A Deep Reinforcement Learning based Online Charging Scheme for Target Coverage and Connectivity in WRSNs

Bui Hong Ngoc

Advisor: Assoc. Prof. Do Phan Thuan

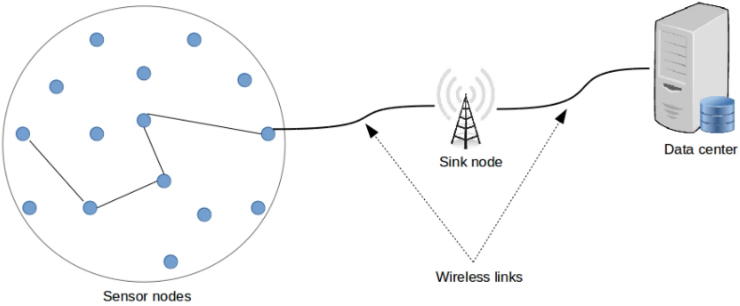
Hanoi, 2021

Hanoi University of Science and Technology

Outline

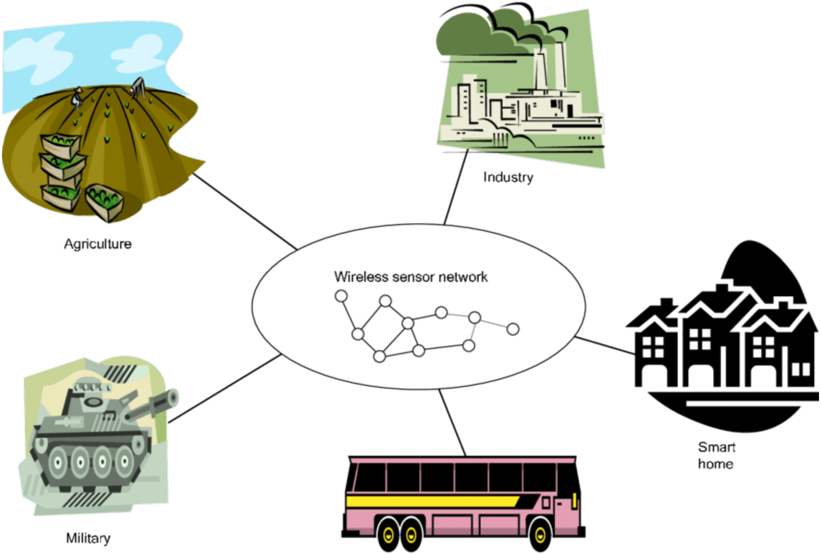
Outline

Wireless Sensor Networks (WSNs)



Source: Bahri (2018)

Wireless Sensor Networks - Applications



Source: Chen et al. (2010)

Fundamental question:

How to prolong the network lifetime?

Fundamental question:

How to prolong the network lifetime?

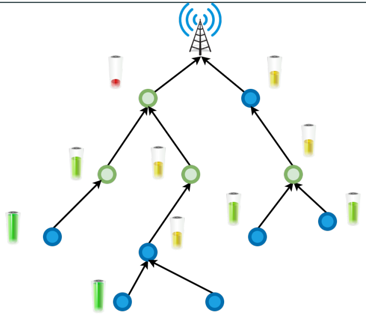
Non-rechargeable	Rechargeable

Wireless Sensor Networks (WSNs)

Fundamental question:

How to prolong the network lifetime?

Non-rechargeable	Rechargeable

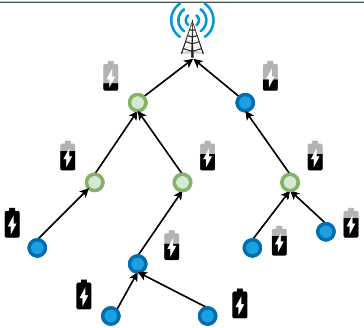


- Data reduction
- Routing protocol

Wireless Sensor Networks (WSNs)

Fundamental question:

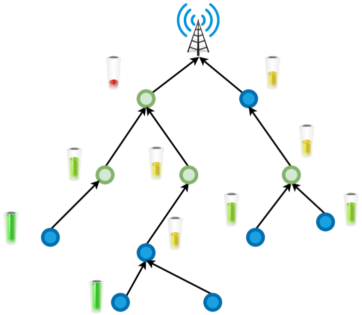
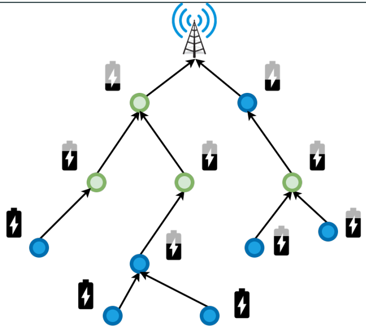
How to prolong the network lifetime?

Non-rechargeable	Rechargeable
	
	• Energy harvesting • Wireless charging

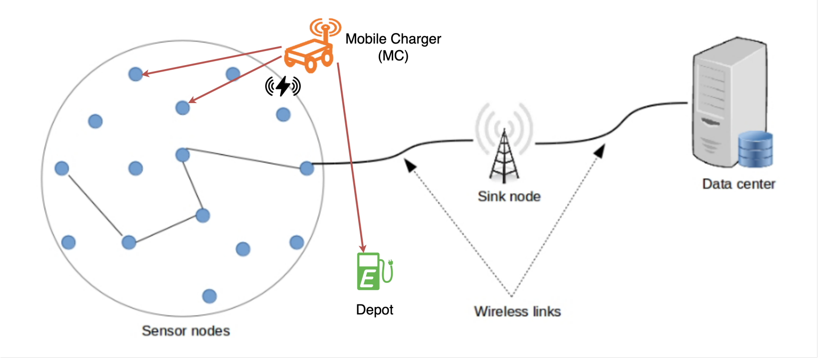
Wireless Sensor Networks (WSNs)

