災害用エネルギーハーベスティング型無線機における深層強化学習を用いた無線アクセス制御方式

研究代表者　　王　瀟岩　　茨城大学　工学部　准教授

1 序論

大規模災害が発生した直後の時間帯において、被災者の安否確認や支援を行うための通信手段の確保が極めて重要である。ただし、災害が発生したエリアに設備の損害や停電により人命救助にも重要な役割を果たす通信インフラが途絶する危険が存在する。例えば、2011 年の東日本大震災では、通信インフラに壊滅的な被害を受け、地震後に約 29000 の基地局が機能できなかった。携帯電話事業者は地震発生直後に通信インフラの復旧へ多大な労力を投入したが、約 1 か月半以上の時間がかかった[1]。

「72 時間の壁」とも言われる災害発生直後の時間帯において、被害を軽減し、命を救うために重要である。したがって、通常の通信インフラが直ちに回復できない場合でも、通信を確保する代替手段が極めて重要と考えられる。そのため、Movable and Deployable Resource Unit (MDRU) と呼ばれる移動式 ICT ユニットを被災地域に配置する仕組みを提案していた[2]。MDRU は衛星や地上回線などの通信インフラを通してインターネットに繋がり、被災者や救急隊などにネットワークアクセスを提供できる。MDRU は車両型のリソースユニットとして、被災地域に迅速に新のローカル無線アクセスネットワークを構築でき、バッテリー駆動で 5 日間以上動作可能になる[3]。

しかし、MDRU の通信サービス範囲は一般的に 500m 程度で制限されている。広い災害地域に通信サービスを回復するため、大量の MDRU の配置は必要となる。しかし、時間とコストの制限により、短時間に多数の MDRU を配置することは困難である。MDRU の通信サービス範囲を拡大するため、マルチホップ通信技術を用いて複数の低価の relay ノードを中継として機能することにより、MDRU のサービスエリアを拡大する仕組みが提案されていた[4]。

本文では、このような災害後ネットワークにおける、エネルギーハーベスティング機能を備える移動ユーザー無線機 (UE: user equipment) が環境情報を収集し、MDRU に報告するシナリオを検討する。このようなシナリオにおける、UE が災害後環境の写真またはビデオ情報を MDRU に転送し、MDRU が収集された情報を活用し、生存者の検出または損傷状態の評価を行える[5]。また、災害地域には一般的に充電が困難になるため、UE の稼働時間を延長するためエネルギーハーベスティング機能[6]を備えることを想定する。このような災害後ネットワークシナリオでは、送受信局間のチャネル状態、UE の位置、パケット生起、およびエネルギー状態が常に変動しているため、UE の最適な無線アクセス制御は困難となる。さらに、災害地域で生き残った通信インフラ分布とトラフィック量は、通常時と比べ、大きく変わる。この事前の統計情報がないため、既存の最適化アルゴリズム[7]の適用が困難になる。この問題を解決するため、本研究は災害後ネットワークにおける、深層強化学習[8]を用いたエネルギーハーベスティング型無線機の最適な無線アクセス制御方式を検討する。具体的には、UE での無線アクセス問題を MDP(Markov Decision Process)問題としてモデル化する。UE の送信タイムイング、ルーティング、送信電力などを環境に適応できるため、UE が周りの無線環境の情報を収集しながら、現在のネットワーク状態における異なる行動に対応するコストを計算し、遅延時間とパケットドロップ率を最小化になる最適な無線アクセスポリシーを学習する。提案手法の有効性は、シミュレーション結果によって検証された。評価結果により、提案手法が従来手法と比較し、遅延とパケット損失率に対する優位性を示した。

2 システムモデル

2-1 ネットワークアーキテクチャ

MDRU は，ICT サービスの提供に必要な装置類を収容した可搬型のユニットである。災害発生直後、MDRU は被災地など情報通信を求める所に設置されることにより、周辺におよそ 500m の WiFi などのローカル災害対応ネットワークを構築し、ネットワークのカバー範囲内のスマートフォンやセンサーなどの UE に最低限の通信サービスを即時に提供できる。UE からの通信データは MDRU が生き残る光ファイバや衛星通信などを活

公益財団法人電気通信普及財団
研究調査助成報告書　第37号　2022年度

用し、データセンターに転送される。災害対応ネットワークと広域ネットワークの間に MDRU が情報通信中継として損害した通信インフラの機能を果たす。MDRU 内にはいくつの可動 WiFi モジュールがあり、5 GHz、2.4 GHz、および 920MHz の周波数の使用が可能となる[9]。MDRU には、短時間で災害地域に通信サービスを回復でき、被災者や救援隊などの情報通信を確保できるため大きな意義がある。MDRU の通信サービス範囲は一般的に 500m 程度で制限されているため、マルチホップ通信技術により複数の低価のリレーノード(relay)が AP 中継として配置することにより MDRU の通信サービスエリアが拡大できる。

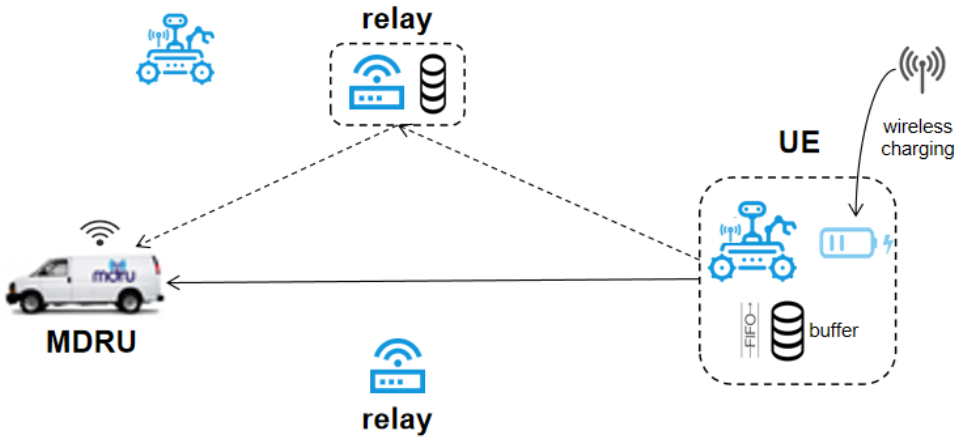


図 1 ネットワークモデル

本文では、図 1 に示すように、一つの MDRU と複数の relay が構成された災害対応無線アクセスネットワークを考慮する。各 relay が異なるチャネルを持ち、UE から受信された情報を MDRU へ転送する。MDRU と relay は同じ周波数帯域幅 B を持つことを仮定する。また、各 relay と MDRU が複数の UE からパケットを受信しているため、先入先出（FIFO）のルールに従って、パケットを処理する。n 番目の relay と MDRU 間のチャネル利得$g^k_{(n),(m)} \in G^k$で表し、以下の式で定義される。

$$g^k_{(n),(m)} = 20\log_{10}\left(\frac{\lambda}{4\pi d}\right) + \omega, \qquad (2.1)$$

