Faculty of Engineering — University of Porto

Railway Smart Meters

Vítor A. Morais

Thesis Research Plan

Doctoral Programme in Electrical and Computer Engineering

Supervisor: António P. Martins

Co-supervisor: José L. Afonso

July 26, 2017

Railway Smart Meters

Vítor A. Morais

Doctoral Programme in Electrical and Computer Engineering

July 26, 2017

Contents

1 Introduction

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

2 Objectives and Contributions

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

2.1 Objectives

This work is focused on measuring the energy flows in the railway system. The aim of this work is to improve the energy efficiency in the railway transportation system (RTS) and reduce the maintenance cost of RTS power systems. The implementation of smart meters (SM) in RTS promote a better overview of power flow and, based on the information of SM, algorithms focusing on energy efficiency can be implemented. The SM requires sensors such as voltage and current sensors. The level of intrusion as well as the level of electric «valor da grandeza electrica» of such sensors implies considerable costs of the sensors. Therefore, the implementation of complex processing on smart meters is of added value. This complex processing can be the implementation of fault monitoring algorithms in SM based on the energy measurements. Framed in the shif2rail, the work is focused on the implementation of a smart meter demonstrator for the RTS. To embrace the entire railway system, the power flow should consider the energy flux from and to the catenary. Therefore, the key point should be the measurement of the energy in the traction substations and in the train power transformer Based on this thesis proposal, the objectives are the following:

1. Research on high-voltage and high-current measurement systems
2. Research of train power transformer and implementation of a simulation model of a train power transformer.
3. Development and implementation of a measuring system with high acquisition and processing capabilities.
4. Research on communication systems and development of a network model in a simulation environment.
5. Research, development and implementation of a fault monitoring system.
6. Research, development and implementation of an energy flow monitoring system.
7. Implementation and validation of SM in a pilot project through real tests.

2.2 Contributions

1. Increase of energy flow information of RTS

2. Reduction of transmission costs of information (no need of LTE, the data are concentrated and transmitted from trains to stations, during passenger exchange, with a high throughput link)

3. Decrease of the Life Cycle Cost (LCC) of

3 Literature Review

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

3.1 Section 1: Smart metering

In this section, special attention is given to smart metering. The smart grid framework is presented to justify the need for smart meters and then, an overview on metering systems is presented.

1. Smart grids and the need for smart meters
2. Metering systems – overview
3. Metering systems in railways
4. Wireless sensor networks – overview
5. Smart metering with WSN in Railway Transportation System (RTS)

3.1.1 Smart grids and the need for smart meters

Smart Grids improves the functionality and concept of traditional electrical grids by obtaining the grid component's data using Information and Communication Technology (ICT). Such grids benefit the reliability and the efficiency of the system with the usage of the acquired data, Mohassel et al. (2014).

Although the smart meters does not have an effective definition, those devices are composed by an electronic box and a communication link, Seppo and Elektraflex (2012). A smart meter is responsible for measuring the energy-related parameters and the user consumption with a given time interval. All those measurements are then transmitted upon a communication network to the utility or to other player with the responsibility of using the meter data. The information obtained from the data is shared with consumer-side devices, to inform the end-users on their related costs and energy usage, Siano (2014).

Smart meters implement a bidirectional communication on top of AMR. They are inherent to smart grid systems.

Similarly to the evolution of electricity meters, the utility grid has evolved from a centralized production and control perspective to a distributed one. The conventional electrical grid is a network with a transmission link connecting power producers and end-user consumers. The control and distribution of electrical power is made in a centralized way. With the increase of power demand, increase of complexity and having more and more decentralized power generation, a migration to the smart grid framework is required, Reddy et al. (2014).

3.1.2 Features of Smart Meters and metering systems

A smart metering system combines several controlling devices, a extensive number of sensors for measuring the parameters and devices responsible for the transferring of the data and the commands. The detection of unauthorized consumption due to electrical energy theft and the improvement of the energy in the distribution are other advantages of smart meters. These devices acts as a gateway by having a communication interface protocol to the database stored by the utility company, Reddy et al. (2014).

The design of an ideal smart grid has to focus on prediction, adaptability and reliability points. Moreover, it requires to cover the demand adjustment, the load handling, flexibility and sustainability and it should incorporate advanced services. In advance, an end to end control capability has to be ensured as well as finding the optimal cost and asses, increase the quality of energy and quality of service. Another features of smart grids are the automatic restoration ans self-healing, being all the previously presented features of the smart grids highly dependent of the role of the smart meters, Mohassel et al. (2014).

Smart-meter types are also distinguished based on features like data-storage, communication type and connection with the energy supplier. The data storage capability allows data to be stored in the meter, being transferred after a few days or weeks to the Meter Data Management System (MDMS) of the utility. Compensations for some power quality deficiencies can be also considered; therefore the future meters should be also capable of register certain basic power quality characteristics. In advance, the design of rate and tariffs of electricity providers determine the requirements such as the period of meter intervals or the temporal resolution (commonly ranging from 15 min to 1 h). During those intervals, the production and consumption of active and reactive power is mandatory to be separately measured, Siano (2014).

3.1.3 Metering systems in railways

Towards na increase of interoperability of the rail system within the Community Council of European Union (2008), the Directive 2008/57/EC specifies the need of Technical Specifications for Interoperability (TSIs), presenting essential requirements in which each rail subsystem should meet to ensure the interoperability of the railway system within the EU. Those TSIs are of the responsibility of European Union Agency for Railways (ERA) and are listed as following:

- Locomotives and passenger rolling stock - 1302/2014;
- Noise - 1304/2014;
- Wagons - 321/2013;
- Infrastructure - 1299/2014/EU;
- Energy - 1301/2014;
- Control command and signalling - 2012/88/EU;

- Persons with reduced mobility - 1300/2014/EU;
- Safety in railway tunnels - 1303/2014;
- Operation and traffic management - 2015/995/EU;
- Telematics applications for freight service - 1305/2014/EU;
- Telematics applications for passenger service - 454/2011;

On the energy field and with the purpose of implementing on-ground energy data collecting systems (data collecting service - DCS), technical specifications for interoperability relating to the 'energy' subsystem of the rail system in the Union are specified in the Commission Regulation (EU) No 1301/2014, Council of European Union (2014b).

The On-Board energy measurement systems are pointed in Appendix D of Commission Regulation (EU) No 1302/2014 Council of European Union (2014a). This regulation appendix presents the requirements for such energy measurement system. The general architecture is defined as following:

- **Energy measurement function (EMF)**, measuring the voltage and current, calculating the energy and producing energy data;
- **Data handling system (DHS)**, producing compiled energy billing data sets for energy billing purposes, by merging data from the EMF with time data and geographical position, and storing it to be sent to on-ground data collection system (DCS) by a communication system;
- **On-board location function**, giving geographical position of the traction unit. Contrary to fixed installation revenue meters, train meters must have the knowledge of time and geographical position Santschi and Braun (2015);

Figure 3.1 presents the general overview of the on-board energy measurement system.

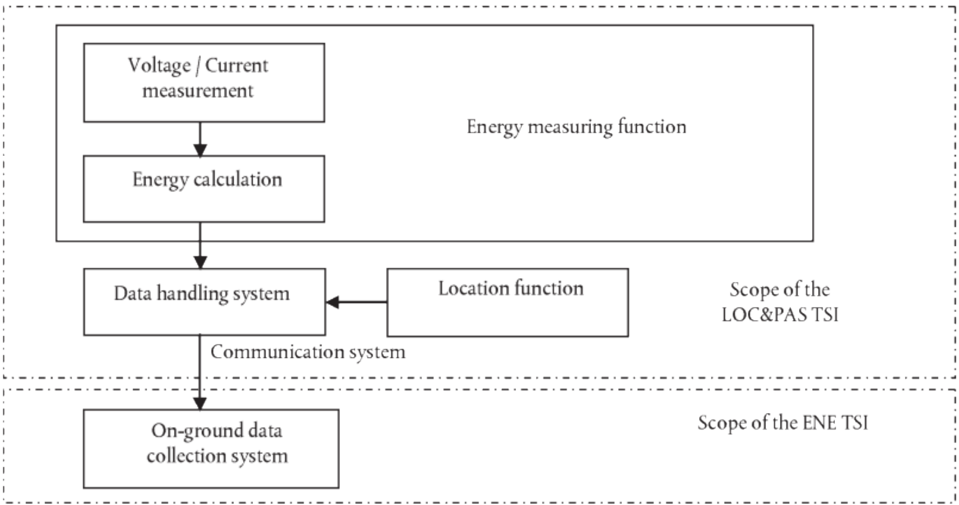

Figure 3.1: Functions, data flow and regulation scope of on-board energy measurement system.