Fundamental question:

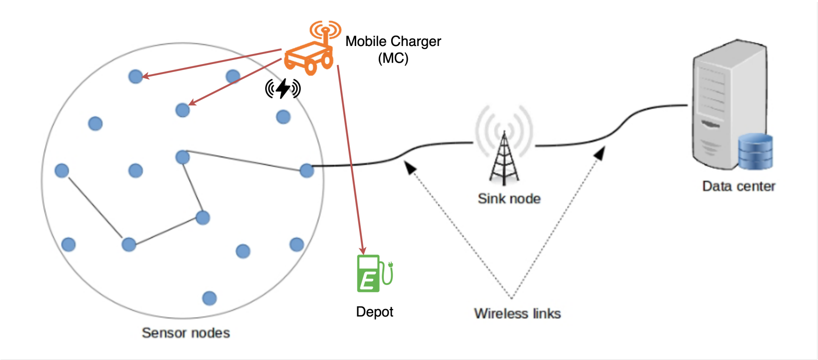
How to prolong the network lifetime?

Non-rechargeable	Rechargeable
	
• Data reduction • Routing protocol	• Energy harvesting • ⇒ Wireless charging

Wireless Rechargeable Sensor Networks (WRSNs)



Wireless Rechargeable Sensor Networks (WRSNs)



Fundamental question:

How to design an effective charging scheme?

Outline

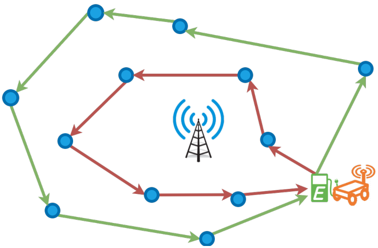
Related works

The current works on WRSNs are divided into two categories:

Periodic	On-demand
The charger moves along a predetermined charging path	MC moves and charges upon receiving requests from the sensors

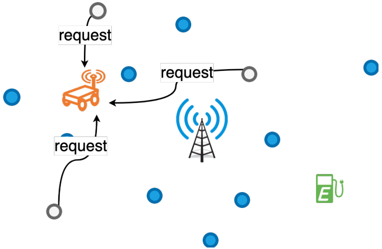
Related works

The current works on WRSNs are divided into two categories:

Periodic	On-demand
The charger moves along a predetermined charging path 	MC moves and charges upon receiving requests from the sensors

Related works

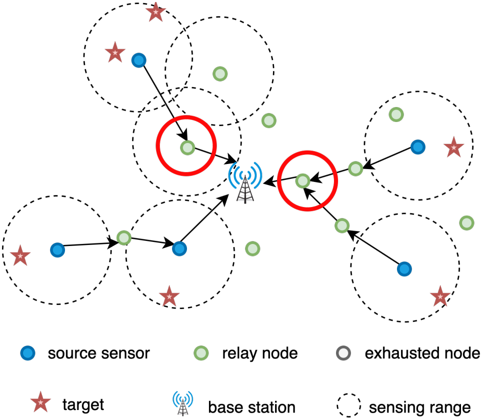
The current works on WRSNs are divided into two categories:

Periodic	**On-demand**
The charger moves along a predetermined charging path	MC moves and charges upon receiving requests from the sensors 

On-demand charging scheme

Study	Name	Algorithm
He et al. (2013)	NJNP	heuristic
Lin et al. (2019)	DWDP	heuristic
Fu et al. (2015)	ESync	TSP-based
Lin et al. (2017)	TSCA	heuristic
Kaswan et al. (2018)	GSA	gravitational search
Zhu et al. (2018)	INMA	heuristic
Cao et al. (2021)	RMP-RL	deep reinforcement learning
La et al. (2020)	Q-charging	Q-learning

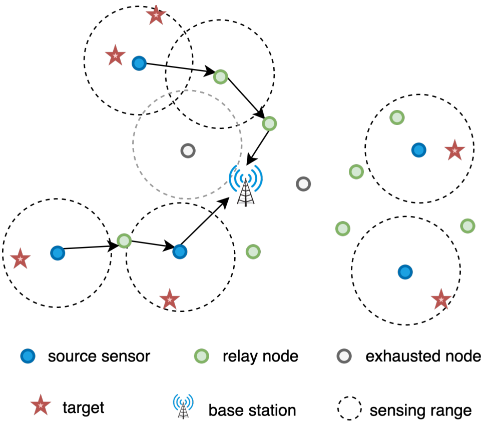
On-demand charging scheme - Drawback 1



Drawback 1

Most of the current approaches consider the role of sensors to be the same.

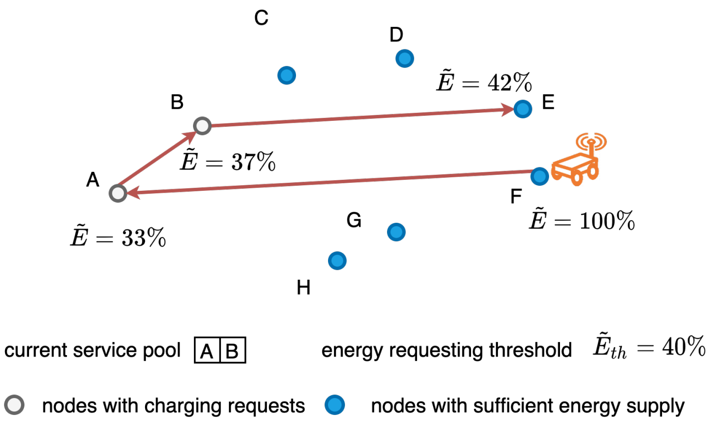
On-demand charging scheme - Drawback 1



Drawback 1

Most of the current approaches consider the role of sensors to be the same.

