Universidade do Minho
Departamento de Informática

Master Course in Informatics Engineering

Distributed Aggregation Algorithms in Smart Meters

Pre-Dissertation Report

Telmo Rafael Rodriges Remondes

Supervised by:

Prof. Carlos Baquero Moreno

Braga, October 2014

Abstract

Abstract

The power grids all over the planet become increasingly bigger leading to problems of energy waste and sustainability. Since the recognition of this kind of problems, new renewable energy resource emerge , as well as the need to integrate them into the grid. The Smart Grids show up to integrate all these new energy sources and to respond to the new demands of the modern grid.
This new grid is a complex system that englobes a new mechanism to collect measurement data from the consumers meters. The new meters, the smart meters, along with the smart metering system enable the overall system to collect fine-granular readings regarding the energy consumed by the costumers. With the aggregation of this data, several other goals could be achieved such as time-adaptive tariffs, load balancing the distribution of energy and saving computation resources since the aggregation enables to summarize the collected data.
In this work we address the problem of smart metering data aggregation. We propose a distributed data aggregation approach, where all the smart meter sense the consumption data and some of them can work as aggregators as well. We also focus in observing how the aggregation algorithms work in the smart grid, collecting the results and evaluating which algorithm suites best.

Resumo

As redes eléctricas por todo o mundo tornaram-se cada vez maiores, levando a problemas de desperdício de recursos e de sustentabilidade. Desde a constatação destes problemas, novas energias renováveis apareceram assim como a necessidade de as integrar dentro da rede. As *Smart Grids* apareceram para fazer face a essa necessidade de integração e para responder as necessidade da rede moderna. A nova rede é um complexo sistema que engloba novos mecanismos para recolher dados das medições dos contadores dos consumidores. Este novos contadores, *smart meters*, assim como o sistema inteligente de medição permitem a todo o sistema colecionar dados de leituras de fina granulação acerca da energia consumida pelos consumidores. Com a agregação dos dados vários outros objetivos podem ser atingidos como tarifas adaptáveis ao longo do tempo, balancear a distribuição de energia e poupar recursos computacionais considerando que a agregação permite sumariar os dados recolhidos.
Neste trabalho sera tratado o problema da agregação de dados de forma distribuída. E proposta uma abordagem distribuída de agregação, onde todos os leitores inteligentes leem o consumo de energia e alguns funcionam como agregadores.

Contents

List of Figures

Abbreviations

AMI	Automated Metering Infrastructure
AMR	Automatic Meter Reading
SG	Smart Grid
SM	Smart Meters
WSN	Wireless Sensor Network

Chapter 1

Introduction

The power grid is a very important infrastructure in the modern world. The energy it provides is considered of main importance and a basic condition to guarantee minimum life quality.Due to its large size, the power grid consumes a enormous amount of natural resources, make it unsustainable in long term leaving to the dawn of new renewable energy sources that claims the need to modernize the grid since it's mandatory to interconnect them. This modernation, requires that the grid became more sophisticated, eco sustainable and integrate all the energy sources to enable efficient electrical power distribution. These urges lead to a new concept of grid called Smart Grid (SG).

SG is a modern power grid that uses computation, information and communication. In an automatic way, SG improves the energy efficiency, sustainability both in power distribution and in electricity production. It enables the grid to become more sustainable because it makes a more efficient management of natural resources. The SG is composed by 'Islands of Automation' interconnected with a communication infrastructure [56].

Smart Meters ((SM) are one of the main components of the Smart Grid. They are devices located in the consumers/costumers houses or in industrial facilities that sense the energy consumption. They read periodically in short intervals that range from minutes to milliseconds. This amount of data can be used for performing statistical analyses that lead to effective consumption forecasting and profiling. This fine grained readings will assist users in achieving a more efficient energy use and adapting to the network status and supply by choosing an appropriate and advantageous tariff [43].

In the next years, the amount of user data collected by the SG is expected to dramatically increase with respect to the current electrical power grid.The amount of *Big Data* collected is important because it leads to a great number of comercial advantages and better energy consumption predictions[45].

In this work, we look at the information collected within the SG. More specifically, the information collected by Smart Meters in the households. This data is very important, not only for billing purposes but also to improve the energy management, enabling it to become more *Smart*.

1.1 Objectives

There are two types of architectures[43] regarding the SM data aggregation: *decentralized* and *centralized*. In a *centralized* architecture, the meters only sense the energy consumption every specific time and send it to a central data aggregator center. In a *decentralized* architecture the meters sense the consumers consumption and they also perform a partial data aggregation themselves. It's called in-network aggregation[43].

In this work, we will focus on the second type of architecture considering it provides more interesting challenges. The main goal of this work is, considering a *decentralized* architecture, evaluate an efficient data aggregation algorithm that provides relevant information to the consumer and to the electricity producer. In order to achieve the main objective, it's important to first understand the various possible *decentralized* architectures and the the role of each component. As we saw in [5] there are some sensors that work as aggregation nodes an others that work as simple nodes.

At first, it is important to know how the SG works, how all components interact together and the status of deployed models . Furthermore, it is important to construct a suitable topology for this work, with several meters collecting information about the consumers consumption and aggregate that data in a distributed way. This topology may be constructed consedering real and deployed examples of a smart mettering system. This is an important part of this work as the study of the current algorithms to performe distributed aggregation.

The study of distributed aggregation algorithms embraces the awareness of their functionalities, advantages and disadvantages. It also requires a implementation of then in familiar topologies to understand in better way how the algorithms behave and also to adquire insight about them.

When we have both the topology and also the algorithms, the next step will be implement them. We are interested in knowing which algorithm provides the best results in time, exchange messages, scalability, resilience, fault tolerence and accuracy. It is also imporant to understand which aggregation funtions are important to compute in this specific context. Function such as $AVERAGE$ or SUM may be important, so it is mandatory to choose algorithm that enable these functions.

In the end, an overall comparison between the algorithms will be presented. Improvements to the algorithms may be required in order to obtain relevant information to the consumer and to the electricity producer. The improvements will occur as we select the kind of data that is important to aggregate and collect.

1.2 Motivation

As stated before, Smart Grid is a new and important concept of grid that is of main importance towards the world energy sustainability. The new needs and urges for integration of the new renewable energy sources make the upgrade of the grid mandatory.

With this concept in mind, an important part, the smart meters and the AMR, of this intelligent grid is studied. The data collected from the meters is one of the main parts of a electrical grid. Not only for billing purposes, as it is said in above section, but to achieve better management(management that enables the grid to spare less resources). Grid management could not be done as long as there is no info about the consumption.

The aggregation is a vital process. Aggregation summarize the overall collected data, reducing the computational power required to process the information. Doing this in a distributed fashion withdraws the need of a central aggregator with a high processing power. It also enables the aggregation to be more resilient, reliable and fault tolerant since it is distributed and cheaper in therms of resources.

1.3 Document structure

In this document, a state of the art regarding the overall work thematic is presented. In chapter 2 it is presented the various definitions of the new grid and the point they converge. It is detailed also the infrastructure and model, how the Smart grid is organized and how the diferente components interact. The communication structure and the technologies used on it is also presented, with the various alternatives to realize communication in the modern grid. The important part for this work, smart meters, is detailed.

In chapter **??** is the definition of *Wireless Sensor Networks*(WSN). Smart Metering System could be consider as a specific implementation of WSN so it is important to understand how WSN work and, more important, how in-network aggregation takes place. Awareness of this aspects is important considering it's helpful to understand aggregation in Smart Metering Systems. WSN are a concept widely study with similarities with Smart Metering, a bridge between the two concepts are also presented. Although very similar, the two networks have their differences that are presented in the same chapter.

In chapter 3 it is referenced the concept of distributed aggregation, some aggregation function and its proprieties. The various aggregation algorithms are referenced with its description. The distributed aggregation within the *smart meters* and WSN is mentioned as well.

Chapter 2

Smart Grid

The Smart Grid is a new concept of grid which introduces new technologies into the traditional power system. They enable power grids to become more efficient, integrate other sources of energy rather than traditional ones, and they increase the overall management performance by using modern information technologies. The SG is capable of delivering power in more efficient way and respond to a wide variety of condition and events [39]. Although there are no SGs fully implemented, there are several SG pilot projects show that the new generation grid pose new opportunities and challenges to both consumers and producers.

There are several definitions for the SG among the literature. For example [39] states that "*SG can be regarded as an electric system that uses information, two-way and cyber-secure communication technologies and computational intelligence in an integrated fashion to achieve a clean, safe, secure, reliable, resilient, efficient and sustainable system*".[44] considers the SG as "*a platform that embraces several multidisciplinary concepts towards computerization of electrical power grids*". The common concept over the literature is that SG main goal is to integrate several components, traditional and new, to achieve better performance, interoperability, energy management and sustainability in long term.

SG creates an environment that introduces a convergence between the infrastructure of generation, transmission, distribution, energy, information technology and digital communication infrastructure that enables the exchange of information and control action among the various segments of the power grid.

As it is possible to notice, these integration means that the SG itself is a very complicated system. Achieving the mentioned goals is a complex task. Due to its variety of problems and challenges, most of the proposed solution and studies regarding the SG focus in some specific aspects. An interesting table that presents a comparison between the traditional grid and the SG is presented in [39]:

Existing Grid	Smart Grid
Electromechanical	Digital
One-way communication	Two-way communication
Centralized generation	Distributed generation
Few sensors	Sensors throughout
Manual monitoring	Self-monitoring
Manual restoration	Self-healing
Failures and blackouts	Adaptive and islanding
Limited control	Pervasive control
Few customer choices	Many customer choices

FIGURE 2.1: Brief Comparison Between the Existing Grid and the Smart Grid

2.1 Smart Grid Model

The SG proposes a new model to the power grid where consumers are no longer passive actors in the grid, but prosumers(insert quote) that can both consume and produce energy in small quantities thanks to the new renewable energy sources, plus the introduction of ICT that means new actors that are now present in the grid, enabling new features into the traditional ones.

Typically, the components in a power grid go one way, in the in SG all the flows of electricity and information go two-ways. These new features enable the operations to become faster and more accurate and the interactions between them are increased resulting that, in the future, everything that happens in the grid can be monitored almost in real time.

In am ideal scenario, the SG's new vision, states that a specific component of the grid, such as a household, can both receive energy from the global grid and in the next moment can disconnect from it and become self-sustainable.

There are several visions and models proposed to the SG. One of the more general and accepted model, based on this vision of actors and their interactions, is the NIST report [31] which proposes a conceptual model providing the main actors towards the SG. Costumers, the end users of electricity, Markets, Service Providers, Electricity Companies, Operations, Managers of the Movement of electricity, Bulk Generation, Generation Centers, Transmission and Distribution of energy. In [39] it is provided a more technical approach where the SG is separated into three major subsystems:

- *Smart infrastructure system* embraces the energy subsystem, information subsystem and communication infrastructure subsystem. The energy subsystem is responsible for advanced electricity generation, delivery and consumption. The information subsystems are responsible for information metering, monitoring and management in the context of the SG. Finally, the communication subsystem is responsible for the communication among the various components and also its connectivity.

