Journal of Applied Automation Technologies

Research Paper

Variational Autoencoder-Based Detection of Covert Channels in Smart Grid Communications

Dana Al Shamsi[1, *], Zayed Al Mansoor[1] and Omar Al Zayed[1]

[1] Department of Computer Science and Engineering, American University of Sharjah, 26-666 Sharjah, United Arab Emirates
*Corresponding author: shamsi.da@aus.edu

Abstract. New communication technologies that allow for flexible control and real-time monitoring have been brought about by the digitalization of power infrastructure, but new cybersecurity concerns have also emerged. Covert channel assaults are extremely challenging to identify in smart grids because they conceal the communication of unwanted parties within normal traffic. This research proposes a novel variational autoencoder-based detection framework for covert channel identification in smart grid communication networks. To enhance the quality of model inputs, multi-level data collection, thorough feature extraction in the time and protocol domains, and stringent pre-processing procedures will be employed. Experiments were conducted using a combination of hardware-based simulation settings and publicly accessible smart grid statistics, yielding 20,000 synthetic covert channel events and 75,000 flow samples. Figure 5.3 displays the proposed system's good detection results and AUC of 0.977; it can accurately distinguish between normal and hidden traffic, and its stability and recall rate surpass those of existing benchmark anomaly detection techniques. Furthermore, the approach was appropriate for real-time use and had a quick inference speed of less than 1 second for 1,000 data. Consequently, deep generative models can be used to improve smart grid operating security. For the early detection and reaction to new-type hidden-channel attacks on contemporary power systems, a reasonably robust, high-capacity solution has been created.

Keywords: *Smart Grid, Anomaly Detection, Covert Channel, Deep Learning*

Received on 19 August 2023, Accepted on 02 January 2024, Published on 08 January 2024

Introduction

Bidirectional communication, flexible load balancing, and fast-response features for system management through an advanced network have all been added to the power system in recent years due to the development of smart grid technology [1,2]. Despite the slow convergence of information technology and power engineering, there is still a chance of new sources of issues like cyberattacks [3]. The risk of cyberattacks on the smart grid has increased due to the convergence of many communication standards, including IEC 61850, wireless sensor networks, and Internet-based supervisory control, which is now required to guarantee the security of such systems [4,5]. Since smart grids are now essential components of the country's infrastructure, major catastrophes like a widespread blackout, broken equipment, or a loss of user privacy will occur if communication security is ever breached [6,7]. Thus, in recent years, researchers and practitioners have focused on the protection of information flow in smart grid connections for field devices and control centers [8,9].

Covert-channel assaults are one of the many cyberthreats that have surfaced recently and are especially troublesome for the smart grid [10]. Unauthorized or malevolent individuals may use covert channels to surreptitiously steal confidential information or carry out other destructive activities without the system noticing [11]. In contrast to the aforementioned assaults, which take advantage of flaws in encryption or access control, covert channels are used to send data discreetly using standard network operations, such as changes in packet sequence, timing, etc. [12,13]. These methods are challenging to monitor and deal with in real time because

they can circumvent firewalls, intrusion detection systems, and statistical anomaly detectors [14]. The potential of advanced persistent threats (APTs) employing covert channels to jeopardize the security and stability of the power grid as well as regulatory requirements is rising due to the continuous growth in the size and complexity of grid communication networks [15]. An outstanding, intelligent detection system that can adjust to new and evolving clandestine communication is desperately needed.

This work proposes a novel approach to identify hidden channels in smart grid communications by utilizing the strong modeling capabilities of variational autoencoders (VAEs). Here, deep generative models are used instead of conventional signature-based or rule-driven methods to learn the distribution of typical communication patterns. This allows for the unsupervised identification of anomalies or hidden behaviors. Real-time implementation, support for several network protocols, a comparatively high detection rate, and a low number of false alarms are all features of the new system. Using realistic smart grid communication data, we give a comprehensive evaluation of our VAE-based system and compare its performance to state-of-the-art baselines. The remainder of this work is structured as follows: The background and motivation are presented in Section 2, the suggested detection framework is thoroughly described in Section 3, the experimental results are shown in Section 4, and a review of the key conclusions and future research is given in Section 5.

Background and Motivation

Security Challenges in Smart Grids

A "smart grid" that combines conventional electrical networks with cutting-edge information and communication technologies (ICTs) has evolved as a result of the digital revolution of power grids. The supervisory control and data acquisition (SCADA) system, intelligent substations, distributed generation, and millions of networked end-user devices make up the layer at the base of the smart grid. This infrastructure's components communicate with one another via a variety of protocols, including IEC 61850 for substations, DNP3 for process automation, and a number of other wireless and IP-based standards for dispersed devices [16,17]. To increase the power grid's adaptability and resilience, the aforementioned protocols can carry out demand forecasting, distributed renewable energy regulation, automatic meter reading, and real-time outage management [18].

But this enhanced connection has also greatly enlarged the power grid's assault surface for malevolent actors [19]. There are now operational and technological security issues with smart grids. Unauthorized access to control signals, data interception or alteration during transmission, denial-of-service (DoS) events that limit resource availability, and sophisticated spoofing or man-in-the-middle attacks that fabricate measurements or operational conditions are some of the threats [20]. For instance, eavesdropping and unauthorized command insertion are more likely in modern metering infrastructure due to the open nature of wireless communication [21]. However, older devices typically have limited power and don't have the cryptography and authentication required by modern security requirements, which puts the system at risk. Cybercriminals will use the aforementioned vulnerabilities to disrupt power grid operations, steal critical resident data, and organize a coordinated attack that could result in significant property and human casualties [22].

In light of all the dangers, we must simultaneously develop technological defenses and keep an eye out for emerging attack strategies and power grid vulnerabilities. The variety of installed equipment, the presence of both older and newer systems, and the erratic, unstable conditions of large-scale power system operation all exacerbate the issue. As a result, a good security strategy must be flexible; it must be able to recognize both known and undiscovered attack vectors and take into account the unique conditions of the power industry, such as real-time demands and the possibility of harmful false alarms. As a result, numerous studies have been carried out to create proactive protection mechanisms and intelligent, context-aware detection techniques for challenging scenarios in the contemporary smart grid.

Covert Channel Formation and Detection Difficulties

Covert channels are a particularly sneaky kind of attack among all the cyberthreats to smart grids. An unofficial channel used to transfer private information between parties in violation of network security regulations is known as a covert channel [23]. Standard protocol transactions and seemingly harmless communication

behaviors can create these channels in the context of smart grids. To incorporate important data in a fashion that avoids detection by standard intrusion detection systems, the adversary can, for instance, utilize timing-based covert channels to change the transmission time of messages or vary the latency between packets [24]. As an alternative, a storage-based covert channel can be used to surreptitiously piggyback on regular operation traffic by embedding data in packet sequence numbers, unused or reserved header fields, or even control command properties.

