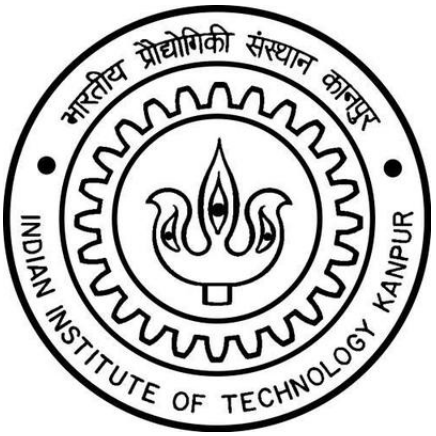

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

CS 631 Semester 2021–2022-I: Project

Anomaly Detection for Cyber Physical Systems

TEAM - Bob the Builder

1 INTRODUCTION

Cyber-Physical System (CPS) is collection of networking, computation and physical processes. Embedded computers and networks monitor and control the physical processes, with feedback loops where physical processes influence computations and vice versa. We cannot underestimate security of these critical Industrial Control Systems (ICS) as affects of any cyber attack is not limited to only financial loss these days. Attacks on critical infrastructures can take human life on stake as well such as attack on nuclear gas plant or chemical plants. We cannot ignore some of past notorious attacks i.e. Stuxnet [1], Triton [2] , the Maroochy water breach [3] and many more. Because of the great rewards that the attack on ICS, these frameworks are becoming appealing focuses for attackers.

2 OBJECTIVES

The goal of this work are following:

- Implement univariate PASAD [4] and multivariate M-PASAD [5] model for chosen dataset and compare both model.
- Develop unsupervised clustering ML models for chosen dataset.
- A comparative analysis between each implemented models.

3 Threat Model

ICS system (Fig:1) may have several threat vector, some of them are mentioned below:

- The communication link between the PLC and SCADA.
- APT Malware threat, Insider threat and Dos attack.
- Network based attack can also change dynamics of the system.
- Need to have some continuous monitoring system to monitor continuous system behaviour.

4 DATASET

The dataset [4] we have chosen to carry out this work is based on **Tennessee-Eastman** process control model [7] and created by developing a set of sophisticated attacks on TE simulation model which simulates real plant-wide chemical process. In a broader view we can understand the process flow of simulation model as; TE process produces 2 liquid products i.e. G and H from 4 gaseous reactants i.e. A,C,D and E as well as an inert B and byproduct F. They have monitored total 41 sensors reading labeled as XMEAS. The controller reads XMEAS data and sends commands to control various process flows. This dataset has been prepared by simulating integrity attacks on both actuators and sensors. Attack has been performed by two ways, either tamper command sent by controller to actuator or tamper process reading transmitted by sensors. Basically they have performed two types of attacks: *(i) Stealth attacks* - objective is to degrade process performance and cause slow damaging perturbations, and *(ii) Direct Damage attack* - aim is to damage equipments such as pipes or reactor, which are necessary to run the process.

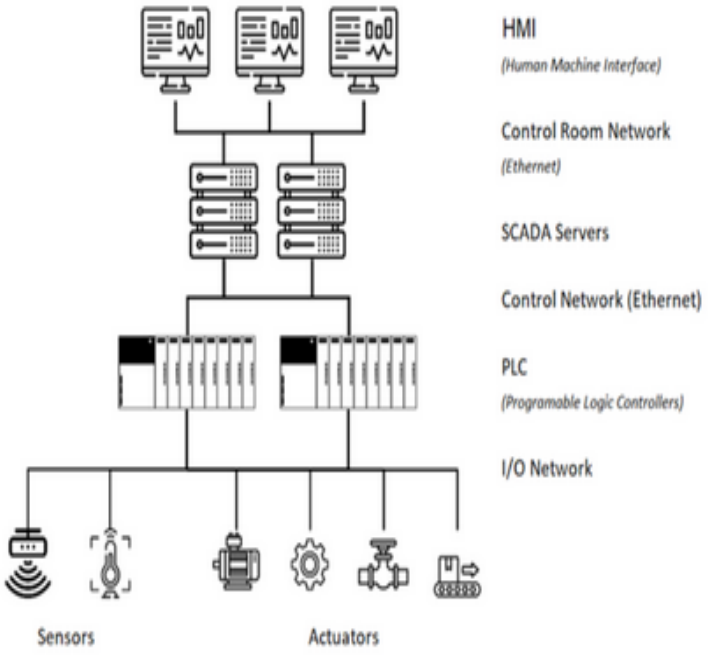

Figure 1: Overview of ICS system [6].

Stealth attacks are usually try to be undetected and keep process reading under a set of thresholds. In this dataset we have total three stealth attacks files i.e. SA1, SA2 and SA3 consisting all 41 sensors reading at different time frame. Direct damage attacks try to interrupt process flow and sabotage euipments. We have total two dataset files i.e. DA1 and DA2 in this dataset. This dataset was prepared by running chemical plant simulation for continuous 40hr without attack and last 8hr under attack. The rate of measurement is 100 measurement in each 1hr. Hence, we have total approx 4800 measurements for each 41 sensors, In which first 4000 measurements are without attack and last 800 measurements are under attack.