ただし、dは n 番目の relay と MDRU 間の距離であり、λは波長を表す。また、フェージングによりチャネル状態の変化を考慮するため、チャネル利得に変数ωを加算する。変数ωは平均 0、分散σ_n^2に従うガウス乱数である。本研究には、短い時間幅を考慮するため、MDRU と relay のエネルギーが無限大と仮定する。また、一般的に relay の処理能力が高いため、バッファサイズの上限を設定しない。

災害エリアにおける、様々なユーザ無線機が災害対応ネットワークを利用できる。例えば、被災者が避難経路や避難情報を求めること[10]や、救援隊が互いに連絡し、リクエストや報告することなどが考えられる。また、より良い救援と避難を行うため、センサーノードの配置により災害地域の環境監視と情報収集も重要である。具体的には、ロボットやドローンなどに搭載されたセンサーにより、画像またビデオ情報を収集し、クラウドサーバに送信する。クラウドがそちらの情報を活用し、効率的な救援活動をサポートする。本研究では、UE が被災地域の環境を監視する移動センサーノードを想定する。

UE には、被災地に連続長時間に稼働することが望ましい。ただ、災害地域に充電することが困難となるため、UE はエネルギーハーベスティング機能[11]を備えることを仮定する。 ここで、E_e^kはタイムスロット k における、UE に収穫されたエネルギーの量を表す。エネルギーの収穫は平均ハーベスト率$\lambda_e \in [0,1]$のポアソン分布に従う。タイムスロット k+1 における、UE のエネルギーの量は式(2.2)により変化する。

$$H^{k+1} = \min\{H^k - E^k + E_e^k,\ \bar{H}\}, \qquad (2.2)$$

ここで、H^kはタイムスロット k における UE のエネルギー量、E^kは送信で消耗したエネルギーの量、$\bar{H}$は UE が保有できる最大のエネルギー量である。

災害エリアの環境をリアルタイムで把握するために、UE はモニタリングされた情報を MDRU に転送する。MDRU は、収集された画像またビデオ情報を活用し、生存者の検出やインフラ損傷の評価などをサポートする。ここで、環境情報のセンシングデータ（パケット）がランダムで生起することを仮定する。UE はパケットの平均生起率$\lambda_p \in [0,1]$のベルヌーイ確率分布に従ってパケットを生起する。P^kはタイムスロット k における、パケット生起を表す指示変数であり、式(2.3)に示すように、パケットを生起する場合$P^k = 1$になり、そうでなければ$P^k = 0$になる。

$$\Pr\{P^k = 1\} = 1 - \Pr\{P^k = 0\} = \lambda_p \tag{2.3}$$

異なるアプリケーションにより、移動している UE を考えられる。タイムスロット k における、UE の位置は$L^k \in L$で表される。UE と n 番目 relay 間のチャネル利得$g^k_{(u),(n)} \in G^k$も式(2.1)で計算される。UE が移動する場合、タイムスロット k に送受信点の距離はd^kで表示する。UE と MDRU 間のチャネル利得$g^k_{(u),(m)} \in G^k$は同じ方法で計算できる。

UE に対して、エネルギーがないまたチャンネル状態が悪い場合、パケットをサイズ$\bar{J}$のバッファに一時的に保存することが可能となる。パケットの再送信も FIFO のルールに従って、生成された新のパケットと保存されたパケットを一括で送信する。タイムスロット k に、$J^k(\leq \bar{J})$はバッファに保存されているパケット数を表す。バッファのパケット数J^kが$\bar{J}$を超える場合、パケットを捨てるしかない。

2-2 UE の無線アクセス制御方式

災害対応ネットワークにおいて、エネルギーハーベスティング機能を備える UE が周りの環境をモニタリングして、生成されたパケットを MDRU に送信する。UE のパケットは一つの relay を経由して MDRU に転送すると、直接に MDRU に送信することが可能である。

次に、UE の具体的な無線アクセス方式を述べる。

タイムスロット k における、UE の無線アクセス方式を$T^k \in \{-1,-2\} \cup \{0\} \cup N$に定義する。$T^k = 0$はパケットを直接 MDRU に送信すること、$T^k = -2$はパケットを送信せずドロップすること、$T^k = -1$はパケットを送信せずバッファに保存すること、$T^k = n, n \in N^+$は n 番目の relay を経由し MDRU に送信することをそれぞれ表す。

続きまして、まず relay を経由し MDRU に送信する場合の送信遅延を計算する。UE が n 番目 relay にパケットを送信するかかる時間$t^k_{(u),(n)}$は以下のように算出される。

$$R^k_{(u),(n)} t^k_{(u),(n)} = (J^k + P^k)\mu, \tag{2.4}$$

ここで、μはパケットサイズ、$R^k_{(u),(n)}$は UE から relay n へのデータレットであり、式(2.5)のように計算される。

$$R^k_{(u),(n)} = \mathrm{B}\log_2\left(1 + I^{-1} g^{k}_{(u),(n)} \frac{E^k}{t^k_{(u),(n)}}\right), \tag{2.5}$$

ただし、Iはバックグラウンドノイズの平均電力である。

タイムスロット k における、n 番目 relay に先に到着していたパケットが存在する場合、UE のパケットが relay から MDRU までに転送する際には、前に到着していたパケットの処理を待たなければならない。待ち時間を含む転送の遅延時間$t^k_{(n),(m)}$は次のように計算される。

$$R^k_{(n),(m)} t^k_{(n),(m)} = (2m^k_n + J^k + P^k)\mu, \tag{2.6}$$

$$R^k_n = \mathrm{B}\log_2\left(1 + I^{-1} g^{k}_{(n),(m)} \frac{E}{t^k_{n,m}}\right), \tag{2.7}$$

ここで、m_n^kはタイムスロットkにおける、n番目のrelay のところ先に到着していたパケット数、R_n^kはn番目のrelayのデータレット、Eはrelayの送信エネルギー、$g_{(n),(m)}^k \in G^k$はn番目relayとMDRU間のチャネル利得である。ここで、簡単化のために、他のUEのパケットの受信時間と転送時間が同じであることを仮定する。

