Deep Autoencoding Gaussian Mixture Model for Detection of Manipulation Attacks on NPP Process Data

Junyong Bae, Ji Hyeon Shin, Seung Jun Lee*
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

Abstract: With recent advances in artificial intelligence, research into the application of these techniques to nuclear power plant operation has flourished over the past decade. One promising application is the detection of anomalies in plant parameters. In cases where anomaly data for training is scarce, such as detecting manipulation of plant process data by cyberattacks, unsupervised learning methods using only normal data are a possible option. In particular, deep autoencoder has gained wide acceptance for several such applications. However, the autoencoding approach, which trains the model to reconstruct normal data and identifies anomalies based on reconstruction errors, underutilizes the compressed information encoded in latent space. To overcome this limitation and develop a model for detecting manipulation attacks on plant process data, we adapt a Deep Autoencoding Gaussian Mixture Model (DAGMM) that exploits both reconstruction errors and latent space features of normal data. The DAGMM has been applied to the scenarios of plant parameter manipulation attacks during emergency operations and the results demonstrate its improved performance in terms of anomaly detection sensitivity and detection times compared to the sole autoencoder model.

Keywords: Cybersecurity, Deep learning, Autoencoder, Data manipulation.

1. INTRODUCTION

The transition from analog to digital instrumentation and control (I&C) systems has been implemented across various infrastructures due to several advantages, including enhanced reliability, expedited computation, ease of function implementation via software, and increased data storage capacity. Accordingly, the I&C systems of nuclear power plants (NPPs), which represent safety-critical infrastructures, have been digitized to manage various plant functions such as reactor protections, component controls, and human-system interactions. For instance, at the Advanced Power Reactor 1400 in Korea, safety-related systems like the reactor protection system and non-safety-related systems such as the process control systems for the nuclear steam supply system have been implemented using programmable logic controllers (PLCs) and distributed control system platforms, respectively. In the main control room (MCR), large display panels indicate measurements of plant parameters, while individual operator consoles display computerized procedures. These systems are interconnected through a network and share their process variables.

This digitization and network interconnection of NPP I&C systems have raised substantial cybersecurity concerns. Although the I&C systems of safety-critical infrastructures are generally isolated from external networks, historical incidents have demonstrated that these systems can be compromised through various attack vectors. A notable example is the Stuxnet attack on Iran's uranium enrichment facility. Initially, Stuxnet infected computers running the Windows operating system and spread through USB sticks or local networks. Upon connection to PLCs from specific manufacturers, Stuxnet injected malicious code into the controllers, which can manipulate control signals. At the enrichment facility, the injected code periodically altered the speed of centrifuge motors from high to low and back to high while mimicking normal sensor outputs to conceal the attack from facility operators. Consequently, this attack mechanically damaged 984 centrifuges, significantly impacting Iran's nuclear development program. Stuxnet exemplifies an advanced persistent threat (APT) employing specialized attack vectors and strategies tailored to the target system to achieve its goal, including causing physical damage to the facilities [1, 2].

In addition to the Stuxnet attack, historical cybersecurity incidents at the Davis-Besse [3], Brown Ferry [4], and Hatch [5] NPPs further underscore the significant possibility of cyberattacks on the digital I&C systems of NPPs, even when isolated from external networks. In terms of consequences, if a cyberattack compromises critical safety functions such as reactivity control and core heat removal, it could result in severe outcomes. Even without core damage or a radiological release, cyberattacks on NPPs could cause considerable economic losses by stopping plant operations.

To address such risks, regulatory agencies have established standards and issued guidelines for NPP cybersecurity. For instance, the U.S. Nuclear Regulatory Commission has published Cyber Security Rule 10 CFR 73.54 [6] and regulatory guide 5.71 (RG 5.71) [2, 7]. Similarly, the Korea Institute of Nuclear Nonproliferation and Control (KINAC) has issued KINAC regulatory standard-015 (RS-015). These guidelines primarily focus on constructing a defensive architecture to mitigate various attack vectors and disruptions caused by intentional cyberattacks. In this architecture, critical digital assets (CDAs) are identified and classified into several security levels according to their importance for the safety, security, and emergency preparedness (SSEP) of the plants. CDAs with high level of security are protected by more rigorous security measures, and the network connections between assets at different levels are secured by firewalls or segregated by a single-direction data flow (i.e., from high to low security levels) [2, 8]. In accordance with these efforts, research is being conducted with the objective of more effectively classifying the CDAs through risk assessment techniques such as Bayesian networks [9] and probabilistic safety assessments [10, 11].

However, if a well-motivated attacker launches APTs or more sophisticated attacks using national-level resources, the attack may infiltrate the I&C system without being stopped and compromise plant functionality in various ways. In such scenarios, plant operators can still mitigate the impact of compromised plant functionality by utilizing multiple manual backup functions within the I&C system. Furthermore, several highly safety-related backup functions are even implemented using hardwired connections, which are inherently immune to cyberattacks.