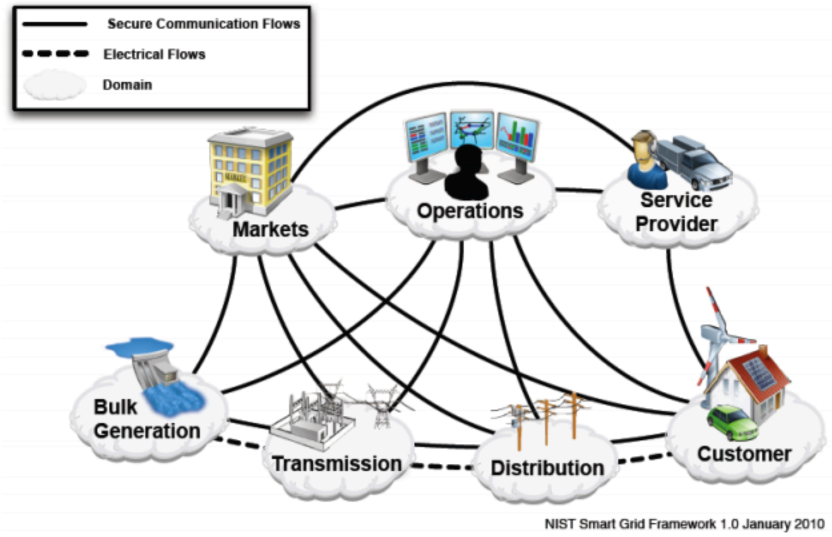

FIGURE 2.2: NIST Conceptual Model for SG

- *Smart management system* Provides advanced management and control services and functionalities, [39] considers this system the key reason why SG can revolutionize the grid. Most of the new grid goals are related to energy efficiency improvement, supply and demand balance, emission control etc. and it is the scope of problems the management systems tries to resolve.
- *Smart protection system* Provides advanced grid reliability analysis, failure protection, security and privacy protection services.

Smart Grids are about improving the current power grid in therms of reliability, energy efficiency and costs while providing a better and more flexibly service to the costumers. These improvements are made possible with the integration of ICT into the power grid, leading to a opportunity for the dawn of new software applications. In [**Andrea**] it is stated that Service Oriented Architectures represent the type of software architecture that satisfies the characteristics needed for a SG software: capable of sustaining a set of systems and applications that are diverse, highly distributed and with constrains for security and timing. In [**Andrea**] it is provided an overall picture that show that show the interaction between these type of software and the physical infrastructure.

2.2 Smart Grid Communication

The most important question regarding the communication is " *what network and communication should be used*"[39]? Since there is no standard communication system in SG, several communication solutions were proposed divided into wired and wireless communication.

Wired solutions are normally more costly to implement than Wireless, mainly because of the need,

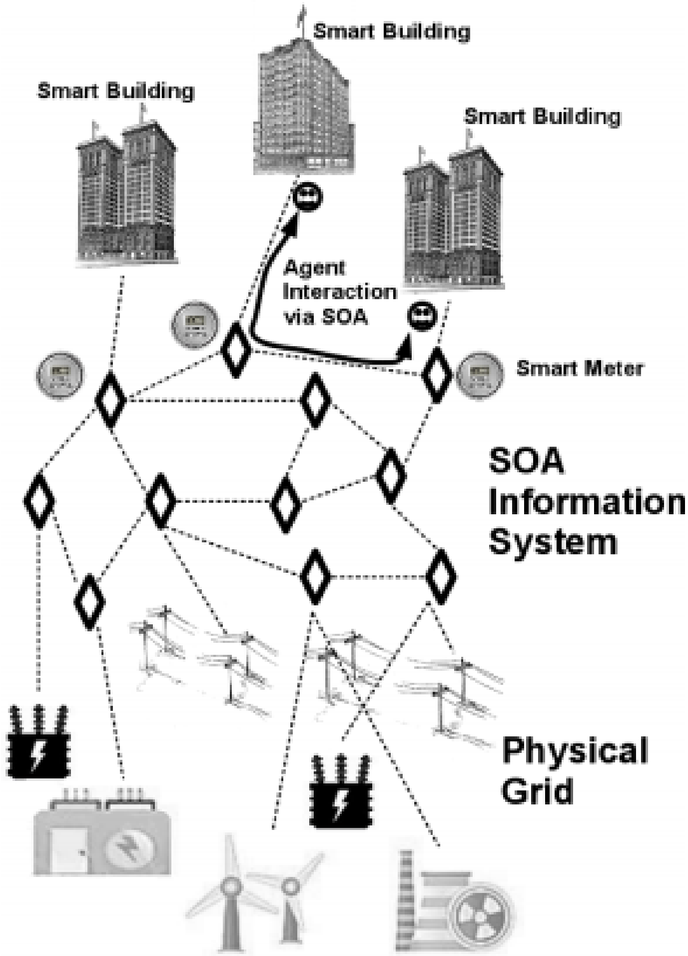

FIGURE 2.3: Smart grid physical and information infrastructure

in some cases, to install or deploy from zero a physical infrastructure like cables to link the components in order to enable communication. Wireless communication can be a better option in terms of cost, time to deploy and furthermore they are normally more suitable for remote end applications [25]. However, they lack of some performance compared to wired solutions, specially in speed. Also, the costs of deploying an wired communication infrastructure can be reduced if they are implemented in the existing infrastructure, case of power line communication that use the power cables.
There are several wireless possibilities for communication.

- *Wireless Mesh Network* (WMN) is a communication network made up for radio nodes organized in a mesh topology[39]. It increases reliability and automatic network connectivity, has large coverage and high data rate.

- *Cellular Communication Systems* GSM and 3G. Useful in case of low computation power devices such as the meters. It is quick and low-cost to obtain data communications coverage over a large geographic area [21]. There several solutions that use a Short Message Service communication to send the meters data.

- *Wireless Communication based on 802.15.4* ZigBee is a wireless communication that is recommended to be used in SG considering the IEEE 802.15.4 protocol stack[25]. ZigBee is designed for radio-frequency applications that require low data rate, long battery life, and secure networking. Selected as the communication technology for the smart metering devices[24] because it provides a standardized platform for exchanging data between smart metering devices and appliances located on costumer premises[39]. WiMax, WirelessHART and ISA100.11a are other examples of wireless communications based on the IEEE 802.15.4 protocol.

Other examples of wireless communication are satellite cognitive radio and microwave communications. Fiber-optic Communications and Power-line Communications are some of the wired communication possibilities. Power-line communication has the advantage of being already installed, so the cost of deployment is less expensive than other wired solutions. Fiber-Optic has also the advantage of being fast but it can be more expensive to deploy because of the need to implement from zero in an infrastructure that lacks cables with that sort of technology.

2.3 Smart Information SubSystem

This part of the SG refers to the whole information that is collected by sensing the consumers consumption and its management . The data collected is often used for billing, grid status monitoring and user appliance control [39]. It is aggregated and collected, afterwards *smart management* is ideally performed on the data.

An important concept in the information subsystem is the *Smart Metering* and the Smart Metering System or Automatic Meter Reading AMR. This system is responsible for collect the data from the measurements that are performed by the SMs.

Other part of the Smart Information SubSystem is the *Smart Monitoring and Measurement* which can be approached by either *sensors* or *phasor measurement units*(PMU). *Sensors* are used for detecting failures, tower collapses, hotspots and extreme mechanical conditions. They can also provide real-time diagnose of the grid status. PMU's are devices that measure the electrical waves on a electrical grid to determinate the health of the system. These systems collect information regarding the status of the grid in order to monitor it and detect failures and outages.

The Smart Metering Systems only collects data from SM's and it only embraces the management of that data.The management refers to the whole information analysis and modeling, integration and optimization.In this specific part of SG there are several areas of research that represent a new set of opportunities.

2.3.1 AMR and AMI

As referred in this document, the smart metering system is composed by smart meters that sense the energy consumption and send their data to a Gateway or a Data Collector. It can also be defined as AMR or AMI. In [43], the AMR is described in more detailed as an "*technology of automatically collecting diagnostic, consumption and status data from energy metering devices and transferring that data to a database for billing troubleshooting and analyzing*".

The Automated Metering Infrastructure is a more sophisticated version of the traditional AMR, it provides two-way communication, enabling a more sophisticated control of a smart meter behavior. Therefore, all of the meter information is available in real time, allowing improved system operations and costumer power demand management[43]. AMI has also the ability of reconfigure from communication failures, perform outage management and reporting, service connect and disconnect and it also enables time stamping of meter data [17]. AMI is built upon AMR.

Current SM enable two-way communication, an important part of the benefits that come from the usage of this new meters, comes from two-way communication, also two-way communication is not only important for behavior control and outage detection, it enables the realization and implementation of in-network algorithms. Now it is more correct to assume that every AMR has an AMI built upon it enabling two-way communication.

In the pilot projects studied, a smart metering system is, of course, composed by the smart meters. The common part is that in a pre-defined period of time, the devices send the consumption data. In projects, a cloud based service is used for the SM to send the data . In other cases, substations that work as data collectors are used to concentrate the consumption information from the SMs connected to it, usually a whole neighborhood, afterwards, the data from the substations is sent to a data center. In small SG, there is no central data center. The substations communicate with each other.

2.3.2 Smart Meters

Smart meters are devices that sense the energy consumption. They are installed in the costumer side, households or in industrial facilities, depending on the costumer nature. Playing a major role in the information subsystem, smart meters present several number of challenges in sensing

and analyzing[43]. SMs, more specifically, the Smart Mettering System has also the denomination of AMR(Automatic Meter Reading). In [34] the AMR is referred as the technology whose goal is to help collect the meter measurement automatically and possibly send commands to the meters.
As referred in the previous section, the main function of a smart meter, and all meters, is sense the consumption in the costumer side. The feature of sending their data, allocate and aggregate the information that comes from many meters allows a company to remotely read the consumers' consumption at each household, without the need to actually go to the premises and without notifying the costumers[56]. Jorge Vasconcelos [18] enlightens in his work the potential benefits of the smart meters, for example, the potential benefits for customers are customer awareness and energy saving, more accurate meter reading, billing, better service quality, greater tariff variety and flexibility, improved conditions for vulnerable customers, easier comparability of offers and it is easier to change supplier. [34] states some benefits of the smart metering system: Real time pricing, power quality measurement, automated Billing, Load management„ Remote Connect/Disconnect, Outage notification and Bundling with water and gas.
Privacy and security are important concerns when dealing with the sensed information. There are many privacy issues considering that external parties access the consumer energy consumption. Some are authorized parties, but there is a risk of an unauthorized access of this data, leaving to some security and privacy dangers. For example, by analyzing the data, one could determinate which devices are plugged in at some specific time, giving for example information about if there is people in home or note. Many pieces of work propose solution to securely store this sensible information. Although privacy and security are out of the scope of this work, it is important to mention this point.

2.3.3 Smart Grid Projects

So far there aren't standards to realize the Smart Grid, as it was aforementioned, not even a complete and specific definition about what is a Smart Grid. Even without a specific definition regarding the standard model and communication, there are common concepts that are well accepted and visions that are transversal . The introduction of communication and information technology into the grid, the idea of a consumer that is not only a consumer but also a small producer that can supply the grid and itself with electrical energy, the remote control of the components like electrical cable and station and more, are ideas that seem to be features that all future SG will have. In order to understand how it is currently the status of the SG and the directions it will take in the future, we analyze several pilot projects and companies that are know a days trying to implement the new grid.

