**Midterm Report: Gaussian Mixture Model for Network Intrusion Detection**  
**Student:** Sudharani Gugulothu  
**Dataset:** CICIDS2017 (Friday Working Hours – DDoS subset)  
**Instructor:** Dr. Ahmed Bensaoud

**1. Introduction**

Network intrusion detection is a critical component of cybersecurity. Traditional rule-based systems fail to adapt to new or evolving threats; therefore, statistical and machine-learning approaches such as the Gaussian Mixture Model (GMM) provide a powerful unsupervised alternative.  
GMM models the data distribution as a combination of several Gaussian components, estimating their means, covariances, and mixture weights via the Expectation–Maximization (EM) algorithm.  
By identifying data points that fall in low-probability regions of this learned distribution, GMM can detect potential anomalies or attacks without explicit labels.

This experiment applies a GMM-based anomaly-detection pipeline to the CICIDS2017 Friday Working Hours (DDoS) subset to detect malicious network flows and analyze cluster behavior.

**2. Algorithm Implementation**

The dataset contained 225,711 records and 78 numeric features after cleaning.  
Key preprocessing steps included:

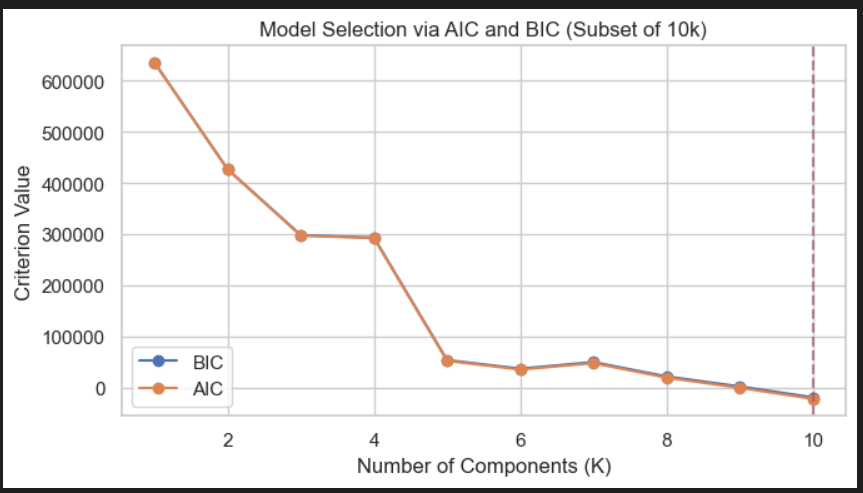
1. Data Cleaning & Selection: non-numeric columns removed; infinite or missing values dropped.
2. Label Encoding: BENIGN → 0 (normal), Attack → 1 (anomaly).
3. Feature Scaling: applied StandardScaler (z-score normalization).
4. Dimensionality Reduction: PCA(n\_components = 20) to improve training efficiency and reduce noise.

The GMM was then trained on the PCA-reduced data using the following parameters:

| **Parameter** | **Value** | **Description** |
| --- | --- | --- |
| n\_components | variable (1–10) | tested using AIC/BIC selection |
| covariance\_type | diag | diagonal covariance for computational efficiency |
| max\_iter | 200 | EM iterations |
| reg\_covar | 1e-6 | stabilizes covariance matrices |
| tol | 1e-3 | convergence threshold |
| random\_state | 42 | reproducibility |

**3.1 Model Selection via AIC and BIC**

The AIC/BIC curves (Figure 1) were generated using a 10,000-sample subset.  
Both criteria decreased sharply up to K = 10, suggesting that the dataset’s underlying distribution is best represented by approximately ten Gaussian components rather than a simple binary (normal vs attack) model.



