MUNI C4E

lasaris LAB OF SOFTWARE ARCHITECTURES AND INFORMATION SYSTEMS

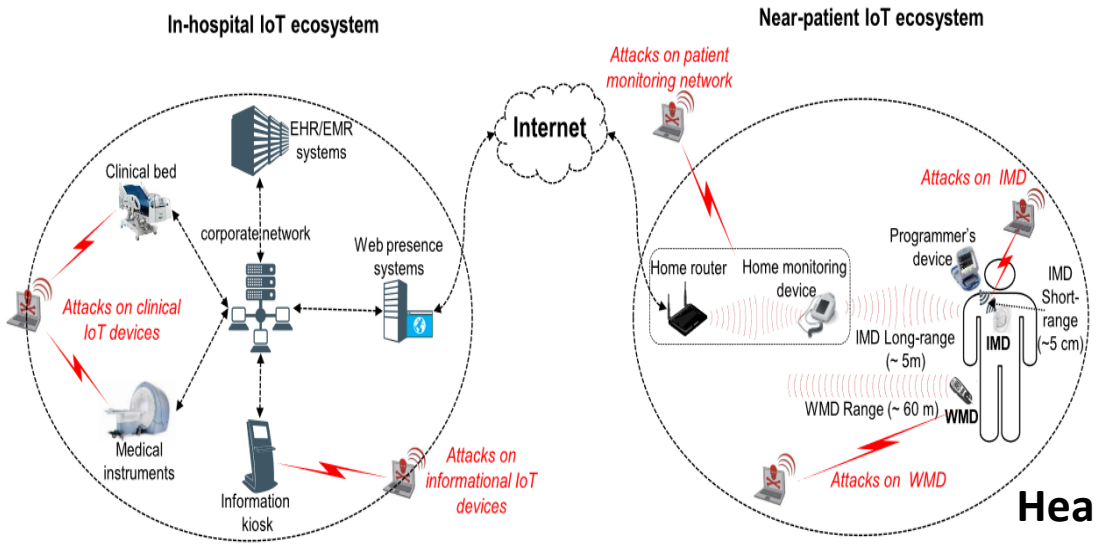
# Machine Learning-Based Data Approximation To Improve Cybersecurity

**Hind Bangui**
Hind.bangui@mail.muni.cz

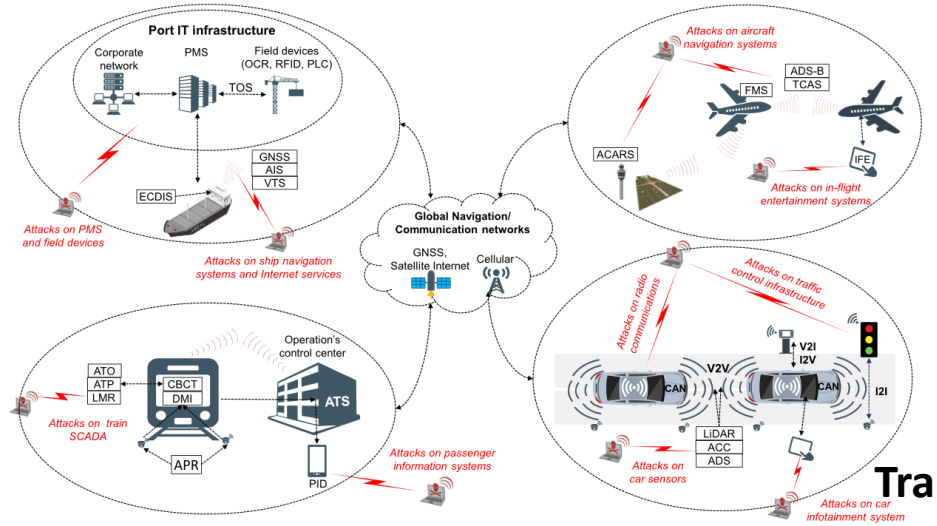
**Barbora Buhnova**
buhnova@fi.muni.cz

# Motivation

- Cyber Physical Systems (CPSs) aim to conduct pre-competitive research on architectures and design, modeling, and analysis techniques for cyber-physical systems, with emphasis on industrial applications.
- The CPS is a system deeply combined with computing and physical system.
- Critical infrastructures are evolving into cyber-physical systems by adopting information and communication technologies.
- Critical infrastructures are more vulnerable now than before to new security threats.



