Use Cases

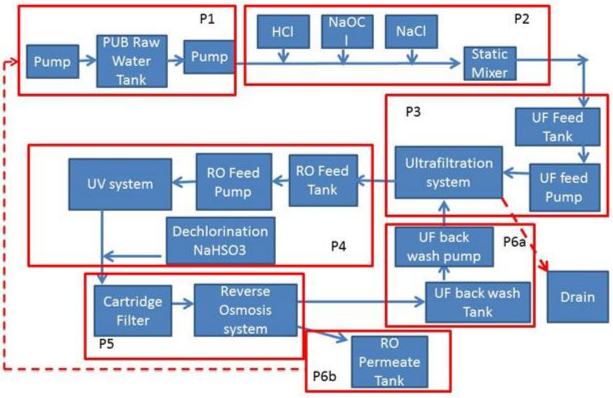
1.- Water quality sensors

1.1.- Use case description

In water treatment plants, it is possible to find several sensors measuring water levels, water quality, valve status, flow rates, etc. Data measured by these IoT sensors can be related in complex, nonlinear ways. For instance, opening a valve might result in changes in pressure and flow rate. We use two sensor datasets based on water treatment physical test-bed systems: SWaT and WADI, where operators have simulated attack scenarios of real-world water treatment plants, recording these as the ground truth anomalies. Here we present results for the Secure Water Treatment (SWaT) dataset that comes from a water treatment test-bed site for cybersecurity research. SWaT was launched in 2015 by the iTrust Centre for Research in Cyber Security of the Singapore University of Technology and Design. It represents a small-scale version of a realistic modern CyberPhysical system, integrating digital and physical elements to control and monitor system behaviors. Such systems are increasingly used in critical areas, including power plants and Internet of Things (IoT), which need to be guarded against potential attacks from malicious attackers.

SWaT is a scaled-down replica of a full-fledged water treatment plant. It represents an heterogeneous scenario where several types of sensors coexist. The testbed simulates a modern water treatment facility consisting of six processes:

P1: Raw water intake
P2: Chemical disinfection
P3: Ultrafiltration
P4: Dechlorination using ultraviolet lamps
P5: Purification by reverse osmosis
P6: Ultrafiltration membrane backwash and cleaning



The dataset contains seven days of data from normal operations, which are used as training data for the model (train.csv). A number of controlled, physical attacks are conducted at different intervals in the following four days, which correspond to the anomalies in the test set (test.csv).

MORE INFO: https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/

1.2.- Ontology

Nº of sensors: 50
SWaT consists of an array of monitoring sensors to ensure its safe operations. These are:

- AITx0y: Analyser Indicator Transmitter
  - o Conductivity (µS/cm)
  - o pH
  - o Oxidation Reduction Potential (mV)
- DPITTx0y: Differential Pressure Indicator Transmitter (kPa)
- FITx0y: Flow Indication Transmitter (m3/hr)
- LITx0y: Level Indication Transmitter (measured in mm)
- MVx0y: Motorised Valve (open/close)
- Px0y: Pump (Start/Stop)

Detailed list of sensors with corresponding name in list.txt

Extract of ontology definition of the model.

:ados-0 a aom:configuration;
aom : model : water-gnn ;
aom : modelVersion : 0.1 ;
aom : location :
ipfs://bafybeiemxf5abjwjbikoz4mc3a3dla6ual3jsgpdr4cjr3oz3evfyavhwq/
aom : idNodes [LIT101, MV101, P101, P102, AIT201, AIT202, AIT203, FIT201, MV201, P201, P202, P203, P204, P205, P206, DPIT301, FIT301, LIT301, MV301, MV302, MV303, MV304, P301, P302, AIT401, AIT402, FIT401, LIT401, P401, P402, P403, P404, UV401, AIT501, AIT502, ……., P603] .
aom : iExecMaxNumWorkers : 10 .

:wt-63672 a ssn : System ;
ssn : hasSubsystem [
:LIT101 a ssn : Sensor, fiemser : CommDevice ;
Fiemser:uses :MQTT
ssn : observes : Flow
] .
dul : hasLocation [
a geo : Point ;

geo : latitude “59.956632” ;
geo : longitude “30.663322”
] .
:MQTT a fiemser : NetProtocol ;
fiemser : hasName “MQTT” ;
fiemser : hasVersion “1.0” .