次、UEがパケットを直接MDRUへ送信する場合、UEとMDRU間のチャンネル利得$g_{(u),(m)}^k \in G_{(u),(m)}$を用いて転送時間$t_{u,m}^k$の計算式は以下になる。

$$R_{(u),(m)}^k t_{(u),(m)}^k = (m_m^k + J^k + P^k)\mu, \tag{2.8}$$

$$R_{(u),(m)}^k = \mathrm{B}\log_2(1 + I^{-1} g_{(u),(m)}^k \frac{E^k}{t_{(u),(m)}^k}) \tag{2.9}$$

ここで、m_m^kはタイムスロットkにおける、MDRUのところ先に到着していたパケット数である。

以上のUEの無線アクセス方式をまとめ、UEからMDRUまでのパケットの転送遅延d_kは以下の式(2.10)で計算できる。

$$d_k = \begin{cases} 0, & T^k = -1, -2; \\ t_{(u),(m)}^k & T^k = 0 \quad ; \\ t_{(u),(n)}^k + t_{(n),(m)}^k & T^k \in N \quad ; \end{cases} \tag{2.10}$$

ただし、計算された転送遅延d_kには、パケットをバッファ内に保存する場合の待ち時間が含まれていない。この待ち時間は、タイムスロット長τにパケットがバッファに保存したタイムスロットの数を掛けることで簡単に取得できる。

本研究の目的は、生起されたパケットのドロップ率と遅延時間を最小化になる、UEの最適な無線アクセスポリシーΦを見つけ出すことである。そのため、タイムスロットkにおける、パケットのドロップ率と遅延時間を評価するコスト関数r_kは以下の式(2.11)に定義する。

$$r_k = d_k + \rho \cdot P^k + \xi \cdot J^k, \tag{2.11}$$

ここで、パケットがドロップされた場合は$P^k = 1$、それ以外の場合は$P^k = 0$。$\rho \cdot P^k$と$\xi \cdot J^k$はそれぞれパケットドロップとパケットをバッファに保存する場合のペナルティーであり、$\xi, \rho \in R^+$は重みである。これで、最適な無線アクセスポリシーΦを見つけ出す問題は長期的なコスト関数r_kを最小化する問題に転換できる。

2-3 無線アクセス制御問題のモデル化

最適な無線アクセスポリシーΦを見つけ出す問題はシングルエージェントのマルコフ過程(MDP)にモデル化できる。

本研究では、環境状態集合$\mathcal{S}$、アクセス制御ポリシーΦと、現在の状態sから次の状態s'までの状態遷移確率関数が次のように定義する。

タイムスロットkにおける、状態s^kは環境状態集合$\mathcal{S} = \mathcal{G} \times \mathcal{L} \times \mathcal{H} \times \mathcal{P} \times \mathcal{T}$に属する。環境状態集合$\mathcal{S}$には、UEの位置集合$\mathcal{L}$、エネルギー量の集合$\mathcal{H}$、チャネル状態集合$\mathcal{G}$、パケット生起の集合$\mathcal{P}$及びバッファに保存しているパケット数の集合$\mathcal{T}$が含まれている。

無線アクセスポリシーΦはパケット処理方式T^kと送信エネルギーE^kで構成され、最適な無線アクセスポリシーΦ*は長期的にコスト関数を最小化になれるポリシーである。具体的に次の式(2.12)のように示される。

$$\Phi(s^k) = (\Phi_T(s^k), \Phi_E(s^k)) = (T^k, E^k), \tag{2.12}$$

ここで、$\Phi_T(s^k), \Phi_E(s^k)$はそれぞれ環境s^kにおいて、UEのパケット処理方式と送信エネルギー制御方式である。

このような無線アクセスポリシーに基づいて、UEの位置L^k、パケットの生起P^kとエネルギーハーベストの量E_e^kに関わる状態遷移確率関数式は(2.13)のように表される。

$$\Pr\{s^{k+1}|s^k, \Phi(s^k)\} = \left(\prod_{n=1}^{N+1} \Pr\left(g_{(u),(n)}^{k+1} | g_{(u),(n)}^k\right)\right) \cdot \Pr\{L^{k+1}|L^k\} \tag{2.13}$$

$$\cdot \Pr\{H^{k+1}|H^k, \Phi(s^k)\} \cdot \Pr\{P^{k+1}|P^k\} \cdot \Pr\{J^{k+1}|J^k, \Phi(s^k)\},$$

状態遷移確率関数と無線アクセスポリシーが与えられる場合、初期状態$\hat{s}$により予想される長期コストは次の式(2.14)ように表す。

$$\mathrm{V}(s, \Phi) = E_{\Phi}\left[(1-\gamma)\sum_{k=1}^{\infty}\gamma^{k-1}r^k|\hat{s}=s\right], \tag{2.14}$$

ここで、$\gamma \in [0,1)$は割引係数、γ^{k-1}は（k−1）番目のコストの割引係数である。式(2.15)により長期的にコスト関数を最小化になれる最適なポリシーΦ^*は次の式で表示できる。

$$\Phi^* = arg\min_{\Phi}\mathrm{V}(s, \Phi) \quad \forall s \in \mathcal{S}, \tag{2.15}$$

この式の解は長期的に割引係数があるシングルエージェント MDP の最適な制御と考えられる。同時に、式（2.16）の解はベルマン方程式の解と同等であり、次式で与えられる。

$$\mathrm{V}(s^k) = \min_{\Phi(s^k)}\left[(1-\gamma)r^k + \gamma\sum_{\forall s^{k+1}\in\mathcal{S}}\Pr\{s^{k+1}|s^k, \Phi(s^k)\}\cdot\mathrm{V}(s^{k+1})\right], \tag{2.16}$$

この式は反復代用の方法で解けるが、移動位置、チャネル状態、エネルギーキュー状態、およびパケットの生起に関する統計データを事前に知るのは必要である。しかし、災害後動的な環境における、事前にこの知識を得るのは不可能であるため、本研究では、深層強化学習を用いた最適な無線アクセス制御法を提案する。

3 深層強化学習に基づく提案手法

最初に、強化学習の３要素である、状態(state)、行動 (action)と報酬 (reward)／コスト(cost)について説明する。

状態$s \in \mathcal{S}$とは、試行錯誤を繰り返して、学習を行う主体エージェントが観測できる「環境」の要素を表す集合である。エージェント自身が観測を行う場合や、外部から観測情報を受け取る場合もある。

