Tanzania Journal of Engineering and Technology 2024, **43**(1):102-114
OPEN ACCESS articles distributed under Creative Commons
Attribution Licence [CC BY-ND]
Websites: https://ajol.org/tjet; https://tjet.udsm.ac.tz

ISSN 1821-536X (**print**); ISSN 2619-8789 (**electronic**)
https://doi.org/10.52339/tjet.v43i1.966

Regular Research Manuscript

Optimal Load Shedding During Service Restoration in Electrical Secondary Distribution Network Based on Reinforcement Learning

Rukia J. Mwifunyi[1†] and Eliah Mbwilo[2]

[1]College of Informatics and Virtual Education, University of Dodoma, P.O. Box 490, Dodoma, Tanzania
[2]College of Information and Communication Technologies, University of Dar es Salaam, P.O. Box 33335, Dar es Salaam, Tanzania.
[†]Corresponding author's e-mail: rukia.mwifunyi@udom.ac.tz.
[†]ORCID: https://orcid.org/0000-0001-7465-3926

ABSTRACT

Increased stress in traditional power systems results in blackouts due to voltage instability attributed to a mismatch between available capacity and load demand, especially in distribution networks. Service restoration schemes are designed to return power supply to the affected parts of the networks. The availability of insufficient supply is a complex problem that requires operational experience or an automatic system. The stochastic nature of load demand significantly impacts service restoration as it results in increased restored demand in case a fault occurs during off-peak hours and helps reduce overload if the fault occurs during peak hours. The study adopts an experimental design methodology to develop the Reinforcement Learning-based service restoration algorithm considering the stochastic nature of load demand. Three reinforcement learning models were used to develop the optimal load shedding model, including Actor-Critic (A2C), a Deep Q Network (DQN) and Proximal Policy Optimization (PPO2), and compared to maximize restored customers, satisfaction of operational constraints, and balancing of power supply and demand. The Particle Swarm Optimization (PSO) algorithm, a metaheuristic algorithm, was also implemented to compare with the proposed approach. The proposed solution has been tested using data from a real electrical secondary distribution network. The proposed solution considered the stochastic nature of load demand, resulting in more restored customers. The computation time during restoration has been improved by 69.8% compared to the metaheuristic approach.

ARTICLE INFO

Submitted: **Nov. 23, 2023**

Revised: **Apr. 3, 2024**

Accepted: **Apr. 15, 2024**

Published: Apr**., 2024**

Keywords: *Reinforcement Learning, Load Shedding, Service Restoration, Power System, Load Demand*

INTRODUCTION

Increased stress in traditional power systems results in blackouts due to voltage instability attributed to a mismatch between available capacity and load demand (Larik et al., 2018). Service Restoration (SR) schemes are designed to return the power supply to the affected part of the network (Mwifunyi *et al.*, 2021). If the source is available in a particular feeder, then a set of loads might be transferred to this feeder. When the available supply is insufficient to save all the customers during SR, optimal load shedding is among the vital operations to mitigate the overloading (Ferreira *et al.*, 2019). Optimal load shedding also helps in avoiding blackouts, maintaining grid stability and reducing stress on power generation (Ahmadipour et al., 2023; Khan et al., 2022; MPEMR, 2023; Rajbhandari et al., 2022). Load shedding during the availability of insufficient supply is a complex problem that requires operational experience or the use of an automatic system. The complexity of the problem is also attributed to the stochastic nature of load demand, which increases search space. An efficient load-shedding mechanism is required to solve the complex problem.

A load-shedding mechanism must be optimal and performed at the right time and location to prevent further voltage instability and avoid losing customer trust in the utility company (Mogaka *et al.*, 2020). Several techniques have been proposed for optimal load shedding during insufficient supply, including Genetic Algorithm (GA) (Luan *et al.*,, 2002; Malekpour and Seifi, 2008), Particle Swarm Optimization (PSO) (Larik et al., 2018), Neural Network (NN) (Zhou *et al.*, 2022), Simulated Annealing (SA), Harmony Search Algorithm (HAS) and Reinforcement Learning (RL) (Ferreira *et al.*, 2019; Chen *et al.*, 2023; Pei *et al.*, 2023). Most metaheuristic approaches take longer than learning-based approaches to reach the final decision.

Advances in RL have shown a dramatic improvement in decision-making problems. Applying RL in Power Grid Systems (PGS) allows for a way of dealing with unanticipated states of the system, which would generally be challenging for manual, metaheuristic, and rule-based control algorithms. This is because RL algorithms can generalize over a large state space, which is the case with PGS. RL has been used in SR problems (Ye *et al.*, 2011; Ghorbani *et al.*, 2016; Ferreira *et al.*, 2019; Kalysh *et al.*, 2019). In (Ye *et al.*, 2011; Ghorbani *et al.*, 2016), RL based on q-learning was used to perform SR without considering load shedding. In (Kalysh et al., 2019), q-learning was used in performing the restoration, considering the load priorities and operation constraints, but load demand variability was not considered. Load shedding and SR using RL have been studied (Ferreira *et al.*, 2019; Vu *et al.*, 2021; Zhang *et al.*, 2021), but fixed load demand was used. Different user types during load shedding have been considered by (Chen *et al.*, 2022); in this study, a Deep Q Network (DQN) with discrete action space was used, which have limitations for large distribution networks. The current study adopted RL using a continuous action space algorithm to perform optimal load shedding if the power supply is insufficient during SR, considering the operational constraints, load priorities, and stochastic nature of load demand. The main contributions of this paper are as follows:

- The use of the RL algorithm with continuous action space in load shedding during service restoration.
- Consideration of the stochastic nature of load demand during service restoration
- Unlike previous service restoration studies, which focused on the primary distribution network, this study focuses on the electrical secondary distribution network, which is more complex.