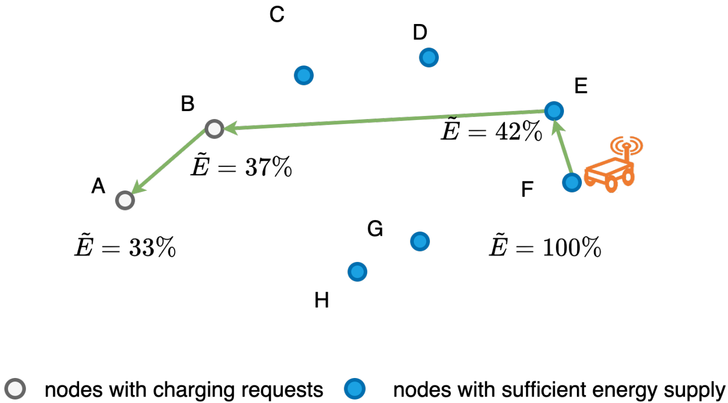
On-demand charging scheme - Drawback 2



Drawback 2

The performance of charging algorithms highly depends on the predefined energy threshold.

On-demand charging scheme - Drawback 2



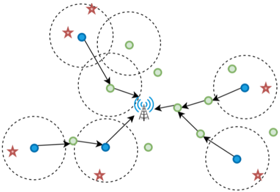
Drawback 2

The performance of charging algorithms highly depends on the predefined energy threshold.

Outline

Target coverage and connectivity in WRSNs

(Zhao and Gurusamy, 2008)



WRSN

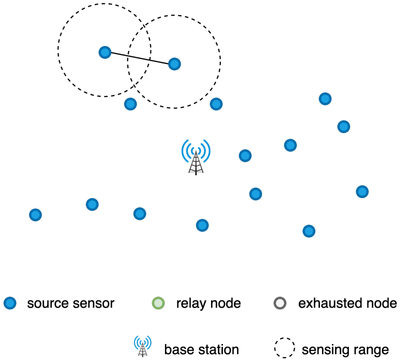
We consider the target coverage and connectivity problem (Zhao and Gurusamy, 2008) in the WRSNs' configuration.

Zhao, Qun, and Mohan Gurusamy. Lifetime maximization for connected target coverage in wireless sensor networks. *IEEE/ACM transactions on networking* 16.6 (2008): 1378-1391.

Target coverage and connectivity in WRSNs

Given

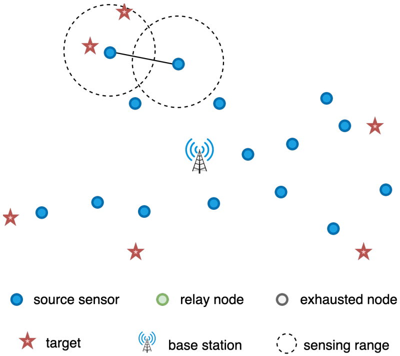
- 1 base station/sink
- n sensors



Target coverage and connectivity in WRSNs

Given

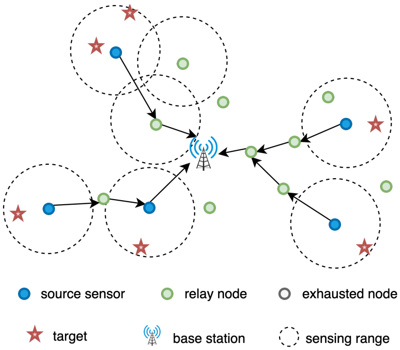
- 1 base station/sink
- n sensors
- m targets



Target coverage and connectivity in WRSNs

Given

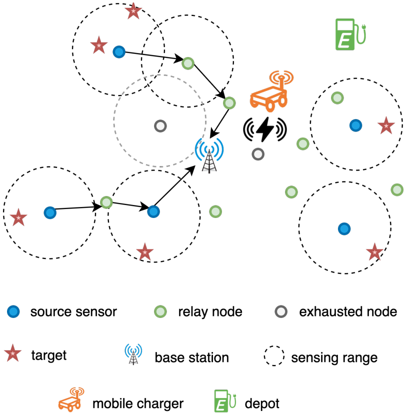
- 1 base station/sink
- n sensors
- m targets

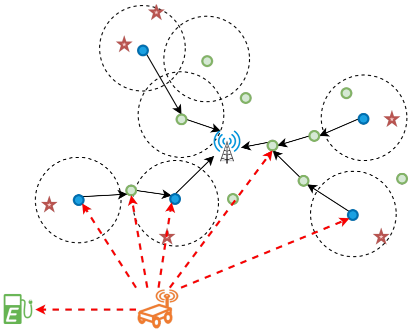


Target coverage and connectivity in WRSNs

Given

- 1 base station/sink
- n sensors
- m targets
- 1 mobile charger (MC)
- 1 depot





Objective

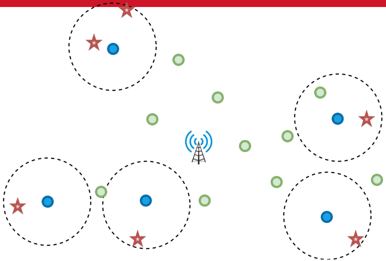
Designing a MC's charging strategy to maximize the network lifetime.

Target coverage and connectivity in WRSNs

Network's lifetime

Network's lifetime is the time interval from when the network starts till the target coverage or the connectivity is not satisfied.