5 METHODOLOGY

We have implemented PASAD [4] and M-PASAD [5] which conserves time series aspects of dataset. Also, we have consider some clustering model such as GMM, K-Means, OPTICS, etc. by removing time series aspects and compared both types of implementation to figure out best one (Shown in Fig:2).

5.1 Time-series based Implementation

5.1.1 PASAD Implementation

PASAD [4] stands for **Process-Aware Stealthy Attack Detection**. The authors have trained the model on normal behaviour and check whether new observations are departing the normal behaviour or not. PASAD has been implemented as univariate and we need to implement it for

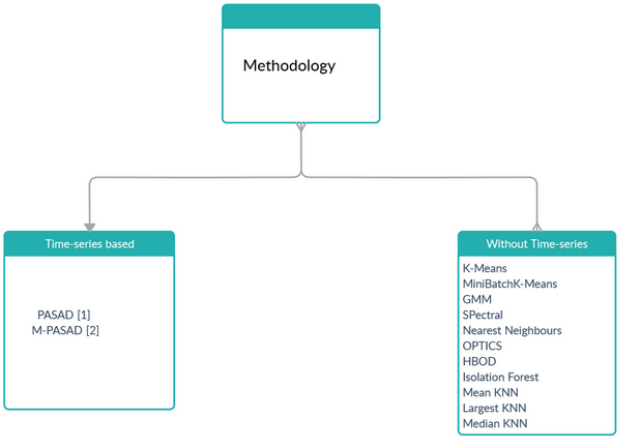

Figure 2: Method Classifications

each sensors in any ICS system. PASAD introduce the concept of Isometic trick i.e.

$$||U^T X|| = ||UU^T X|| \tag{1}$$

Distance calculated by euclidean distance between the most recent test vector and the centroid of the cluster. and centroid will be calculated by below formulae:

$$\hat{c} = Pc \tag{2}$$

Where,
$\hat{c}$ is centroid
P is projection matrix
c is mean of sample

The steps followed in PASAD implementation is shown in Fig 3. Where θ is the threshold value for departure score. If departure score of test sample exceeded the θ then attack alarm will be generated. Result of PASAD for SA and DA attach is shown in Fig 4, 5 respectively. Accuracy score for both type of attacks are 94.30% and 93.94% respectively.

5.1.2 M-PASAD Implementation

M-PASAD [5] is a **multivariate** extension of PASAD. It also reduces computation cost in terms of time complexity and space complexity compared to PASAD. During implementation we just need to deploy M-PASAD at SACADA server and we are up to go after that. No need to implement for each sensor as described in PASAD. Though M-PASAD also add some overhead (Stacked hankel matrix calculation and L2 norm calculation before testing) but that is affordable overhead. Steps followed by M-PASAD is described below and illustrated in Fig 6.

