CANLock

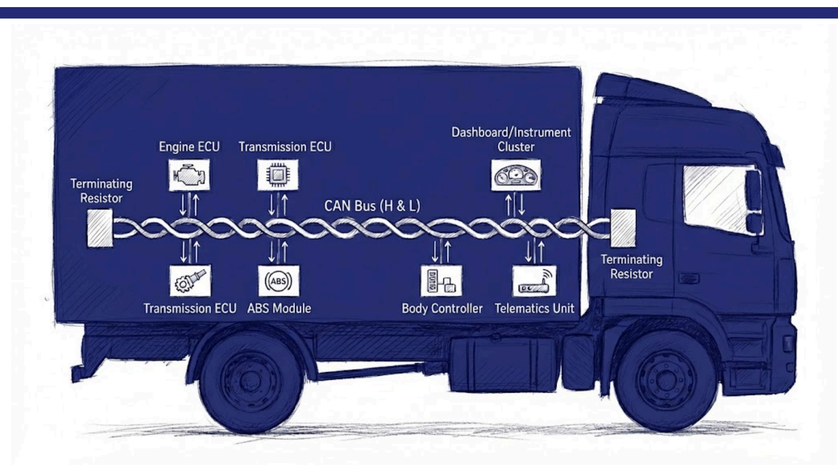
Équipe 1:
Marin Kerboriou, Michal Naumiak, Nadine Slimani, Ouissal Ezzarouali, Ismaël Debiche, Yazid El Yaakoubi, Youssed Ezzat

Équipe 2:
Yanny Edvard Lafleur, Aya Bensekhria, Elina Govi, Fassilatou Dipama, Teddy Kana

Team Lead: Loïc Baret

Introduction

- Assurer la sécurité globale du véhicule, et par extension, celle de ses occupants.
- Les véhicules modernes intègrent des unités électroniques embarquées, échangeant en permanence des données critiques pour le fonctionnement du véhicule:
  - Capteurs
  - Actuateurs
- Ces échanges sont assurés par le bus CAN, un protocole de communication standardisé dans l'industrie automobile.



Problématique

- Un attaquant accédant au réseau CAN du véhicule peut injecter, modifier ou supprimer des messages, donc altérer le comportement du véhicule: freinage, direction, affichages, etc...
- Les modèles les plus récents peuvent atteindre des taux d'erreur inférieurs à 1%
- Plusieurs centaines de milliers de messages sont échangés par minute.
- Un taux d'erreur de 1% = des milliers de faux positifs par minute.
- Essentiel de concevoir des approches de détection plus fiables et contextuellement pertinentes, capables de distinguer des anomalies réellement critiques tout en réduisant drastiquement le taux de faux positifs.

Avancement

- **Semaines 1 à 3**: Prise en main de l'environnement de développement et familiarisation avec le sujet.
- **Semaine 4**: Présentation de *Simon Bellemare* sur le protocole J1939 et le bus CAN.
- **Semaines 5 à 7**: Installation de la base de données et découverte des données.
- **Semaine 8**: Semaine de lecture.
- **Semaines 9 à 11**: Développement d'un pipeline de décodage des données.

- **À venir**: Résolution du problème en construisant des modèles d'apprentissage automatisé.

Données

- Données de bus CAN collecté depuis un poids lourd de la marque Renault.
- Environ 490 millions d'enregistrements.
- Respectent le protocole bus CAN J1939.
- Nous avons accès à 3 blocs d'un message CAN (= 128 bits):
  - Identifier (29 bits): Un identifiant unique ainsi que sa priorité.
  - DLC (Data Length Code - 4 bits): La longueur du data field en octets.
  - Data field (0 à 64 bits): Les données à transmettre.