Nevertheless, if a cyberattack simultaneously targets the situational awareness of plant operators, such recovery responses may be hindered. For instance, if the information processing system is compromised, the information displayed to operators in the MCR can be manipulated. This data manipulation attack can prevent operators from recognizing the current cyberattack situation and ultimately hinder timely recovery responses [12]. In fact, in 2003, the safety parameter display system of the Davis-Besse NPP was disabled for nearly five hours due to the Slammer worm [13, 14]. Although there was no immediate threat to reactor safety, this cybersecurity provides evidence of the potential for manipulation attacks and their impact on plant operators.

Furthermore, these data manipulation attacks have the potential to induce human error among MCR operators. This vulnerability has been discussed in several studies, including [10, 11, 15, 16]. In situations where timely responses from operators are necessary, such as during emergencies or abnormal operations, cyberattack-induced human errors can be especially critical. Although inadequate information was not caused by a cyberattack, the Three Mile Island (TMI) accident demonstrates that this inadequate information (i.e., pressurizer water level) can result in operator confusion and human error (i.e., termination of emergency core cooling system) in emergency situations, ultimately leading to core damage.

Therefore, alongside defensive architecture, a cyberattack detection system, including monitoring of process data manipulation, needs to be deployed concurrently. Several studies have proposed such detection systems. Zhang et al. proposed a system composed of three detection modules in accordance with the defense-in-depth concept [2, 17]. In the situation that the defensive architecture has been penetrated, detection modules 1 and 2 (M1 and M2, respectively) attempt to identify the cyberattack based on network data. Specifically, M1 classifies current network patterns into known patterns from the cyberattack database, utilizing machine learning classification models such as k-nearest neighbor, bootstrap aggregating, and random forest [17]. In the case of M2, the module compares current network flow patterns with trained normal patterns using an auto-associative kernel regression (AAKR) model, which predicts network flow data by training on statistical data of network flow in an unsupervised manner [2]. To address process data manipulation attacks, module 3 (M3) is employed. Similar to M2, AAKR model is utilized and trained on normal process data under normal operating conditions [17]. The multilayer detection system was tested on a cyber-physical system containing a loop facility to mimic the thermal-hydraulic system of a typical two-loop light water reactor and a supervisory control and data acquisition system that receives sensor data from and sends commands to the loop facility [2, 17]. Zhang et al. also deploy the AAKR model and additionally, a support vector regression model in a localized cyberattack detecting module for PLCs [18].

Similarly, Lee et al. proposed a detection method for manipulation attacks on NPP process data to secure the situational awareness of MCR operators during emergency operations [14]. In this method, the Kalman filter algorithm estimates the current process state variables, and the manipulated variables are identified by

comparing the observed values with the estimated values. The proposed method was tested on a hardware-in-the-loop (HIL) system, which consisted of an NPP simulator, a physical system with water pumps, valves, and storage tanks, and control components [19]. As a result, the method was able to not only detect manipulation on a safety-related process variable but also successfully predict the current value of the variable.

As demonstrated in these studies, detection methods that use data-driven models trained on normal operating records to predict estimated values of plant parameters and then compare these estimates with current values have been intensively researched. Such methods have also been utilized for anomaly detection during plant operation. In this domain, autoencoders have been intensively utilized as prediction models [20-22]. Like other models, including the AAKR model and Kalman filter algorithm, training an autoencoder requires only normal operating records. This is beneficial in cases where abnormal records are scarce or hard to obtain from the plant simulator, including manipulation attacks, as there are numerous potential manipulations depending on the intentions of the attacker. In addition, as deep-learning models, autoencoders can effectively consider the relationships between parameters and monitor dozens of parameters simultaneously. The aforementioned advantages have led to the deployment of an anomaly detection system based on deep autoencoders at the research reactor in Korea [23].

This study proposes the methods for detecting anomalies in the plant process data that may be manipulated by cyberattack. Building on the previous studies on anomaly detection for NPPs, we also employ the autoencoding approach. However, the proposed method utilizes a deep autoencoding Gaussian mixture model (DAGMM), which monitors not only the residuals between the estimates and current values, but also the latent space of the autoencoder [24]. The proposed DAGMM-based anomaly detection model was evaluated scenarios of plant parameter manipulation attacks during emergency operations. The results demonstrated its enhanced performance in terms of anomaly detection sensitivity and detection times compared to the sole autoencoder model.

This paper is structured as follows. Section 2 provides related concepts, including autoencoder, Gaussian mixture model, and DAGMM. Section 3 presents our investigation of a latent space of trained autoencoder and introduces a DAGMM-based anomaly detection model. Section 4 explains the assumed manipulation attack scenario for model testing and the result. Finally, conclusions and perspectives are presented in Section 5.