- Perform Normalization using below formula

$$\frac{x_i^j - \mu^j}{\sigma^j} \tag{3}$$

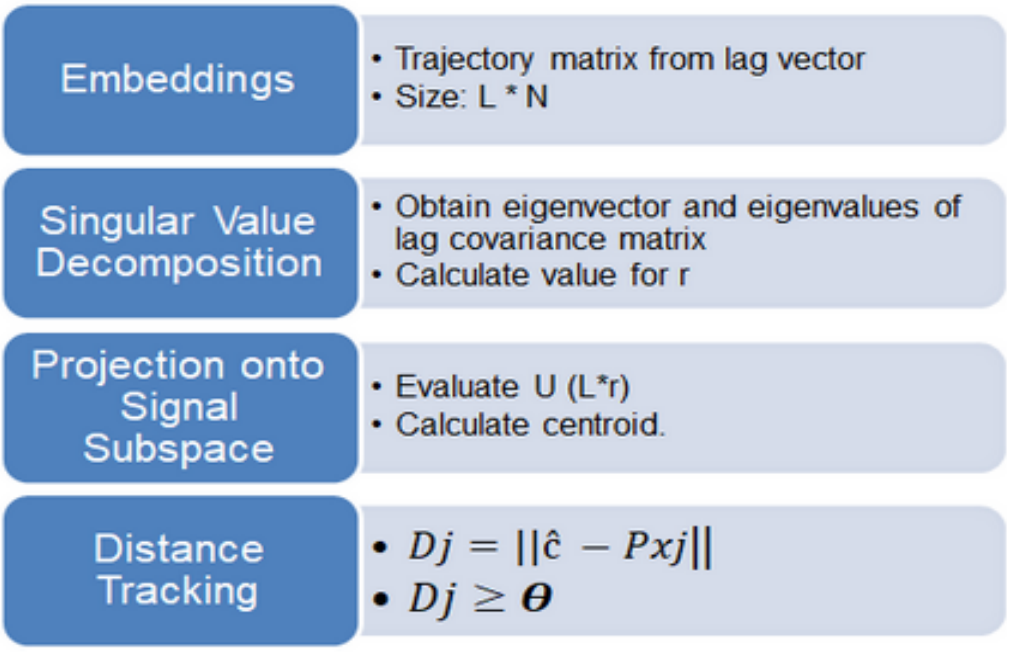

Figure 3: Steps followed for PASAD implementation

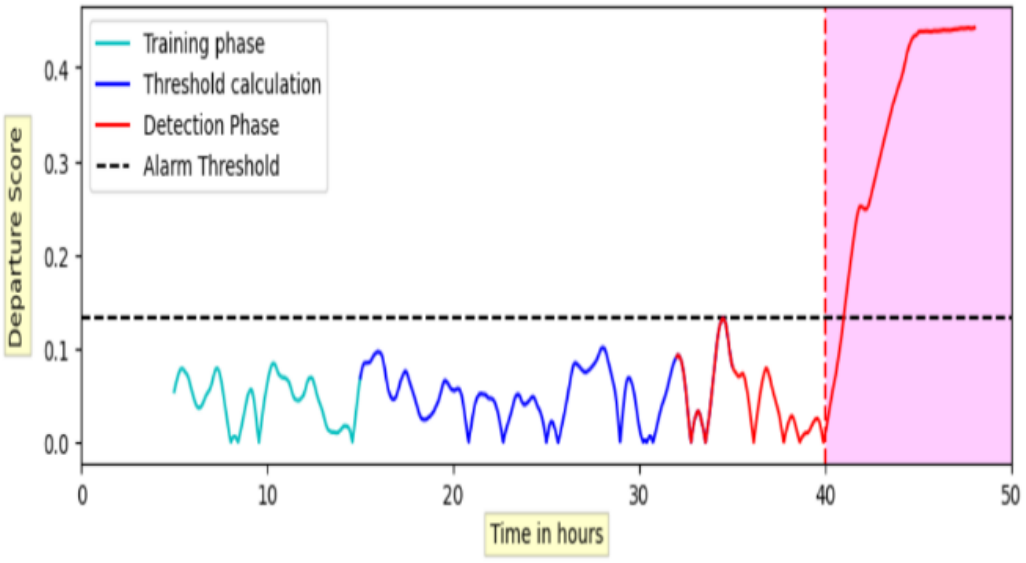

Figure 4: PASAD can detect Stealth attacks

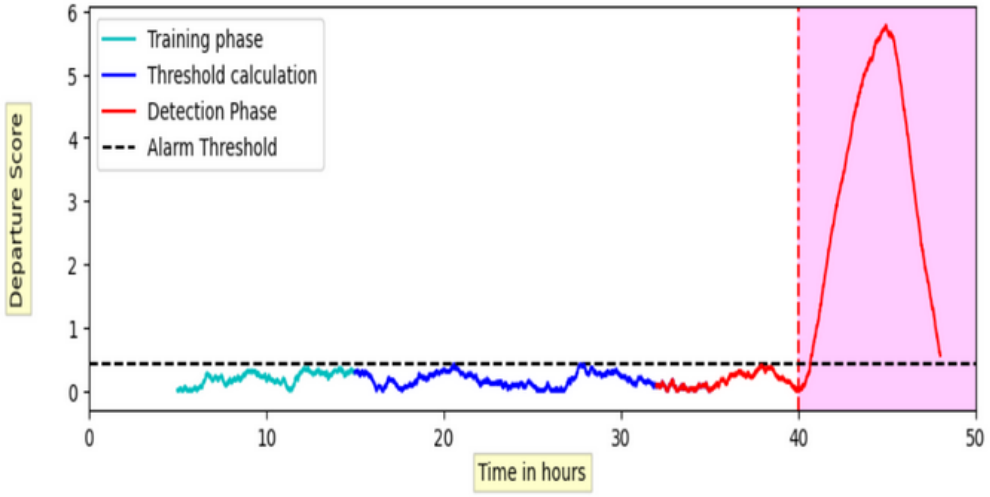

Figure 5: PASAD can detect Direct damage attacks

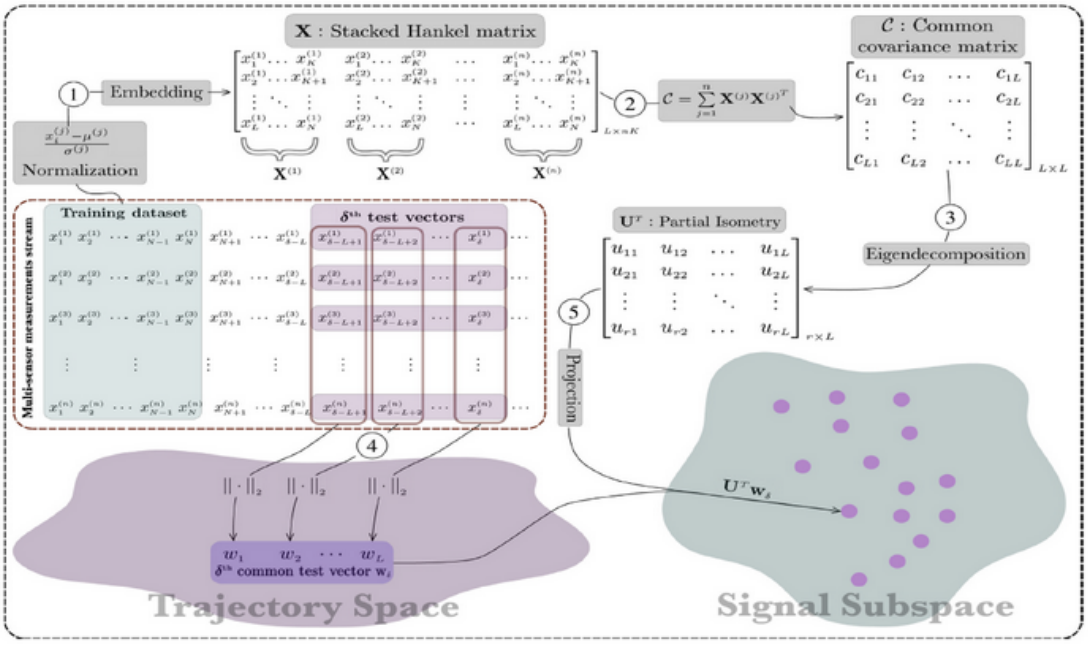

Figure 6: Workflow of M-PASAD [5].