Target coverage and connectivity in WRSNs

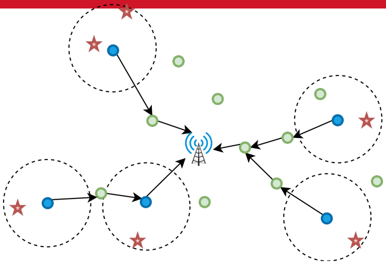


Network's lifetime

Network's lifetime is the time interval from when the network starts till the target coverage or the connectivity is not satisfied.

- *Coverage*: each target be covered by at least one sensor.

Target coverage and connectivity in WRSNs

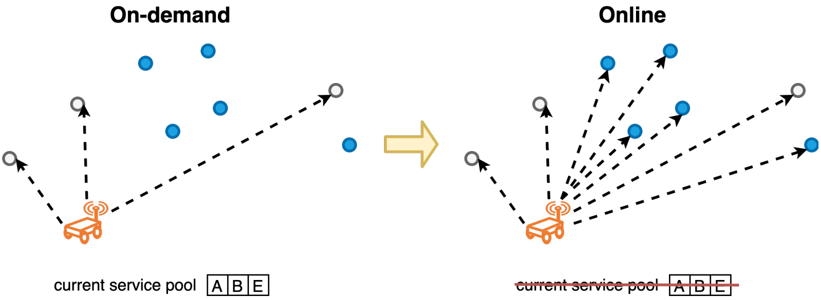


Network's lifetime

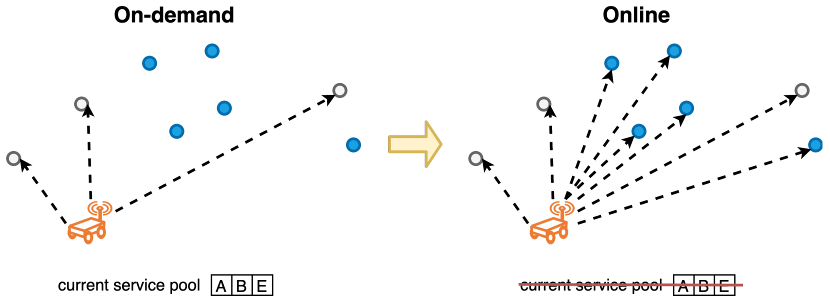
Network's lifetime is the time interval from when the network starts till the target coverage or the connectivity is not satisfied.

- *Coverage*: each target be covered by at least one sensor.
- *Connectivity*: from each source sensor to sink, there must exist at least one route traversing through only the active sensors.

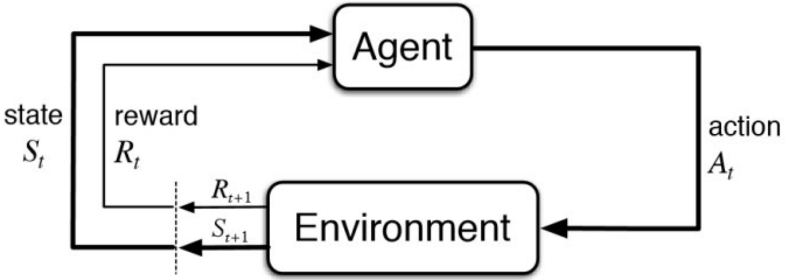
Outline



- Proposing a **novel online charging scheme** by omitting the energy requesting threshold.

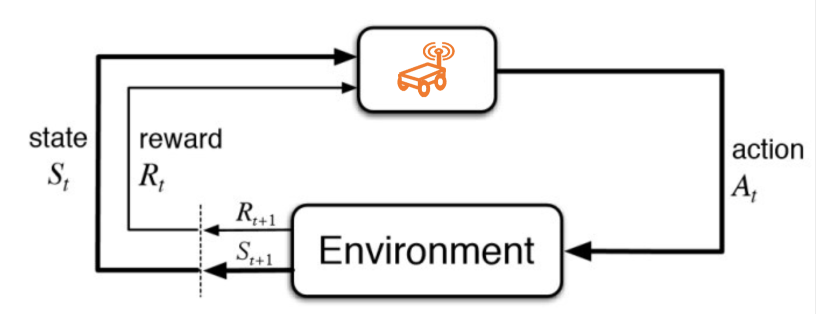


- Proposing a **novel online charging scheme** by omitting the energy requesting threshold.
- Using **deep reinforcement learning** to model a charging scheme.



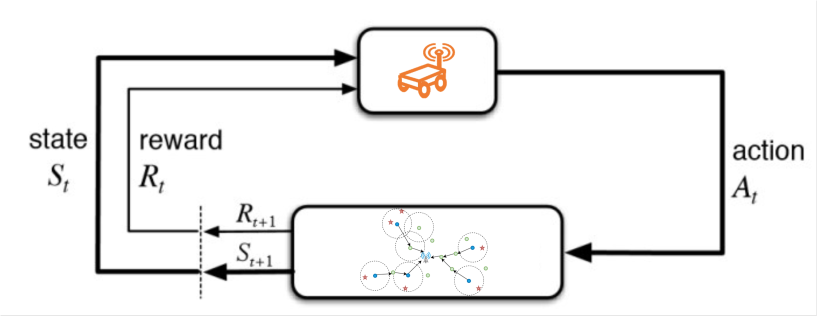
Given an agent interacting with an environment, reinforcement learning problem is to learn good strategy to maximize cumulative rewards.