2. RELATED CONCEPTS

2.1. Autoencoder

An autoencoder is a type of artificial neural network designed to learn compressed representations of input data, typically for the purposes of dimensionality reduction or anomaly detection. Figure 1 illustrates the structure of an autoencoder. The network comprises two main components: the encoder and the decoder. The encoder compresses the input data (X) into a lower-dimensional representation (Z) in a latent space, while the decoder reconstructs the original data ($\hat{X}$) from this compressed representation (Z). This process forms a "bottleneck" structure, , where the latent space in the middle has reduced dimensionality compared to the input and output layers. This bottleneck forces the network to distill the essential aspects of the data into the latent space representation (i.e., dimensionality reduction)

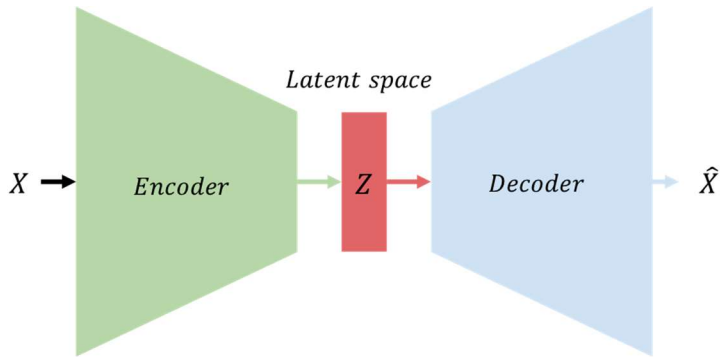


Figure 1. Structure of autoencoder

The training process of an autoencoder involves the minimization of the reconstruction error, which is the difference between the input data (X) and the reconstructed data ($\hat{X}$). When the input data closely resembles the training data, the autoencoder is capable of accurately reconstructing it. However, if the input data significantly differs from the training data, the reconstruction error will be higher. By organizing the training data with typical or normal data, the unusual or anomalous data can be identified by monitoring the difference between the input and reconstructed data (i.e., anomaly detection).

2.2. Gaussian Mixture Model

A Gaussian mixture model (GMM) is a probabilistic density function represented as a weighted sum of multiple Gaussian distributions, as shown in Eq. (1),

$$p(X) = \sum_{k=1}^{K} \phi_k \, \mathcal{N}(X \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

where X is data point, ϕ_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mixing coefficient, mean vector and covariance matrix, respectively, for k^{th} Gaussian distribution [25]. As illustrated in Fig. 2., GMMs can be utilized to represent the unknown complex distribution of given data points. Typically, GMMs are fitted to the unknown data distribution by the expectation-maximization (EM) algorithm. In this algorithm, the parameters ϕ_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ for each Gaussian distribution are iteratively adjusted to maximize the expectation [i.e., sum of likelihood $p(X)$ for entire data points] for the training data sets. The trained GMM can be employed as an anomaly detection model since the likelihood of the data points that deviate from the trained ones will be low.

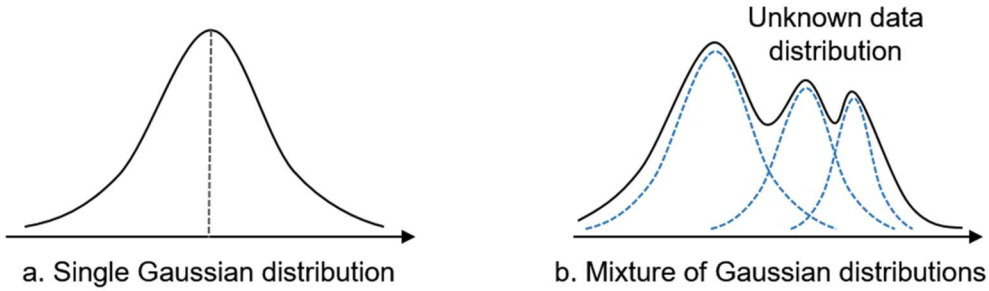


Figure 2. Mixture of multiple Gaussian distributions for representing unknown complex distribution.

2.3. Integration of Autoencoder and Gaussian Mixture Model

Clustering, including GMM, is a robust anomaly detection method that does not require human supervision. However, it can be limited when the input data is high-dimensional and multivariate. To address this, dimensionality reduction followed by density estimation has been widely adopted. In this approach, clustering is conducted on the low-dimensional representation in the latent space of the dimensionality reduction model, rather than on the input data itself. However, the performance of this composite method can be suboptimal because the dimensionality reduction and clustering models are trained independently, without any information about their ultimate objective.

To overcome this limitation and improve the performance of the composite approach, DAGMM has been proposed [24], In this model, the autoencoder and GMM are trained jointly with unified learning objectives. The model consists of two deep-learning components: the compression network (i.e., autoencoder) and the estimation network. The compression network produces a low-dimensional representation of the input data $Z = [z_c, z_r]$, as shown in Fig. 3. Unlike a typical autoencoder, the representations produced by the DAGMM include not only the distance between the input and reconstruction z_r measured by Euclidean distance or cosine similarity, which is typically used for anomaly detection with a standalone autoencoder, but also the latent space z_c.