- Embedding and generate Stacked Hankel matrix of size L*nk, Where n is the no of sensor in dataset.

$$X = [X^{(1)} : X^{(2)} : X^{(3)} : ..... : X^{(n)}] \tag{4}$$

- Calculate common covariance matrix of size L*L by using below fromula

$$C = XX^T = \sum_{j=1}^{n} X^{(j)} {X^{(j)}}^T \tag{5}$$

- Perform eigen decomposition and generate U with the help of eigenvectors having r leading eigenvalue. Hence, U^T has been reduced to r*L from L*L and it saves the computational completexity with the help of **isometric tricks**.
- Calculate centroid using below formulae:

$$\hat{c} = Pc \tag{6}$$

  Where,
  $\hat{c}$ is centroid
  P is projection matrix
  c is mean of sample

- Calculate L2 norm of each test sample by using below formula

$$\omega_i = \sqrt{(x_i^{(1)})^2 + (x_i^{(2)})^2 + (x_i^{(3)})^2 + ... + (x_i^{(n)})^2} \tag{7}$$

- Project trajectory space to signal subspace by using

$$U^T w_\delta \tag{8}$$

- Check projected space have departure score more than threshold or not by using below formula

$$D_\delta = ||U^T w_\delta - c||^2 \tag{9}$$

M-PASAD has resulted with best accuracy compared to PASAD and also time take for testing is 1.9 times quicker than PASAD. The performance of M-PASAD is illustrated in FIg 7 and 8

5.2 Without Time-series Implementation

In this implementation we have removed time series aspects from dataset and implemented dimensionality reduction and implemented various Ml model on same such as GMM (Gaussian Mixture Model), K-Means, Spectral, OPTICS, HBOD, Isolation Forest, and variants of KNN (k-Nearest Neighbour).