Reinforcement learning



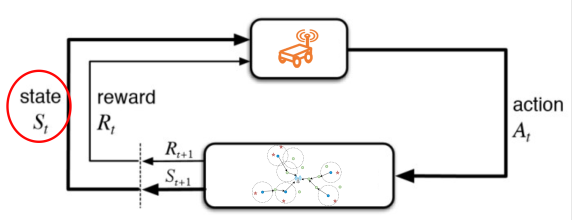
Agent $\Rightarrow$ Mobile charger

Reinforcement learning



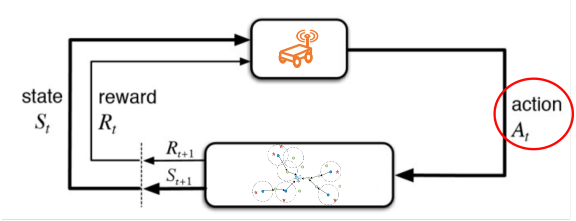
Environment ⇒ Wireless sensor network (WSN)

A deep reinforcement learning-based mobile charging scheme



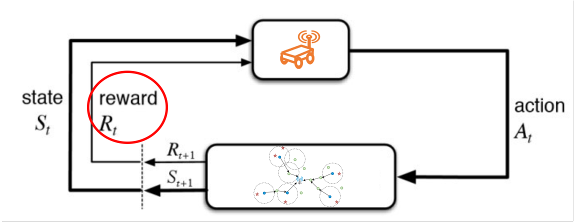
- **State ($\mathscr{S}$):** the current status of sensors and the MC itself (e.g. current position, current energy, current energy consumption rate, ...).

A deep reinforcement learning-based mobile charging scheme



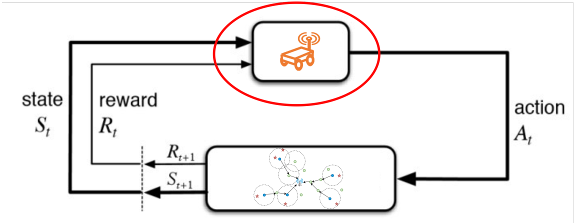
- **Action ($\mathscr{A}$):** a next charging destination. $a_t = i$ is to charge the i-th sensor and $a_t = 0$ is to go back to the depot and recharge itself.

A deep reinforcement learning-based mobile charging scheme

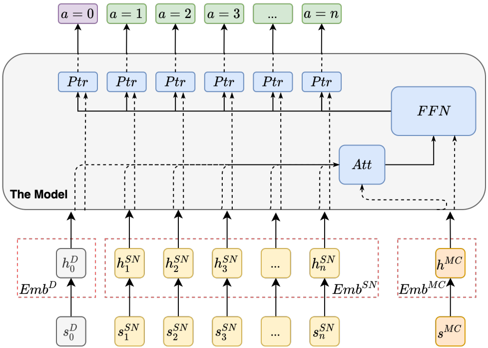


- **Reward function ($\mathscr{R}$):** $R(s,a) = t$ is a time interval of doing charging action.

The model



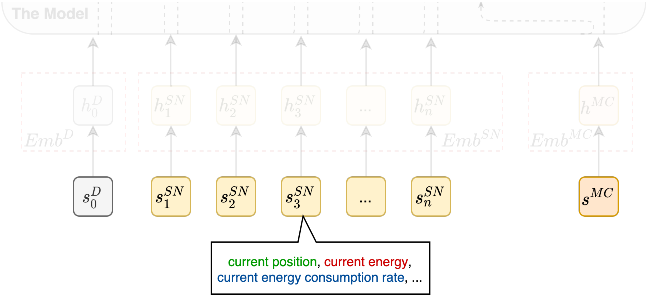
The model



Adapted from the model of André and Kevin (2020) proposed for Capacitated Vehicle Routing Problem (CVRP).

Hottung André and Tierney Kevin. Neural Large Neighborhood Search for the Capacitated Vehicle Routing Problem. *Frontiers in Artificial Intelligence and Applications*, 325(ECAI 2020), 443–450.

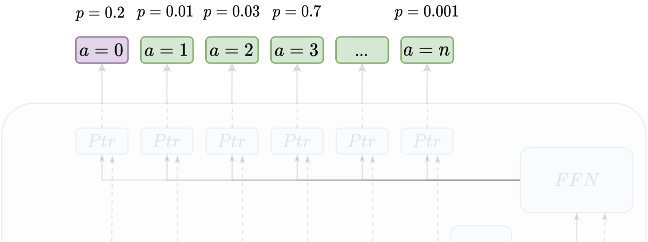
The model



Input:

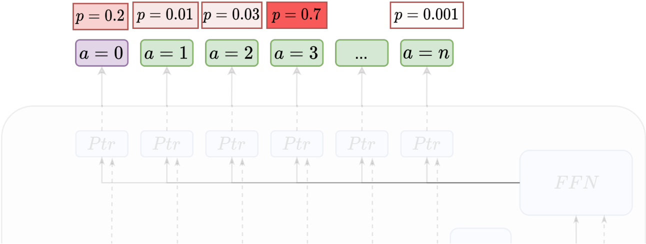
- s_0^D: state of the depot.
- s_i^{SN}: state of the i-th sensor.
- s^{MC}: state of the MC.

The model



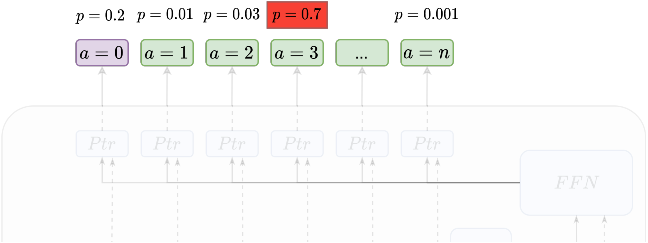
Output:

- The distribution over actions



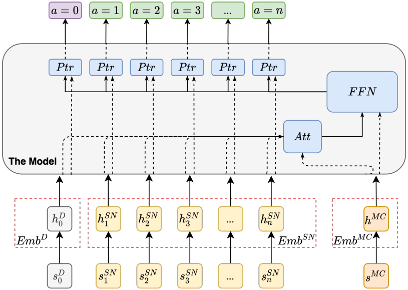
Training: The next action will be drawn based on the probability of each action.