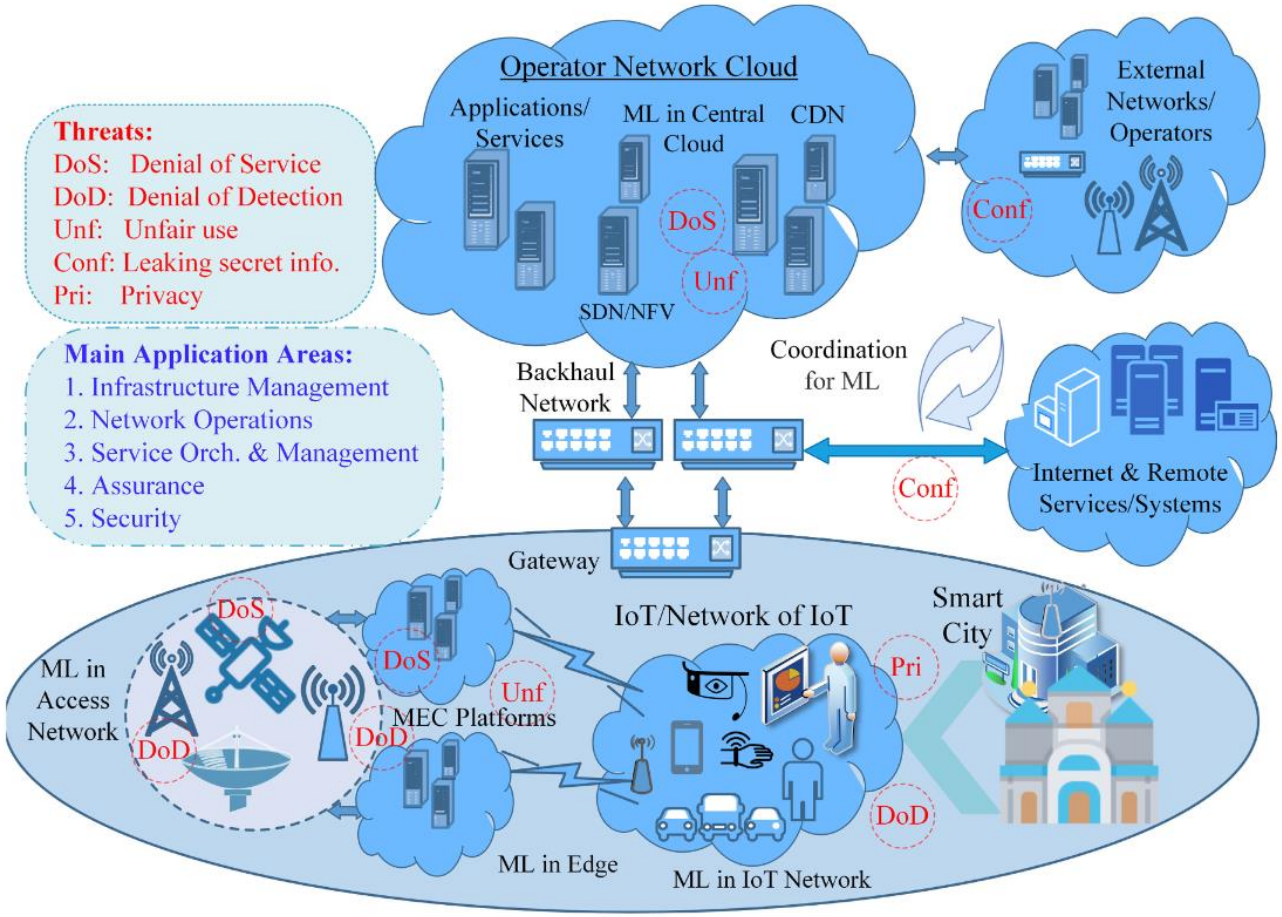

**Healthcare**




**Transportation**

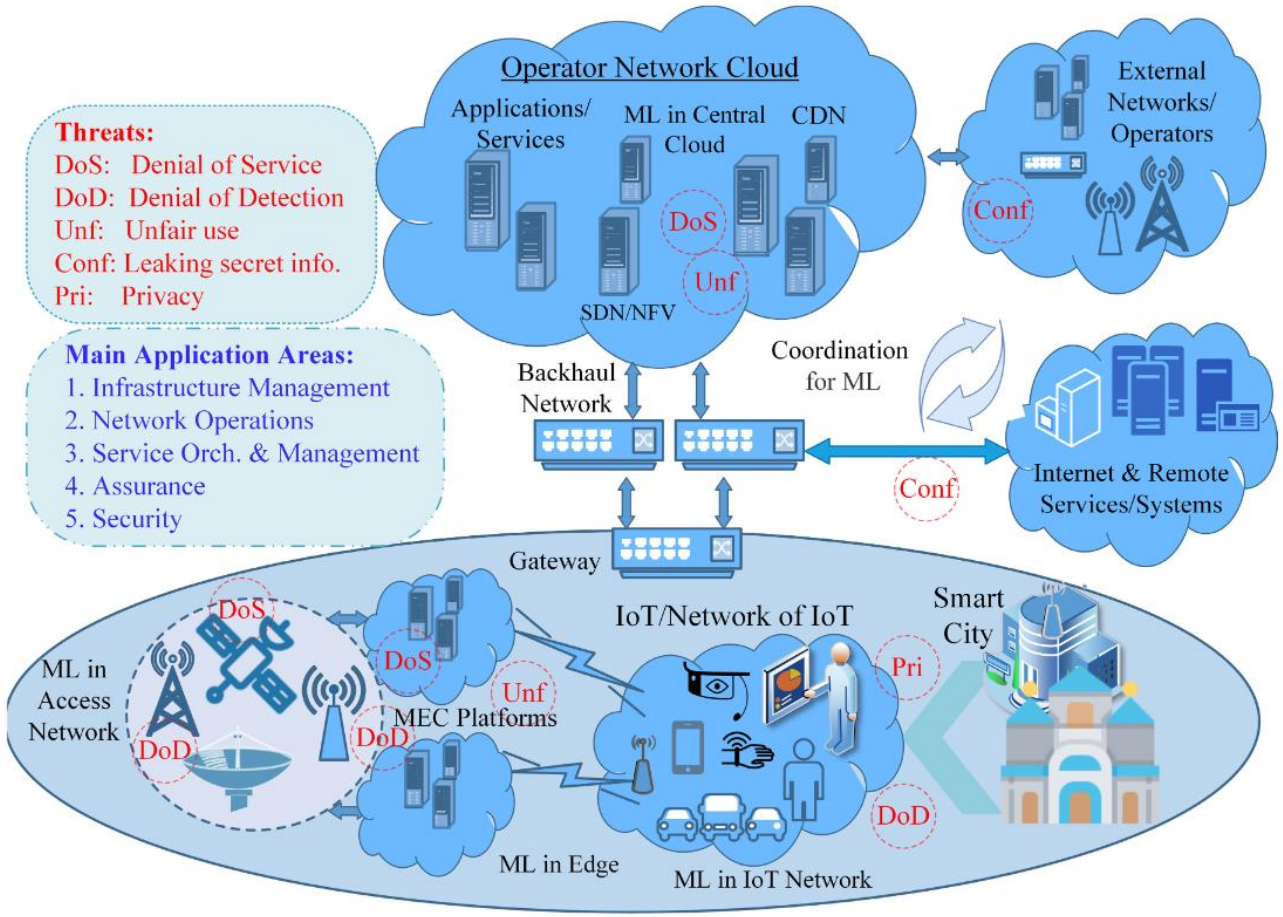
# Motivation

- Advances in Articial Intelligence (AI) techniques show promise in enabling cybersecurity experts to counter the ever-evolving threat posed by adversaries.
- Critical systems demand for solutions with low latency, energy efficiency, and high bandwidth.
- Edge computing has gained attention for bringing compute and storage resources in user proximity.
- Traditional cybersecurity solutions are becoming inadequate at detecting and mitigating emerging cyberattacks.



# Motivation

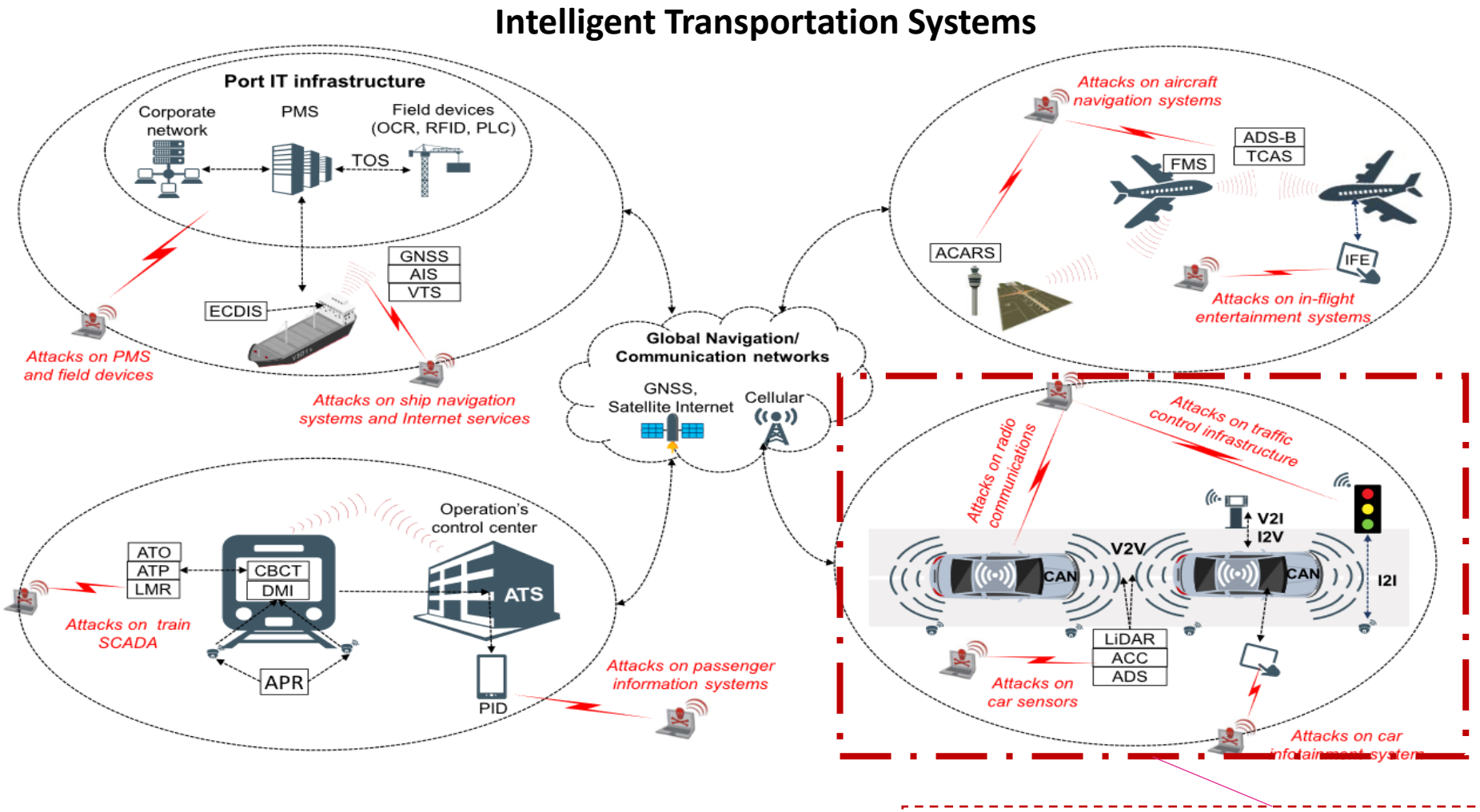
- Machine learning-based cybersecurity solutions should meet the edge computing requirements:
  - Support dynamic and ever-changing circumstances and requirements.
  - Meet the demand of low latency and energy consumption.
  - Deal with intensive computing.
  - Learn unbounded streaming data.