The rest of the paper is organized as follows: Section II describes the material and methods used in developing the optimal load-shedding algorithm and an overview of the RL algorithm. Section III provides the result and discussion; finally, section V concludes the paper.

METHODS AND MATERIALS

Study Approach

The study adopts an experimental design methodology to develop the RL-based SR algorithm. The environment under RL was designed to involve the definition of actions, states, and reward functions. Tanzania's real power distribution network was used to train and test the model. During training, the electrical network was exposed to different loading levels. Three RL models were used to develop the optimal load shedding model, including Actor-Critic (A2C), DQN and Proximal Policy Optimization (PPO2), and compared to maximize restored customers, satisfaction of operational constraints, and balancing of power supply and demand. A learning rate of 0.005 was found to perform well in training the optimal load-shedding model regarding the ability to converge. RL models successfully converge after 6000-time steps. During training, the simulation starts at each episode with the same initial state in which all loads are served, and there is no voltage violation. Then, the different loading levels are randomly induced into the network at different simulation times. The random selection of loading level guarantees the network to be subjected to normal loading and overloading.

Reinforcement Learning and Load Shedding

RL is a machine learning algorithm that is different from supervised and unsupervised learning. In RL, learning is done by mapping situations to actions to maximize a numerical reward signal (Sutton and Barto, 2017). The learner is not told which actions to take but must discover which ones yield the most reward by trying them. In RL, an agent, the learner, learns independently from experience gathered by trial and error (On-Policy Learning) or experience gathered by others (Off-Policy Learning). After learning, the knowledge obtained controls its actions in the environment for maximum reward (Jin & Ma, 2017).

There are different types of RL algorithms, including Q-learning-based, SARSA-based, and deep Q-learning algorithms. The commonly used RL algorithm in power restoration is Q-learning-based (Kalysh et al., 2019; Ribeiro et al., 2017; Shirazi & Jadid, 2019). In the Q learning algorithm, the decision is evaluated using the reward function, which expresses the effectiveness of the chosen control action in terms of power loss minimization and minimization of energy not served customers. The parameters used in the Q learning algorithm are learning rate, discount factor, and greedy, denoted by α , γ ε respectively. The Q-function (1) and its update function are presented in (2).

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \tag{1}$$

$$Q_{(t+1)}(s_t,a_t) = (1-\beta)Q_t(s_t,a_t) + \beta[r_t + \gamma V(s')] \tag{2}$$

where $Q(s,a)$ is a Q function that depends on the action a taken by an agent in the state s, r is a reward gained by an agent upon taking action a, and $Q(s',a')$ is the maximum Q value obtained from optimal action and state.

A major limitation of Q-learning is that it only works in environments with discrete and finite state and action spaces. Applying Q-learning to a large state space slows down convergence. Function Approximation of the state is used to learn a reduced state representation. Large-scale problems lead to memory explosion, and more training is needed to develop a good solution. As a result, Deep Reinforcement Learning (DRL), which is the combination of RL and deep learning technologies, has been devised to solve problems faced by Q-learning (Huang et al., 2020). Several DRL algorithms have been devised in literature, including A2C, DQN, PPO2, and Soft Actor-Critic (SAC). DQN is the most used algorithm in power system control, which uses a neural network (NN) with weights θ to estimate Q values. The DQN uses a target network ($\hat{Q}$) beside the Q network and experiences replay during training, which increases efficiency and stability. The agent's experience $e_t = (s_t, a_t, r_t, s_{t+1})$

is stored in dataset D each time step for performing experience replay. A Q-network is trained using mini-batches randomly drawn from D by minimizing a sequence of the loss function (3).

$$L_i(\theta_i) = \mathrm{E}_{(s,a)\sim p}[(y_i - Q(s,a;\theta_i))^2] \quad (3)$$

where y_i is the target Q-value for iteration i computed by Q', and p is the probability distribution of the state and action pair (s,a). NN weights can be updated by stochastic gradient descent with the gradient calculated by (4).

$$\nabla_{\theta i} L_i(\theta_i) = \mathrm{E}_{(s,a)\sim p}[(y_i - Q(s,a;\theta_i))\nabla_{\theta_i} Q(s,a;\theta_i)] \quad (4)$$

DQN performs well in problems with discrete action space, while A2C and PPO2 have shown good performance in problems with continuous action space (Hojmark & Gyllensten, 2019). A2C updates the actor-network, which represents the policy, and the critic network, which represents the value functions. Over time, in A2C, the actor learns to get better actions and critic evaluates those actions. Proximal Policy Optimisation combines having many agents in A2C with exploring the policy region in the optimisation problem (Yang et al., 2020).