The model



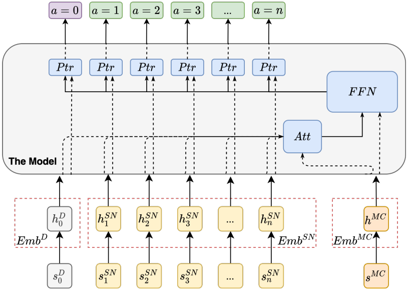
Testing: Action with highest probability will be selected.

The model - Remark



- The model complexity does not depend on input size n (thanks to attention and pointing mechanism).

The model - Remark



- The model complexity does not depend on input size n (thanks to attention and pointing mechanism).
- MC can be deployed on the fly (using FFN instead of GRU).

Our objective is to maximize the expected total reward:

$$J(\theta) = \mathbb{E}_{\tau \sim p_\theta(\tau)} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \tag{1}$$

where τ is the charging trajectory and γ is the discounted factor.

Training - Policy gradient method

Our objective is to maximize the expected total reward:

$$J(\theta) = \mathbb{E}_{\tau \sim p_\theta(\tau)} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \tag{1}$$

where τ is the charging trajectory and γ is the discounted factor. Applying the REINFORCE algorithm, the gradient is given by:

$$\nabla J(\theta) = \mathbb{E}_{\tau \sim p_\theta(\tau)} \left[\sum_{t=0}^{\infty} \nabla_\theta \log(\pi_\theta(a_t | s_t)) \hat{\mathscr{A}}_t^{GAE(\lambda)} + \beta \nabla_\theta \mathscr{H}(\pi_\theta(\cdot | s_t)) \right] \tag{2}$$

where $\mathscr{H}$ is entropy function, β is a hyperparameter controlling the strength of the regularization, and $\hat{\mathscr{A}}_t^{GAE(\lambda)}$ is the Generalized Advantage Estimated function.

Outline

Settings

Parameter	Value	Unit	Comment
$W \times H$	200×200	$m \times m$	sensor field
n	$20 \sim 30$	–	number of deployed sensors
m	$10 \sim 20$	–	number of critical targets
B_{MC}	500	J	battery capacity of the MC
ω_{move}	0.04	J/m	battery capacity of a sensor
v	5	m/s	velocity of the MC
B_s	10	J	battery capacity of a sensor
r_s	40	m	sensing range
r_c	80	m	communication range
μ	0.04	J/s	charging rate

Baselines

We mainly compare our proposed with three baselines:

- **Random**: The agent chooses the next charging destination at random.
- **NJNP** (He et al., 2013): chooses the spatially closest requesting node as the next charging destination.
- **INMA** (Zhu et al., 2018): Similar to NJNP but aims to minimize the invalid nodes caused by an action.
- **DRL-TCC**: The proposed method.

Performance Metrics

- **Network's lifetime**: the time interval from when the network starts till the target coverage or the connectivity is not satisfied.
- **Sustainability**: the ratio of the number of the network instances so that the mobile charger (MC) can elongate to a sustained state (still active and guarantee coverage and connectivity after a large number M of charging actions of the MC).
- **Travel distance**: accumulate travel distance of MC.

Research questions

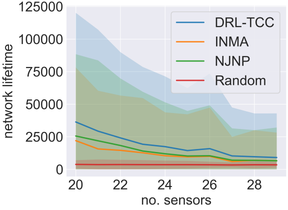
- **Scenario 1:** Evaluating the impacts of the number of sensors. Varying the number of sensors from 20 to 30.
- **Scenario 2:** Evaluating the impacts of the number of targets. Varying the number of targets from 10 to 20.
- **Scenario 3:** Evaluating the impacts of the packet generation probability.

Datasets

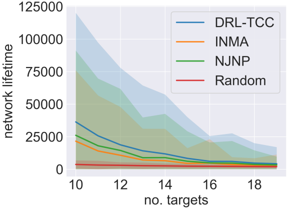
The positions of sensors are drawn in a square area $200m \times 200m$ according to the uniform distribution.

- Training set: 10000 network instances with 20 sensors and 10 targets.
- Testing set: 1000 network instances for each configuration.

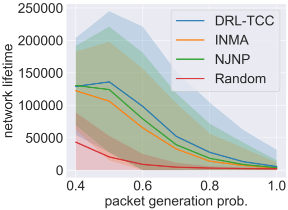
The network's lifetime



(a) Varying the number of sensors



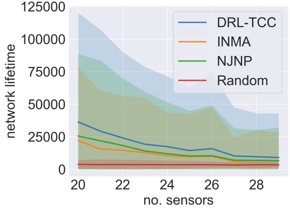
(b) Varying the number of targets



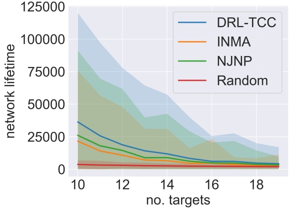
(c) Varying the packet generation prob.

DRL-TCC >> NJNP >> INMA >> Random.

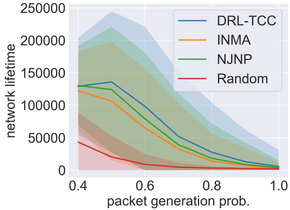
The network's lifetime



(a) Varying the number of sensors



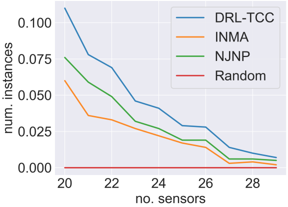
(b) Varying the number of targets



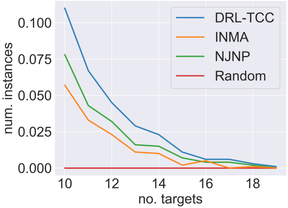
(c) Varying the packet generation prob.

