INL/CON-23-71647-Revision-0

Operation Optimization using Reinforcement Learning with Integrated Artificial Reasoning Framework

July 2023

Junyung Kim, Daniel Mark Mikkelson, Xingang Zhao, Xinyan Wang

Changing the World's Energy Future

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-23-71647-Revision-0

Operation Optimization using Reinforcement Learning with Integrated Artificial Reasoning Framework

Junyung Kim, Daniel Mark Mikkelson, Xingang Zhao, Xinyan Wang

July 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Operation Optimization Using Reinforcement Learning with Integrated Artificial Reasoning Framework

Junyung Kim[1], Daniel Mikkelson[1], Xinyan Wang[2], Xingang Zhao[3], and Hyun Gook Kang[4]

[1]Idaho National Laboratory, Idaho Falls, ID,
[2]Massachusetts Institute of Technology, Boston, MA,
[3]Oak Ridge National Laboratory, Oak Ridge, TN,
[4]Rensselaer Polytechnic Institute, Troy, NY

[leave space for DOI, which will be inserted by ANS]

ABSTRACT

In large and complex systems, operational decision-making requires a systematic analysis with a vast amount of data from both process parameters and component status monitoring. In this paper, we present an integrated artificial reasoning approach for system state transition models that can help operational decision-making with explainable and traceable reasoning. The integrated artificial reasoning framework is a physics-based approach of defining the system structure in a Bayesian network, so we leveraged it in a Markov decision process (MDP) for finding optimal operational solutions. In our proposed framework, the MDP is implemented on a dynamic Bayesian network (DBN), which represents causalities in a system. The multilevel flow modeling was utilized in order to extract these causalities in a more efficient and objective manner. Since multilevel flow modeling is based on the fundamental energy and mass conservation laws, the target system is decomposed into several mass, energy, and information structures, which serve as the basis for a DBN. The MDP consists of the processes of finding a solution for the Bellman equation, which can be derived from the conditional probability equations of the constructed DBN. System operators can capture stochastic system dynamics as multiple subsystem state transitions based on their physical relations and uncertainties coming from the component degradation process or random failures. We analyzed a simplified example system to illustrate finding an optimal operational policy with this approach.

Keywords: Reinforcement learning, Markov decision process, high-temperature gas reactor, Modelica

1. INTRODUCTION

The coordination of equipment in a nuclear power plant to safely and reliably generate power falls to the instrumentation and control systems. These systems handle many variables and provide operators with important information to make informed decisions. To maximize the usefulness of nuclear power plants, instrumentation and control systems need to simultaneously adapt to changing power demands and support non-electric applications. The challenge is managing the vast amounts of data these systems generate to make optimal decisions.

While machine learning algorithms are helpful for managing complex systems with numerous variables, their black box nature makes it difficult to understand the reasoning behind their decisions. Supervisory control systems in regulated industries, like nuclear power, require explainable machine learning results, so operators understand the algorithm's decision-making process. Reinforcement learning (RL) is a practical

method for automating the search for optimal points. Model-based RL formulates the optimization problem as a Markov decision process (MDP) [1], allowing for traceable state transitions and rewards for a clear understanding of the physical reasoning behind RL solutions. A cell-to-cell mapping approach [2] is one way to construct state transition models, which divides the system space into multiple state cells to address uncertainties. Using probabilistic mappings of the discretized system space allows for the quantification of probabilistic system evolution over time and the tracking of fault propagation [3–5]. To effectively map state transitions, it is essential to establish clear and comprehensible definitions of state cells and to control the number of system states. An integrated artificial reasoning framework (IARF) [6] transforms the physical representation of a system into a format that can be used for MDP. This study presents a decision-theoretic approach to optimizing system control logic by connecting artificial reasoning and decision-making to the state transition and reward models of MDP. We have extended our previous research, encapsulating system dynamics in the MDP models.

