Power Beacon Deployment Optimization based on Deep Q-Learning for Wireless Energy Transfer

Abellard Christley Mainviel

***Abstract*—The efficient placement of Power Beacons (PBs) is critical for optimizing energy harvesting in Internet of Things (IoT) networks. This study investigates various optimization techniques with a focus on Deep Q-Learning (DQN) ; other technique such as Particle Swarm Optimization (PSO), and Genetic Algorithms (GA) are used to enhance the energy collection by IoT devices from PBs and compare with DQN. An environment simulation is created to model the energy harvesting process, considering factors such as transmission power, channel characteristics, and the spatial distribution of PBs and IoT devices.**

The DQN approach employs a neural network to learn the optimal PB positions through interaction with the environment. PSO and GA, on the other hand, use population-based heuristic methods to explore and exploit the search space for optimal configurations. The performance of these algorithms is compared against random positioning strategies.

Experimental results demonstrate that the DQN, PSO, and GA methods significantly improve the total collected energy compared to random placement. DQN converges to a stable solution after several episodes, while PSO and GA effectively explore the search space to find high-fitness solutions. This comparative analysis highlights the effectiveness of these advanced optimization techniques in enhancing IoT energy harvesting through optimal PB placement.

I. Introduction

Wireless Energy Transfer (WET) is a transformative and sustainable technology, crucial to the growing Internet-of-Things (IoT) ecosystem. This innovation allows IoT devices to recharge wirelessly, addressing the significant challenge of battery life extension in low-cost devices and, in some cases, eliminating the need for traditional batteries altogether. This is especially important in large-scale deployments or extreme environments where regular maintenance is impractical or impossible.[1]

Moreover, WET leverages renewable energy sources such as solar panels, wind turbines, and radio-frequency (RF) signals. Solar panels are eco-friendly and scalable, though their effectiveness depends on weather and space availability, and they have high installation costs and environmental impacts. Wind turbines offer clean energy and efficient space use but face issues like inconsistent power generation, noise pollution, and high initial costs. RF signals, abundant from sources like cellular networks, provide a novel energy harvesting method particularly useful for IoT devices in hard-to-reach areas. However, RF energy density is low, making it suitable only for low-power needs, and its efficiency is affected by distance and environmental barriers.[2]

In conclusion, while solar and wind energy remain significant for large-scale power production, RF energy harvesting is particularly advantageous for small, wireless applications. This method enhances energy diversification and sustainability, especially for emerging technologies and remote applications. RF energy harvesting ensures a continuous power supply to IoT devices, enabling their deployment in previously inaccessible areas. Consequently, WET is set to expand IoT applications, supporting innovations in smart cities, environmental monitoring, and healthcare, and ushering in a new era of connected, intelligent, and sustainable technologies.[2]

WET, while promising, faces a myriad of challenges, particularly when it is implemented across diverse and large-scale IoT environments. One of the primary issues it encounters is the efficient and strategic distribution of energy. This involves carefully planning the placement of Power Beacons (PBs) throughout the network. However, this placement strategy must be carefully crafted to adhere to certain energy outage constraints, which significantly adds to the complexity of the overall deployment strategy. It's through such relentless efforts and exploration that the full potential of WET can be harnessed effectively in the ever-evolving IoT ecosystem. [2].

A. Related Works

Research (WET) focuses on optimizing power harvesting and system coverage. [1] discusses the strategic deployment of power beacons, emphasizing the efficiency of clustered arrangements in energy transfer. Using mathematical modeling, simulations, and heuristic approaches, the study evaluates various deployment scenarios to enhance network performance. Similarly, [3] explores power beacon placement, highlighting non-uniform deployment strategies and their potential to improve network efficiency using the Sequential Convex Approximation (SCA) method while Mozaffari et al.[4] optimized the 3D deployment of UAV-BS to maximize coverage area with minimum transmit power, considering user mobility and varying QoS requirements.

