Traitement des données : Application aux C-ITS

Télécommunication
Reseaux, Sécurité et Objets Connectés
ENSEIRB-MATMECA

Secil Ercan
sercan@bordeaux-inp.fr
https://github.com/ercansec/DataProcessing_CITS

Novembre 2023

C-ITS
Cooperative Intelligent Transport Systems

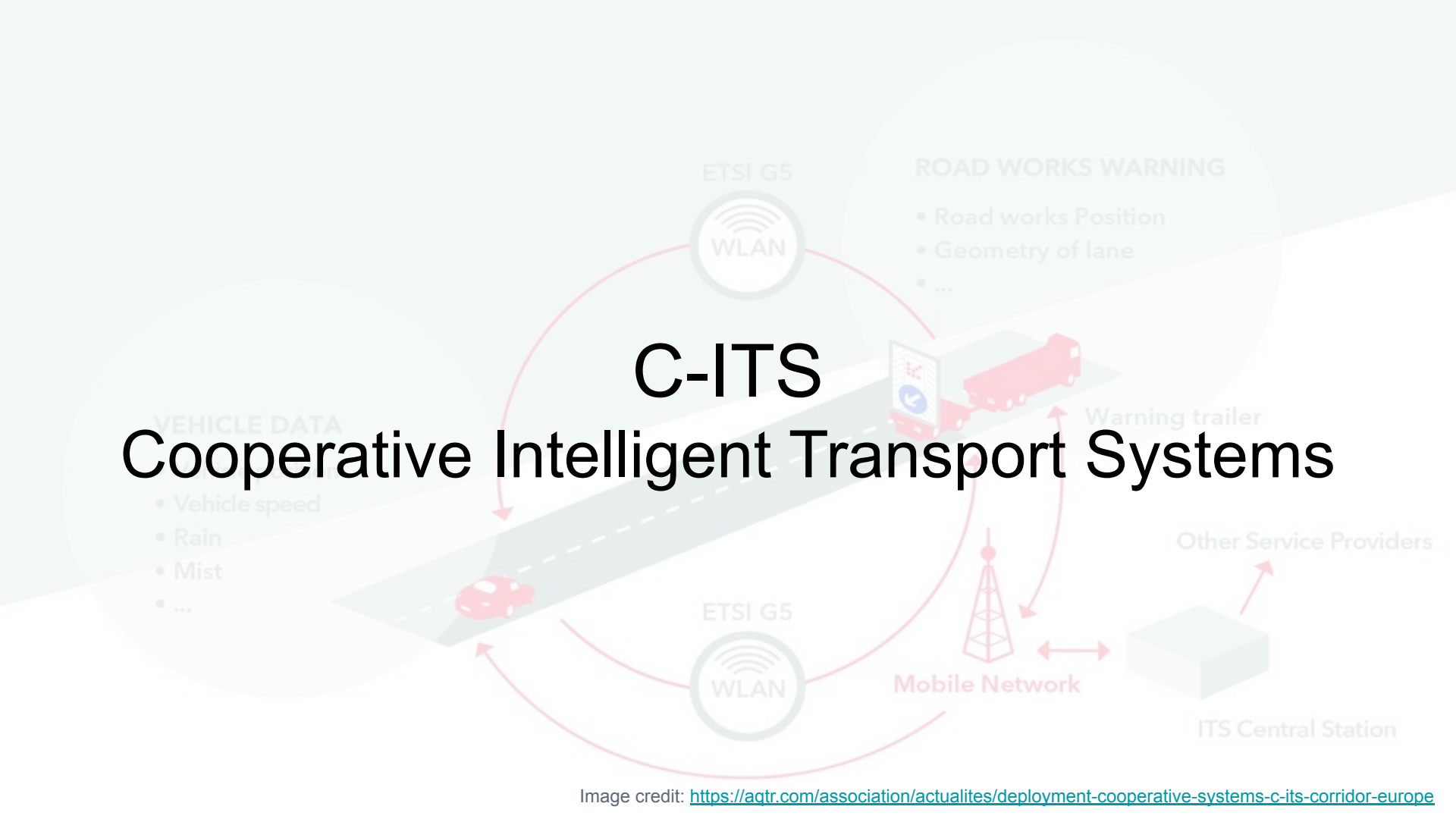


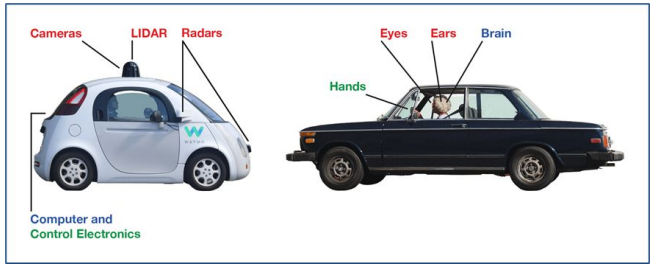
Image credit: https://aqtr.com/association/actualites/deployment-cooperative-systems-c-its-corridor-europe

Main Objectives in ITS

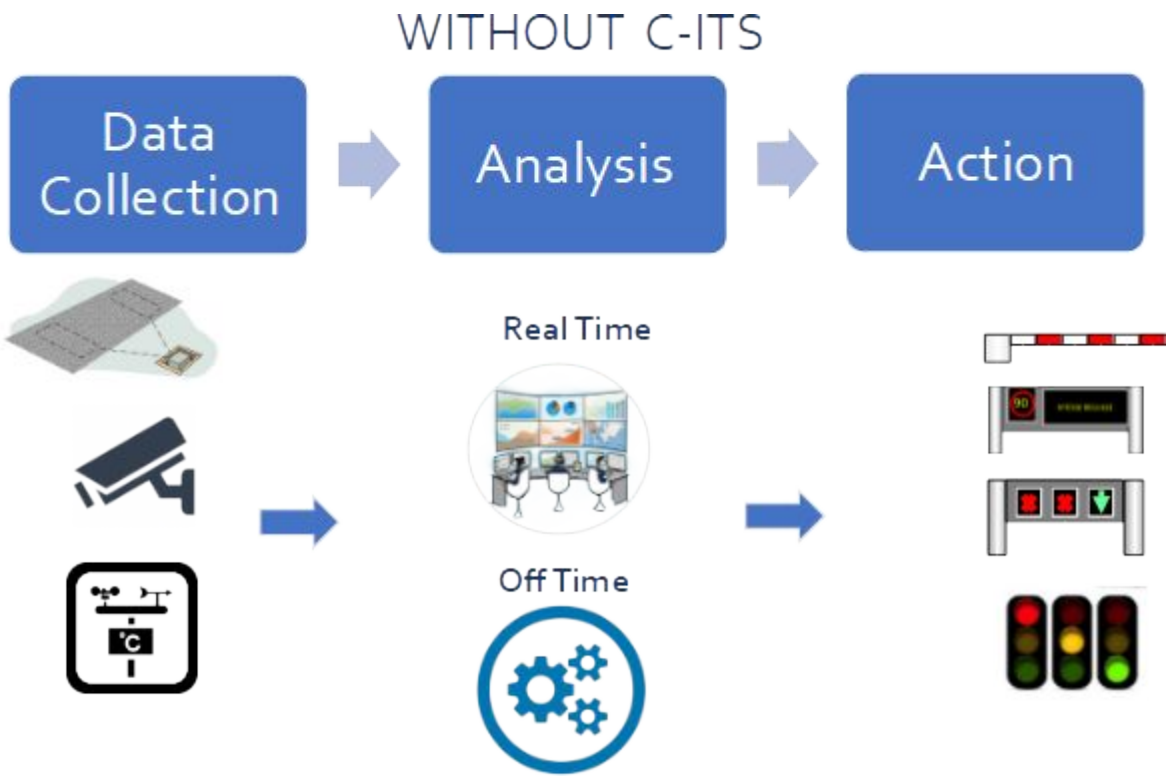
- Road safety
- Traffic efficiency
- Comfort
- Sustainability

Moreover in C-ITS:

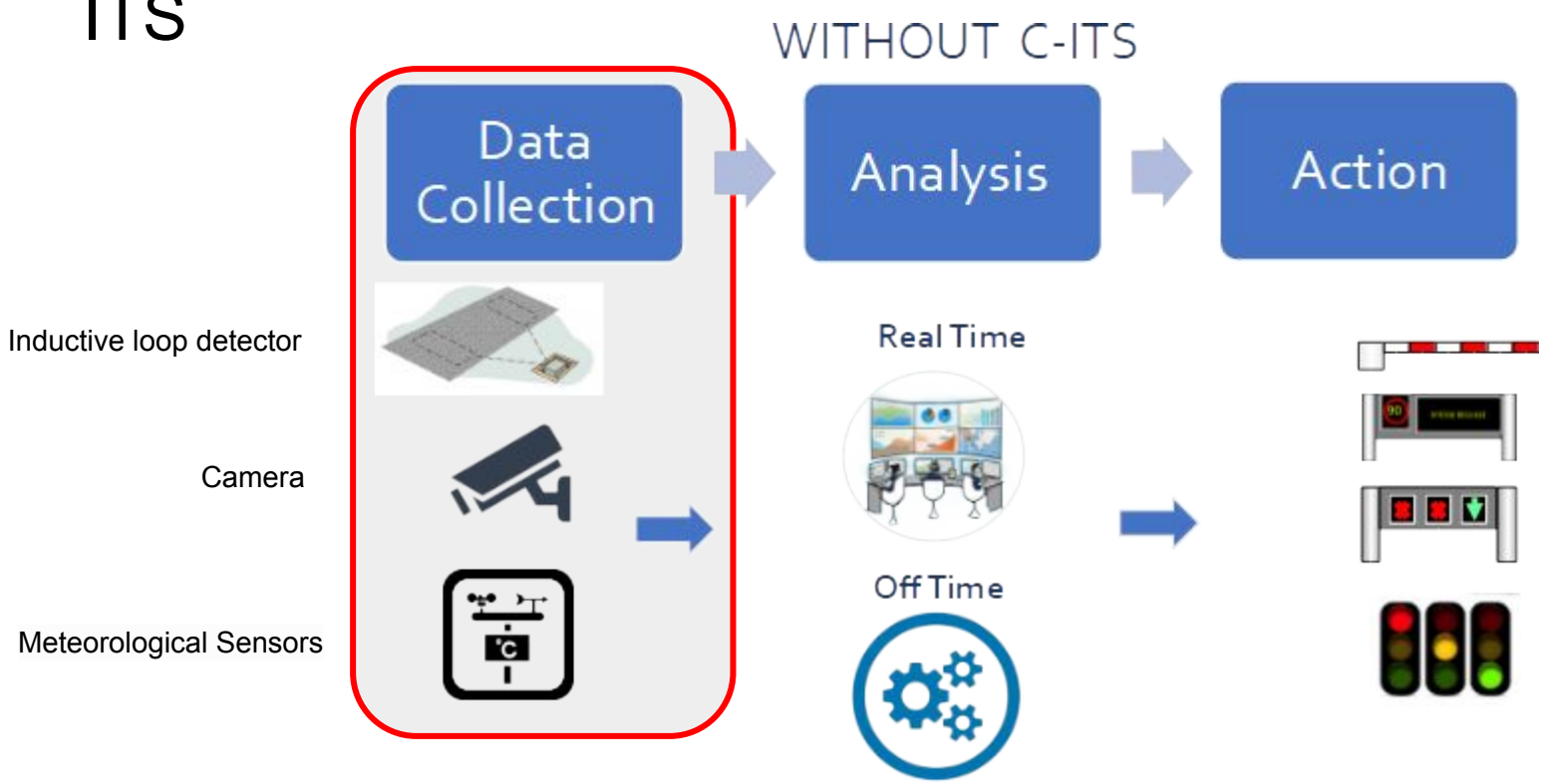
- Being prepared for future challenges in automation