The paper structure is: Section 2 provides an introduction to IARF and MDP basics; Section 3 presents multidomain system modeling in Modelica language. Section 4 presents an example of our proposed approach, which involves solving an optimal operational decision-making problem of a high-temperature gas reactor (HTGR) and balance of plant (BOP) model taking into account of target components degradation; and finally, Section 4 outlines the study's conclusions.

2. MARKOV DECISION PROCESS WITH INTEGRATED ARTIFICIAL REASONING FRAMEWORK

The MDP structure is well-suited for representation with a DBN. At each time step in an MDP, there are typically three types of DBN nodes: reward ($R^{(t)}$), system state ($S^{(t)}$), and action ($A^{(t)}$), where *t* represents time step. In this study, we split the system state nodes into two categories: process variables of the system ($P^{(t)}$) and component status ($I^{(t)}$), where the component status can affect process variables. We separate these nodes to explicitly consider component failure under certain actions in the modeling. The typical MDP model's structure is shown in Figure 1. The reward node ($R^{(t)}$) at each time step only depends on the node corresponding to the current process variables. The action taken at the current time step will change the component status node's state in the next time step, while the component status nodes affect the process variable nodes in their own time step, with no temporal delay between the control input and process variable change.

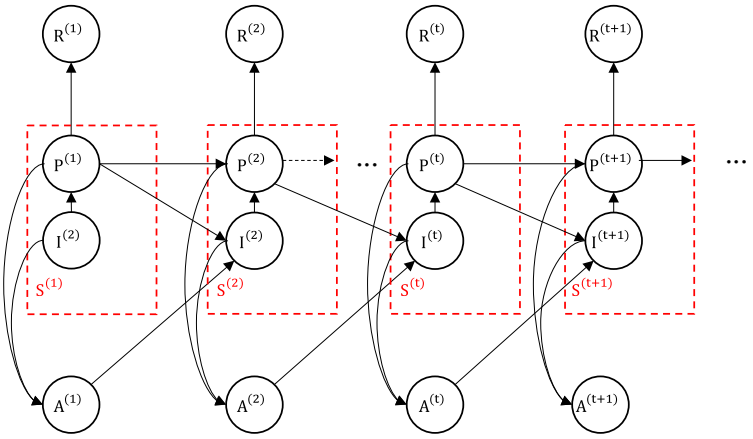


Figure 1. MDP model as a DBN, where *R*, *P*, *I*, and *A* denote state of reward, state of process variable, state of components' status, and state of actions, respectively, the superscripts are for time points, and Box *S* represents the system state.

The MDP is based on the set of Bellman equations [7]. The goal of solving the Bellman equation is to find an optimal operational policy, denoted by π, which comprises a sequence of actions from the set of actions (A), that provides the maximum state values at each time step:

$$v_\pi\left(S^{(t)} = \left\{P^{(t)}, I^{(t)}\right\}\right) = r^{(t)} + \gamma \cdot \max_a \cdot \sum_{i^{(t+1)}} \sum_{p^{(t+1)}} \Pr\left(i^{(t+1)} | a^{(t)}, p^{(t)}, i^{(t)}\right) \cdot \Pr\left(p^{(t+1)} | p^{(t)}, i^{(t+1)}\right) \cdot v_\pi\left(S^{(t+1)} = \{P^{(t+1)}, I^{(t+1)}\}\right) \quad (1)$$

where $v_\pi\left(S^{(t)}\right)$ implies a state value of $S^{(t)}$ given the optimal policy π. Since $v_\pi(S^{(t)})$ depends on $v_\pi(S^{(t+1)})$, the MDP agent will iterate over all the states and actions to solve the equation.

DBN structures can be designed using various data-driven approaches [8,9]. In this study, we utilized the IARF [6], which is a physics-based approach that selects key process variables based on causal relations among subsystems and the laws of physics to design the MDP structure. The IARF approach involves transforming system schematics into an MDP. The MDP state transition matrices aim to capture two critical aspects of a system's state: process parameters and component health. Process parameters comprise measurements such as temperature, mass flow rates, and system pressures. Meanwhile, the health state is a measure that quantifies the remaining useful life of a component. Furthermore, these representations of the system state are divided into smaller matrices of subsystems that form the entire system.