Metaheuristic Algorithm in Load shedding

Some of the existing metaheuristic algorithms in load shedding for power systems are glowworm swarm optimisation (GSO) (Mageshvaran & Jayabarathi, 2015), GA (Guichon et al., 2012), PSO (He et al., 2009; Mwifunyi et al., 2021), and grasshopper optimisation algorithm (GOA) (Ahmadipour et al., 2023). While the GA-based algorithms are more accurate than other metaheuristic algorithms, they take longer to find the best answer. PSO, GSO, and GOA can experience unpredictable convergence speeds and local minimum.

Load Demand Forecasting Model During Service Restoration

This study's load demand forecasting model was created for our earlier investigation (Mwifunyi et al., 2020). In this study, Long Short-Term Memory (LSTM) was utilised for service restoration since it demonstrated superior performance, with an accuracy of 96.4%, when compared to Recurrent Neural Network and Gated Neural Network. The goal of Mwifunyi et al. (Mwifunyi et al., 2020) was to create a forecasting model for usage in SR; in this work, the actual application was realised and utilised in the SR considering the stochastic nature of load demand and load shedding. Following the voltage instability, the model is run using the 48 load demand values before the occurrence of the incidence. The historical dataset was used to create the model.

Problem Formulation

RL has been designed based on the Markov Decision Process for decision-making during the SR process. RL requires the definition of an Environment to accomplish the decision-making process. The environment for RL differs based on the application to be used; for the power SR, the environment has been designed by adopting the approaches presented by Huang et al. (Huang et al., 2020). Environment design involves the definition of states, actions, and reward functions. The state of the environment is based on the status of the power system network topology. The network topology is defined as the graph with edges and vertices. In the vertices, four parameters are used to determine the environment state: priority level, status, demand, and voltage. Action to be taken involves switching the switches by opening or closing a switch. A string of bits represents actions, and their size depends on the number of buses or zones within the power distribution system. The reward function is computed under four main goals, which are as follows:

i) Maximization of restored customers as presented by (5);

$$r_d = \sum_{i \in buses} w_i * L_i * y_i \tag{5}$$

where L_i is the load at the bus i, y_i: status of the load at the bus i, w_i: priority level of the load at the bus i, r_d: restored demand.

ii) Minimization of the voltage deviation from the threshold values;

iii) Minimization of power loss and

iv) Backup feeder transfer capacity limit.

The power of load supplied from the backup feeder must not exceed the feeder's available capacity, as seen in (6).

$$P_{available,t} \geq \sum_{d=1}^{D} P_{d,t} y_{d,t} \tag{6}$$

where $P_{available}$: power from the source, and P_d: power of the load, and y_d: status of the load.

The total reward was computed as, $R = r_d + r_v + r_p + r_{bf}$

where r_d: reward due to restored load maximization, r_v: reward due to voltage deviation, r_p: reward due to power loss, and r_{bf}: reward due to backup feeder limit.

Proposed Solution

After representing the restoration as a Markov Decision Process, an RL algorithm called Actor-Critic, a policy gradient method that learns the agent's policy given the environment instead of learning value functions, has been implemented. DQN and PPO2 were also considered in this study. The deep Q network was implemented with discrete action space, while the A2C and PPO2 networks had continuous action space.

The basic idea is that given that the network in the power grid system is represented as a graph that could have any number of nodes, the state space is infinite. There is a need to find a way to generalize the information in this large state space. The network is then represented in an environment with a continuous state and action space. The network information is used as input to the agent at each state, and the agent decides based on the policy it has learned how best to achieve optimal reward given the current state. The relationship between the learning environment and power grid environment is as follows: Agent: switch or load; Actions: Open/Close load switch; State: System topology; Reward: Minimize the unrestored customers while fulfilling the operational constraints; and Environment: Electrical Power Distribution Network.

An environment is needed to train an RL agent. The environment receives the agent's actions and returns to the agent a new state, as well as rewards based on the goal of the agent and the quality of the state in which the agent is. For the problem of optimal load shedding, the environment has been designed with two main modules. The first is the RL module, developed using OpenAI Gym (Brockman et al., 2016). The second module is the Power Flow Analysis Module, developed based on the Direct load Flow method and programmed with Java to provide real-time responses to actions performed by agents. The two modules communicate using a communication bridge called Py4J, as seen in Figure 1. With this environment developed using OpenAI Gym, running different RL algorithms without changing the environment is possible.

Figure 1: Interface between RL algorithms and power system

Tanzania Secondary Distribution Network: Case Study

The data for conducting experiments was acquired from the Tanzanian Secondary Distribution network using three transformers, which were selected as the case study. The selected transformers are the BBQ Village SS1, Kimweri Avenue SS2, and Abiudi Street SS2, as seen in Figure 2. BBQ Village SS1 has 77 buses, Abiudi Street SS2 has 25 buses, and Kimweri Avenue has 20 buses.