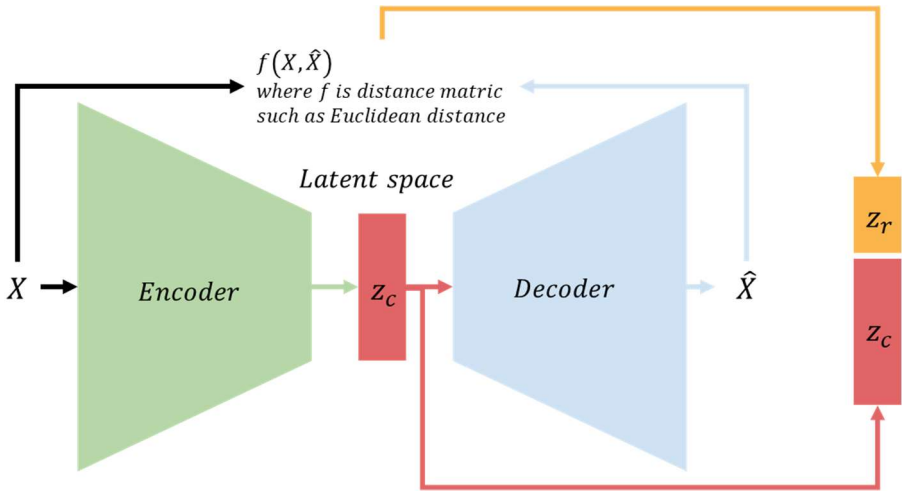


Figure 3. Structure of compression network of DAGMM.

To achieve joint training instead of using the EM algorithm, this method leverages a separate estimation network that can indirectly infer the GMM parameters ϕ_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$. As illustrated in Fig. 4, this network predicts the memberships, γ, for each Gaussian distribution given the low-dimensional representations.

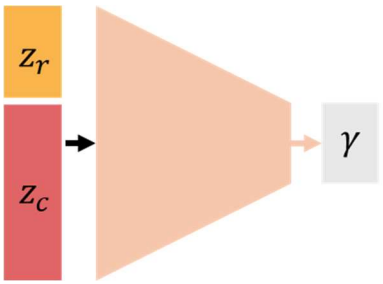


Figure 4. Structure of estimation network of DAGMM.

For example, if the output vector is $[0.7, 0.2, 0.1]$, it indicates that the current input is attributed to 70% of Gaussian distribution A, 20% to B, and 10% to C. By using these predicted memberships for all training data points, the GMM parameters can be inversely calculated, as shown in Eqs (2-4),

$$\phi_k = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N} \tag{2}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} Z_i}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \tag{3}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} (Z_i - \boldsymbol{\mu}_k)(Z_i - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \tag{4}$$

where N is the number of data points, and $\hat{\gamma}_{ik}$ is the predicted membership of the i^{th} data point for the k^{th} distribution. Using these parameters, the negative log-likelihood, or "energy", for any data point can be computed using Eq. (5). Like the EM algorithm, by minimizing this energy across all training data points, anomalies that deviate significantly from the distribution of the training data can be identified as they typically exhibit higher energy values.

$$E(Z) = -\log p(Z) = -\log \left[\sum_{k=1}^{K} \phi_k \, \mathcal{N}(Z | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right] \tag{5}$$

These two networks are jointly trained with a single objective (i.e., loss function), as shown in Eq. (6), where θ_C and θ_E represent the trainable parameters of the compression and estimation networks, respectively, $L(X_i, \hat{X}_i)$ denotes the reconstruction error, $P(\mathbf{\Sigma}_k)$ is the penalization term aimed at preventing overfitting of the distribution to several data points, and λ_1 and λ_2 are weights assigned to each loss term. In this manner, the autoencoder compresses the input data X into Z n a manner that is more suitable for fitting the GMM.

$$J(\theta_C, \theta_E) = \frac{1}{N}\sum_{i=1}^{N} L(X_i, \hat{X}_i) + \frac{\lambda_1}{N}\sum_{i=1}^{N} E(Z_i) + \lambda_2 P(\mathbf{\Sigma}_k) \quad (6)$$

In addition, during the implementation phase with the learned GMM parameters, anomalies are identified by estimating the energy of each sample. The data points with high energy values, exceeding a pre-defined threshold, are labeled as anomalies.

3. DEEP AUTOENCODING GAUSSIAN MIXTURE MODEL FOR DETECTING DATA MANIPULATION ATTACKS

In this research, we applied a DAGMM based on our investigation into anomaly detection models using an autoencoder. We trained the autoencoder model on operational data from NPP simulators during emergency operations, which included operator actions [26] and tested the trained autoencoder on assumed data manipulation attack scenarios. The reconstruction error was computed using mean squared error, and the detection threshold was set to achieve 99% specificity on the training data.

The model successfully detected most manipulations, but there were instances of detection failure, as shown in Fig. 5. During these instances (highlighted in orange shading), the autoencoder model reconstructed the manipulated signals quite accurately. Over time, however, as the manipulation persisted, the reconstruction error eventually exceeded the threshold, leading to the detection of the manipulation.

