# A Comparative study on anomaly detection techniques for smart city wireless sensor networks

- Software documentation -

# Table of contents

# Description

This software is part of the work in the article "A comparative study on anomaly detection techniques for smart city wireless sensor networks", which aims at evaluating different techniques for detecting anomalies in wireless sensor networks (WSN) in the context of the smart cities.

The software has two main components. On the one hand, "Smart City Sensor" is an application module for the WSN simulator Castalia which allows the use of actual sensor readings in the simulations. On the other hand, this software contains the R source code that compares different anomaly detection techniques using data extracted from the simulation with Castalia using "Smart City Sensor".

# Smart City Sensor

“Smart City Sensor” está incluído en la carpeta “Castalia”. Este es un módulo de aplicación para el simulador de WSN Castalia en el que se le proporcionan los datos recogidos por sensores reales y el momento de envío de esos datos y el simulador reproduce el escenario concreto. A continuación describimos el contenido de esta carpeta y el uso de los componentes:

- ApplicationPacket.msg, ApplicationPacket_m.cc and ApplicationPacket_m.h: source code to subsitute this code in the Castalia directory Castalia-3.3/src/node/application
- smartCitySensor: application module to send the values of a from each node instead of simulated values. This folder must be included in Castalia-3.3/src/node/application.

Once these files have been copied to the Castalia directory, the simulator must be recompiled as it is indicated in the Castalia documentation.

# Anomaly analysis with R

The comparative among anomaly detection techniques is performed with R. The source code and the dataset for this analysis is included in the folder R. This folder contains:

- anomaly_analysis_smart_city_wsn.Rmd R code that generates reports to analyze different algorithms to detect outliers. The begining of this document contains the parameters that lets the user choose the algorithm to compare. This requires the installation of wsnanomalies_0.1.tar.gz package.
- reports: folder with the reports generated with anomaly_analysis_smart_city_wsn.Rmd
- wsnanomalies_0.1.tar.gz: R package with the main functions for the analysis.
- processed_smart_city_wsn_dataset.csv: dataset to train and test the detection algorithms. Each row of the dataset corresponds to a 30 minute samples generated by the Castalia Simulator using data from a real WSN implementation. The simulation was executed 8 times: without attacks, with 4 selective forwarding attacks (30%, 50%, 70% and 90% dropping rates) and 3 jamming attacks (near the base station, near node 4 and near node 9). Each sample has a label that indicates with which of the 8 types of simulations it was executed. The columns of the dataset are: the label, type of mac protocol, interval number, hour of the day, number of received application packets in the base station from each node, number of sent application packets from each node, consumed energy for each node, remaining energy in each node, TX packets from each node, MAC Beacons at each node, MAC ACK at each node, MAC RTS at each node, MAC SYNC at each node, radio TX packets failed with no interference from each node, radio TX packets failed with interference from each node, radio TX packets failed below sensitivity from each node, radio TX packets failed because non RX state from each node, radio TX packets received despite interferences from each node, radio TX packets received without interferences from each node, for each hour of the day a boolean value indicating if the sample was taken in that hour, a boolean value indicating if the WSN was configured with tmac, a boolean value indicating if the WSN was configured with 802.15.4, number of lost application packets from each node