As global requirements, all active and reactive energy should be measured, the measurement equipments should be rated to match the traction unit current and voltage rating, the system should be protected from intrusion, and finally, the loss of power in the measurement system should not affect the data stored in EMS.

Complementary to the previously presented system requirements, each component of the energy measurement system has specific requirements as listed as following:

Energy measurement function (EMF) Has specific metrological requirements to specify the accuracy of the sensors and it has the requirement of having the reference period of 5 minutes defined by the UTC clock time (shorter measuring period is allowed in the case that the data is aggregated into 5 minutes time reference period);

Data handling system (DHS) Should compile the data (without corrupting them), using the same time reference as EMF. This system should store compiled data of, at least, 60 days' continuous work and should have an alternative method of accessing the data. Finally, this system should produce compiled energy billing data sets (CEBD) by including an EMS identification number, a timestamp, a location data and the consumed/regenerated active and reactive energy.

Location function Has specific location requirements to define the latitude and longitude, as well as a accuracy of 250m and the location data information should have the same timestamp.

On-board to ground communication The specification related to interface protocols and transferred data format are an open point.

Santschi and Braun (2015) lists the standards that should be considered to the certification of Energy Measurement Systems on on-board trains:

EN 50463-1 General;
EN 50463-3 Data Handling System;
EN 50463-4 Communication;
EN 50463-5 Conformity Assessment.

3.1.4 Smart metering with WSN in Railway Transportation System (RTS)

3.2 Section 2: Wireless networks

In this section, communication networks are presented, with special attention to wireless technologies. An historic overview is presented

1. Network technologies – historic overview
2. Current technologies and standards:
3. Emerging technologies and standards
4. Network Simulators and Network Emulators

3.2.1 Network technologies – historic overview

3.2.2 Current Wireless technologies and standards

The following sections will cover the wireless communications in smart metering systems, starting with the low-rate and low-power communications applied on smart meters and ending with the high-rate communications (and consequently higher costs and power than low rate communications). With the increasing on demand for higher bandwidth, broadband technologies such as mobile WiMAX, IEEE 802.16e and broadband PLC are expected to be considered and used in newer installations, Mohassel et al. (2014).

3.2.2.1 IEEE 802.11 (Wireless LAN (WLAN) or Wi-Fi)

IEEE 802.11 is the standard for the information exchange between systems and for the telecommunications. The coverage area of this technology is on local and metropolitan area networks (LANs and MANs). The specific requirements are on the Medium Access Control and on Physical Layer. The most popular versions of this standard is the IEEE 802.11b and IEEE 802.11g, that differs in the modulation technique (Direct Sequence Spread Spectrum (DSSS) technique versus Orthogonal Frequency Division Multiplexing (OFDM) modulation technique). The data rates are, respectively, 11 Mbps and 54 Mbps, Usman and Shami (2013); IEEE (2016).

3.2.2.2 IEEE 802.15.4 (ZigBee)

The standard IEEE 802.15.4 imposes conditions in the physical layer and media access control focusing on low-rate (up to 300 kHz) wireless personal area networks. Developed by the Zigbee Alliance and covering the specifications of the IEEE 802.15.4 on the physical layer and the medium access control, Zigbee is a commonly used for low power wireless communication technology. It operates on the ISM bands of 868 MHz, 915 MHz and 2.4 GHz adopting direct sequence spread spectrum (DSSS), Usman and Shami (2013).

3.2.2.3 DASH7

On the low-rate field of research, an alternative to Zigbee is the DASH7. Using the ISO/IEC 18000-7 standard to support this wireless sensor network technology, DASH7 is developed to reach active Radio Frequency Identification Devices (RFIDs) and operates at 433MHz band. The advantage is the typical range of 250m (could achieve 5 km) and has a typical and maximum data rates of 28 kbps and 200 kbps, being in this specifically designed for Smart Grid and for applications in Smart Energy.

3.2.2.4 6LoWPAN

IEEE 802.15.3: IEEE 802.15.3 [46] is a physical and MAC layer standard for high data rate WPAN. It is designed to support real-time multi-media streaming of video and music. IEEE 802.15.3 operates on a 2.4 GHz radio and has data rates starting from 11 Mbps to 55 Mbps

3.2.2.5 Wibree

Wibree [47] is a wireless communication technology designed for low power consumption, short-range communication, and low cost devices. Wibree allows the communication between small battery-powered devices and Bluetooth devices. Small battery powered devices include watches, wireless keyboard, and sports sensors which connect to host devices such as personal computer or cellular phones. Wibree operates on 2.4 GHz and has a data rate of 1 Mbps. The linking distance between the devices is 5–10 m. Wibree is designed to work with Bluetooth. Bluetooth with Wibree makes the devices smaller and more energy-efficient. Bluetooth–Wibree utilizes the existing Bluetooth RF and enables ultra-low power consumption. Wibree was released publicly in October 2006.

3.2.2.6 Industrial Wireless Communications: WirelessHART and ISA100.11a

Launched by HART Communication Foundation in September of 2007, WirelessHART is an open wireless communication standard designated specifically for the process measurement and control applications, Song et al. (2008). This standard is specifically designed to comply with industrial requirements, such as stricter timing requirement, higher security concern, immunity to harsher interferences and obstacles and enough scalability to be used in large process control systems.

Similarly, ISA100.11a aims to provide secure and reliable wireless communication for non-critical monitoring and control applications, Petersen and Carlsen (2011).

3.2.2.7 IEEE 802.16 (WiMAX)

On the field of the broadband wireless communication there is the Worldwide Interoperability for Microwave Access (WiMAX) under the IEEE 802.16 standard. It is specifically developed aiming the point-to-multipoint communications being applied in fixed and mobile applications and it has data rates up to 70 Mbps over a distance of 50 km. Framed into the smart grid systems, this communication technology is considered as a solution for high data rate communication link to be applied at the backbone of the utilities, Usman and Shami (2013).

3.2.2.8 Broadband communications: GSM/GPRS and LTE/LTE-Advanced

Operating at 900 MHz and 1800 MHz, the Global System for Mobile communications (GSM) is the most used cellular network all over the world. The modulation technique is the Gaussian Minimum Shift Keying (GMSK) and it achieves transfer rates up to 270 kbps. Its architecture consists of four components: the Operation Support Substation, the Network Switching Substation, the Base Station Subsystem and the Mobile handset. Due to its level of development around the world being present in remote locations, this advantage makes this an interesting technology to be applied in Smart Grid applications, Usman and Shami (2013).

Long Term Evolution (LTE) is a recent standard for wireless technology that allows high data rates with high capacity and low latency and with a good Quality of Service (QoS). The improved version of this technology, the LTE-Advanced, admit higher capacity with expanded peak data rate of 1 Gbps for the downlink and 500 Mbps for the uplink, obtained on the increase of the spectral efficiency, higher number of active subscribers connected at the same time, and better performance at cell edges, Mohassel et al. (2014). This technology, for the Smart Metering environment where the high bandwidth and good QoS are mandatory at some communication points.