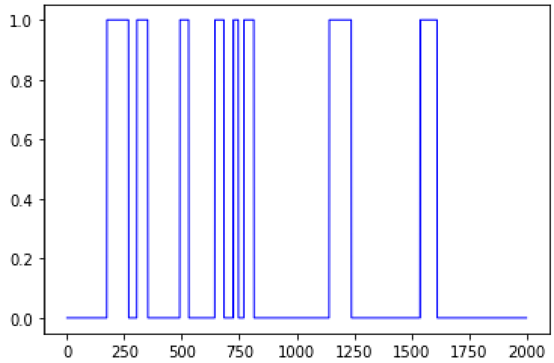
:obs-0-resultvalue a ssn : ObservationValue ,
qud: QuantityValue ;
qud : numericValue “mm”^^xsd:float ;

1.3.- Validation

The validation process for this use case was based on a dataset that consists of two assets, the training and validation train.csv file, and a test.csv file used for model evaluation. The testing test.csv file had already introduced a set of attacks on a set of sensors. The following graph shows the timestamps in which attacks have occurred, being these with anomalous values (constant values) within the normalized range of the sensors [0.1].

On this dataset an evaluation has been performed on the parameters that provided the best results. The variation is not large among them, although the best values are shown below.

Translated with www.DeepL.com/Translator (free version)
training and validation: train.csv 47521 data x 50 sensors
testing: test.csv 44990 data x 50 sensors



SWAT attack data

Model: best_04 28-07 30 48.pt
F1 score: 0.7762419006479481
precision: 0.9540748606318025
recall: 0.6542872747132714
BATCH_SIZE=32, SLIDE_WIN=5, DIM=64, SLIDE_STRIDE=1,
OUT_LAYER_INTER_DIM=64, TOPK=5, EPOCH=10, SEED=5

Model: best_04 28-07 55 14.pt
F1 score: 0.8193926478423015
precision: 0.9879239465570401
recall: 0.6999817950118332
BATCH_SIZE=32, SLIDE_WIN=5, DIM=64, SLIDE_STRIDE=1,
OUT_LAYER_INTER_DIM=128, TOPK=5, EPOCH=15, SEED=5

Model: best_04 29-10 54 16.pt
F1 score: 0.7876698438450618
precision: 0.8889906157015335
recall: 0.7070817403968688
BATCH_SIZE=64, SLIDE_WIN=5, DIM=64, SLIDE_STRIDE=1,
OUT_LAYER_INTER_DIM=64, TOPK=5, EPOCH=20, SEED=5

2.- Noise pollution

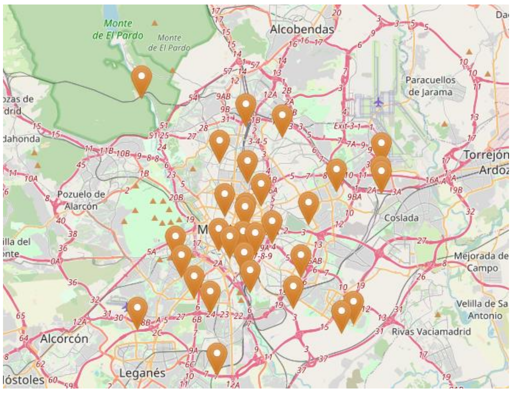
2.1.- Use case description

Acoustic sensors: in smart cities, acoustic changes can reveal changes in car traffic, weather conditions, etc.. Acoustic sensors are used to monitor the sound pressure level of the cities, related to noise pollution that affects the quality of life of people living there. As part of our use cases, we will test our architecture with a dataset collected in Madrid (Spain).

In order to control environmental noise levels, the Department of Noise Control of the General Directorate of Sustainability and Environmental Control manages the fixed noise pollution control network, consisting of a set of remote stations capable of capturing information on the acoustic conditions of their surroundings, which is transmitted to a central station where the data acquired is processed.

This network, which began operating in 1994 with 6 stations, currently has a total of 37 stations that are continuously measuring environmental noise levels 24 hours a day, every day of the year. In this use case, every IoT device are of the same type (sonometer), so we are now in an homogeneous scenario.