High variance.

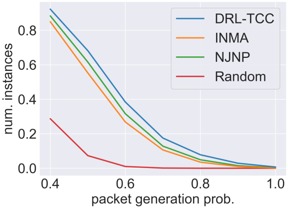
Sustainability



(a) Varying the number of sensors



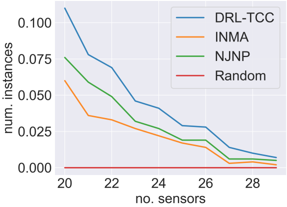
(b) Varying the number of targets



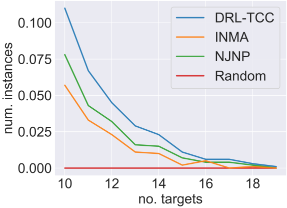
(c) Varying the packet generation prob.

Massive deterioration of all algorithms when increasing energy consumption.

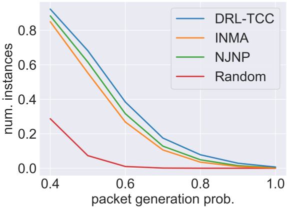
Sustainability



(a) Varying the number of sensors



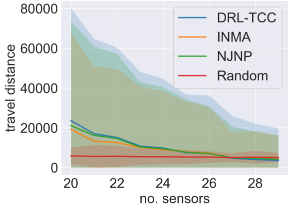
(b) Varying the number of targets



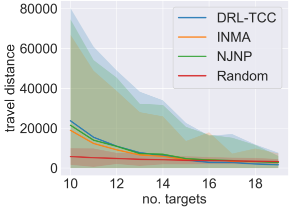
(c) Varying the packet generation prob.

At $p = 1$, DRL-TCC result is **twice** as that of NJNP and INMA (2.9% compared to 1.5% and 1%, respectively)

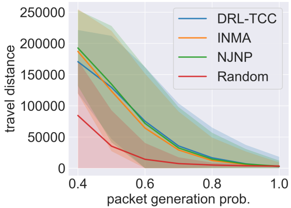
Travel distance



(a) Varying the number of sensors



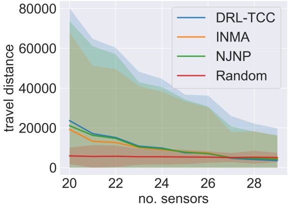
(b) Varying the number of targets



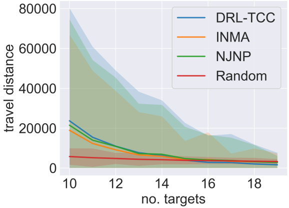
(c) Varying the packet generation prob.

DRL-TCC >> NJNP >> INMA >> Random.

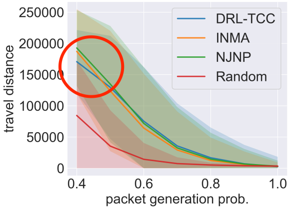
Travel distance



(a) Varying the number of sensors



(b) Varying the number of targets



(c) Varying the packet generation prob.

Moving less when the network is light ⇒ **Adaptability**

Outline

Conclusion

- We investigated the target coverage and target connectivity problem in WRSNs.

Conclusion

- We investigated the target coverage and target connectivity problem in WRSNs.
- We proposed a novel online charging schesme in which the requesting energy threshold is omitted.

Conclusion

- We investigated the target coverage and target connectivity problem in WRSNs.
- We proposed a novel online charging schesme in which the requesting energy threshold is omitted.
- We proposed DRL-TCC to tackle the target coverage and connectivity problem in WRSNs.

Conclusion

- We investigated the target coverage and target connectivity problem in WRSNs.
- We proposed a novel online charging schesme in which the requesting energy threshold is omitted.
- We proposed DRL-TCC to tackle the target coverage and connectivity problem in WRSNs.
- We conducted extensive experiments to demonstrate the superiority of our algorithm.

Future works

- Evaluating the algorithms in large and dense networks.

Future works

- Evaluating the algorithms in large and dense networks.
- Introducing *idle* state as one of the MC's actions. It will reduce unnecessary charging action.

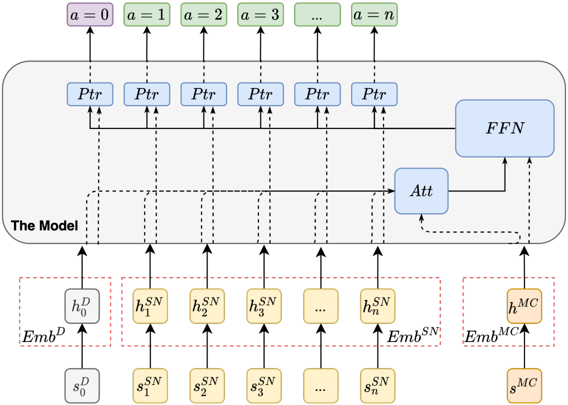
Future works

- Evaluating the algorithms in large and dense networks.
- Introducing *idle* state as one of the MC's actions. It will reduce unnecessary charging action.
- Using Graph neural network to embed the network's state.

Thank you for listening.

