Conceptual Framework for Real-time Threat Detection in NPP Control Systems using Logic-based Anomaly Monitoring and Dynamic Bayesian Network

Jae Hwan KIM*, Kwang Seop SON, Jae Gu SONG, Ju Hyung SONG, Young Jun LEE
Security R&D Team, Korea Atomic Energy Research Institute, 111, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea
*Corresponding author: jhkim85@kaeri.re.kr

Keywords : instrumentation and control (I&C) systems, cybersecurity, logic-based anomaly monitoring (LAM), dynamic bayesian network (DBN)

1. Introduction

Instrumentation and control (I&C) systems in nuclear power plants are divided into safety and non-safety categories [1]. While research has traditionally focused on safety systems, non-safety controls can also play a decisive role in initiating reactor trips. Stable operation under disturbances such as load rejection, turbine trip, or auxiliary system faults depends on the proper functioning of these non-safety controls. Their malfunction may inadvertently trigger reactor protection actions.

The Steam Bypass Control System (SBCS) is a representative case. It regulates main steam pressure during normal operation and, in transients, controls Turbine Bypass Valves (TBVs) to dump steam into the condenser, thereby maintaining the thermal stability of the steam generators and the Nuclear Steam Supply System (NSSS). Failure of the SBCS may cause excessive steam pressure rise, MSSV actuation, or reactor trip conditions [2].

Recent concerns over cyber security highlight the risk of intentional manipulation of control logic or signals in non-safety systems [3]. Existing monitoring schemes mainly hardware checks or single-device anomaly detection are insufficient to discriminate between equipment faults and cyberattacks. To address this gap, this study proposes a logic-based anomaly detector for SBCS demand signals, combined with controller diagnostic signals within a Dynamic Bayesian Network (DBN) framework. The DBN provides real-time posterior probabilities of controller states (Normal, Fault, and Attack). This work aims to present a high-reliability concept for real-time anomaly detection, supporting enhanced cyber security of nuclear I&C systems.

2. Target System and Functions

The Steam Bypass Control System (SBCS) is a subsystem of the NSSS Process Control System (NPCS) in the APR-1400, classified as a non-safety I&C system. Its primary role is to regulate main steam pressure and thereby maintain the energy balance of the NSSS. The system controls Turbine Bypass Valves (TBVs), ensuring stable steam pressure during normal operation and providing critical functionality for rapid load change management during transients.

This study focuses on four control functions of the SBCS that have direct potential to induce reactor trips:

- TBV Quick-Open DEMAND
- Reactor Power Cutback (RPC) DEMAND
- Automatic Motion Inhibit (AMI) DEMAND
- Automatic Withdrawal Prohibit (AWP) DEMAND

The TBV Quick-Open DEMAND is activated during sudden load rejection or turbine trip events, commanding multiple TBVs to open rapidly and prevent excessive steam pressure rise. Its failure or delay may result in MSSV actuation or reactor trip by the protection system. The RPC DEMAND reduces reactor power swiftly through the Reactor Power Cutback System (RPCS) when load reduction exceeds TBV bypass capacity; its malfunction increases the likelihood of reactor protection actuation. The AMI DEMAND inhibits automatic control rod motion under certain operating conditions, preventing unnecessary rod movement. Anomalies in this signal may cause reactor power fluctuations and spurious trips. The AWP DEMAND blocks automatic control rod withdrawal during transients, thereby avoiding inappropriate positive reactivity insertion. Its failure may lead to rapid power escalation and subsequent protection system intervention.

All four functions are tightly linked to reactor trip prevention. Therefore, ensuring integrity and reliability not only at the system level but also at the functional signal level is essential. This study proposes a methodology to detect inconsistencies between each DEMAND signal and its intended reference logic, classifying the cause as either a controller fault or a cyber threat.

3. Process Logic-based Anomaly Detection

The first key element proposed in this study is the Logic-based Anomaly Monitor (LAM). The LAM replicates the design logic of the control system in an independent, isolated platform. It receives the same input signals as the actual controller and executes the logic in parallel. By comparing the outputs of the reference logic

with those of the controller, the LAM detects inconsistencies that indicate abnormal behavior.

Unlike conventional signal-level monitoring, this approach treats the full control logic as a white-box reference model, thereby enabling detection at the functional level. For instance, in the SBCS, a TBV Quick-Open DEMAND signal must be generated when specific conditions (e.g., rapid steam pressure rise, turbine trip) are met. The LAM checks whether the outputs of the reference logic and the actual controller align under identical inputs. Any discrepancy is flagged immediately as an anomaly.