To explore the possibility of reducing detection time, we investigated the latent space of the trained autoencoder. The three-dimensional graph on the right side of Fig. 5 visualizes this latent space: blue dots represent low-dimensional representations of the entire training data, and red dots represent instances of detection failure. As shown in the graph, the red dots form an independent cluster far away from the clusters of training data. Based on this investigation, we adopted DAGMM to improve detection time efficiency, as it monitors not only the reconstruction error but also the dynamics of the latent space.

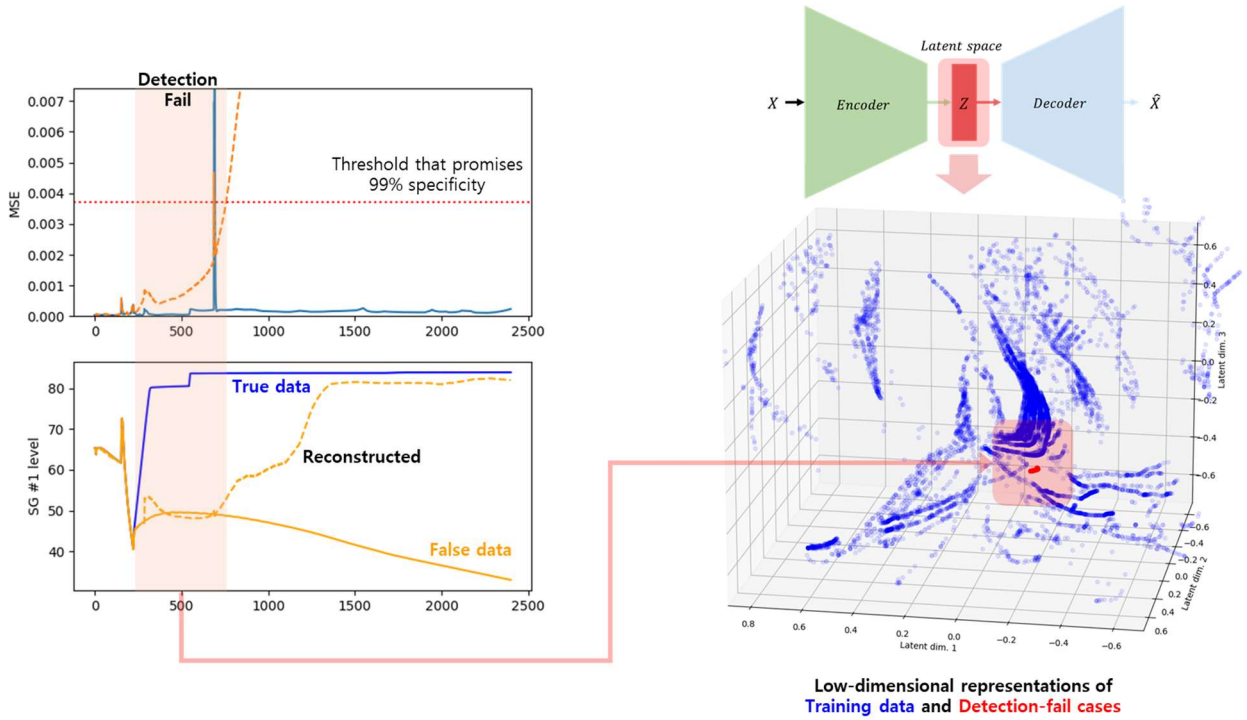


Figure 5. An example of instances where the detection fails and low-dimensional representations for these instances

4. CASE STUDY

As a case study, we implemented autoencoder and DAGMM-based cyberattack detection models to detect data manipulation attacks during emergency operations of NPPs. Specifically, we assumed manipulation attacks on parameters critical to monitoring the critical safety functions (CSF) of the plants since inaccurate display of these parameters could potentially lead to misunderstandings about the current situation and, ultimately, human errors with a relatively high probability.