[5] expands on large-scale WET implementations, crucial for powering IoT devices in the forthcoming 6G era. This research stresses the importance of innovative WET approaches to meet future wireless network demands. Key optimization methods include Energy Beamforming (EB), which requires precise Channel State Information (CSI), and various resource scheduling techniques to manage network performance. Hybrid schemes integrating different EB strategies and CSI-limited/free approaches are also discussed, addressing scenarios with challenging CSI acquisition.

[6] examines energy source deployment strategies in cellular networks to ensure consistent energy supply for mobile communications. The study evaluates three main optimization methods: energy harvesting from full-duplex base stations, symmetrically deployed power beacons with isotropic and directed modes, and a hybrid approach combining power beacons with distributed antenna elements (DAEs). These methods are assessed based on their impact on Signal-to-Noise Ratio (SNR) outage probability and spectral efficiency.Additionally, research on UAV-based aerial base stations (UAV-BS) using (DQN) highlights the application of DQN to optimize the 3D deployment of aerial-BSs for capacity enhancement, considering dynamic network topology and user distribution [7]. [2] addresses optimal power beacon placement in large IoT networks, utilizing optimization techniques and nature-inspired algorithms to enhance network lifetime and sensor endurance, and improve coverage with multiple antennas, thereby reducing energy outage probability. In recent years, various studies have proposed different algorithms for aerial-BS placement. For instance, Lyu et al. developed a polynomial-time algorithm for optimizing UAV-BS placement to enhance user coverage while minimizing transmit power. Alzenad et al. provided a framework for evaluating the three-dimensional location of UAV-BS to maximize coverage using minimal power [8], [9]. Each study contributes significantly to advancing WET in Wireless-Powered Communication Networks (WPCNs) push the boundaries of current knowledge and pave the way for future innovations in that field.

B. Main Goal

This proposal tackles the strategic placement of Power Beacons (PBs) in Wireless Energy Transfer (WET) systems for IoT networks. The main challenge is efficiently distributing energy to IoT devices amidst network complexities, dynamic environmental conditions, and resource limitations. The goal is to maximize energy harvesting and ensure uninterrupted operation of IoT devices, regardless of location. Our scalable and adaptable solution leverages Deep Q-learning to continuously optimize PB placement, improving WET system efficiency and reliability in IoT applications.

Deep Q-learning, a reinforcement learning technique using deep neural networks, effectively handles complex, high-dimensional state spaces. It is particularly useful in wireless communications for optimizing WET in extensive IoT networks. Applying Deep Q-learning to PB deployment enhances energy distribution efficiency, maximizing average incident power and minimizing energy outages across IoT networks, even under changing conditions.[2]

We aim to implement a Deep Q-learning algorithm to optimize PB positioning in WET systems, improving energy distribution in IoT networks. The model is expected to outperform traditional methods by adaptively learning optimal PB placements, maximizing energy coverage, and minimizing outages in dynamic network conditions. This will significantly advance WET system efficiency and reliability for IoT applications, ensuring sustainable and continuous energy supply.

C. Methodology

To investigate whether Deep Q-learning can effectively optimize the positioning of PBs in WET systems, we propose a comprehensive methodology. Initially, we will define the problem and set parameters, considering the IoT network layout and key WET factors. secondly, we built a kind of dataset with previously learned data to train the model. Then, we will develop a Deep Q-learning model, defining state and action spaces, and a reward function focused on maximizing energy coverage. The model will be integrated into a simulated IoT network environment for training, using diverse scenarios to enhance its adaptability. We plan to rigorously evaluate the model against traditional methods like random positioning , particle swarm optimisation or genetic algorithm , focusing on metrics like total energy harvested or coverage. This will be followed by testing the model under various hypothetical scenarios to assess its robustness and make necessary adjustments. The final step involves a detailed analysis of the simulation data to validate our hypothesis and a comprehensive report outlining the methodology, results, and potential areas for further research. This methodology aims to blend detailed simulation-based testing with robust analysis, ensuring a thorough exploration of Deep Q-learning's capabilities in optimizing PB placements in WET systems.

II. SYSTEM MODEL

In this system, PBs are used to provide energy to IoT devices distributed in a 2D plane. The objective is to optimize the placement of these PBs to maximize the total collected energy by the IoT devices. This optimization is achieved using a Deep Q-Learning algorithm.