2.3.3.1 Opower 4

Opower[54] iis a company that promises to help costumers to reduce their energy consumption. They provide a cloud based service to gather data regarding the costumers information about their energy consumption, and using big data and behavioral science they provide reports to the costumer with their consumption history in the time period that report is about. Also, the reports give tips and advices where the consumer can reduce the energy consumption, and therefore, reduce the energy bill. One of the version of the promised platform, one of the most recent, is called Opower 4. Opower 4 works as a service platform, is a *Software as a Service* platform. The model is like the general model for the SG Information subsystem described in section 2. Households using this services have a smart meter installed, every 15 minutes, the device send the data to a cloud through a cloud based service. The collection of the data is only made in one point, in a Data Center that concentrates all data. There is no reference regarding the number of data centers that the company uses. Big Data is performed in the data. Mainly, as it was stated, the platform exists for billing proposes and to raise awareness in the costumers so that they reduce their consumption with reports that have statistics of each household. For example, one costumer comparison with the neighborhood and what devices are consuming more or less.

2.3.4 DEHEMS Project

The DEHEMS project [55]t is an infrastructure that reasons about the household's energy behavior and tests various persuasive techniques effectiveness. The system receives energy information from several sensing devices, including the ones that sense electrical consumption. The special feature about the DEHEMS project is that it uses Informix TimeSeries [52] that is a builtin feature of Informix, a database type of IBM, that adds support for managing time series (timestamped) data, this feature is specially important in the management of data regarding the reading of the meter.
This project operates in the distribution grid. The model also includes de household, where several sensors are installed, not only for electricity, but also for gas and water. Each sensor, with a 433Mhz Radio, takes reading every 6 seconds and sends it to a DEHEMS Gateway that aggregates all the information about the house. The data of all the Gateways is concentrated in a Informix database so that big data operations can be performed on it.
The goal of this project is the same as the aforementioned project, raise awareness in the costumers by generating statistics abou each household consumption(CO2 emissions, cost of the energy, history of consumption and comparison with the other consumers) and send it to the consumer.

2.3.4.1 Pecan Street Project

The Pecan Street Project[53] is a research project / is researching a project?? in Smart Grids by Pecan Street Inc., a University of Texasbased research organization. Started in Austin an then expanded to other cities and states. The focus of the research is mainly in the information subsystem of Smart Grid. One of the project goals is to understand how to lower the carbon emissions by learning how energy is being used among homes. But understanding the â€œhowâ€ is only half of the challenge: Pecan Street also seeks to understand what homeowners need in order to manage their energy use.
As the other, it operates in the distribution grid and it works in a similar way as the DEHEMS project. In each house, there are several sensors installed to sense the consumption in each device, for example the Air Condition System. The sensors send their data every 2 seconds to the gateway and then to the smart meter every 15s, the meter sends the collected information from the gateway to a data center every 15 minutes. In the gateway it is performed an estimated average of all devices connected to a sensor. The consumption data is in the end used for statistical analysis to produce results about the consumer energy consumption. With this information, the Pecan Street Project staff pretends to lead their costumers to use their energy more efficiently.

2.3.4.2 Smart Meter Data Stream in the Cloud

This is a solution proposed to handle the SM data in a distribution electrical grid which is in [36] for real time pricing. The simulation considers a set of 1 million meters connect via TCP/IP to a data center/Cloud. Every second, each SM sends a package containing information about the electrical consumption of an household. The cloud model is composed by layers. Since every moment a new package arrives, it is like a stream, so several stream tasks are created in the lowest layer to handle the incoming data. In the upper level, within the cloud, aggregation tasks are created and they work in parallel to handle the information that comes from the lower level, the stream tasks. As the traffic increases or decreases, more aggregation tasks are created or deleted accordingly. In the paper simulation, 2 aggregation task were created. In the highest lawyer there is one real time pricing task that has the role of updating the energy price according to the amount of the energy consumed.
This paper offers a solution to handle smart meters data to provide a realtime pricing policy to balance the demand of energy and also to continuously monitoring of the meter, mainly to prevent outages and blackouts during peak time.

2.3.4.3 Inovgrid/InovCity

Inovgrid[46] is a project powered by EDP, Energias de Portugal, that pretends to modernize the portuguese electrical grid, more specifically, the distribution grid, in other words, the project aims to transform the traditional grid into a smart grid by adding information and communication technology. This is still a pilot project, there is no mass scale attempts yet to fully implement.So far, there is only a pilot project called the InovCity in the city of Évora that consists of a smart grid small experiment, with the installation of several smart meters and sensors in some of Évora households.
In further detail, the InovCity model can be explained by dividing the grid into three smaller networks: a Home Area Network(HAN), a network in each house, whereas each device has a sensor that communicates with the smart meter installed in the house. In the set of devices that composes the HAN, electrical vehicles are also included. Local Area Network, a set of households, a neighborhood connect to a DTC/substation that communicates through the electrical cables, PLC Prime and LMS protocol. Finally, the wide area network that embraces all the other minor networks
In terms of number, in InovCity 300 000 Smart Meters and 300 DTC/Substations were installed. The Smart Meteres communicate the consumption of an household every 15 minutes to a Substation which therefore communicates to other substation and finally to a central facility. Each substation has the capability of performing data analysis and process data function, so, depending on the type of analysis, the substation can perform it locally. Basically, the whole collection of data works in an hierarchical way, the data is collected in every smart meter regarding the information about every device with a sensor, a substation aggregates information about the houses connected to it, the upper level of substation collects the information of the other substation connected to it in lower levels, and finally the central facility concentrates all, working as a ”sink”.
There are 3 goals the company claims to achieve with the InovCity architecture. More energy efficiency by raising awareness in the clients with detailed information about their consumption. Increase Operations efficiency and reduce its costs by remotely perform any needed operations from a central station instead of doing it locally. Finally, commercial benefits by having a real time consumption instead of an estimated one and more accurate control by having a real time alarm of a failure in a SM..

2.3.4.4 PowerMatching City

PowerMatching city[51] is a pilot project of a self sustainable micro smart grid implemented and tested in Hoogkrek, a town in the north of Netherlands. Opposite to the other examples, in this case there were grid in considerable size and they were more focused on reducing the consumption and improving the performance of the grid, in this case, the goal is to create a selfsustainable city when it comes to energy consumption. In this city, the costumers can buy and sell their energy. They buy it from a market that is composed by small producers in the city that can generate energy through renewable sources, and, therefore, they can sell it too. This way, the city becomes independent from major electrical companies.

In PowerMatching, each household has a smart meter that has the information about the consumption of each device and also about the energy produced. The information is sent by the smart meter to a coordinator/data collector through an VPN communication infrastructure. Also, an ADSL communication channel is used between the coordinator and the houses connected to it to prevent the occurrence of faults. The coordinator is responsible for collecting the information about the energy consumed and produced, and generating the prices accordingly, working as a market. This idea can scale adding more coordinators that are connected and communicate among each other to work as a whole market.

In the implementation in Hoogkrek, 25 Household had an SM installed to the PowerMatching city network with, at least, one collector/coordinator. Data is collected in every coordinator station, therefore, the process in the station works in a lawyered process, the lower levels receive and collect the information, send it to the upper level in the bif format, to buy or to sell energy that are communicated to every house connected. It is not mentioned what is the interval by which the prices of the energy are changed, but we can admit that it is not about the time, but in terms of supply and demand as in all markets.

Also, the system contemplates three web portals for data: user Portal where the user can check her/his stats about energy consumption/production, operator Portal which is mainly used for operations(monitoring and detecting failures, data analysis that generates reports used mainly for research proposers and for the development of the project.

The main goal is to organize a market whereas all community is independent from global companies. The measured data is used and aggregated mainly for price proposes, i.e., following the market rules, the data is used to give selling and buying prices. Also, in terms of singular house, the data is used also by the system to buy or sell the energy. If a house has low energy supply plus high demand, the systems should buy it, on the other hand, if there is a surplus, ideally it should be sold it to the market.

There is other examples of other smaller SG models that represents more and idea. Keita Suzuki *et al* [48] presents a particular case in a office building in Japan(Heating ventilation and air conditioning facilitie,HVAC) where existis the need to aggregate power curtailments from hundred or thousands of distributed HVAC facilities. Several smart meters where placed, connected to a Gateway that receives the consumption data for daily or monthly billing. The Gateways are connected to a central ADR, Aggregation Cloud, which aggregates all the consumption.

Another work using a *de-centralized* way is in Rottondi *et al*[38]. The smart meters generate the energy consumption measurements, the Gateways securely aggregate the metering data and external parties access the aggregation results. Each meter is directly connected to a Gateway, receiving data from a limited number of meters. At regular time intervals, 15 min in this case, the meter generate a measurement and send it to the Gateway.

Chapter 3

Distributed Data Aggregation

3.1 Definition

Data aggregation is a technique that , on its basis, consists in reducing the amount of data collected, reducing the resources needed to process it. According to [33], data aggregation is considered a subset of information fusion, that aims at reducing the handled data volume. A more precise definition is given in the same report:

Definition 3.1. An aggregation function f takes a multiset of elements from a domain I and produces an output of a domain O.

$$f : \mathbb{N}^I \to O$$

The order in which the elements are aggregated is irrelevant and a given value may occur several times. The main goal of data aggregation, "*the aggregation function aims to summarize information. The result of an aggregation takes less space than the inputed multiset (element from* $\mathbb{N}^I$ *)*".
Distributed data Aggregation or *in-network* aggregation tends to distribute de computation of an aggregation funtion among several nodes in the network. In contrary of a *centralized* architecture, where a central node compute all the data and performs the aggregation function, a *de-centralized* aggregation distribute the data, hence the effort to compute the aggregation function is reduced.

3.2 Wireless Sensor Network

Wireless Sensor Networks(WSN) are *ad-hoc* networks composed by tiny devices with limited computation and energy capacities. These tiny devices, sensors, are called tiny because of their low capability of computation, communication and storage. The WSN low-cost sensors monitor physically on environmental conditions, such as temperature, sound, vibration, pressure, monitor pollutants and to cooperatively pass their data through the network to a main location(sink node) via multi-hop wireless links[42] or to their peers.

WSNs act under severe technological constraints: individual sensors have severely limited computation, communication and power(battery) resources and need to operate in settings with great spatial and temporal variability.The ad-hoc nature of a WSN implies that sensors are also used in the network infrastructure, i.e., not just sending their own data and receiving direct instructions but also forwarding data for other sensors. Modern networks are bi-directional, enabling control of sensor activity but some WSN could not have bi-direccional communicatin due to low computation power of the sensors. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance.

Today, WSN networks are used in many industrial and consumer applications like industrial process monitoring and control, machine health monitoring and so on. Some of WSNs requirements are: large number of nodes, low energy use, network self organization, collaborative signal processing and querying ability.