The process in the IARF involves functionally decomposing physical representation of system through multilevel flow modeling (MFM). MFM is a qualitative modeling approach that breaks down a system into its constituent subsystems and their respective flows (mass, energy, and information between system components) [10]. The MFM model is based on conservation laws and causal relationships, and its structure is then transformed into a DBN with nodes representing the mass, energy, and information states of the subsystems. The connections between nodes in the DBN, which were established based on physical laws and reasoning, can be used to explain the optimal solutions obtained. Figure 2 shows an IARF workflow.

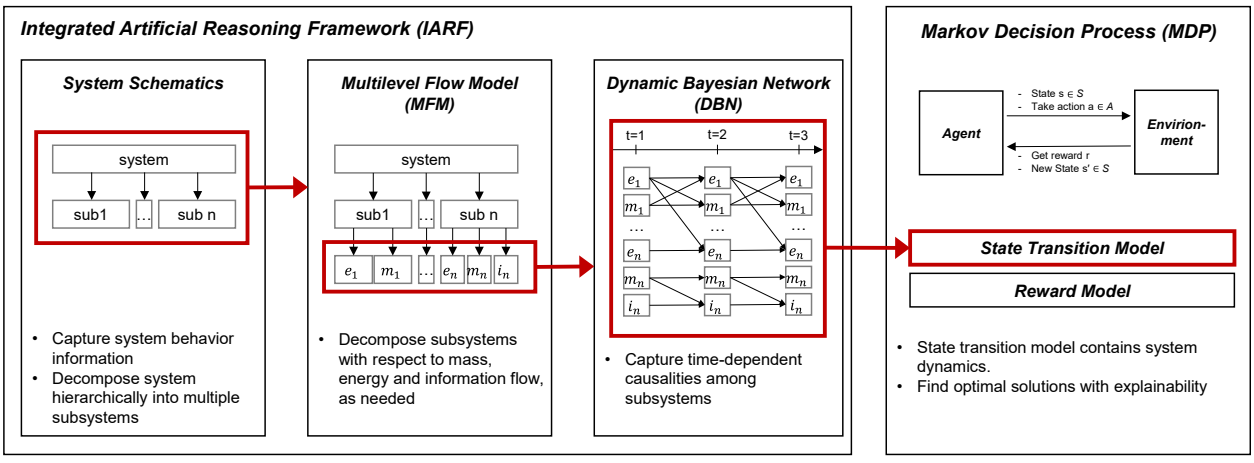


Figure 2. IARF workflow.