3.2.3 Protocols and standards

In computer networks, a protocol is a set of rules that ensure a communication of specific set of information between two machines. A standard is a document that specifies several aspects of something that has the overwhelming agreement and support of an entity (the standards making body). In the networking area, several protocols are supported by standards. In this section is presented some of the protocols that ensure a coherent communication among the sensor networks.

3.2.3.1 IEEE 802.15

The standard family defines the topologies and network roles. In particular, it defines the physical (frequency and channels, spectrum handling, modulation and bit rate) and MAC (packet formats, operational modes, timing aspects, topologies) layers of the OSI model (Hackmann (2006)).

Standard	Description	Initial Release / Revision Date	Amendments
IEEE 802.15.1 (Bluetooth)	MAC and PHY Layer Specifications for Wireless Personal Area Networks (WPANs)	2002 / 2005	Bluetooth Core Configuration v4.0 and Bluetooth Low Energy (2009)
IEEE 802.15.2	Coexistence of Wireless Personal Area Networks With Other Wireless Devices Operating in Unlicensed Frequency Bands	2003	In hibernation since 2011.
IEEE 802.15.3	MAC and PHY Layer Specifications for High Rate Wireless Personal Area Networks (HR-WPANs)	2003	802.15.3b (2006): Amendment to MAC Sublayer 802.15.3c (2009): Millimeter-wave-based Alternative Physical Layer Extension
IEEE 802.15.4	MAC and PHY Layer Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)	2003 /2006/ 2011	802.15.4.a (2007): PHY Layer Extension to Chirp Spectrum Techniques and UWB systems 802.15.4c (2009): Alternative PHY Extension to support one or more of the Chinese 314-316 MHz, 430-434 MHz, and 779-787 MHz bands 802.15.4d (2009): Alternative PHY Layer Extension to support the Japanese 950 MHz bands 802.15.4e (2012): Amendment 1: MAC sub-layer 802.15.4f (2012): Active Radio Frequency Identification (RFID) System PHY 802.15.4j (2013) – Alternative PHY Extension to support Medical Body Area Network (MBAN) services operating in the 2360-2400 MHz band
IEEE 802.15.5	Mesh Topology Capability in Wireless Personal Area Networks	2009	-
IEEE 802.15.6	Wireless Body Area Networks	2012	-

Figure 3.2: Members of the 802.15 family. Adapted from Panousopoulou (2014) lecture presentation slides.

Based on figure 3.2, the 802.15.4 is the standard that defines the PHY and MAC layers for low rate wireless personal area networks. It supports **full-function devices** (device capable of being the network coordinator or simple node and can have implemented complex network functionalities) and **reduced-function device** (limited devices with low-bandwidth limitations and limited or no-network intelligence). The possible network topologies are the following:

Star — Each device in the network communicates with the full-function device network coordinator;

Peer-to-peer — All devices communicate with each other (if they are in the communication range). Sufficiently flexible to implement more complex topologies such as multi-hoping, cluster trees and mesh topologies;

Multi-hopping — This is a technique that allows the usage of two or more wireless nodes to convey data from a source to a destination;

Cluster trees Topology to reduce the routing complexity where each node knows its parent node and all it's child nodes. It has always only one single path between two nodes.

Wireless mesh This technique allows data to be propagated along a path by hopping from node to node until it reaches its destination.

3.2.3.2 802.15.4-based wireless standards

Radmand et al. (2010) presents a comparison of wireless sensor standards for industrial applications. In figure 3.3 is present the overall schema of the wireless standards.

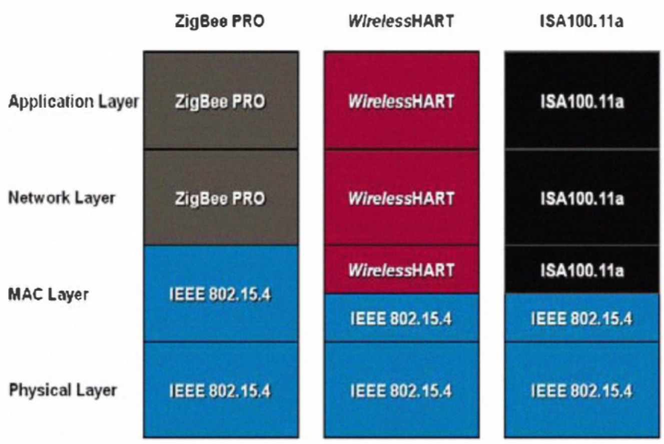

Figure 3.3: Overall schema of wireless standards. Adapted from Radmand et al. (2010).

3.2.4 Communication KPI's

According to Radmand et al. (2010), it is identified the following "Industrial Requirements" for the WSN's:

Reliability — Reliability is a measurement of the transmission accuracy, in percentage, that evaluates the amount of data that reach its destination. This measurement uses the properties of data communication, acknowledge-based usually.

Latency — Latency is the measurement of the time delay and is defined as the time that a data packet takes to be transmitted from the source to the destination. The latency is directly related to the link quality and a high latency link is result of a link with high signal-to-noise ratio.

Sensor Data Update Rates — This KPI is not directly related with the communication link. However the update rate of the sensor data affects the power consumption due to the increase of the processing effort. In a SYNC-based update rate, this KPI is related to the frequency of the SYNC event.

Wireless Transmission Range — This KPI is the maximum distance that a communication link supports the data transfer with a given reliability and in specific conditions (indoor/outdoor; line-of-sight or LOS).

Power consumption — The power consumption is a measurement of the combination of the computational effort of the nodes and the transmission effort. It is directly related with the update rate as well as with the link quality and, if it exists, the routing activity in each node.

3.2.5 Emerging Technologies and Research Trends

ver livro khan

3.2.6 Network Simulators and Network Emulators

Several techniques have been introduced for performance evaluation of protocols and algorithms in USN, including analytical modeling, simulation, emulation, testbed, and real-world experimentation (Imran et al., 2010). Analytical models are a set of equations that represent the performance of a system. Although analytical models simplify the modeling procedure, they cannot accurately represent the inherent complexity of sensor networks (Krop, 2007). Simulation has been cited as the most frequent and effective method for designing and developing network protocols and algorithms (Imran et al., 2010). By using simulators various scenarios of the real environment can be modeled. Also, they provide the possibility of testing and debugging protocols at any stage of design. Emulation, as a hybrid method, is a combination of hardware and software components accompanying simulation possibilities for network modeling (Kiess and Mauve, 2007). Emulators use

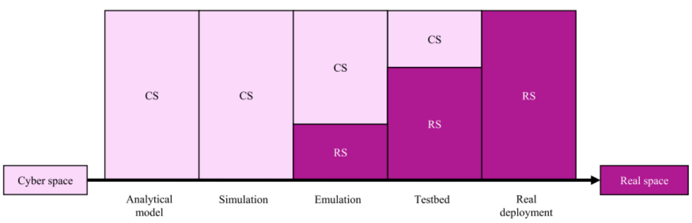

Figure 3.4: FIGURE TO BE RENAMED.

An extensive review on simulation tools, made by Nayyar and Singh (2015), is presented as following:

NS-2 (network simulator-2) Object orientated discrete event simulator tool, based on two languages (C++ and OTcl).

NS-3 (network simulator-3) Written in pure C++, this simulator has been extensively explored in the literature with various modules like 802.15.4, 6LoWPAN and RPL available.

OMNET++ This simulator is based in C++ in the basic modules and uses Network DEscription (.NED) scripts to connect and assemble the simulation basic modules.

J-Sim JavaSim Simulator has been developed in Java and it has several advantages to carry out large scale WSN simulations.

Mannasim Is based on NS-2 to perform WSN simulations.

SensorSim Similarly to Mannasim, this is a framework for WSN based on NS-2 simulator. However it is not available, currently, to the public.

NRL Sensorsim This extension to NS-2 simulator is focused on WSN, similarly to SensorSim and Mannasim frameworks. Currently, it is no longer in development and does not have support.