**The main *purpose of research***

How to **redesign** the existing cybersecurity solutions with different **AI algorithms**, like **machine learning**, to make them **compatible** to edge computing?

# Scenario

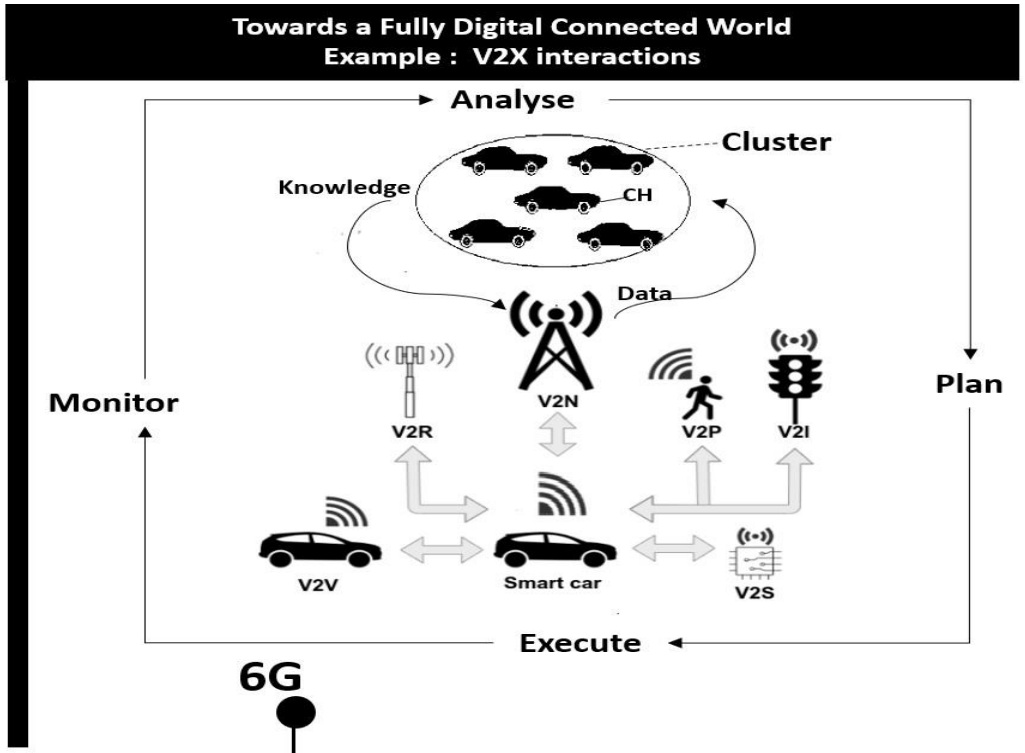


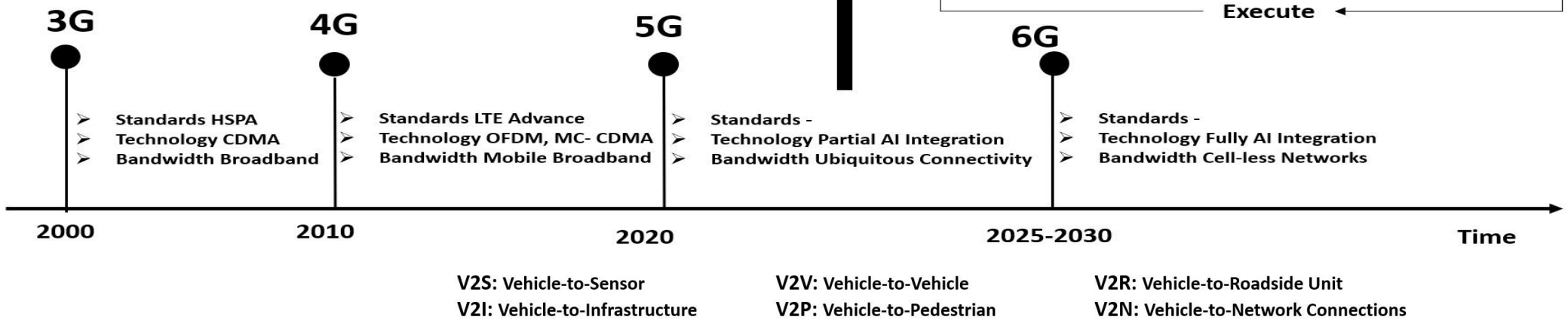

- The aim of Intrusion Detection Systems is detecting attacks in VANETs.
- IDS= Intrusion Detection System

**Intrusion Detection Systems in VANETs (vehicular ad hoc networks )**

# Scenario

- Smart mobility refers to the use of ICT in modern transport technologies to improve urban traffic.
- Vehicular Ad-hoc Network (VANET) is a typical smart mobility system.
- VANET comes under the subgroup of conventional Mobile Ad hoc Network (MANET).