行動$a \in \mathcal{A}$とは、エージェントが環境に対して働きかける行動 a の集合である。どんな行動を取ったら環境にとって何の変化が生じるか、ということは試行錯誤の中で学習できる。

報酬（またコスト）については、ある状態においてエージェントが行動を起こした結果、その結果を評価する関数である。強化学習手法の報酬／コストは直前でもらえる報酬／コスト(即時報酬・即時コスト)だけではなく、将来に渡る報酬／コスト(遅延報酬・遅延コスト)も合計する。単純な二つ報酬／コストの和を取ると、無限ステップの行動で獲得の報酬／コストが無限に発散してしまうため、割引率$\gamma \in (0,1)$(discount factor)を導入して、以下の式(3.1)で割引報酬和／コスト和を考える。

$$R_k = \sum_{\tau=0}^{\infty}\gamma^{\tau}r_k = r_k + \gamma r_{k+1} + \gamma^2 r_{k+2}\cdots, \tag{3.1}$$

ここで、r_kは即時報酬／コスト、残りの部分は遅延報酬／コストである。割引報酬和／コスト和には未来の不確実性を報酬／コストから割り引いた意味と考えられる。割引率が小さいと短期的な収益、大きいと長期的な収益を重視することになる。強化学習手法は、この報酬（またコスト）に関連した目的関数を最大化（また最小化）する方策を学習する問題に帰着するため、学習した方策の良さを示す行動価値関数が式(3.2)で表す。

$$\mathrm{Q}(s_k, a_k) = E[R_k|s_k = s, a_k = a], \tag{3.2}$$

本研究では、前述のように最適な無線アクセスポリシーΦを見つけ出す問題をベルマン方程式に定式化した。Q ラーニングには、環境sの定義は説明した状態s^kと同じであり、行動aは無線アクセスポリシーΦの行動

ペア(T^k, E^k)と定義される。また、本研究には割引コスト和の最小化を目的にし、即時コスト関数を式(2.11)のように定義される。

Q ラーニングにおける、式(3.2)の行動価値関数は Q 値また Q 関数とも呼ばれ、更新では下式(3.3)を用いる。

$$\mathrm{Q}(\mathrm{s}^k,(\mathrm{T}^k,E^k)) \leftarrow \mathrm{Q}(\mathrm{s}^k,(\mathrm{T}^k,E^k)) + \alpha^k\left(r^k + \gamma min\, E[\mathrm{Q}(\mathrm{s}^{k+1},(\mathrm{T}^{k+1},E^{k+1}))] - \mathrm{Q}(\mathrm{s}^k,(\mathrm{T}^k,E^k))\right), \tag{3.3}$$

ここで、$\alpha^k \epsilon [0,1)$は時間とともにどんどん減少する学習率である。本研究の Q 値はアクセス制御のコストを表すため、ターゲットとしての$r^k + \gamma min\, E[\mathrm{Q}(\mathrm{s}^{k+1},(\mathrm{T}^{k+1},E^{k+1}))]$部分は Greedy 方策で一番大きいではなく、最も小さい Q 値の行動を選択する。

しかし、状態行動空間が大きくなると、Q テーブルは超高次元になり、行動価値関数の計算時間が非常に長くなり、現実に応用できない、次元の呪いと呼ばれる問題が存在する。この問題を解決したのは深層強化学習アルゴリズムの一つである DQN (Deep Q Network)となる。DQN とは、行動価値関数をニューラルネットワークで近似し、強化学習を行う手法である

本研究では、逆伝播による確率的勾配降下法 (stochastic gradient descent)を用いて、ニューラルネットワークを更新する。確率的勾配降下法とは、訓練テストに対してコストが最小になるように、モデルパラメータを少しずつ操作し、モデルを訓練テストに対して適合したパラメータに収束させる方法である。

具体的には、DQN の目的関数は、行動価値関数をニューラルネットワークで近似し、損失関数を表すと以下の式で定義される。

$$\begin{aligned}\mathrm{L}(\theta^{k+1}) = E_{\{s_k,(T^k,E^k),r_k,s_{k+1}\}\in\tilde{\varphi}}\Big[&((1-\gamma)r_k \\ &+ \gamma Q(s_{k+1}, \arg\min_{(T^{k+1},E^{k+1})} Q(s_{k+1},(T^{k+1},E^{k+1}),\tilde{\theta}^k),\theta^k) \\ &- Q(s_k,(T^k,E^k),\theta^{k+1}))^2\Big],\end{aligned} \tag{3.4}$$

ただし、集合$\tilde{\varphi}$はメモリに保存した各ステップの経験 (experience) である。損失関数の最小化により最適方策Φ^*が得られる。

本研究では環境に適応する最適な無線アクセス制御を実現するため、深層強化学習を用いた手法を提案する。状態空間s_kから行動(T^k, E^k)を選択できるため、深層強化学習の Q 値の更新は$Q(s_k,(T^k,E^k))$にニューラルネットワークの重みθ^kをつけ、次式で与えられる。

$$Q(s_k,(T^k,E^k),\theta^k) \leftarrow Q(s_k,(T^k,E^k),\theta^k) + \alpha\left\{\begin{matrix}(1-\gamma)r_k \\ +\gamma \min\limits_{(T^{k+1},E^{k+1})} Q\left(s_{k+1},(T^{k+1},E^{k+1}),\tilde{\theta}^k\right) - Q(s_k,(T^k,E^k),\theta^{k+1})\end{matrix}\right\} \tag{3.5}$$