NCTUns 6.0 On the field of network emulators, this software has the main advantage of using the real world Linux TCP/IP stack and implements almost every IEEE network standards. Last release was on 2010.

SSFNet Java simulator designed especially for WSN's. Last release in 2004.

GloMoSim Non-commercial simulator (the commercial version is *"QualNet"*) designed for wireless and wired network systems.

QualNet 7.0 + EXata 5 Considered one of the most advanced simulator platform these days, QualNet enables a high fidelity virtual model of network, with an advanced GUI. It has a free academic version.

sQualNet Simulator Is an extension of QualNet for sensor network specific models.

OPNET Modeler Suite Is a powerful collection with an interactive GUI to build network scenarios. It has a free academic edition.

SENSE Is a powerful sensor network simulator and emulator.

DRMSim Is a Java-based software that enables large-scale network simulation.

NetSim Is a simulator designed for protocol modeling, network research and development and defense applications.

UWSim This simulator is designed for marine robotics research.

Visual Sense This modeling and simulation software is designed for wireless and WSN applications, as an extension of Ptolemy II.

Viptos Is an interface/bridge between TinyOS and Ptolemy II.

PTOLEMY II Is an open source simulation software, based on Java and with actor-oriented design (where actors are software components, have a concurrent execution and communicate via interconnected ports)

SENS This specific framework for WSN's simulator and emulator that uses a simplified sensor model.

SHAWN Is a customizable sensor network simulator focused on the simulation of the effect caused by a phenomenon (not the phenomenon itself) with scalability and support for extremely large networks. According to SHAWN development repository, last contribution was on 2013.

SIDnet-SWANS This Java-based simulator was made to provide a simulation and proof-of-concept platform for application of WSN's. Latest version was released in 2011.

WSim/Worldsens Simulater/WSNet Simulator This simulator states for being an event-driven simulator for large scale wireless networks. Latest version was released in 2009.

WSN Localization Simulator This is a WSN location simulator stating for being easy, scalable and extendable to many/different localization schemes. It was released in 2013.

NetTopo Simulator This framework is an open source simulator designed in Java. Its main objective is to analyze various algorithms in WSN's.

SIDH Is a Java-based simulator focused on the simulation of thousand of sensor nodes.

PROWLER The Probabilistic Wireless Sensor Network Simulator is a framework that runs under Matlab and is focused to TinyOS applications.

Matlab/Simulink With extensive usage by the research community, Matlab and Simulink software provides resourceful toolboxes for simulation of communication networks, being possible to build a complete WSN model system.

PiccSIM This simulation platform uses Matlab/Simuling and NS-2 for networked control systems (in particular wireless)

LabVIEW Various toolboxes to simulate WSN's are available with LabVIEW.

3.3 Section 3: Energy sensors

1. Sensor overview – historic perspective

2. Current transducers and voltage transducers a. Commonly used technologies and principles b. New breakthroughs

3. High power measurement challenges in RTS

4. Energy measurement technologies in RTS

3.3.1 Sensor overview – historic perspective

3.3.2 Current transducers and voltage transducers

a. Commonly used technologies and principles b. New breakthroughs

3.3.3 High power measurement challenges in RTS

3.3.4 Energy measurement technologies in RTS

3.4 Section 4: Power system of RTS

1. Overview of existing worldwide power systems

2. Overview in the perspective of production-distribution-consumption a. Traction substation (Production) b. Catenary (Distribution line) c. Rolling stock (Consumption/load)

3. Traction substation transformer overview

4. Catenary (?)

5. Train power transformer

6. Train motor and power converter

7. Auxiliary loads

3.4.1 Overview of existing worldwide power systems (25kV/15kV/DC 50Hz/16,6Hz)

3.4.2 Overview in the perspective of production-distribution-consumption

3.4.2.1 Traction substation (Production)

3.4.2.2 Catenary (Distribution line)

3.4.2.3 Rolling stock (Consumption/load)

3.4.3 Traction substation transformer overview

3.4.4 Train power transformer

3.4.5 Train motor and power converter

3.4.6 Auxiliary loads

3.5 Section 5: Decision Support Systems (DSS)

1. Overview/definition 2. Eco-driving – driving assistant 3. Timetable scheduling 4. Maintenance support

3.5.1 Overview/definition

3.5.2 Eco-driving – driving assistant

3.5.3 Timetable scheduling

3.5.4 Maintenance support

3.6 Section 6: Outlier detection in RTS energy measurement

An important contribution of a wireless sensor network in the railway system is the availability of useful knowledge about energy consumption to the decision support systems.

Therefore, such acquisition systems are required to provide accurate data regardless of the quality of the acquisition sensors, electromagnetic influences (EMI), sensor supply fluctuations, among other error sources.

Through computational algorithms, the increase of communication reliability and fault tolerance is possible. Those computational algorithms detect outliers or, in the scope of this PhD, detect erroneous data that will disturb the outcomes of decision support systems.

1. Definition of outlier detection in RTS energy measurement perspective
2. Literature review of Outlier detection in WSN

a. Motivation

b. Research areas

c. Challenges

3. Taxonomy of outlier detection techniques

a. Classification based

b. Statistical based

c. NN-based

d. Clustering based

e. Spectral decomposition-based

3.6.1 Definition of outlier detection

Outlier detection is a computational task to detect and retrieve information from erroneous data values. The definition is usually close to anomaly detection or deviation detection.

Branch et al. (2006) identifies outlier detection as an essential step to either suppress or amplify outliers and precedes most data analysis routine. Abid et al. (2016) points the need of detecting aberrant data and sensors within an WSN. Zhuang and Chen (2006) extends the outlier definition to the case where the outliers are introduced in sensing queries and in sensing data analysis.

In the scope of the PhD, an outlier is a data value or a data instance that do not represent the correct consumption status.

The threshold of what is an outlier or not (or a value that do represent the correct consumption status or not) is given by the output of the subsystem that is immediately after the acquisition of consumption status subsystem, the decision support subsystem, gave a correct output or not. Figure 3.5 illustrates the integration of the consumption acquisition subsystems with the decision support subsystem.

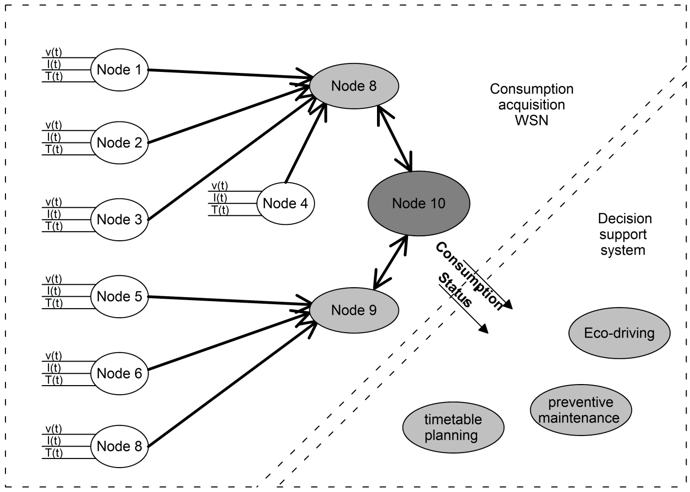

Figure 3.5: Integration of the WSN with a decision support system.

Without an outlier detection mechanism, the decision support subsystem may have the following outputs:

Input deviation from real value lower than a threshold The Decision Support Subsystem output is in accordance to the real consumption conditions.

Input deviation from real value greater than a threshold The Decision Support Subsystem output is not in accordance to the real consumption conditions.

The problem of taking decisions based on wrong considerations of the consumption status is that it may lead to loss in desirable efficiency or increase of costs.

Let us consider a simple and hypothetical example where the DSS will provide an output towards suggesting an action in preventive maintenance based on the usage of a component. Considering that the usage of the component is depending on the counting of situations that the component is working above the nominal conditions. Without an outlier detection mechanism, the outliers will induce the DSS in counting situations of overcharge of the component, where the measurement is not related to working nominal conditions but is related to external influences, such as EMI or temperature. The output of DSS may suggest a preventive maintenance on a component that is working in proper conditions.