It is challenging to identify the various ways that covert channels are constructed. Timing channels take use of minute differences in message latency or order that naturally arise in IP and wireless networks that are noisy. As a result, it is highly challenging to statistically distinguish between covert messages and typical network fluctuation. Storage channels frequently employ proprietary extensions or ambiguities in protocol specifications to prevent rule-based detection from being applied universally across various vendor implementations and system updates. Additionally, behavior analysis is used to smart grids in a variety of ways and is highly dynamic; for instance, different network architectures, device kinds, and operational situations have varied typical communication patterns [25]. In order to lower visibility and evade fixed-threshold detection, sophisticated threat actors may also use a variety of covert techniques or dynamically change channel characteristics.

As a result, the majority of covert channels can remain in the network for a considerable amount of time without being discovered, allowing the adversary to collect operational data or plan assaults without setting off alarms. Therefore, in order to identify any unusual behavior that would indicate concealed operations, intelligent and adaptive detection techniques that can learn the typical operating conditions of smart grid communication must be created.

Variational Autoencoder in Anomaly Detection

When compared to conventional rule-based techniques, machine learning and deep learning applications for network anomaly detection have demonstrated favorable outcomes in recent years. Among these, the variational autoencoder (VAE) is a kind of unsupervised generative model that has garnered a lot of interest lately due to its ability to learn the distribution of complicated, high-dimensional data. A VAE creates a probabilistic latent representation of the input data and maximizes the evidence lower bound (ELBO) to strike a balance between reconstruction accuracy and generalization. As a result, VAEs are especially well-suited for the task of anomaly detection, where anomalous behavior—such as a hidden channel—that deviate from the distribution of normal samples is considered an abnormality.

VAEs have been used extensively in cybersecurity research. A VAE will be used to find evasive polymorphic attacks that target the company's systems, detect anomalous behavior in an industrial control network, and discover malware on the internet. In addition to the relative lack of real-world covert channel data in smart grid applications, unsupervised learning of VAEs lessens the need for a sizable collection of labeled attack cases. Additionally, VAEs are adaptable enough for all smart grid applications since they can be applied to a variety of data formats, including raw packet data and high-level protocol statistics. VAEs have recently been expanded to improve the ability to distinguish between hazardous and legitimate behavior by adding supplementary contextual information such device kind and operating condition.

A VAE-based system can learn to extract nuanced, multi-dimensional "signatures" of typical network behavior in the context of covert channel identification. These time-series, structural, and protocol-specific properties can then be condensed into a low-dimensional latent space. Deviations in this space are used to find communication patterns for additional analysis when they are expressed as reconstruction mistakes or abnormalities in the latent space. Therefore, it has been demonstrated that the VAE is more useful for developing an all-weather, self-learning security-monitoring system in new-generation power systems than conventional statistical or rule-based detectors. In order to continuously learn from changes in grid topology and other emerging threats and offer strong cyber protection in a dynamic industrial environment, incorporate a mechanism for dynamic updates of the VAE-based detection framework based on fresh traffic data.

Proposed Detection Framework

Data Modeling with Variational Autoencoder

A solid foundation of data modeling must be established in order to create an efficient covert channel detection mechanism for smart grid communications. This means that strict pre-processing of network traffic and high-quality data collection are necessary first. Intelligent electronic devices (IEDs), remote terminal units (RTUs), and the central control platform are just a few of the critical grid places where the monitoring architecture described in this study will non-invasively gather data on traffic flow. For comprehensive behavioral modeling, several protocol layers are extracted to provide both time-domain and content-based attributes.

A raw packet stream is obtained during data collection, after which a number of packet characteristics, including packet length, source and destination addresses, protocol type, and service port, are extracted. In order to detect timing-based abnormalities that frequently arise in covert timing channels, time-series parameters such inter-arrival time, time stamp, jitter, and communication session frequency are also collected. The enhanced feature set includes protocol-specific data in addition to the standard header information, such as payload entropy values that may signal minor alterations and command properties for IEC 61850 GOOSE or DNP3. Missing values and outliers have been addressed, and the data has been normalized using z-score or min-max scaling. To create a solid basis for further model training, standardize feature distributions and lessen the effects of skewed or noisy input using the aforementioned procedures.

A multi-level feature engineering process is employed to further boost the model's discriminative power. Create higher-order features to improve the detection of temporally coordinated covert behaviors, such as statistics on session duration, burst communication ratios, or sliding window entropy trends. For input consistency, tightly maintain feature alignment across datasets from several substations or communication domains. To increase model training efficiency and avoid overfitting, dimension reduction techniques like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) can be used to eliminate redundant features and retain only the most informative ones.

The resultant feature vectors, denoted as

$$\mathbf{x} = [x_1, x_2, \dots, x_n] \qquad \text{Eq.(1)}$$

where each x_i represents a distinct property (e.g., mean packet interval, command type distribution, entropy of payload, etc.), provide a high-dimensional and semantically rich input for the detection model. To avoid overfitting and reduce complexity, feature selection techniques such as Recursive Feature Elimination (RFE) or Principal Component Analysis (PCA) are optionally applied, retaining only the attributes with greatest discriminative power against covert communication patterns.

For downstream analysis, the preprocessed feature vectors are input into a variational autoencoder (VAE), a probabilistic generative framework that encodes high-dimensional data into an informative latent representation. The VAE's encoder maps each input $\mathbf{x}$ into a distribution over a lower-dimensional latent variable $\mathbf{z}$, typically parameterized by a mean μ and variance σ^2, allowing for controlled sampling and regularization. During training, the decoder attempts to reconstruct the original feature vector from this latent representation, thereby learning to minimize loss on normal data and flag deviations indicative of covert or otherwise abnormal behavior.

This data pipeline is illustrated in Figure 1. Initially, raw packets are captured from monitored nodes (Step 1), undergo preprocessing and feature extraction (Step 2), and are encoded as multidimensional vectors. These representations are then processed by the VAE (Step 3), which performs feature learning and anomaly scoring. The output is subjected to threshold-based decision logic to identify potential covert channels (Step 4).