WSNs are becoming increasingly popular in many spheres of life [9], they also have the capability of forming the sensor web services which can be considered as an extension of the future internet towards smart devices, Internet of Things(IoT)[42].

3.2.1 WSN and Smart Grids

Considering the overall appliances, WSNs has also several applications in the SG. Furthermore, the AMR could be considered as a specific example regarding the appliance of WSN. It can be implemented the proposed WSN solutions for data aggregation in AMR .

Recently, WSN has been widely recognized as a vital component of the electric power system[40]. WSN contains a large number of low cost and multifuncional sensor nodes which "*can be of benefit to electric system automation application, especially in urban areas*"[29]. The collaborative and context-awareness nature of WSN brings several advantages over traditional sensing including great fault tolerance, improved accuracy, larger coverage area and extraction of localized features [40]. Sensor nodes can monitor the overall network.

WSN could apply to several features in the SG: basis measurement, smart voltage sensors, smart capacitor control, smart sensors for outage detections and weather condition sensors, distributed generation, smart gird storage and, referenced before and more importantly for this work, WSN for AMI(Advanced Metering Infrastructure) or AMR. A specific example is in [40] where a WSN could apply

perfectly to a household or House Area Network(HAN) . ZigBee is a communication technology often choosed in Smart Grids due to its reliable wide area coverage and predictable latencies, it is also a suitable choice for a Local Area Network such as a household or a neighborhood. As a example in [40], a WAMR(Wireless Automatic Meter Reading) can determinate real-time energy consumption of the customers by sensing each device that has a wireless sensor on it. The smart meter within the household perform an interface that translates, summarizes and aggregates data of power usage and presents it to the power utility.
Other examples of WSN appliances in SG are found in [40]. WSN could apply in Power Delivery and in Power Generation as well since the sensors can monitor the deliver systems, in the first case, and monitor the energy generated in the second case.
Although very similar, there are some differences between WSN and Automatic Meter Reading. Such diferences are stated in [34]. For example, individual consumption measurments must preserve its information. In WSN, sink doesn't care about inidivudal data but in AMR, aggregation nodes must preserve the unique measurments, plus, the meters must have a unique indentifier that links the smart meter to a household/costumer/producer. Futhermore, Smart meters have fixed positions on contrary to some WSN, base stations may need to disconnect/connect to a specific costumer. Even in security, there are some differences. The main security concern in WSN is to preserve the privacy of data, in SM, altough privacy is an important issue, integrety of data is the main concern.
WSN, even considering the diferences to AMR, it provides a variety of solutions and gives some insight to understand and comprehend the problem of distributed aggregation in AMR since WSN is a well studied subject.The topology we can find in some WSN can apply to the ones in the AMR. So, even with differents communication infrastrutures or different computation powers, from the topological view, both networks are very similar.

3.3 Distributed Data Aggregation Algorithms

Distributed Data Aggregation Algorithms are protocols used to compute aggregation function in a *decentralized* way. They are used and more suitable when the network lacks a node or component that have the computonational capacitie to process large ammounts of data. The case of WSN , were all the nodes are tiny devices with low storage and proccessing capacity.
In [33] is also presented a simple taxonomy of the existing algorithms that performe distributed data

aggregation. First it is analyzed the algorithms from the communication perspective, i. e., the routing protocols and the intrinsic topologies, afterwards, it is analyzed the computation issues, how the aggregation functions are computed by the algorithms.

3.3.1 Communication

3.3.1.1 Hierarchy-based approaches

Traditionally, existing aggregation algorithms operate on a hierarchy- based communication scheme. This is *structured* communication scheme. It is required to know in advance the topology of the network. A hierarchy communication tree is constructed, with several levels of nodes. In the root of the tree there is a main repository of all data, denominated as sink. Besides the sink, other special nodes can be defined to compute intermediate aggregates, working as aggregation points that forwards their results to upper level nodes. There are generally two main phases, *request* phase, corresponding to an aggregation request spreading through all the nodes, an the *response* phase where all the nodes respond to the request sending their aggregation results. Some specific examples of these kind of communication are presented.

TAG The Tiny Aggregation algorithm that suits for ad-hoc networks described in [1]. This algorithm requires the previous creation of a tree-based routing topology, and the continuous maintenance of such routing structure in order to operate over mobile networks. TAG provides a SQL-like declarative language to the users. The algorithms consists of two main phases, the *distribution* phase, i which a aggregation query is disseminated through all the spanning tree, and a *collection* phase, where the values are aggregated. A waiting time is required to conclude this two phases.

DRINA DRINA is a cluster based protocol described in [49] that denominates the algorithm as *lightweight and reliable routing approach for in-network aggregation in wireless sensor networks.* Considers four roles for each node: textitcolaborator, a node that detects an event and reports the gathered data to coordinator one,*coordinator* a node that also detects an event and collected all the gathered data sent by collaborator nodes, aggregating them and sending the result to upper levels,*sink*, a node that receives all the data from a set of coordinators and finnaly a *relay*, a node that just forward the data towards the sink. The algorithm works in three phases: First the hop tree from the sensor nodes to the sink node is built. In this phase, the sink node starts building the hop tree that will be used by Coordinators for data forwarding purposes, second the cluster formation and cluster-head election, third phase is responsible for setting up a new route and updating the hop tree.

DAG An aggregation scheme fro WSN is proposed in [14] that aims to reduce the number of message losses. For each node, multiple parents are set but only one is chosen to aggregate intermediate values. The most common parent's parent(grandparent) are chosen among the list received as the destination aggregator. Messages are aggregated if the receiving node corresponds to the destination, forwarded if correspond to its parent or discarded otherwise.

Sketches Algorithm proposed in [4] that uses small sketches. Based on the probabilistic counting sketches technique that estimates the number of distinct elements in a data collection and it is described in further detail in [16]. Like other algorithms of this type, is uses two phases: the sink propagates the aggregation request across the network and then the results are collected back to the sink. In the first phase, all nodes compute their distances to the root, in the second phase the partial aggregates are computed across the routing structure, using the adapted counting sketch scheme, and send to the upper levels in successive rounds.

I-LEAG Cluster-based aggregation approach designated as I-LEAG is in [11]. The routing structure of this algorithm is composed by a hierarchy of clusters or partitions. A single pivot is designated for each cluster and the root is the pivot of the upper level cluster. This structure can be consider similar as we can see in networks with *super-peers*, but organized in a tree structure. The algorithm works as follows: each cluster check local conflicts that are reported to the pivot, then the pivot computes the new aggregate and multicast the result, each node must forward the received result to the nodes outside the cluster.

Tributary-Delta This approach mixes the traditional use of tree and multi-path routing schemes, dividing the network in two routing regions: *delta*(multipah) and *tributary*(tree). Use tributaries in regions with low rate of message losses to take advantage of traditional tree schemes and delta in regions with higher rate of message losses(mostly regions near the sink with the aggregate of several nodes).

3.3.1.2 Gossip-based approaches

This type of approach is referred as an *unstructured* approach, contrary of the aforementioned *structed* approaches. In this type of scheme there is no previous knowledge of the topology of the network or any specific structure. The information or messages are commonly disseminated across the network without following any specific topology, the information it is passed node to node, or nodes, like a infectious disease or a gossip,i.e., an "infected" node sends a message to a random subset of nodes. This type scheme tends to allow a robust(fault tolerant) and scalable information dissemination all over the network[33]

Push-Sum Protocol Push-sum protocol is described in [3] and it is a gossip-based protocol. [33] describes the algorithm function : along discrete times t, each node i maintains and propagates information of a pair of values $(s_t i, w_t i)$ where s represents the sum of the exchanged values and w the weight associated. In each iteration, a neighbor is chosen uniformly at random and half of the actual values are sent to the target node and the other half to the node itself. Upon received, the local values are updated, adding each value from a received pair to its local component.

3.3.1.3 Hybrid approaches

Hybrid approaches propose a solution that merge both hierarchic and gossip-based approaches, using the high accuracy and efficiency of the hierarchic based schemes and the robustness of the gossip approaches. In the disadvantages of one approach , the other one has it as an advantage, Hybrid approaches aim to merge the advantages of both schemes to eliminate both disadvantages.

Chitnis et al, 2008 Chitnis et al.[15] proposed an hybrid approach, using TAG as an hierarchy-based approach and Push-Sum as a gossip-based protocol. This hybrid approach divides the network node in groups. Inside each group, a gossip-based protocol is used. In each group, a leader is elected to further perform a hierarchic communication with other leaders nodes regarding the aggregation results from the gossip group.

3.3.2 Computation

3.3.2.1 Hierarchical

The input is separated into groups so it can be computed in a distributed hierarchical way. It depends on the previous formation of a communication structure such as tree or cluster. Some node work as *forwarders*, just forward data to upper levels of the hierarchy, and others work as *aggregators*, apply the aggregation function directly to all received input and then works as a normal *forward* node. This class of algorithms allows any decomposable function with high accuracy without the presence of faults. Algorithms of this class were aforementioned.

3.3.2.2 Averaging

This class of computation scheme is based on an iterative computation of partial aggregates, where all nodes share their results among the network and all of them contribute for the final result. This scheme provides high accuracy, considering that all nodes converge to the same result.However, in order to converge to the correct result, the algorithms must respect an important principle commonly designated as ”mass conservation”. [33] describes ”mass conservation” as an invariant, stating that the sum of the aggregated values of all network nodes must remain constant along time. One example of algorithms of this class, is the ones with gossip base communication scheme, since the results of the aggregates could be share randomly with the neighbor nodes. Due to its nature, Averaging algorithms tend to be highly robust, i.e., tolerant to faults on opposite of the structured algorithms. Decomposable and duplicate sensitive functions can be be computed in this class.

Push-Pull Gossiping Similar to the aforementioned *push-sum protocol*, the push-pull gossiping[6] performs an averaging process. This algorithm executes an epidemic protocol to perform a pari-wise exchange of aggregated values among neighbor nodes[33]. In periodic intervals of time, a node send its value to a randomly selected node and waits to receive a result back, the response from the selected node. Afterwards, an average with the new value and the present value its performed in order to calculate and store a new one. When a node receives a value from another node, the same process is performed, send the current value and calculate a new one from the average of the received value and the current one.

DRG(Distributed Random Grouping) This approach [13] randomly creates groups across the network in which aggregates are successfully computed. There are three modes a node can perform:*leader,idle* and *member* which corresponds to three phases. First every node is in *idle* mode, then every node broadcasts a Group Call Message, pretending to be a group leader(with a pre-defined probability associated) and waits for members. The nodes who receives the group call, responds to the first one received with a JACK(Joining Acknowledgment) tagged with their aggregated value becoming a member of the group. Finally, the *leader* gathers all the aggregated values, computing the aggregation function($AVERAGE$) and broadcasts a Group Assignment Message with the final result. Every group member waits until it receives the result from the leader to update its local value and then returns to *idle* mode.

3.3.2.3 Sketches

Algorithms based on the use of an auxilary data structure with a fixed size that holds a *sketch* of all network values. Input values are used to create *sketches* that aggregated across the network, using specific operations to update and merge them. The aggregation could be done using multiple paths. This type of algorithms enables operations of order and enables duplicate insensitive. The computational cost of this class depends mainly on the resources used to produce the result by the estimator and the complexity of the operations to produce the *sketches.* This kind of algorithms tend to be very fast, depending on the dissemination protocol used to propagate the sketches, but lack accuracy because they are based on probabilistic methods.