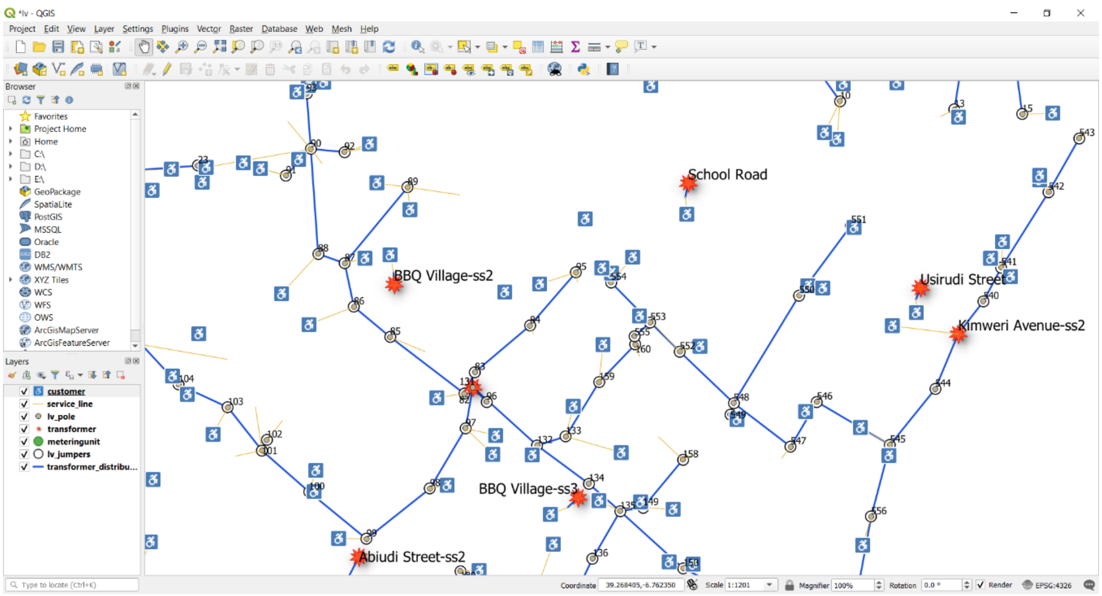

Figure 2: Secondary distribution network with pilot site.

RESULTS AND DISCUSSIONS

Optimum Sizing Results using PSO

RL has been implemented using Python based on the Stable Baseline framework (Hill et al., 2018). A stable baseline has been selected as it is widely documented, implemented several deep reinforcement learning algorithms, including A2C, DQN, PPO2, and SAC, and is frequently maintained. The OpenAIGym environment has been customized to make decisions on the power system. The RL algorithm was implemented and integrated with the Java environment to confirm the decision made using the power flow analysis algorithm. The state of the environment is also obtained from a Java environment. Three RL algorithms have been implemented, namely, A2C, PPO2, and DQN. Results for the reward function during the training of these algorithms are presented in Figure 3, Figure 4, and Figure 5. All algorithms successfully converged after at least 2000 timesteps. Initially, all three algorithms were trained with discrete action space, and the performance was slightly poor due to the large action space (Hill et al., 2018) of the restoration problem. Then, the continuous action space was used, which resulted in improved high-priority restored customers. A2C and PPO2 were used with continuous action space, while the DQN was utilized in discrete action space as it always maximizes over discrete action selection. The observation space at a given time is presented by the voltage level at every node and the actions taken for every bus, as seen in Figure 6. Trained models have been used to perform load shedding during SR. The trained models have been used to perform load shedding after load transfer through network reconfiguration from the BBQ village network to the Abiudi street network, which caused the voltage limit violation. Load shedding successfully improved the Abiudi Street network's voltage level following the network reconfiguration, as seen in Figure 7.

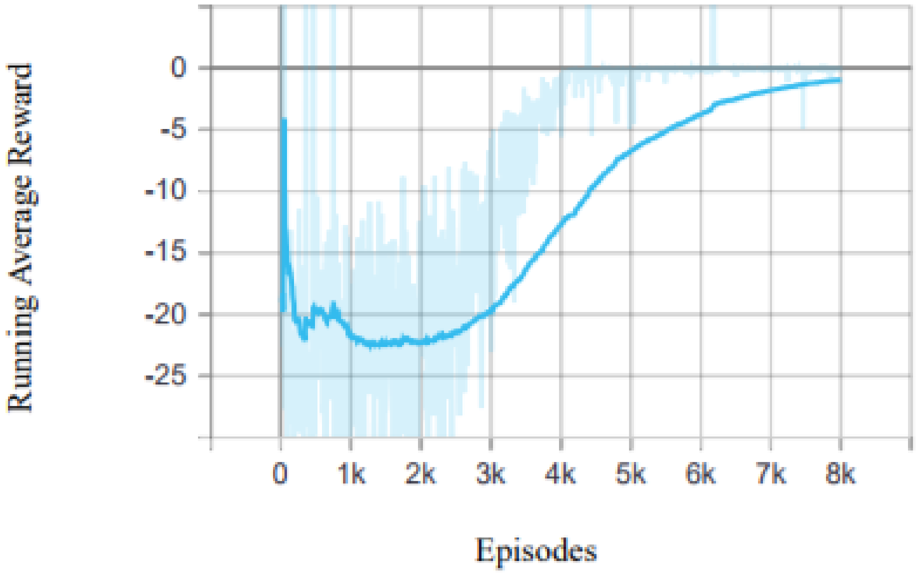

Figure 3: Running average reward during training using A2C.