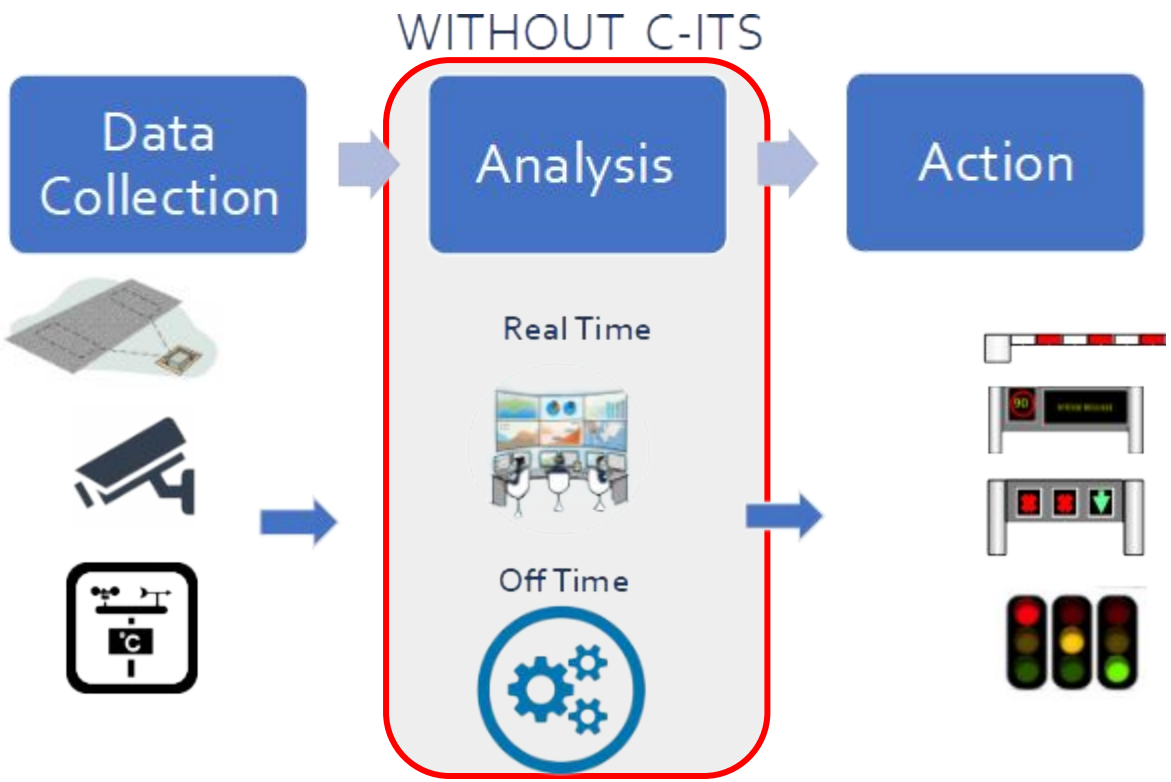
ITS


WITHOUT C-ITS
Data Collection
Analysis
Action
Real Time
Off Time

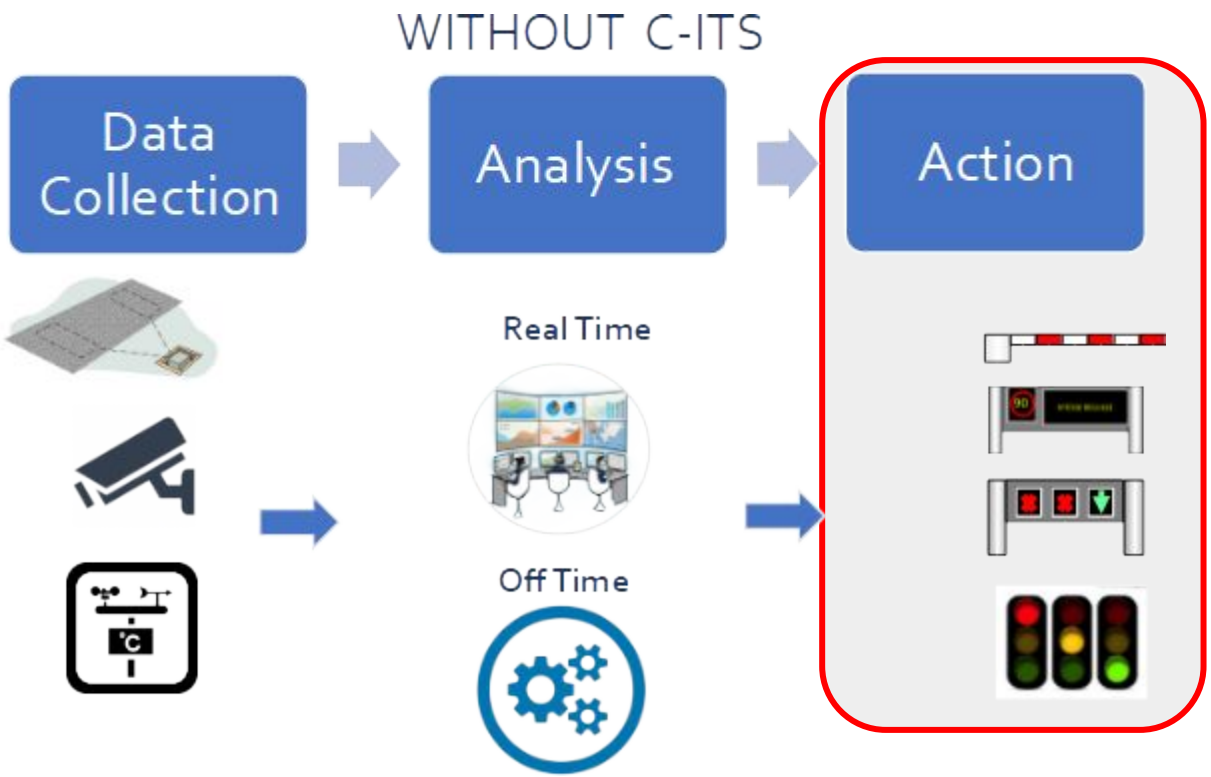
ITS


WITHOUT C-ITS
Data Collection
Analysis
Action
Inductive loop detector
Camera
Meteorological Sensors
Real Time
Off Time

