Power Increase Automation of Small Modular Reactor Using Proximal Policy Optimization and a Rule-based System

Hee-Jae Lee[1], Daeil Lee[2], San Kang[3], Jonghyun Kim[1]*
[1]*Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*
[2]*Korea Atomic Energy Research Institute, 111, Daedeok-daero 989 Beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea*
[3]*Chosun University, 309 Pilmun-daero, Dong-gu, Gwangju, 61452, Republic of Korea*
[*]*Corresponding author: jonghyun.kim@kaist.ac.kr**

***Keywords: Small Modular Reactor, Autonomous Operation, Power Increase, Reinforcement Learning, Proximal Policy Optimization**

1. Introduction

Global interest in small modular reactors (SMRs) is steadily increasing, driven by their advantages in enhanced safety design (e.g., passive safety systems) and operational flexibility (e.g., electricity demand-adjustment capability). Given that SMRs are intended for multi-module operation concept, operators are tasked with monitoring and controlling multiple modules from a single control room. In the experiment performed by J. Hartmann et. al. [1], it is observed that the number of reactors that a single operator can safely operate highly dependent on the states of the reactors. Furthermore, when manual intervention is required, the situation becomes increasingly challenging a single operator to manage and increase operator's workload.

Power increase operation is a typical example where manual intervention is dominant among the operational modes in nuclear power plants (NPPs). First, the need of decision making is increased such as selection of power operation target and determination of the control strategy based on guidelines [2]. Second, many manual actions are required due to extensive maintenance and monitoring plant parameters [2]. Lastly, procedure description may be relatively insufficient in some procedural steps, providing only the operational goal without detailing the specific tasks for the operator.

One effective strategy to reduce operator's workload during power increase operation is to enhance the level of automation. In recent years, the utilization of AI has become a prominent trend across various industrial fields, with its ability to process numerous variables and to make complex decisions autonomously. AI presents a promising alternative for developing intelligent controllers specifically designed for power increase operations [2-4].

In this regard, this study focuses on developing an algorithm to fully automate the power increase operation of SMR. To achieve this, a simulator of an integral Pressurized Water Reactor (iPWR), a type of SMR currently being developed in several countries, was utilized. The algorithm aims at increasing the reactor power from 0% to 100%. To identify the targets for automation and derive appropriate control strategies, task analysis on the power increase operation procedure was performed. Based on the control strategies, proximal policy optimization (PPO) and 'if-then' logic were applied respectively for continuous and discrete actions.

2. Power increase operation of small modular reactor

2.1. Overview of the power increase operation

The iPWR simulator used in this study is designed for a SMR with a thermal power of 150 MW and an electrical power of 45 MWe. In this simulator, the operational range covers from the hot standby condition, as defined by commercial large-scale nuclear power plants, to 100% reactor power. The operation is broadly divided into two phases: 1) low-power operation from the hot standby condition to turbine synchronization, and 2) high-power operation from turbine synchronization to 100% reactor power. The initial and final conditions for the power increase operation are presented in Table I.

Table I. Major parameter status at initial and final conditions of power increase operation

Major parameter	Initial condition	Final condition
Reactor neutron power	0 %	100 %
Generator power	0 %	45 MWe
Reactor coolant system (RCS) average temperature	250.11 ℃	287.42 ℃
Turbine revolutions per minute	0 RPM	3600 RPM
Turbine load demand	0 MWe	45 MWe
Rod position	0 Step (A Bank) 0 Step (B Bank) 0 Step (C Bank)	80 Step (A Bank) 80 Step (B Bank) 49 Step (C Bank)
Synchronous connection	Disconnected	Connected

2.1. Task analysis

This section describes a task analysis for the power increase operation, utilizing the power increase operation procedure of the iPWR simulator. Through the analysis, this study identified the targets for automation, control implementation strategies, and the key input and output parameters for the algorithm.