To conclude, with an outlier detection mechanism in the consumption acquisition subsystem the decision support subsystem may know if the value of consumption is an outlier or not and, with that information, the DSS output will be more accurate with the real conditions of operation.

3.6.2 Literature review of Outlier detection in WSN

Wireless sensor networks (WSNs) has been widely used in several applications in several domains such as industrial, scientific, medical and others. Those applications have been supported by the advances in wireless technologies as well as in the evolution of microcontroller technologies, with enhanced processing capabilities associated with reduced energy consumption.

3.6.2.1 Motivation

Rajasegarar et al. (2007) points an important motivation for the inclusion of computational algorithms, i.e. outlier detection algorithms, to reduce the transmission of erroneous data, since in WSNs, the majority of the energy consumption occurs in radio communication. In particular, they present the case of Sensoria sensors and Berkeley motes, where the energy consumption in communication exceeds, in ranges from 1000 to 10000, the energy consumption of computation.

Thus, this raises a research opportunity to reduce the communication usage of μCs, by adding processing features, where the small increase in the computation will significantly reduce the energy consumption in the transmission. These processing features are, among others, the outlier detection algorithms.

On the field of the quality of the data acquired by WSNs, the motivation of detecting outliers in data acquired from WSNs has been extensively presented in the literature. The need for acquire data from harsh or "highly dynamic" environments as well as the need to validate and extract knowledge from the acquired data are one of the main points in the motivation to study the outlier detection in WSNs, Zhang et al. (2010); Chandola et al. (2009); Ghorbel et al. (2015); Martins et al. (2015).

3.6.2.2 Research areas

Zhang et al. Zhang et al. (2010) identifies the outlier detection research areas in three domains:

- Intrusion detection: Situation caused by malicious attacks, where the detection techniques are query-driven techniques;
- Fault detection: Situation where the data suffer from noise and errors and where the detection techniques are data-driven ones;
- Event detection: Situation caused by the occurrence of one atomic or multiple events and where the majority of the research has been developed due to its complexity.

Based on the division of this three domains, the upcoming research is intended to be focused on the event detection and fault detection techniques. Specifically, the main goal for this research will be the event detection algorithms.

3.6.2.3 Challenges

The challenges of outlier detection in WSNs are related to the quality of the acquisition of the sensors, the reliability of the modules in terms of energy or environmental susceptibility, and the communication requirements and restrictions.

Zhang et al. Zhang et al. (2010) lists the challenges as the following:

- Resource constraints;
- High communication costs;
- Distributed streaming data;
- Dynamic network topology,
  frequent communication failures,
  mobility and heterogeneity of nodes;

- Large-scale deployment;
- Identifier outlier sources;

Branch et al. (2006) identifies an important challenge, where the probability of occurrence of outlier events are extremely small. Abid et al. (2016) as well as Sheng et al. (2007) identifies the large amount of data as the main challenge for outlier detection in WSN. Zhuang and Chen (2006) identifies the inexpensive and low fidelity sensors as the main reason for the error generation and the main challenge is identified on the distributed streaming data among a large amount of sensors. Ghorbel et al. (2015) points the main challenge as the processing of data from sensors that generate continuously data, that is uncertain and unreliable.

To conclude, the main challenge will be the usage of inexpensive and low fidelity sensors that will be affected by the rush railway environment. Complementary, the main challenge of using outlier detection mechanisms in the railway WSN is the balance between the detection accuracy and the influence that undetected data-instances will induce in other subsystems (in particular in decision support systems dependent on data from the WSN). In addition, the detection accuracy is directly related with the memory usage, computational requirements, communication overhead, etc.

3.6.3 Taxonomy of Outlier Detection Techniques

The study of detection techniques requires a well-defined taxonomy framework that addresses the related work on the different areas. This taxonomy is well defined and solid in the literature, where the works of Zhang et al. (2010) and Chandola et al. (2009) reflect a similar approach on presenting a taxonomy for outlier detection techniques.

In the following sections, a coverage in relevant techniques is presented:

- Classification based techniques.
  - Bayesian Networks
  - Rule-based techniques
  - Support Vector Machines
- Statistical based techniques.
  - Parametric — Gaussian based
  - Non-parametric — Histogram based
  - Non-parametric — Kernel function based
- Nearest Neighbor-based techniques.
  - Using distance
  - Using relative density
- Clustering based techniques.
- Spectral Decomposition based techniques.

3.6.4 Classification based techniques

Classification based techniques are based on systematic learning approaches which use sets of data. The supervised approaches require previous knowledge to train a model (or classifier) from a set of data instances (or training data) and classifies a new data instance as normal or as outlier. The unsupervised approaches do not require knowledge and learn the boundary around normal instances, declaring the new instance as normal or as outlier depending if the data instance is outside of the boundary of the previous data sets.

The classification based techniques are listed as the following:

- Neural Networks-based;
- Bayesian Networks-based;
- Rule-based;
- Support Vector Machines-based.

Neural networks-based approaches are interesting strategies for outlier detection where a given neural network might be trained with only normal data-sets. At testing stage, the data instances that are similar to the training data-set are accepted by the neural network and then considered as normal. The remaining data-sets are rejected by the neural network due to their lack of similarity with normal data-sets. Thus, those data instances are considered as outliers. These techniques are classified as semi-supervised due to their need for normal data-sets for the training stage.

Bayesian networks-based approaches are identified as prominent techniques for outlier detection in WSNs, being the reason why they are extensively covered further on in 3.6.4.1. Those techniques use probabilistic graphical models to detect outliers based on the interdependencies of different variables.

Rule-based techniques, presented in 3.6.4.2, classify an outlier based on a confidence value related to the number of the training instances correctly classified by a given rule and the total number of training instances covered by the same rule. For each test instance, all the rules are tested and the confidence value is ordered. The output of this outlier detection technique is given by the inverse of the confidence value of the rule that better captures the test instance.

Support Vector Machine (SVM) techniques are used for outlier detection to classify a given instance based on the fitness of a hyper-sphere to the data in a higher dimensional space. The hyper-sphere is obtained with a linear optimization algorithm where the objective function of this linear optimization problem is to minimize the radius R that cover the majority of the image vectors. The output of the SVM applied to OD is the classification of the image vectors as outliers if they are outside of the hyper-sphere. The SVM techniques are presented in 3.6.4.3.

3.6.4.1 Bayesian Networks

Zhang et al. (2010) divide the bayesian network based techniques in three categories:

- Naïve Bayesian Networks;
- Bayesian Belief Networks;
- Dynamic Bayesian Network Models;

All those approaches use probabilistic graphical models to represent a set of variables and their probabilistic interdependencies. This graphical model aggregates the information from different variables and provides an estimate on the expectancy of an event to belong to the learned class.

Xiang et al. (2016) illustrates an application to measure the concentration of NO_2, CO and O_3 pollutants, using a bayesian network. All the three variables are all correlated and also depends on the temperature as presented in figure 3.6. The real measurements acquired by the microcontroller are represented with (s) and the representations in (t) refers to the real concentration of those pollutants.

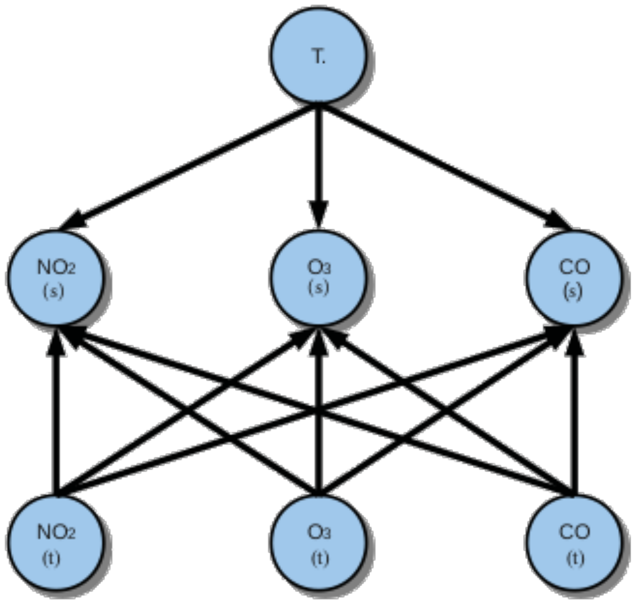

Figure 3.6: Application of a Bayesian Network to an atmospheric measurement system.