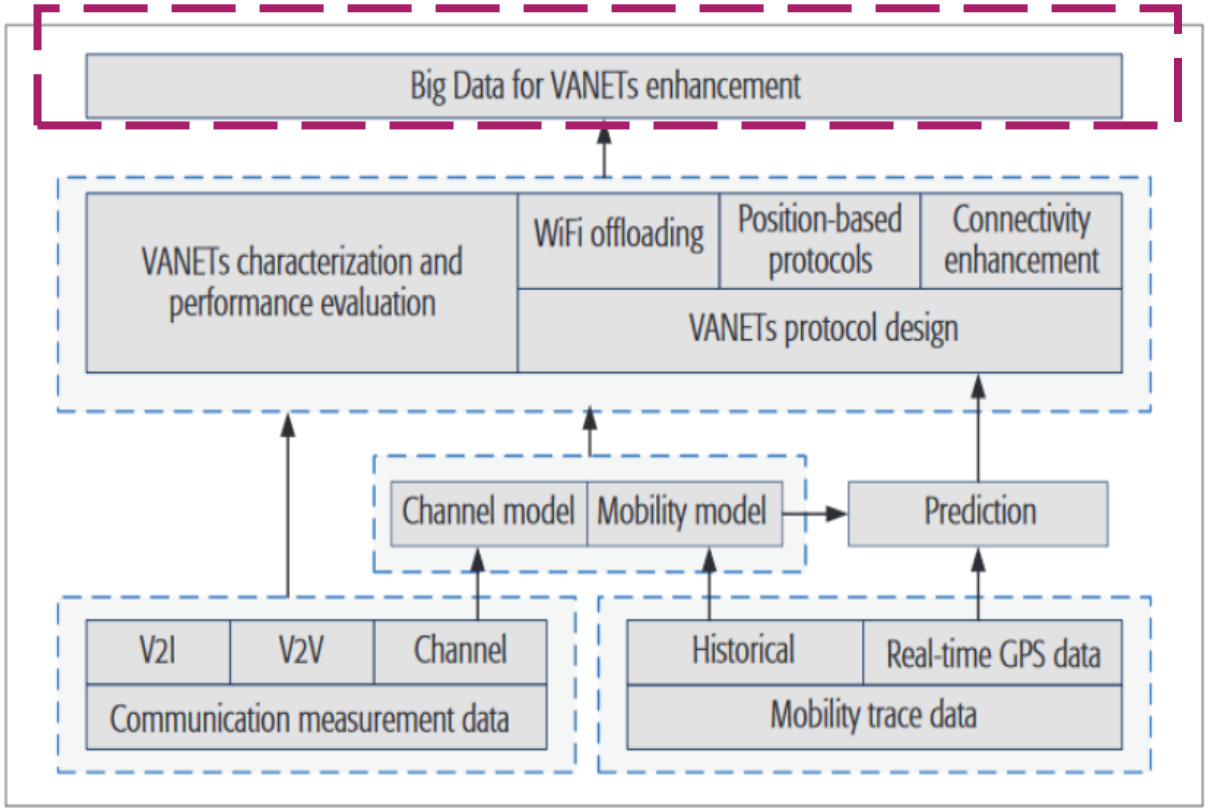



**V2S:** Vehicle-to-Sensor
**V2I:** Vehicle-to-Infrastructure
**V2V:** Vehicle-to-Vehicle
**V2P:** Vehicle-to-Pedestrian
**V2R:** Vehicle-to-Roadside Unit
**V2N:** Vehicle-to-Network Connections

# Scenario

- Data in VANETs can well match the “6Vs” of big data characteristics:
  - **Volume:** Refers to data generated in a large quantity from various sources
  - **Variety:** Refers to structured, semi-structured, and unstructured data with various dimensions.
  - **Velocity:** The data is generated with a speed which leads to new challenges in terms of management.
  - **Value:** Represents the meaningful contribution of big data in making decisions.
  - **Veracity:** Refers to the reliability and quality of big data generated from various sources.
  - **Variability:** This refers to the fact that data changes from time to time



# Scenario

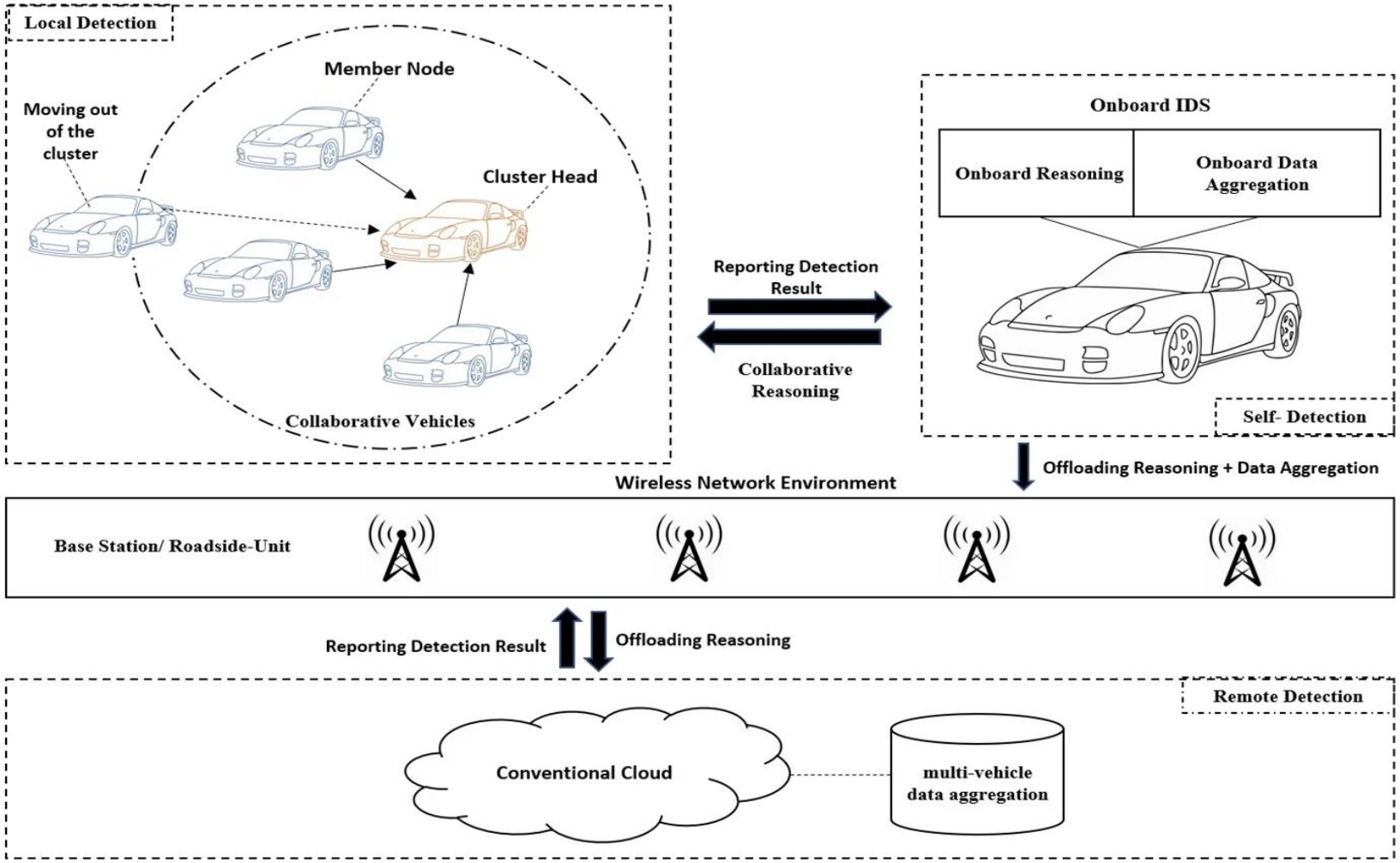
| VANET Applications | Applications examples |
|---|---|
| **Infotainment** | Cooperative Local Services, Global Internet Services, Point Of Interest Notification, Local Electronics Commerce, Insurance Services, Fleet Management, Multimedia Downloading |
| **Active Road Safety** | Emergency Vehicle Warning, Lane Change Warning, Wrong Way Warning, Emergency Break Warning, Pre-Crash Warning, Collision Warning, Traffic Condition Warning, Signal Violation Warning |
| **Traveler Assistance** | Post-Crash Notifications, Pedestrian Crossing, Parking Spot Locater |
| **Public Services** | Stolen Vehicle Tracker, Traffic Flow Control, Approaching Emergency Vehicle Warning |
| **Driving Alert** | Curve Speed Alert, Lane Change Alert, Lane Merging Alert, Emergency Break Alert. Overtaking Vehicle Warning |
| **Road Condition Alert** | Vehicle Based Alert, Infrastructure Based Alert, Blind Spot Alert |
| **Collision Alert** | Wrong Way Driver Alert, Intersection Collision Alert, and Traffic Signal Violation Alert |
| **Urban Sensing** | Data Collection, Security in data communication, Photographs, Road Conditions |
| **Traffic Efficiency And Management** | Real Time and Reliability, Traffic Flow, Road Conditions, Danger on the Road. Speed Limit Notification, Synchronized Platooning, Adaptive Cruise Control |
| **Comfort** | Real Time, Free Parking Space, Music, Videos, Resting Places |

