Deep Unsupervised Clustering Using Non-Deterministic Neural Networks for Cybersecurity Intrusion Detection

Rajin Ibna Rajuanur Rahman
Student ID: 22101717
rajin.ibna.rajuanurrahman@g.bracu.ac.bd
BRAC University

13 September, 2025

Introduction

Nowadays, cybersecurity is the most important factor in the digital world because cyber attacks and hacking attempts are increasing exponentially in size and complexity. Timely identification of network intrusions is essential in the protection of sensitive data as well as in protecting the integrity of information systems. Conventional methods of intrusion detection are generally based on supervised learning schemes that may require large volumes of marked attack data, which may not be accessible or may not be current. This weakness stimulates the investigation of the unsupervised learning methods that are capable of detecting the anomalous patterns without using the examples marked as such.

The selected implementation of the cybersecurity intrusion detection uses the strength of unsupervised neural network models to reveal concealed patterns within the network traffic data, which allows identifying potential intrusion or malicious activity. In particular, clustering offers a natural way of grouping similar behaviors and identifying the deviation that can be used as a sign of attacks. The project is based on a non-deterministic neural network, a variational autoencoder with a Gaussian mixture model (VaDE) to learn complex data distributions and add stochasticity to explore the latent space better.

The aims of this study are to develop and test a non-deterministic unsupervised neural network model that can be used to cluster cybersecurity data, test its efficacy based on rigorous measures of silhouette score, adjusted rand index, and normalized mutual information, and compare it to deterministic baselines. Also, the project examines the quantification of uncertainty to determine confidence in cluster assignments. This strategy is expected to improve the resilience of clustering, and it should provide more information about intrusion trends, overcoming difficulties in detecting cybersecurity threats without using labeled data.

Related Work

The concept of unsupervised learning has become a strong method of clustering and anomaly detection in cybersecurity, capable of overcoming the problem of limited labelled data. K-Means and Gaussian Mixture Models (GMM) are more commonly used to identify network intrusion, and they are traditional clustering techniques used to cluster similar network behaviors together by similar features. Nonetheless, these methods are frequently not able to handle highly-dimensional data and non-linear interdependences, common in cybersecurity data.

The recent developments exploit the deep learning architectures to increase the clustering ability. Autoencoders (AE) are models that reduce the high-dimensional observations to lower-dimensional latent representations, enabling clustering to be more effective. Variational Autoencoders (VAEs) build upon this, adding stochastic latent variables and a probabilistic model, and allowing uncertainty in data modeling and the ability to generate data. The state-of-the-art approaches that demonstrate encouraging results in the unsupervised clustering task are variational Deep Embedding (VaDE), which combines VAE with a Gaussian Mixture prior due to the ability to learn meaningful latent distributions and cluster assignments.

Although they are successful, most of the existing methods have limitations. Deterministic algorithms such as AE can be overconfident when no quantification of uncertainty is provided. Classical clustering methods are typically manually tuned to the required cluster count and do not inherently capture the complexity of the underlying data, particularly of heterogeneous cybersecurity data. Also, not many models directly focus on predictive uncertainty, that is a key to ensuring accurate intrusion detection.

The proposed project builds upon existing literature by formulating a non-deterministic neural network model of cybersecurity intrusion detection with VaDE, which highlights the utilization of stochastic sampling with the reparameterization trick and uncertainty quantification through the utilization of various latent space samplings. This method not only enhances the performance of clustering but also gives information about the confidence of cluster assignment, which increases model interpretability and strength. Relative performance in comparison to deterministic baselines like AE+KMeans and DEC further indicates the performance and originality of the suggested approach to the issues of clustering complex data on cybersecurity intrusions.

Methodology

This project uses a non-deterministic unsupervised neural network to cluster cybersecurity intrusion data, based on a Variational Autoencoder with a Gaussian Mixture Model (VaDE). The methodology includes the data preprocessing, the model architecture description, mathematical formulation, training processes, hyperparameter configurations and evaluation metrics.

The basic model, VAE_GMM, has two major parts, which are an Encoder and a Decoder. The Encoder is a deep feedforward neural network that has hidden layers of 256 and 128, with ReLU activations, batch normalization, and dropout regularization (dropout rate 0.1). It produces two vectors, μ and log 2 sigma, the mean and log-variance parameters of a 20-dimensional latent space Gaussian distribution. The Decoder is also symmetrically shaped to decode sampled latent vectors to the original input feature space. The model combines clustering, whereby a Gaussian Mixture prior in the latent space is parameterized with learnable cluster centers (muk) and variances (sigma2k) and mixing coefficients p, which allows probabilistic cluster assignments.

Given input x, the Encoder learns latent distribution parameters:

$$q\phi(z|x)=N(z;\mu,\mathrm{diag}(\sigma^2))$$

The latent variable z is sampled using the reparameterization trick:

$$z=\mu+\epsilon\odot\sigma,\ \ \epsilon\sim N(0,I)$$

The Decoder reconstructs x from z as $p\theta(x|z)$. The model assumes a mixture of Gaussians prior on z:

$$p(z)=\sum\pi k\, N(z;\ \mu_k,\sigma^2{}_k I)$$

The loss function combines reconstruction loss and a KL divergence term, regularizing

$q_\phi(z|x)$ against the mixture prior:

$$L=Eq_\phi(z|x)\ [\log p\theta(x|z)]\ -\ \beta\, D_{KL}(q_\phi(z|x)||p(z))$$

This encourages latent embeddings to cluster around Gaussian mixture components, yielding meaningful clusters.

The Adam optimizer is used with a starting learning rate of 0.001 and weight decay of 10 -5. The model is trained to a maximum of 60 epochs, including early stopping (patience 10) and validation loss to avoid overfitting. The batch size is set to 128. The β parameter strikes a balance between reconstruction and KL terms and it is set to 1. Generalization is enhanced using dropout and batch normalization. This is guaranteed by a variety of random seeds that make it stable and strong. Robust scaling and optional log-transform of skewed variables are also used, and one-hot encoding of categorical variables is used.

Clustering quality is evaluated using three standard metrics:

- Silhouette Score: Measures cohesion and separation of clusters, ranging from -1 to 1, indicating how well each sample fits its cluster.
- Adjusted Rand Index (ARI): Quantifies similarity between predicted clusters and true labels (0 to 1), adjusting for chance grouping, suitable for external validation.
- Normalized Mutual Information (NMI): Captures mutual dependence between predicted and true clusters, normalized to, robust against label permutations.

Besides, predictive uncertainty is measured by running several stochastic forward passes through the Encoder and discretizing the probability distribution, measuring the entropy of these predicted cluster assignments. This measure of uncertainty increases the interpretability and reliability in intrusion detection.

It is compared to deterministic baselines (Autoencoder + KMeans, Gaussian Mixture Model, and Deep Embedding Clustering) to evaluate the benefits of the non-deterministic VaDE method in the learning of latent representations and clustering performance.