The three categories presented by Zhang et al. (2010) differs between them where the first category captures the sensor nodes correlations on spatio-temporal domain; The second one considers not only the spatio-temporal correlations but also the conditional dependence of sensor attributes; The third category proposes the measurement of state variables at a current time instance.

Janakiram et al. (2006) proposes the detection of outliers in sensor streamed data by capturing the conditional dependencies among the observation of it's attributes. this is made in three phases:

Training Phase Phase where the Bayesian Belief Network is trained to capture the spatio-temporal correlations.

Testing Phase Phase where the trained BBN is tested on the level of accuracy and, if needed, the learned parameters are updated.

Inference Phase Phase where the missing values are inferred and the remaining streamed data are tested to detect if it is an outlier or not.

Janakiram et al. (2006) also defined the BBN, where the BBN is a directed graph, together with an associated set of probabilistic tables. The graph is divided in nodes and arcs, where the nodes represents the variables and the arcs are the representation of the casual/influential relationship among variables.

The main contribution of BBN is the possibility to have a model that, with the dependency between uncertain variables (by filling a node probability table), it is possible to describe complex probabilistic reasoning about uncertainty.

Janakiram et al. (2006) describes their process in three steps:

- Constructing the Bayesian Belief Network

  IF a few variables have direct dependencies

  AND many of the variables are conditionally independent

  THEN all the probabilities can be computed from the joint probability distribution.

- Learning Bayesian Belief Networks

  IF the network structure is given

  AND all variables are fully observable in the training examples

  THEN estimating the conditional probabilities is enough

  IF the network structure is given

  AND some of the variables are observable

  THEN apply neural network using Gradient Ascent Procedure

  IF the network structure is unknown

  THEN Use heuristic search

  OR Use constraint-based technique to search through potential structures

- Inferring from Bayesian Belief Networks

  THESIS The probability distribution of certain attributes might be inferred

  PROOF Given the fact that the values that other attributes can take are known

Paola et al. (2015) proposes an adaptive distributed Bayesian approach for detecting outliers in data collected by a WSN. The focus of the proposed algorithm is the optimization of outlier classification accuracy, time and communication complexity and also considering externally imposed constraints on conflicting goals. The proposed algorithm is intended to run in each sensor node.

From the individual sensor node point of view, this algorithm consists in two phases:

Outlier detection Where, based on sensor readings and on the collaboration with neighbors, is made the probabilistic inference where the results are evaluated in three metrics: classification accuracy, time complexity and communication complexity.

Neighborhood selection Where the best neighbors are identified and selected to cooperate with, and, in addiction, to correspond to a reconfiguration of the Bayesian Network structure.

In the global point of view, if there is a high number of cooperating nodes, the classification is naturally higher with the drawback if increasing the processing time and communication complexity (thus resulting in increased detection delay and increase of energy consumption).

Xiang et al. (2016) proposes the addition of recover and recalibrate the drifted sensors simultaneously based on the usage of a Bayesian network. The authors have applied their algorithm to the measurement of the variables in the sensor readings of the NO_2, CO and O_3 pollutants, as previously presented in figure 3.6. Based on the correlations of the sensor readings and on the temperature influence, the algorithm itself detects the outliers, recover valid information and adjust the BBN to automatically recalibrate the sensor.

3.6.4.2 Rule-based techniques

Rule based is another classification based technique for outlier detection. Similarly, this technique is based on a training stage from a data-set and a model generation to detect new data-instances based on history values.

Rule based techniques depends on two steps Chandola et al. (2009):

- **1. Learn rules from the training data-set**
  Using a learning algorithm (i.e. RIPPER, Decision Tree, etc.)
  Where each rule has an associated confidence value proportional to the ratio:
  $\text{Confidence Value} = \frac{\text{number of training instances correctly classified by the rule}}{\text{number of total training instances covered by the rule}}$
- **2. Find for each test instance the rule**
  That better capture the given test instance.
- $\Rightarrow$ **The anomaly score is**
  The inverse of the confidence value for the rule that better capture the test instance.

Islam et al. (2016) proposes an algorithm for outlier detection inserted in rule-based taxonomy. They propose a new belief-rule-based association rule, with the focus on handling various types of uncertainties.

Due to the nature of the sensor data, a traditional inference mechanism cannot be used. Therefore, they propose a new inference mechanism for the rule-based algorithm that consists of an input transaction database that is converted into the following:

- belief transaction database;
- support calculation;
- belief matrix;
- confidence calculation;
- belief association rule discovery.

3.6.4.3 Support Vector Machines

Rather than performing outlier detection in the central node, Rajasegarar et al. (2007) proposes a distributed approach to:

- performs detection on local data at each node
- and communicates to the parent node only the summary information to perform at upper layer the global classification of the data.

Their proposal is based on a one-class quarter sphere SVM and is divided into 2 parts:

- **Anomaly detection algorithm**

  The OD is supported by previous works where, with the fitness approach of a hypersphere to the data in a higher dimensional space, and by applying a linear optimization to the problem of fitting the hypersphere with minimal radius R, having the center fixed at the origin and encompassing the majority of the image vectors.

  The result of the linear optimization problem is the classification of the image vectors as:

  → **Support Vectors**, if inside the sphere;

  → **Outliers**, otherwise.

- **Distributed anomaly detection**

  1^{st} **step:** Each sensor node runs the entire AD algorithm on local data;

  2^{nd} **step:** The resulting radius is sent to the parent node;

  3^{rd} **step:** Each parent computes the global radius;

  4^{th} **step:** Parents sends the radius to children nodes;

  5^{th} **step:** Children compares global radius with local one and updates parameters.

Xu et al. (2012) proposes a KNN-SVM which is a Support Vector Machine based on K-Nearest Neighbor Algorithm.

Despite KNN taxonomy is presented further on in 3.6.6, in a synthesis the KNN is a distance-based approach that detect outliers in data-instances lying in the sparsest regions or lying in the outside of a given model boundary of the feature space.

Considering the Quarter sphere SVM technique proposed by Rajasegarar et al. (2007) the KNN-SVM combine the origin and the radius R that contain most of the samples and introduces kernel functions to make the optimization region more tighten.

3.6.5 Statistical based techniques

Chandola et al. (2009) identifies the statistical techniques for anomaly detection based on the assumption that, in a stochastic model, the most common data instances occur in high probability regions and the anomalies occur in low probability regions.

To detect anomalies, parametric techniques are suggested, since those techniques **assume** the knowledge of the underlying distribution and **estimate** the parameters from a given data set. Non-parametric techniques differ from parametric ones without the need of assuming the knowledge of the distribution.

Andrade et al. (2016) lists some statistical parametric techniques:

Peirce's Criterion This statistical parametric method is based on a normal distribution.

Chauvenet's Criterion Is based on the assumption that a given arbitrary measurement may be rejected, if the probability of having the deviation for the average value is lower than the inverse of the double of the number of measurements.

3.6.5.1 Parametric — Gaussian based

Zhang et al. (2010) summarizes parametric techniques as an anomaly detection strategy based on the following steps:

- The available knowledge is generated from a known distribution;
- The distribution parameters is then estimated from the given data.

The usage of Gaussian models allows the spatial correlation of readings towards distinguishing between outlying sensors and event boundary.