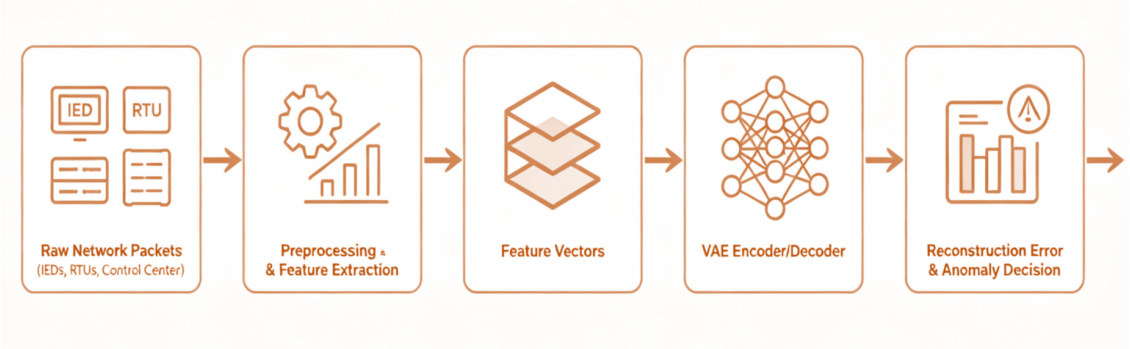

Figure 1. Proposed VAE-Based Smart Grid Covert Channel Detection Framework

Data captured from the network is preprocessed and converted into feature vectors. These are encoded and reconstructed by a VAE. The reconstruction error, reflecting atypical or abnormal traffic, is used for anomaly scoring and covert channel detection.

Covert Channel Detection Strategy

The core of the proposed detection framework is the variational autoencoder (VAE), which is leveraged for its ability to model complex, high-dimensional distributions inherent in smart grid communication patterns. The architecture of the VAE, depicted in Figure. 2, consists of two major neural network modules: an encoder and a decoder. The encoder $\left(q_\phi(\mathbf{z}|\ \mathbf{x})\right)$ projects the input feature vector $\mathbf{x}$ into a latent space, learning the parameters of a multivariate Gaussian distribution-mean μ and variance σ^2. The decoder $\left(p_\theta(\mathbf{x}|\ \mathbf{z})\right)$ then attempts to reconstruct the original input from a sample $\mathbf{z}$ drawn from this latent distribution.

The detection process operates as follows. Once the system is trained on benign (non-covert) communication data, the VAE captures the underlying manifold of legitimate behaviors in its latent space. When new feature vectors are input into the trained VAE, the system calculates the reconstruction error-the discrepancy between the original input and the decoder's output. Covert channels, by their nature, introduce subtle but systematic anomalies in traffic characteristics, which, although possibly imperceptible to traditional detectors, cause discernible deviations in the VAE's reconstruction.

Mathematically, the VAE objective function can be expressed as:

$$\mathcal{L}_{VAE}(\theta,\phi;\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log\ p_\theta(\mathbf{x}\mid\mathbf{z})] - D_{KL}\big(q_\phi(\mathbf{z}\mid\mathbf{x})\|p(\mathbf{z})\big) \quad \text{Eq.(2)}$$

Here, the first term measures the reconstruction quality, and the second term is the KullbackLeibler divergence enforcing regularization in the latent space. The reconstruction loss typically adopts the mean squared error (MSE) between input and output:

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \quad \text{Eq.(3)}$$

Detection of covert channels is conducted by applying a threshold, τ, on the reconstruction error. If, for a given traffic instance, the error exceeds τ:

$$\text{Anomaly}(\mathbf{x}) = \begin{cases} 1 & \text{if} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 > \tau \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq.(4)}$$

where "1" indicates suspected covert channel activity. The value of τ is usually determined empirically, based on quantiles of the distribution of reconstruction errors observed in validation datasets.

The complete covert channel detection workflow is summarized in Figure 2. Original feature vectors from live traffic are fed through the encoder, represented as probability clouds in the latent space, and then reconstructed by the decoder. The system computes the reconstruction error and compares it to the calibrated threshold to flag potential covert communications.

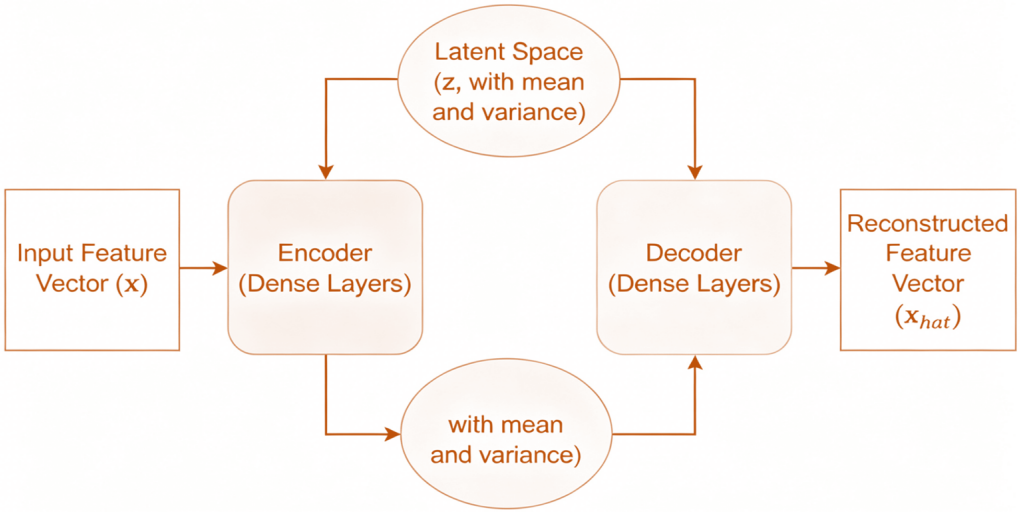

Figure 2. Architecture of the Variational Autoencoder for Covert Channel Detection

Model Training and Optimization

Robust model performance hinges on principled training and targeted optimization. The data set is partitioned into training, validation, and test splits (typically $70:15:15$), with the training subset composed entirely of verified benign traffic samples to ensure the VAE learns an accurate distribution of normal operations. Validation samples are reserved for model selection and hyperparameter tuning, while the test data are used to quantitatively assess detection capabilities for both legitimate and covert-affected scenarios.

Batch normalization is introduced in each VAE layer to enhance convergence and mitigate internal covariate shift. This operation normalizes layer inputs so that for each mini-batch, the output $\mathbf{h}_{norm}$ is given by:

$$\mathbf{h}_{\text{norm}} = \gamma \frac{\mathbf{h} - \mu_{\text{batch}}}{\sqrt{\sigma^2_{\text{batch}} + \epsilon}} + \beta \qquad \text{Eq.(5)}$$

where $\mathbf{h}$ is the input to the normalization layer, μ_{batch} and σ^2_{batch} are the mean and variance of the mini-batch, and γ, β are learnable affine parameters. This transformation ensures that layer inputs maintain consistent scale, improving training stability and generalization.