Such inconsistencies may arise from different root causes, including:

- Hardware faults: unresponsive I/O modules, CPU processing delays, communication link failures.
- Software faults: logic mismatches between redundant controllers, timing errors, internal variable miscalculations.
- Cyber threats: forced insertion or blocking of DEMAND signals, sequence distortion, timestamp manipulation.

The LAM itself does not attempt to classify whether an anomaly originates from a fault or a cyberattack. Instead, it generates a residual signal, representing the degree of mismatch between reference logic and controller outputs. This residual, together with additional diagnostic signals, serves as a key input to the subsequent DBN-based state estimation module, which categorizes the controller state into Normal, Fault, or Attack.

Formally, the residual can be expressed as:

$$r(t) = f\big(y_{ref}(t), y_{ctrl}(t)\big) \tag{1}$$

Where, $y_{ref}(t)$: output of the reference logic (LAM), $y_{ctrl}(t)$: output of the actual controller, $f(\cdot)$: comparison function depending on the signal type.

Specifically, for Boolean signals, residuals are defined as:

$$r(t) = \begin{cases} 0, y_{ref}(t) = y_{ctrl}(t) \\ 1, y_{ref}(t) \neq y_{ctrl}(t) \end{cases} \tag{2}$$

and for REAL/SINT signals, residuals are defined as:

$$r(t) = y_{ref}(t) - y_{ctrl}(t) \tag{3}$$

In this study, the residual definitions were limited to Boolean, REAL, and SINT signals, as these represent the actual data types generated by the target controller and accessible through the engineering workstation. Other data types (e.g., INT, DINT, DOUBLE) are not used in the system, and thus were not considered in this conceptual framework.

The anomaly decision is determined by applying a threshold θ:

$$|r(t)| > \theta \rightarrow \text{Anomaly Detected} \tag{4}$$

For Boolean signals, a fixed threshold (θ = 0.5) ensures immediate detection of mismatches, whereas for REAL/SINT signals, the threshold is set considering sensor precision and allowable tolerance.

Figure 1 illustrates the block diagram of the LAM concept. The controller inputs are simultaneously provided to both the reference logic and the actual controller. Their outputs are compared to produce a residual signal, which is then supplied to the DBN-based state estimator along with other diagnostic information.

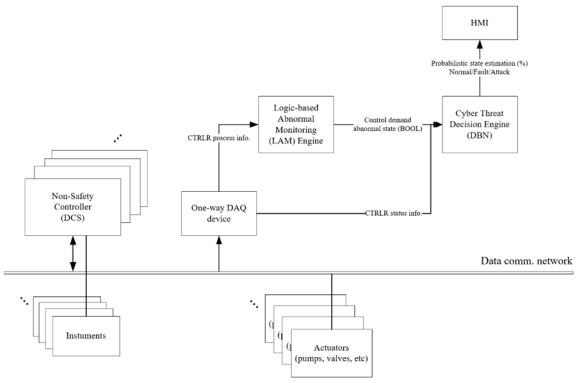

Fig. 1. Block diagram of LAM and DBN for Real-time controller threat state estimation

4. DBN-based Real-time Threat State Estimation

In this study, a DBN was employed to enable real-time estimation of controller states. A DBN provides a probabilistic framework to model the temporal relationship between hidden states and observable evidence, making it particularly suitable for signal interpretation under various sources of uncertainty [4].

In our formulation, the hidden state variable is defined with three categories: Normal, Fault, and Attack. At each time step, the DBN computes the posterior probability of the controller's state based on the set of observations.

4.1 Modeling Environment and Tools

The DBN model was implemented using the GeNIe/SMILE platform. GeNIe supports both visual model construction and probabilistic inference, enabling intuitive design even for complex I&C systems with a large number of observation variables [5]. In this study, the DBN input nodes were configured to include multiple categories of evidence related to the SBCS controller: internal diagnostic signals, sensor quality indicators, and anomaly residuals generated by the LAM.

4.2 Observation Nodes

The observation nodes of the DBN can be classified into three major groups:

Logic-based anomaly residuals: Discrepancies between the reference logic output and the actual controller output for each DEMAND function.

Internal diagnostic signals: Indicators such as CPU utilization, logic mismatch between redundant controllers, communication port status, memory integrity, and operator authority transitions.

Sensor Quality (SQ) signals: Quality states of key instrumentation, including pressurizer pressure, steam generator flow, main steam pressure, and turbine bypass valve positions.

