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, February 2014

Abstract

Abstract

The modern and globalized world become more and more dependent on energy sources such as electricity or oil. The power grids all over the planet become increasingly bigger leading to problems of energy waste and sustainability. The smart grids appear to address this issues and also to interconnect the various renewable energy sources like solar or wind since they have different requirements than the traditional ones. This new concept of grid introduce ICT and computation and that's why they are called 'smart'. One of the biggest advantages of smart grids is the ability to remotely read fine-granular measurements from each smart meters, which enables the grid operators to balance load efficiently and offer adapted time dependent tariffs. With data aggregation from the smart meters it's possible to reduce the energy consumption and change it to be more efficient. 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 observe how the aggregation algorithms work in the smart grid, collecting the results and evaluate which algorithm suites the best.

Resumo

O mundo moderno e globalizado tornou-se cada vez mais dependente em energias como a electrecidade ou o petroleo. As redes de energia por todo o planeta tornaram-se cada vez maiores levando a problemas de perda de energia e de sustentabilidade. As *Smart Grids* apareceram para fazer face a esses problemas assim como interligar as varias fontes de energia renovaveis como a solar ou a eolica pois elas tem diferentes requisitos das tradicionais. Este novo conceito de rede introduz computacao e TIC e e por isso que elas sao denominadas de *smart*. Uma das grande vantagens das *Smart Grids* e a habilidade para ler dados de fina granularidade de maneira remota dos leitores inteligentes, permitindo aos operadores de rede balancear de forma eficiente a energia e oferecer tarifas adaptaveis. Com a agregacao dos dados dos leitores inteligentes possivel reduzir o consumo de energia e muda-lo para ser mais eficiente.

Neste trabalho sera tratado o problema da agregacao de dados de forma distribuida. E proposta uma abordagem distribuida de agregacao, onde todos os leitores inteligentes leem o consumo de energia e alguns funcionam como agregadores.

Contents

List of Figures

Todo list

Chapter 1

Introduction

The power electrical 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. As important as it is and thanks to its large size, the power grid consumes a enormous amount of natural resources, make it unsustainable in long term. The dawn of new renewable energy sources also increase the need to modernize the grid since it's mandatory to interconnect them to the traditional Grid. The introduction of ICT and computation in the grid is trying to change it to became more sophisticated, eco sustainable and integrate all the energy sources to enable efficient electrical power distribution. This new concept of grid is called Smart Grid.

Smart Grid is a modern power grid that uses computation, information and communication. In an automatic way, Smart Grid improves the energy efficiency, sustainability both in power distribution and in electricity production. It enables the grid to become more sustainable as it makes a more efficient management of natural resources. The Smart Grid is composed by 'Islands of Automation' interconnected with a communication infrastructure [33].

Smart meters 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 [29].

In the next years, the amount of user data collected by the Smart Grid 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[31].

In this work, we look at the information collected within the Smart Grid. 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

Are the Objectives clear enough?

There are two types of Smart Meters network architectures[29] namely *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 too. They also perform a partial data aggregation themselves, It's called in-network aggregation[29].

In this work, we will focus on the second type of architecture which provides more interesting challenges.The purpose 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 goal, 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 Smart Grid works, how all components work together in a integrated way and the status of deployed models . The chose of suitable topology for our work, with several meters collecting information about the consumers consumption and aggregate that data in a distributed way, is an important part of this work as the study of the existed algorithm to perfume distributed aggregation.

The study of chosen 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 in a Smart Grid topology. We are interested in knowing which algorithm provides the best results in time, exchange messages, scalability, resilience, fault tolerance and, most important, which one provides the most relevant information.

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

Maybe some more motivation

As stated before, Smart Grid is a new and important concept of grid that is of main importance towards the world energy sustainability. Also, 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 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

review the chapters ref

In this document, a state of the art regarding the overall work thematic is presented. In chapter **??** 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 os *Wireless Sensor Networks*(WSN), an importante concept because the AMI it is considered in this work as a specific case of the implementation of this kind of the networks, with the same limitation, problems and challenges such as the low capability devices or *in-network* aggregation, which is the main point to consider this network in the scope of this thesis. The usage of this sensors and the bridge between the two concepts, smart grid and WSN, is also presented.