For the task analysis, the decomposition method was used. This method involves breaking down complex tasks into smaller, more manageable subtasks, thereby facilitating a systematic understanding of each step within the operating procedure. Task decomposition was conducted based on the following categories: step, main task, sub-task, action verb, monitoring parameter, expected response, graphical user interface (GUI) sheet, manipulation, and control type.

Table II presents an example of the task analysis results, identifying a total of 19 control actions. These actions were categorized into two types: discrete and continuous controls. Discrete control involves the direct setting of a target value based on specific, predefined conditions. In such systems, the control actions are triggered by distinct conditions, resulting in a step-wise change in the system state. For example, "Push OUT until Bank C reaches 49 steps". For discrete controls, the strategy was set to apply 'if-then' logic. In contrast, continuous controls involve adjusting component states to achieve specified operational goals, where the adjustments cannot be easily defined by simple logic. For example, "Introduce the new boron concentration setpoint (until neutron power stabilizes at 8%)". Reinforcement learning is employed for continuous control, as it requires real-time monitoring of parameter states, adaptive decision-making, and ongoing adjustments to respond effectively to dynamic conditions.

3. Development of power increase automation algorithm

This section describes the design and development process of the power increase automation algorithm The control modules designed for automation are categorized into two types: discrete and continuous, as illustrated in Figure 1.

3.1. Discrete control module using if-then logic

The discrete control module was designed for the discrete controls identified through the task analysis, with 'if-then' logic. For example, in Step 7, the control rods in bank B need to be withdrawn. The automation of this step is implemented using 'if-then' logic in the following sequence: 1) switch the bank selector to B, 2) click the 'OUT' button until all control rods in bank B are fully withdrawn, and 3) once the rods are fully withdrawn (at 80 steps), proceed to the next step in the procedure.

3.2. Continuous control module using PPO algorithm

PPO algorithm is a policy-based reinforcement learning method based on the Actor-Critic framework. In this approach, the Actor network determines which action to take given a specific state, while the Critic network calculates the advantage of the action taken by the Actor. The key strength of the PPO algorithm is stability in learning convergence with a clipping mechanism. During the process of updating the policy by the Actor network, there is a risk that the new policy may deviate, potentially leading to divergence in the learning process. To prevent this risk, the PPO algorithm applies clipping to ensure that the new policy does not exceed a predefined threshold, thereby improving the stability.

Step 9 is a representative example of a continuous control operation, where boron dilution is required to achieve reactor criticality. The goal of this step is to raise reactor power and stabilize it at 8%. During this process, the boron concentration must be adjusted within 20 ppm per batch, and the start-up rate (SUR) must be maintained below 0.5 dpm. To automate Step 9, reinforcement learning was applied to enable continuous control that adapts to varying conditions.

Table II. Example of task analysis on the power increase operation procedure using the decomposition method.