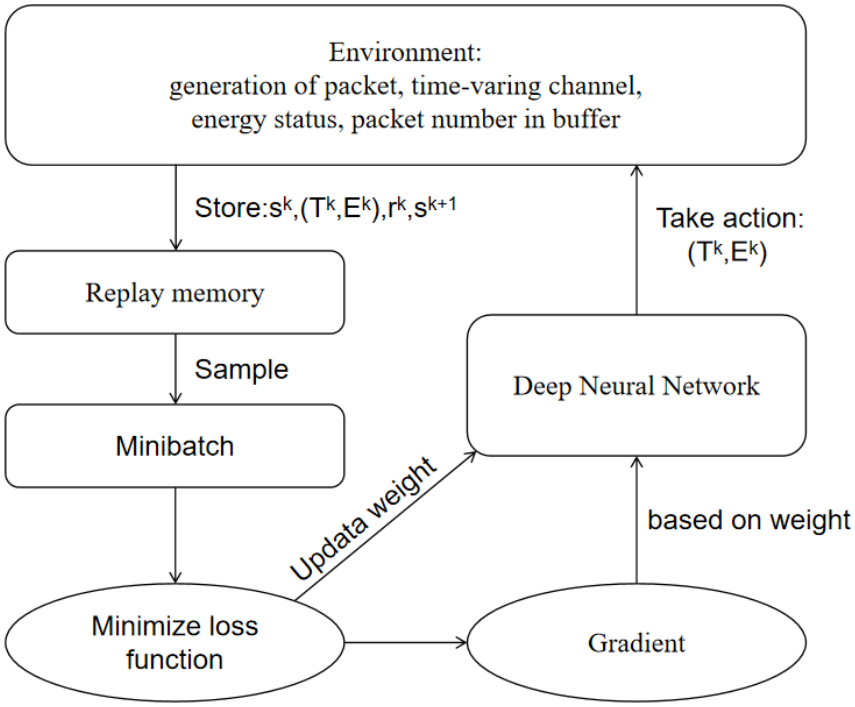


図 2　深層強化学習を用いた提案手法

提案した深層強化学習のモデルは図２を示す。深層強化学習がリプレイメモリを用いて、過去のs_k、(T^k, E^k)、r_k、s^{k+1}を保存する。リプレイメモリ$\varphi=\{m^{k-U+1}, \cdots, m^k\}$から、$\tilde{\varphi} \in \varphi$をサンプルして損失関数(3.4)を用いて勾配を計算し、ニューラルネットワークの重みθ^{k+1}を更新する。学習を安定させるため、メインのネットワークとは別の Target network を作り、ネットワーク更新時に次の状態における行動はメインネットワークで決めるが、その行動の価値評価は別のネットワークで行う。Target Network の重みは$\tilde{\theta}^k$となり、$\tilde{\theta}^k$は一定の頻度でθ^{k+1}から更新する。式(3.4)の計算結果が収束したら、ニューラルネットワークが環境に適応する最適な無線アクセスポリシーの学習を終わる。また、損失関数(3.4)の勾配は式(3.6)で与えられる。

$$\begin{aligned}\nabla_{\theta^{k+1}} \mathrm{L}(\theta^{k+1}) = E_{\{s_k,(T^k,E^k),r_k,s_{k+1}\}\in\tilde{\varphi}}\Big[&((1-\gamma)r_k \\ &+\gamma Q(s_{k+1}, \arg\min_{(T^{k+1},E^{k+1})} Q(s_{k+1},(T^{k+1},E^{k+1}),\tilde{\theta}^k),\theta^k) \\ &-Q(s_k,(T^k,E^k),\theta^{k+1}))\nabla_{\theta^{k+1}} Q(s_k,(T^k,E^k),\theta^{k+1})\Big]\end{aligned} \tag{3.6}$$

4 シミュレーション評価

4-1 シミュレーション設定

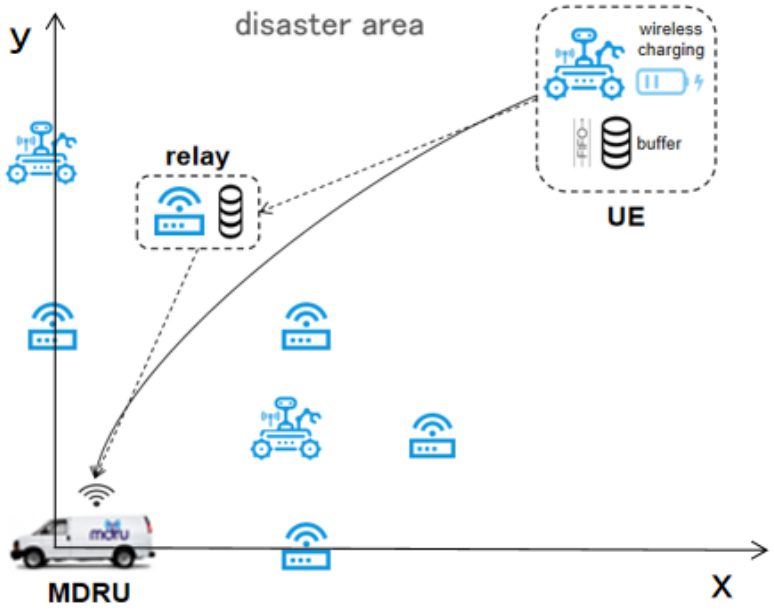


図 3　シミュレーションシナリオ

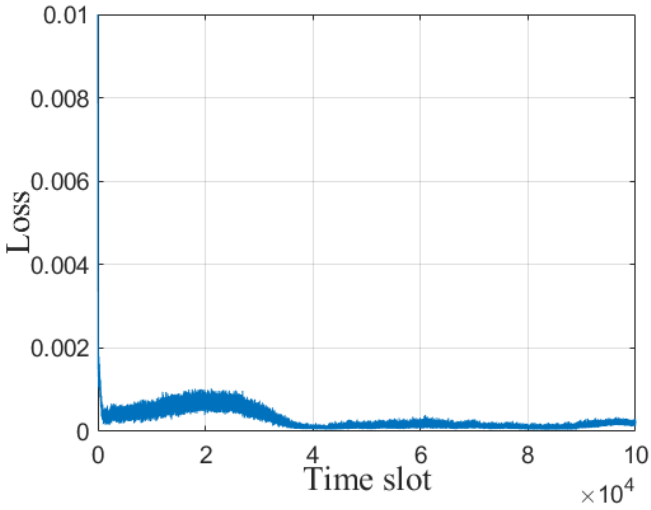
図３に示すような 500m×500m の災害対応ネットワークを考える。この災害対応ネットワークにおける、ゲートウェイとして一つの MDRU と MDRU のサービス範囲を拡大する六つの relay を配置されることを想定する。