As it is shown in Figure X, the acoustic nodes are evenly distributed throughout the city, but the city center concentrates the largest number of nodes. Each node captures sound pressure of its location in a continuous mode, 24 h/7 days a week, using an equivalent sound pressure level in dBA during 1 minute.



The data-set used in this use case provides a long-term analysis, using 30 days of data from normal operations (01/09/2019-30/09/2019). From these 30 days, we used 24 days (80%) as training data for the model (train.csv). For obtaining a test file (test.csv) we used a period of 6 days (20% of the total dataset). Then, we implemented a code to simulate puntual anomalies, i.e. malfunctions of devices at different intervals .

[MORE INFO]

2.2.- Ontology

Nº of sensors: 37
Wireless acoustic sensor network in Madrid consists of a set of remote sonometers that continually monitor the sound pressure level (SPL) in decibels A (dBA).

Detailed list of sensors with corresponding name in list.txt

Extract of ontology definition of the model.

:ados-0 a aom:configuration;
aom : model : noise-gnn ;
aom : modelVersion : 0.3 ;
aom : location :
ipfs://bafybeiemxf5abjwjbikoz4mc3a3dla6uwetertdgsgdsgsdfgsdfgdsfget/
aom : idNodes
[A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15,A16,A17,A18,A19,A20,A21,A22,
A23,A24,A25,A26,A27,A28,A29,A30,A31,A32,A33,A34,A35,A36,A37] .
aom : iExecMaxNumWorkers : 5.

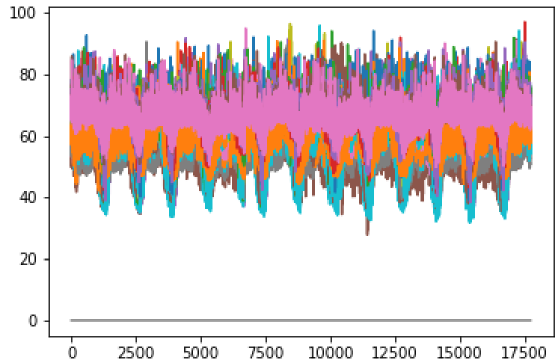
:wt-63345 a ssn : System ;
ssn : hasSubsystem [
:A1 a ssn : Sensor, fiemser : CommDevice ;
Fiemser:uses :MQTT
ssn : observes : SPL
] .
dul : hasLocation [
a geo : Point ;
geo : latitude “59.956632” ;
geo : longitude “30.663322”
] .
:MQTT a fiemser : NetProtocol ;
fiemser : hasName “MQTT” ;
fiemser : hasVersion “1.0” .

:obs-0-resultvalue a ssn : ObservationValue ,
qud: QuantityValue ;
qud : numericValue “dB”^^xsd:float ;

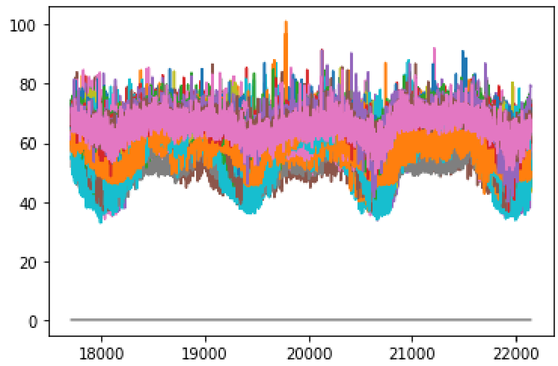
2.3.- Validation

In the case of the sound sensors, since all the sensors are of the same type, no normalization has been applied, i.e., the values are measured in dB. However, as some of these sensors had missing data, to avoid distorting the training, the times with sensors without NaN values have been removed from the training file.