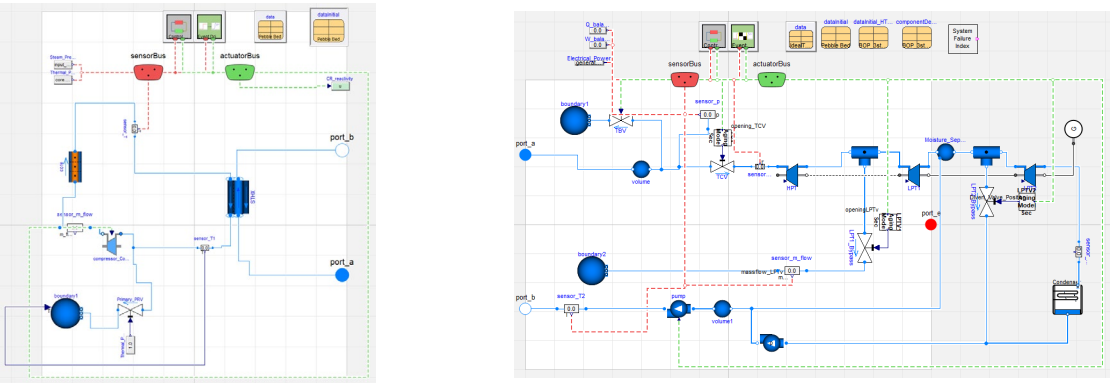
3. MULTIDOMAIN HIGH-TEMPERATURE GAS REACTOR BALANCE OF PLANT

MODELING IN MODELICA

3.1. System Models

In this research, we used the Modelica programming language to model the target system. Modelica is an open-access object-oriented programming language aimed at multidomain mathematical and physics modeling [11]. Model development with Modelica begins by defining components using their governing equations, either by hand or using preexisting commercial or open-source libraries, and then assembling components into subsystem-level models, which can then be nested further into entire systems. Commercial software, such as Dymola, which was used in this work, is a tool that provides different differential-algebraic equation solvers to solve the system of equations created by assembling components into a system, thus solving for the time evolution of the system. The library we used for HTGR modeling was HYBRID, a Modelica library of high-fidelity process models in the Modelica modeling language [12]. Within this library is an HTGR example model, which we modified to act as the steady-state base model for the demonstration cases.

Figure 3 shows the Modelica models used in this study: an HTGR reactor model and a BOP model with a three-staged turbine. The HTGR coolant is blown through the core and into a He-water heat exchanger. The coolant path exits the core and is directed to the heat exchanger. The water is boiled on the other side and is directed to a steam turbine to produce power. The BOP model with a three-staged turbine was developed when the primary modeling concern was to create a BOP model capable of capturing changes in turbine power demand [13]. After the first stage of the turbine, a T-junction pipe directs steam to either the first bypass valve (LPV-1) or the first low-pressure turbine stage. After the second stage, a moisture separator directs any liquid content to mix with and heat the feed flow sourced from the condenser. A valve (LPV-2) downstream of the moisture separator but upstream of the third turbine stage is controlled to bypass additional flow to preheat the feed flow to the desired temperature. To meet external electricity demands, the control for the HTGR-BOP must be able to function during cyclical operation ramping up and down of turbine power. The BOP control methods are summarized in Table 1, which matches operating conditions found in literature for an older iteration of X-Energy's Xe-100 design [14]. The turbine control valve operates to maintain the steam pressure at the steam generator outlet. To maintain the steam temperature, the feedwater control pump increases or decreases its speed in order to increase or decrease the pumping power. Note that the turbine power is governed by LPV-1 and that the feedwater temperature is what triggers LPV-2 to open and close.



(a) HTGR model. **(b) BOP model.**

Figure 3. HTGR reactor model [12] and BOP model with three-staged turbine [13].

Table I. Summary of Control Setpoints.

Label	Name	Controlling	Setpoint
1	Turbine control valve	Steam pressure	140 bar
2	Feedwater control pump	Steam temperature	540°C
3	Low-pressure turbine bypass valve-1 (LPV-1)	Turbine power	—
4	Low-pressure turbine bypass valve-2 (LPV-2)	Feedwater temperature	208°C

3.2. Valve Degradation Modeling

As motor operated valves (MOVs) degrade, they deviate from what the valve opening should be. For instance, a degraded turbine control valve (TCV) will only open 45% even if the control signal indicates it should open 50% (i.e., setpoint drift). This deviation may lead the steam pressure and mass flow rate to be deviated from its setpoint value. The valve health state will be defined based on the deviation level. Setpoint drift includes cases where the actuator output is outside the specified output requirement. Actuator output can change for a variety of reasons without any physical adjustment. For example, changes in the stem friction coefficient caused by stem lubricant aging can result in a reduction in actuator output [15]. We modeled valve performance degradation in Dymola assuming that we only considered the MOV setpoint drift as an indicator of component health index (i.e., degradation level) and that the health index is proportional to the cumulative hazard function of a valve-fail-to-operate event. Table 2 shows the valve degradation modeling parameters used in the model.

Table II. Probability density function of MOV's fails-to-operate event.