チャネル状態は五つの可能な値に量子化され、シャドーイングにより分散σ_n^2は 6[dB]と設定する。UE がマンハッタングリッド移動モデルに基づき、10m/s の一定速度で災害エリア内で移動する。UE の位置変化により、チャネル利得の変化範囲が大きくなるため、32 個の可能な値に量子化される。また、シナリオ２における、二種類の UE 分布を考慮する。一つ目は一様分布であり、平均送信待ちパケット数τ_nは{1.25, 0.75, 1, 0.95, 0.95, 0.85}と設定する。二つ目は、人々の避難行動による特定な地域に人が集まる UE の集中分布である。この場合、平均送信待ちパケット数τ_nは{1.25, 0.75, 1, 10, 10, 0.85}に設定する。UE のバッファサイズ$\bar{J}$を0,1,2の三つのケースを考慮する。$\bar{J}=0$の場合は低遅延リアルタイム向けのネットワークアプリケーション、$\bar{J}=1,2$は遅延耐性ネットワークアプリケーションと考えられる。二つのネットワークに対して、コスト関数の重みをそれぞれ$\bar{J}=0$の$\rho=1,\xi=0$と$\bar{J}=\{1,2\}$の$\rho=5,\xi=0.6$と設定する。また、リプレイメモリーサイズを 10000、ミニバッチサイズを 512 に設定する。Target Network のパラメータは 150 タイムスロットごどに更新する。

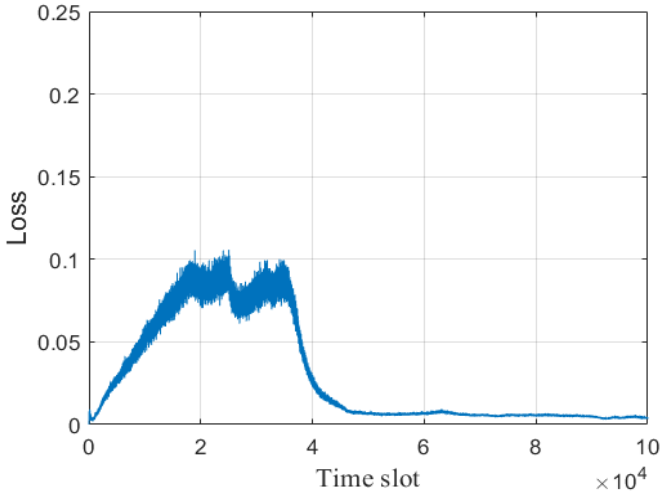
4-2 シミュレーション結果

まず、損失関数の結果を用いて、提案手法の収束性を検証する。ここで、ニューラルネットワークのニューロン数 512 と隠れ層層数 1 に設定した。

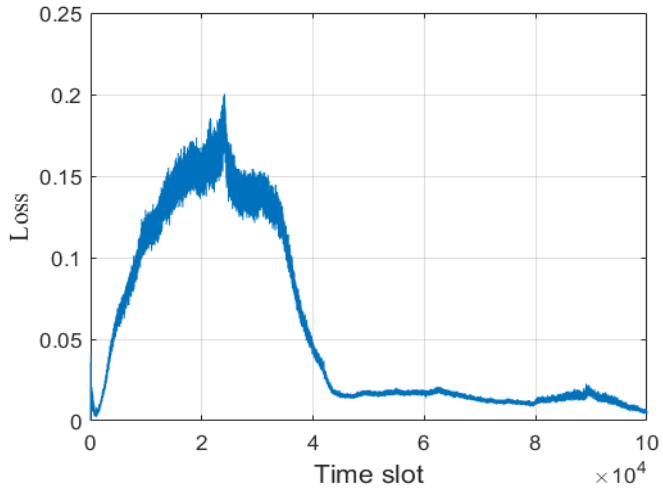
エネルギーハーベスト率$\lambda_e=0.3$とパケット生起率$\lambda_p=0.4$の場合の収束特性を図 4 に示す。結果により、図 4(a)に示したバッファなしの場合は、およそ 40000 タイムスロットで収束できることを確認した。また、バッファありの場合は、図 4(b)と図 4(c)に示したように収束する前に損失関数の値は大幅な上昇したが、およそ 42000 タイムスロットのとき収束できることを確認した。$\bar{J}=2$は$\bar{J}=1$より loss 上昇の幅が大きい、収束もう少し遅いことが分かった。



（a）　$\bar{J} = 0$



（b）　$\bar{J} = 1$



（c）　$\bar{J} = 2$

図 4　シナリオ 2 における収束特性

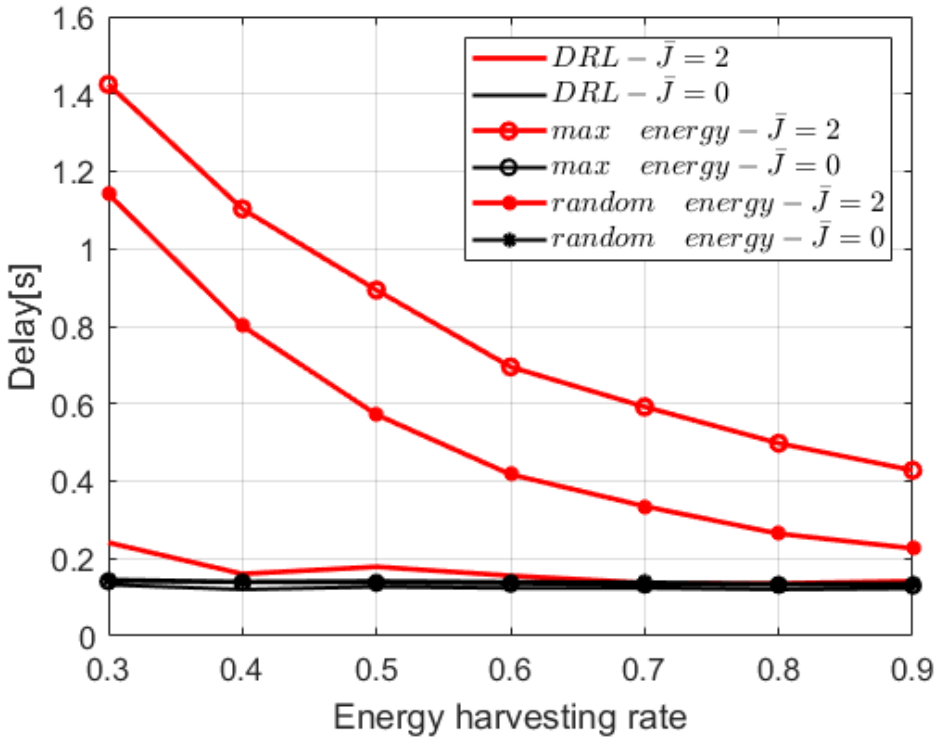
提案手法の性能を評価するため、提案手法は以下の 2 つのベースライン手法と比較する。

- Max energy：パケットが生起すると、UE は最大エネルギーを用いて一番近い MDRU/relay に送信する。
- Random energy：パケットが生起すると、UE は送信エネルギーをランダムに割り当て一番近い MDRU/relay に送信する。