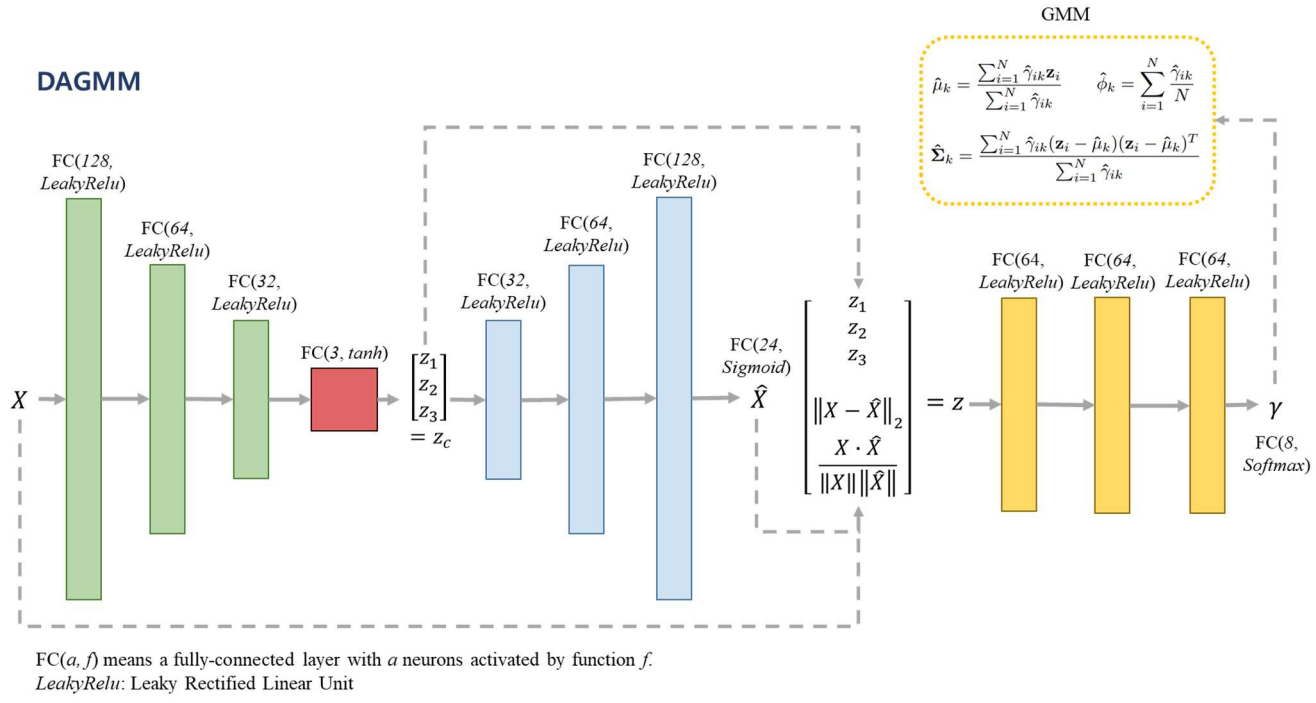
We considered three types of accidents: loss of coolant accidents (LOCA), steam generator tube rupture (SGTR), and spurious reactor trips. We collected 1,153 operation records totaling more than 2,220,000 seconds from simulators. The details of this data can be found in [26].

Table 1 lists the 24 plant parameters selected as monitoring variables for the detection models, and Fig. 6 depicts the structure of the DAGMM-based cyberattack detection models. The dimension of latent space was three and two reconstruction errors (i.e., Euclidean distance and cosine similarity) were measured. Therefore, the dimension of Z was five. In addition, we assume 8 Gaussian distributions. The structure of the autoencoder is analogous to that of the compression network within the DAGMM-based model.

Table 1. Monitoring parameters and their related CSFs

Critical Safety Functions	Monitoring Parameters
Subcriticality	Power range percent power Start-up rate Intermediate range neutron level
Core cooling	Core outlet temperature Hot-leg temperatures of loop 1, 2, and 3 Pressurizer pressure*
Heat Sink	Steam generator narrow levels of the generator 1, 2, and 3 Steam generator pressures of generator 1, 2, and 3 Auxiliary feedwater flows of generator 1, 2, and 3
Integrity	Cold-leg temperature of loop 1, 2, and 3 Pressurizer pressure*
Containment	Containment pressure Containment radiation Sump water level
Inventory	Pressurizer level

*Monitored for both core cooling and integrity



FC(a, f) means a fully-connected layer with a neurons activated by function f.
LeakyRelu: Leaky Rectified Linear Unit

Figure 6. Structure of the implemented DAGMM

After these models had been trained, the specificities according to the detection thresholds were evaluated as shown in Fig. 7. We set the mean squared error and energy producing 99% specificity as the thresholds for autoencoder and DAGMM models, 6.75e-4 and 1.7, respectively.

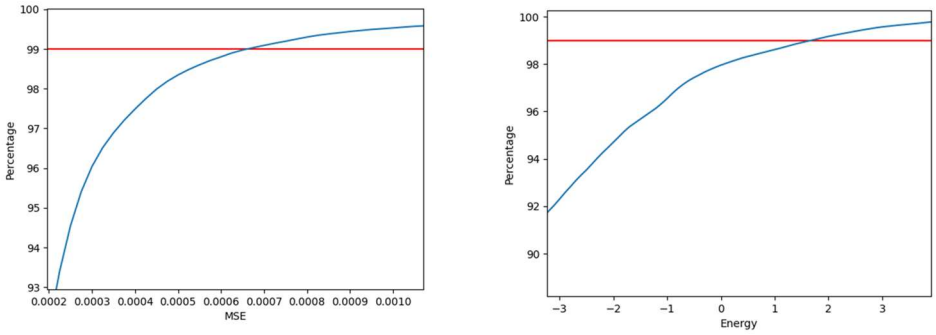


Figure 7. The specificities according to the detection thresholds. left: autoencoder, right: DAGMM

Figures 8 and 9 illustrate the detection results of the sole autoencoder and DAGMM-based models. The blue line represents the true trend of the parameter, while the dashed, orange line represents the manipulated trend of the parameter. In the same context, the blue lines and orange lines on the right side of the two figures illustrate the trend of the MSE (i.e. the monitoring parameter for the sole autoencoder model) and energy (i.e., monitoring parameter for the DAGMM) over time.