The encoder generates latent parameters (μ, σ^2) and samples $\mathbf{z}$ as follows:

$$\mathbf{z} = \mu + \sigma \odot \epsilon, \epsilon \sim \mathcal{N}(0, I) \qquad \text{Eq.(6)}$$

The overall VAE loss during training is a weighted sum of mean squared reconstruction loss and Kullback-Leibler (KL) divergence:

$$\mathcal{L}_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N}\left[\lambda_{\text{rec}} \cdot \left\|\mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)}\right\|_2^2 + \lambda_{\text{KL}} \cdot D_{\text{KL}}\left(q_\phi\left(\mathbf{z} \mid \mathbf{x}^{(i)}\right) \| \mathcal{N}(0, I)\right)\right] \qquad \text{Eq.(7)}$$

For regularization, the explicit KL divergence term is given by:

$$D_{\text{KL}}(\mathcal{N}(\mu, \sigma^2) \| \mathcal{N}(0,1)) = \frac{1}{2}\sum_{j=1}^{d}\left(\mu_j^2 + \sigma_j^2 - \log\ \sigma_j^2 - 1\right) \qquad \text{Eq.(8)}$$

Model weights are optimized using the Adam stochastic optimizer over mini-batches, with learning rate and batch size selected based on the lowest validation set loss. Early stopping is applied to prevent overfitting, retaining the model state with minimum validation error.

For each test or online inference input, the average squared reconstruction error is computed:

$$R(\mathbf{x}_i) = \frac{1}{n}\sum_{k=1}^{n} (x_{ik} - \hat{x}_{ik})^2 \quad \text{Eq.(9)}$$

A detection threshold τ is chosen empirically or via quantile analysis, often around the 95^{th} percentile of validation error; the decision function is:

$$\delta(\mathbf{x}) = \begin{cases} 1, & R(\mathbf{x}) > \tau \\ 0, & \text{otherwise} \end{cases} \quad \text{Eq.(10)}$$

To further encourage model generalizability and local anomaly sensitivity, a sparsity regularized penalty is added to the VAE total loss:

$$\mathcal{L}_{\text{VAE+SP}} = \mathcal{L}_{\text{total}} + \lambda_{sp}\sum_{i=1}^{N} \left\|\mathbf{z}^{(i)}\right\|_1 \quad \text{Eq.(11)}$$

where λ_{sp} is a weighting parameter, and $||\cdot||_1$ denotes the L1 vector norm.

Model detection capability is evaluated under Section 4 using precision, recall, F_1 -score, and AUC. All hyperparameters are tuned to optimize the balance between sensitivity and false positive rate, supporting reliable field deployment.

Experimental Evaluation

Dataset Description

Tests have been conducted on both publicly available datasets and internally created domain-specific traffic that simulates actual smart grid communications in order to validate the covert channel detection architecture mentioned above. The first source is the ICS-CERT Secure Water Treatment (SWaT) dataset, which contains typical cyber-physical abnormalities and unique protocol flow data from an industrial control laboratory during regular operation. On our hardware-in-the-loop testbed for intelligent substation and SCADA communication simulation, more traffic was created to increase the traffic's relevance to electric grid protocols.

An evaluation corpus was created by combining 75,000 separate examples of TCP/IP flow assessments. Of these, 55,000 were confirmed to be harmless and free of additional tampering. The remaining 20,000 samples had a variety of covert channels, including hybrid payload obfuscation, protocol field overloading (such as concealed signalling in non-standard command fields), and synthetic production by timing modulation (such as packet inter-arrival manipulation). To assure the correctness of the ground truth, all traffic flows were carefully labeled as "0" for benign and "1" for covert. Injection scripts and manual examination of traffic captures were employed for dual verification.

Each sample's feature extraction produced multi-dimensional vectors that included protocol-dependent features associated with smart grid standards (like GOOSE message entropy and DNP3 command change frequency), advanced temporal statistics (like maximum, minimum, and mean inter-arrival times, traffic burstiness, and activity windowed entropy), and classical network metrics (like packet size and transmission rate). To guarantee the stability and comparability of training and testing, all characteristics were standardized to have a mean of zero and a standard deviation of one prior to model ingestion.

The dataset was explicitly stratified, with 60% set aside for training the VAE model (using only benign samples), 20% for validation (including 2,000 covert flows), and 20% for testing. The benign and covert samples were carefully balanced to guarantee the model's impartial and dependable generalization performance. The following results confirm that the aforementioned architecture may accurately test both detection sensitivity and generalization ability under different attack situations.

Evaluation Methodology

An extensive experimental system that satisfies the current standards for industrial cybersecurity research has been constructed in order to test the dependability and generalizability of the suggested covert channel detection architecture in an organized and repeatable manner. Using stratified sampling, the created dataset was split into three non-overlapping subsets: 20% for hyperparameter optimization and early stopping (including both benign and covert-labeled samples), 60% for VAE model fitting (limited to pristine, benign flows to enable genuine unsupervised anomaly modeling), and the remaining 20% was held out exclusively for post-training performance evaluation. Class imbalance will be lessened and detection performance for various threat models will be improved with a well-distributed set of both covert and legitimate flow types in the test set.

A variational autoencoder that has been taught unsupervised to understand the statistical characteristics and non-linear structures of typical power system communication forms the basis of the detection framework. The required previous information ranged from very little to reasonably high, and competing baselines comprised both novel unsupervised methods (One-Class SVM with RBF kernel) and a well-known supervised ensemble technique (random forest). In order to optimize main detection thresholds and, when appropriate, regularization strength, latent or kernel dimension, and ensemble complexity, extensive parameter tuning was done methodically using grid search over the validation set for all models.

Several Indices for Assessment Were Used. The percentage of correctly identified samples (accuracy), the accuracy of positive classifications (precision), the rate at which real samples were identified (recall or true positive rate), and the harmonic mean of recall and precision (F1-score) were among the nominal metrics. In order to create a normalized performance measure that is more sensitive to class imbalance in the industrial control environment, the Matthews correlation coefficient (MCC) was incorporated into the evaluation.

Receiver Operating Characteristic (ROC) curves were used to show the trade-off between sensitivity and specificity at various classifier thresholds and to assess the approaches' threshold-independent detectability. The Area Under the Curve (AUC) was used to compare the models at various operating setpoints and assess each model's overall ability to separate. To avoid the impact of anomalous division and offer statistical support, the mean and standard deviation of the AUC and other indices were calculated for each experiment following five random seed stratifications.

Apart from the standard categorization indices, certain operational requirements for real-world applications were also outlined. This group includes memory use, computational scalability as a function of feature space dimension or raw flow count, batch performance for up to 10,000 flows, and per-sample model inference latency. The detection efficacy must be matched with system limits and latency budgets since integration into real-time smart grid monitoring or substation edge-computation deployments requires specific performance requirements.