A. IoT Devices and PBs

The environment is a 2D plane with dimensions 10×10 units. K IoT devices are randomly positioned within this plane. m PBs are also randomly positioned within the same plane.

B. State Representation

The state of the environment is represented by the concatenated positions of all PBs and IoT devices. The state vector S includes the x and y coordinates of each PB and IoT device, denoted as [7]:

$$S = \{(x_{pb_1}, y_{pb_1}, x_{pb_2}, y_{pb_2}, \ldots, x_{pb_m}, y_{pb_m}, x_{iot_1}, y_{iot_1}, \ldots, x_{iot_K}, y_{iot_K})\}$$

C. Action Space

The action space consists of possible movements for each PB in the 2D plane. Possible actions include moving up, down, left, right, or staying in the current position [7]:

$$A = \{\text{up}, \text{down}, \text{left}, \text{right}, \text{stay}\}$$

D. Reward Function

The reward is defined as the total energy collected by all IoT devices from the PBs. The goal is to maximize the total collected energy, which is influenced by the distance between the PBs and the IoT devices.

E. Path Loss Model

The path loss model characterizes the signal attenuation between PBs and IoT devices, which affects the power received by the IoT devices. The path loss β_k is calculated as:

$$\beta_{m,k} = \left(\frac{\lambda}{4\pi d_{m,k}}\right)^2 \text{ [10]}$$

where λ is the wavelength, and d_k is the distance between a PB and an IoT device.

F. Channel Model

The channel model includes both line-of-sight (LoS) and non-line-of-sight (NLoS) components, modeled using a Rice fading channel. The received power P_k at an IoT device from a PB is given by:

$$P_k = \beta_{m,k} \left| \sum_{n=1}^{N} h_k \psi_k \right|^2 \text{ [10]}$$

where h_k represents the channel gain, ψ_k represents the transmitted signal, and N is the number of antenna elements.

G. Energy Harvesting Model

The energy harvested by an IoT device from the received power is modeled using a sigmoid function:

$$E_k = \frac{\mu}{1 + \exp(-a(P_k - b))} - \frac{\mu\Omega}{1 - \Omega} \text{[10]}$$

where μ is the energy harvesting efficiency, a and b are parameters of the sigmoid function, and Ω is a pre-calculated constant.

H. Optimization Objective

The optimization objective is to adjust the positions of the PBs to maximize the total energy collected by the IoT devices. This is formulated as:

$$\max \sum_{k=1}^{K} E_k$$

III. DQL

In this section, we introduce a 2D position planning algorithm (that can be seen in algorithm 1) for (PBs) that uses (DQN) to optimize the total energy harvested by IoT devices. DQN integrates deep learning through Convolutional Neural Networks (CNN) and Q-learning, a type of reinforcement learning. Within this framework, the agent (PBs) adapts and refines its placement strategy based on the feedback (rewards or penalties) it receives from the environment, thereby enhancing its performance over time. The reinforcement learning model is characterized by the ¡A, S, R¿ tuple: Action (A) represents the set of possible moves for the PBs, State (S) denotes the positions of the PBs and IoT devices, Reward (R) quantifies the effectiveness of the current placement in terms of energy collection. the proposed dqn architecture is proposed in 1

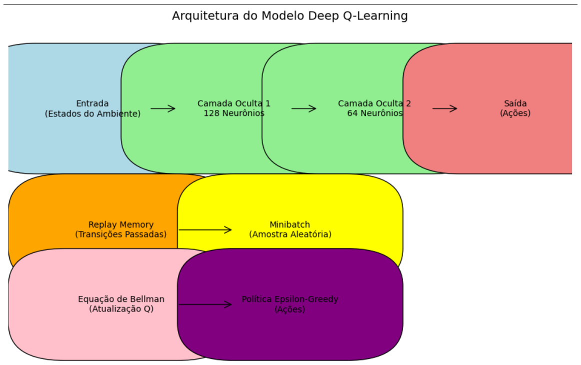

Figure 1: Environment Layout with random positions of 4 Power Beacons (PBs) and 20 IoT devices. The color intensity represents the energy harvested by each IoT device.