training and validation: train.csv 17712 data x 37 sensors



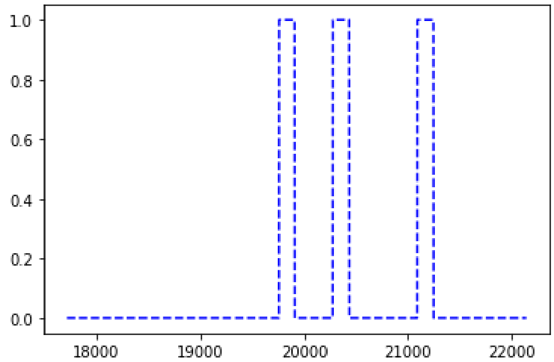
testing: test.csv 4428 data x 37 sensor

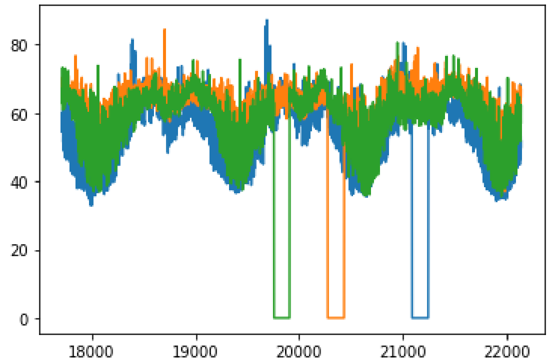


A series of tests have been carried out to simulate sensor failures following the usual behavior of these sensors when failures occur as a result of previous data analysis.

Specifically, a set of faults consisting in the sensor showing a constant measurement value over an instant of time is fixed on the testing dataset. In this case, we propose a failure duration of 150+D time intervals where D is a uniform random variable between 0 and 20.

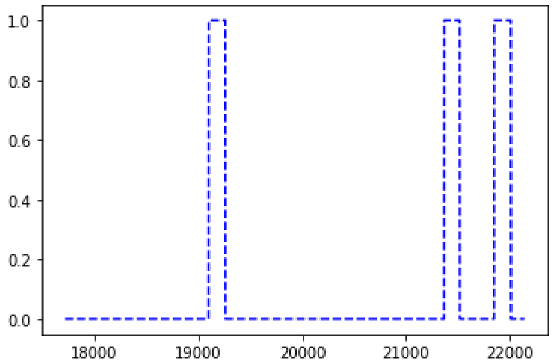
Depending on the value set for the attacks, the validation results change. In case the sensor shows a value of 0 dB, the following is obtained.

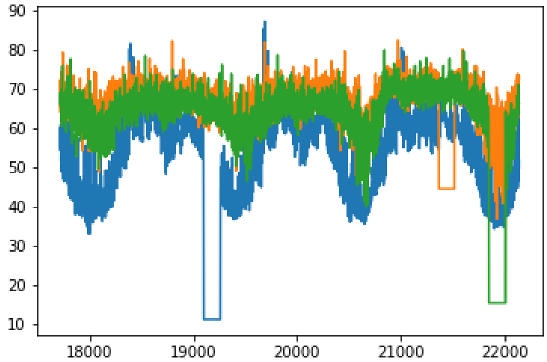




Model:	best_06|23-09:52:48.pt
F1 score:	0.9892933618843683
precision:	0.9850746268656716
recall:	0.9935483870967742

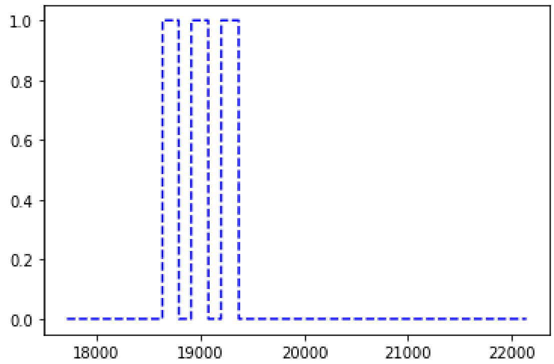
We then set some values for sensor faults, as a constant, but random value within the range of the sensors. Obtaining these results:

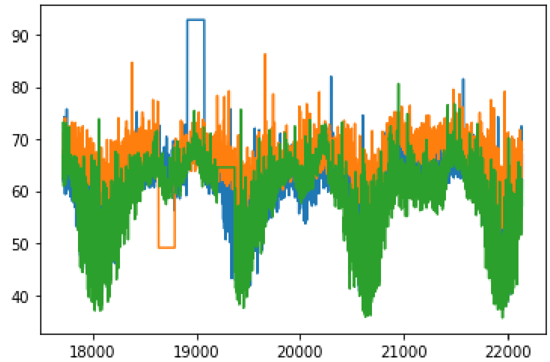




Model:	best_06|23-10:08:22.pt
F1 score:	0.8568075117370892
precision:	0.9656084656084656
recall:	0.770042194092827