Component	Probability Distribution		
	Type	Parameters	
		Alpha	Beta
TCV	Gamma	2	7.87E+3
Low-pressure turbine bypass valve 1 (LPV-1)	Gamma	1.5	8.33E+3
Low-pressure turbine bypass valve 2 (LPV-2)	Gamma	2.4	8.00E+3

3.2. Steam Turbine Degradation Modeling

Steam turbine performance degradation can be expressed by the drop of the isentropic efficiency (η) with time [16]. We assumed that the isentropic efficiency decreases over time exponentially, as described in Eq. (x), and each turbine has different degradation speed, which implies different λ (HPT: 3.0e-10; LPT1: 1.0e-9; LPT2: 9.0e-10).

$$\eta^{(t+1)} = \eta^{(t)} \times \exp(-\lambda t) \tag{2}$$

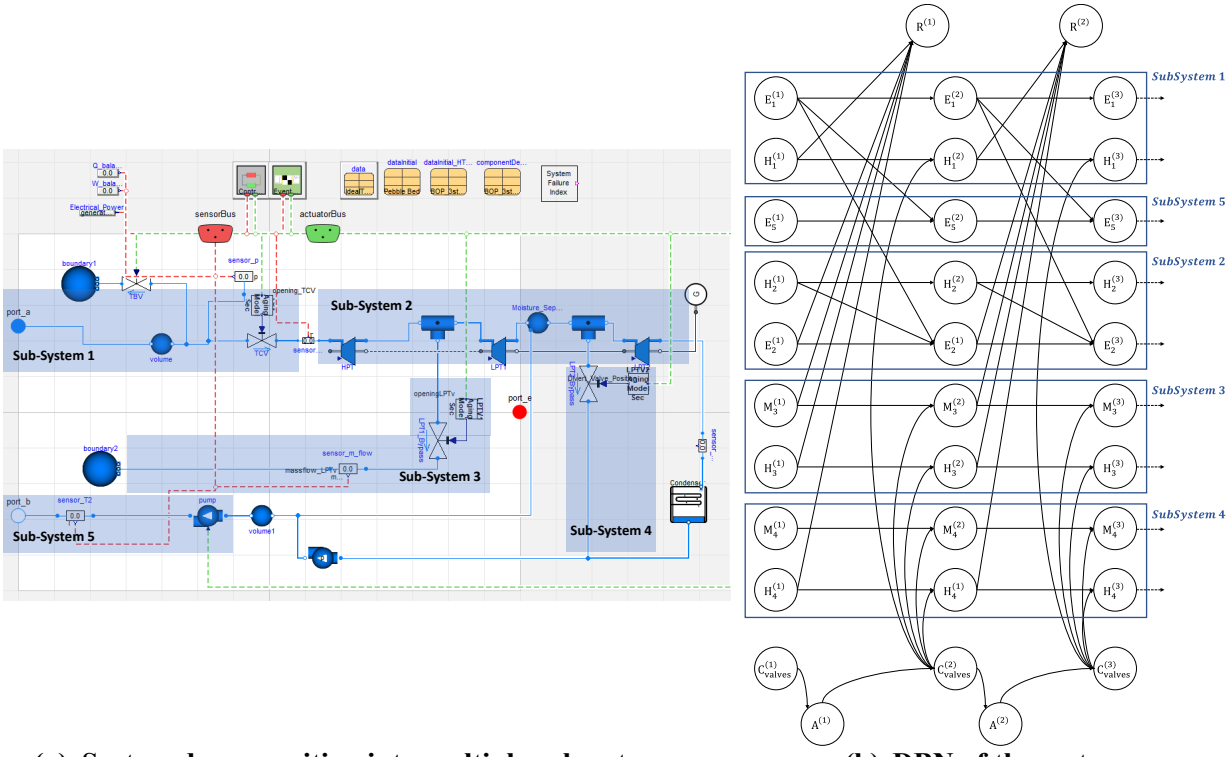
4. OPERATIONAL STRATEGY OPTIMIZATION OF A HIGH-TEMPERATURE GAS REACTOR

In this study, we defined the optimization problem as finding an optimal HTGR operational strategy considering component degradation. The objective is to maximize economic benefit from operating the reactor during 42 months of operation while considering component degradation, which may lead to

component failure and system shutdown accordingly. The system can send steam to either an electrolysis facility or steam turbine, which will generate hydrogen and electricity, respectively. The economic benefit is dependent on the amount of steam passed to those facilities and the prices of hydrogen and electricity.