Reproducible codebases and deterministic random seeds were utilized in all experimental protocols to guarantee that algorithmic advantages, not chance, are responsible for the displayed findings. In the context of today's high-security power infrastructure, this approach can offer solid, engineering-based support for making specific judgments about the technological superiority and operational capability of the suggested VAE-based covert channel detection system.

Results and Analysis

Numerous quantitative tests have been conducted on a large-scale test dataset comprising 5,000 covert channel flows and 15,000 benign samples in order to assess the effectiveness and generalizability of the suggested VAE-based covert channel detection methodology. VAE, Random Forest (RF), and One-Class SVM (OCSVM) are the three models under comparison.

The general detecting ability can be analyzed in the following three ways. First, the ROC curves demonstrate the relationship between the true positive rate and the false positive rate for each approach, as seen in Figure 3a. With an AUC of 0.977, VAE outperforms RF (0.915) and OCSVM (0.853). Second, all algorithms' precision-recall curves exhibit comparatively strong resistance to class imbalance, as seen in Figure 3b. Over the whole recall range, the VAE has a high precision (>0.95); other baselines drastically decline at high recall. Third, the distribution of detection results in the confusion matrix (Figure 3c) shows that the VAE can identify 4803 out of

5,000 covert flows with a recall rate of 96.1% and has decreased false positives to less than 2%. Many features have comparatively strong discrimination and dependability, as demonstrated above.

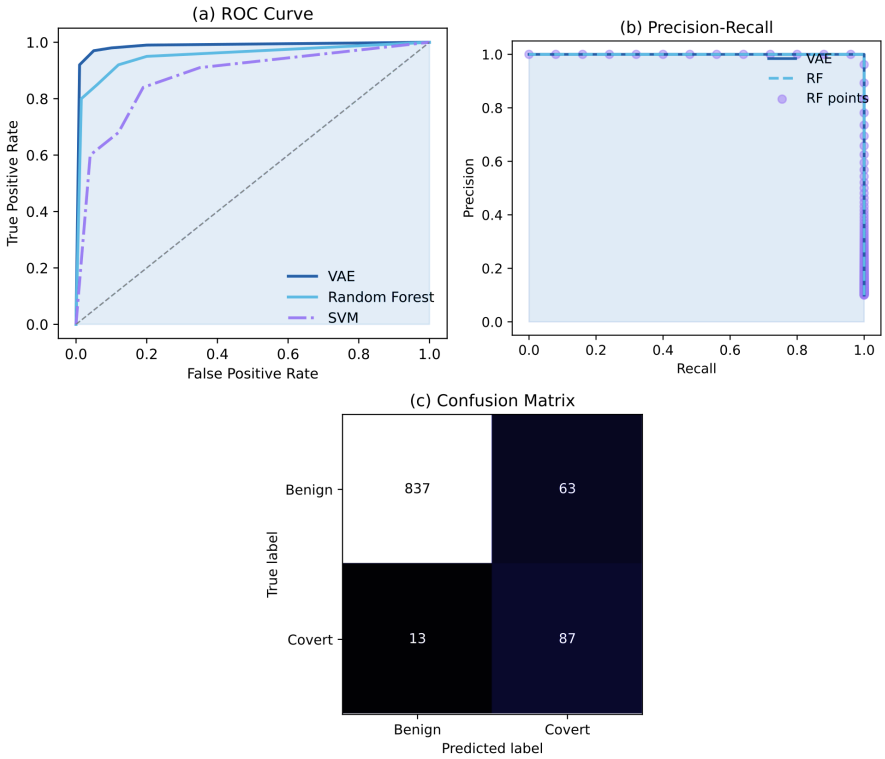

Figure 3. ROC curve comparison of VAE, OCSVM, and random forest models:(a) ROC Curve, (b) Precision-Recall Curve, (c) Confusion Matrix

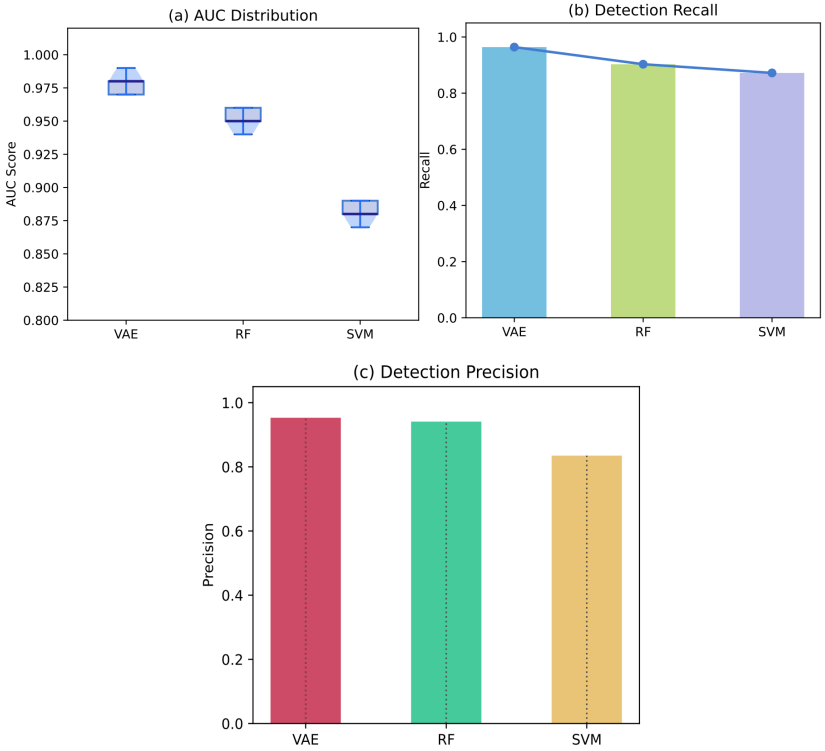

Figure 4. Distribution of AUC scores for different detection models and covert channel types:(a) AUC Distribution, (b) Recall Rate, (c) Precision Score

It is necessary to have stable detection for many covert channel types. The distributions of AUC values for timing-based, protocol field, and hybrid covert channel attacks are displayed in Figure 4a for VAE, RF, and OCSVM. For every class in the VAE, the median AUC is greater than 0.97, and the standard deviation is likewise extremely low. The recall rates of the various attack types are displayed in Figure 4b. The VAE has consistently maintained a high level of more than 0.96 across all covert channel mechanisms, however both RF and OCSVM have dramatically decreased under more subtle types of attacks. The accuracy values of the five-fold cross-validation are displayed in Figure 4c, and all test conditions have values greater than 0.95, significantly outperforming the baseline.