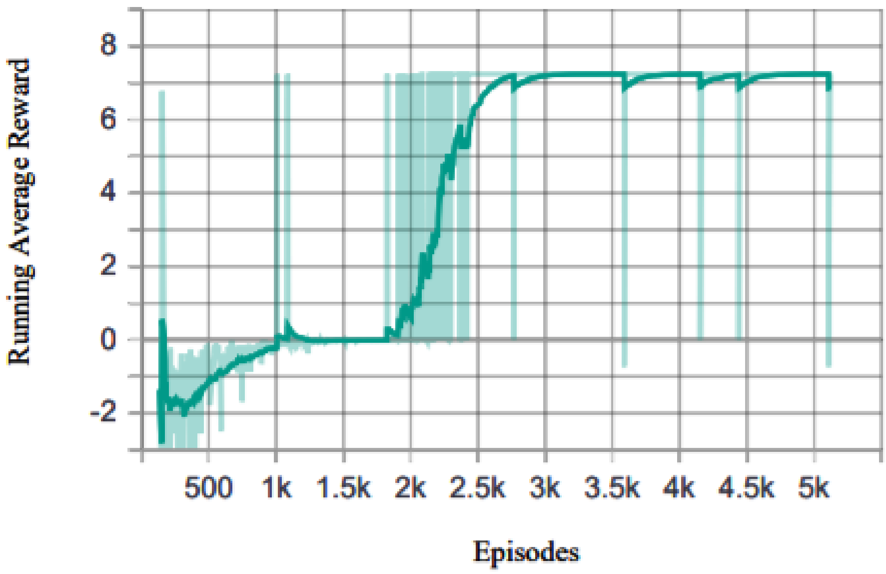

Figure 4: Running average reward during training using PPO2.

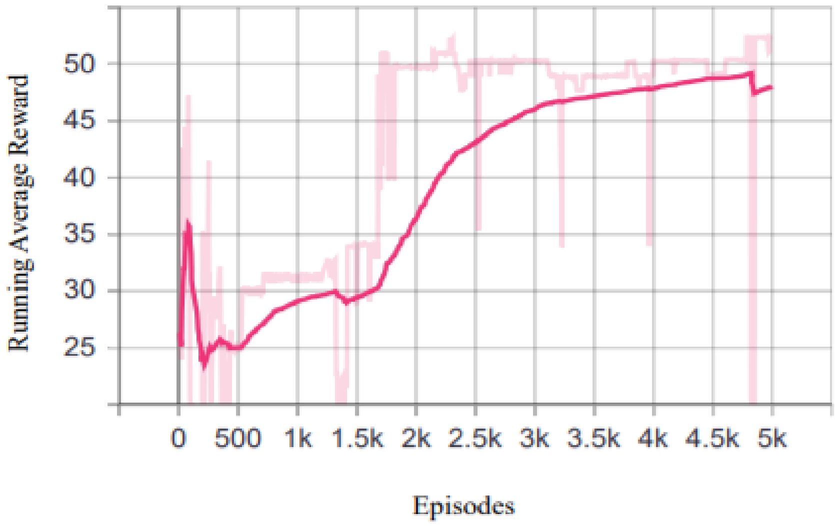

Figure 5: Running average reward during training using DQN.

```
OBSERVATIONS: [1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0. 1. 1. 1. 1. 0. 0. 0.
 1. 1. 1. 1. 1. 1. 1. 1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1.]
[1.         0.99969623 0.99946801 0.99941558 0.99658849 0.99367159
 0.99190169 0.99162273 0.99183004 0.99029469 0.98944955 0.99021809
 0.98949072 0.98949072 0.99941558 0.99947019 0.9934937  0.98872441
 0.98337288 0.97340691 0.96545861 0.96539685 0.95923595 0.9533369
 0.94759566 0.94337508 0.93855464 0.93298296 0.92774609 0.92394047
 0.92033912 0.92032964 0.91592958 0.91459334 0.9099093  0.90673216
 0.90616995 0.90432913 0.90416543 0.9030435  0.90262117 0.90254644
 0.90200496 0.90161654 0.90062763 0.90062038 0.90062038 0.90059848
 0.9005781  1.         0.99568424 0.99505577 0.99167178 0.98908929
 0.98554502 0.98052384 0.98028222 0.97978257 0.97788877 0.97624178
 0.97292162 0.97410484 0.97286818 0.9709859  0.96944369 0.96874502
 0.96844389 0.98826756 0.98522987 0.98152189 0.97945106 0.97769953
 0.97665371 0.97629018 0.98047216 0.98026252 0.99486405]
```

Figure 6: Observation Space.