3.6.5.2 Non-parametric — Histogram based

Sheng et al. (2007) proposes a histogram-based method to reduce the communication cost on sensor networks. The main objective of this proposal is to collect hints (in a form of histogram) about the data distribution and, with the knowledge from these hints, unnecessary data is filtered and potential outliers are detected

Complementary, Wang and Li (2013) introduces clusters on incremental histogram scheme based on a divide and conquer strategy:

- The wireless network is divided in clusters (based on adjacent nodes and data correlated strategy);
- The cluster head and cluster members updates the histogram incrementally and compares histograms in the form of Kullback-leibler divergence differentially (Kullback-leibler divergence is a convenient and robust method of measuring the difference between two data sets in a statistical sense.)

3.6.5.3 Non-parametric — Kernel function based

Zhang et al. (2010) synthesizes the concept of Kernel function non-parametric approaches as methods for estimating the probability distribution function for normal instances (and a new instance that occurs on a low probability area is declared an outlier). Later on in section 3.6.8 is presented the Kernel Principal Component Analysis (KPCA) used by Ghorbel et al. (2014) for outlier detection in WSN's.

Andrade et al. (2016) identify some kernel regression techniques:

Marzullo's Fault Tolerant Sensor Averaging (FTA) Simple method for sensor fusion where the data assumed as anomalous is deleted.

Elmenreich's Confidence-Weighted Averaging (CWA) The sensor's confidence are correlated with the sensor's variance.

CWA+FTA method This method combines both methods where the confidence-weighted average is calculated and two-thirds of the anomalous data is removed.

3.6.6 Nearest Neighbor-based techniques

A promissory technique is extensively explored in the literature with the concept of neighborhoods, based on the key assumption that normal instances occurs in dense neighborhoods and anomalies occurs far from their closest neighbors.

3.6.6.1 Using distance

Branch et al. (2006) proposes algorithms that implements nearest neighbor-based techniques for outlier detection in WSN's. The proposed unsupervised anomaly detection techniques use the following different algorithms:

- The distance to the k^{th} nearest neighbor;
- The average distance to the k nearest neighbors;
- the inverse of the number of neighbors, within a distance α.

Abid et al. (2016) bases the detection technique on the distance between the current measurement and its neighbors. A synthetic database is generated based on the insertion of random values into a real database (in particular the Intel Berkeley lab WSN database).

The procedure is divided in two steps:

- **Step 1a)** For a given time-slot, the data values are sorted in a vector;
- **Step 1b)** After that, for a given point in the vector, Euclidean distance between the predecessor and successor is calculated and stored in a second vector;
- **Step 1c)** Based on the smallest distance between the current point and the predecessor or successor, the current point is linked;
- **Step 2** If the point in the vector is not linked (due to its distance between current point and predecessor/successor higher than a threshold), is declared an outlier;

3.6.6.2 Using relative density

Chandola et al. (2009) defines the NN technique using relative density as a technique that estimates the density of the neighborhood of all data instances. Depending if the instance corresponds to a dense neighborhood or a low density one, the data is declared as outlier or normal.

3.6.7 Clustering based techniques

Chandola et al. (2009) synthesizes the clustering techniques in three categories based on three different assumptions:

- The normal data instances are part of a cluster in the data and the outliers does not fit any cluster;
- The normal data instances are present close to its closest cluster centroid and the outliers lies far away from their closest cluster centroid;
- The normal data instances are part of large dense clusters and the outliers are part of small or sparse clusters.

Rajasegarar et al. (2006) uses a technique to minimize the communication overhead by using clusters among the sensor readings. In a further step, it merges the clusters before the data is sent to other nodes.

Andrade et al. (2016) presents a methodology to apply clustering and statistical techniques. The clusters are grouped according to the spatial position of the sensors and a k-means nearest-neighbor technique is used to provide a better understanding of the sensed environment. The proposed methodology follows a two-step procedure, starting with the usage of clustering information and followed by a statistical-based method. The statistical method is Elmenreich's Confidence-Weighted Averaging (CWA), where the sensor's confidence is correlated with the sensor's variance.

Cenedese et al. (2017) considers the network decomposition (i.e. the communication network topology) together with the data clustering measurements. They propose two algorithms: a centralized clustering algorithm (CCA) and a distributed clustering algorithm (DCA).

3.6.8 Spectral Decomposition-based approach

The usage of Principal Component Analysis (PCA) technique is inherent to the spectral decomposition-based approach. Proposed by Chatzigiannakis et al. (2006), this technique efficiently models the spatio-temporal data correlations, in a distributed approach and, the local outliers are evaluated with the correlation among the sensor nodes.

Zhang et al. (2010) evaluates the Spectral Decomposition-based techniques in two outcomes:

- The PCA-based techniques are of interesting usage where it captures the normal pattern of data;
- However, it is computationally very expensive due to the need of selecting suitable principle components (needed to estimate a correlation matrix of normal patterns).

Gil et al. (2016) lists the steps of a PCA-based approach:

Robust recursive location estimator The PCA requires the estimation of the mean at each sampling time (the measurement vector x is centered).

Subspace tracking approach To avoid the need of extensive calculation of the eigendecomposition, the authors takes advantage of subspace tracking (which recursively tracks the signal subspace spanned by the major principal components)

Recursive eigendecomposition computation The eigenstructure associated to an underlying space is recursively estimated;

Robust recursive detection criteria Two measures to compare the distance between a value and the remaining time-series are used

Robust subspace tracking Having an updating procedure to affect the signal subspace, if an outlier is detected, this updating procedure is skipped.

4 Methodology and Work Plan

Note: not in chronological order

4.1 RTS wireless network

4.1.1 Purpose

Model and simulate a WSN for energy measurement of RTS rolling stock, with an advanced network infrastructure (englobing both train WSN and station AP's)

4.1.2 Contribution

- An energy measurement system in rolling stock does not require a broadband real-time/continuous communication (such as LTE), being possible to collect and store data in train data concentrator and, while the train is waiting at station for passenger exchange (which lasts for less than one minute), the data is transferred between train and station AP (and then to a remote server).

- Therefore, the contribution will be the cost reduction of information transmission of energy sensor network data.

4.1.3 Methodology

- Modeling of energy sensor network of rolling stock: sensor nodes and data concentrator
- Modeling of infrastructure: train concentrators, station AP, station data "buffer" and station internet connection
- Implementation in simulation environment of such models, using NS3 simulator or similar
- Definition of "sensor data rate" as function of the line length-between-stations (?)

4.2 Non-intrusive self-powered sensor node

4.2.1 Purpose

In the scope of Shift2Rail, is expected to develop a smart meter for railways. The purpose is to model, simulate and implement a series of sensor nodes for current measurement in the transformer´s secondary windings. Assuming that the railway environment requires non-intrusive measurement devices and, if possible, self-powered, a set of requirements is then identified for the sensor node:

- Electrically non-intrusive (using hall-effect, rogowsky or current transformer principles; without the need for mechanically changing the windings)
- Self powered, if the current transformer has sufficient power capabilities
- With high processing capabilities, high acquisition frequency and sufficient amount of memory

    Variable acquisition in tens of samples per second (according to the power quality standard of 15kHz <?>)

    Frequency analysis capability

    Capable of implement outlier detection algorithms

4.2.2 Contributions

- New advanced sensor node for high current measurement
- Possible contribution: Given the measurement characteristics, a self powered wireless sensor node can implement features of high processing.

4.2.3 Methodology

4. Methodology to be defined

4.3 Rolling-stock traction transformer model

4.3.1 Purpose

Model the train transformer with two perspectives:

- Efficiency estimation based on secondary measurements
- Evaluation of transformer operation towards fault detection

4.3.2 Contributions

- Possible contribution: an accurate model for train transformer, capable of efficient estimation of energy consumption based on secondary windings current measurements
- Possible contribution: assuming that the influence of transformer in the power life cycle cost is relevant (see note), the contribution will be the operation monitoring towards maintenance cost reduction.

