC Wkshps), pp.1–6IEEE, 2019.
- [2] Zhang.Yue, and Dai.Lin, "Joint optimization of placement and coverage of access points for ieee 802.11 networks," ICC 2020-2020 IEEE International Conference on Communications (ICC), pp.1–7IEEE, 2020.
- [3] D. Amir, Beyranvand.Hamzeh, and S. A., "Optimal placement of access points in cellular visible light communication networks: An adaptive gradient projection method," IEEE Transactions on Wireless Communications, vol.19, no.10, pp.6813–6825, 2020.
- [4] Zhou.Baoding, Tu.Wei, Mai.Ke, Xue.Weixing, Ma.Wei, and Li.Qingquan, "A novel access point placement method for wifi fingerprinting considering existing aps," IEEE Wireless Communications Letters, vol.9, no.11, pp.1799–1802, 2020.
- [5] Zhao.Lei, Sun.Wen, Shi.Yongpeng, and Liu.Jiajia, "Optimal placement of cloudlets for access delay minimization in sdn-based internet of things networks," IEEE Internet of Things Journal, vol.5, no.2, pp.1334–1344, 2018.
- [6] Qiu.Shuwei, Chu.Xiaowen, Leung.Yiu-Wing, and N.K. Yin, "Joint access point placement and power-channel-resource-unit assignment for 802.11 ax-based dense wifi with qos requirements," IEEE INFOCOM 2020-IEEE Conference on Computer Communications, pp.2569–2578IEEE, 2020.

- [7] Korb.Piotr, Hausman.Sławomir, and D. Barba.Paolo, "Application of evolutionary approach for multi-objective improvement of indoor access point placement," 2019 13th European Conference on Antennas and Propagation (EuCAP), pp.1–4IEEE, 2019.
- [8] Alsmady.Abdulsalam, and Awad.Fahed, "Optimal wi-fi access point placement for rssi-based indoor localization using genetic algorithm," 2017 8th International Conference on Information and Communication Systems (ICICS), pp.287–291IEEE, 2017.
- [9] I.S. 802.11k 2008, "Ieee standard for information technology - local and metropolitan area networks - specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 2: Fast basic service set (bss) transition," Amendment to IEEE Std 802.11-2007, pp.1–244, 2008.
- [10] I.S. 802.11r 2008, "Ieee standard for information technology - local and metropolitan area networks - specific requirements - part 11: Wireless lan medium access control (mac)and physical layer (phy) specifications amendment 1: Radio resource measurement of wireless lans," Amendment to IEEE Std 802.11-2007, pp.1–126, 2008.
- [11] K. Zhu, D. Niyato, and P. Wang, "Network selection in heterogeneous wireless networks: Evolution with incomplete information," 2010 IEEE Wireless Communication and Networking Conference, 2010.
- [12] M.A. Khan, U. Toseef, S. Marx, and C. Goerg, "Game-theory based user centric network selection with media independent handover services and flow management," 2010 8th Annual Communication Networks and Services Research Conference, May 2010.
- [13] H. Pervaiz, "A multi-criteria decision making (mcdm) network selection model providing enhanced qos differentiation to customers," IEEE Communications Surveys and Tutorials, May 2010.
- [14] R. Trestian, O. Ormond, and G.M. Muntean, "Game theory-based network selection: Solutions and challenges," IEEE Communications Surveys and Tutorials, vol.14, no.4, pp.1212–1231, February 2012.
- [15] S. Wang, H. Liu, P.H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," IEEE Transactions on Cognitive Communications and Networking, vol.4, no.2, pp.257–265, February 2018.
- [16] O. Naparstek, and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," IEEE Transactions on Wireless Communications, vol.18, no.1, pp.310–323, November 2018.
- [17] A.C.C. Hsu, D.S.L. Wei, and C.C.J. kuo, "A cognitive mac protocol using statistical channel allocation for wireless ad-hoc networks," 2007 IEEE Wireless Communications and Networking Conference, pp.105–110, March 2007.