There are two operational options: flexible operation following electricity demand and steady-state operation (no electrical power changes over time). Assuming that electricity prices change in a day, the economic benefit from selling electricity can be greater in a flexible rather than steady-state operational mode. We assumed that the hydrogen price is based on a long term contract and that it will not change in this case study. During flexible operation, MOVs open and close over time, which leads them to degrade. During steady-state operation, valve position does not change. The entire system must stop when a valve (i.e., TCV, LPV-1, and LPV-2) fails to operate. We will assign a negative reward for such an event in the MDP. During the steady-state operation, no valve position changes. In such case, a valve-fail-to-operate event does not need be counted. Operational decisions choosing an operational action will be made at 6, 18, 30, 42, and 54 months after the system starts running: flexible operation or steady-state operation with either 70% electricity to 30% hydrogen or 60% electricity to 40% hydrogen.

Figure 4 shows how the target system is decomposed into multiple subsystems and DBN of system is made. Constituent variables of each node in Figure 4 (b) can be found in Table 3.



(a) System decomposition into multiple subsystems. **(b) DBN of the system.**

Figure 4. IARF for MDP implementation.

Table 3. Constituent variables in each subsystem in the BOP model.

Subsystem	Constituent variables	Subsystem	Constituent variables
Energy flow of Subsystem 1 (E_1)	Steam pressure Steam temperature	Mass flow of Subsystem 3 (M_3)	Mass flow rate of LPV1

Information flow of Subsystem 1 (H_1)	TCV health status	Information flow of Subsystem 3 (H_3)	LPV1 health status
Energy flow of Subsystem 5(E_5)	Feedwater temperature	Mass flow of Subsystem 4 (M_4)	Mass flow rate of LPV2
Information flow of Subsystem 2 (H_2)	Turbine health status	Information flow of Subsystem 4 (H_4)	LPV2 health status
Energy flow of Subsystem 2 (E_2)	Entropy of HPT inlet Entropy of LPT2 outlet Enthalpy of HPT inlet Enthalpy of LPT2 outlet		

Based on the DBN of system, the Bellman equation is derived in Eq. (3) in a way that the conditional probability distribution of the system state in Eq. (1) (i.e., $\Pr\left(p^{(t+1)}|p^{(t)}, i^{(t)}\right)$) extends to the multiplication of multiple conditional probability distributions of subsystems. Causality information among subsystems captured in DBN is represented when one calculates the state value using the Bellman equation.

$$
\begin{aligned}
&\nu_\pi\left(P^{(t)} = \left\{M_1^{(t)}, E_1^{(t)}, H_1^{(t)}, E_2^{(t)}, H_2^{(t)}, M_3^{(t)}, H_3^{(t)}, M_4^{(t)}, H_4^{(t)}, E_5^{(t)}\right\}, I^{(t)} = \left\{C_{valv}^{(t)}\right\}\right) \\
&= r^{(t)} + \gamma \cdot \max_a \sum_{C_{valv}^{(t+1)}} \Pr(c_{valv}^{(t+1)}|a^{(t)}, h_1^{(t)}, h_3^{(t)}, h_4^{(t)}, c_{valv}^{(t)}) \\
&\quad \textstyle\sum_{M_1^{(t+1)}} \dots \sum_{E_5^{(t+1)}} \Pr(e_1^{(t+1)}|e_1^{(t)}) \cdot \Pr(h_1^{(t+1)}|h_1^{(t)}, c_{valv}^{(t+1)}) \cdot \\
&\quad \Pr(e_5^{(t+1)}|e_1^{(t+1)}) \cdot \\
&\quad \Pr(h_2^{(t+1)}|h_2^{(t)}) \cdot \Pr(e_2^{(t+1)}|e_1^{(t+1)}, h_2^{(t+1)}, c_{valv}^{(t+1)}) \cdot \\
&\quad \Pr(m_3^{(t+1)}|c_{valv}^{(t+1)}) \cdot \Pr(h_3^{(t+1)}|h_3^{(t)}, c_{valv}^{(t+1)}) \cdot \\
&\quad \Pr(m_4^{(t+1)}|e_5^{(t+1)}) \cdot \Pr(h_4^{(t+1)}|h_4^{(t)}, c_{valv}^{(t+1)}) \cdot \nu_\pi\left(P^{(t+1)}, I^{(t+1)}\right)
\end{aligned}
\tag{3}
$$