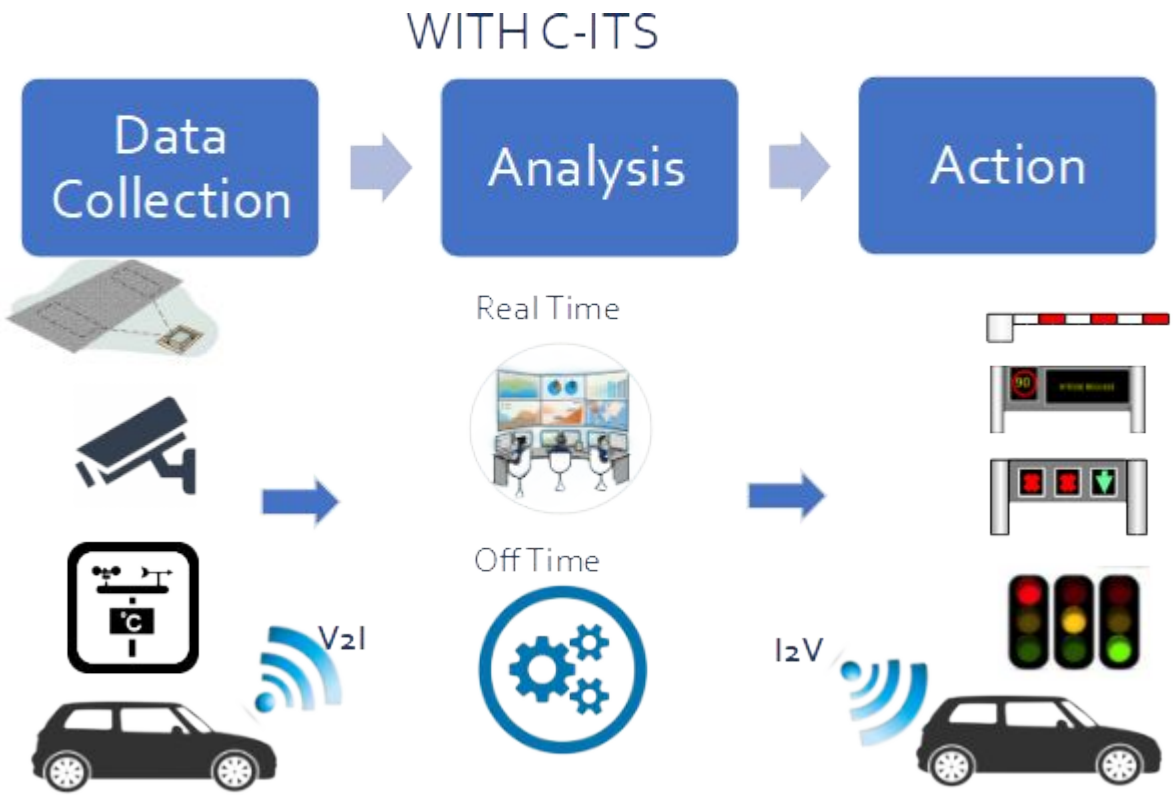
ITS


WITHOUT C-ITS
Data Collection
Analysis
Action
Real Time
Off Time

ITS

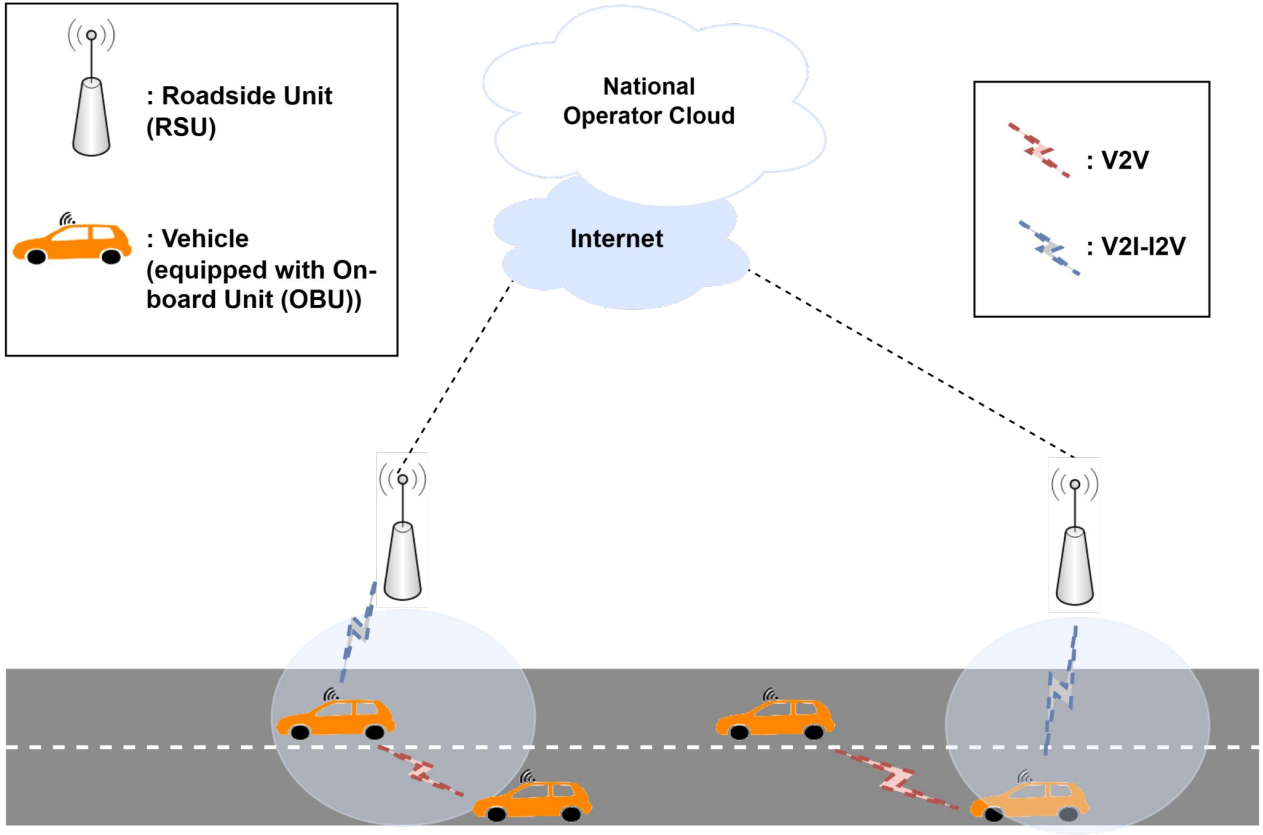

WITHOUT C-ITS
Data Collection
Analysis
Action
Real Time
Off Time

C-ITS



Exchange of data

C-ITS


: Roadside Unit (RSU)
: Vehicle (equipped with On-board Unit (OBU))
National Operator Cloud
Internet
: V2V
: V2I-I2V

C-ITS

Different communication modes:

- V2V (Vehicle-to-Vehicle)
- V2I (Vehicle-to-Infrastructure)
- I2V (Infrastructure-to-Vehicle)
- V2P (Vehicle-to-Pedestrian)
- V2X (Vehicle-to-Everything), etc.

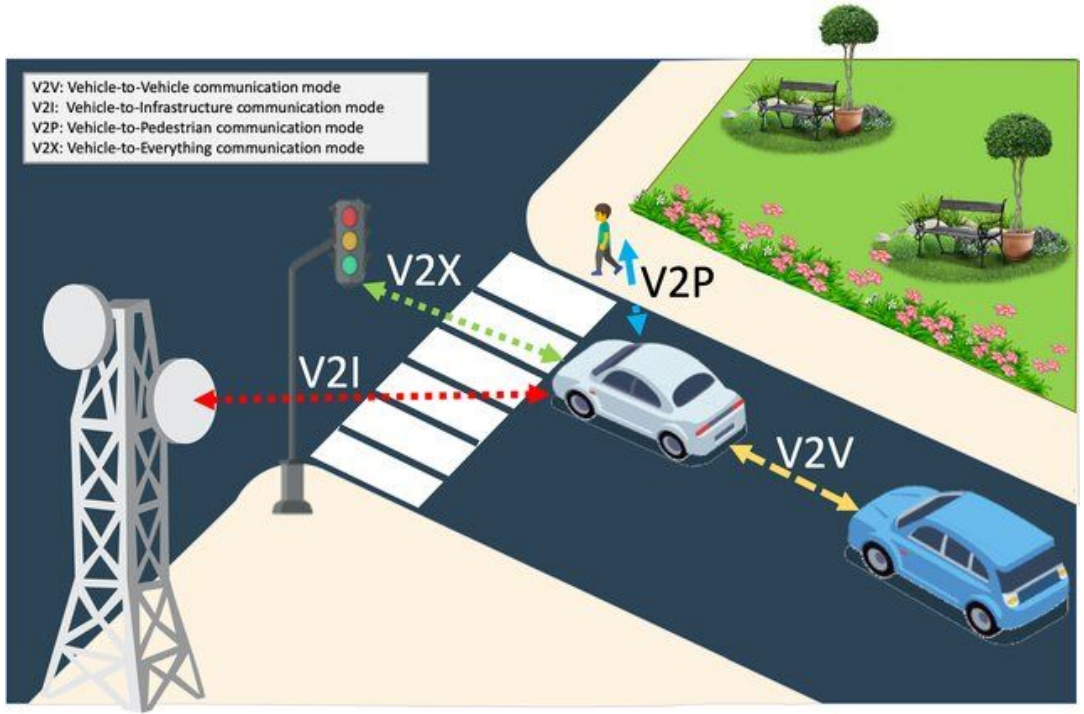


Image credit: Zeadally, S., Guerrero, J. & Contreras, J. “A tutorial survey on vehicle-to-vehicle communications”, *Telecommun Syst* 73, 469–489 (2020). https://doi.org/10.1007/s11235-019-00639-8

C-ITS

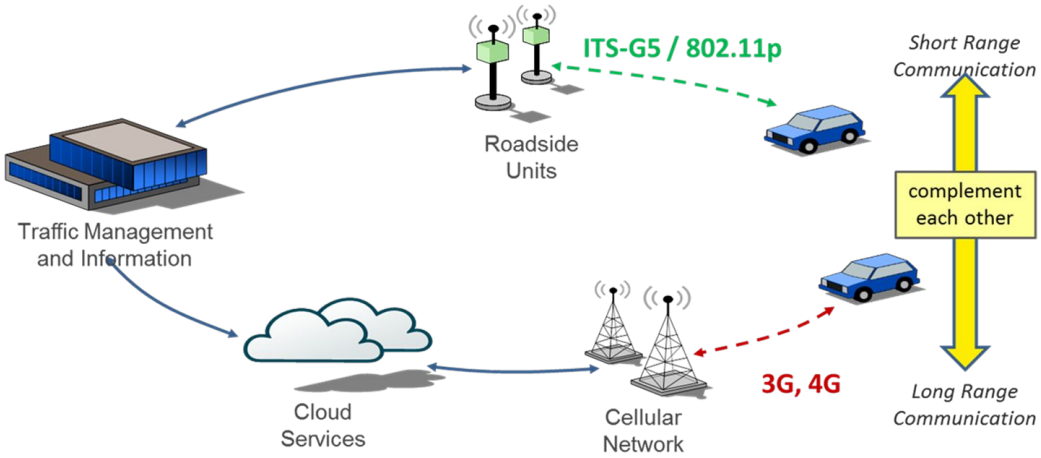
Different types of messages:

- CAM (Cooperative Awareness Message)
- DENM (Decentralized Environmental Notification Message)
- IVI (In-Vehicle Information)
- CPM (Collective Perception Message)
- TLM (Traffic Light Maneuver), etc.

C-ITS

Different Radio Access Technologies (RAT):

- ITS-G5 (IEEE 802.11p),
- Cellular (3G/4G/5G, etc.), etc.



C-ITS

Examples of services for data collection - V2I

- Basic PVD (Probe Vehicle Data): Speed, position, heading, etc.
- Extended PVD: Information on road events, etc.

