Electric Vehicle Routing for Emergency Power Supply with Deep Reinforcement Learning

Daisuke Kikuta  
NTT Corporation  
Tokyo, Japan  
daisuke.kikuta@ntt.com

Hiroki Ikeuchi  
NTT Corporation  
Tokyo, Japan  
hiroki.ikeuchi@ntt.com

Kengo Tajiri  
NTT Corporation  
Tokyo, Japan  
kengo.tajiri@ntt.com

Yuta Toyama  
NTT DOCOMO  
Tokyo, Japan  
yuuta.toyama.zg@nttdocomo.com

Masaki Nakamura  
NTT DOCOMO  
Tokyo, Japan  
masaki.nakamura.ua@nttdocomo.com

Yuusuke Nakano  
NTT Corporation  
Tokyo, Japan  
yuusuke.nakano@ntt.com

ABSTRACT
To maintain telecom services even during power outages, maintaining the power of the base stations is essential. Here, we consider a solution where Electric Vehicles (EVs) go around to directly supply their power to the base stations whose power is continuously decreasing. The goal is to find EV routes that minimize both total travel distance and the number of downed base stations. In this paper, we formulate this routing as a new variant of the Electric Vehicle Routing Problem (EVRP) and propose a solver that combines a rule-based vehicle selector and a reinforcement learning-based node selector. We evaluate our solver on synthetic datasets and real datasets. The results show that our solver outperforms baseline solutions in terms of the objective value and computational time. See https://ntt-diku.github.io/rl-evrpeps for details (full paper, code, visualization, etc).

KEYWORDS
Electric Vehicle Routing Problem; Emergency Power Supply; Deep Reinforcement Learning

1 INTRODUCTION
With natural disasters increasing [4], maintaining infrastructures during disasters is becoming more critical. As a telecom company, our company has a mission to maintain telecom services even during disasters caused by disasters. One of the most fundamental challenges here is maintaining the battery of telecoms base stations. Although each base station has a backup battery, maintaining the base station’s power over extended periods requires power supply from external sources.

This paper addresses this challenge by leveraging Electric Vehicles (EVs) as the external source: EVs go around to supply their power to the base stations directly. In contrast to existing approaches [3, 7–9], this approach works without Vehicle-to-Grid systems, whose installation cost is high, and is effective for the case where the number of EVs is less than that of base stations.

Here, we formulate the base station relief as a new variant of the Electric Vehicle Routing Problem (EVRP) [2], termed EVRP for Emergency Power Supply (EVRP-EPS). We also propose a solver that combines a rule-based vehicle selector and a reinforcement learning-based node selector. In contrast to existing EVRPs, EVP-R-EPS additionally considers the battery discharge of EVs, as well as mandatory details such as preparation/cleanup time and EV discharge limit. With reinforcement learning, our solver enables deriving reasonable routes within a short time.

2 PROBLEM SETTING
Objective. Given a time horizon \( T \) (i.e., expected blackout duration) and sets of base stations, charge stations, and EVs, the objective is to maintain as many base station batteries as possible during the time horizon while minimizing the total travel distance of all EVs. Formally, the objective function below is minimized.

\[
L = \sum_{k} \sum_{a=1}^{A_k} \frac{d(x_{nk(a)}, x_{nk(a+1)})}{N_{ev}} + \alpha \frac{1}{T} \int_{t=0}^{T} \sum_{i} I(b_{b, i} = 0) dt,
\]

where \( A_k \) is the number of \( k \)-th EV’s actions, \( d(\cdot, \cdot) \) is the distance between two points, \( n_k(a) \) is the index of node visited by \( k \)-th EV at \( a \)-th action, \( \alpha \) is the positive weighting factor, \( b_{b, i} \) is the \( i \)-th base station’s battery at the time \( t \), \( N_{bs/ev} \) is the number of base stations/EVs, and \( I(\cdot) \) is the Boolean indicator function.

Action Space and Sub-actions. EVs cycle through an action (move) and three subsequent sub-actions (prepare, discharge/charge, and clean-up). The action space here is to determine which node EVs move to from the current nodes. The sub-actions are deterministically and automatically conducted depending on the state when EVs arrive at base/charge stations.
where
\[ \text{Xfmr} \]
Therefore we employ a rule-based vehicle selector that always selects an EV finishing clean-up the soonest from the current time. In the action/sub-action cycle, EVs can start selecting an EV finishing clean-up the soonest from the current time.

The parameterized policy \( p_{\theta} \) is trained by REINFORCE [6] with a greedy rollout baseline [1] so that Eq. (1) is minimized. For the decoding (i.e., route generation), we employ the sample decoding.

### 4 EXPERIMENTS

#### Setups

We evaluate our solver on synthetic datasets (Syn-ev6, Syn-ev12) and real datasets (Real-ev6, Real-ev12). The baselines in the evaluation are two naive approaches (Greed and Rand) and a constraint programming solver on a time-space network (Tsn). Greed and Rand replace the node selector of our solver with a greedy node selection that selects a base station with the lowest battery and a random node selection, respectively. Tsn solves sub-problems divided by a heuristic clustering, meaning that solutions derived by Tsn are near-optimal. We set the time limit for solving EVRP-EPS to 30 minutes according to the actual requirements.

#### Results

Table 1 shows the evaluation results. We report the results of Tsn with two discrete time resolutions \( \Delta t = 0.5 \) h and those of our solver with different decoding: greedy (G) and sampling decoding (S=#samples). Overall, our solver consistently outperforms the baselines in terms of the objective value and computational time. In particular, S=12800 provides the minimum objective value in all cases. Regarding computational scalability, Tsn exhibits an exponential increase in computational time as the time horizon is doubled, whereas our solver restricts the escalation to a linear increment. Our solver also shows the scalability for the increase of the number of nodes and EVs (see the full paper), demonstrating the capability of handling larger-scale situations (i.e., a longer time horizon and more nodes/EVs) within a short time.

### 5 DISCUSSION

The experimental results reveal that our solver (reinforcement learning-based solver) is effective in solving the complicated EVRP within a limited time. On the other hand, some limitations remain, including the balance of travel distance among EVs and the estimation of travel time. We will address them by introducing a route-balancing strategy and considering actual travel distance and uncertainty of travel time due to traffic situations. In terms of unavailable roads due to a disaster, we may handle them by obtaining real-time road conditions from a provider and masking unavailable roads. We plan to validate the ideas in a demonstration experiment.
REFERENCES