Model:	best_06|23-10:10:39.pt
F1 score:	0.9842271293375394
precision:	0.9811320754716981
recall:	0.9873417721518988





Model:	best_06|23-10:16:49.pt
F1 score:	0.7852760736196319
precision:	0.9846153846153847
recall:	0.6530612244897959

Model:	best_06|23-10:19:06.pt
F1 score:	0.7726708074534161
precision:	0.9873015873015873
recall:	0.6346938775510204

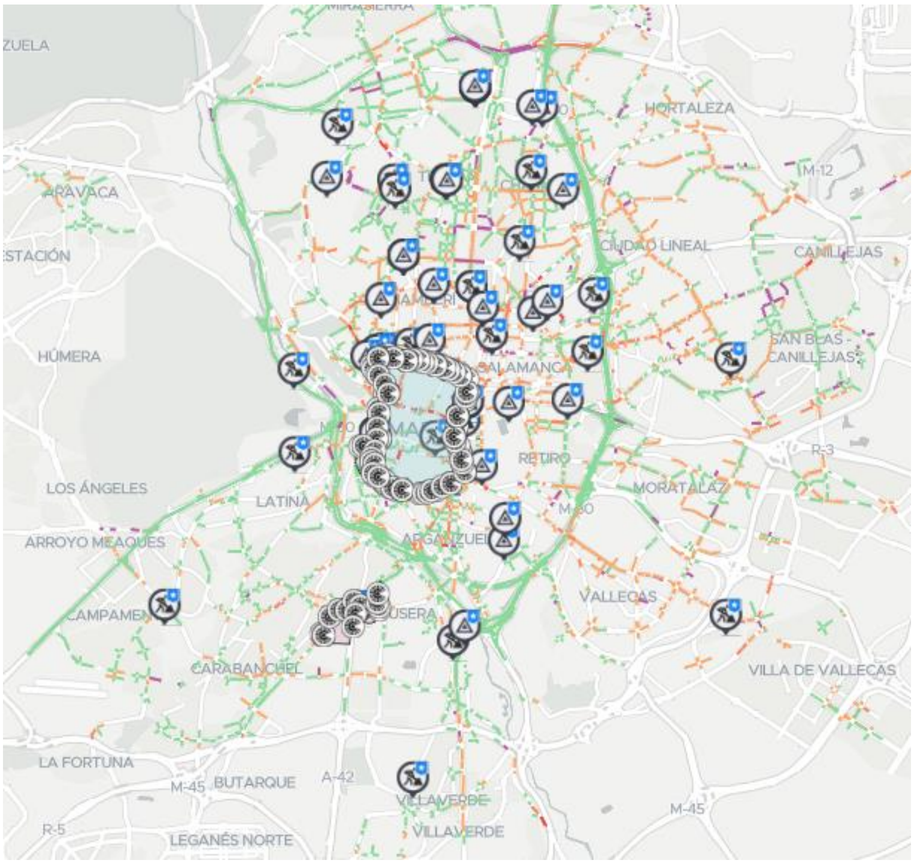
3.- Traffic

3.1.- Use case description

Transportation sector: especially regarding the upcoming industry of autonomous vehicles, our plan is to understand the existing dynamics arising in scenarios of car traffic in order to understand how vehicular mobility of several vehicles can affect each other. Understanding the dynamics of vehicles is a major issue in smart cities evolution. Impacts are preponderant in user guidance, fuel consumption, pollution, but also in instrumentation efforts to deploy the infrastructures of communication.

In this case, we will use databases describing the traffic of big cities, since these are the most challenging scenarios for road traffic. The database on which the system will be tested will be the traffic database of the city of Madrid, provided as opendata by the city council. This data is captured by vehicle detectors and more than 4000 measurement points, providing traffic intensity, load and average speed data in 5-minute periods since 2013. Traffic sensing is carried out by means of a variety of equipment that allows the counting of vehicles and the obtaining of the degree of road occupancy. These detection systems are mostly electromagnetic loops that are placed under the pavement and detect the metallic mass of the vehicles under the pavement and detect the metallic mass of the vehicles passing over them. These systems are of high quality and precision.