C-ITS

Examples of services for action - I2V

- GLOSA (Green Light Optimal Speed Advisory)
- RWW (Road Works Warning)
- IVSL (In vehicle speed limit)
- IVS (In vehicle signage - virtual DMS)

C-ITS

Example:

RWW (Road Works Warning) message

Image credit: https://www.c-roads.eu/downloads.html

Coopits - Interactive Map

https://coopits.developpement-durable.gouv.fr/carte-interactive-des-services/

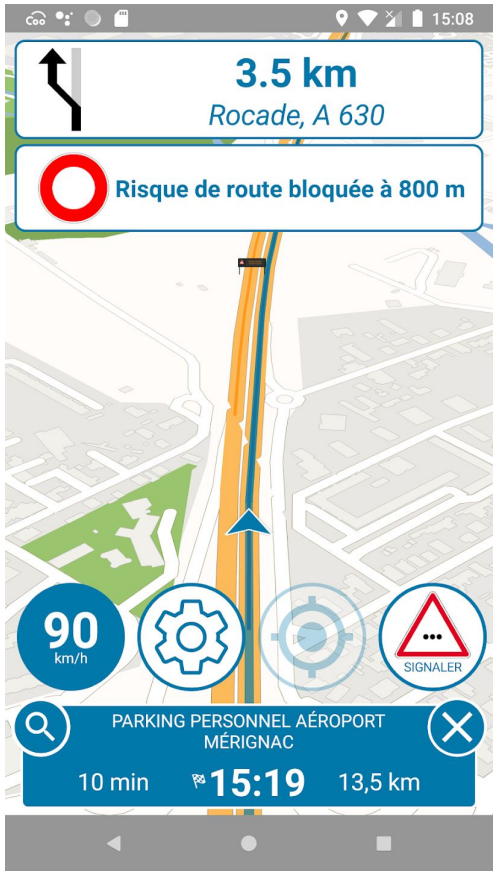
Different actors

Different services

*“**Coopits** est une application innovante et sécurisante qui s’appuie sur les outils numériques dernière génération pour vous aider à mieux conduire : aide à la décision, écoconduite, etc.”*

https://coopits.developpement-durable.gouv.fr/

Coopits - Mobile Application



C-ITS Evaluation

- Latency
- Message size
- Message volume (mean by hour/day/week/month/weekend)
- Events triggered by smartphones or road operators
- Validity duration of event messages
- Message reception
- Message display
- Speed limit
- Slow down/stop after displaying an event
- Maneuver analysis

C-ITS Evaluation

CAM
log_timestamp **log_stationid** position_latitude position_longitude speedvalue headingvalue

DENM
log_timestamp **log_stationid** originatingstationid sequencenumber referencetime event_latitude event_longitude causecode

DENMEVENT
log_timestamp **log_stationid** originatingstationid sequencenumber eventid referencetime causecode

DENMACTION
log_timestamp **log_stationid** eventid eventmodelid eventactionid

C-ITS Problems

- Traffic prediction
- Intrusion detection
- Driver behavior detection
- Accident prediction/detection
- Obstacle detection
- Vehicle movement prediction
- Routing protocol
- Advanced toll system
- Smart parking
- Spectrum sensing
- Energy efficient
- Lane changing assessment
- Entertainment
- Advertising
- etc.

Data Processing

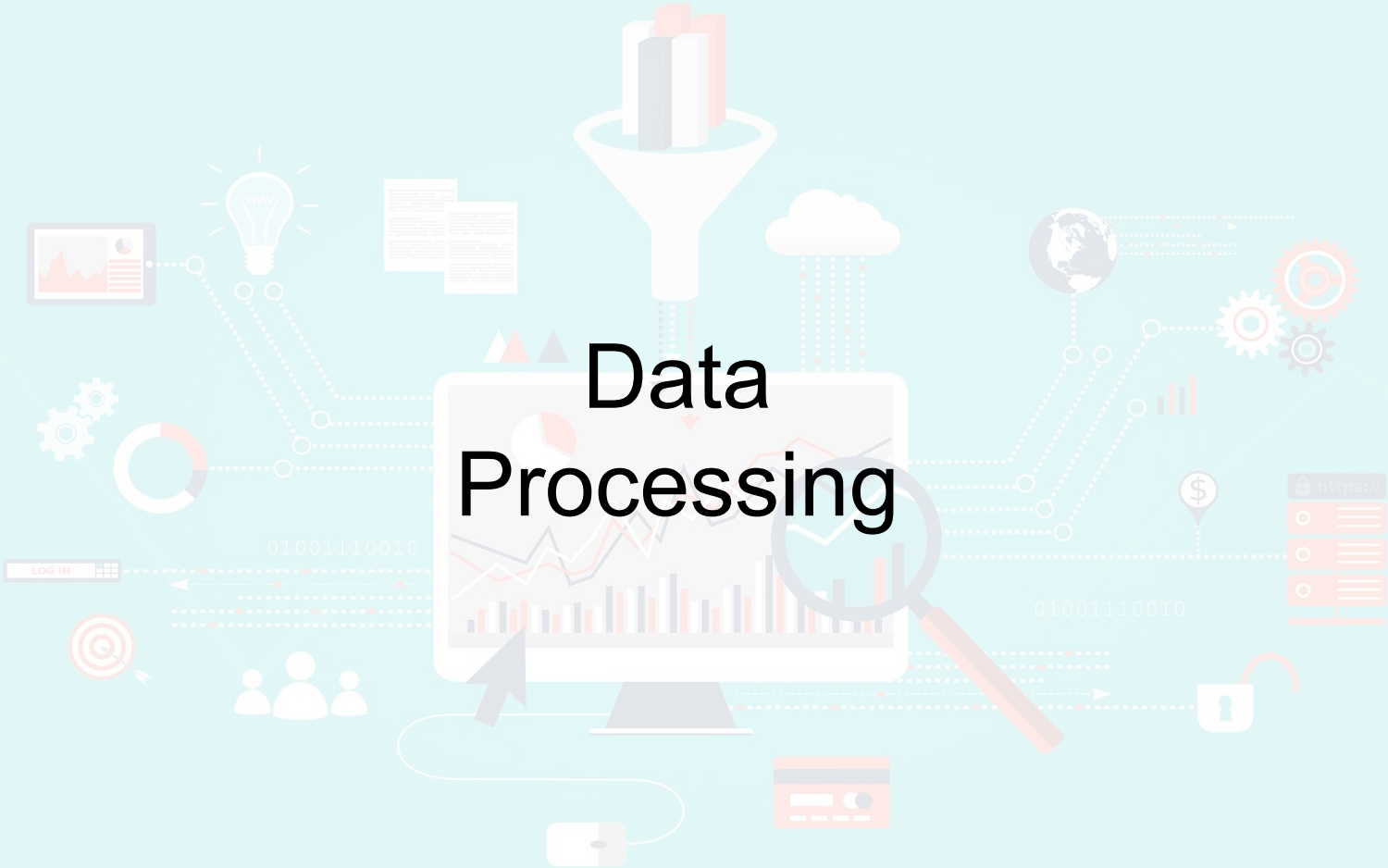


Image credit: Big Data processing with Scalding on Amazon EMR | by Krzysztof Grajek | SoftwareMill Tech Blog