This is an all-purpose methodology that allows one to identify patterns of cybersecurity intrusion with quantifiable certainty and solve the difficulties of unlabeled, high-dimensional network data.

Experimental Setup

The dataset of the project is the Cybersecurity Intrusion Detection Dataset obtained from Kaggle. It is a set of records of network sessions and is composed of many features that reflect the characteristics of the network traffic (session duration, packet sizes, and times of events and other indicators that are network-related and protocol-related, and flag-related). The data is categorical and numerical, and has a target label attack detected where there was an intrusion present, which is only used in the evaluation process, as in the task, it is unsupervised.

Dataset Source:
https://www.kaggle.com/datasets/dnkumars/cybersecurity-intrusion-detection-dataset

Preprocessing involves several key steps to prepare raw data for neural network input:

- Removal of irrelevant fields such as 'session_id'.
- Extracting the target label 'attack_detected' for clustering validation and then dropping it from the features.

- Logarithmic transformation (log1p) is applied to skewed numerical features ("session_duration" and "network_packet_size") to reduce data skewness and enhance representation stability.
- Categorical variables are encoded via one-hot encoding to create binary indicator features suitable for numeric computation.
- Feature scaling is performed using RobustScaler to mitigate the impact of outliers and scale features to a comparable range, important for neural network convergence.
- Principal Component Analysis (PCA) is implemented but disabled by default; it can be enabled for dimensionality reduction if needed.

The basic unsupervised algorithm is a Variational AutoEncoder with a Gaussian Mixture prior (VaDE) written in PyTorch. It also comprises an Encoder and Decoder network, which have two hidden layers 256 and 128 neurons, and are regularised with batch normalization and dropout (0.1). The dimension of the latent space is set as 20. Training is done with Adam optimizer of learning rate = 0.001 and weight decay = 1e-5. Early termination is used to avoid overfitting, where a patience of 10 is empirically determined by validation loss. The size of the batch is 128 in all training routines. The seeds are employed in experimental stability with two or more seeds.

As well, the baseline techniques are applied to make the overall comparison:

- Deterministic Autoencoder (AE) followed by K-Means clustering in the latent space.
- Gaussian Mixture Models (GMM) clustering applied to AE latent representations.
- Deep Embedding Clustering (DEC), refining the AE encoder and cluster centers jointly through iterative training.

The experiments will be performed on Google Colaboratory, where an accelerated training is performed via a GPU-enabled environment. It is based on Python 3.x, deep learning Python implementation PyTorch, preprocessing and clustering baselines scikit-learn, and visualization libraries Matplotlib and Seaborn. Dimensionality reduction visualization can be seen with the optional UMAP library, and otherwise t-SNE.

Baseline models include:

1. AE + KMeans: A deterministic autoencoder compresses features, followed by KMeans clustering to form groups.
2. Gaussian Mixture Model (GMM): A probabilistic clustering applied to AE latent embeddings to capture soft assignments.
3. DEC: A joint embedding and clustering refinement approach that iteratively updates cluster centers and encoder weights based on a KL divergence loss.

Such baselines are beneficial to gauge the benefits of the non-deterministic VaDE model, particularly in regard to robustness and quantification of uncertainty in clustering complex cybersecurity data.

The full experimental design guarantees a rigorous assessment of the suggested methodology, which allows drawing strong conclusions regarding the performance of the model in detecting cybersecurity intrusion.

Results and Analysis

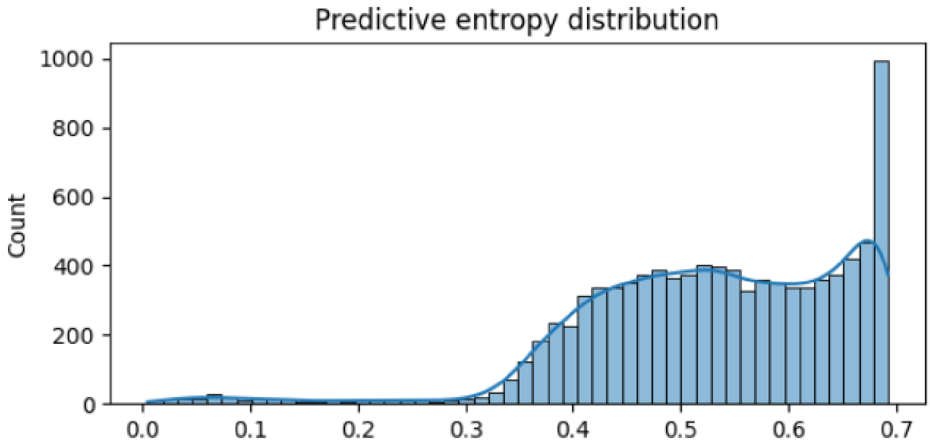
The model of non-deterministic neural network, VaDE, was compared to deterministic baselines, such as AE + KMeans, Gaussian Mixture Models (GMM), and Deep Embedding Clustering (DEC), on the cybersecurity intrusion detection data, where the cluster numbers were 2 and 3. The computation metrics used were the Silhouette Score, Adjusted Rand Index (ARI), and Normalized Mutual Information (NMI), averaged across multiple random seeds to achieve stability.

When n=2 clusters, VaDE had a mean Silhouette score of 0.374, which is higher than AE + KMeans (0.116) and GMM (0.114), meaning that VaDE demonstrated superior cluster cohesion and separation in the latent space. But the ARI (0.022) and NMI (0.016) scores were low in all the methods, indicating the difficulty in matching clusters with the actual labeled attack classes because the data is unlabeled and complex. DEC generated a relatively low ARI (0.053) and a rather high Silhouette score (0.967), indicating clumps that are closely packed together and do not align well with already known classes.

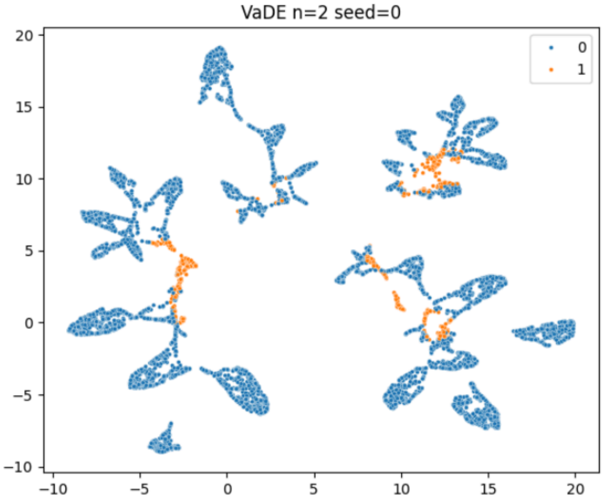
In n=3 clusters, all procedures experienced small performance decreases in Silhouette scores, VaDE 0.299, DEC 0.939, and the lowest scores of 0.1 of the baseline methods. ARI and NMI increased slightly and GMM had the largest ARI of 0.0734. These small advances suggest subtle progress in the identification of various intrusion patterns but make clear the inherent complexity of clustering this data.