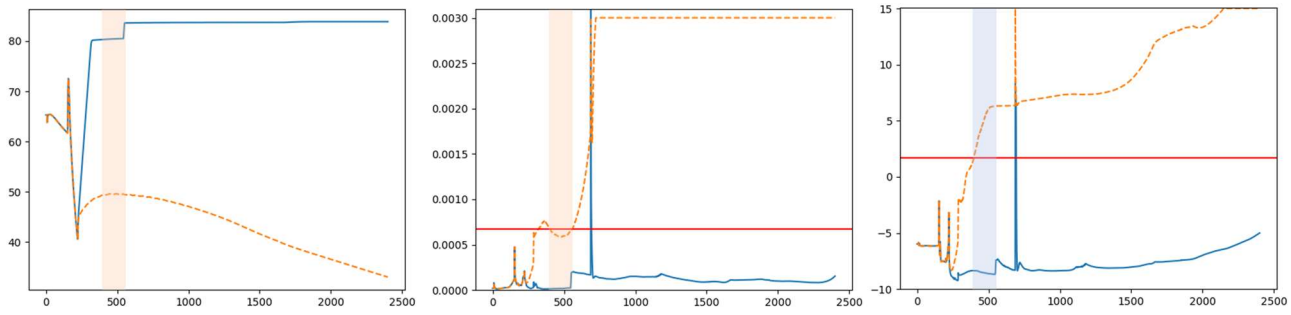


Figure 8. Detection result when water levels of normal and damaged steam generators are manipulated.

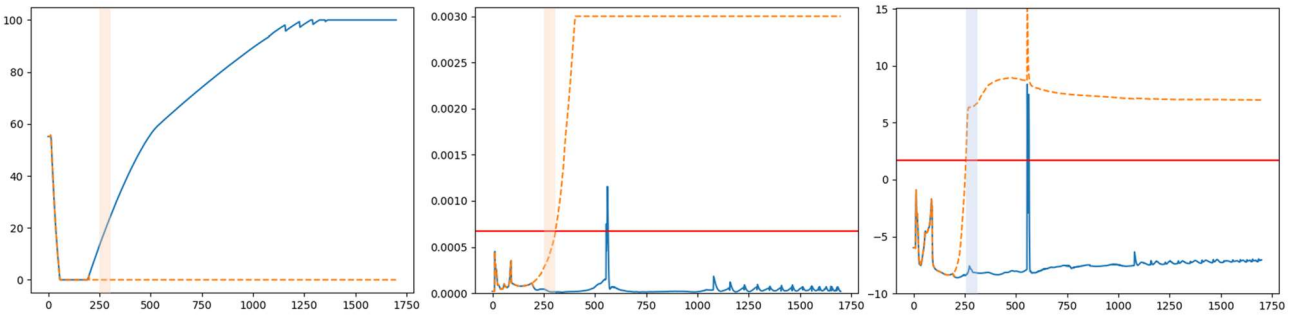


Figure 9. Detection result when pressurizer water level is manipulated to 0% during spurious trip.

When a manipulation attack swapped the levels of intact and damaged steam generators, as shown in the Fig. 8(a), the energy exceeded the detection threshold 100 seconds earlier (the color-shaded area) than the MSE did. Since the detection thresholds for both models were set to have the same specificities for the training data, it can be said that the DAGMM outperformed the autoencoder in this case. Likewise, the DAGMM achieved an earlier detection than the autoencoder when the pressurizer water level was manipulated to be stuck at 0%, as shown in Figure 9.

Considering our investigation in Section 3, the outperformance of the DAGMM may be due to the fact that the model monitors not only the reconstruction error, but also the patterns in the latent space. However, since this case study only tested the DAGMM for two cases of manipulation attacks, further studies with additional manipulation attack scenarios should be conducted.

5. CONCLUSION

The digitization of NPP I&C systems has led to increased cybersecurity concerns. In particular, process data manipulation attacks can affect the situational awareness of plant operators, potentially leading to human errors that could be critical in situations requiring timely responses, such as emergencies. To address this vulnerability, we proposed cyberattack detection models for NPP process data manipulation attacks using advanced machine learning techniques. Specifically, we introduced a DAGMM tailored to improve anomaly detection performance by integrating dimensionality reduction and density estimation on an autoencoder. The DAGMM exploits both reconstruction errors and latent space features extracted by the autoencoder, enabling the model to detect manipulation when the sole autoencoder model fails. As a case study, we trained the model with emergency operation records obtained from NPP simulators. The results showed that the DAGMM-based model reduced the detection time by identifying the attack when the autoencoder reconstructed the manipulated signals.