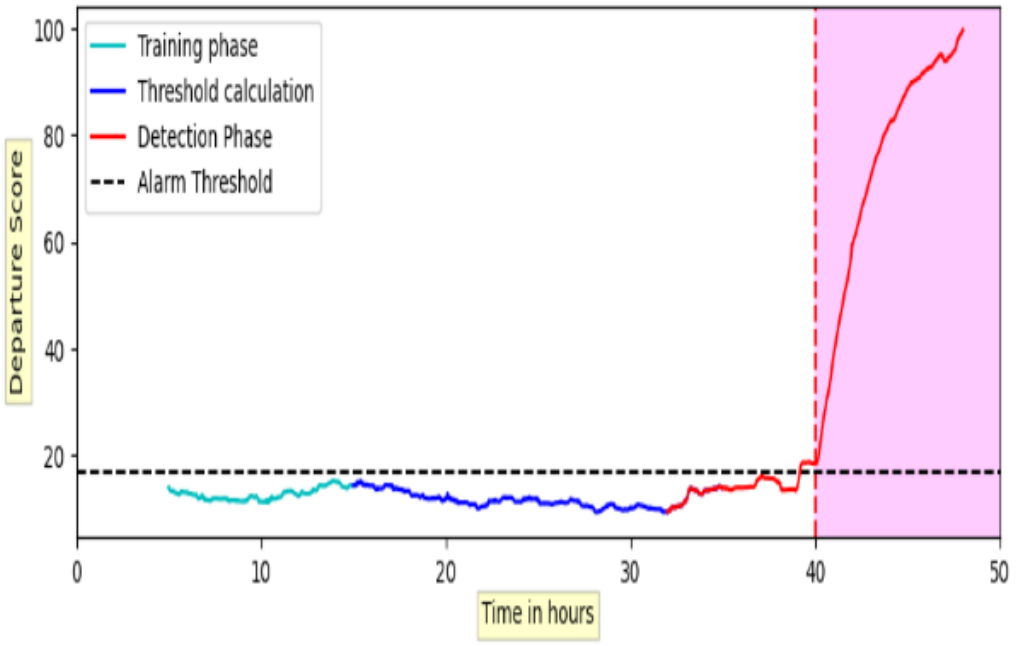

Figure 7: M-PASAD can detect Stealth attacks

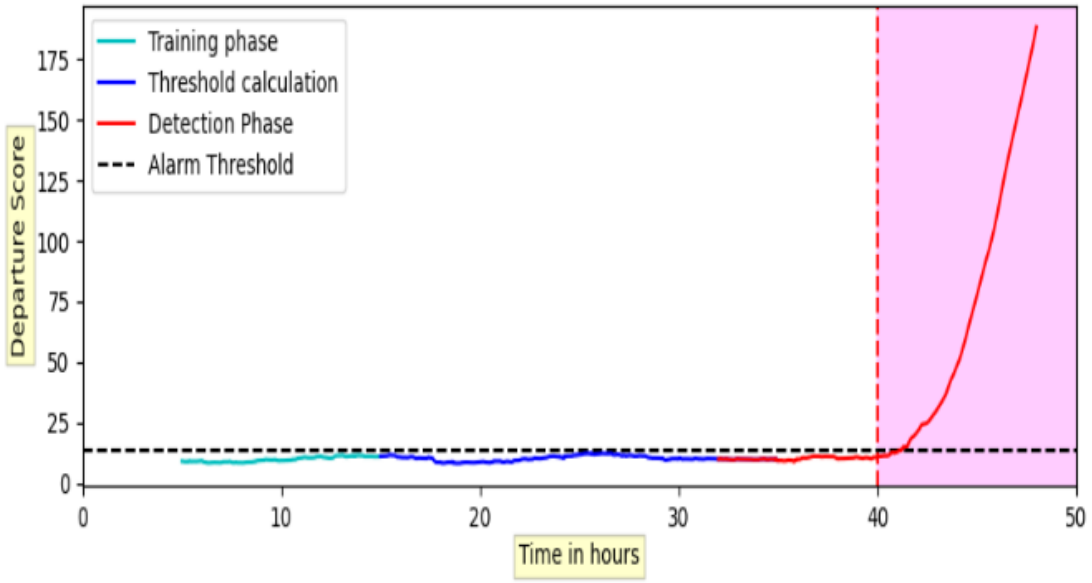

Figure 8: M-PASAD can detect Direct damage attacks

5.2.1 Implemented dimensionality reduction techniques

Dimensionality reduction alludes to strategies for reducing the no of features in training data.When managing high dimensional information, it is frequently valuable to diminish the dimensionality by projecting the information to a lower dimensional subspace which capture the "essence" of the information. Lesser input dimensions mean correspondingly less parameters or a less complex structure in the ML model, alluded to as degree of freedom. A model with an excessive number of freedom is probably going to overfit the training dataset and subsequently may not perform well on new data. There are several dimentionality reduction techniques are available but we have use two main techniques like t-SNE (t-distributed Stochastic Neighbor Embedding) and PCA (Principal COmponent Analysis).

- **t-SNE** [8]: It is a statistical methodology for picturing high-dimensional information by giving each data point a location in a few dimensional map. An altered dataset with decreased dimentionality on Latent space is created by utilizing t-SNE to project on latent space to show correlation. The Gaussian mixture model (GMM) endeavors to find a mixture of multi-dimensional Gaussian probability distributions that best model any input dataset. GMMs can also be utilized for finding clusters in the same manner as kmeans
- **PCA** [8]: Principal component analysis is a fast and flexible unsupervised method for dimensionality reduction in data, Using PCA for dimensionality reduction involves zeroing out one or more of the smallest principal components, resulting in a lower-dimensional projection of the data that preserves the maximal data variance.

  $$minimize_{W,C}\{\Sigma_{i=1}^{N}\Sigma_{j=1}^{D} log(1 + exp(-x_{ij}.\Sigma_{k=1}^{K} W_i k C_j k))\}$$

  where, W = loading matrix, C = top K components, N=no. of samples and D = no. of features. In this feature set, a value of $K = 41$ was chosen. This decision was based over the visualization provided by the scree plot. 41 features accounted for $\sim$97% of the total variance found in the dataset. The dataset was created using the R library logistic PCA where the parameters of $K = 41$ and $m = 2$ was considered (the value of m is chosen for which the negative log likelihood is the least after applying the method cv.lpca over multiple values of m).

5.2.2 Models Utilized

Various machine learning classifiers were explored, such as logistic regression with L2 regularization, support vector machines with radial basis function kernel, neural network and ensemble learning techniques such as extreme gradient boost, and stacking. For every drug, the problem is a single binary classification problem, i.e to predict if the MTB isolate is resistant or susceptible to that particular drug. The models were applied using popular libraries: sklearn, tensorflow and keras.

K-Means Clustering [8]: The k-means algorithm searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers.

cluster-center : arithmetic mean of all the points belonging to the cluster. Each point is closer to its own cluster center than to other cluster centers. kmeans involves an intuitive iterative approach

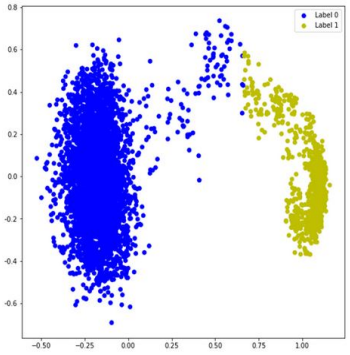

Figure 9: Spectral clustering result

known as *expectation–maximization.
1. Guess some cluster centers
2. Repeat until converged

- *E-Step*: assign points to the nearest cluster center
- *M-Step*: set the cluster centers to the mean