Data Types

- Unstructured data
- Semi-structured data
- Structured data

Unstructured data

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

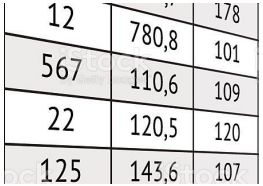
Semi-structured data

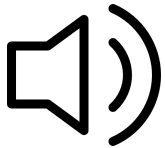
```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

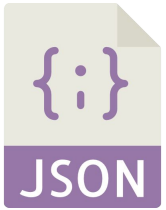
- Numeric data
- Video/Image data
- Text data
- Audio data
- etc.





Data Processing Steps

- Collection
  - JSON, CSV, XML, PNG, etc.
- Preparation
- Input
- Processing
- Output
  - Performance indicators
  - Display results
- Storage



Data Processing Methods

- Single user programming
- Multiple programming
- Real-time processing
- On-line processing
- Time sharing processing
- Distributed processing
- etc.

Data Processing Tools

- Python - Pandas/ Sklearn/ TensorFlow/ Keras
- Spark
- Hadoop
- etc.

- Anaconda
- Jupyter notebook
- Python

Data in C-ITS

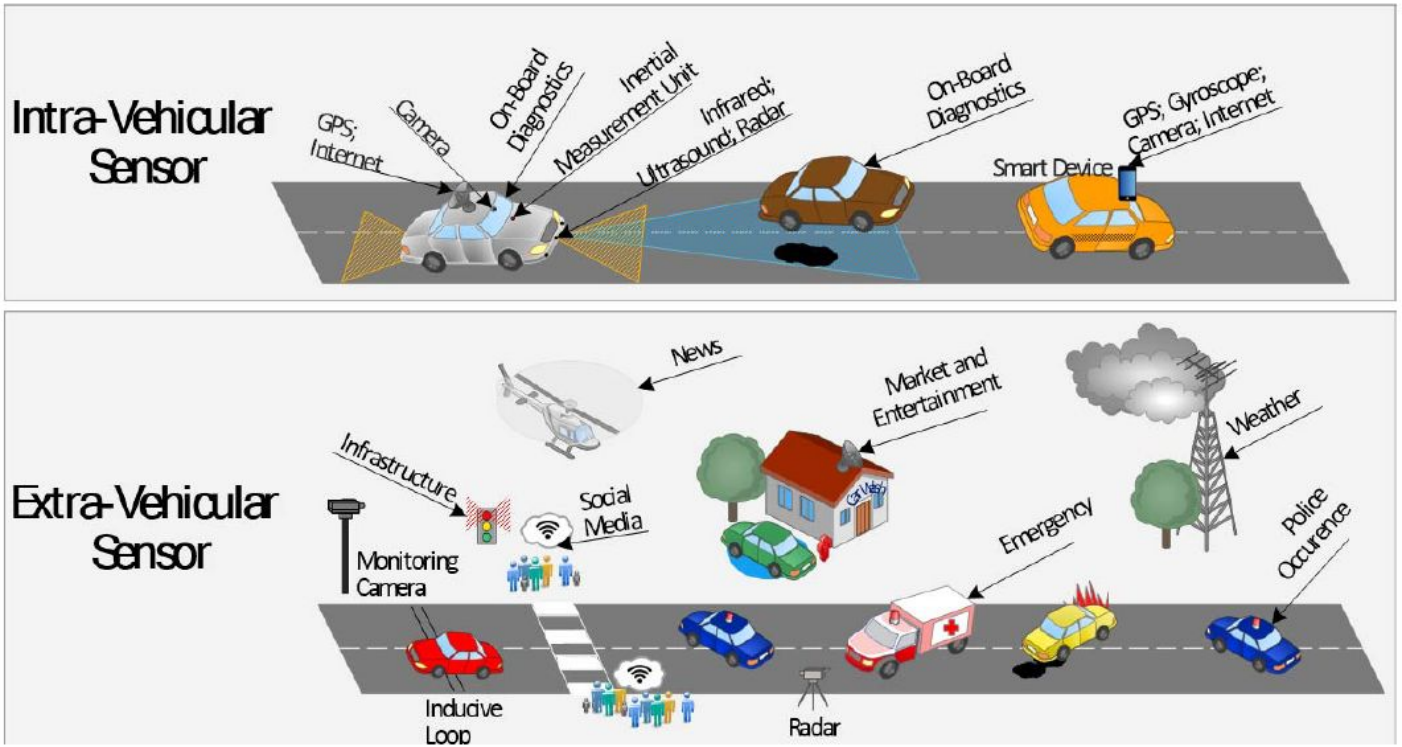


Image credit: P. H. Rettore, G. Maia, L. A. Villas and A. A. F. Loureiro, "Vehicular Data Space: The Data Point of View," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2392-2418, thirdquarter 2019, doi: 10.1109/COMST.2019.2911906.

Data collection in C-ITS

Sensors:

- LiDARs
- Cameras
- Radars
- Onboard sensors
- etc.

Collection of sensor data:

- from vehicles
- from infrastructures

Data collection in C-ITS

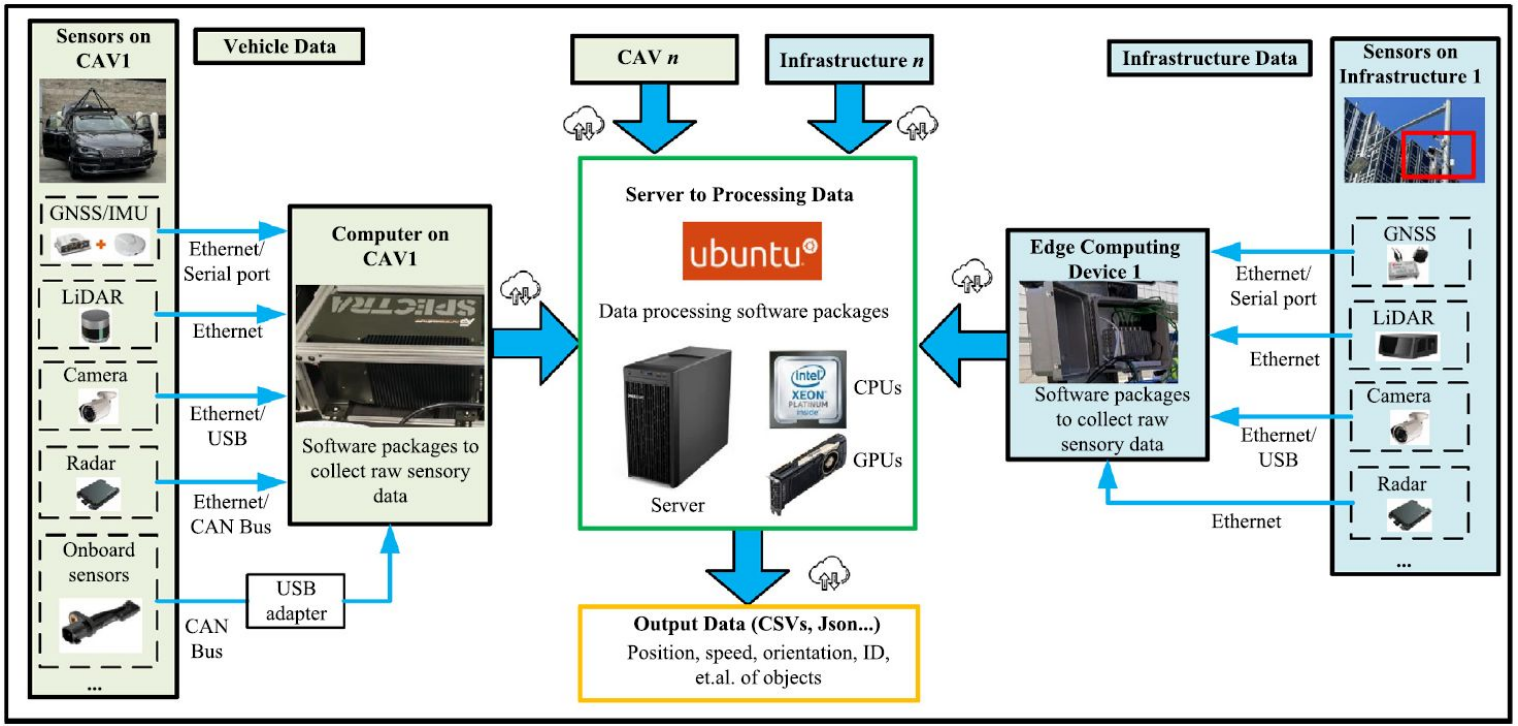


Image credit: Tak S., Lee J.-D., Song J., Kim S. *Development of AI-based vehicle detection and tracking system for C-ITS application*, J. Adv. Transp. (2021)

Data preparation in C-ITS

Processing of sensor data (data including position, speed, ID, etc. through CSVs or JSONs)

Example: Onboard sensors → Communication logs

ASN1DATA:

“0202F069BCBDE08B4059F3DFB24DB431F7200000000033BD1E00924FC0007EBFE9ED0737FEEBFFF600014200007FFFF1CE201C180F1B1CE1FDA78155B1CE1FD648140B1CE1FD00816A71CE1FC7B818371CE1FC5A818BB1CE1FA43810371CE1FEE57E3471CE1FF267E2931CE1FE6F7E4A71CE1FC5A7EC331CE1FB607EFDB1CE1FC277EF131CE1FD757EE471CE1FED37EBF31CE1FFBD7E1871CE200F97DE231CE202467E3171CE202BB7E84F1CE202AA7E8CF1CE202477EC371CE2030E7E7C71CE202AA7EA5F1CE2031F7E6FF1CE2030F7E88F1CE203517E67B1CE202DC7E8D31CE2030E7E9DB1CE202367EECF1CE202577EAE71CE202577E6371CE2029A7EA1F1CE202577EF531CE202997EE8F1CE202AB7EBAF1CE202357EC771CE202897EC771CE202257ECFB1CE2029A7E6FF1CE0”

Fields in .csv:

protocolversion double precision	messageid double precision	stationid double precision	generationdeltatime double precision	stationtype double precision	referencetposition_postgis geography
2	2	4033461437	57483	5	0101000020E6100000A1061DBE96710F40A22EADE17DCA4540

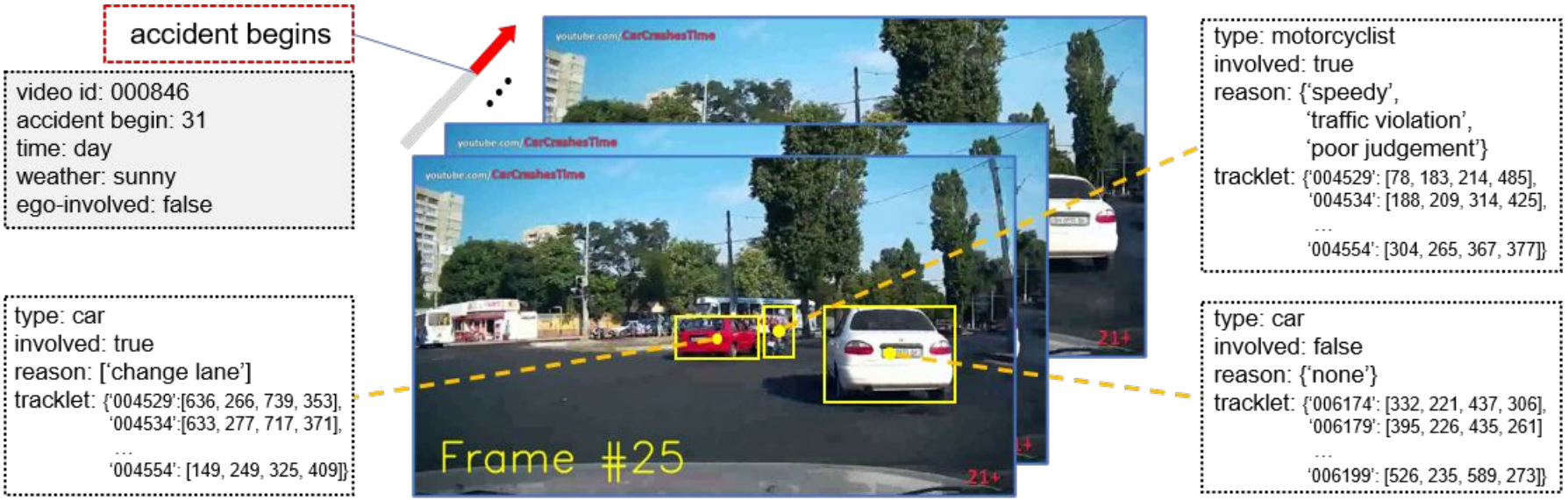
referencetposition_latitude double precision	referencetposition_longitude double precision	referencetposition_semimajorconfidence double precision	referencetposition_semiminorconfide double precision
43.5819666	3.9304633	0	

referencetposition_semimajororientation double precision	referencetposition_altitudevalue double precision	referencetposition_altitudeconfidence altitudeconfidence	headingvalue double precision	headingconfidence double precision	speedvalue double precision	speedconfidence double precision	drivedirection drivedirection	vehiclelengthvalue double precision	vehiclelengthconfidenceindication vehiclelengthconfidenceindication	vehiclewidth double precision	longitudinalaccelerationvalue double precision	longitudinalaccelerati double precision
0	5960	unavailable	2340	[null]	0	[null]	unavailable	[null]	unavailable	[null]	[null]	

Data preparation in C-ITS