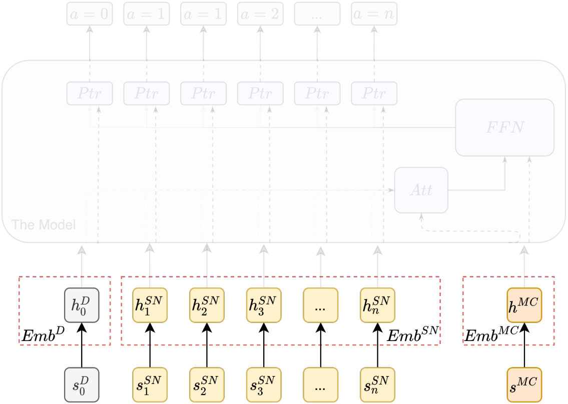
Question & Answer

The model



Adapted from the model of André and Kevin (2020) proposed for Capacitated Vehicle Routing Problem (CVRP).

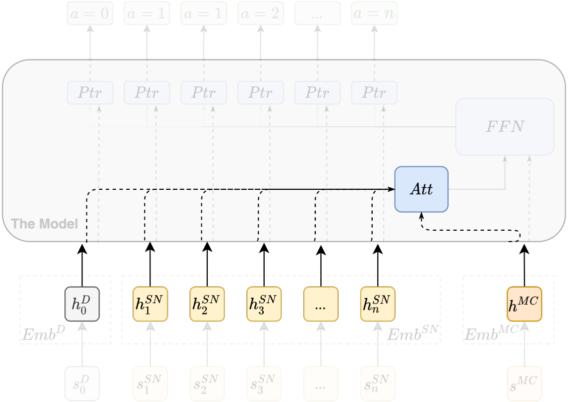
The model



Embedding state $s_i^{<\cdot>}$ to d_h-dim representation vector $h_i^{<\cdot>}$.

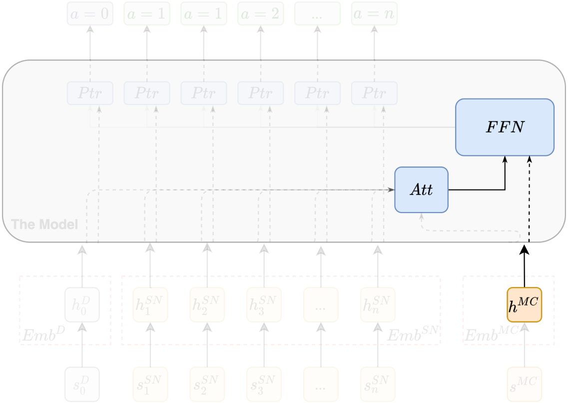
- s_0^D: state of the depot.
- s_i^{SN}: state of the i-th sensor.
- s^{MC}: state of the MC.

The model



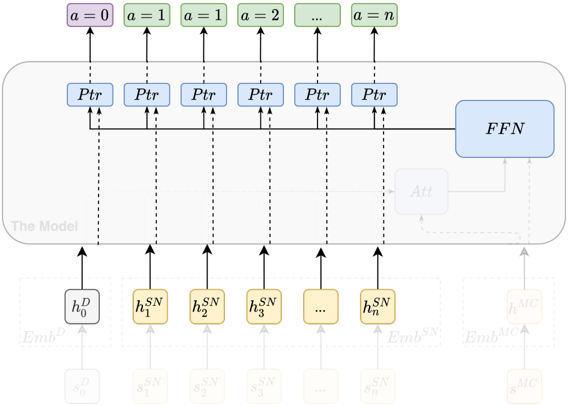
Using attention mechanism to compute the context vector c.

The model



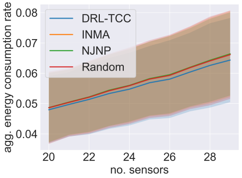
Two layers feed-forward network is used to produce a single, compatible vector q.

The model

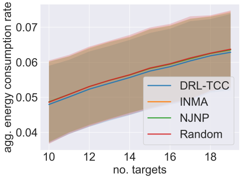


The pointing mechanism is leveraged to produce the distribution of the policy over all actions.

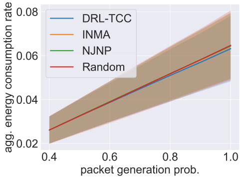
Discussion on the self-organizing capability



(a) Varying the number of sensors



(b) Varying the number of targets

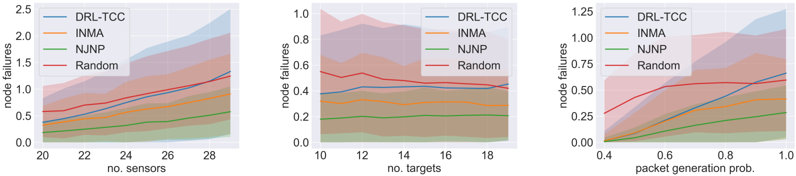


(c) Varying packet generation probability

The comparison of the aggregated energy consumption rate.

The aggregated energy consumption rate of DRL-TCC strategy is slightly lower than that of the others.

Discussion on the self-organizing capability



(a) Varying the number of sensors

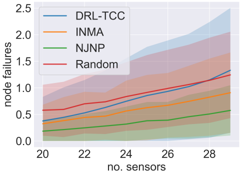
(b) Varying the number of targets

(c) Varying packet generation probability

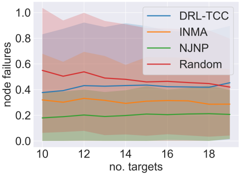
The comparison of the number of node failures.

- at $p = 0.4$, DRL-TCC ~ NJNP and INMA.
- p ↑, the number of node failures of DRL-TCC ↑, and the aggregated energy consumption rate ↓.

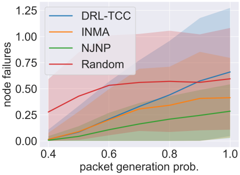
Discussion on the self-organizing capability



(a) Varying the number of sensors



(b) Varying the number of targets



(c) Varying packet generation probability

The comparison of the number of node failures.

It demonstrates the **adaptability** and **generalization** to various scenarios of the proposed model.