4.3 Hidden States and State Transition Modeling

In the proposed DBN, the hidden state variable represents the operational status of the SBCS controller, which can take one of three mutually exclusive categories: Normal, Fault, or Attack. These states are not directly observable, but are inferred from the probabilistic relationship between the hidden states and the observation nodes introduced in Section 4.2.

The temporal dynamics of the hidden state are captured by the state transition probability matrix. In general, the model assumes high self-persistence (e.g., Normal → Normal with high probability), while transitions to abnormal conditions (Normal → Fault or Normal → Attack) are modeled with lower probabilities. Similarly, recovery transitions (e.g., Fault → Normal) are allowed but occur with relatively low likelihood, reflecting realistic controller behavior. Transitions between Fault and Attack are also possible but are defined separately, since hardware/software faults and intentional cyber intrusions typically exhibit distinct temporal signatures.

The observation model links the hidden state to the three groups of observation nodes:

- Logic-based anomaly residuals are strongly indicative of deviations from intended control logic, and their abnormal patterns increase the likelihood of either Fault or Cyberattack.
- Internal diagnostic signals primarily reflect hardware/software malfunctions, and thus contribute more strongly to distinguishing Fault from Cyberattack.
- Sensor Quality (SQ) signals provide contextual evidence that helps confirm whether anomalies are consistent with physical process faults or artificially injected disturbances.

The posterior probability of each hidden state at time t is computed using standard DBN inference algorithms as:

$$P(X_t|O_{1:t}) \propto P(O_t|X_t)\sum_{X_{t-1}} P(X_t|X_{t-1})P(X_{t-1}|O_{1:t-1}) \quad (5)$$

where X_t is the hidden state (Normal, Fault, and Attack), and O_t represents the set of observation nodes at time t. This recursive formulation enables real-time estimation of controller status, continuously updating the belief state as new evidence becomes available.

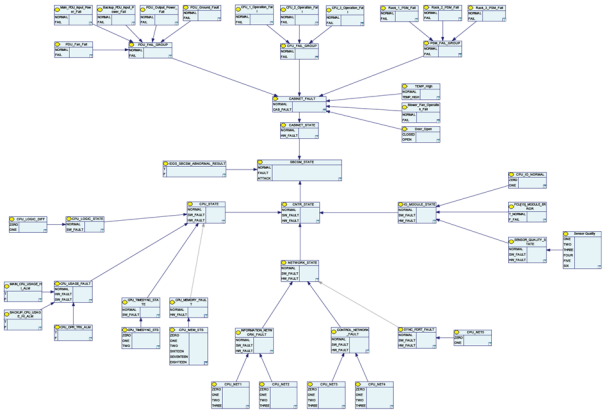

Fig. 2. Conceptual DBN state estimation model for SBCS threat detection

5. Conclusions

This study proposes a conceptual framework for detecting controller faults and cyber threats in the Steam Bypass Control System (SBCS) of APR-1400. By combining a LAM with a DBN, the approach enables integrity verification of critical DEMAND signals and probabilistic classification of controller states into normal, fault, or attack.

The proposed method extends beyond simple signal monitoring by leveraging process logic for anomaly detection and probabilistic reasoning for real-time state estimation, thereby offering potential support for operator decision-making. This study conceptually proposes a real-time threat detection framework combining LAM and DBN. Future work will enhance practicality through data-driven learning and validation.

REFERENCES

[1] R. T. Wood, R. A. Joseph III, K. Korsah, M. D. Muhlheim, and J. A. Mullens, Classification Approach for Digital I&C Systems at U.S. Nuclear Power Plants, Letter Report, LTR/NRC/RES/2012-001, Oak Ridge National Laboratory, prepared for the U.S. Nuclear Regulatory Commission, February 2012.

[2] Korea Electric Power Corporation and Korea Hydro & Nuclear Power Co., Ltd., APR1400 Design Control Document Tier 2, Chapter 7: Instrumentation and

Controls, APR1400-K-X-FS-13002, Rev. 0, September 2013.

[3] S. L. Eggers, The Nuclear Digital I&C System Supply Chain Cyber-Attack Surface, INL/CON-20-57213-Revision-0, Idaho National Laboratory, June 2020.

[4] N. Zhou, Z. Huang, D. Meng, S. Elbert, S. Wang, and R. Diao, Capturing Dynamics in the Power Grid: Formulation of Dynamic State Estimation through Data Assimilation, PNNL-23213, Pacific Northwest National Laboratory, March 2014.

[5] BayesFusion, LLC, *GeNIe Modeler (Version 5.0)*, Pittsburgh, PA: BayesFusion, LLC, 2020.