- [18] M. Li, P. Li, X. Huang, Y. Fang, and S. Glisic, "Energy consumption optimization for multihop cognitive cellular networks," ieee trans. mobile comput., IEEE Transactions on Mobile Computing, vol.14, no.2, pp.358–372, February 2015.
- [19] C. Xiong, G.Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in ofdma networks," IEEE Transactions on Communications, vol.60, no.12, pp.3767–3778, September 2012.
- [20] Huang.Hao, Peng.Yang, Yang.Jie, Xia.Wenchao, and Gui.Guan, "Fast beamforming design via deep learning," IEEE Transactions on Vehicular Technology, vol.69, no.1, pp.1065–1069, 2019.
- [21] Alkhateeb.Ahmed, Alex.Sam, Varkey.Paul, Li.Ying, Qu.Qi, and Tujkovic.Djordje, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," IEEE Access, vol.6, pp.37328–37348, 2018.
- [22] Wu.Qingqing, and Zhang.Rui, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," 2018 IEEE Global Communications Conference (GLOBECOM), pp.1–6IEEE, 2018.
- [23] Wu.Qingqing, and Zhang.Rui, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," IEEE Transactions on Wireless Communications, vol.18, no.11, pp.5394–5409, 2019.
- [24] S. Tekinay, and B. Jabbari, "Handover and channel assignment in mobile cellular networks," IEEE Communications Magazine, vol.29, no.11, pp.42–46, November 1991.
- [25] 赤岩芳彦, デジタル移動通信技術のすべて, コロナ社, 東京, 2013.
- [26] A.Mishra, M.Shin, and W.Arbaugh, "An empirical analysis of the ieee 802.11 mac layer handoff process," SIGCOMM ComputerCommunication Review, vol.33, pp.93–102, April 2003.
- [27] M. Sanchez, and AzzedineBoukerche, "On ieee 802.11k/r/v amendments: Do they have a real impact?," IEEE Wireless Communications, vol.23, pp.48–55, February 2016.
- [28] P. Jaesung, and L. Yujin, "A handover prediction model and its application to link layer triggers for fast handover.," Wireless Personal Communications, vol.52, pp.501–516, February 2010.
- [29] J.McNair, and F.Zhu, "Vertical handoffs in fourth-generation multinet network environments," IEEE Wireless Communications, vol.11, pp.8–15, 2004.
- [30] J. Zhang, F. Yu, S. Wang, T. Huang, Z. Liu, and Y. Liu, "Load balancing in data center networks: A survey," IEEE Communications Surveys and Tutorials, vol.20, no.3, pp.2324–2352, March 2018.
- [31] T. Han, and N. Ansari, "A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources," IEEE/ACM Transactions on Networking, vol.24, no.2, pp.1038–1051, March 2015.
- [32] S. Singh, H.S. Dhillon, and J.G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," IEEE Transactions on Wireless Communications, vol.12, no.5, pp.2484–2497, April 2013.
- [33] H. Kim, G. d. Veciana, X. Yang, and M. Venkatasubramanian, "Distributed alpha - optimal user association and cell load balancing in wireless networks," IEEE/ACM Trans. Networking, vol.20, no.1, pp.177–190, Feb 2012.
- [34] C. Raiciu, S. Barre, C. Pluntke, A. Greenhalgh, D. Wischik, and M. Handley, "Improving datacenter performance and robustness with multipath tcp," ACM SIGCOMM Computer Communication Review, vol.41, no.4, pp.266–277, August 2011.
- [35] J. Zhang, F. Yu, S. Wang, T. Huang, Z. Liu, and Y. Liu, "Hula: Scalable load balancing using programmable data planes," Proc. ACM Symp. SDN Res., no.3, pp.1–12, March 2016.
- [36] N. Katta, M. Hira, A. Ghag, C. Kim, I. Keslassy, and J. Rexford, "Clove: How i learned to stop worrying about the core and love the edge," Hotnets '16: Proceedings of the 15th ACM Workshop on Hot Topics in Networks, pp.155–161, November 2016.