IV. SIMULATION RESULTS

in our simulation, we consider a 2D area of 10 units by 10 units where IoT devices are randomly distributed; we deploy (PBs) within this area and maximize the energy harvested by the IoT devices. To validate our work we compare the energy harvested to other methods of positioning . This strategy aligns with the goal of PBs to cater to areas with fluctuating energy demands.

The parameters used in the simulation are presented in Table I.

Parameter	**Value**
Number of IoT Devices (K)	20
Number of Beacons (m)	2, 4, 6
Number of Antenna Elements (N)	8
Transmission Power (P_T)	2 W
Operating Frequency (f)	915 MHz
Speed of Light (c)	3×10^8 m/s
Path Loss Exponent (α)	2.5
Rice Factor (κ)	1.5
Energy Harvesting Efficiency (μ)	0.01
Sigmoid Function Parameter (a)	0.23
Sigmoid Function Parameter (b)	5.36

Table I: Parameters Used

Figure 2 shows the random positions of 4 Power Beacons (PBs) for 20 IoT devices. The layout of the environment is visualized, where IoT devices are represented by purple circles and PBs are represented by red triangles. The color intensity of the IoT devices indicates the amount of energy harvested.

in Figure 3, we compare our algorithm with other methods . we can observe that the DQN algorithm- significantly outperforms the other methods across the range of the number of PBs. Specifically, the total collected energy increases with the number of PBs when using DQN, demonstrating its efficiency in learning optimal positions. In contrast, the PSO, Random Positions, and Genetic Algorithm show relatively static performance, indicating their limitations in adapting to the increasing number of PBs. This highlights the superiority of reinforcement learning

Algorithm 1 DQN Network and Benchmark Comparison

State size, action size, learning rate
Optimal DQN and performance comparison
Initialize replay memory D to capacity N
Initialize action-value function Q with random weights θ
Initialize target action-value function $\hat{Q}$ with weights $\theta^- = \theta$

episode = 1 to N Initialize the number of users and distribution of users in sequence s
Initialize 2D position coordinates of the PBs in sequence s and preprocessed sequence $\phi_1 = \phi(s_1)$

$t = 1$ to T choose an action based on epsilon-greedy policy
Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D
Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D

$$\text{Set } y_j = \begin{cases} r_j & \text{if ends at } j+1 \\ r_j + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases}$$

Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ with respect to the network parameters θ
every C steps reset $\hat{Q} = Q$
each number of PBs in $\{2, 4, 6\}$ Initialize environment with K IoT devices and m PBs
Initialize DQN agent with state size and action size
episode = 1 to 200 Reset environment and get initial state
Initialize total reward to 0
step = 1 to n_step Choose action using epsilon-greedy policy
Execute action and observe next state, reward, and done flag
Perform learning step with current state, action, reward, and next state
Update current state to next state
Accumulate total reward
Log episode duration and total reward
Check for convergence criteria
Store maximum reward obtained
Run PSO algorithm and store best fitness
Run random positions algorithm and store mean energy
Run Genetic Algorithm and store best fitness
Plot and compare results of DQN, PSO, Random Positions, and Genetic Algorithm

approaches, particularly DQN, in dynamic and complex environments.

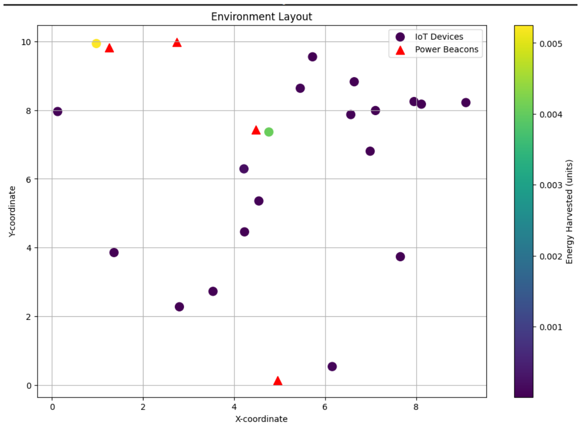

Figure 2: Environment Layout with random positions of 4 Power Beacons (PBs) and 20 IoT devices. The color intensity represents the energy harvested by each IoT device.