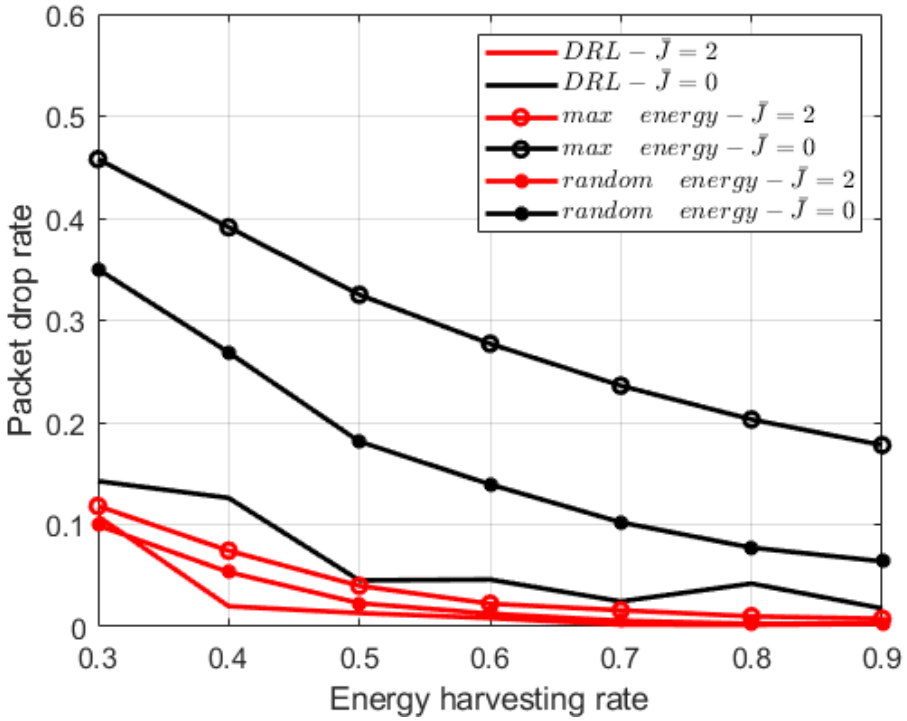
UE が一様分布と集中分布の比較結果をそれぞれ図 5、6 に示す。パケット生起率λ_p＝0.7 となる。

まず、UE が一様分布の比較結果を図 5 に示す。バッファなしの場合、エネルギーハーベスト率に関わらず、三つの手法は同じ遅延が達成した。パケットドロップ率に関しては三つの手法ともエネルギーハーベスト率の増加により顕著的に減少する。また、提案手法のパケットドロップ率は二つのベースライン手法と比べ、大幅に低減することが分かった。バッファありの場合、エネルギーハーベスト率の増加によって、三つの手法とも平均遅延時間とパケットドロップ率を大幅に改善できることを分かる。また、ベースライン手法と比べ、提案手法は明らかに良いパフォーマンスが達成した。

UE が集中分布の比較結果は図 6 に示す。バッファありとバッファなしの場合の結果とも一様分布とほぼ同じ傾向であることが確認できる。ただ、バッファありの場合において、ベースライン手法の遅延は一様分布より劣化したが、提案手法が同等な性能を維持した。