Processing of sensor data (data including position, speed, ID, etc. through CSVs or JSONs)

Example: Cameras → Videos

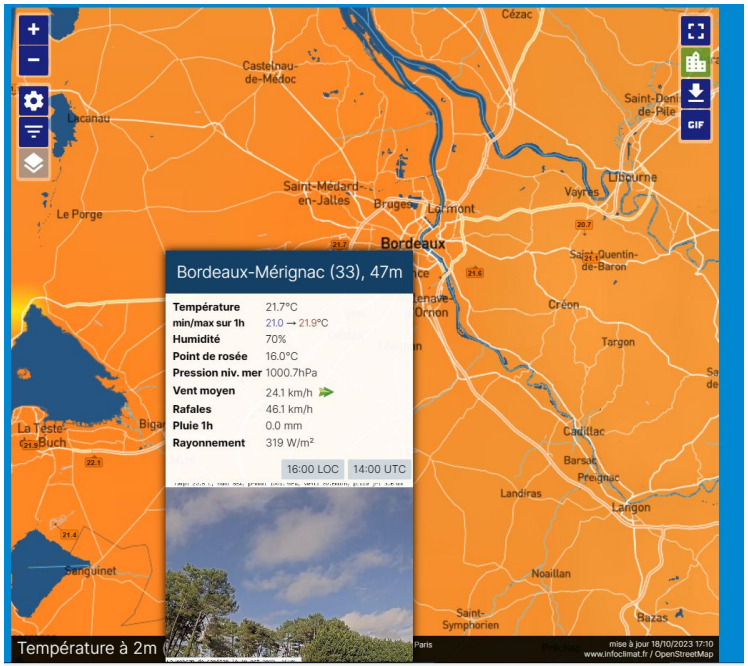


Car Crash Dataset, available at https://github.com/Cogito2012/CarCrashDataset

Data preparation in C-ITS

Processing of sensor data (data including position, speed, ID, etc. through CSVs or JSONs)

Example: Meteorological sensors

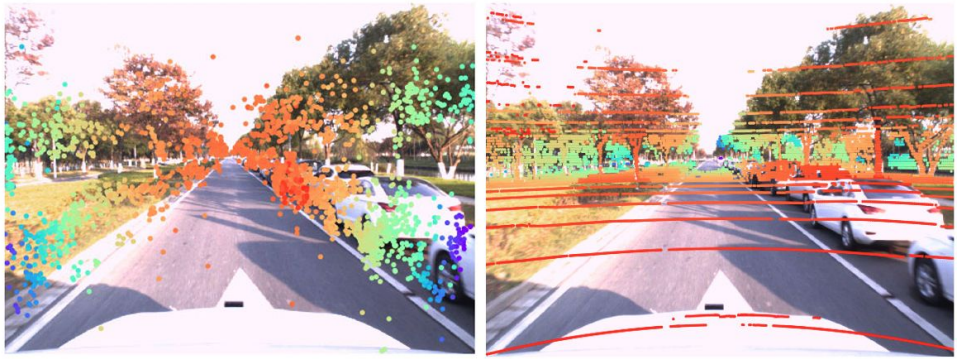


Cartes en temps réel des stations météorologiques mondiales, available at
https://www.data.gouv.fr/fr/reuses/cartes-en-temps-reel-des-stations-meteorologiques-mondiales/

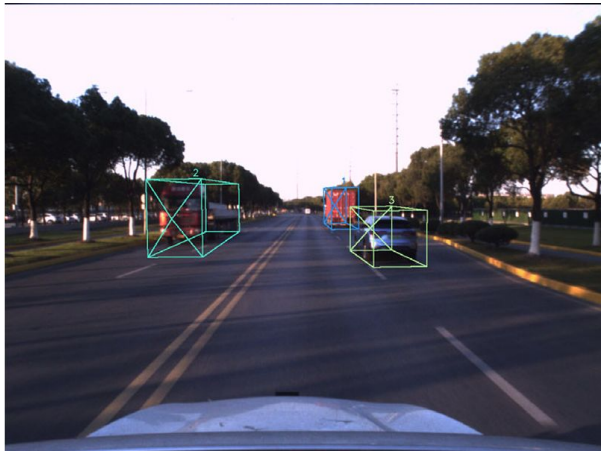
Data preparation in C-ITS

Processing of sensor data (data including position, speed, ID, etc. through CSVs or JSONs)

Example: Radars, Lidars



(a) 4D radar point clouds (b) Lidar point clouds



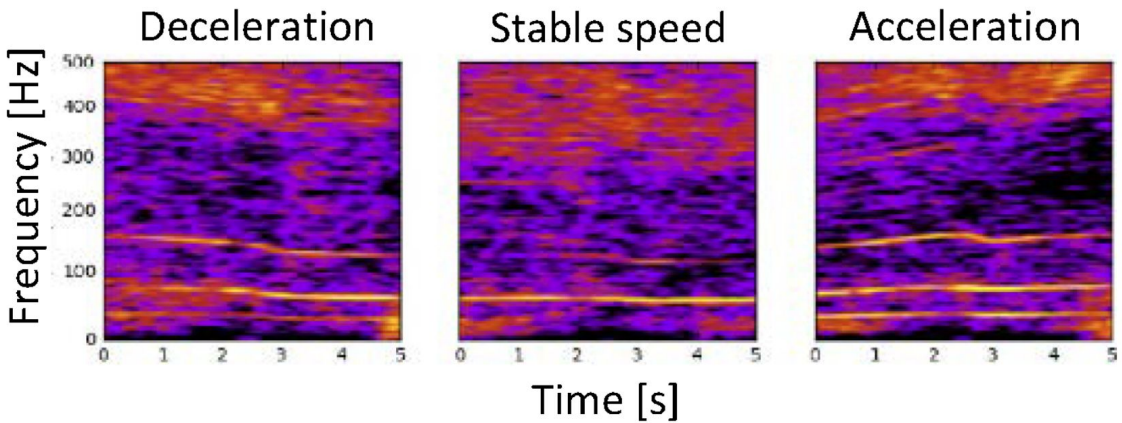
3D bounding box of each object includes the center point (*x,y,z*), length, width, height (*l,w,h*), and orientation angle (yaw)

L. Zheng *et al*., "TJ4DRadSet: A 4D Radar Dataset for Autonomous Driving," *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, Macau, China, 2022, pp. 493-498, doi: 10.1109/ITSC55140.2022.9922539.

Data preparation in C-ITS

Processing of sensor data (data including position, speed, ID, etc. through CSVs or JSONs)

Example: Audio sensors → Spectrogram

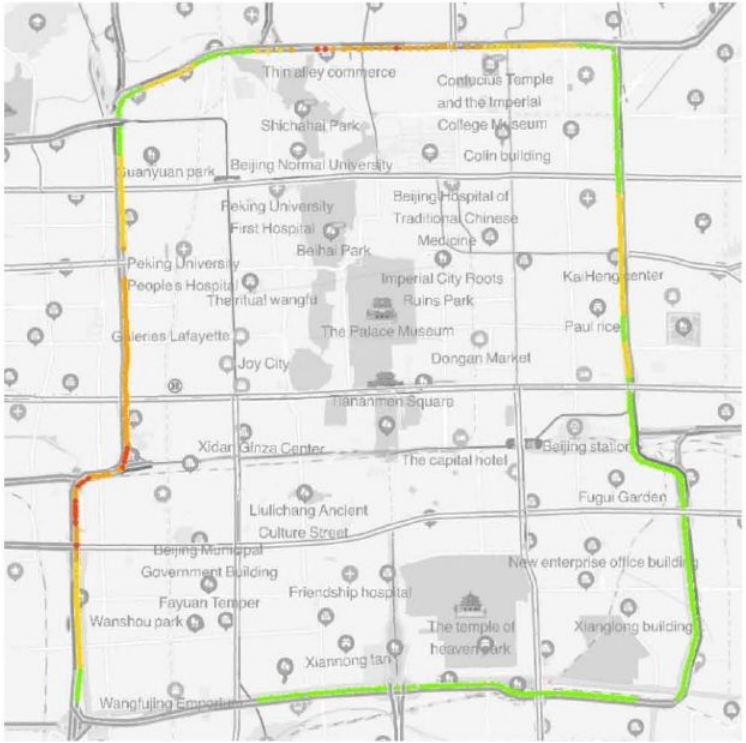


Zero Crossings Rising, Audio Spectrum Envelope (SE), SumSE, MaxSE, IndMaxSE (index of the SE maximum), Audio Spectrum Flatness, Energy (energy of the entire spectrum), Audio Spectrum Centroid, Audio Spectrum Spread

Kubera E, Wieczorkowska A, Kuranc A, Słowik T. Discovering Speed Changes of Vehicles from Audio Data. *Sensors*. 2019; 19(14):3067. https://doi.org/10.3390/s19143067

C-ITS Problems

Traffic prediction



400 points
Between 7:00-20:00
During 1 month
Pekin/China

Features: month, day, weekday, hour, minute, and index

Traffic density classes:

- Unimpeded
- Slow
- Congested
- Seriously congested

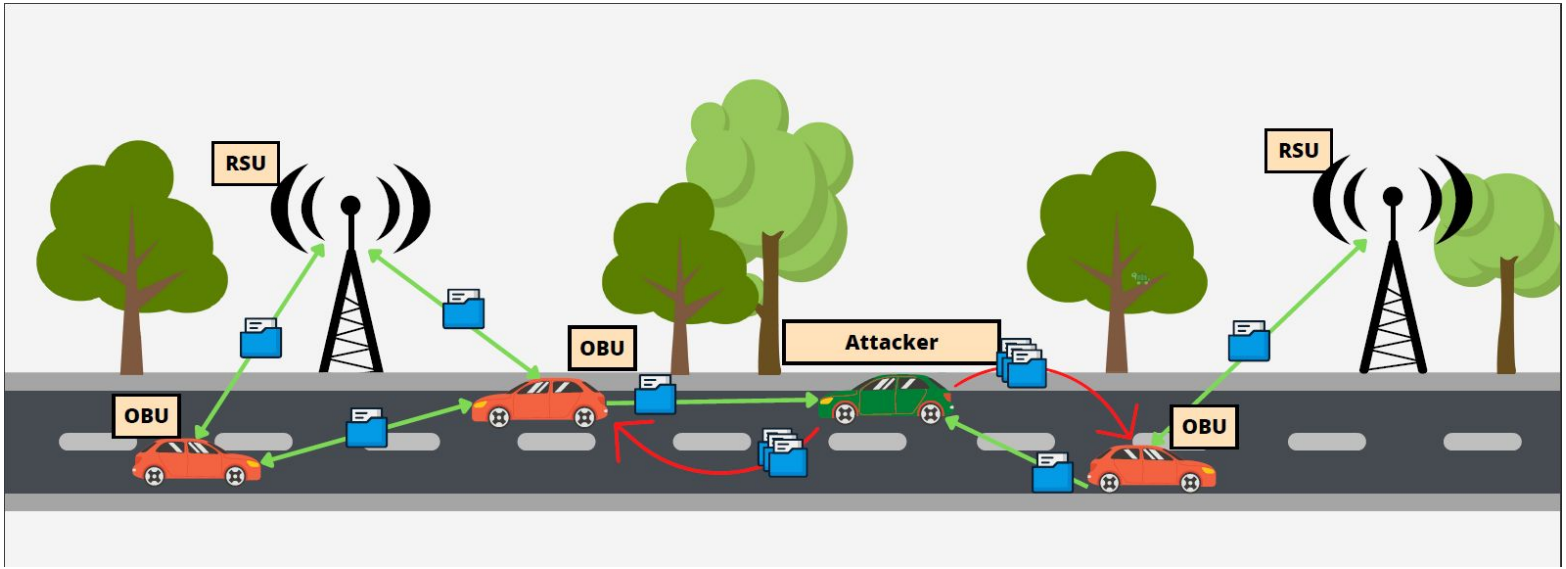
H. Wang, X. Wei, J. Yao and Y. Zhang, "Traffic Flow Prediction Using Machine Learning Methods," *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, China, 2021, pp. 30-35, doi: 10.1109/MLBDBI54094.2021.00014.

C-ITS Problems

Intrusion detection

Different attack types:

- DoS (denial-of-service)
- DoS Random
- DoS Disruptive
- DoS Random Sybil
- DoS Disruptive Sybil



S. Ercan, L. Mendiboure, L. Alouache, S. Maaloul, T. Sylla and H. Aniss, "An Enhanced Model for Machine Learning-Based DoS Detection in Vehicular Networks," 2023 IFIP Networking Conference (IFIP Networking), Barcelona, Spain, 2023, pp. 1-9, doi: 10.23919/IFIPNetworking57963.2023.10186390.

C-ITS Problems

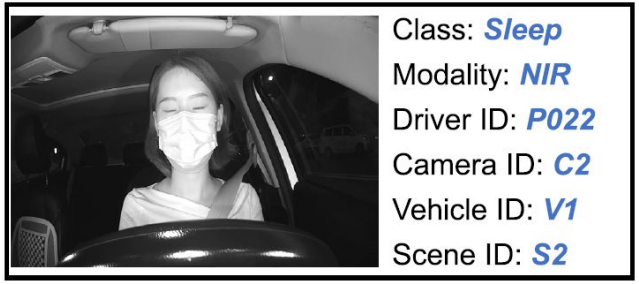
Driver behavior detection

4 cameras for each driver (front-left, front, front-right, and side-right)

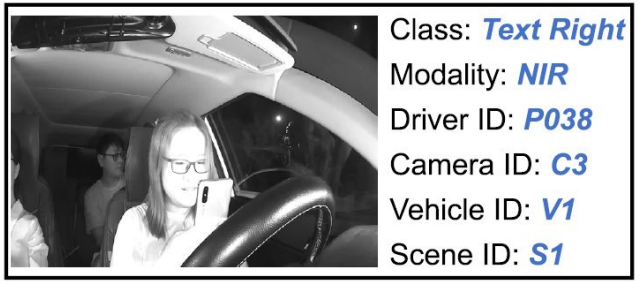
> 470K images

Class: ***Talk Left***
Modality: ***RGB***
Driver ID: ***P007***
Camera ID: ***C1***
Vehicle ID: ***V3***
Scene ID: ***S3***

Class: ***Eat Right***
Modality: ***RGB***
Driver ID: ***P008***
Camera ID: ***C4***
Vehicle ID: ***V1***
Scene ID: ***S2***



Class: ***Sleep***
Modality: ***NIR***
Driver ID: ***P022***
Camera ID: ***C2***
Vehicle ID: ***V1***
Scene ID: ***S2***



Class: ***Text Right***
Modality: ***NIR***
Driver ID: ***P038***
Camera ID: ***C3***
Vehicle ID: ***V1***
Scene ID: ***S1***

No.	Behavior
1	Normal driving
2	Sleeping
3	Yawning
4	Talk with cellphone (left)
5	Talk with cellphone (right)
6	Texting (left)
7	Texting (right)
8	Hair / makeup
9	Looking left
10	Looking right
