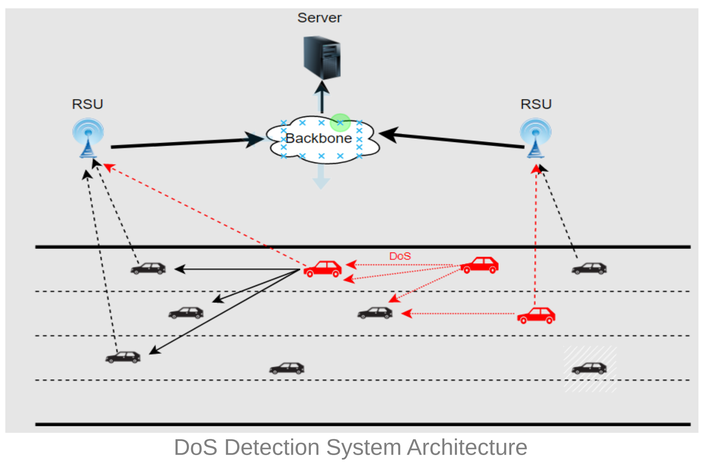
Unsupervised Learning Algorithms for Denial of Service Detection in Connected Vehicles

Connected Cars Context

- 237 millions connected cars in the world in 2021 and over 400 millions by 2025.
- Global ITS **(Intelligent Transportation System)** market from $22.91 billion in 2021 to $42.80 billion in 2028
- Cybersecurity of the connected cars is highly critical.
- **Denial of Service** (DoS) attacks jam and flood the network by mass of malicious messages.

- Proposed DoS detection system architecture:
  - CAM DoS attack generated by one or more malicious vehicles.
  - CAM messages contain information (position, velocity, etc.).
  - CAM messages are sent to a central or distributed servers for analysis.
  - **Clustering-based** detection algorithms are applied to separate benign and malicious data.



DoS Detection System Architecture

Parties prenantes

Institut Mines-Télécom

Auteurs

Ali EL ATTAR
Rida Khatoun
Fadlallah Chbib
Ahmad Fadllalah

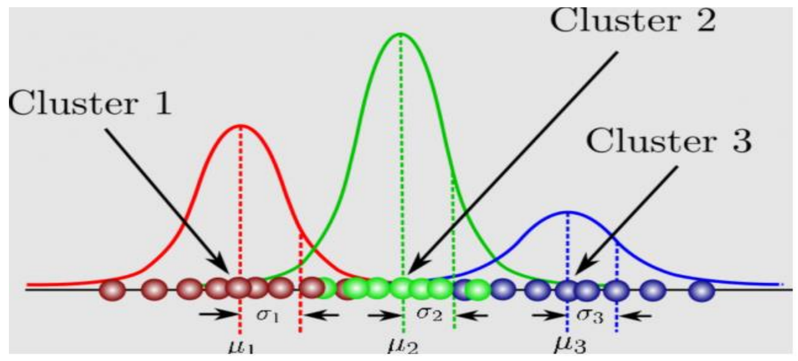
Partenaires

InDiD Infrastructure Digitale de Demain

DoS Clustering-based Model Detection

Several Clustering algorithms are applied:

- Centroid-based clustering (e.g. **K-Means, CLARA**)
- Density-Based clustering (e.g**. DBSCAN**)

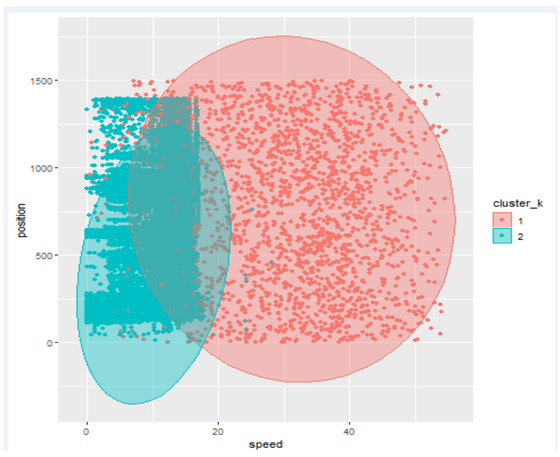


Example of clustering with a Gaussian Mixture Model of three components

- Probabilistic Model-based Clustering (e.g. **GMM**). EM (Expectation-Maximization) algorithm to estimate the parameters of GMM, it alternates between two steps:
  - **E step:** compute the log likelihood expectation using the current parameter estimation.
  - **M-step :** estimate of the parameters by maximizing the expected log likelihood found in the (E) step.

Results and Analysis

- Evaluation based on Vehicular Reference Misbehavior **(VeReMi)** dataset, containing the message logs for each vehicle in the simulation:
  - Position, velocity, acceleration and heading vectors.
  - Position noise, velocity noise, acceleration noise.
  - Reception timestamp, transmission time, sender, ID message, etc.



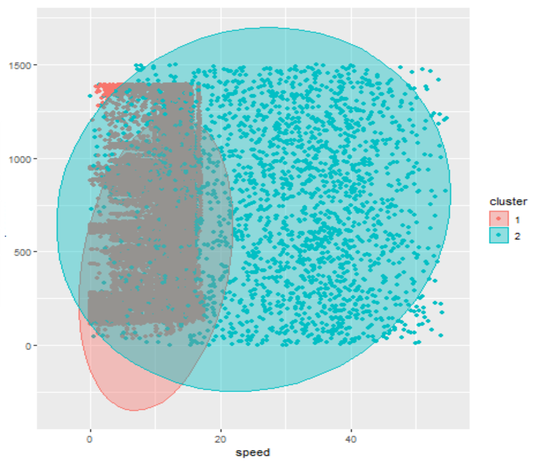
- GMM produces the best results in terms of **F-measure**. It is approximately 95% for the three subsets (**DoS (D1), random DoS (D2) and disruptive DoS (D3)**), with high precision and recall values.

Criterion	K-means	GMM	Clara	DBSCAN
Specificity	0.6179	0.8103	0.5079	0.0647
Sensitivity	0.9856	0.9541	0.5049	0.8993
Precision	0.8823	0.936	0.7489	0.7365
Recall	0.9856	0.9541	0.5049	0.8993
F-Measure	0.9311	0.945	0.6031	0.8098

Detection Results on (B,D1,D2,D3)

Criterion	K-means	GMM	Clara	DBSCAN
Specificity	0.5084	0.5137	0.4965	0.0806
Sensitivity	0.5061	0.9714	0.5037	0.8961
Precision	0.8785	0.9334	0.8754	0.8725
Recall	0.5061	0.9714	0.5037	0.8961
F-Measure	0.6422	0.952	0.6394	0.8841

Detection Results on (B,D1,D3)



Results of GMM clustering (cluster 1:malware data, cluster 2: benign data

Contact : ali.elattar@telecom-paris.fr

October 2022 Colloque de l'Institut Mines Télécom Paris