where $C_{valv}^{(t)}$, valves' status, can be valves position of corresponding to either flexible operation, steady-state operation (steam dispatch: 60% electricity to 40% hydrogen), steady-state operation (steam dispatch: 70% electricity to 30% hydrogen), or fail to operate.

The state transition flowchart is in Figure 5. For plot cleanliness, states related to the optimal policy, which is highlighted by the red-colored blocks and arrows, and some selected states are included. Prior to 18 months of operation, flexible operation gives us a higher reward than other operational modes. Thirty months after the system runs, flexible operation does not have any merit over steady-state operation due to valve failure.

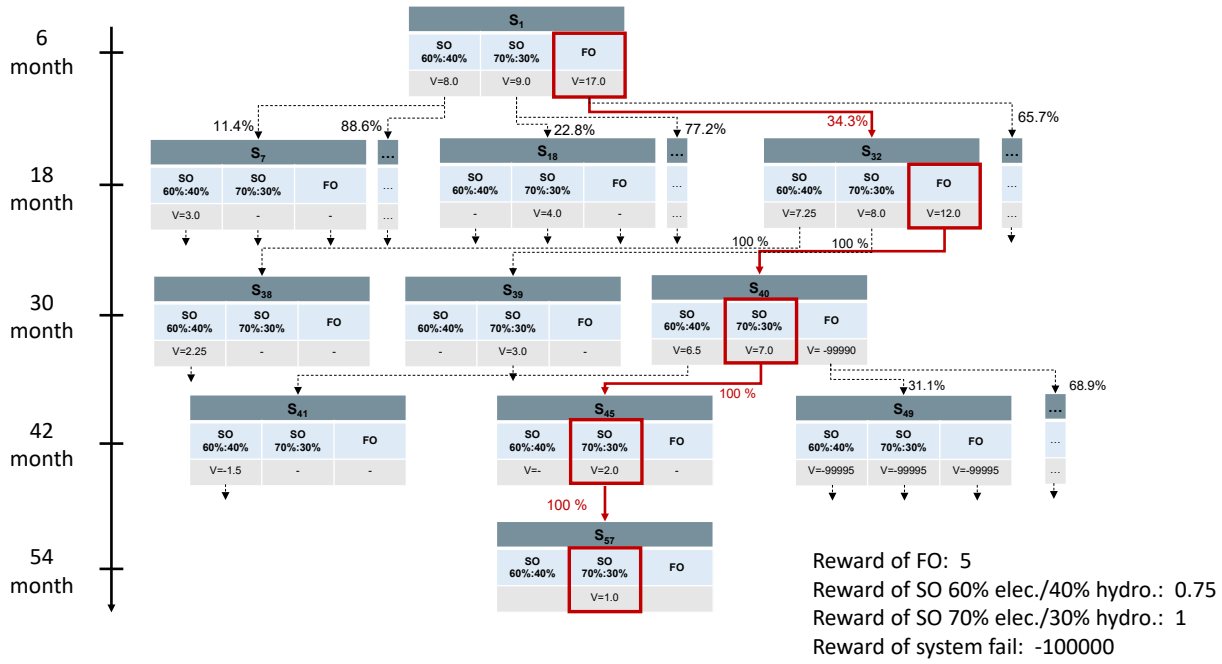


Figure 5. State transition flowchart and state-action values (V) corresponding to actions taken.

5. CONCLUSIONS

This research demonstrates the use of MDP formulated with IARF in a relevant case study for the nuclear power systems community: optimal system operation considering component degradation. The case studies have two implications that demonstrate the value of this approach beyond optimizing economic gain.