11	Looking up
12	Looking down
13	Smoking (left)
14	Smoking (right)
15	Smoking (mouth)
16	Drinking / Eating (left)
17	Drinking / Eating (right)
18	Adjusting radio
19	Operating GPS / entertainment system
20	Reaching behind
21	Hands off the steering wheel
22	Talking to passengers

J. Wang *et al*., "100-Driver: A Large-Scale, Diverse Dataset for Distracted Driver Classification," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7061-7072, July 2023, doi: 10.1109/TITS.2023.3255923.

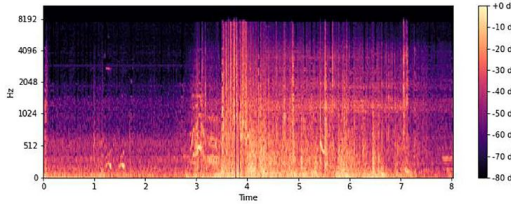
C-ITS Problems

Accident detection

Video data

Audio data

- audio features (energy, entropy, spectral flux, etc.)
- spectrogram images



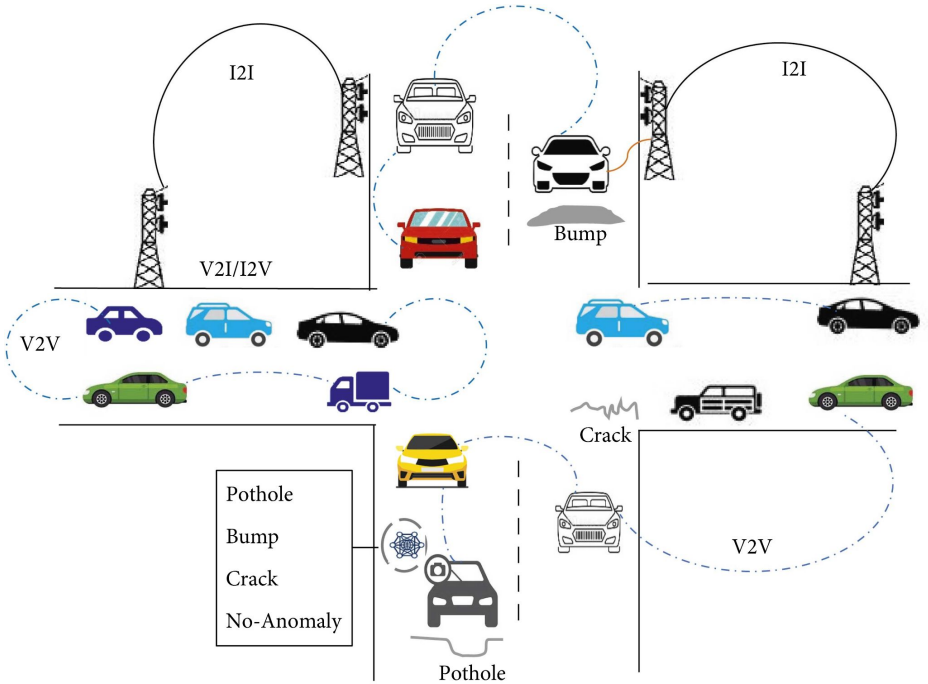
J.G. Choi, C.W. Kong, G. Kim, S. Lim, Car crash detection using ensemble deep learning and multimodal data from dashboard cameras, *Expert Syst. Appl.* (2021) 115400.

C-ITS Problems

Obstacle detection

Obstacles:

- Bump
- Pothole
- Crack

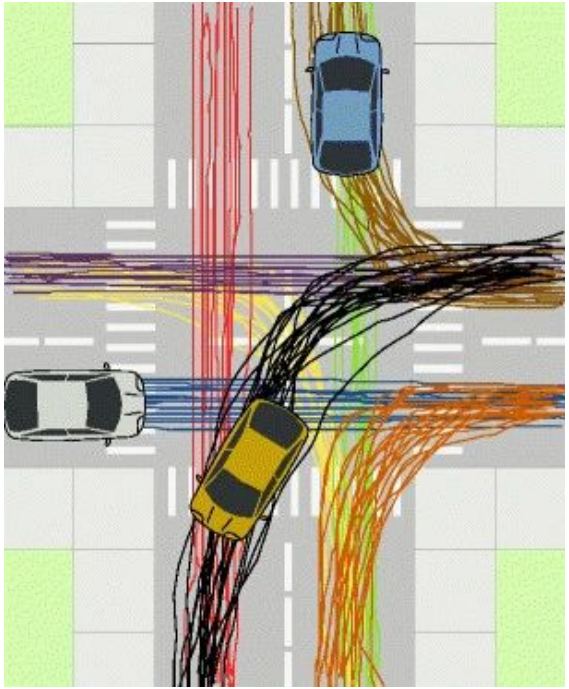


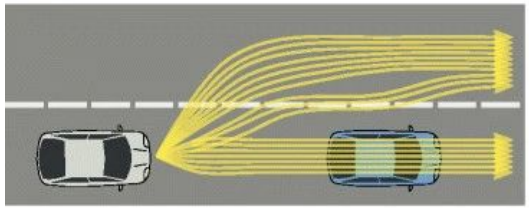
Images:

R. Bibi, Y. Saeed, A. Zeb, T.M. Ghazal, T. Rahman, R.A. Said, et al., "Edge AI-based automated detection and classification of road anomalies in VANET using deep learning", *Computational intelligence and neuroscience*, 2021.

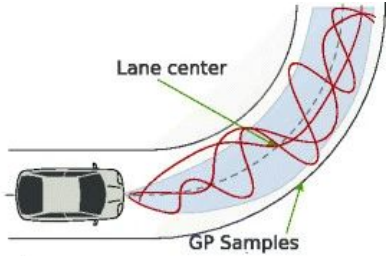
C-ITS Problems

Vehicle movement prediction

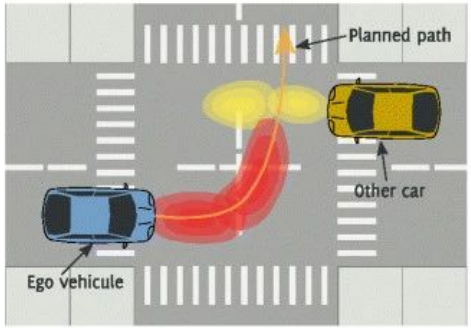




a



b



c

Maneuver classes:

- Go straight
- Left turn
- Right turn
- Stop

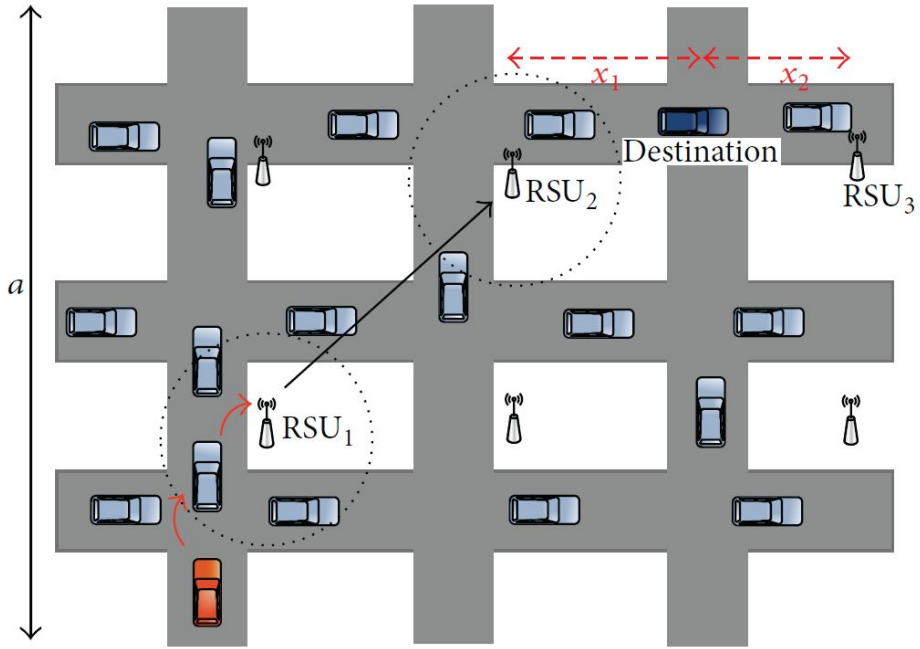
Features:

- distance to intersection,
- position,
- speed,
- longitudinal acceleration,
- distance to lane center,
- steering angle,
- orientation,
- road curvature,
- turn signal

Lefèvre, S., Vasquez, D. & Laugier, C. A survey on motion prediction and risk assessment for intelligent vehicles. *Robomech J* 1, 1 (2014). https://doi.org/10.1186/s40648-014-0001-z

C-ITS Problems

Routing protocol



Prediction mechanism of vehicle location

Classes

- go straight,
- turn left,
- turn right

Features:

- lane number,
- direction,
- speed,
- traveling roads

Evaluation mechanism of transmission capacity (collision probability)

Evaluation mechanism of forwarding direction

W. K. Lai, M.-T. Lin and Y.-H. Yang, "A machine learning system for routing decision-making in urban vehicular ad hoc networks", *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 3, pp. 374-391, 2015.

Application

Installation

- For ubuntu users: https://github.com/lmendiboure/DP_TP/blob/master/installation-guide-updated.md

- Download **anaconda**:

  https://www.anaconda.com/download

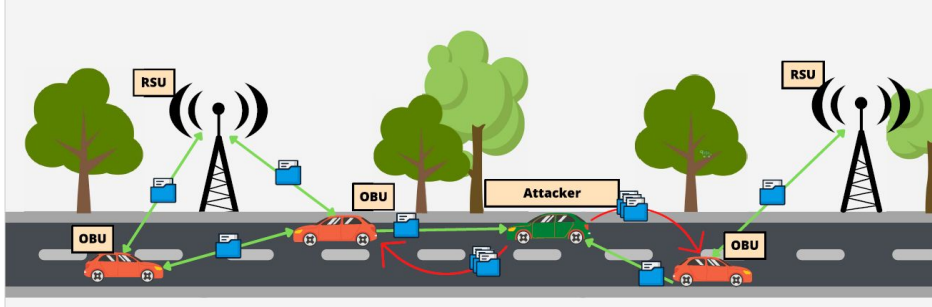
- Download **Python**:

  https://www.python.org/downloads/

- Install necessary libraries:

  pip install tensorflow

Intrusion Detection problem



- Anomaly based Intrusion Detection System
- Different attack types: Denial-of-Service (DoS), Sybil, position falsification, black-hole and gray-hole attack
- An open public dataset: VeReMi
  - Vehicular Reference Misbehavior (built specifically for testing V2X security)
  - A simulated dataset, generated using LuST and VEINS (based on OMNET++ and SUMO)

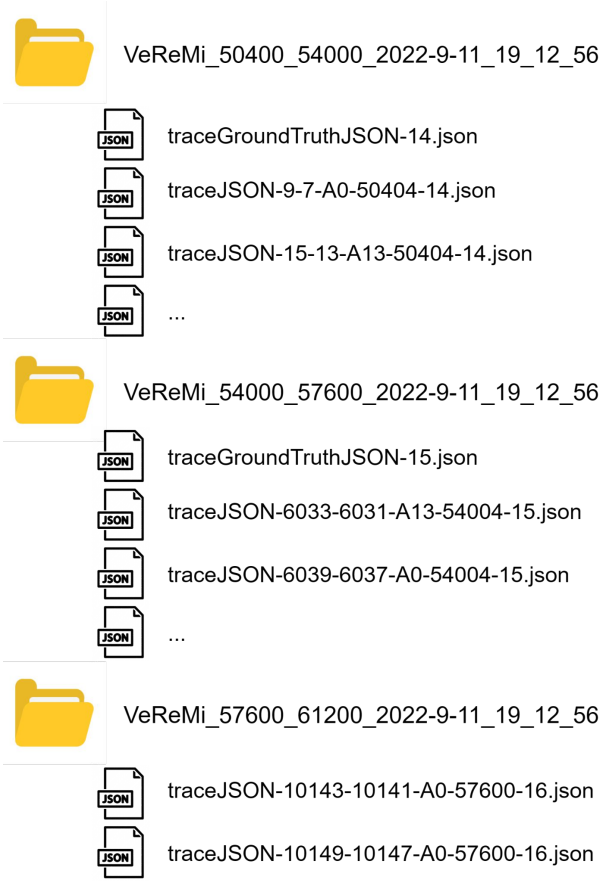
- Download dataset (only the zip file: **VeReMi_54000_57600_2022-9-11_19.12.56.zip**)

  https://mega.nz/folder/z0pnGA4a#WFEUISyS5_maabhcEI7HQA/folder/a1QxhaqC

- Python file for the application

  https://github.com/ercansec/DataProcessing_CITS

Intrusion Detection problem - Dataset



- Any json file
  - Type = 2 (GPS)
    - Sent messages
  - Type = 3 (BSM-Basic Safety Message)
    - Received messages
- Ground Truth json file