# Scenario

| Attacks | Types | Communication model in VANETS |
|---|---|---|
| **Wireless Interface** | Location Tracking, DoS, DDoS, Sybil , Malware and spam, Tunneling, Black hole, Grey hole, MiM, Brute force | V2V |
| **Hardware and Software** | DoS, Spoofing and forgery, Cheating with position information (GPS spoofing), Message suppression/alteration/fabrication, Replay, Masquerade, Malware and spam, MiM, Brute force | V2V, V2I |
| **Hardware and Software** | Sybil, Injection of erroneous messages (bogus info), Tampering hardware, Routing, Black hole, wormhole and Grey hole, Timing | V2V |
| **Sensors input in vehicle** | Cheating with position info(GPS spoofing), Illusion attack, Jamming attack | V2V |
| **Infrastructure** | Session hijacking, DoS, DDoS, Unauthorized access, Tampering hardware, Repudiation, Spoofing, impersonation or masquerade | V2I and V2V |

- The aim of Intrusion Detection Systems is detecting attacks in VANETs
- IDS= Intrusion Detection System

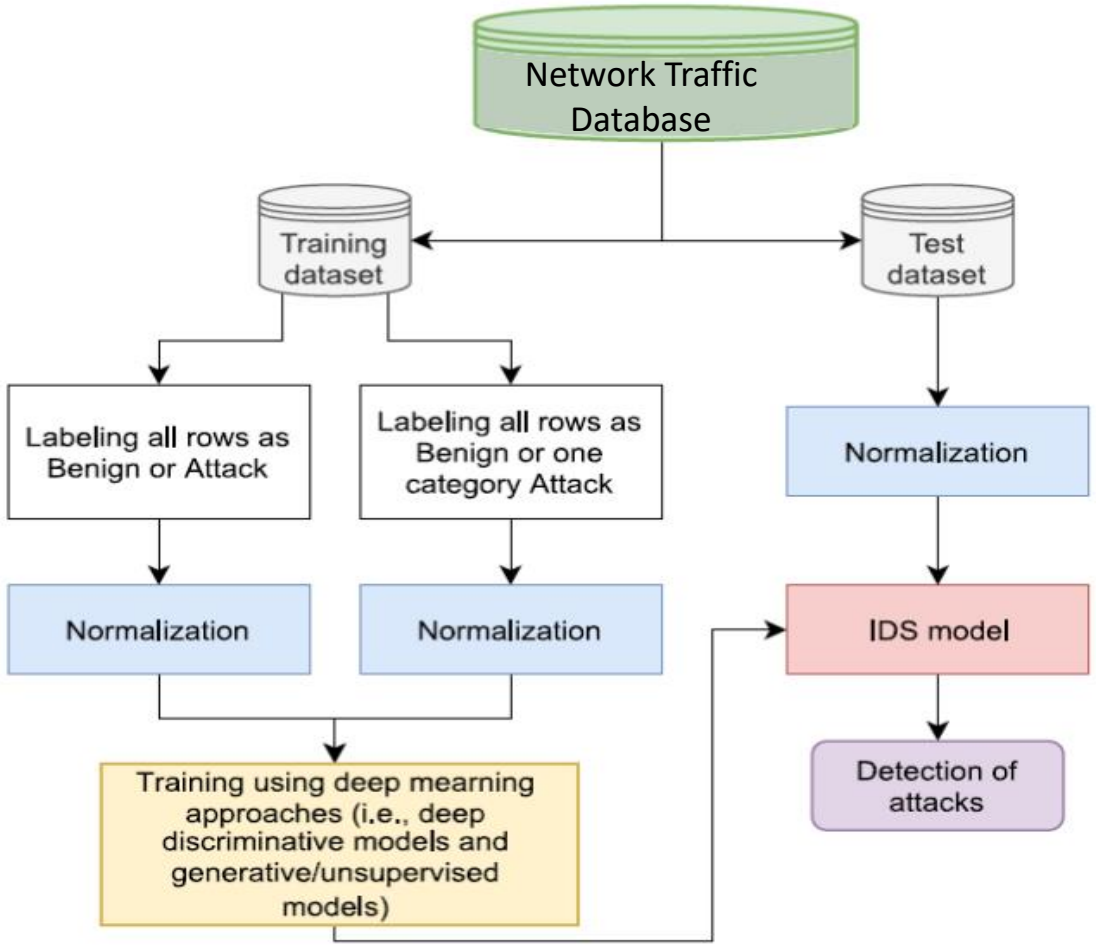
# What we have and what we want to do




- The aim of Intrusion Detection Systems is detecting attacks in VANETs
- IDS= Intrusion Detection System

# What we have and what we want to do

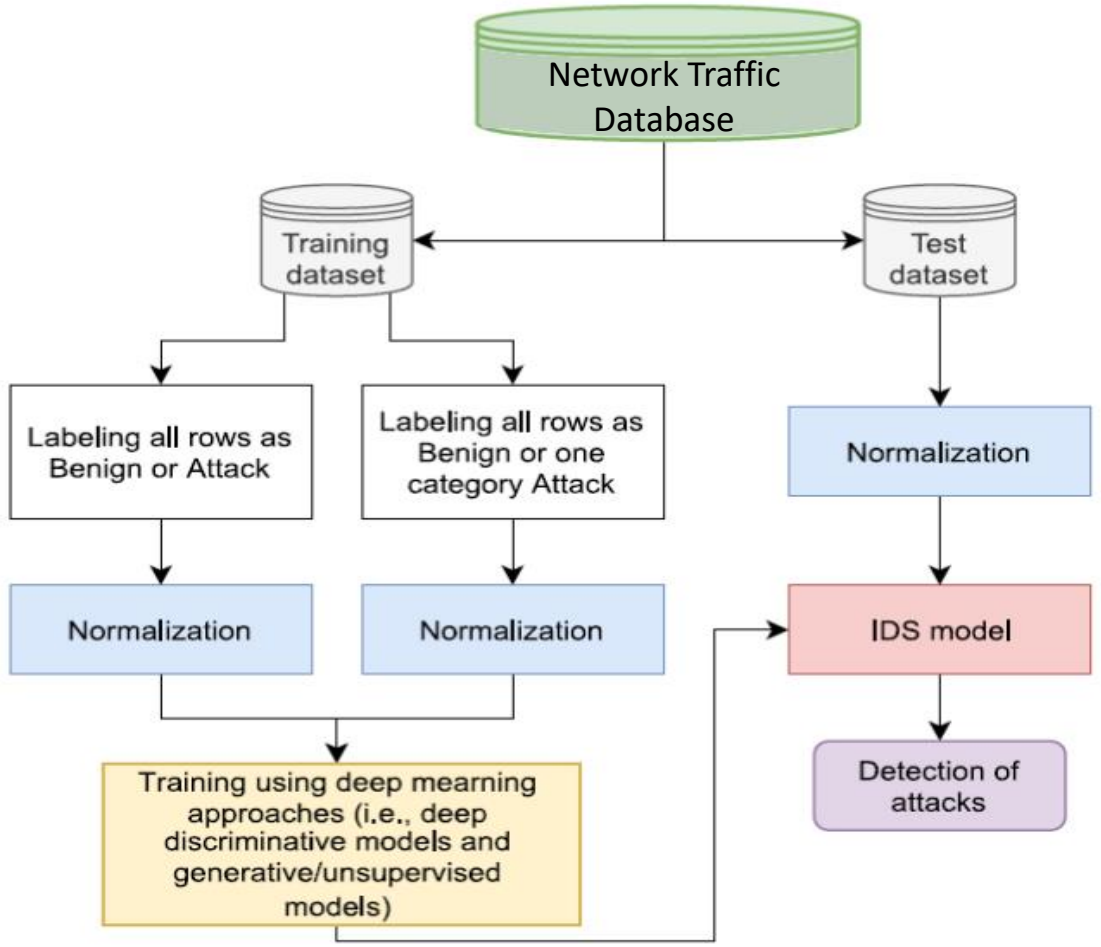
- A mount of vehicular data will be generated
  - Huge size,
  - Complicated structure,
  - Continuously generated,
- Large-scale datasets tend to provide rich knowledge to the learning process at the cost of very high computing time.
- Traditional machine learning algorithms cannot handle modern systems that require parallel real-time computations of infinite distributed streams.