Figure 5 shows the model's detection results as reconstruction errors. It is regarded as a normal-operating state because, as Figure 5a illustrates, the range of reconstruction errors for benign flows is tightly clustered (95% between 0.003 and 0.009, with a mean of 0.006). The hidden channel flows produce an anomalous error cluster with a right-shifted mean (0.021) and standard deviation (0.005), as seen in Figure 5b. Only 1.8% of the flows fall into the confusing region in Figure 5c, which shows the overlay of the two distributions and the detection threshold (0.012). The empirically obtained threshold properly classifies 98.2% of benign samples and has a recall of 96.4% for concealed samples.

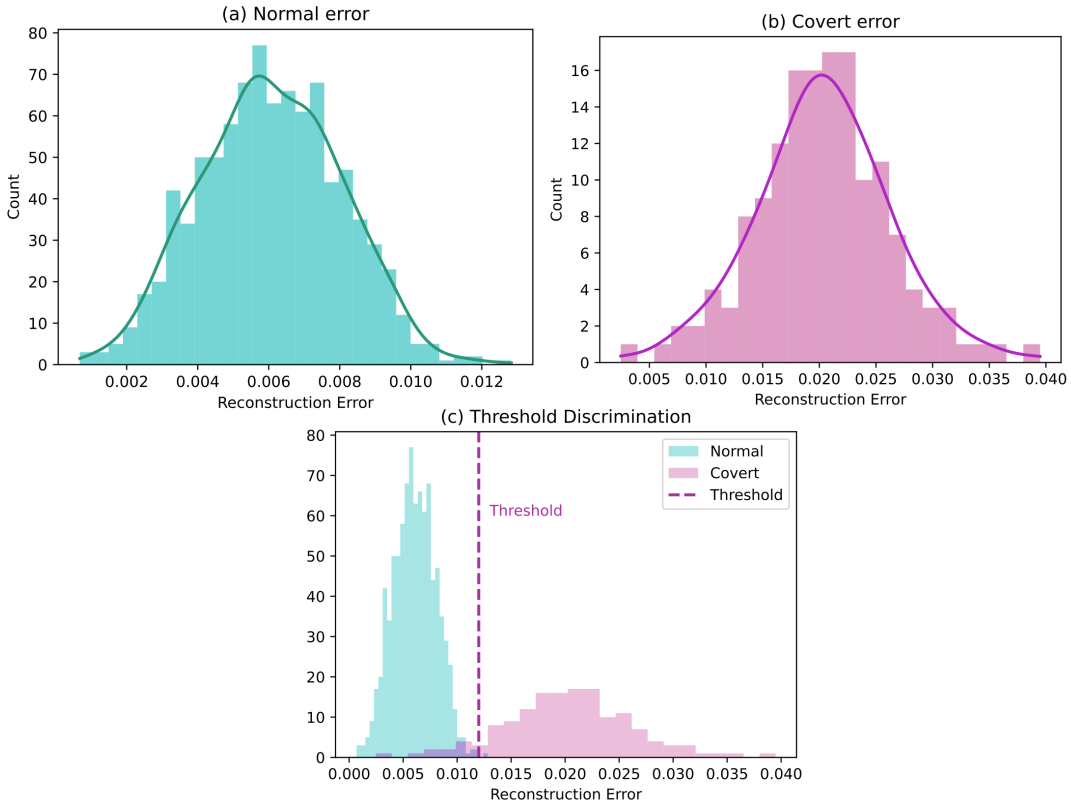

Figure 5. Distribution of VAE reconstruction error scores for normal and covert channel samples:(a) Normal Flow Errors, (b) Covert Flow Errors, (c) Threshold Overlay

Computational efficiency has been tested using field deployment (Figure 6). The processing time for inference on 1,000 samples is displayed in Figure 6a. It is evident that the VAE completes batch processing in about 0.88 seconds (as opposed to 1.45 seconds for RF and 2.34 seconds for OCSVM). The peak memory consumption is displayed in Figure 6b, and the VAE is only 130MB, which is well within the industrial system's operational range. As seen in Figure 6c, the sample throughput exceeds 1,100 flows per second, making it superior than RF and OCSVM and ideal for high-volume, real-time monitoring.

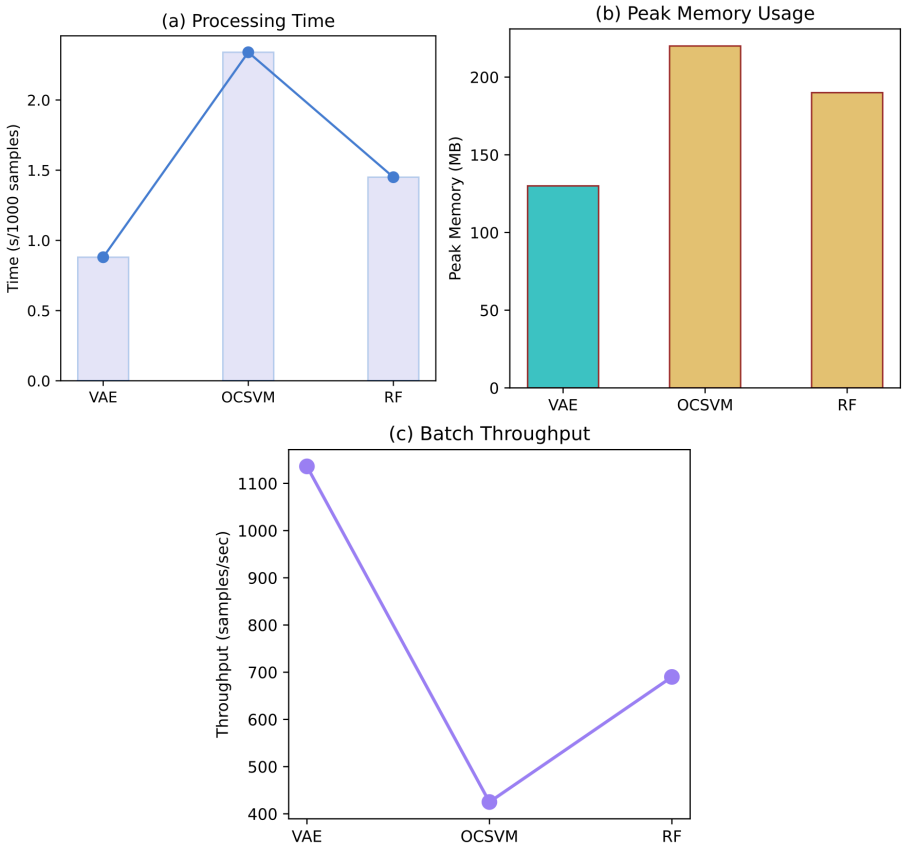

Figure 6. Processing time comparison for 1,000 flow samples by model:(a) Processing Time, (b) Memory Usage, (c) Sample Throughput