However, the proposed detection model has a limitation. Since the model cannot distinguish between intentional manipulation and the deviations due to sensor failure, calibration, or real phenomena, the detection result needs to be cross-verified by other models as proposed by Zhang et al. [2, 17, 18]. In particular, the suspicious network flow detection model can serve as a complement to the process parameter-based detection model considered in this research. In addition, the model needs to be tested on a real physical system not only on NPP simulators. If these limitations can be addressed, the proposed DAGMM-based detection model can be a barrier in the defense-in-depth against cyberattacks on NPPs.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2022-00165231).

References

[1] Langner, R., Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security & Privacy, 2011. 9(3): p. 49-51.
[2] Zhang, F., J.W. Hines, and J.B. Coble, A Robust Cybersecurity Solution Platform Architecture for Digital Instrumentation and Control Systems in Nuclear Power Facilities. Nuclear Technology, 2020. 206(7): p. 939-950.
[3] Poulsen, K., Slammer worm crashed Ohio nuke plant network. SecurityFocus, 2003, 2003.
[4] Lemos, R., Data storm" blamed for nuclear plant shutdown. SecurityFocus, May, 2007. 18: p. 84.
[5] Carr, J. and S. Goel, Project Grey Goose Report on Critical Infrastructure: Attacks, Actors, and Emerging Threats. Grey Logic, January, 2010. 21.
[6] Lee, J.-W., et al. Cyber security considerations in the development of I&C systems for nuclear power plants. in Proceedings of the International Conference on Security and Management (SAM). 2011. Citeseer.
[7] Commission, U.N.R., Cyber security programs for nuclear facilities. 2010: US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research.
[8] Kim, I.-k., Y.-e. Byun, and K.-h. Kwon, Analysis of the application method of cyber security control to develop regulatory requirement for digital assets in NPP. Journal of the Korea Institute of Information Security & Cryptology, 2019. 29(5): p. 1077-1088.
[9] Shin, J., H. Son, and G. Heo, Cyber Security Risk Evaluation of a Nuclear I&C Using BN and ET. Nuclear Engineering and Technology, 2017. 49(3): p. 517-524.
[10] Park, J.W. and S.J. Lee, Probabilistic safety assessment-based importance analysis of cyber-attacks on nuclear power plants. Nuclear Engineering and Technology, 2019. 51(1): p. 138-145.

[11] Park, J.W. and S.J. Lee, A quantitative assessment framework for cyber-attack scenarios on nuclear power plants using relative difficulty and consequence. Annals of Nuclear Energy, 2020. 142: p. 107432.
[12] Lee, C., Y. Ho Chae, and P. Hyun Seong, Development of a method for estimating security state: Supporting integrated response to cyber-attacks in NPPs. Annals of Nuclear Energy, 2021. 158: p. 108287.
[13] Levy, E., Crossover: Online pests plaguing the offline world. IEEE Security and Privacy, 2003. 1(6): p. 71-73.
[14] Lee, C., et al., Development of a method for securing the operator's situation awareness from manipulation attacks on NPP process data. Nuclear Engineering and Technology, 2022. 54(6): p. 2011-2022.
[15] Song, J.-G., et al., AN ANALYSIS OF TECHNICAL SECURITY CONTROL REQUIREMENTS FOR DIGITAL I&C SYSTEMS IN NUCLEAR POWER PLANTS. Nuclear Engineering and Technology, 2013. 45(5): p. 637-652.
[16] Kim, H.E., et al., Systematic development of scenarios caused by cyber-attack-induced human errors in nuclear power plants. Reliability Engineering & System Safety, 2017. 167: p. 290-301.
[17] Zhang, F., et al., Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data. IEEE Transactions on Industrial Informatics, 2019. 15(7): p. 4362-4369.
[18] Zhang, F. and J.B. Coble, Robust localized cyber-attack detection for key equipment in nuclear power plants. Progress in Nuclear Energy, 2020. 128: p. 103446.
[19] Song, J.-g., et al. Development of Hardware In the Loop System for Cyber Security Training in Nuclear Power Plants. 2019.
[20] Lee, S. and J. Kim, Design of computerized operator support system for technical specification monitoring. Annals of Nuclear Energy, 2022. 165: p. 108661.
[21] Cancemi, S.A., et al., Unsupervised anomaly detection in pressurized water reactor digital twins using autoencoder neural networks. Nuclear Engineering and Design, 2023. 413: p. 112502.
[22] Mena, P., R.A. Borrelli, and L. Kerby, Detecting Anomalies in Simulated Nuclear Data Using Autoencoders. Nuclear Technology, 2024. 210(1): p. 112-125.
[23] Ryu, S., et al., Development of deep autoencoder-based anomaly detection system for HANARO. Nuclear Engineering and Technology, 2023. 55(2): p. 475-483.
[24] Zong, B., et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. in International conference on learning representations. 2018.
[25] Reynolds, D.A., Gaussian mixture models. Encyclopedia of biometrics, 2009. 741(659-663).
[26] Bae, J., G. Kim, and S.J. Lee, Real-time prediction of nuclear power plant parameter trends following operator actions. Expert Systems with Applications, 2021. 186: p. 115848.