# What we have and what we want to do

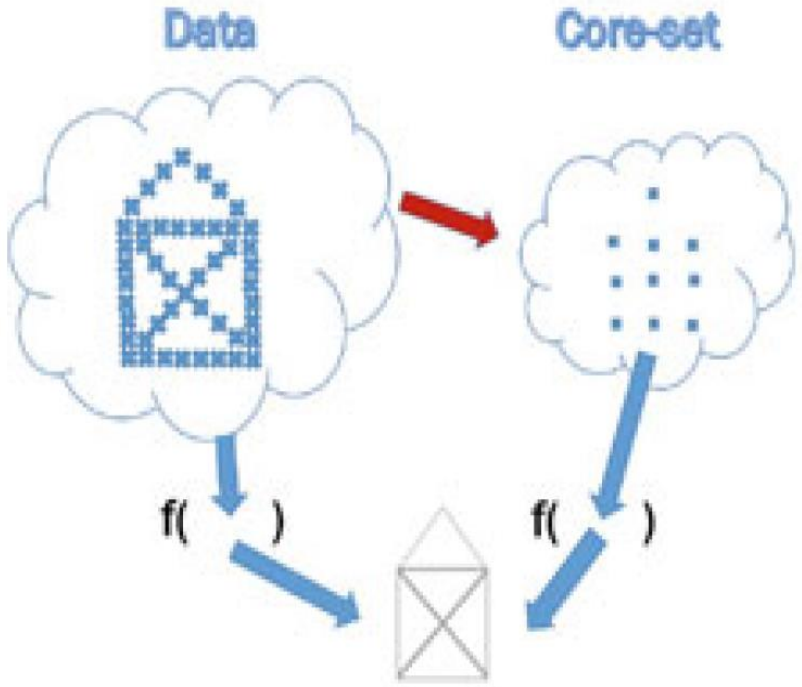
**Solutions:**

- Design new Lightweight algorithms with better running times.
- Use the same algorithm over a reduced version of the input data **=>** the same algorithm may complete faster.

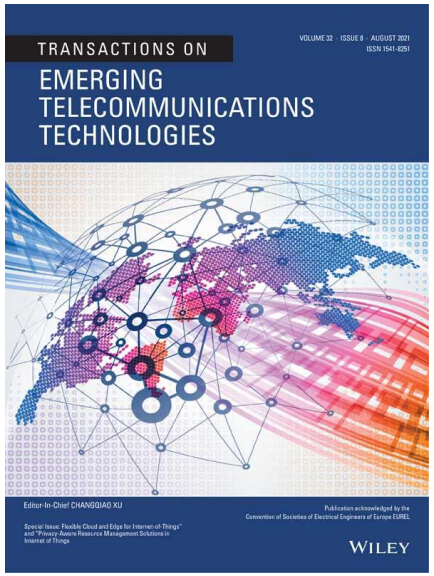


# Aim of the research

- Coreset is a “small data” summarization of the input “big data,” where every possible query has approximately the same answer on both data sets.
  - Coreset concept [Agarwal et al., 2004]
- Generic techniques enable efficient coreset maintenance of streaming, distributed, and dynamic data.
- Traditional algorithms can be applied on coresets to maintain the approximated optimal solutions.

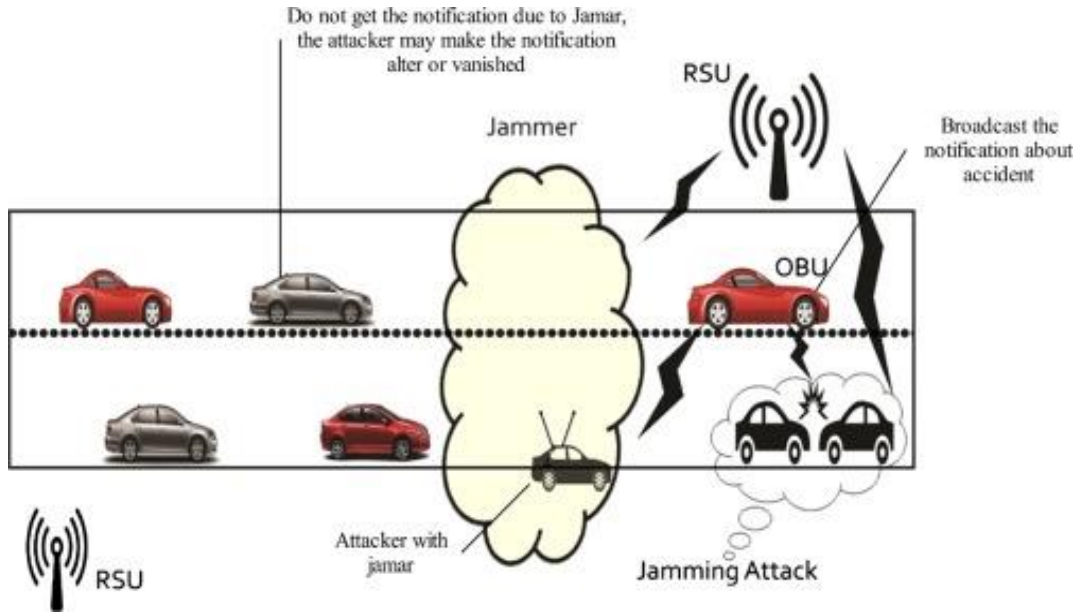


# Publications



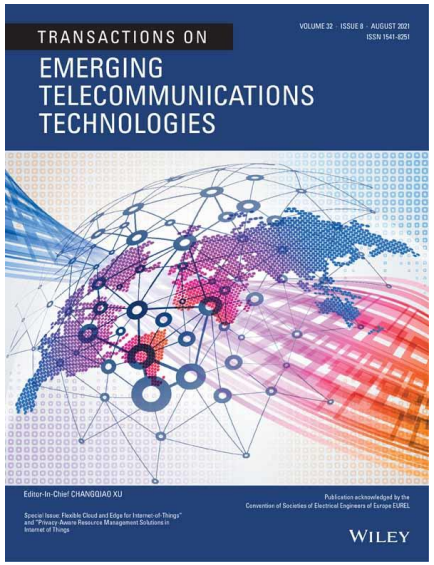

**Towards faster big data analytics for anti-jamming applications in vehicular ad-hoc network**

**Authors:** Hind Bangui, Mouzhi Ge, Barbora Buhnova, Hong Trang
**Date of Publication:** 2021
**Journal of Transactions on Emerging Telecommunications Technologies, IF:** 2.638




Jamming attack is defined as radio signal emission that aims to disturb the transceiver operation.