RIA-LC/DC Algorithm proposed in [16], a multi-path routing aggregation approach. The algorithm consists of two phases. First an aggregation request is sent by the sink throughout the whole network, creating a multi-path routing hierarchy. Second, starting in the lower levels, each node generates a *sketch* correspondent to its current state and sends it to the nodes in the upper level. The node that receives the *sketch*, creates a new one combining tis current value and the received *sketch* and sends it to the upper node until the top is reached where the sink computes the aggregation estimate.

Extrema propagation This approach reduces the computation of an aggregation function[33]. A vector x_i of k random number is created at each network node i. Random numbers are generated according to a known random distribution, using the node initial value as an input parameter. The execution of the algorithm *"consists of the computation of the point wise minimum between all exchanged vectors"*[33]. At each node, the obtained vector is used as a sample to produce an approximation of the aggregation result. This algorithm is focused on obtaining a fast estimate, rather than an accurate one.

3.3.2.4 Digests

This class of algorithms allowd the computation of more complex functions like median or mode than the normal aggregation function such as SUM or $AVERAGE$. This algorithms produces a *digest*, data structure with a bounded size that holds an approximation of the statistical distribution of input values in the whole network, that summarizes the system data distribution, an histogram. The accuracy of this class of algorithms depends mostly on the quality and size of the obtained *digest*. Usually it requires more resources.

Q-Digest This aggregation scheme allows the approximation of complex aggregation function in WSN is proposed in [7]. It uses an hierarchical routing topology to build and disseminate quantile digests. Each node maintains a quantile digest of the data availabe, which is built in a button-up fashion by merging received digest from lower nodes(children nodes). This new quantile digests are compressed according to a specific compression factor. Aggregation functions are computed by manipulating and traversing the quantile structure according to a specific criteria.

Equi-Depth Gossip-based approach described in [2]. The scheme executes a gossip protocol and merges specific function on the exchanged data. Each node keeps a list of k value or *digests*, initially set to its input value. Each node randomly chooses a neighbor to exchange the digest to merge with its own. This round is executed several number of times, producing an approximation of the network distribution of values. There are four merging techniques *swap, concise, equi-with histograms* and *equip-depth histograms* that are detailed in [33].

Adam2 Adam2 is a gossip based algorithm to estimate the statistical distribution of values across a decentralized system[26]. Each node can decide to start an instance of Adam2 where each instance is uniquely identified by its starting node. The starting node i initializes the interpolation set H_i(composed of k pairs of values (x_k, f_k) where x_k represents an interpolation point and f_k the fraction of nodes with value less or equal to x_k). The interpolation is initialized by setting f_k to 1 if the node attribute reading v_i is less or equal than the corresponding interpolation value x_k, 0 otherwise. Node stores a set of interpolation points for each running algorithm instance. A new node that, learning about the new instance, performs a initialization and then starts participating in the protocol. The sets are exchanged like push-pull, the sets are merged by averaging the fraction at each interpolation point. After a predefined number round the CDF is approximated by interpolating the point of the resulting set.

A overall taxonomy table is presented in [33]

	Advantage	Disadvantage	Requirements
Hierarchical	- accurate (without faults); - very efficient (messages);	- result at a single node; - not fault-tolerant;	- specific routing structure (e.g. spanning tree);
Sketches	- very fast; - result at all nodes; - fault-tolerant;	- less accurate;	- local knowledge of neighbor IDs, or global UIDs; - source of randomness;
Averaging	- accurate; - result at all nodes; - fault-tolerant; - churn support;	- less efficient (messages);	- local knowledge of neighbor IDs;
Sampling	- efficient (messages);	- not accurate - result at a single node; - not fault-tolerant	- global UIDs; - source of randomness;
Digests	- computation of complex aggregates; - result at all nodes;	- less accurate; - resources needed (e.g. larger messages);	- local knowledge of neighbor IDs;

FIGURE 3.1: Summary of the characteristics of main data aggregation classes

3.4 Distributed Data Aggregation in WSN

Distributed Data Aggregation in WSN is an widely study subject, with several works and proposed solutions. Distributed aggregation adquires a special importance in WSN, since the sensor are low resources devices so the effort distribution is quite mandatory. The aggregation techniques reduce the amount of data communicated within a WSN and thus conserve battery power [9].Periodically, as measurements are recorded by individual sensors, they are been collected and processed to produce data representative of the entire WSN. An natural approach is consider that the sensor send the measured data to special sensor nodes, i.e., aggregator nodes [9]. In *in-network* aggregation nodes forward the aggregated data to a sink that store it.

An example of in-network aggregation in WSN is in [9]. In this model, it is assumed that all nodes

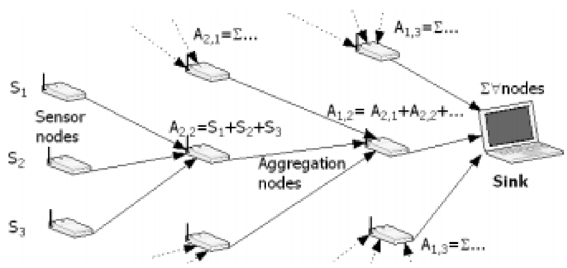

FIGURE 3.2: Principle of in-network aggregation

are potential aggregators and that data gets aggregated as they propagate towards the sink. The aggregation is set as an obligation of being simple not involving any expensive or complex computation. The aggregation requires all sensors to send their data to the sink within the same sampling period so there is a need for a global so that all node can synchronize. Another study is in [5], where a special kind of distributed aggregation is proposed, *Concealed Data Aggregation*. This type of aggregation is defined as an approach than promises the combination os end-to-end security and *in-network* aggregation. In [12] it is assumed a general multi-hop network with a set $S = s_1....s_n$ of n sensor nodes anda a single base station R. The aggregation is performed over an *aggregation tree* which is the directed tree formed by the union of all the paths from the sensors nodes to the base station. Another WSN distributed aggregation scenario is presented in [20]. The network model consist of a n sensor nodes and one base station that is also called a sink. Each sensor node can send or receive data to or from all directions. It is assumed that all nodes have the same transmission range for simplicity. A node can either receive or send data at a time and it can receive a data packet correctly when it hears only this packet at that moment.

3.5 Smart Metering Aggregation Model

There are two main architectures for smart metering considering data aggregation are *centralized* and *distributed* or *decentralized*[43]. In *centralized* fashion, the meters just sense the data, afterwards, it is sent to a central aggregator with higher computation power that holds a central database. In a *decentralized* way, the aggregation role is distributed among several meters, not all of then. This type of aggregation is also called *in-network* aggregation [5][8]. The aggregation node in this scheme communicate the calculated energy consumed to an appropriate party such as a energy producer. Typically, this communication occurs once per billable period [43]. As introduced before, the architecture chosen for this work is *de-centralized* due to the nature of the aggregation algorithms.
Several examples of SG projects and models were given in chapter [**chap:sg**]. Usually, the *centralized* approach is composed by a cloud service provided to the costumers to store their consumption data. These approaches make use of a data center and uses this architecture so they are capable of storing big quantities of data and perfome Big Data techniches on it, providing useful information for both energy companies and costumers.
One example of the *de-centralized* architecture for aggreagation is in the PowerMatching City of Hoogkrek. Because the goals are different than providing statistycall information regarding the costumers consumption, in this case the goal if forming a market in microgird that enables the household to be self-sufficient in therms of energy. They use the *de-centralized* archictecture to form several points

of this market wherer households sell and buy their energy, the nodes that concentrates the offers communicate with each creating the city energy market. Another example is in Rottondi *et al*[38], the overall scheme is presented in 3.3, where a set of meters are connected to gateway sharing information, the central station only works to set the aggregation rules.

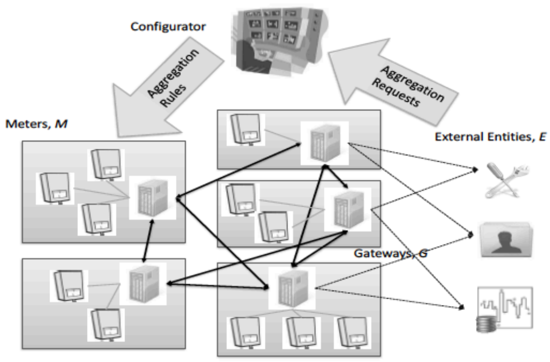

FIGURE 3.3: The functional nodes of the architecture

Chapter 4

Network Topology

In order to test the aforementioned algorithms, it is necessary to build and evaluate a topology that represent an SG network. The algorithms work with data, their function is to gather and collect it and compute aggregation functions in a distributed way. Therefore, our focus to build a testing network is the SG information subsystem. As it was stated in the chapter2, the AMI system is part of the information subsystem where meter information, the one we are interested in, is treated. So, our topology must represent an AMI Network, with the smart meter data used to compute the aggregation function. We are interested in the collection part of that data and the process made to update the prices accordingly to production/consumption of the network. In the following section, we will characterize more precisely what scenario we assume and which communication assumptions we make.

4.1 Communication

When building a topology or a network, is important to know and consider the communication technology used and what are the assessments we make. In AMI networks, there are many options, both wired and wireless, for as many needs. A suitable choice for a Smart Meter data collector could not be a suitable choice for a central facility substation flow because of the different needs and requirements. Since in this work focused in the data collected by the aggregators and the communication between s, the choice of a communication infrastructure must be among the technologies used in the Low Voltage grid, i.e., where the metering is performed. Therefor, wee assume an AMI network that uses a PowerLine Communication Infrastructure, used, for example, in EDP InovGrid[46] and Iberdrola pilot projects [41]

4.1.1 PowerLine Communication

PowerLine Communication(PLC) is a communication technology which uses the installed electrical cables to transmit data. It has been the first choice for communication with the electricity meter due to the direct connection with the meter and successful implementation of AMI in urban areas where other solutions struggle to meet the need of utilities[32].

The typical application of the PLC is to connect the smart meter to the data concentrator. The SM communicates the consumption or production data to the collecting device through the power line. After the collection from the meters, normally, the data aggregator send the stored data to a central facility, that is owned in most cases by the electrical company, through a more fast technology like optic fiber. In Europe the majority of the transformers serves 200 customers or more, so the data collectors are located in the LV side, in the transformers and using PLC to communicate with the meters. In the USA, the concentrator or aggregator is often in the MV side of the grid, due to the low number of end points per transformer. The large number of end-points per transformer in Europe does not require to locate the concentrator up in the substation or on the MV side, so it can be conveniently located on the LV section of the grid. REMPLI project tells us that unlike solution based on ZigBee os Wifi, PLC-based AMI have a proven track record of being able to avoid network congestion when cooperative schemes are employed[30]

As it was stated before, one of mains advantages of using a PLC based communication infrastructure is the cost to deploy. The fact that the cables are already installed makes PLC a more appealing technology compared to other wired solutions. Considering only the costs, PLC can even compare to the wireless proposed solutions[32].