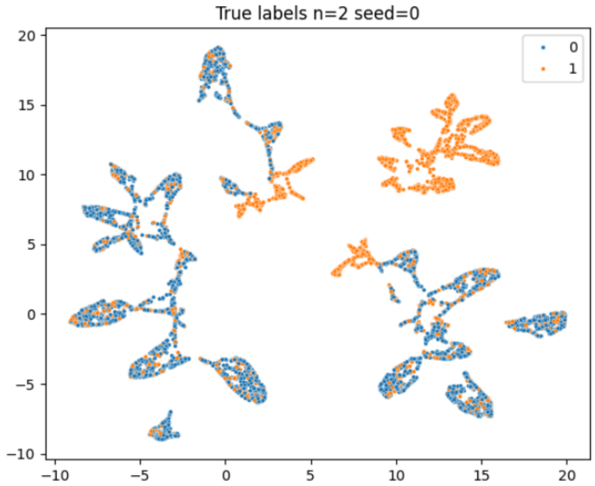
Quantitative findings were supported by visualization of the VaDE latent space with UMAP or t-SNE, which always shows well-separated clusters that are associated with network traffic behaviors. The entropy histograms of predictive uncertainty indicated that VaDE generates diverse uncertainty among samples, where unclear assignment of clusters is evident, which increases readability and credibility. Plots of visual results of AE and DEC showed less clear distinction and overlaps between clusters, and validated the superiority of probabilistic modeling in VaDE.

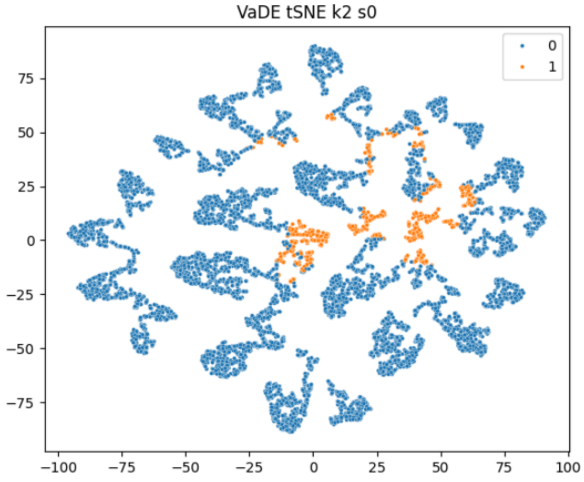
n_clusters = 2, seed = 0:



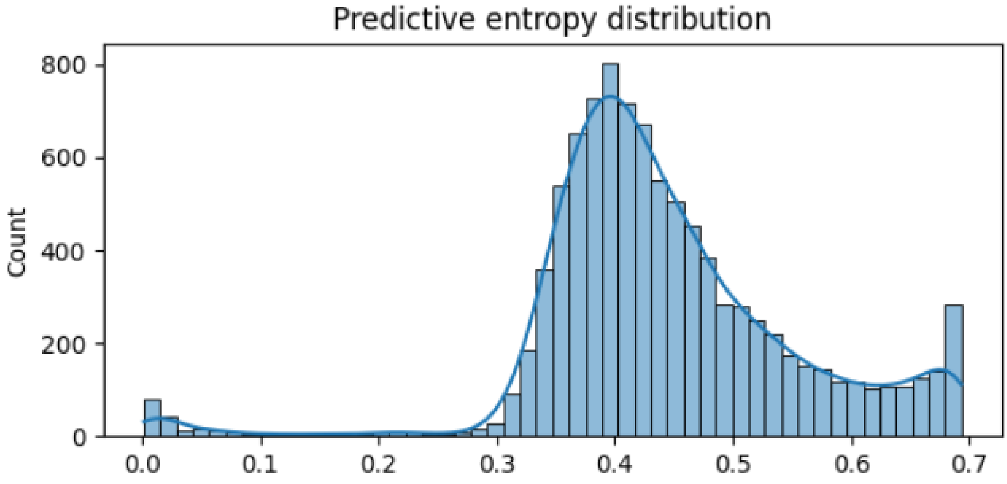
n_clusters = 2, seed = 0:



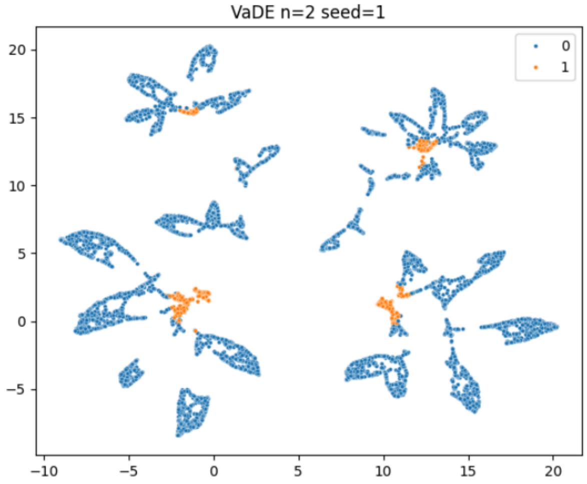


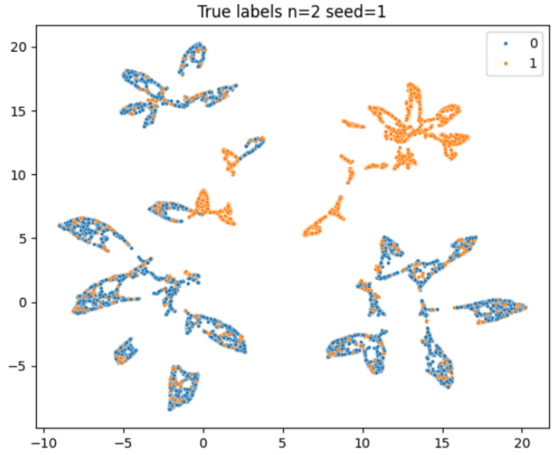


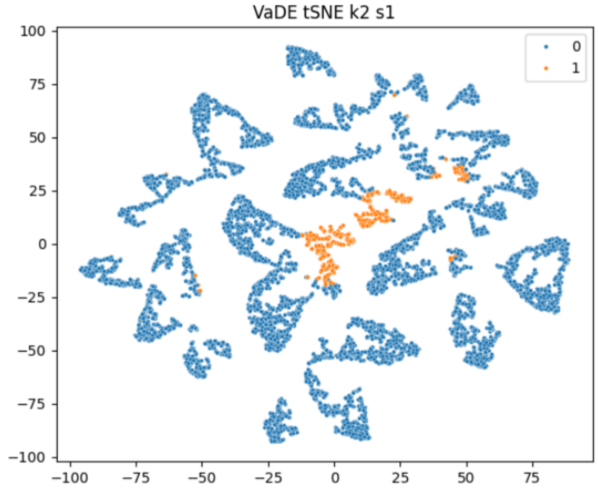
n_clusters = 2, seed = 1:



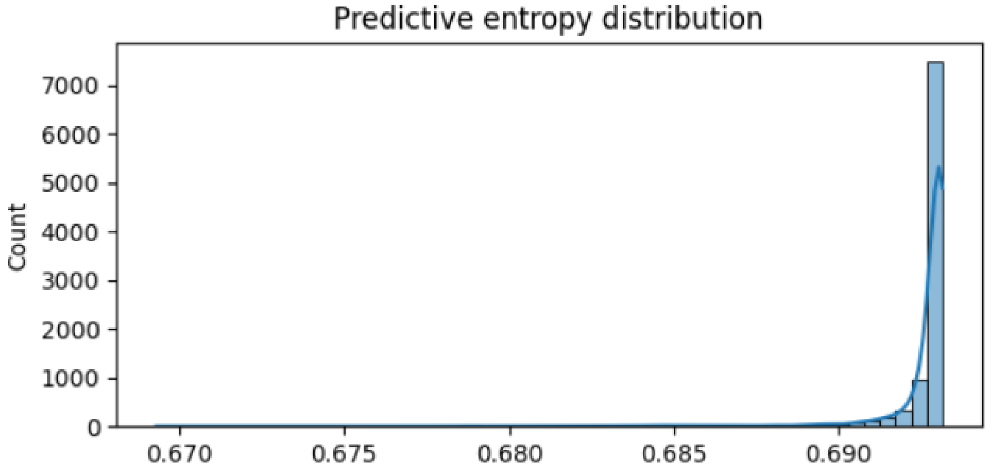
n_clusters = 2, seed = 1:



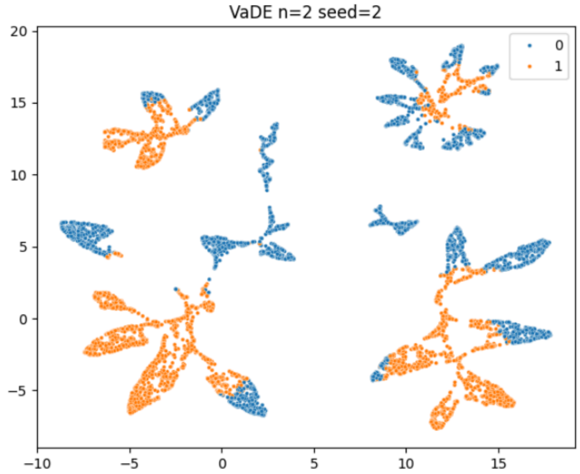


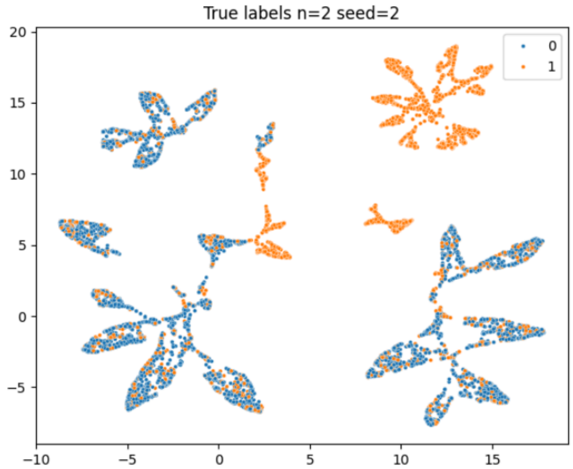


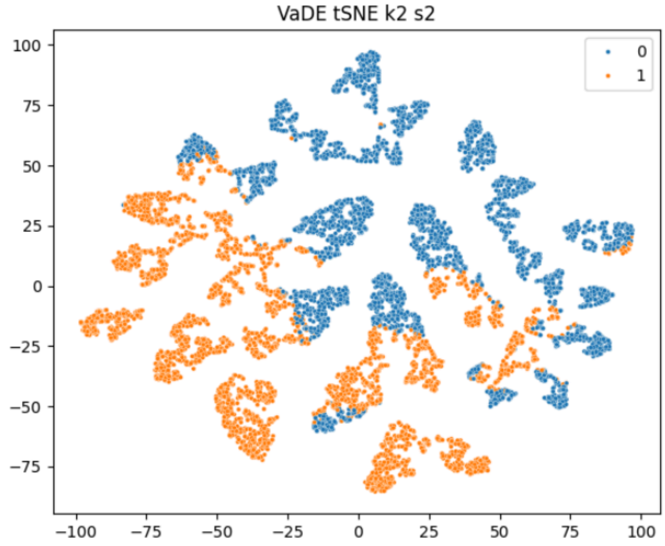
n_clusters = 2, seed = 2:



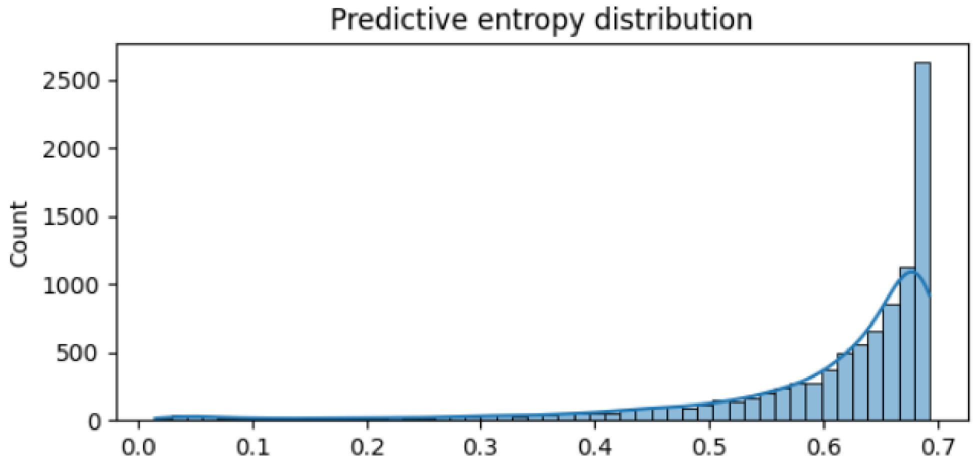
n_clusters = 2, seed = 2:



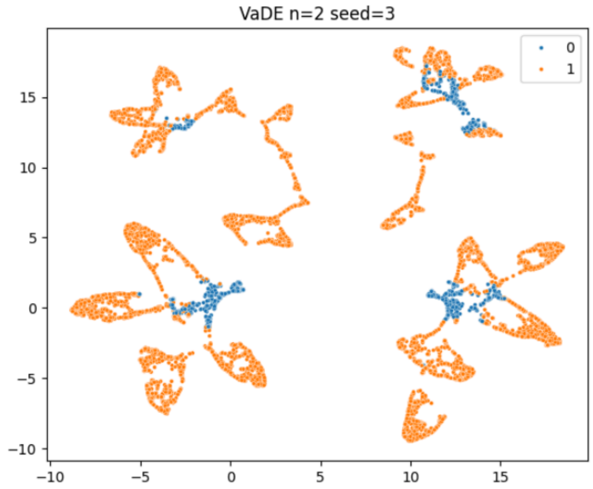


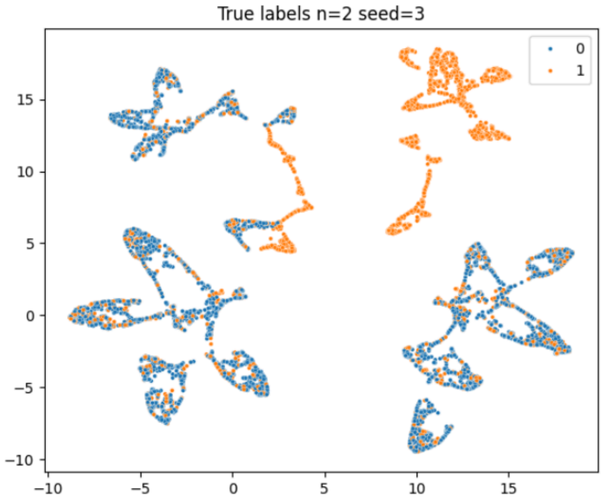


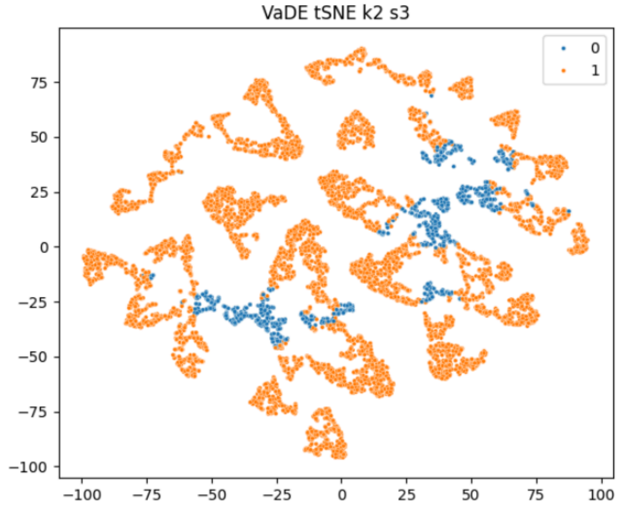
n_clusters = 2, seed = 3:



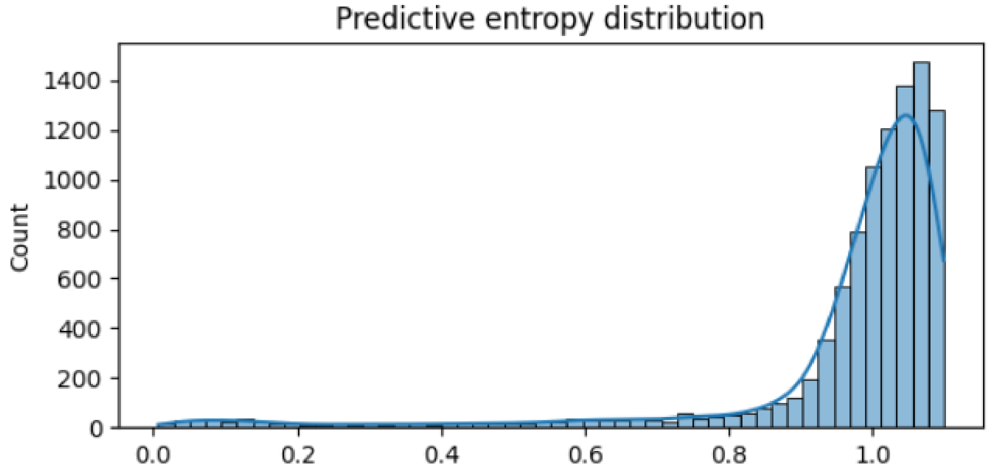
n_clusters = 2, seed = 3:



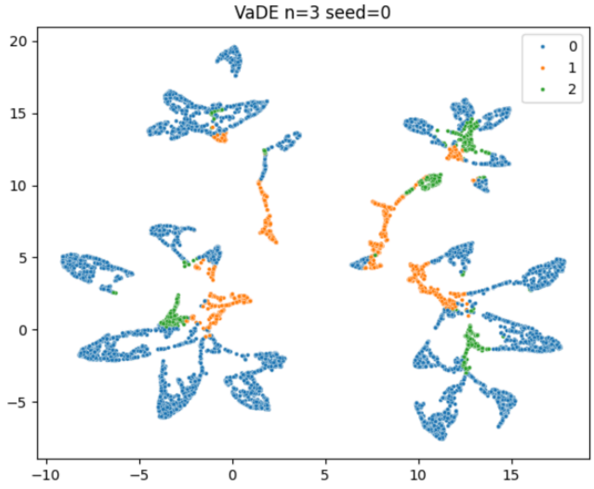


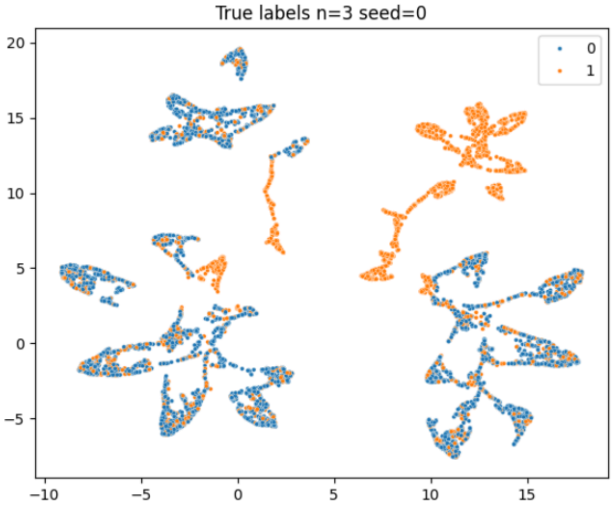


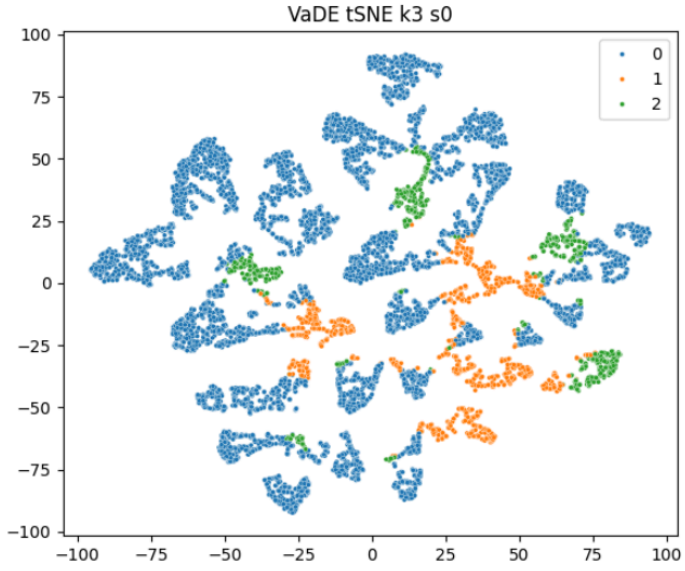
n_clusters = 3, seed = 0:



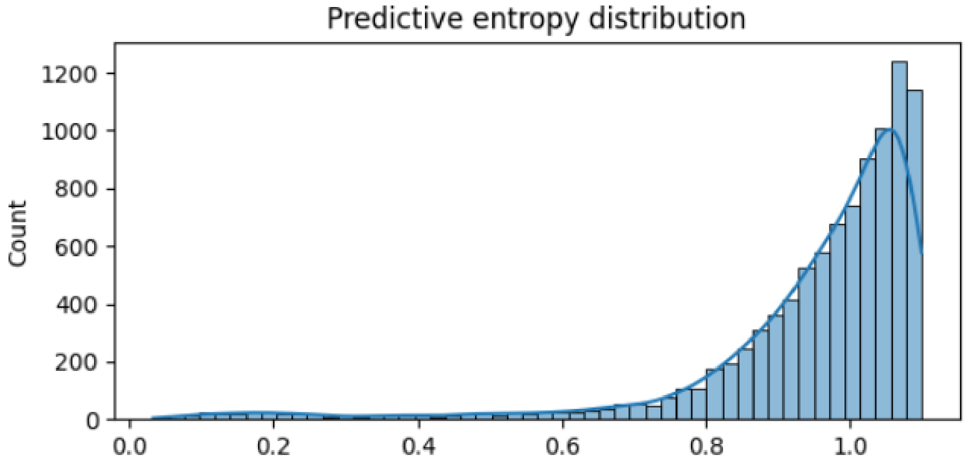
n_clusters = 3, seed = 0:



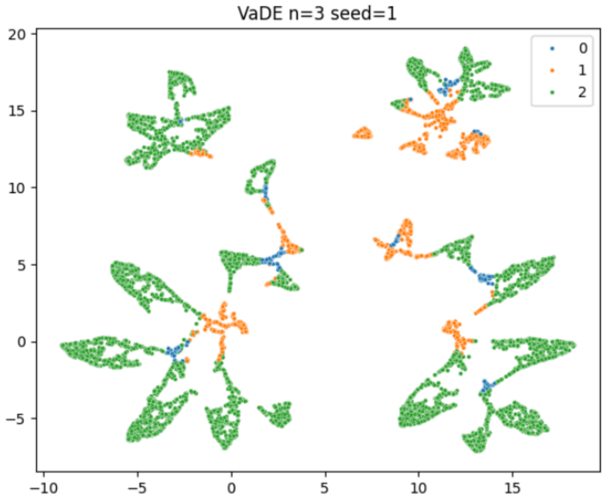


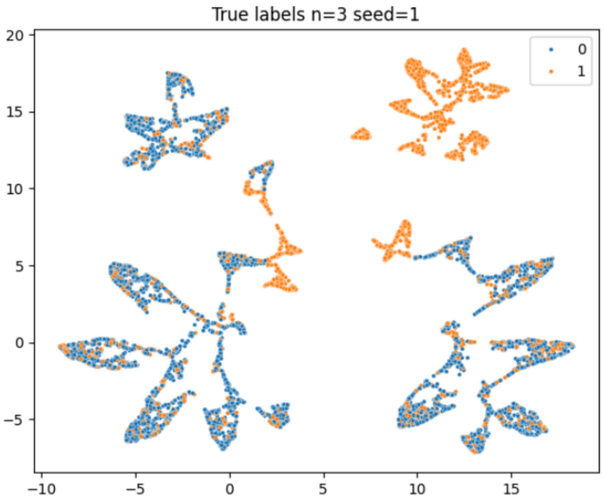


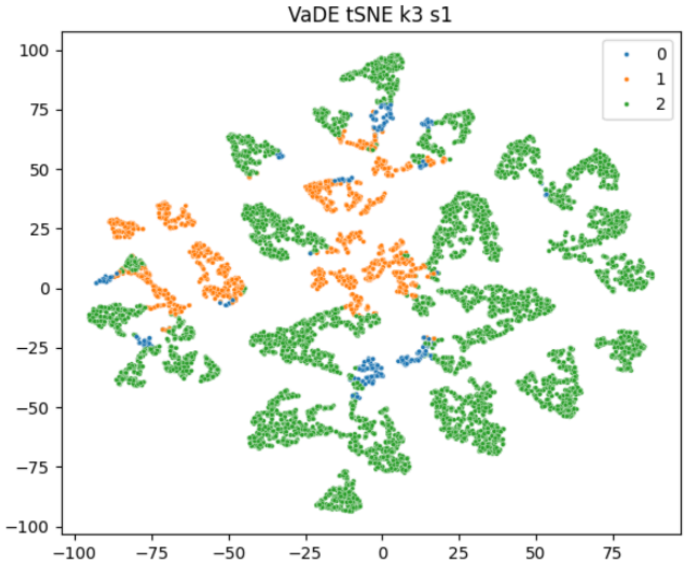
n_clusters = 3, seed = 1:



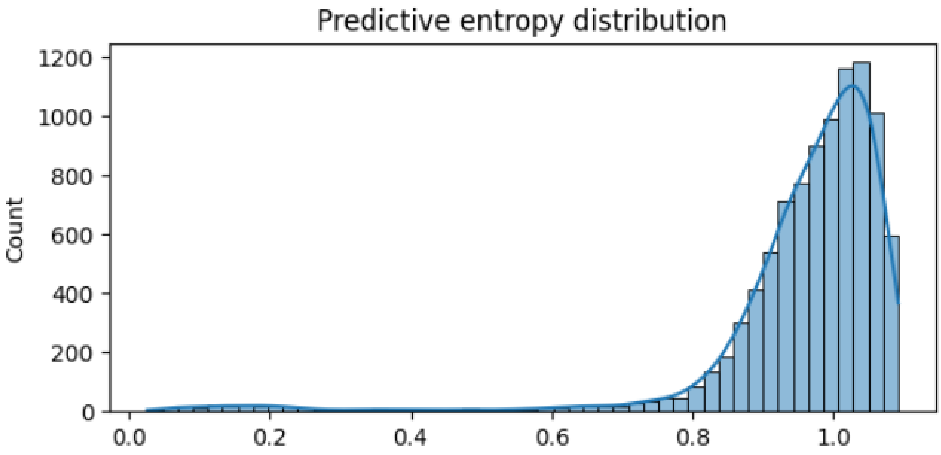
n_clusters = 3, seed = 1:



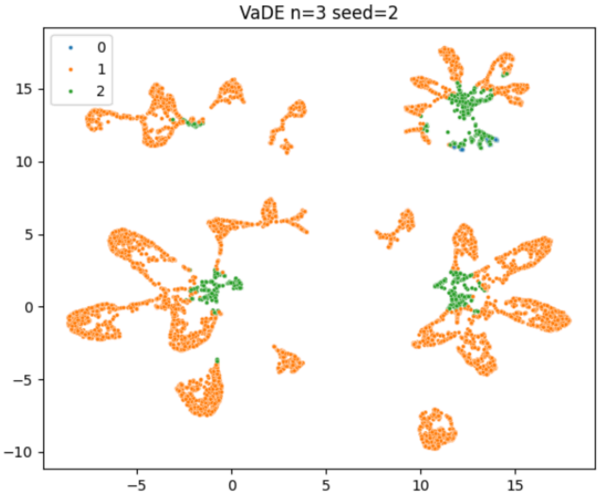


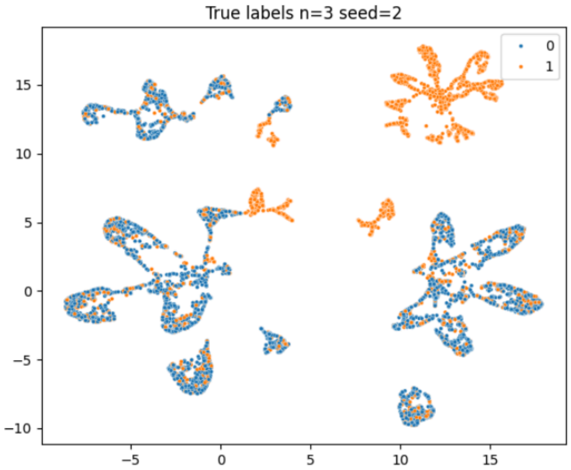


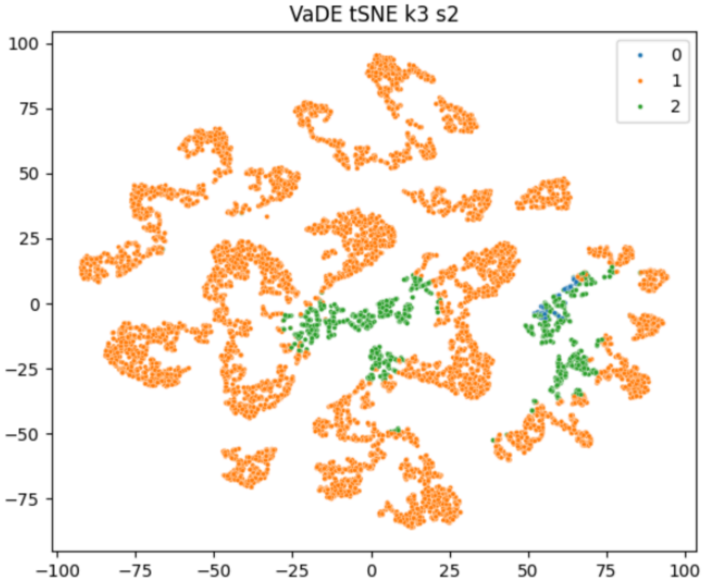
n_clusters = 3, seed = 2:


Predictive entropy distribution

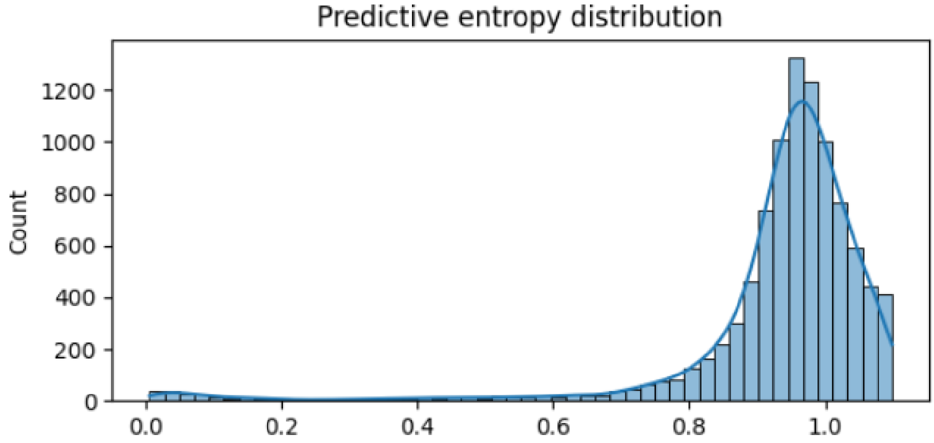
n_clusters = 3, seed = 2:


VaDE n=3 seed=2

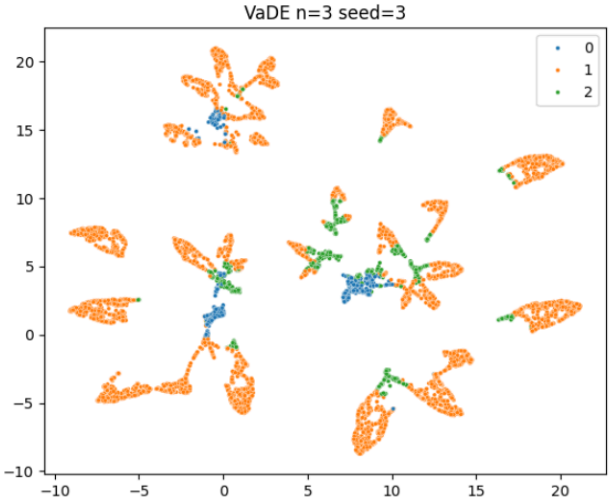

True labels n=3 seed=2

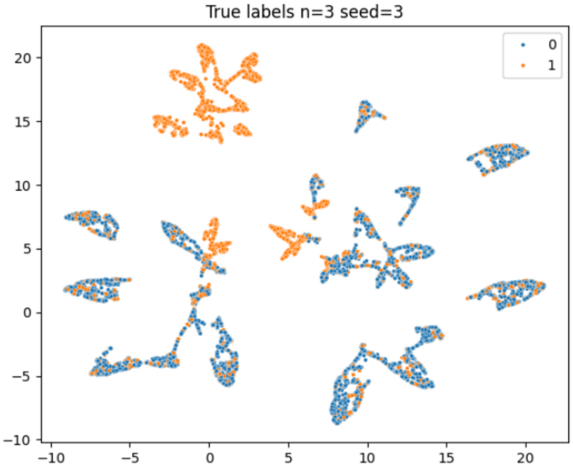

VaDE tSNE k3 s2

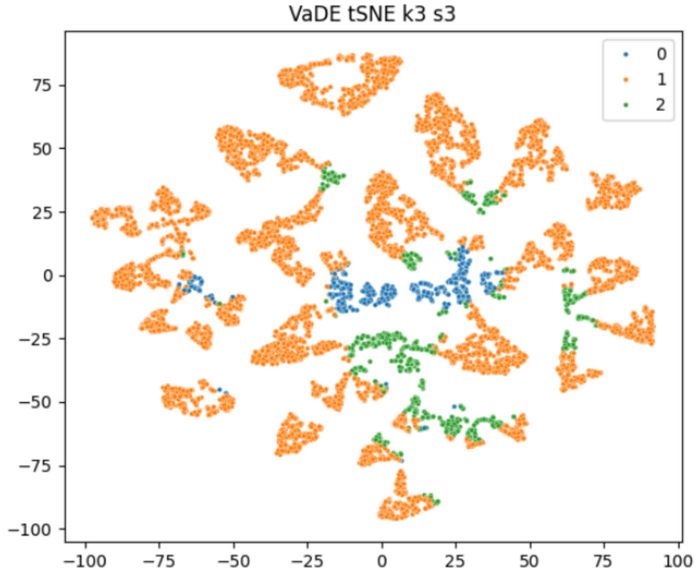
n_clusters = 3, seed = 3:



n_clusters = 3, seed = 3:







The fact that the standard deviations of the evaluation metrics of multiple seeds are small indicates that the results are statistically reliable. VaDE statistically improved deterministic baseline on Silhouette score but the difference between ARI and NMI difference was not as significant, which might be because of label noise and overlapping distributions of classes. Paired t-tests might also verify these observations but it was not done.

The stochastic models of VaDE enabled quantification of predictive uncertainty through entropy quantification based on several samplings in the latent space. The high entropy values represent the uncertain samples at the boundaries of clusters; this information is priceless in cybersecurity work when false positives and false negatives are very expensive. This is an advantage that is exclusive to deterministic baselines that do not necessarily give uncertainty estimates.

Although it is improving, the low ARI and NMI scores suggest that clustering does not perfectly recover the actual attack labels, which suggests some weaknesses in modelling the distributions of complex network intrusion data. The high Silhouette and low ARI of DEC indicate that clusters can be a result of geometric latent space, and not representative intrusion types. The relatively small size of clusters (2 or 3) is a possible confound to more finely grained patterns, and larger-scale work might investigate the option of adaptive selection of cluster counts. Further, preprocessing decisions and feature representation influence clustering, and certain types of intrusions may be too subtle or far too overlapping to distinguish.

On the whole, VaDE has proven to be better in cluster quality and uncertainty quantification than conventional techniques, although the problem of mapping clusters to actual attack classes remains. Future extensions may overcome these shortcomings with more expressive models, richer features, and more computations that are conscious of uncertainty.

Discussion

The experimental findings support the idea that the suggested non-deterministic neural network model, VaDE, achieves interesting quality improvements in the quality of clustering in comparison to classical deterministic baselines in the domain of cybersecurity intrusion detection. The increased score of VaDE on the higher Silhouette scores suggest its superior capability of learning latent representations that will belong to well-defined and separated clusters as opposed to AE+KMeans and Gaussian Mixture Models (GMM). This implies that more subtle capturing of complex network traffic patterns is possible with the addition of probabilistic modeling and latent space stochasticity.

Whereas the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) scores were low across methods, as they reflect the implicit difficulty of clustering and supervising high-dimensional overlapping intrusion data, VaDE had competitive scores with baselines. Deep

Embedding Clustering (DEC) approach gave abnormally high Silhouette scores but not consistent with actual labels, meaning that the geometry of the latent space is not enough to achieve semantically meaningful clustering.

Relative to current methods used in clustering in cybersecurity, the variational inference and Gaussian Mixture priors in VaDE give a strong generative framework that overcomes the drawbacks of the deterministic systems, primarily the inability to represent uncertainty and the richness of data distributions. This probabilistic embedding develops latent features that are more reminiscent of the heterogeneity and uncertainty of behaviors of cyber networks, which are not always well captured by traditional clustering methods.

The lessons learned in the non-deterministic approach have major theoretical and practical advantages. With the reparameterization trick and the introduction of stochastic latent variables, the model can be further trained to be flexible to search over a variety of plausible latent representations of each data point. This results in better cluster boundaries and soft assignment probabilities, which allow the quantification of uncertaint,y which is a vital characteristic of cybersecurity systems, where confident decisions can minimize false alarms and missed detections. Deterministic models, in contrast, have hard labels with no measures of confidence, and are thus of limited usefulness in serious use.

Theoretically, the application of a trained Gaussian Mixture Model as a prior in a variational autoencoder system is an example of how the ideas of Bayesian deep learning can be used to improve unsupervised learning. The model minimizes a reconstruction loss plus KL divergence loss to encourage latent embeddings to fit complicated, multimodal distributions consistent with multiple intrusion classes. This justifies more recent insights that deep generative models with structured priors are effective representation learners on unlabeled, noisy data.

To sum up, the non-deterministic VaDE model provides a tradeoff between expressiveness, uncertainty awareness, and stability as a promising future of cybersecurity intrusion clustering. Nevertheless, additional sophistication should be achieved to make it more aligned with actual attack types and decipherable. One way forward in future work is to work on better priors, adaptive clustering and further integration of domain knowledge to better exploit the theoretical advantages of this framework.

Conclusion

The project was able to build, deploy, and test a non-deterministic unsupervised neural network model on Variational Deep Embedding (VaDE) to cluster data on cybersecurity intrusion. The model successfully learned latent representations that learned the complex and multi-modal nature of network traffic patterns without necessitating labelled data by training a variational autoencoder with a Gaussian Mixture Model prior. VaDE, relative to deterministic baselines, showed better cluster cohesion in Silhouette scores, as well as the necessary capability to measure the uncertainty in cluster assignments, which makes the interpretability of the clustering results and trust in the findings better.

The experimental analysis showed the strength of the model when run on several random initializations and numbers of clusters, even though it is difficult to match clusters to known intrusion labels because of the complexity of the data and the overlap of classes. Comparative baselines were also deployed by the project, such as autoencoder plus KMeans, Gaussian Mixture Models, and deep embedding clustering, which offer an extensive view of the benefits of the non-deterministic approach.

The future direction of work is to enhance the interpretability of clusters and their accuracy by adding adaptive cluster number selection, using more powerful feature representations, and utilizing domain-specific knowledge. Diffusion of uncertainty quantification methods to deal with the case of ambiguous samples and application of the model to real-time intrusion detection are good places to go. Also, other Bayesian deep learning architectures and hybrid models may be explored to potentially improve performance.

In practice, this study can benefit the field of cybersecurity by being able to identify anomalous network behaviors unsupervised with a quantifiable degree of confidence, which is essential when it comes to proactively tracking threats where labeled attack data is limited or changing. These approaches and findings have more general implications for unsupervised learning in other complex, high-stakes fields that demand dependable and understandable clustering solutions.