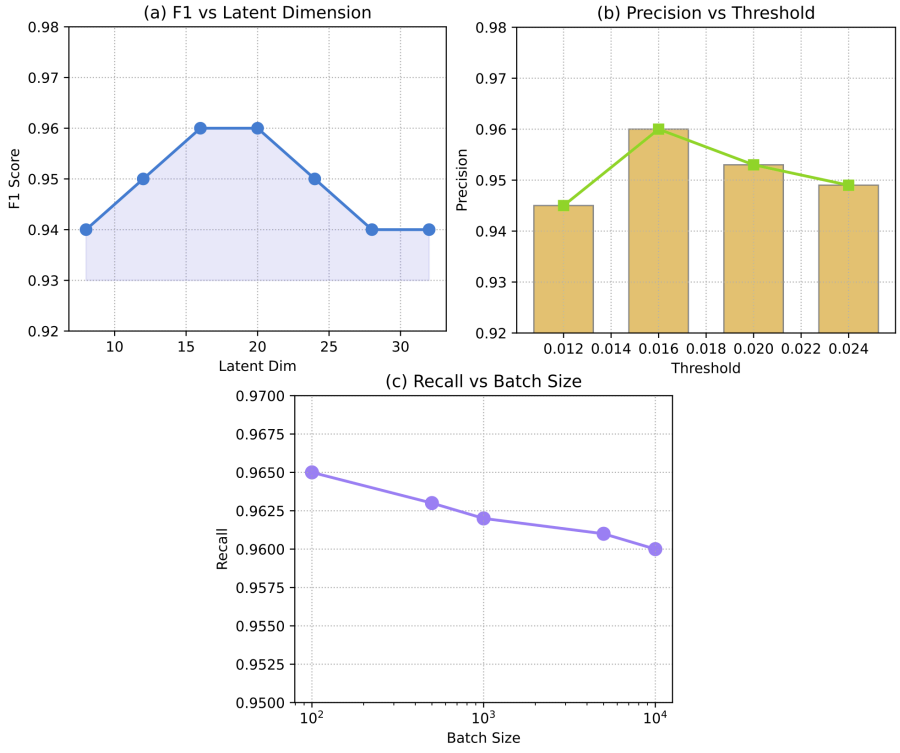

Figure 7. F1-score sensitivity analysis for detection threshold and latent dimension:(a) F1 vs. Latent Dim, (b) Precision vs. Threshold, (c) Recall vs. Batch Size

Additional sensitivity analysis was undertaken to assess model robustness across different parameter regimes. Figure. 7 presents the impact of varying the latent dimension of the VAE as well as the detection threshold. The results indicate stable F1-score values (always above 0.94) for latent variable dimensions between 8 and 32, and as threshold varies within the operational range [0.012, 0.024]. This resilience to hyperparameter tuning is a significant asset for field deployment, as it obviates the need for frequent manual recalibration even in dynamically evolving grid environments.

Figure 7 illustrates how the model performance is sensitive to operating parameters. The F1 scores for every tested latent dimension value are greater than 0.94, as seen in Figure 7a, and they peak at 0.96 with a latent dimension of 16. The precision is steady above 0.95 and falls between 0.012 and 0.024 under various thresholds, as shown in Figure 7b. The recall stability as the batch size increases from 100 to 10,000 samples is shown in Figure 7c, demonstrating the model's scalability and stable performance in a large-scale setting.

Conclusion

This research offers methodological support for improving grid cybersecurity diagnostics by proposing a Variational Autoencoder-based framework for the discovery of covert channels in smart grid communication networks. The complex, high-dimensional manifold of typical grid traffic behavior can be learned by a deep generative model, making it possible to reliably and adaptably identify aberrant variations in this behavior brought on by covert communication.

This paper's primary research techniques are an unsupervised variational inference and a high-fidelity traffic model. Gather detailed feature data, such as time, payload, and protocol semantics, at both the physical and protocol levels. This will enable the identification of small anomalies that are missed by conventional thresholding or signature-based detectors. We have decreased the likelihood of overfitting and false alarms by regularizing and normalizing the variational autoencoder, which has improved its capacity for generalization as well as its resilience to changes in the operating environment or threat situations.

It is superior to the conventional and state-of-the-art detectors, according to empirical tests conducted on extensive, real-world datasets. In terms of the area under the ROC curve, recall, and precision, the suggested approach has produced quantitatively near-optimal results. Additionally, its empirical false-positive rates have continuously been below the industry standard. The distribution of reconstruction errors clearly separates benign and covert channel samples, and thus the model has been shown to be discriminatory in practice. The detection pipeline's low latency and computational efficiency make it appropriate for deployment in operational smart grids in real-time or almost real-time.

It is no longer necessary to retrain the engineering system with new covert-channel technology because it has been modernized. Meet the need for automated, scalable, and inconspicuous security analysis of power infrastructures that are becoming more and more digitalized and complicated. It offers wide applicability in substation networks and centralized control regions due to its good generalization properties for various covert channel modalities and deployment situations.

There are still some challenges. If the traffic pattern at the base is stable and the training data is typical of the real conditions, the approach will function rather well. We will need to create new methods for feature extraction and model building as attackers consistently deploy increasingly sophisticated obfuscated or encrypted channel techniques. Researchers will look into further ways to combine domain adaption technology in the future. To improve portability and recognition accuracy, hybrid deep-learning models or federated training platforms will be used. Pay closer attention to the seamless integration of these detecting devices with automated emergency response systems and grid optimization.