Gaussian Mixture Model [9]: Finite Mixture Models is an algorithm for clustering and its purpose is to determine the inner structure of data when no information other than the observed values is available. complicated clustering algorithm which has a better quantitative measure of the fitness per number of clusters As we saw in the previous section, given simple, well-separated data, kmeans finds suitable clustering results. By eye, we recognize that these transformed clusters are non-circular, and thus circular clusters would be a poor fit. You might imagine addressing these weaknesses by generalizing the kmeans model A Gaussian mixture model (GMM) attempts to find a mixture of multi-dimensional Gaussian probability distributions that best model any input dataset. In the simplest case, GMMs can be used for finding clusters in the same manner as kmeans But because GMM contains a probabilistic model under the hood, it is also possible to find probabilistic cluster assignments—in Scikit-Learn this is done using the 'predict-proba' method. The result of this is that each cluster is associated not with a hard-edged sphere, but with a smooth Gaussian model. Another means of correcting for over-fitting is to adjust the model likelihoods using some analytic criterion such as the [Akaike information criterion (AIC)]or the [Bayesian information criterion (BIC)] Notice the important point: this choice of number of components measures how well GMM works as a density estimator, not how well it works as a clustering algorithm.

Spectral Clustering: Spectral clustering algorithms embed the data points by projection onto a few eigenvectors of (some form of) the graph Laplacian matrix and use this spectral embedding to find a clustering. This technique has been shown to work on various arbitrarily shaped clusters and, in addition to being straightforward to implement, often outperforms traditional clustering algorithms such as the k-means algorithm. Output is shown below:

OPTICS Clustering: Ordering Points To Identify the Clustering Structure, in short, OPTICS, is closely related to DBSCAN, finds core sample of high density and expands clusters from them. Unlike DBSCAN, it keeps cluster hierarchy for a variable neighborhood radius. It is better suited for usage on large datasets. Clusters are then extracted using a DBSCAN-like method (cluster_method

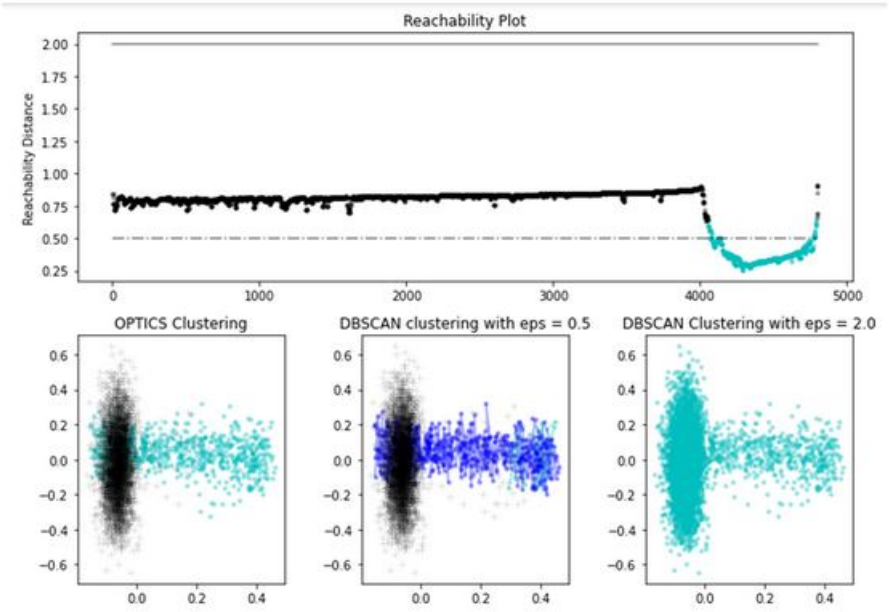

Figure 10: OPTIC clustering result.

= 'dbscan') or an automatic technique proposed in (cluster_method = 'xi').

Histogram-base Outlier Detection It is a fast unsupervised method that assumes feature independence and calculates outlier scores by constructing histograms. It is significantly faster than multivariate methods. The algorithm presupposes that the characteristics are totally independent of one another (zero multicollinearity).

Isolation Forest It employs the scikit-learn library and performs data partitioning via a set of trees. Isolation Forest calculates an anomaly score based on how isolated a certain location is within the structure. The anomaly score is then used to distinguish outliers from the rest of the data. Isolation Forest performs well with multi-dimensional data.

k-Nearest Neighbors The outlying score for any data point could be defined as the distance to its k^{th} nearest neighbour and kNN algorithm gives the output using this outlying score. There are three variants of kNN used in this work:

- Largest kNN : The outlier score is calculated using the distance of the k^{th} neighbour.
- Mean kNN : The outlier score is calculated using the average of all k neighbours.
- Median kNN : The outlier score is calculated using the median of the distance to k neighbours.