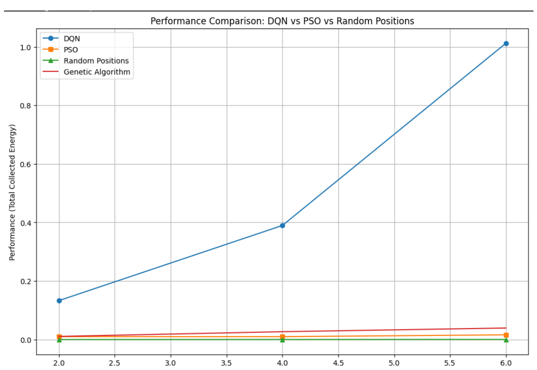

Figure 3: Performance Comparison: DQN vs PSO vs Random Positions. The plot shows the total collected energy as a function of the number of PBs for different algorithms.

V. Conclusion

In this work , we study the way to optimise the position os PBs to maximise the energy colected by iot devices. first of all we define the environment then we define our deep reinforcement learning algorithm DQN to otimise the position then we compare it to other methods. The results shows that our algorithm is having a lot better perfomance than the others.Future research could extend this study by exploring multi-agent reinforcement learning (MARL) for distributed PB optimization, integrating renewable energy sources to ensure a sustainable energy supply, and developing real-time adaptive systems for dynamic PB positioning based on environmental changes. Additionally, addressing security and privacy concerns, investigating hybrid beamforming techniques, and optimizing machine learning models such as Proximal Policy Optimization (PPO) or Soft Actor-Critic (SAC) could enhance system performance. Scalability to large-scale IoT networks, improving energy harvesting efficiency through advanced rectenna designs, and cross-layer optimization considering physical, MAC, and network layers are also promising areas for further investigation. These efforts will contribute to more efficient and sustainable wireless energy transfer systems, paving the way for advanced IoT applications. [1]

References

[1] S. Zhou, H. Dai, H. Sun, G. Tan, and B. Ye, "On the deployment of clustered power beacons in random wireless powered communication," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2424–2438, 2023.

[2] O. M. Rosabal, O. L. A. López, H. Alves, S. Montejo-Sánchez, and M. Latva-Aho, "On the optimal deployment of power beacons for massive wireless energy transfer," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 531–10 542, 2021.

[3] K. Liang, L. Zhao, G. Zheng, and H.-H. Chen, "Non-uniform deployment of power beacons in wireless powered communication networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1887–1899, 2019.

[4] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1207–1210, 2016.

[5] O. L. A. López, H. Alves, R. D. Souza, S. Montejo-Sánchez, E. M. G. Fernández, and M. Latva-Aho, "Massive wireless energy transfer: Enabling sustainable iot toward 6g era," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8816–8835, 2021.

[6] H. Tabassum and E. Hossain, "On the deployment of energy sources in wireless-powered cellular networks," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3391–3404, 2015.

[7] J. Guo, Y. Huo, X. Shi, J. Wu, P. Yu, L. Feng, and W. Li, "3d aerial vehicle base station (uav-bs) position planning based on deep q-learning for capacity enhancement of users with different qos requirements," in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, 2019, pp. 1508–1512.

[8] J. Lyu, K. Huang, and S. Zhou, "Placement optimization of uav-mounted mobile base stations," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 1390–1403, 2017.

[9] M. Alzenad, M. Mozaffari, H. Yanikomeroglu, A. El-Keyi, and M. Youssef, "3-d placement of an unmanned aerial vehicle base station (uav-bs) for energy-efficient maximal coverage," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 1339–1351, 2017.

[10] V. D. P. Souto, O. M. Rosabal, S. Montejo-Sánchez, O. L. A. López, R. D. Souza, and H. Alves, "Otimização do tempo de recarga com transmissão de energia sem fio usando beamforming analógico baseado nas estatísticas do canal," in *XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2023)*. São José dos Campos, SP, Brasil: SBrT, October 2023.

[1] link do projeto :https://github.com/MacUpr/tp-558/tree/main