The standardization efforts on PLC networks, the cost effective, ubiquitous nature and widely available infrastructure can be the reason for its strength and popularity. It is also well suited for urban areas for smart grid applications, besides AMI application like metering, PLC can also be used to monitoring and control because, once again, the involved areas are already covered.

Although it popularity, PLC still represents some disadvantages mainly because of the nature of the channel and because of some technical problems. The communication channel is a harsh and noisy environment that makes it difficult modeled. Furthermore the network topology, the number and type of the devices connected to the power lines, wiring distance between transmitter and receiver, all, affect the quality of signal that is transmitted over the power lines. However, recent modulation techniques and technologies are overwhelming the mentioned difficulties, mitigating the noisy environment by reducing the data rates and the bandwidth.

4.2 Data Collector/Aggregator

Data collectors or data Aggregators are devices installed in the Low Voltage side of the Grid, in the European case, or in the Medium Voltage side of the grid in the American case, because the ratio SM per transformer it lower in USA than in Europe. They normally serve as data storage where a set of SMs are directly connected to it sending the consumption or production of a household.

Data Collectors normally serve as a storage but this is not a rule that apply to all, there are cases where the data aggregators perform more operations. In the PowerMatching City project case, the data collectors receive data in bid format, to buy or to sell and then calculate the price of the energy to others buy and sell. They work in market way, not only collecting information, but also making decisions about prices. In the case of InovGrid, the data collectors just collect data. Other projects lack the existence of this devices, using instead a cloud based service, others use it inside the house to collect data from different electrical devices to know which ones are consuming the most. As it was presented in chapter 2, SG projects which don't require the use of any data collector propose a architecture that enable the smart meters to directly send their data to a data center through a cloud based service. These different architecture exist because there are different goals in the different pilot projects. The ones which use big data to perform detailed statistical report to raise awareness in the costumers require the maximum amount of data as possible. In the cases where the goal is achieve a ideal strategy to buy/sell energy, provide a reasonable Demand Side Management and useful information to the operation side don't require a big amount of data, so they use an mid level set of devices to aggregate the data.

Review this idea

4.3 Demand Side Management and Real Time Pricing

Demand Side Management refers to a serie of strategies and policies to load balance the demand efficiently in order to prevent very high demand peak hours, preventing the overload of the production and the occurrence of blackouts or faults. Although applied in the SG and AMI scenario, it is not a recent trend, demand side management has been considered since the early 80's [27].

One of the application of Demand Side Management in the Low-Voltage Grid it Demand Response(DR) which refers to the ability to make demand able to respond to the varying supply of generation that cannot be scheduled deterministically. DR is a means to alleviate peak demand and also to bring more awareness on energy usage to the consumer. It is believed that DR will allow a better control of peak power conditions, maximize the use of available power and increase power system efficiency.

A more specific example of how DR can be applied is by real time pricing which is one of the strategies to achieve and realize Demand Side Management. Real time pricing is a pricing policy, basically it consist of upgrading the prices accordingly to the demand and the supply, in other words, the prices vary as the demand increases or decreases along periods of time. This policy tend to persuade the costumer to reduce their energy consumption during the peak hours, since the price is higher, and increase slightly outside the peak hour, when the price is lower.

In [28] is realized a research regarding the aforementioned policies in UAE. The experience states that costumers react better to peak time adaptive tariff. Also, it is stated that with these pricing policies, costumers are more aware of their consumption and, for consequence, more willing to reduce their consumption during peak hours. The authors also point that consumption demand is reduced during the peak hour while balancing it.

[50] proposes an optimal pricing policy for aggregators. The aggregator "buys" the electrical energy from the supplier, and regarding the necessary demand, defines the price. The time is spited in K time intervals, in the k interval the aggregator must make decisions based on the $k-1$ interval information about demand and energy available from the electrical company. The goal is to maximize the profit, but it can also be applied to reduce the demand during peak hours if the decisions are made for reducing the demand when the supply is low and the demand high. The behavior is similar to the PowerMatching city, but the aggregator have instead information about bids to buy or sell energy.

We assume a real time pricing scenario, where each aggregator should have a behavior similar to [50] and the PowerMatching city, i. e., the aggregator make decisions regarding the energy price based on the supply from the electrical company, on the total consumption from the costumers and also from their production by renewal energy sources. The pricing policy followed is outside the scope of this work, we only consider the problem of a given time, evaluate the overall consumption or production, so the aggregators can make the pricing decision accordingly.

4.4 Case Study

In this work we do consider the existence of data collectors. We don't make any assumption regarding the number of households connected per data collector. Instead we assume that a certain number of smart meters are connect to a data collector, therefore, the device contains information regarding the connected houses consumption and production.

Definition 4.1. For a data collector D, a set H of households with a smart meter are connect to it. H is defined as

$$H = \{h_1, h_2, h_3, \ldots, h_n\}$$

Where n is the size of H and $n >= 1$

Definition 4.2. At a given time k, a data collector D_k contains a consumption C_{D_k} where is defined:

$$C_{D_k} = \sum_{i=1}^{n} c_i$$

where n is the size of the set H and c_i is the consumption of the i household connected to the data collector D_k at time k

Also, we assume that each data collector is located not only on the MV/LV transformer, but in every substation of the network which contains electric equipments (e.g., rails, bus bars, electrical switches, etc). Although we don't cover the security aspects of the data collectors, not only information security,privacy and integrity, but also physical information, there are in the current projects security measures to avoid the violation/steal of the data collector and encryption techniques and ciphers are also applied to provide security information guarantees. Once again, security issues are not cover in this work.

In the InovGrid case [46] and in Iberdrola projects [41], the data collectors only communicate the data to the electrical company, and not among each other. We took the example of the PowerMatching city, and other proposed SG architecture seen in 2, and consider not a hierarchical network composed by households, data collector substations, and the data center as a sink, but a network of data collectors that communicate with each other.Although they communicate the overall network consumption/production with the electrical company, they share their data in order to all the collectors in order to compute an aggregation function without the need of a central device.

To complete our scenario, we assume one more similar aspect to the PowerMatching city, but, instead of the data collector receive the data in a bid format to buy or sell energy, we consider consumption and production values. Therefore, the price of the energy should be calculate according to this 2 variables, as in a PowerMatching market. We will not cover the market aspect, the problem we focus on it knowing the consumption or production of a data collector network using Distributed Aggregation Algorithms.

4.4.1 Network Architecture

This work focus on the Low Voltage power grid network, with voltage $V_{LV} <= 10kV$, and evaluate a Smart Grid scenario. We start by choosing a Low Voltage power grid network sample and define a Smart Grid graph. For the low voltage power grid sample graph, we choose one from the work of [47] and define a Smart Grid graph.

Definition 4.3. A Smart Grid graph is a graph $G(V, E)$ such that each element $v_i \in V$ is either a substation with electrical equipment or a transformer os a physical power grid with a data collector installed, There is a edge $e_i, j = (v_i, v_j) \in E$ between two nodes if there is physical cable that enables a PowerLine communication connecting directly the element represented by v_i and vj

With the definition of our Smart Grid graph, we need to bring and specify some important properties into the graph definition. Each edge $e_i, j \in E$ represent a physical cable, enabling a PowerLine Communication between the data collectors. Our Smart Grid graph is a weighted graph, we consider channel capacity in bps a suitable weight for the edges. We calculate the capacity of the channel by the channel capacity function defined in [22] , assuming that all channel are Gaussian Channels. We also assume that the communication uses the CENELEC A band.

Definition 4.4. Channel Capacity is defined by the function M_{ij}

$$M_{ij} = \int_{B_1}^{B_2} C\left(\frac{S_T|H_{ij}(f)|^2}{N_0\Gamma}\right) \mathrm{d}f$$

Where $H_{ij}(f)$ represent the transfer function between nodes v_i and v_j, Γ is a gap factor to account for practical coding schemes, $B_1 = 9kHz$ and $B_2 = 95kHz$ are the lower and upper frequency of the CENELEC A band, S_T represent the power spectral density, N_0 is the background noise modeled as white Gaussian noise and $C(\gamma)$ is the capacity at signal-to-noise power ratio (SNR) of γ

We assume the same reference values in [22], $S_T = -60dBV^2/Hz$, $N_0 = -138dBV^2/Hz$ and $\Gamma = 10$. $C(\gamma)$ is defined as:

$$C(\gamma) = \log_2(1 + \gamma)$$

We assume that all the cables are of the type NAYY150SE, the cable lengths are the ones provided the samples used in [47] and the other physical characteristics are taken from [35].
The figure 4.1 represent illustrate the graph represent the network used for the tests.

4.5 Used Aggregation Algorithms

In this section is presented the aggregation algorithms used. In Chapter 3 there is only a brief description of all the Algorithms, here we formalize the algorithms used. The problem the algorithms need to solve is a evaluate the consumption of all the data collectors, so the algorithms tested are the

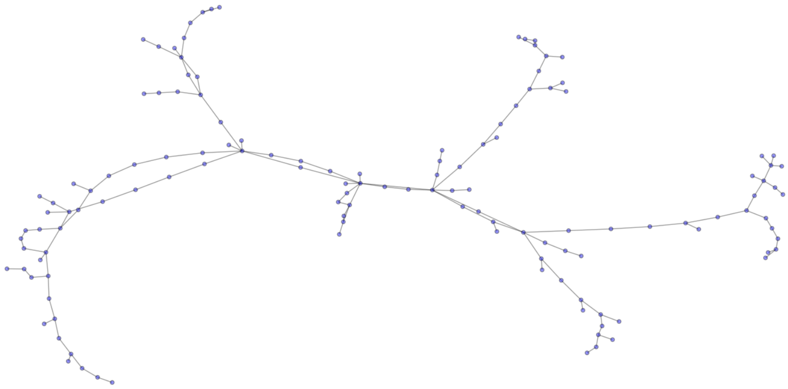

FIGURE 4.1: Smart Grid Graph

ones who allow us to do so. We choose the *Averaging* algorithms and based on *Sketches* because they are designed to evaluate the function *AVERAGE*, and most important, the function *SUM*. Other algorithms are more suitable to understand a distribution of values in a network or to evaluate its size.

4.5.1 Push-Sum

Algorithm 1 Push-Sum Algorithm

Let $\{(s_r, w_r)\}$ be all pairs sent to i in round $t-1$
Let $s_{t,i} := \sum_r s_r$, $w_{t,i} := \sum_r w_r$
Choose a target $f_t(i)$ uniformly at random
Send the pair $\left(\frac{1}{2}s_{t,i}, \frac{1}{2}s_{t,i}\right)$ to $f(i)$ and i (yourself)
$\frac{s_{t,i}}{w_{t,i}}$ is the estimate of the average in step t

In the Push-Sum algorithm [3], initially, each node generates a pair (s, w) where s is its value to aggregate and w its weight, initiated to one. At every round t, each node i evaluate s_r and w_r which are the SUM of all the pairs $(s, w$ send to i in the previous round. After evaluating (s_r, w_r), send half the $\frac{1}{2}s_r$ and $\frac{1}{2}w_r$ to a randomly choosed neighbor and to itself. The $AVERAGE$ in the round t is the SUM of all received s in the round i plus s_r divided to the e SUM of all received w in the same round i plus w_r. To evaluate the total SUM of the values instead of the $AVERGE$, the algorithm suffers one slightly difference. Every node starts with initial weight equal to 0, except for one node, which its initial weight is 1.