Step	Main task	Sub-task	Action verb	Monitoring parameter	Expected response	GUI sheet	Mani-pulation	Control type
1	Check the following variables are stable	Thermal power	Check	Thermal power	1.59 MWt	Overview	x	
		Neutron power	Check	Neutron power	0%	Overview	x	
		SUR	Check	SUR	0 dpm	Rod Position Controls	x	
		Primary temperature	Check	Primary temperature	~ 248 ℃	Overview	x	
		Primary pressure	Check	Primary pressure	15.5 MPa.	Overview	x	
		Primary level	Check	Primary level	23%	Overview	x	
		FW flow	Check	FW flow	0.02 kg/s	Overview	x	
2	Ensure proper status of the main plant controls	Plant mode in Reactor Leading mode	Ensure	Plant mode	Plant Mode selector in Reactor leading mode.	RodPositionControl	o	Discrete
		Rod Control in Manual	Change	Rod control	Rod control selector in Manual mode	RodPositionControl	o	Discrete
		MSB is controlling Steam pressure	Ensure		-	-	x	
		FW system in Auto	Ensure		FWSV13/ FWSV14 in Auto.	Feedwater Control	x	
		CW and CNR systems in service	Ensure		CW system in service	Systems	x	
		FW control adjusting FW flow	Ensure	FW control valve	FWSV13/FWSV14 in Auto	Systems	x	
3	Check alarm conditions exist	Reactor trip	Check		Alarm blinking	Alarms	x	
		Turbine trip	Check		Alarm blinking	Alarms	x	
		Turning gear	Check		Alarm blinking	Alarms	x	
		Generator Breaker open	Check		Alarm blinking	Alarms	x	
4	Reset reactor trip		Reset		Reactor trip alarm deactivates	Trips	o	Discrete
5	Check Boron concentration stable		Check			Core	x	
6	Withdraw shutdown bank A while monitoring neutron flux	Monitor neutron flux	Monitor			Rod Position Control	x	
		Ensure Bank A is selected	Push		Back A is selected.	Rod Position Control	o	Discrete
		Push OUT until Bank A is fully withdrawn	Push	Bank A position	Shutdown rod position changes	Rod Position Control	o	Discrete
			Monitor			Rod Position Control	x	
7	Withdraw shutdown bank B while monitoring neutron flux	Monitor neutron flux	Monitor	Source range power SUR	Source Range power (RCSNP03_TR) begins to rise but stabilises or falls when rod movement stops.	Rod Position Control	x	
		Ensure Bank B is selected	Push	Bank B position	Back B is selected.	Rod Position Control	o	Discrete
		Push OUT until Bank B is fully withdrawn	Push		Control Rod position changes.	Rod Position Control	o	Discrete
8	Withdraw shutdown bank C while monitoring neutron flux	Monitor neutron flux	Monitor	Source range power SUR	Source Range power (RCSNP03_TR) begins to rise but stabilises or falls when rod movement stops.	Rod Position Control	x	
		Ensure Bank C is selected	Push	Bank C position	Back C is selected.	Rod Position Control	o	Discrete
		Push OUT until Bank C is fully withdrawn	Push		Control Rod position changes.	Rod Position Control	o	Discrete
9	Dilute RCS for Criticality	Introduce the new Boron concentration setpoint (max 20 ppm per badge).	Introdu-ce	Boron concentration SUR	Introduced values appear correctly.	Core	o	Discrete
		Place selector in 'Dilution'	Place	Selector state SUR	Boron concentration begins to lower	Core	o	Discrete
		Check Boron concentration decreases to the Boron concentration setpoint.	Check	Boron concentration SUR	Boron concentration begins to lower	Core	x	
				Neutron power SUR	When critical, neutron power RCSNP01_TR) rises slowly without Control Rod movement.	Core	x	
		Repeat changing concentration setpoint above until Neutron flux slowly rises without Control Rod movement.	Repeat	Boron concentration Neutron power Reactivity SUR	Criticality observed Point of Adding Heat occurs when Intermediate Range ~ 10E-8. SUR should not exceed a sustained 0.5 dpm.	Core Rod Position Control	o	Continuous
		Raise Reactor power to 8%	Raise	Boron concentration	Neutron power stabilises	Core	o	Continuous

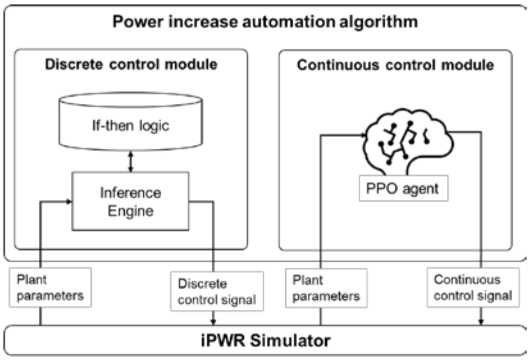

Figure 1. An overview of the power increase automation algorithm.