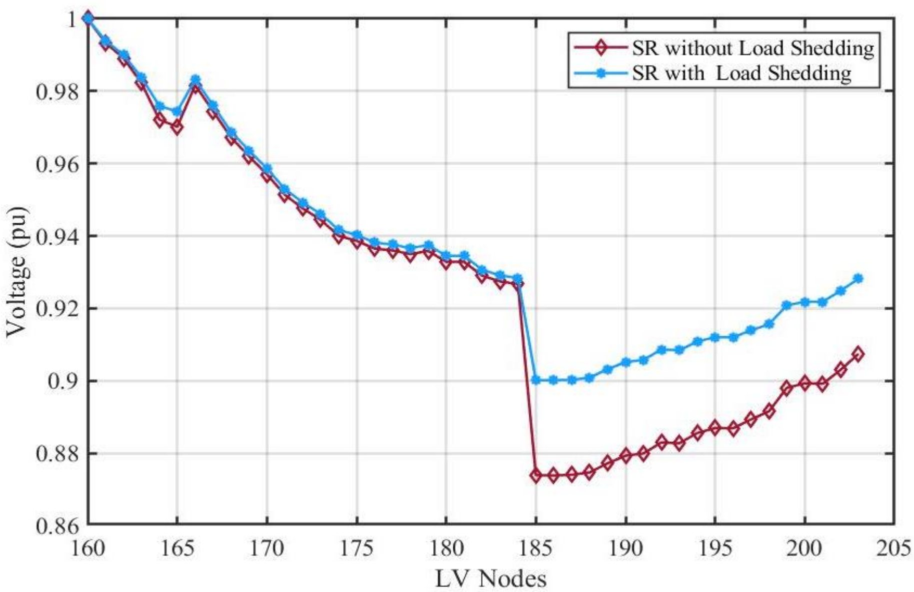

Figure 7: Service restoration with load shedding.

Restoration Efficiency

Restoration efficiency for the proposed optimal load shedding mechanism has been measured based on the Restoration Efficiency of Important Loads (REI). In power systems, essential loads include hospitals, schools, and the military. The restoration efficiency for the developed RL-based load-shedding operation was collected, analyzed, and compared. The quality of results is also assessed by considering the objective functions and the constraints stated in section 2.2. The priority level of customers restored during the restoration process and power loss of the power network after restoration was considered. Comparative results on the restoration efficiency comparing the network status after restoration and priority levels are presented in Table 1 in which one represents the restored network part and zero, the unrestored network part. It has been revealed that the RL algorithm successfully restores high-priority customers and sheds only low-priority customers when the power supply is insufficient.

Table 1: Comparative results for restoration efficiency

Priori	1	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.1	.	.	.	1
PPO2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
A2C	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
DQN	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1

Comparison of Reinforcement Learning and Binary Partial Swarm Optimization (BPSO)

Moreover, BPSO was conducted to compare RL-based algorithms and metaheuristic algorithms, as presented in Table **2**. Three RL algorithms, namely A2C, DQN, and PPO2, have been used in which PPO2 takes the least time but achieves a higher final reward than others. It has been revealed that all RL algorithms take less time to achieve the optimal result than BPSO. The study proposes using RL in case the knowledge base is enough to make the decision and using metaheuristic methods only when there is less knowledge base. Restoration time has improved by almost 69.8% with RL compared to BPSO.

Table 2: Performance comparison of RL and BPSO

Agent	Training Time (Seconds)	Running Time (Seconds)	Average Reward
ACTOR CRITIC	2087.55	0.290	6.68433
DQN	1888.3461	0.2186	5.1514139
BPSO	Not applicable	1.703	6.3326801
PPO2	2056.8336	0.296	6.9177437

Consideration of Variable Load Demand During Service Restoration

The stochastic nature of load demand has also been considered in this study by using a well-established Long Short Term Memory (LSTM) forecasting model with a forecasting accuracy of 96.43% (Mwifunyi *et al.*, 2020). During SR, the forecasted values were calculated based on the time of the day. The stochastic nature of load demand has reduced the load shedding significantly once a fault occurs in low peak hours, in which load can be restored up to 100% compared to peak demand load, which ended up to 89.77%. Comparison results for load shedding with RL considering the stochastic nature of load demand are presented in Table 3 which faults were simulated at the same point at different times of the day. It has finally revealed a different load shedding level based on the time of occurrence of a fault.

Table 3: Restoration with variable load demand

Fault Time	% Restored	Shedding Bus	Priority
0000	98.04	48	0.1
1200	100	-	-
0600	97.58	42,48	0.1
1900	97.44	47,48	0.1
2100	89.77	89,94,116,121,122,125,127,128,138	0.1,0.5

The Use of Forecasted Demand and Peak Demand with Load Shedding during Service Restoration

Load shedding during SR as the result of using forecasted demand and peak demand was also compared, as presented in Figure 8. With the use of forecasted demand, there were no voltage limit violations, and all load was restored with a minimum voltage of 0.9037 p.u. Using peak demand resulted in a restored demand of 92.8% with a minimum voltage of 0.9003 p.u.