4.5.2 Flow Updating

Algorithm 2 Flow Updating

State Variables:
$f_{ij}, \forall j \in D_i$, flow, initially $f_i j = 0$
$e_{ij}, \forall j \in D_i$, estimates, initially $e_{ij} = 0$
v_i, input value
message-generation function:
$\text{msg}(i, j) = (f_{ij}, e_{ij}), \forall j \in D_i$
state-transition function:
for all (f_{ji}, e_{ji}) **received do**
  $f_{ij} \leftarrow -f_{ji}$
  $e_{ij} \leftarrow e_{ji}$
end for
$e_i \leftarrow \frac{(v_i - \sum j \in D_i f_{ij}) + \sum j \in D_i e_{ij}}{|D_i|+1}$
for all $j \in D_i$ **do**
  $f_{ij} \leftarrow f_{ij} + (e_i - e_{ij})$
  $e_{ij} \leftarrow e_i$
end for

In Flow Updating[19], each node i initializes its state variables, a set of pair (f_{ij}, e_{ij}) where $f_{ij} = 0$ and $e_{ij} = 0$, a pair correspondent to each neighbor, contain the flow and an estimate. Also, the node holds an input value v_i, the value to aggregate.

Alinhar isto

At every round, a node generates an send a message to each neighbor j, the node i send its correspondent flow and estimate (f_{ji}, eij).

The next step, the state transition function, each node starts by updating the local flows and estimates with the corespondent received one from the corespondent neighbor. Thereafter, each computes a new prediction of the aggregation value e_i by averaging the received estimates and the one locally calculated by the equation bellow, that evaluates the overall estimate $AVERAGE$ of the network. It updates after its state accordingly: the new estimates equals to the one previous estimate calculated and the flow f_{ij} is added the difference between the new estimate e_i and the received estimate from$j$.

difference to calculate the SUM

$$a_i = v_i - \sum j \in D_i f_{ij}$$

4.5.3 PushPull

Algorithm 3 Push-Pull Active Thread

$q \leftarrow getneighbour()$
send s_p to q
$s_q \leftarrow receive(q)$
$s_p \leftarrow update(s_p, s_q)$

Algorithm 4 Push-Pull Passive Thread

$s_q \leftarrow receive(*)$
send s_p to $sender(s_q)$
$s_p \leftarrow update(s_p, s_q)$

In Push-Pull protocol [10], the nodes work with two *threads*. The active *thread* runs once at each round. Selects a neighbor q at random and send to it its value to aggregate s_p, afterwards, expects to receive the value s_q from the neighbor q. Update them the value s_p by averaging s_p and s_q.
Each node runs the passive *thread* in background, all the time. This background process basically sends the hold value s_p to every requested neighbor q. After sending it, the node updates it value s_p the same way as the active *thread*.
To calculate the SUM, instead of the $AVERAGE$, a pair is exchanged instead of only a value to average. One of the value is the value to average an the other value of the pair is the value to calculate the size. The overall SUM will evaluated by multiplying the size to the average. Note that to evaluate the size, every node should start the value to calculate the size as 0, except for one node that initializes with 1, the same principle as in Flow Updating and Push-Sum.

4.5.4 Extrema Propagation

Algorithm 5 Extrema Propagation

const K
var $n, x[1..K]$
Require: Init
$n \leftarrow neighbours(self)$
for all $l \in 1..K$ **do**
$x[i] \leftarrow rExp(1)$
end for
Send x to every p $p \in n$
Require: Receive $m_1..m_j$ from all $p \in n$
for all $l \in 1..j$ **do**
$x \leftarrow pointwisemin(x, m_l)$
end for
Send x to every $p \in n$
Require: Query
return $\hat{N}$

Extrema Propagation [37] is based on *sketchs*, each node holds and shares a *sketch*, an auxiliary data structure to calculate the desired aggregation function. Each node holds an array x with dimension K, and initializes every $x_i \in [1..K]$ equal to a random value calculated by the function $rExp(1)$, which returns a random number with and exponential distribution of rate parameter 1. n is initialized as the set of all neighbors to a node i, thereafter, the each node send the array x to every neighbor from n.

At each round, every single node from each message l from $m1..m_j$ updates the array x begin equal to the $pointwisemin(x, m_l)$. After updating x, send the updated array to each neighbor from n.
To calculate the estimate of the size of the network $\hat{N}$ in each round, each node computes the equation:

$$\hat{N} = \frac{K}{\sum i = 1^K x[i]}$$

To calculate the SUM instead of the size of the network, the $rExp()$ function takes as argument the value to aggregate instead of 1.

4.5.5 RIA LC/DC

RIA LC/DC [23] is also based on *sketchs*. Each node initially holds a *sketch*, an array of zeros of size m. Each node initializes the array by mapping, using a random function that gives the index of the array, the value to aggregate into the sketch, bit by bit. For example, if the value to aggregate is 3, the node should map the values 1, 2 and 3. After initialized the sketch, at each round the nodes share with their neighbor the sketch and merging the received ones with the sketch hold locally by using *XOR*. In order to calculate the estimate $\hat{n}$ of the SUM is calculated by

$$\hat{n} = -m * ln(V_n)$$

Where V_n is equal to the division of the number of zeros in the *sketch* and the size of m.

Chapter 5

Conclusion

Until now, much of the work done focused on the analysis of the state of the art in Smart Grids, AMR and in-network aggregation in both AMR and WSN and also the various existing algorithms.

After the state of the art complete, the next phase will be analyze in further detail the selected in-network aggregation algorithm. This analysis will require implementation in simple topologies in order to gain more insight about them. A performance analysis may take place in order to reduce the number of algorithms to be implemented in a SG or AMR topology. Afterwards a selection of a AMR topology will take place. Considering the existing pilot projects, the future topology may be as close to real implementations as possible. The next phase will be implement the algorithms and collect the results. Some improvements to the algorithms should occur so that a better performance may be achieved and aso the better results to the overall grid.

Appendix A

Aggregation Functions

A.0.6 Decomposable functions

For some aggregation function, one node may need to perform a single computation operation involving all the elements of the multiset, requiring more resources than the ideal ones. So, in order the distributed the effort to compute the multiset, there are some aggregation function that are decomposable. Meaning that, the effort could be done in a distributed way. A definition for decomposable function is also given in [33]:

Definition A.1. An aggregation function $f : \mathbb{N}^I \to O$ is said to be self decomposable if, for some (merge) operator $\diamond$ and all non empty multisets X and Y:

$$f(X \uplus Y) = f(X) \diamond f(Y)$$

The $\uplus$ denotates the standard multiset sum. The operator $\diamond$ is commutative and associative [33]. Some functions that are self-decomposable:

$$\begin{aligned} SUM(x) &= x, \\ SUM(X \uplus Y) &= SUM(X) + SUM(Y). \end{aligned}$$

$$\begin{aligned} COUNT(x) &= x, \\ COUNT(X \uplus Y) &= COUNT(X) + COUNT(Y). \end{aligned}$$

$$MIN(x) = x,$$
$$MIN(X \uplus Y) = MIN(X) \sqcap MIN(Y).$$

Definition A.2. An aggregation function $f : \mathbb{N}^I \rightarrow O$ is said said to be decomposable if for some function g and a self-decomposable aggregation function h, it can be expressed as:

$$f = g \circ h$$

As the definition above, stated in [33], self decomposable functions are a subset of the decomposable functions. One example of a decomposable functions $AVERAGE$:

$$AVERAGE(X) = g(h(X)),$$
$$h(x) = (x, 1),$$
$$h(X \uplus Y) = h(X) + h(Y),$$
$$g((s, c)) = s/c.$$

Another example is the $RANGE$ which gives the difference between the maximum and minimum value.

A.0.7 Duplicate sensitiveness and idempotence

For some functions, the presence of duplicate results does not affect the result. Examples of this aggregation functions are MAX and MIN, where" *the result on only depend on its support set(obtained by removing all duplicates)*"[33]. Others, like SUM or $COUNT$, the duplicate numbers are relevant. This propriety is called duplicate sensitiveness, it is relevante in distributed aggregation. Using an idempotent binary operator on the elements of the multiset helps obtaining fault tolerance [33].

Definition A.3. An aggregation function f is said to be duplicate insensitive if for all multiset M, $f(M) = f(S)$, where S is the support set of M.

A taxonomy table of aggregation is in [33] and it is presented below.

	Decomposable		Non-decomposable
	Self-Decomposable		
Duplicate insensitive	MIN, MAX	$RANGE$	$DISTINCT, COUNT$
Duplicate sensitive	$SUM, COUNT$	$AVERAGE$	$MEDIAN, MODE$

Bibliography

[1] Samuel Madden, Michael J Franklin, Joseph M Hellerstein, and Wei Hong. "TAG: A tiny aggregation service for ad-hoc sensor networks". In: *ACM SIGOPS Operating Systems Review* 36.SI (2002), pp. 131–146.

[2] Keren Horowitz and Dahlia Malkhi. "Estimating network size from local information". In: *Information Processing Letters* 88.5 (2003), pp. 237–243.

[3] David Kempe, Alin Dobra, and Johannes Gehrke. "Gossip-based computation of aggregate information". In: *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on.* IEEE. 2003, pp. 482–491.

[4] Jeffrey Considine, Feifei Li, George Kollios, and John Byers. "Approximate aggregation techniques for sensor databases". In: *Data Engineering, 2004. Proceedings. 20th International Conference on.* IEEE. 2004, pp. 449–460.

[5] Joao Girao, Markus Schneider, and Dirk Westhoff. "CDA: Concealed Data Aggregation in Wireless Sensor Networks". In: *ACM Workshop on Wireless Security.* Poster presentation. WiSe 2004. Philadelphia, USA, Oct. 2004. URL: `http://www.ece.cmu.edu/~adrian/wise2004/`.

[6] Márk Jelasity and Alberto Montresor. "Epidemic-style proactive aggregation in large overlay networks". In: *Distributed Computing Systems, 2004. Proceedings. 24th International Conference on.* IEEE. 2004, pp. 102–109.

[7] Nisheeth Shrivastava, Chiranjeeb Buragohain, Divyakant Agrawal, and Subhash Suri. "Medians and beyond: new aggregation techniques for sensor networks". In: *Proceedings of the 2nd international conference on Embedded networked sensor systems.* ACM. 2004, pp. 239–249.

[8] Claude Castelluccia. "Efficient aggregation of encrypted data in wireless sensor networks". In: *In MobiQuitous.* IEEE Computer Society, 2005, pp. 109–117.

[9] Claude Castelluccia, Einar Mykletun, and Gene Tsudik. "Efficient aggregation of encrypted data in wireless sensor networks". In: *Mobile and Ubiquitous Systems: Networking and Services, 2005. MobiQuitous 2005. The Second Annual International Conference on.* IEEE. 2005, pp. 109–117.