3.2.1. Design of network architecture using PPO

In this study, the PPO algorithm was specifically applied to Step 9. Figure 2 illustrates the network architecture of the PPO algorithm, which is composed of a critic and an actor agent, both constructed using linear layers. The algorithm takes four key variables as inputs: 1) reactor power, 2) boron concentration, 3) SUR, and 4) total reactivity. These variables are essential for monitoring during the power increase operation. The PPO algorithm analyzes the current states of these variables and selects the action with the highest probability of successfully achieving the operational goal. The available actions for managing the boron dilution process include: 1) maintaining the current state (Stay), 2) decreasing boron concentration by 1 ppm, 3) decreasing by 2 ppm, and 4) decreasing by 5 ppm. To minimize the frequency of adjustments and maintain stability, control signals are generated at 30-second intervals.

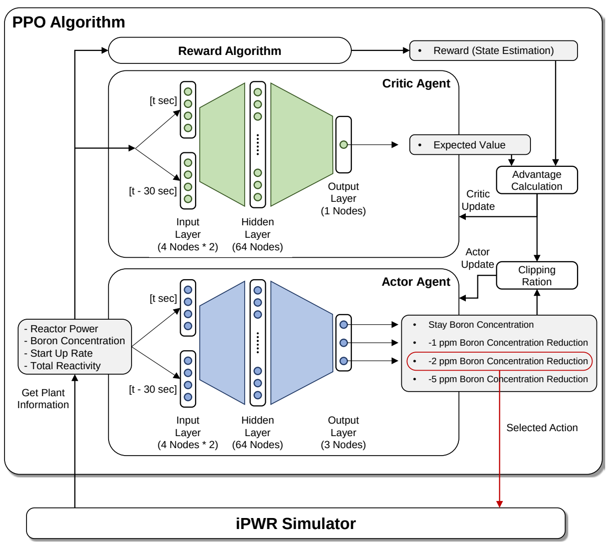

Figure 2. A network architecture of PPO algorithm for the boron dilution operation (Step 9).

3.2.2. Training of PPO algorithm

The training of the PPO algorithm was conducted over 2,500 episodes, as shown in Figure 3. Initially, the rewards exhibit significant variability, reflecting the learning agent's exploration phase. As training progresses, the rewards generally increase, indicating the agent's improving performance. While training, it was observed that the PPO algorithm successfully maintained reactor power at 8% when it achieved a cumulative reward of approximately 1,000 points. As the episodes progressed, the PPO algorithm learned that maintaining the reactor power at 8% for a longer duration resulted in higher rewards. As a result, the PPO algorithm learned the strategy to reach 8% quickly while keeping SUR, thereby optimizing its performance over time.

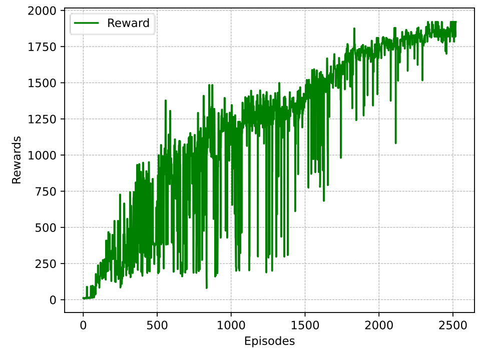

Figure 3. The cumulative reward of the PPO algorithm per episode.

3.3. Experiment Results

This section presents the results of real-time testing conducted by integrating the developed power increase automation algorithm with the iPWR simulator. Figure 4 illustrates the changes in reactor and electrical power during the automated power increase operation from 0% to 100% reactor power. Figure 5 shows the control signals generated by the automation algorithm throughout the operation.

In the initial phase of the power increase operation, from 0 to approximately 2,000 seconds, the control rods were fully withdrawn by the If-then logic. From 2,000 to 15,000 seconds, the reinforcement learning algorithm took over the operation, increasing the reactor power to around 8% while maintaining SUR within the operational constraint of 0.5 dpm. Once the reactor power exceeded 10%, the If-then logic resumed control again, ensuring a stable increase to 100% reactor power and 45 MWe electrical output.