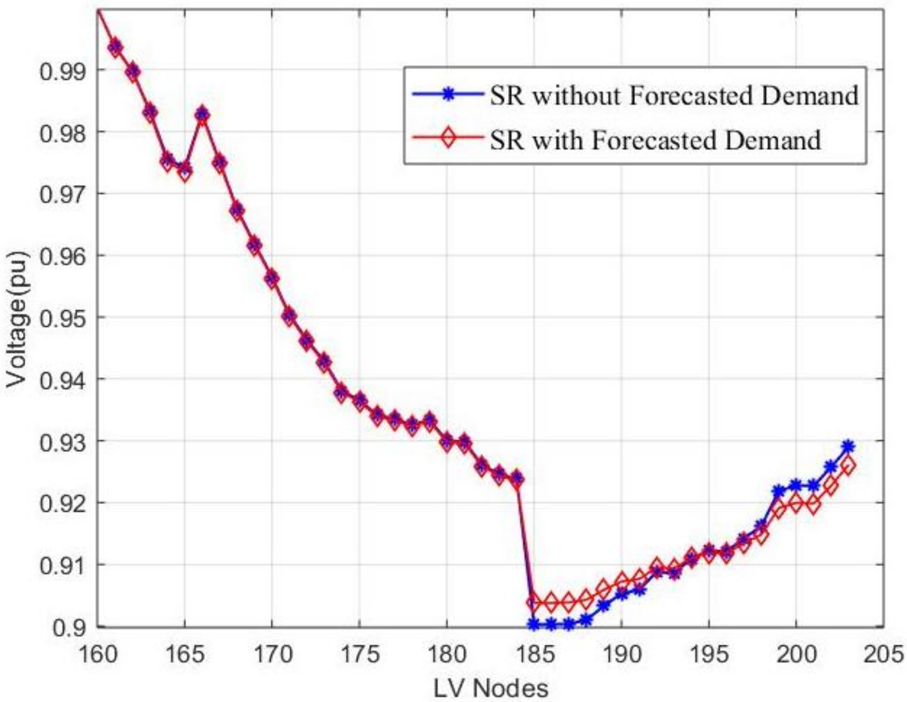

Figure 8: Load Shedding due to the Use of Forecasted Demand and Peak Demand

CONCLUSION

Conclusion

This study focused on developing the optimal load shedding mechanism using RL. The study used the reinforcement learning algorithm with continuous action space in load shedding during service restoration with consideration of the stochastic nature of load demand. The data from the real secondary distribution network were used. It has been revealed that using RL improves the decision-making time by 69.8% compared to the time taken by metaheuristic approaches with almost similar performance. The RL-based algorithm can also shed only low-priority customers once there is a voltage limit violation. The use of variable load demand has shown improvement in the number of restored customers compared to fixed load demand during SR using RL. Network dynamic as the result of network reconfiguration is omitted for future work.

Acknowledgment

This research is part of the capacity building iGrid-Project at the College of Information and Communication Technologies (CoICT) at the University of Dar es Salaam (UDSM) under the sponsorship of Swedish International Development Agency (SIDA).

REFERENCES

Ahmadipour, M., Murtadha Othman, M., Salam, Z., Alrifaey, M., Mohammed Ridha, H., & Veerasamy, V. (2023). Optimal load shedding scheme using grasshopper optimization algorithm for islanded power system with distributed energy resources. *Ain Shams Engineering Journal*, *14*(1), 101835. https://doi.org/10.1016/j.asej.2022.101835

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). *OpenAI Gym*. ArXiv:1606.01540.

Chen, H., Zhuang, J., Zhou, G., Wang, Y., Sun, Z., & Levron, Y. (2023). Emergency load shedding strategy for high renewable energy penetrated power systems based on deep reinforcement learning. *2nd International Joint Conference on Energy and Environmental Engineering*, *9*, 434–443. https://doi.org/10.1016/j.egyr.2023.03.027

Chen, Y., Liao, S., & Xu, J. (2022). Emergency load-shedding optimization control method based on reinforcement learning assistance. *Energy Reports*, *8*, 1051–1061. https://doi.org/10.1016/j.egyr.2022.02.140

Ferreira, L. R., Aoki, A. R., & Lambert-Torres, G. (2019). A Reinforcement Learning Approach to Solve Service Restoration and Load Management Simultaneously for Distribution Networks. *IEEE Access*, *7*, 145978–145987. https://doi.org/10.1109/ACCESS.2019.2946282

Ghorbani, M. J., Choudhry, M. A., & Feliachi, A. (2016). A Multiagent Design for Power Distribution Systems Automation. *IEEE Transactions on Smart Grid*, *7*(1), 329–339. https://doi.org/10.1109/TSG.2015.2453884

Guichon, M., Melo, M., Nieto, A. C., Vignolo, M., & Yedrzejewski, N. (2012). Automatic Load Shedding Calculated with genetic algorithms - DAC-CMAG. *2012 Sixth IEEE/PES Transmission and Distribution: Latin America Conference and Exposition (T&D-LA)*, 1–7. Available at: https://doi.org/10.1109/TDC-LA.2012.6319121.

He, F., Wang, Y., Chan, K. W., Zhang, Y., & Mei, S. (2009). Optimal load shedding strategy based on particle swarm optimization. *8th International Conference on Advances in Power System Control, Operation and Management (APSCOM 2009)*, 146–146. https://doi.org/10.1049/cp.2009.1795

Hill, A., Raffin, A., Maximilian, E., Gleave, A., Kanervisto, A., Traore, R., Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S., & Wu, Y. (2018). *Stable Baselines*. GitHub. Available at: https://github.com/hill-a/stable-baselines.

Hojmark, A., & Gyllensten, R. (2019). *Control temperature of room with reinforcement learning*. Gothenburg University.