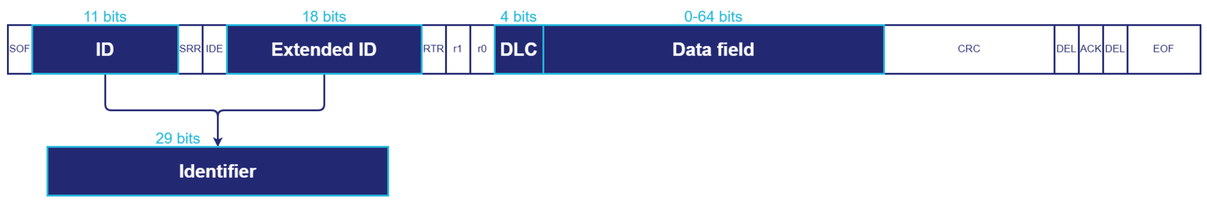


Figure 1: Schema du découpage d'un message CAN

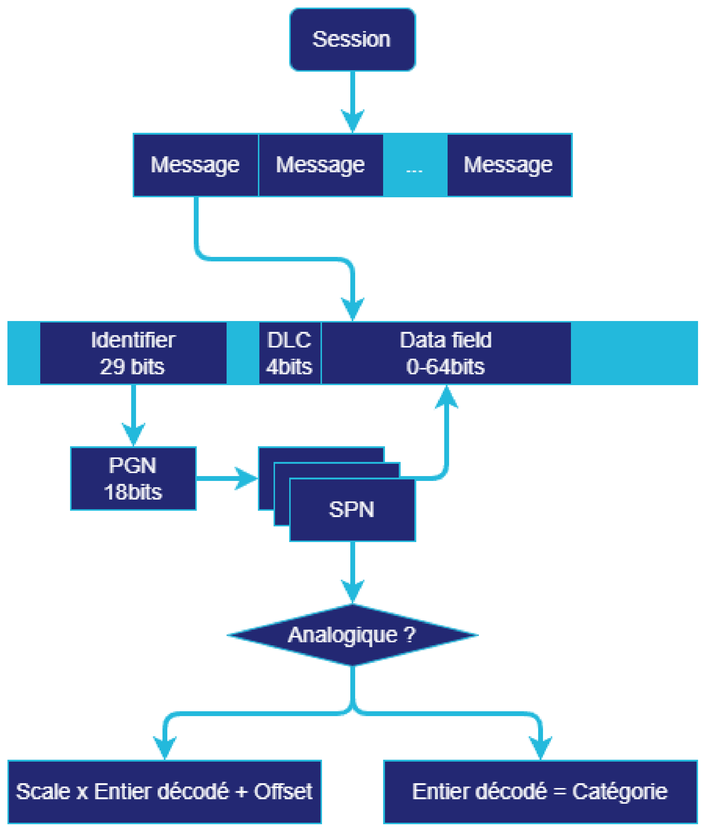


Figure 2: Schema du décodage des données

Modélisation

- Meilleur modèle en ce moment: 2 transformers en parallèles (papier à venir).
- Idée principale: Exploiter la cross-corrélation entre les signaux.
- 2 équipes = 2 approches:
  - VAE (Variational Auto-Encoder): Particulièrement utilisé en détection d'anomalies.
  - RNN (Recurrent Neural Network) + Clustering: Aspect séquentiel et temporel des données.

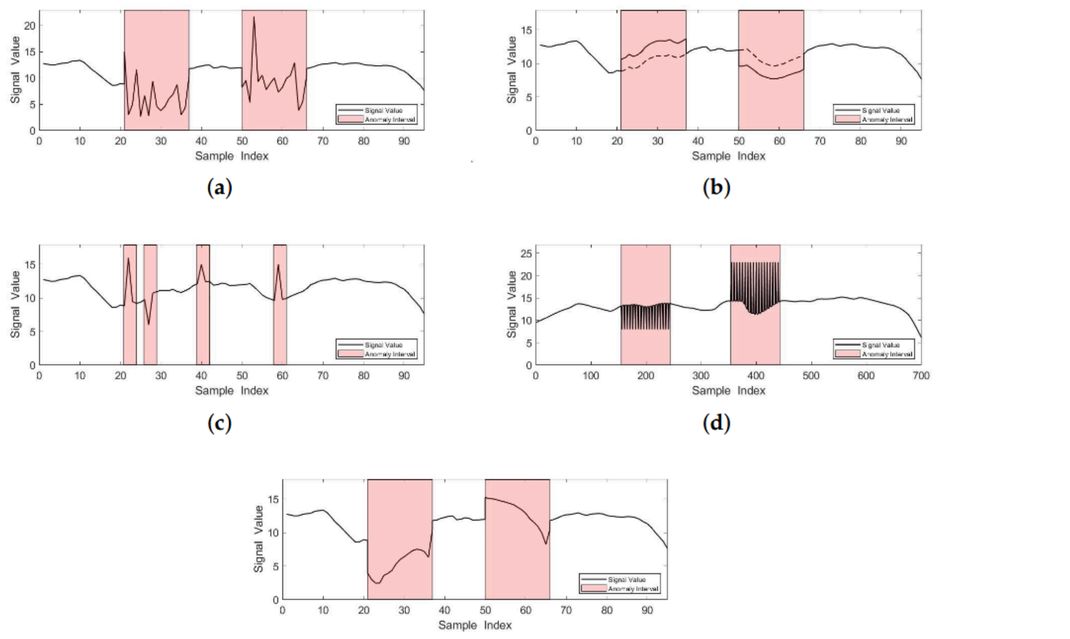


Figure 3: Exemples d'attaques sur les signaux: (a) hijack, (b) biais, (c) injection, (d) DOS et (e) Replay*

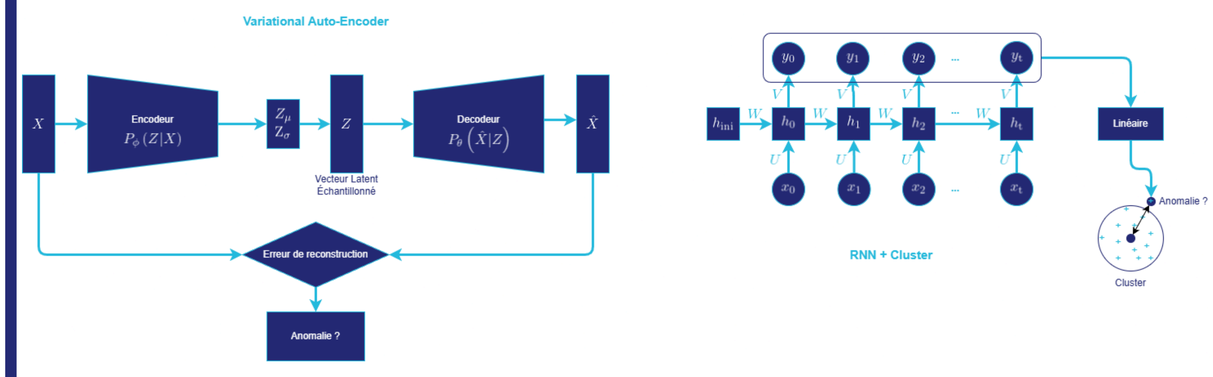


Figure 4: Modélisation envisagée, à gauche le VAE, à droite le RNN + Clustering

*Wang, L., & Zhang, X. (2023). Anomaly Detection for Automated Vehicles Integrating Continuous Wavelet Transform and Convolutional Neural Network. Applied Sciences, 13(9), 5525.