  - Type = 4 (Ground truth)
    - Received messages

traceJSON-6033-6031-A13-54004-15.json

- 6033: receiver
- 6031: OMNET++ module ID
- A13: attacker type

Intrusion Detection problem - Dataset

type =2 (GPS)

- rcvTime
- pos & noise
- spd & noise
- acl & noise
- hed & noise

type = 3 (BSM)

- rcvTime
- sendTime
- sender
- senderPseudo
- messageID
- pos & noise
- spd & noise
- acl & noise
- hed & noise

type = 4 (Ground Truth)

- sendTime
- sender
- senderPseudo
- messageID
- pos & noise
- spd & noise
- acl & noise
- hed & noise

Data Processing Application

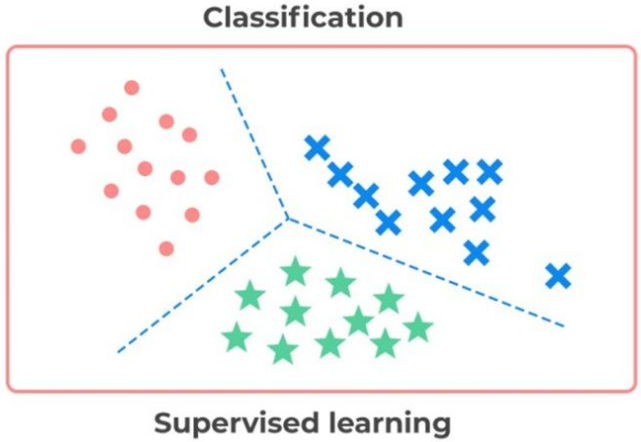
- Import data
- Descriptive statistics
- Check null values, missing data
- Data split for learning
- Implementation of ML methods
  - Classification
  - Clustering
- Output performance indicators
- Save results

Classification

Supervised Learning - with labels (classes)

Examples of Classification Methods:

- Decision Tree (DT)
- Support Vector Machine (SVM)
- k-Nearest Neighbor (kNN)
- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)
-

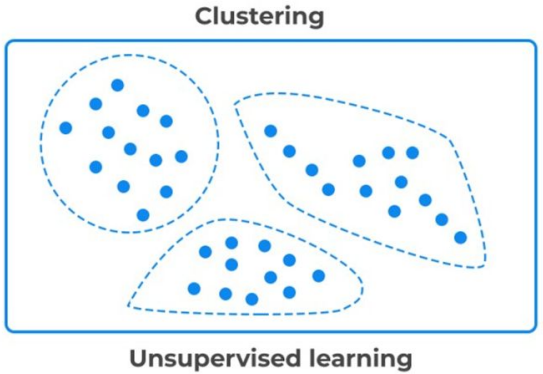


Clustering

Unsupervised Learning - without labels (classes)

Examples of Clustering Methods:

- Self Organizing Map (SOM)
- k-Means
- Gaussian Mixture Model (GMM)
-



Performance Indicators

Confusion Matrix:

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

TN: True Negative

TP: True Positive

FN: False Negative

FP: False Positive

Evaluation Indicators

- Accuracy = (TP+TN)/(TP+TN+FP+FN)
- Precision = TP/(TP+FP)
- Recall = TP/(TP+FN)
- F1-Score = 2*Precision*Recall/(Precision+Recall)

Homework

Homework

Driver Behavior Dataset

- Download dataset (**features_14.csv**):

https://www.kaggle.com/datasets/shashwatwork/driving-behavior-dataset?resource=download

https://data.mendeley.com/datasets/jj3tw8kj6h/2

Context

Driver behavior is one of the most important aspects in the design, development, and application of Advanced Driving Assistance Systems (ADAS) and Intelligent Transportation Systems (ITS), which can be affected by many factors. If you are able to measure the driving style of your staff, there is a lot of actions you can take in order to improve fleet safety, global road safety as well as fuel efficiency and emissions.

Content

- Dataset for modeling risky driver behaviors based on accelerometer (X,Y,Z axis in meters per second squared (m/s2)) and gyroscope (X,Y, Z axis in degrees per second (°/s)) data.
- Sampling Rate: Average 2 samples (rows) per second
- Cars: Ford Fiesta 1.4, Ford Fiesta 1.25, Hyundai i20
- Drivers: 3 different drivers with the ages of 27, 28 and 37
- Driver Behaviors:

1.Sudden Acceleration (Class Label: 1)
2.Sudden Right Turn (Class Label: 2)
3.Sudden Left Turn (Class Label: 3)
4.Sudden Break (Class Label: 4)

- Best Window Size: 14 seconds
- Sensor: MPU6050
- Device: Raspberry Pi 3 Model B

Acknowledgements

Yuksel, Asim; Atmaca, Şerafettin (2020), "Driving Behavior Dataset", Mendeley Data, V2, doi: 10.17632/jj3tw8kj6h.2

Homework

A FAIRE :

- Importez les données
- Nettoyez le jeu de données (si besoin)
- Découvrez les variables
- Implémentez une méthode de Machine Learning pour la classification (Choisissez une méthode et comparez avec XGBoost comme on a fait)
- Utilisez les indicateurs accuracy, precision et recall pour évaluer la performance de votre méthode de classification et celle de XGBoost
- Commentez les résultats

Travaillez individuellement et envoyez votre fichier Python et compte-rendu de vos résultats avec votre commentaire à **sercan@bordeaux-inp.fr**
La date limite de dépôt : 8 décembre 2023 à 23h59