This information is updated almost in real time, with a periodicity of about 5 minutes, which is the minimum time of several traffic light cycles, necessary to give a real measurement, and that the measurement is not affected by whether the traffic light is open or closed.



https://informo.madrid.es/#/realtime?panel=live

In order to define this use case, only the traffic intensity is computed, so it is again an homogeneous scenario. However, the number of sensors are increased, using data from 125 sensors, to create a high volume data scenario in comparison to the others.

MORE INFO

3.2.- Ontology

Nº of sensors: 125
Fixed traffic sensor network in Madrid consists of a set of remote stations that continually monitor the intensity of the number of vehicles per hour. A negative value implies the absence of data.

Detailed list of sensors with corresponding name in list.txt

Extract of ontology definition of the model.

:ados-0 a aom:configuration;
aom : model : noise-gnn ;
aom : modelVersion : 0.5 ;
aom : location :
ipfs://whjtyjtdghfghfguhuk45625abjwjbikoz4mc3a3dla6tryrty34gdsgsdfgsdfgdsfget/
aom : idNodes
[1001,1002,1003,1006,1009,1010,1011,1012,1013,1014,.......,3472,3473,3474,3475,3476,3477,3478,3479,3480,3481] .
aom : iExecMaxNumWorkers : 15.

:wt-63345 a ssn : System ;
ssn : hasSubsystem [
:A1 a ssn : Sensor, fiemser : CommDevice ;
Fiemser:uses :MQTT
ssn : observes : intensity
] .
dul : hasLocation [
a geo : Point ;
geo : latitude “59.956632” ;
geo : longitude “30.663322”
] .
:MQTT a fiemser : NetProtocol ;
fiemser : hasName “MQTT” ;
fiemser : hasVersion “1.0” .

:obs-0-resultvalue a ssn : ObservationValue ,
qud: QuantityValue ;
qud : numericValue “veh/h”^^xsd:float ;

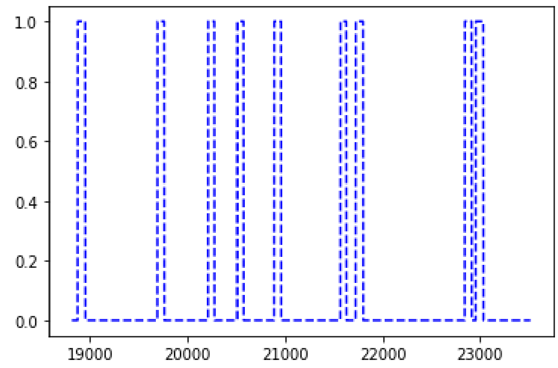
3.3.- Validation

The validation process for this use case was based on a dataset that provides traffic measurement data. A data file has been built by joining data from several months, specifically from 03-2021 to 10-2021, this dataset provides 1 data every 15 minutes.

These data have been normalized between [0,1] to fit the characteristics of the model. From this dataset, 80% of the time_stamps have been extracted and stored in the train.csv file used for training and validation, and 20% stored in the train.csv file used for validation.

To the train.csv file, attacks have been introduced according to the attached graph. During the time of the attack one of the sensors fails and provides a constant traffic value within the range of the data. This type of values has been observed in natural failures of these sensors in the indicated data set.

Translated with www.DeepL.com/Translator (free version)
training and validation: train.csv 18813 data x 126 sensors
testing: test.csv 4704 data x 126 sensors



Traffic attack data

Model: best_06|22-18:24:19.pt
F1 score: 0.027906976744186046
precision: 0.8181818181818182
recall: 0.014195583596214511

Model: best_06|22-18:27:56.pt
F1 score: 0.2112676056338028
precision: 0.9868421052631579
recall: 0.11829652996845426

Model: best_06|22-18:31:42.pt
F1 score: 0.2105263157894737
precision: 0.8636363636363636
recall: 0.11987381703470032