# Publications



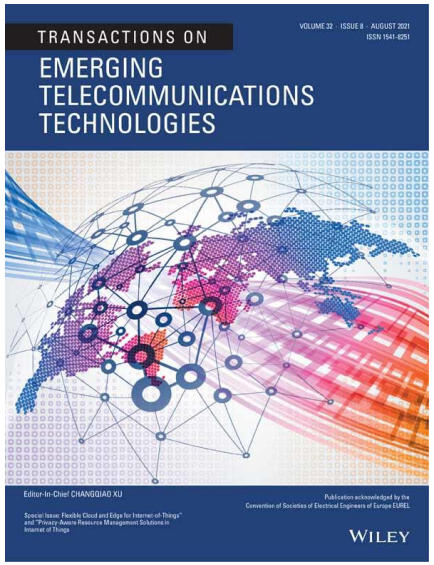

**Towards faster big data analytics for anti-jamming applications in vehicular ad-hoc network**

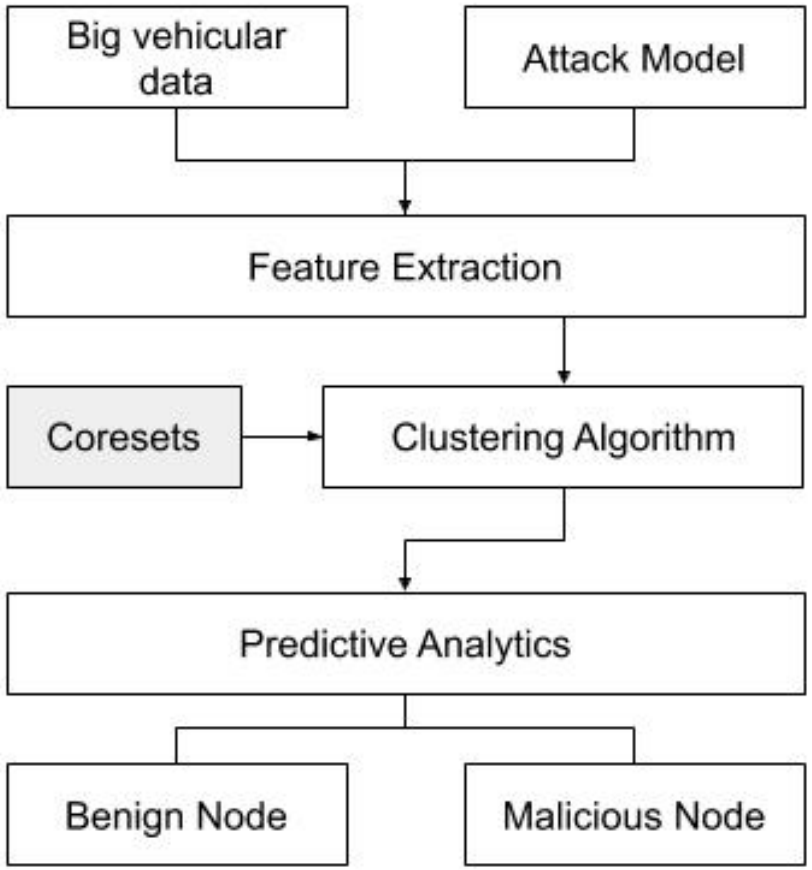
**Authors:** Hind Bangui, Mouzhi Ge, Barbora Buhnova, Hong Trang
**Date of Publication:** 2021
**Journal of Transactions on Emerging Telecommunications Technologies, IF:** 2.638

| Category | Type | Description |
|---|---|---|
| Behavior | Constant | Constant jammers transmit radio signal continuously to prevent legitimate nodes from accessing communication channels. |
| | Random | Random jammer change state between jamming and sleeping in random interval to maximize the jamming period and conserve energy. |
| | Reactive | Reactive jammer block or interfere with communication only upon detection of a certain level of energy. |
| Mobility | Stationary | Jammers alternate between on and off mode and affect the same area. |
| | Target mobility | Jammers target at a specific node. |
| | Random mobility | Jammers move around in the network while emitting signals randomly. |

**Jamming Attacks**

# Publications



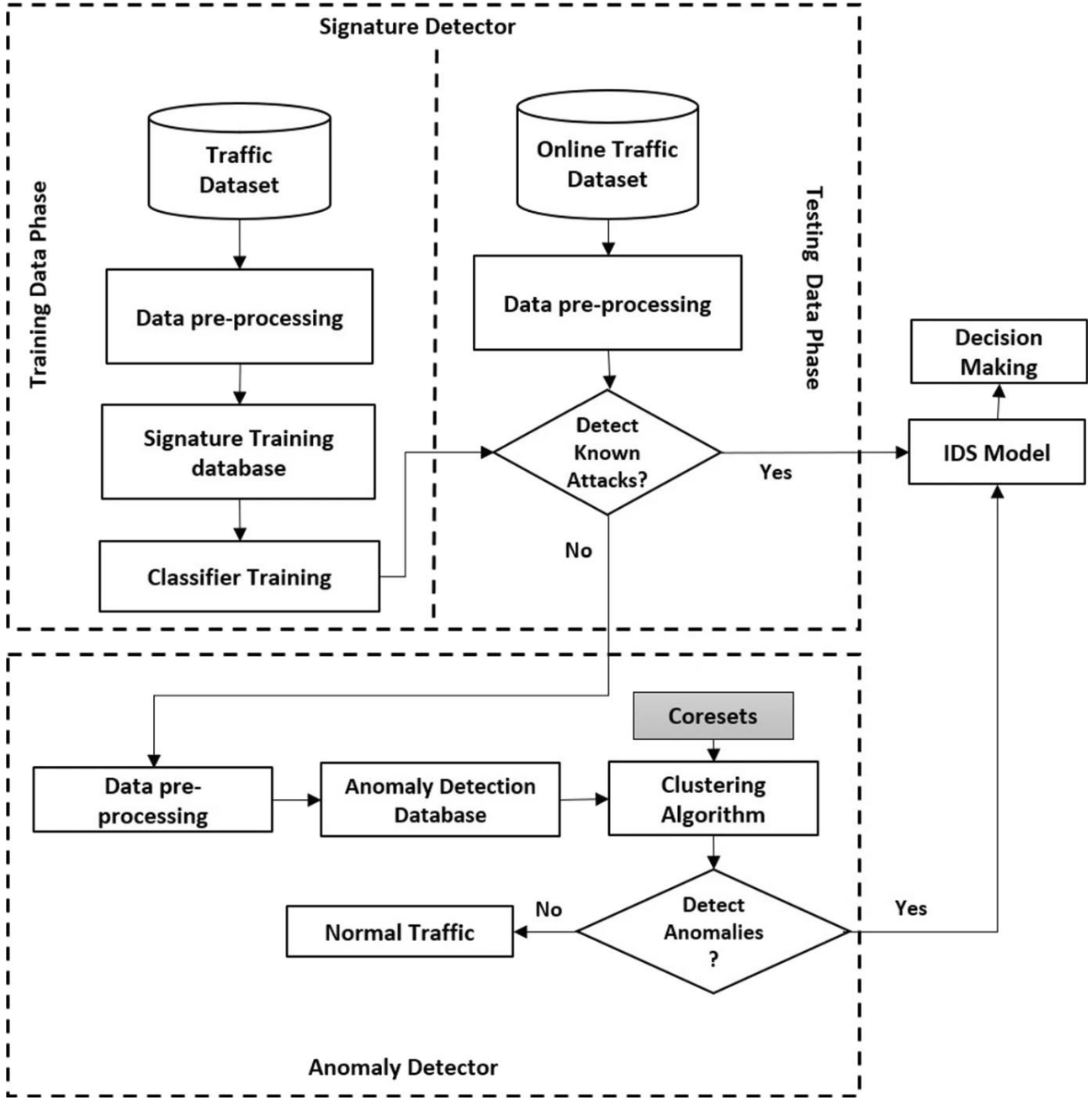

**Towards faster big data analytics for anti-jamming applications in vehicular ad-hoc network**

**Authors:** Hind Bangui, Mouzhi Ge, Barbora Buhnova, Hong Trang
**Date of Publication:** 2021
**Journal of Transactions on Emerging Telecommunications Technologies, IF:** 2.638