5.3 Performance Comparison

The Accuracy achieved by each implemented model is shown in Fig 11 and 12 for both type of attack files. We can see that M-PASAD has highest accuracy achieved , hence we M-PASAD is the

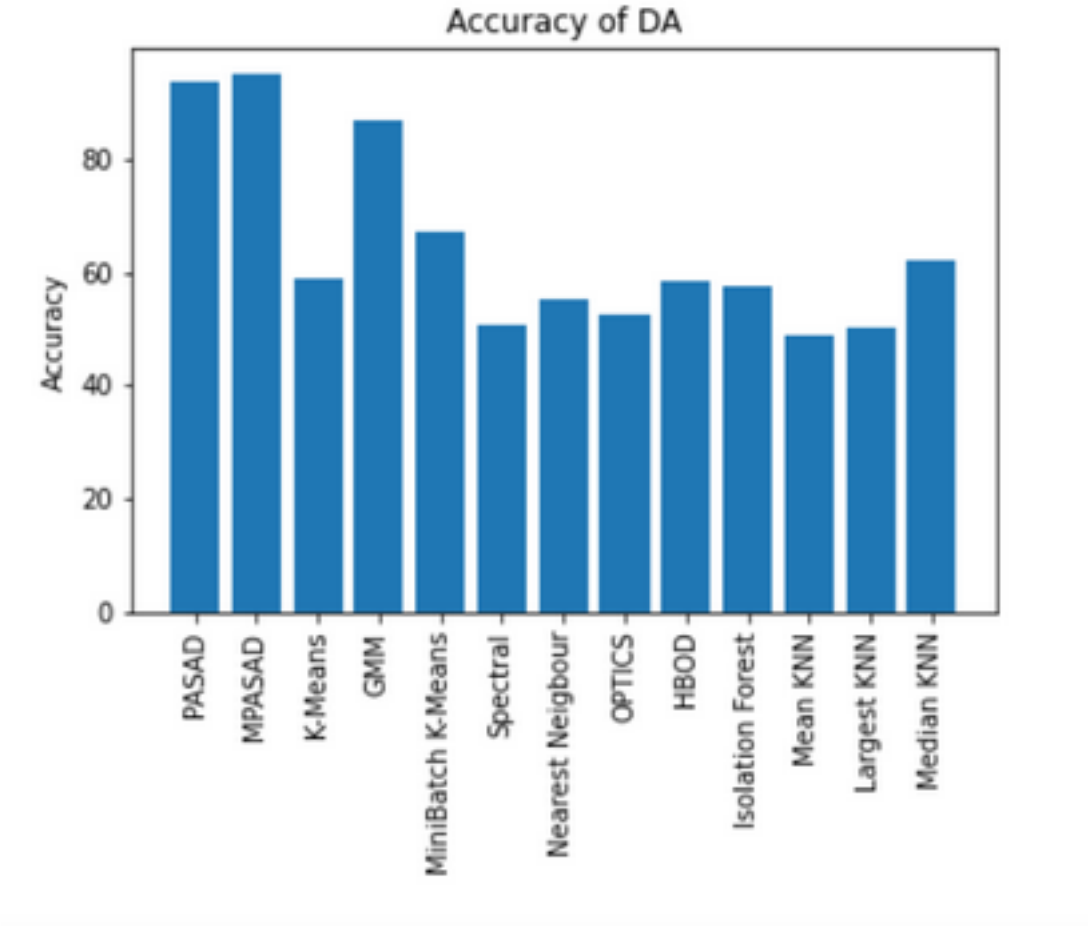

Figure 11: Accuracy comparison for DA attacks

best among all implemented models. Another performance metrics such as precision, recall and f1-score is also illustrated for best model i.e. M-PASAD model in Table 1.

6 Conclusions

- M-PASAD (time-series based) method found to be best among all implemented methodology.
- Capable of detecting sophisticated attacks by monitoring time series of sensor measurements for structural changes in their behavior.
- Less computational cost such as less space and time complexity
- Unlike PASAD, M-PASAD is a multivariate.
- Time-series aspects of data cannot be ignored.
- M-PASAD can detect both Stealthy and Direct damage attack and having best accuracy