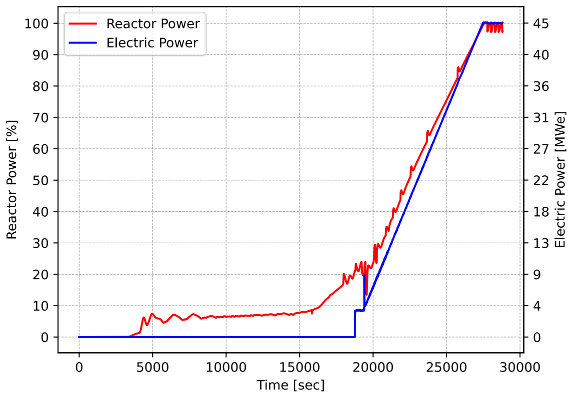

Figure 4. Changes in reactor and electrical output from 0% to 100%.

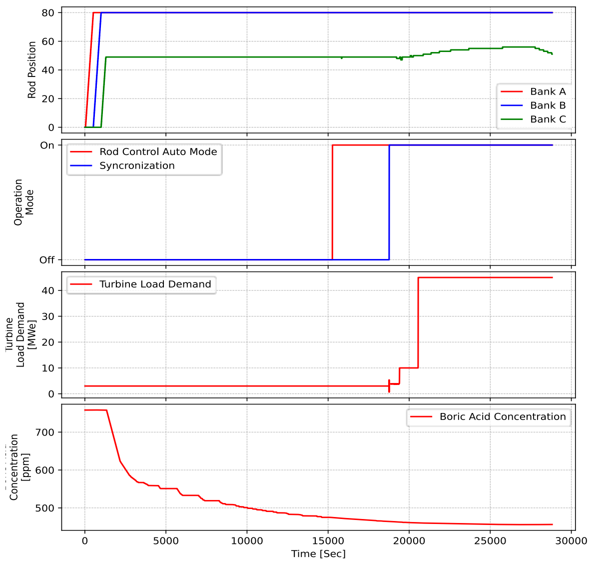

Figure 5. Discrete/continuous control results from 0% to 100%.

3. Conclusion

In this study, the power increase automation algorithm for SMR was suggested and validated using an iPWR simulator. The automation strategy was divided into discrete and continuous control modules. The discrete control tasks, which were triggered by specific conditions, were automated using If-then logic, while continuous control tasks, requiring real-time adjustments and adaptive decision-making, were managed using the Proximal Policy Optimization (PPO) algorithm. The validation of the proposed automation algorithm was demonstrated through a series of real-time simulations.

The results showed that the discrete control module successfully managed initial control tasks, such as the full withdrawal of control rods, while the continuous control module, powered by reinforcement learning, optimized reactor power increases, maintaining stability and adhering to operational constraints. This successful integration and performance of the power increase automation algorithm highlight its potential to enhance SMR operations and reduce operator workload. Building on the findings of this study, future research will focus on further optimizing the algorithm for the multi-module power increase operations in SMRs.

Acknowledgements

This work was supported by the Innovative Small Modular Reactor Development Agency grant funded by the Korea Government (MIST) (No. RS-2023-00258052).

REFERENCES

[1] J. Hartmann et al., The operator and the seven small modular reactors — An estimate of the number of reactors that a single reactor operator can safely operate, Nuclear Engineering and Design, Vol. 418, 2024.
[2] D. Lee et al., Algorithm for Autonomous Power-Increase Operation Using Deep Reinforcement Learning and a Rule-Based System, IEEE, Vol. 8, pp. 196727-196746, 2020.
[3] H. Zhaojun et al., Monte Carlo tree search-based deep reinforcement learning for flexible operation & maintenance optimization of a nuclear power plant, Journal of Safety and Sustainability, Vol 1, pp. 4-13, 2024.
[4] J. Park et al., Control automation in the heat-up mode of a nuclear power plant using reinforcement learning, Progress in Nuclear Energy, Vol. 145, 2022.