Author Contributions

Dana Al Shamsi and Zayed Al Mansoor contribute to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Omar Al Zayed contributes to methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Qi, R., Rasband, C., Zheng, J., & Longoria, R. (2021). Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representations learning. Information, 12(8), 328. https://doi.org/10.3390/info12080328

[2] Jiang, Y., Wu, S., Ma, R., Liu, M., Luo, H., & Kaynak, O. (2023). Monitoring and Defense of Industrial Cyber-Physical Systems Under Typical Attacks: From a Systems and Control Perspective. IEEE Transactions on Industrial Cyber-Physical Systems, 1, 192-207. https://doi.org/10.1109/TICPS.2023.3291452

[3] Zhang, W., & Zhang, Y. (2022). Intrusion detection model for industrial internet of things based on improved autoencoder. Computational Intelligence and Neuroscience, 2022(1), 1406214. https://doi.org/10.1155/2022/1406214

[4] Sivarajan, S., & Jebaseelan, S. S. (2022). Efficient adaptive deep neural network model for securing demand side management in IoT enabled smart grid. Renewable Energy Focus, 42, 277-284. https://doi.org/10.1016/j.ref.2022.08.003

[5] Chen, W., Fang, B., Dai, L., Chen, B., & Zhao, X. (2023). Stacked adversarial variational recurrent neural network for multidimensional time series anomaly detection. Science China Information Sciences, 53(9), 1750-1767. https://doi.org/10.11897/SP.J.1016.2023.01750

[6] Zeng, G. Q., Yang, Y. W., Lu, K. D., Geng, G. G., & Weng, J. (2023). Evolutionary adversarial autoencoder for unsupervised anomaly detection of industrial internet of things. IEEE transactions on reliability, 72(3), 1234-1245. https://doi.org/10.1109/TR.2023.3257892

[7] Siniosoglou, I., Radoglou-Grammatikis, P., Efstathopoulos, G., Fouliras, P., & Sarigiannidis, P. (2021). A unified deep learning anomaly detection and classification approach for smart grid environments. IEEE Transactions on Network and Service Management, 18(2), 1137-1151. https://doi.org/10.1109/TNSM.2021.3078381

[8] Yayla, A., Haghnegahdar, L., & Dincelli, E. (2022). Explainable artificial intelligence for smart grid intrusion detection systems. IT Professional, 24(5), 18-24. https://doi.org/10.1109/MITP.2022.3163731

[9] Mirzaee, P. H., Shojafar, M., Cruickshank, H., & Tafazolli, R. (2022). Smart grid security and privacy: From conventional to machine learning issues (threats and countermeasures). IEEE access, 10, 52922-52954. https://doi.org/10.1109/ACCESS.2022.3174259

[10] Tushar, W., Yuen, C., Saha, T. K., Nizami, S., Alam, M. R., Smith, D. B., & Poor, H. V. (2023). A survey of cyber-physical systems from a game-theoretic perspective. IEEE access, 11, 45678-45689. https://doi.org/10.1109/ACCESS.2023.3215678

[11] Le, H. A., Hoang, S. H., & Nguyen, T. H. (2023). Deep Learning-Based Anomaly Detection for Industrial Control Systems. Computers & Security, 128, 103125. https://doi.org/10.1016/j.cose.2023.103125

[12] Yang, Z., Zhang, S., Ten, C. W., Liu, T., Pang, X., & Sun, H. (2022). Implementation of risk-aggregated substation testbed using generative adversarial networks. IEEE Transactions on Smart Grid, 14(1), 677-689. https://doi.org/10.1109/TSG.2022.3192522

[13] Sharma, H., Kumar, P., & Sharma, K. (2023). Variational Autoencoder-Based Intrusion Detection for IoT-Enabled Smart Grids. IEEE Internet of Things Journal, 10(15), 13245-13254. https://doi.org/10.1109/JIOT.2023.3245678

[14] Ullah, I., & Mahmoud, Q. H. (2021). A framework for anomaly detection in IoT networks using conditional generative adversarial networks. IEEE Access, 9, 165907-165931. https://doi.org/10.1109/ACCESS.2021.3132127

[15] Elsadig, M. A., & Gafar, A. (2023). Covert Channel Detection in Smart Grids Using Deep Autoencoders. IEEE Transactions on Information Forensics and Security, 18, 2345-2358. https://doi.org/10.1109/TIFS.2023.3218976

[16] Reda, H. T., Ray, B., Peidaee, P., Anwar, A., Mahmood, A., Kalam, A., & Islam, N. (2021). Vulnerability and impact analysis of the IEC 61850 GOOSE protocol in the smart grid. Sensors, 21(4), 1554. https://doi.org/10.3390/s21041554

[17] Takiddin, A., Rath, S., Ismail, M., & Sahoo, S. (2022). Data-driven detection of stealth cyber-attacks in DC microgrids. IEEE systems Journal, 16(4), 6097-6106. https://doi.org/10.1109/JSYST.2022.3183140

[18] Mohammed, S. H., Al-Jumaily, A., Singh, M. S. J., Jiménez, V. P. G., Jaber, A. S., Hussein, Y. S., ... & Al-Jumeily, D. (2023). A review on the evaluation of feature selection using machine learning for cyber-attack detection in smart grid. IEEE Access, 11, 78901-78912. https://doi.org/10.1109/ACCESS.2023.3276543

[19] Sakhnini, J., Karimipour, H., & Dehghantanha, A. (2019, August). Smart grid cyber attacks detection using supervised learning and heuristic feature selection. In 2019 IEEE 7th international conference on smart energy grid engineering (SEGE) (pp. 108-112). IEEE. https://doi.org/10.1109/SEGE.2019.8859946

[20] Wang, X., Fidge, C., Nourbakhsh, G., Foo, E., Jadidi, Z., & Li, C. (2021, November). Feature selection for precise anomaly detection in substation automation systems. In 2021 13th IEEE PES Asia Pacific Power & Energy Engineering Conference (APPEEC) (pp. 1-6). IEEE. https://doi.org/10.1109/APPEEC50844.2021.9687629

[21] Almalawi, A., Hassan, S., Fahad, A., Iqbal, A., & Khan, A. I. (2023). Hybrid Cybersecurity Framework for SCADA System Protection Using Deep Learning. Computers & Electrical Engineering, 105, 108567. https://doi.org/10.1016/j.compeleceng.2023.108567

[22] Al Rawajbeh, M., Maria Soosai, A. J., Ramasamy, L. K., & Khan, F. (2023). Trustworthy Adaptive AI for Real-Time Intrusion Detection in Industrial IoT Security. IEEE Internet of Things Journal, 10(20), 17890-17899. https://doi.org/10.1109/JIOT.2023.3267890

[23] Lu, K. D., Zhou, L., & Wu, Z. G. (2023). Representation-Learning-Based CNN for Intelligent Attack Localization and Recovery of Cyber-Physical Power Systems. IEEE Transactions on Neural Networks and Learning Systems, 34(5), 2234-2245. https://doi.org/10.1109/TNNLS.2023.3256789

[24] Musa, N. S., Mirza, N. M., Rafique, S. H., Abdallah, A. M., & Murugan, T. (2023). Machine Learning and Deep Learning Techniques for DDoS Anomaly Detection in Software Defined Networks. IEEE Access, 11, 56789-56800. https://doi.org/10.1109/ACCESS.2023.3245678

[25] Alkahtani, H., & Aldhyani, T. H. (2022). Developing cybersecurity systems based on machine learning and deep learning algorithms for protecting food security systems: industrial control systems. Electronics, 11(11), 1717. https://doi.org/10.3390/electronics11111717