[10] Márk Jelasity, Alberto Montresor, and Ozalp Babaoglu. "Gossip-based aggregation in large dynamic networks". In: *ACM Transactions on Computer Systems (TOCS)* 23.3 (2005), pp. 219–252.

[11] Yitzhak Birk, Idit Keidar, Liran Liss, Assaf Schuster, and Ran Wolff. "Veracity radius: capturing the locality of distributed computations". In: *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*. ACM. 2006, pp. 102–111.

[12] Haowen Chan, Adrian Perrig, and Dawn Song. "Secure hierarchical in-network aggregation in sensor networks". In: *Proceedings of the 13th ACM conference on Computer and communications security*. ACM. 2006, pp. 278–287.

[13] J-Y Chen, Gopal Pandurangan, and Dongyan Xu. "Robust computation of aggregates in wireless sensor networks: distributed randomized algorithms and analysis". In: *Parallel and Distributed Systems, IEEE Transactions on* 17.9 (2006), pp. 987–1000.

[14] Shinji Motegi, Kiyohito Yoshihara, and Hiroki Horiuchi. "DAG based in-network aggregation for sensor network monitoring". In: *Applications and the Internet, 2006. SAINT 2006. International Symposium on*. IEEE. 2006, 8–pp.

[15] Laukik Chitnis, Alin Dobra, and Sanjay Ranka. "Aggregation methods for large-scale sensor networks". In: *ACM Transactions on Sensor Networks (TOSN)* 4.2 (2008), p. 9.

[16] Yao-Chung Fan and Arbee LP Chen. "Efficient and robust sensor data aggregation using linear counting sketches". In: *Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE International Symposium on*. IEEE. 2008, pp. 1–12.

[17] David G Hart. "Using AMI to realize the Smart Grid". In: *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*. IEEE. 2008, pp. 1–2.

[18] Vasconcelos and Jorge. *Survey of Regulatory and Technological Developments Concerning Smart Metering in the European Union Electricity Market*. EUI-RSCAS Working Papers 1. European University Institute (EUI), Robert Schuman Centre of Advanced Studies (RSCAS), Sept. 2008. URL: `http://ideas.repec.org/p/erp/euirsc/p0193.html`.

[19] Paulo Jesus, Carlos Baquero, and Paulo Sérgio Almeida. "Fault-tolerant aggregation by flow updating". In: *Distributed Applications and Interoperable Systems*. Springer. 2009, pp. 73–86.

[20] Bo Yu, Jianzhong Li, and Yingshu Li. "Distributed data aggregation scheduling in wireless sensor networks". In: *INFOCOM 2009, IEEE*. IEEE. 2009, pp. 2159–2167.

[21] BA Akyol, Harold Kirkham, S Clements, and M Hadley. "A survey of wireless communications for the electric power system". In: *Prepared for the US Department of Energy* (2010).

[22] Mauro Biagi and Lutz Lampe. "Location assisted routing techniques for power line communication in smart grids". In: *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE. 2010, pp. 274–278.

[23] Yao-Chung Fan and Arbee LP Chen. "Efficient and robust schemes for sensor data aggregation based on linear counting". In: *Parallel and Distributed Systems, IEEE Transactions on* 21.11 (2010), pp. 1675–1691.

[24] Hassan Farhangi. "The path of the smart grid". In: *Power and Energy Magazine, IEEE* 8.1 (2010), pp. 18–28.

[25] Palak P Parikh, Mitalkumar G Kanabar, and Tarlochan S Sidhu. "Opportunities and challenges of wireless communication technologies for smart grid applications". In: *Power and Energy Society General Meeting, 2010 IEEE*. IEEE. 2010, pp. 1–7.

[26] Jan Sacha, Jeff Napper, Corina Stratan, and Guillaume Pierre. "Adam2: Reliable distribution estimation in decentralised environments". In: *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*. IEEE. 2010, pp. 697–707.

[27] Pedram Samadi, A-H Mohsenian-Rad, Robert Schober, Vincent WS Wong, and Juri Jatskevich. "Optimal real-time pricing algorithm based on utility maximization for smart grid". In: *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE. 2010, pp. 415–420.

[28] Wasim M Taqqali and Nidhal Abdulaziz. "Smart grid and demand response technology". In: *Energy Conference and Exhibition (EnergyCon), 2010 IEEE International*. IEEE. 2010, pp. 710–715.

[29] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni. "Smart meters for power grid: Challenges, issues, advantages and status". In: *Renewable and Sustainable Energy Reviews* 15.6 (2011), pp. 2736–2742. URL: `http://EconPapers.repec.org/RePEc:eee:rensus:v:15:y:2011:i:6:p:2736-2742`.

[30] Stefano Galli, Anna Scaglione, and Zhifang Wang. "For the grid and through the grid: The role of power line communications in the smart grid". In: *Proceedings of the IEEE* 99.6 (2011), pp. 998–1027.

[31] U.S. Government. *Nist Framework and Roadmap for Smart Grid Interoperability Standards, Release 1 0*. General Books, 2011. ISBN: 9781234541682. URL: `http://books.google.pt/books?id=-C7UygAACAAJ`.

[32] Vehbi C Gungor, Dilan Sahin, Taskin Kocak, Salih Ergut, Concettina Buccella, Carlo Cecati, and Gerhard P Hancke. "Smart grid technologies: communication technologies and standards". In: *Industrial informatics, IEEE transactions on* 7.4 (2011), pp. 529–539.

[33] Paulo Jesus, Carlos Baquero, and Paulo Sérgio Almeida. "A Survey of Distributed Data Aggregation Algorithms". In: *CoRR* abs/1110.0725 (2011). URL: `http://dblp.uni-trier.de/db/journals/corr/corr1110.html#abs-1110-0725`.

[34] Tarek Khalifa, Kshirasagar Naik, and Amiya Nayak. "A survey of communication protocols for automatic meter reading applications". In: *Communications Surveys & Tutorials, IEEE* 13.2 (2011), pp. 168–182.

[35] Lutz Lampe and AJ Vinck. "On cooperative coding for narrow band PLC networks". In: *AEU-International Journal of Electronics and Communications* 65.8 (2011), pp. 681–687.

[36] Björn Lohrmann and Odej Kao. "Processing smart meter data streams in the cloud". In: *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on.* IEEE. 2011, pp. 1–8.

[37] Carlos Baquero, Paulo Sérgio Almeida, Raquel Menezes, and Paulo Jesus. "Extrema propagation: Fast distributed estimation of sums and network sizes". In: *Parallel and Distributed Systems, IEEE Transactions on* 23.4 (2012), pp. 668–675.

[38] Giacomo Verticale Cristina Rottondi and Christoph Krauß. "Implementation of a Protocol for Secure Distributed Aggregation of Smart Metering Data". In: *International Conference on Smart Grid Technology, Economics and Policies (SG-TEP 2012).* IEEE, Nov. 2012.

[39] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. "Smart Grid - The New and Improved Power Grid: A Survey." In: *IEEE Communications Surveys and Tutorials* 14.4 (2012), pp. 944–980. URL: `http://dblp.uni-trier.de/db/journals/comsur/comsur14.html#FangMXY12`.

[40] Yide Liu. "Wireless Sensor Network Applications in Smart Grid: Recent Trends and Challenges." In: *IJDSN* 2012 (2012). URL: `http://dblp.uni-trier.de/db/journals/ijdsn/ijdsn2012.html#Liu12`.

[41] Alberto Sendin, Inigo Berganza, Aitor Arzuaga, Anssi Pulkkinen, and Il Han Kim. "Performance results from 100,000+ PRIME smart meters deployment in Spain". In: *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on.* IEEE. 2012, pp. 145–150.

[42] Omar Asad, Melike Erol-Kantarci, and Hussein T Mouftah. "A Survey of Sensor Web Services for the Smart Grid". In: *Journal of Sensor and Actuator Networks* 2.1 (2013), pp. 98–108.

[43] Zekeriya Erkin, Juan Ramón Troncoso-Pastoriza, Reginald L. Lagendijk, and Fernando Pérez-González. "Privacy-Preserving Data Aggregation in Smart Metering Systems: An Overview." In: *IEEE Signal Process. Mag.* 30.2 (2013), pp. 75–86. URL: `http://dblp.uni-trier.de/db/journals/spm/spm30.html#ErkinTLP13`.

[44] S. Ghosh, Manisa Pipattanasomporn, and Saifur Rahman. "Technology deployment status of U.S. smart Grid projects - Electric distribution systems." In: *ISGT*. IEEE, 2013, pp. 1–8. ISBN: 978-1-4673-4894-2. URL: `http://dblp.uni-trier.de/db/conf/isgt/isgt2013.html#GhoshPR13`.

[45] Dejan Ilić, Stamatis Karnouskos, and Martin Wilhelm. "A Comparative Analysis of Smart Metering Data Aggregation Performance". In: *IEEE 11th International Conference on Industrial Informatics (INDIN), Bochum, Germany.* July 2013. URL: `http://diktio.dyndns.org/files/2013_INDIN_aggregationPerformance.pdf`.

[46] Pedro Godinho Matos, Antonio Aires Messias, Pedro Ricardo Daniel, Manuel Sao Miguel Oliveira, Ana Margarida Veiga, and Paulo Libano Monteiro. "Inovgrid, a smart vision for a next generation distribution system". In: *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on.* IET. 2013, pp. 1–4.

[47] Giuliano Andrea Pagani and Marco Aiello. "From the Grid to the Smart Grid, Topologically". In: *arXiv preprint arXiv:1305.0458* (2013).

[48] Keita Suzuki, Chuzo Ninagawa, Hiroki Yoshida, Seiji Kondo, Junji Morikawa, Taiga Kanbe, and Takao Aoki. "Smart grid ADR aggregation delay model on large-scale distributed building HVAC facilities". In: *ISGT Europe.* 2013, pp. 1–5.

[49] Leandro Villas, Azzedine Boukerche, H Ramos Filho, H Oliveira, Regina Araujo, and A Loureiro. "DRINA: a lightweight and reliable routing approach for in-network aggregation in wireless sensor networks". In: (2013).

[50] Xue Lin, Yanzhi Wang, and Massoud Pedram. "Designing the Optimal Pricing Policy for Aggregators in the Smart Grid". In: *Green Technologies Conference (GreenTech), 2014 Sixth Annual IEEE.* IEEE. 2014, pp. 75–80.

[51] *Hoogkrek, PowerMatching City.* URL: `http://www.powermatchingcity.nl/`.

[52] IBM. *Informix TimeSeries.* URL: `http://www-01.ibm.com/software/data/informix/timeseries/`.

[53] Pecan Street Research Institute. *Pecan Street Project.* URL: `http://www.pecanstreet.org/`.

[54] OPower. *OPower 4*. URL: `http://opower.com/company/news-press/press_releases/67`.

[55] *The DEHEMS Project*. URL: `http://www.dehems.eu/`.

[56] G. N. Ericsson. "Cyber Security and Power System Communication—Essential Parts of a Smart Grid Infrastructure". In: *IEEE Transactions on Power Delivery* 25.3 (July 2010), pp. 1501–1507. ISSN: 0885-8977. DOI: `10.1109/MSP.2010.49`. URL: `http://dx.doi.org/10.1109/MSP.2010.49`.