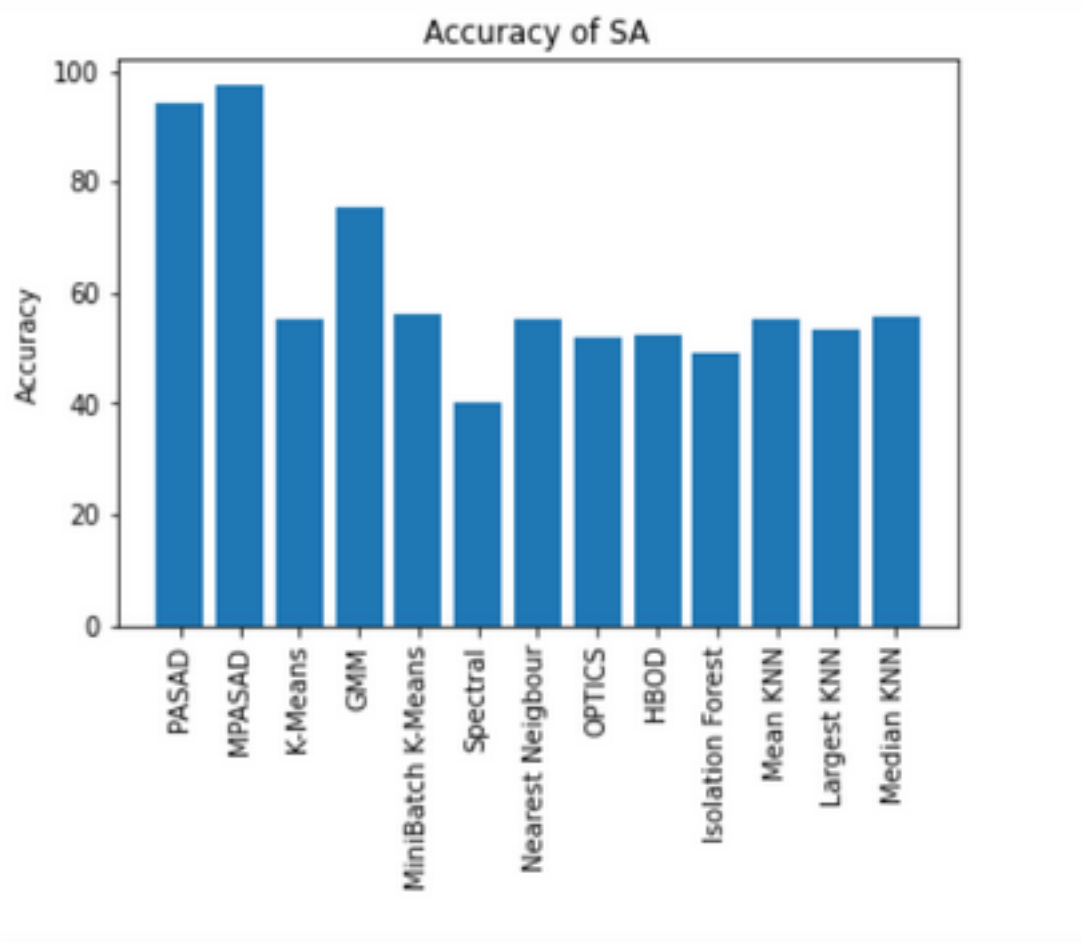

Figure 12: Accuracy comparison for SA attacks

	Precision	**Recall**	**F1-score**	**Accuracy**
SA1	0.89	1.0	0.94	94.88
SA2	1.0	0.73	0.84	82.27
SA3	1.0	0.94	0.97	97.31
DA1	1.0	0.65	0.79	73.59
DA2	1.0	0.86	0.92	92.34

Table 1: M-PASAD Performance Metric

tends to 98% on our chosen dataset.

- Time complexity of PASAD is approx 1.9 times time taken by M-PASAD

7 MEMBER CONTRIBUTIONS

- **Harshvardhan Pratap Singh**:
  Roll no.: 20111410
  Programme: MS(R) CSE

  - Developed and designed distance based clustering algorithm to find anomaly.
  - Experimented with KMeans and MiniBatchKMeans for the clustering purpose
  - Experimented with Non time-series aspects to test model efficiency using clustering.
  - Implemented Gaussian Mixture Model on TE dataset.

- **Nanda Rani**:
  Roll no.: 21111265
  Programme: PhD CSE

  - Experimented with time-serise aspects to test model efficiency.
  - Implemented Process-Aware Stealthy Attack Detection (PASAD) to detect Anomalies.
  - Experimented with DBScan and Agglometry clusetring algorithms to uncover non-time series aspects from dataset
  - Implemented multivariate extension of Process-Aware Stealthy Attack Detection (M-PASAD) to detect anomalies in the system
  - Compared and performed details analysis of performance of each implemented algorithms.

- **Trishie Sharma**:
  Roll no.: 21111270
  Programme: PhD CSE

  - Applied Isolation Forest and k Nearest Neighbor algorithms on all the dataset and identified that Isolation Forest algorithm gives better accuracy out of the two.
  - On all the datasets of stealth and direct attacks, histogram-based outlier detection and multiple variants of the kNN algorithm are applied and their accuracy scores are measured.

- **Harsika Diksha**:
  Roll no.: 20111021
  Programme: MTech CSE

  - Implemented spectral and OPTICS clustering algorithm for finding anomalies in the dataset, used PCA for dimensionality reduction
  - After finding the accuracy, we got to know that spectral performs better in anomaly detection and OPTICS forms very nice visualization on detection of anomalous data.

- **Nikhil Molugu**:
  Roll no.: 20111415
  Programme: MS(R) CSE

  - Applied spectral clustering for finding the outliers among the data set.
  - Applied PCA to reduce the dimensions.

References

[1] Thomas M. Chen and Saeed Abu-Nimeh. Lessons from stuxnet. *Computer*, 44(4):91–93, apr 2011.

[2] Marina Krotofil Dan Scali Nathan Brubaker Blake Johnson, Dan Caban and Christopher Glyer. Attackers deploy new ics attack framework "triton" and cause operational disruption to critical infrastructure, 2017.

[3] Marshall Abrams and Joe Weiss. Malicious control system cyber security attack case study—maroochy water services, australia. 2008.

[4] Wissam Aoudi, Mikel Iturbe, and Magnus Almgren. Truth will out: Departure-based process-level detection of stealthy attacks on control systems. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, page 817–831, New York, NY, USA, 2018. Association for Computing Machinery.

[5] Wissam Aoudi and Magnus Almgren. A scalable specification-agnostic multi-sensor anomaly detection system for iiot environments. *Int. J. Crit. Infrastructure Prot.*, 30:100377, 2020.

[6] Patrick Berry. Processing scada alarm data offline with elk, 2021.

[7] James Downs and Ernest Vogel. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 1993.

[8] *KMeans Clustering : Python Data Science Handbook*.

[9] In depth: Gaussian mixture models. *Journal of Artificial Intelligence Research*.