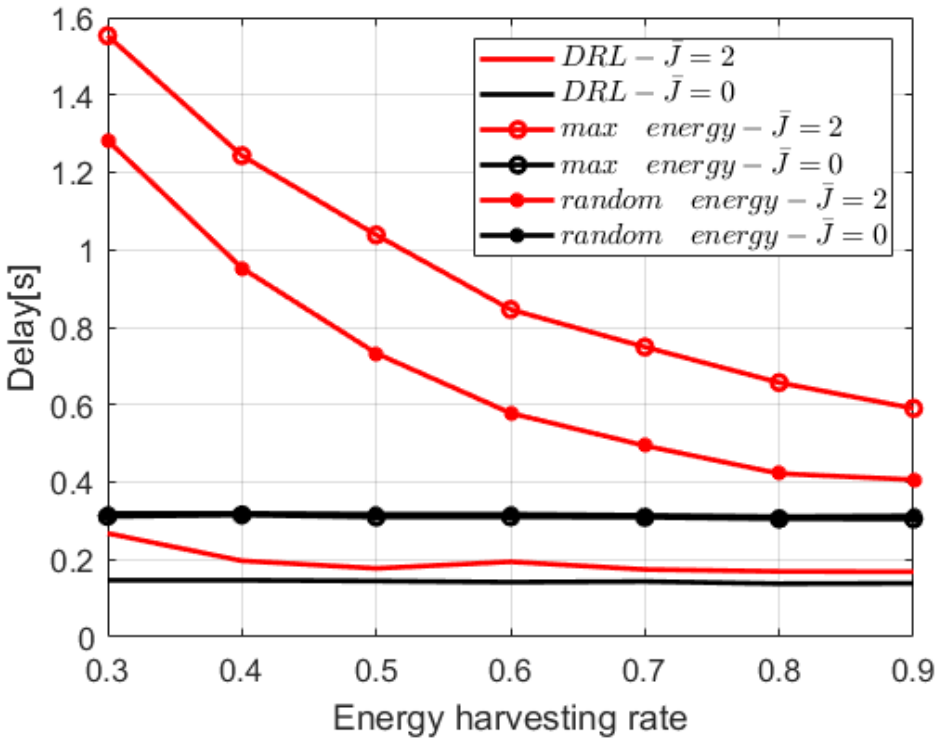


（a） 平均遅延

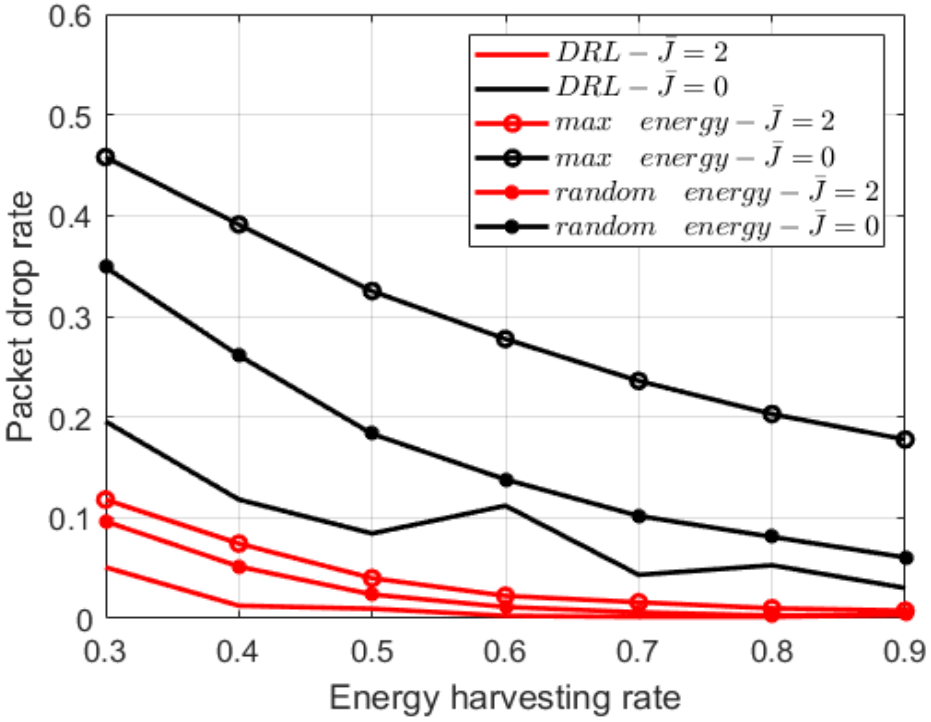


（b）　パケットドロップ率

図 5　UE 一様分布の場合異なる手法の比較



（a）　平均遅延



（b）　パケットドロップ率

図 22 UE 集中分布の場合異なる手法の比較

5 結論

大規模災害が発生した後、通信インフラの損害による通信サービスが途絶する危険がある。通信サービスを確保するため、災害地域に MDRU と relay を配置することにより、災害用エネルギーハーベスティング型無線ネットワークの構築が広く検討されていた。このような災害対応ネットワークにおける様々なユーザー無線機が通信サービスを利用できる。本論文では、災害エリアの環境変化を監視する固定と移動 UE を考慮し、UE がセンシングされた情報を MDRU に転送し、このデータパケットの遅延とドロップ率の最小化は本研究の目的である。動的な環境に適応できる、UE の最適なパケット処理方式と送信エネルギーを制御するため、本研究には深層強化学習を用いた無線アクセス制御方式を提案した。提案手法の性能を評価するため、シミュレーションを行い、ベースライン手法と比べ、提案手法の優位性が確認できた。

【参考文献】

[1] X. Wang, F. Jiang, L. Zhong, Y. Ji, S. Yamada, K. Takano and G Xue, "Intelligent Post-Disaster Networking by Exploiting Crowd Big Data," IEEE Network, 2020.

[2] T. Sakano et al., ”Disaster-resilient networking: a new vision based on movable and deployable resource units,” IEEE Network, vol. 27, no. 4,pp. 40-46, July-August 2013.

[3] T. Sakano, S. Kotabe, T. Komukai, T. Kumagai, Y. Shimizu, A. Takahara, T. Ngo, Z. M. Fadlullah, H. Nishiyama, and N. Kato, “Bringing Movable and Deployable Networks to Disaster Areas: Development and Field Test of MDRU,” IEEE Network, vol. 30, pp. 86–91, Jan. 2016.

[4] Q. T. Minh, K. Nguyen, C. Borcea and S. Yamada, "On-the-fly establish-ment of multihop wireless access networks for disaster recovery," in IEEE Communications Magazine, vol. 52, no. 10, pp. 60-66, October 2014.

[5] N. Chaudhuri and I. Bose, “Application of image analytics for disaster response in smart cities,” in Hawaii International Conference on System Sciences (HICSS), 2019, pp. 1–10.
[6] J. Yang and S. Ulukus, “Optimal task scheduling in an energy harvest-ing communication system,” IEEE Trans. Commun., vol. 60, no. 1, pp.220–230, Jan. 2012.
[7] H. Zhou, Y. Ji, X. Wang and S. Yamada, ”eICIC Configuration Algo-rithm with Service Scalability in Heterogeneous Cellular Networks,” inIEEE/ACM Transactions on Networking, vol. 25, no. 1, pp. 520-535, Feb.2017.
[8] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G.Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S.Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D.Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, Feb.2015.
[9] 小田部悟士, 小向哲郎, 清水芳孝, 東條 弘 ,“移動式 ICT ユニットによる耐災害 ICT の実現”, 電子情報通信学会論文誌 C Vol.J100-C No.3 pp.141-148
[10] 渡邊博之, 成田祐一, 大山勝徳, 加瀬澤正, “モバイル端末を活用した災害時最短避難経路提示システムの開発”, 情報処理学会論文誌, vol 53, N0.7 , pp.1768-1773(2012).
[11] 川原 圭博, 塚田恵 佑, 浅見徹 ,“放送通信用電波からのエネルギーハーベストに関する定量調査”,情報処理学会論文誌, Vil.51 , No.3, pp.824-834(Mar. 2010).

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Wireless Access Control in Edge-Aided Disaster Response: A Deep Reinforcement Learning-based Approach	IEEE Access	2021 年 3 月
Reinforcement Learning for Joint Channel/Subframe Selection of LTE in the Unlicensed Spectrum	Wireless Communications and Mobile Computing	2021 年 6 月
Deep Reinforcement Learning based Usage Aware Spectrum Access Scheme	International Symposium on Wireless Personal Multimedia Communications	2021 年 12 月
Reinforcement Learning based Joint Channel/Subframe Selection Scheme for Fair LTE-WiFi Coexistence	International Conference on Mobility, Sensing and Networking	2020 年 12 月
Deep Reinforcement Learning based Access Control for Disaster Response Networks	IEEE Global Communications Conference	2020 年 12 月
セルラーネットワークにおける深層強化学習を用いたアナログビームフォーミング制御方式の検討	電子情報通信学会スマート無線研究会	2022 年 1 月
コグニティブ無線ネットワークにおける深層強化学習を用いたセカンダリユーザー送信電力制御の一検討	電子情報通信学会スマート無線研究会	2022 年 1 月
深層強化学習を用いたスペクトルアクセス制御法の提案	電子情報通信学会短距離無線通信研究会	2021 年 11 月
LTE と WiFi の公平な共存のための強化学習型チャネル/サブフレーム選択手法	電子情報通信学会スマート無線研究会	2021 年 1 月
環境に適応できる深層強化学習を用いた無線アクセス制御法	電子情報通信学会スマート無線研究会	2021 年 1 月