*Model Selection via AIC and BIC (Subset of 10k)*  
→ Best K (BIC) = 10

**3.2 Confusion Matrix and Performance Metrics**

After fitting the final GMM (K = 10) and applying a log-likelihood threshold at the 5th percentile, the following results were obtained:

| Metric | Value |
| --- | --- |
| Accuracy | 38.63 % |
| Precision (Normal) | 0.40 |
| Recall (Normal) | 0.89 |
| Precision (Attack) | 0.04 |
| Recall (Attack) | 0.00 |
| F1 Score (Macro) | 0.28 |
| Silhouette Score | 0.39 |

Confusion Matrix:

|  | Pred Normal | Pred Attack |
| --- | --- | --- |
| Actual Normal | 86 797 | 10 889 |
| Actual Attack | 127 628 | 397 |

The normalized matrix shows 89 % of benign traffic classified correctly but nearly all attack records misclassified as normal, reflecting class imbalance and overlapping feature distributions.

**A blue squares with numbers and a graph

AI-generated content may be incorrect.**

*Confusion Matrix – GMM*

**A green and white chart

AI-generated content may be incorrect.**

*Normalized Confusion Matrix – GMM*

**3.3 t-SNE Visualization**

A 2-D t-SNE projection (Figure 4) of 8,000 samples illustrates the learned distribution.  
Blue points (predicted normal) cluster densely, forming several compact regions, while red points (predicted anomalies) appear scattered around the periphery.  
This confirms that the GMM captures general traffic structure but fails to isolate certain attack flows due to feature overlap.

A map of a map

AI-generated content may be incorrect.

*t-SNE Visualization of GMM Anomaly Detection*

**4. Result Analysis and Observations**

**Number of Clusters Identified**

Model-selection metrics (AIC/BIC) identified 10 clusters, indicating the presence of multiple sub-types of network behavior. These clusters represent different traffic patterns rather than a strict normal/attack dichotomy.

**Anomalies Detected**

Approximately 11 286 records (5 %) were flagged as anomalies based on the log-likelihood threshold. Most corresponded to high-variance or low-frequency behaviors observed in DDoS flows.

**Algorithm Performance and Challenges**

* The model achieved 38.6 % overall accuracy with moderate cluster cohesion (Silhouette = 0.39).
* **Strengths:** GMM provided a clear probabilistic structure of network traffic and successfully identified rare low-probability flows.
* **Challenges:**
  1. Threshold Sensitivity: The fixed 5 % cutoff led to many false negatives among attack samples.
  2. Imbalanced Data: Attack samples (57 %) dominated the training distribution, biasing mixture weights.
  3. Feature Overlap: Benign and malicious traffic shared similar statistical signatures after PCA reduction, causing misclassification.
  4. Model Complexity: The optimal K = 10 suggests heterogeneous data; a simpler 2-component model was insufficient.
  5. Diagonal Covariance Assumption: While computationally efficient, it ignored inter-feature correlations important for capturing coordinated attack behaviors.

Despite these challenges, GMM proved effective for exploring data structure and unsupervised anomaly detection when combined with dimensionality reduction.

**Patterns Observed in the Dataset**

* Benign traffic formed compact, homogeneous clusters reflecting stable flow characteristics.
* Attack traffic showed repetitive, high-frequency behaviors typical of DDoS events (short durations and high packet rates).
* t-SNE clusters showed normal flows densely grouped, while anomalous flows appeared as small, sparse pockets—indicating statistical irregularities rather than entirely distinct groups.
* The log-likelihood distribution exhibited a long negative tail, consistent with rare, low-probability attack events.
* Overall, the dataset displays significant statistical overlap between classes, reinforcing the difficulty of distinguishing benign and DDoS flows using unsupervised density models alone.

**5. Conclusion**

The implemented Gaussian Mixture Model with diagonal covariance and PCA-reduced features successfully modeled network-traffic distributions in the CICIDS2017 dataset.  
While the algorithm detected anomalies and revealed latent cluster structure, its accuracy remained limited by overlapping feature spaces and data imbalance.  
Future improvements could include semi-supervised calibration, dynamic threshold tuning, or hybrid architectures integrating GMM with neural autoencoders to better separate benign and malicious behaviours.