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., & Huang, Z. (2020). Adaptive Power System Emergency Control Using Deep Reinforcement Learning. *IEEE Transactions on Smart Grid*, *11*(2), 1171–1182. https://doi.org/10.1109/TSG.2019.2933191

Jin, J., & Ma, X. (2017). A multi-objective multi-agent framework for traffic light control. *2017 11th Asian Control Conference (ASCC)*, 1199–1204. https://doi.org/10.1109/ASCC.2017.8287341

Kalysh, I., Kenzhina, M., Kaiyrbekov, N., Kumar Nunna, H. S. V. S., Dadlani, A., & Doolla, S. (2019). Machine Learning-based Service Restoration Scheme for Smart Distribution Systems with DGs and High Priority Loads. *2019 International Conference on Smart Energy Systems and Technologies (SEST)*, 1–6. https://doi.org/10.1109/SEST.2019.8849002

Khan, N., Shahid, Z., Alam, M. M., Bakar Sajak, A. A., Mazliham, M. S., Khan, T. A., & Ali Rizvi, S. S. (2022). Energy Management Systems Using Smart Grids: An Exhaustive Parametric Comprehensive Analysis of Existing Trends, Significance, Opportunities, and Challenges. *International Transactions on Electrical Energy Systems*, *2022*. https://doi.org/10.1155/2022/3358795

Larik, R., Mustafa, M. W., Aman, M., Jumani, T., Sajid, S., & Panjwani, M. (2018). An Improved Algorithm for Optimal Load Shedding in Power Systems. *Energies*, *11*(7), 1808. https://doi.org/10.3390/en11071808

Luan, W. P., Irving, M. R., & Daniel, J. S. (2002). Genetic algorithm for supply restoration and optimal load shedding in power system distribution networks. *IEE Proceedings-Generation, Transmission and Distribution*, *149*(2), 145–151. https://doi.org/10.1049/ip-gtd:20020095

Mageshvaran, R., & Jayabarathi, T. (2015). GSO based optimization of steady state load shedding in power systems to mitigate blackout during generation contingencies. *Ain Shams Engineering Journal*, *6*(1), 145–160. https://doi.org/10.1016/j.asej.2014.09.010

Malekpour, A. R., & Seifi, A. R. (2008). An optimal load shedding approach for distribution networks with DGs considering capacity deficiency modelling of bulked power supply. *International Journal of Electrical, Computer, Electronic and Communication Engineering*, *2*(10), 1–7. Available at: https://doi.org/10.1016/j.asej.2014.09.010.

Mogaka, L. O., Nyakoe, G. N., & Saulo, M. J. (2020). Power Prioritization and Load Shedding in an Island with RESs Using ABC Algorithm. *Journal of Engineering (United Kingdom)*, *2020*. https://doi.org/10.1155/2020/8131952

MPEMR, B. (2023). *Integrated Energy and Power Master Plan (IEPMP) 2023* (Issue July).

Mwifunyi, R. J., Kissaka, M. M., & Mvungi, N. H. (2020). Short-Term Load Forecasting for Improved Service Restoration in Electrical Power Systems: A Case of Tanzania. *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, 1–5. https://doi.org/10.1109/AISP48273.2020.9073069

Mwifunyi, R. J., Mvungi, N. H., & Kissaka, M. M. (2021). Multiagent based fault localisation and service restoration in Tanzanian secondary distribution network. *Cogent Engineering*, *8*(1), 1975900. https://doi.org/10.1080/23311916.2021.1975900

Pei, Y., Yang, J., Wang, J., Xu, P., Zhou, T., & Wu, F. (2023). An emergency control strategy for undervoltage load shedding of power system: A graph deep reinforcement learning method.

IET Generation, Transmission and Distribution, *17*(9), 2130–2141. https://doi.org/10.1049/gtd2.12795

Rajbhandari, Y., Marahatta, A., Shrestha, A., Gachhadar, A., Thapa, A., Gonzalez-Longatt, F., Guerrero, J. M., & Korba, P. (2022). Load prioritization technique to guarantee the continuous electric supply for essential loads in rural microgrids. *International Journal of Electrical Power and Energy Systems*, *134*(November 2020), 107398. https://doi.org/10.1016/j.ijepes.2021.107398

Ribeiro, R., Enembreck, F., Guisi, D. M., Casanova, D., Teixeira, M., De Souza, F. A., & Borges, A. (2017). An Advanced Software Tool to Simulate Service Restoration Problems : a case study on Power Distribution Systems. *International Conference on Computational Science*, *108*, 675–684. https://doi.org/10.1016/j.procs.2017.05.248

Shirazi, E., & Jadid, S. (2019). A multiagent design for self-healing in electric power distribution systems. *Electric Power Systems Research*, *171*, 230–239. https://doi.org/10.1016/j.epsr.2019.02.025

Sutton, R. S., & Barto, A. G. (2017). *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press.

Vu, T. L., Mukherjee, S., Huang, R., & Huang, Q. (2021). Barrier Function-based Safe Reinforcement Learning for Emergency Control of Power Systems. *2021 60th IEEE Conference on Decision and Control (CDC)*, 3652–3657. https://doi.org/10.1109/CDC45484.2021.9683573

Ye, D., Zhang, M., & Sutanto, D. (2011). A hybrid multiagent framework with Q-learning for power grid systems restoration. *IEEE Transactions on Power Systems*, *26*(4), 2434–2441. https://doi.org/10.1109/TPWRS.2011.2157180

Zhang, Y., Yue, M., & Wang, J. (2021). Adaptive Load Shedding for Grid Emergency Control via Deep Reinforcement Learning. *2021 IEEE Power & Energy Society General Meeting (PESGM)*, 01–05. https://doi.org/10.1109/PESGM46819.2021.9638058

Zhou, Y., Park, J., & Zhu, H. (2022). Scalable Learning for Optimal Load Shedding Under Power Grid Emergency Operations. *IEEE Power and Energy Society General Meeting*, *2022-July*. https://doi.org/10.1109/PESGM48719.2022.9916594