4.3.3 Methodology

- Study failure rates of trains/transformers;
- Model in a simulation environment the power transformer
- Identify and model transformer failures
- Implement in sensor nodes an energy estimation mechanism based on the loss model of the transformer and sensor nodes measurements
- Implement in sensor nodes a frequency analysis towards operation monitoring
- Prepare and implement results in field operation

Bibliography

Abid, A., A. Kachouri, and A. Mahfoudhi (2016, mar). Anomaly detection through outlier and neighborhood data in wireless sensor networks. In *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. Institute of Electrical and Electronics Engineers (IEEE).

Andrade, A. T. C., C. Montez, R. Moraes, A. R. Pinto, F. Vasques, and G. L. Da Silva (2016). Outlier detection using k-means clustering and lightweight methods for wireless sensor networks. In *IECON Proceedings (Industrial Electronics Conference)*, pp. 4683–4688.

Branch, J., B. Szymanski, C. Giannella, R. Wolff, and H. Kargupta (2006). In-network outlier detection in wireless sensor networks. In *26th IEEE International Conference on Distributed Computing Systems (ICDCS06)*. Institute of Electrical and Electronics Engineers (IEEE).

Cenedese, A., M. Luvisotto, and G. Michieletto (2017). Distributed clustering strategies in industrial wireless sensor networks. *IEEE Transactions on Industrial Informatics 13*(1), 228–237.

Chandola, V., A. Banerjee, and V. Kumar (2009, jul). Anomaly detection. *ACM Computing Surveys 41*(3), 1–58.

Chatzigiannakis, V., S. Papavassiliou, M. Grammatikou, and B. Maglaris (2006). Hierarchical anomaly detection in distributed large-scale sensor networks. In *11th IEEE Symposium on Computers and Communications (ISCC'06)*. Institute of Electrical and Electronics Engineers (IEEE).

Council of European Union (2008). Directive 2008/57/EC on the interoperability of the rail system within the Community.
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0057.

Council of European Union (2014a). Commission regulation (EU) no 1301/2014 on a technical specification for interoperability relating to the 'rolling stock — locomotives and passenger rolling stock' subsystem of the rail system in the european union.
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.356.01.0228.01.ENG.

Council of European Union (2014b). Commission regulation (EU) no 1301/2014 on the technical specifications for interoperability relating to the 'energy' subsystem of the rail system in the union.
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.356.01.0179.01.ENG.

Ghorbel, O., M. Abid, and H. Snoussi (2014, may). Kernel principal subspace based outlier detection method in wireless sensor networks. In *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, pp. 737–742. Institute of Electrical and Electronics Engineers (IEEE). cited By 0.

Ghorbel, O., W. Ayedi, H. Snoussi, and M. Abid (2015, jun). Fast and efficient outlier detection method in wireless sensor networks. *IEEE Sensors Journal 15*(6), 3403–3411.

Gil, P., H. Martins, and F. Januário (2016, may). Detection and accommodation of outliers in wireless sensor networks within a multi-agent framework. *Applied Soft Computing 42*, 204–214.

Hackmann, G. (2006). 802.15 personal area networks.

IEEE (2016). *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE Std. 802.11-2016.

Islam, R. U., M. S. Hossain, and K. Andersson (2016, nov). A novel anomaly detection algorithm for sensor data under uncertainty. *Soft Computing*.

Janakiram, D., V. Reddy, and A. Kumar (2006). Outlier detection in wireless sensor networks using bayesian belief networks. In *2006 1st International Conference on Communication Systems Software & Middleware*. Institute of Electrical and Electronics Engineers (IEEE).

Martins, H., L. Palma, A. Cardoso, and P. Gil (2015, may). A support vector machine based technique for online detection of outliers in transient time series. In *2015 10th Asian Control Conference (ASCC)*. Institute of Electrical and Electronics Engineers (IEEE).

Mohassel, R. R., A. Fung, F. Mohammadi, and K. Raahemifar (2014). Electrical Power and Energy Systems A survey on Advanced Metering Infrastructure. *INTERNATIONAL JOURNAL OF ELECTRICAL POWER AND ENERGY SYSTEMS 63*, 473–484.

Nayyar, A. and R. Singh (2015). A comprehensive review of simulation tools for wireless sensor networks (wsns). *Journal of Wireless Networking and Communications 5*(1), 19–47.

Panousopoulou, N. (2014). Notes on wireless sensor networks: Applications, standards and programming tools. lecture presentation slides.

Paola, A. D., S. Gaglio, G. L. Re, F. Milazzo, and M. Ortolani (2015, may). Adaptive distributed outlier detection for WSNs. *IEEE Transactions on Cybernetics 45*(5), 902–913.

Petersen, S. and S. Carlsen (2011). Wirelesshart versus isa100.11a: The format war hits the factory floor. *IEEE Industrial Electronics Magazine 5*(4), 23–34. cited By 131.

Radmand, P., A. Talevski, S. Petersen, and S. Carlsen (2010). Comparison of industrial wsn standards. In *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*, pp. 632–637. IEEE.

Rajasegarar, S., C. Leckie, M. Palaniswami, and J. Bezdek (2006). Distributed anomaly detection in wireless sensor networks. In *2006 10th IEEE Singapore International Conference on Communication Systems*. Institute of Electrical and Electronics Engineers (IEEE).

Rajasegarar, S., C. Leckie, M. Palaniswami, and J. C. Bezdek (2007, jun). Quarter sphere based distributed anomaly detection in wireless sensor networks. In *2007 IEEE International Conference on Communications*. Institute of Electrical and Electronics Engineers (IEEE).

Reddy, K. S., M. Kumar, T. K. Mallick, H. Sharon, and S. Lokeswaran (2014). A review of Integration , Control , Communication and Metering (ICCM) of renewable energy based smart grid. *Renewable and Sustainable Energy Reviews 38*, 180–192.

Santschi, C. and J.-P. Braun (2015). The certification of railway electricity meters. Technical report, METAS - Federal Institute of Metrology.

Seppo, K. and O. Elektraflex (2012). Task XVII: Integration of Demand Side Management, Distributed Generation, Renewable Energy Sources and Energy Storages — Subtask 5, Report No. 5. *International Energy Agency DemandSide Management Programme*.

Sheng, B., Q. Li, W. Mao, and W. Jin (2007). Outlier detection in sensor networks. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc 07*. Association for Computing Machinery (ACM).

Siano, P. (2014). Demand response and smart grids — A survey. *Renewable and Sustainable Energy Reviews 30*, 461–478.

Song, J., S. Han, A. Mok, D. Chen, M. Lucas, M. Nixon, and W. Pratt (2008). Wirelesshart: Applying wireless technology in real-time industrial process control. In *Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS*, pp. 377–386. cited By 246.

Usman, A. and S. H. Shami (2013). Evolution of communication technologies for smart grid applications. *Renewable and Sustainable Energy Reviews 19*, 191 – 199.

Wang, Y. and G. Li (2013, aug). Incremental histogram based anomaly detection scheme in wireless sensor networks. In *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)*. Institute of Electrical and Electronics Engineers (IEEE).

Xiang, Y., Y. Tang, and W. Zhu (2016, feb). Mobile sensor network noise reduction and recalibration using a bayesian network. *Atmospheric Measurement Techniques 9*(2), 347–357.

Xu, S., C. Hu, L. Wang, and G. Zhang (2012, sep). Support vector machines based on k nearest neighbor algorithm for outlier detection in WSNs. In *2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*. Institute of Electrical and Electronics Engineers (IEEE).

Zhang, Y., N. Meratnia, and P. Havinga (2010). Outlier detection techniques for wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials 12*(2), 159–170.

Zhuang, Y. and L. Chen (2006). In-network outlier cleaning for data collection in sensor networks. In *In CleanDB, Workshop in VLDB 2006*, pp. 41–48. APPENDIX.