To model the system, we used the object-oriented programming language Modelica, which is gaining attention in the nuclear community. Modelica allowed an integration of detailed subsystem models to form an overall system with easily replaceable subsystems. The flexibility of this approach was demonstrated by replacing models of typical and ideal components with models of degrading behavior. This approach simplifies the modeling and data generation process for complex systems.

We have shown that IARF is an effective means of understanding complex system behaviors. Under this formulation, the system was decomposed into mass, energy, and information flows, which were converted into a probabilistic map in the form of state transition probabilities and matrices. This decomposition allows any system of complexity to be represented compactly, making it easier to interpret. This is particularly relevant in the heavily regulated and safety-oriented discipline of nuclear power.

Future work utilizing and further developing the methods demonstrated in this paper will include a more complex application and changes to the MDP formulation. Applications that relax many of the simplifying assumptions made for demonstration cases can be developed, permitting use for more realistic scenarios.

REFERENCES

1. R. E. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, USA (1957).

2. C. S. Hsu, “Cell-to-cell mapping: a method of global analysis for nonlinear systems,” *Springer Science & Business Media* **64** (2013).
3. C. Hsu, "A theory to cell-to-cell mapping dynamical systems," *Journal of Applied Mechanics*, **47**, pp. 931-939 (1980).
4. T. Aldemir, "Computer-assisted Markov failure modeling of process control systems," *IEEE Transactions on Reliability*, **36 (**1), pp. 133-144 (1987).
5. J. Kim, A. U. A. Shah, and H. G. Kang, "Dynamic risk assessment with Bayesian network and clustering analysis," *Reliability Engineering & System Safety*, **201** (2020).
6. J. Kim, X. Zhao, A. U. A. Shah, and H. G. Kang, "System risk quantification and decision making support using functional modeling and dynamic Bayesian network," *Reliability Engineering & System Safety*, **215**, pp. 107880 (2021).
7. I. H. Haff, K. Aas, A. Frigessi, and V. Lacal, "Structure learning in Bayesian Networks using regular vines," *Computational Statistics & Data Analysis*, **101**, pp. 186-208 (2016).
8. A. M. Hanea, D. Kurowicka, R. M. Cooke, and D. A. Ababei, "Mining and visualising ordinal data with non-parametric continuous BBNs," *Computational Statistics & Data Analysis*, **54** (3), pp. 668-687 (2010).
9. M. Lind, "An introduction to multilevel flow modeling," *Journal of Nuclear Safety and Simulation*, **2**, pp. 22-32 (2011).
10. R. Bellman, "On the theory of dynamic programming," *Proceedings of the National Academy of Sciences* (1952).
11. M. Tiller, ed., “Introduction to physical modeling with Modelica,” *Springer Science & Business Media*, **615** (2012).
12. D. M. Mikkelson, et al.. “Status report on FY2022 model development within the integrated energy systems HYBRID repository.” INL/EXT-21-65432-Rev000, Idaho National Laboratory, Idaho Falls, ID (2021).
13. D. M. Mikkelson, et al.. “Status Report on Thermal Extraction Modeling in HYBRID.” INL/EXT-23-03062-Rev000, Idaho National Laboratory, Idaho Falls, ID (2021).
14. Y. Brits, F. Botha, H. v. Antwerpen, and H.-W. Chi, “A control approach investigation of the Xe-100 plant to perform load following within the operational range of 100 – 25 – 100%,” *Nuclear Engineering and Design*, **329**, pp. 12-19 (2018).
15. T. Wierman, D. Rasmuson, and N. Stockton, “Common-Cause Failure Event Insights: Motor-Operated Valves (NUREG/CR-6819 Vol.2),” Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID (2003).
16. Najjar, Yousef SH, Osama FA Alalul, and Amer Abu-Shamleh. “Steam turbine bottoming cycle deterioration under different load conditions.” *Thermal Science and Engineering Progress* **20**, pp. 100733 (2020).