# Publications


**A hybrid machine learning model for intrusion detection in VANET**

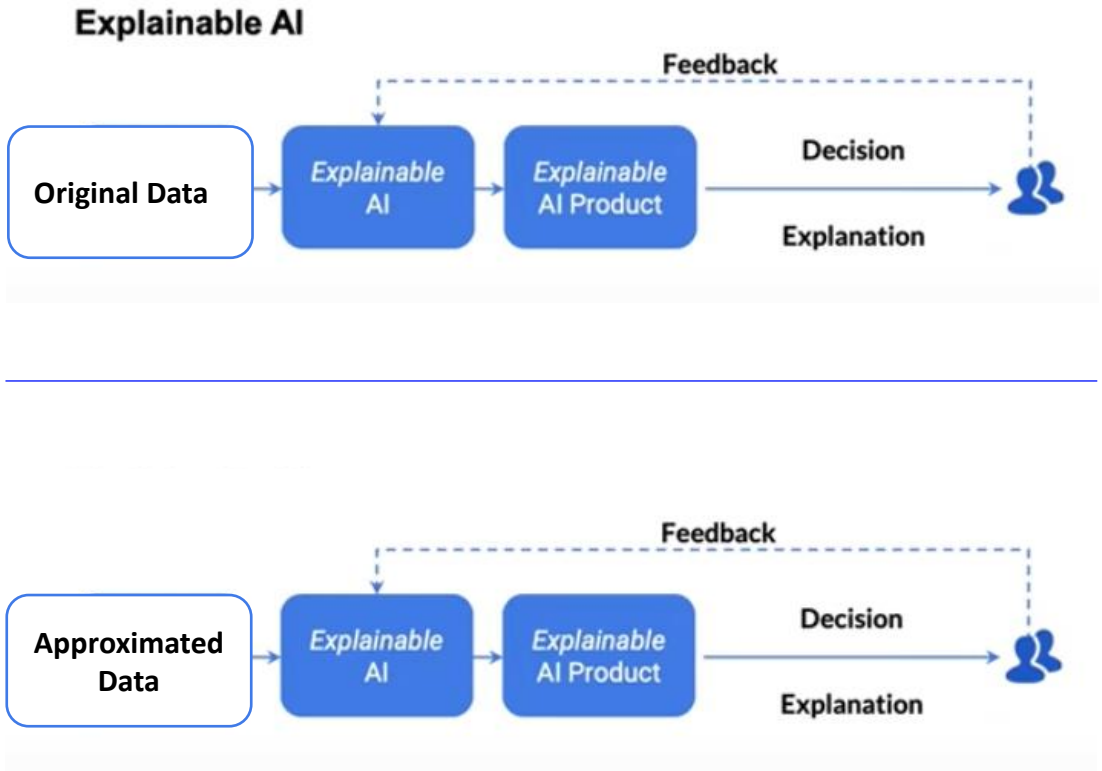
**Authors:** Hind Bangui, Mouzhi Ge, Barbora Buhnova
**Date of Publication:** 2021
**Computing Journal, IF:** 2.495

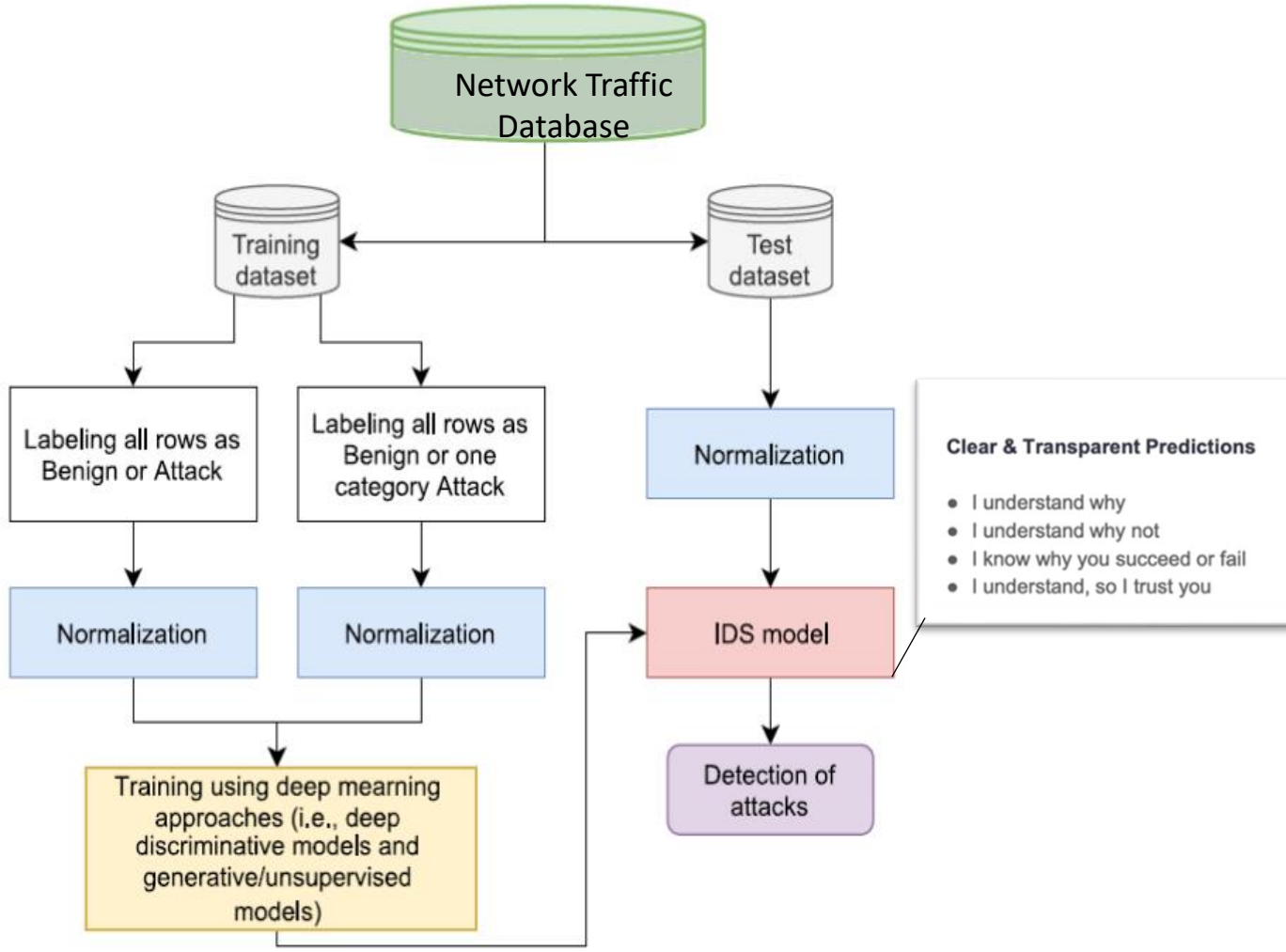


# Publications

- **“Recent Advances in Machine-Learning Driven Intrusion Detection in Transportation: Survey."**
  - Bangui, Hind, and Barbora Buhnova. *Procedia Computer Science* 184 (**2021**): 877-886.
- **"A Hybrid Data-driven Model for Intrusion Detection in VANET."**
  - Bangui, Hind, Mouzhi Ge, and Barbora Buhnova. *Procedia Computer Science* 184 (**2021**): 516-523.
- **"Improving big data clustering for jamming detection in smart mobility."**
  - Bangui, Hind, Mouzhi Ge, and Barbora Buhnova. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pp. 78-91. Springer, Cham, **2020**.
- "**Scaling big data applications in smart city with coresets."**
  - *Trang, Le Hong, Hind Bangui, Mouzhi Ge, and Barbora Bühnová. (**2019**).*

# Research Roadmap



# Thank you for your attention!