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

Chapter 2

Smart Grid

The *Smart Grid* (SG) is a new concept of grid which introduces new technologies into the common power system. They enable power grids to become more efficient, integrate other sources of energy than traditional ones such as renewable energies and increase the overall management performance By using modern information technologies, the SG is capable of delivering power in more efficient way and responding to wide ranging condition an events [26]. There are several definitions for the SG among the literature. For example, Xi Fan *et al*[26] states that SG can be regarded as an electric system that uses information, two-way, cyber-secure communication technologies and computational intelligence in a integrated fashion to achieve a clean, safe, secure, reliable, resilient, efficient and sustainable system. S. Gosh *et al* [30] 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. Correct definition?

SG creates an environment that introduces a converge 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 is it 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 is variety of problems and challenges, most of the proposed solution and studies regarding the SG focus in some specific aspects.

2.1 Smart Grid Model

The components in a traditional grid go one way, contrary to what happens in SG where all the model of all grids?

flows of electricity goes two-way. So the role of each component are different, for example a consumer can both consume energy from the grid and provide it too considering that he has a device that produces renewable power. The NIST report [22] proposes a conceptual model providing the main actors towards the SG. Costumers, the end users of electricity, Markets, Services Providers, Electricity Com-

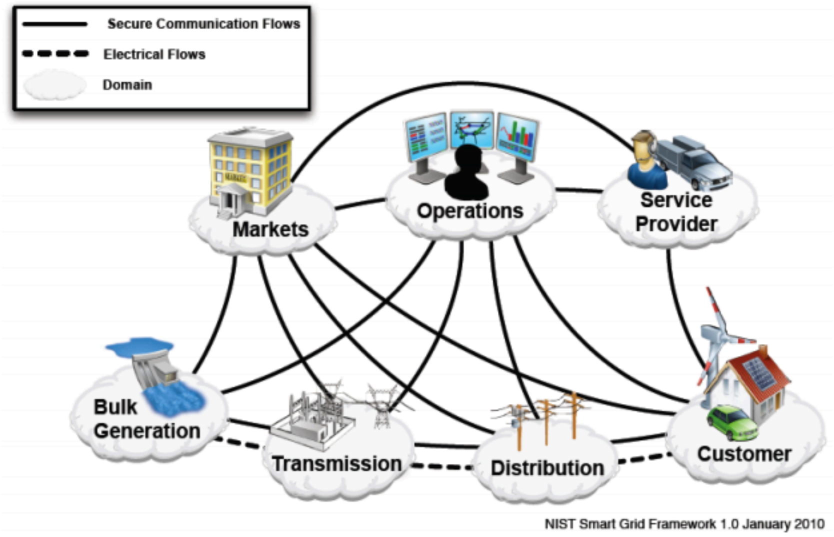

FIGURE 2.1: NIST Conceptual Model for SG

panies, Operations, managers of the movement of electricity, Bulk Generation, Generation Centers, Transmission and Distribution. In [26] provides a more technical approach where the SG is separated into three major subsystems:

- *Smart infrastructure system* embraces the energy, information and communication infrastructure. 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.

- *Smart management system* Provides advanced management and control services and functionalities, [26] considers the key reason why SG can revolutionize the grid Most of the new grid goals ate related to energy efficiency improvement, supply and demand balance, emission control etc. and thats the scope of problems the management systems tries to achieve.

- *Smart protection system* Provides advanced grid reliability analysis, failure protection, security and privacy protection services.

The focus of this work is in the smart information subsystem. We are interested in the information that comes from the smart meters. In the next section is detailed the SG information subsystem.

2.2 Smart Information SubSystem

This part of the SG refers to all the information that is collected by sensing the consumer consumption and its management . The data collected is often used for billing, grid status monitoring and user appliance control [26]. Then, it is aggregated and *smart management* is ideally performed on it.

Enough information regarding the information susbsystem?

An important concept in the information subsystem sensing is *Smart Metering*. It is basically a mechanism that sense and monitor consumption data from end users. In the next chapter we take a look at the main device called the smart meter.

Other aspect is the *Smart Monitoring and Measurement* which can be approached by either *sensor* 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 measures the electrical waves on a electrical grid to determinate the health of the system. The management refers to all the information analysis and modeling, integration and optimization.

In this specific part of SG there are several areas of research that represent a all new set of opportunities.

2.3 Smart Grid Communication

The most important question in the communication is what network and communication should be used[26]. Since there is no standard communication system in SG, several solutions were proposued.

There are two types of communication, wired and wireless communication. Wired communication are harder to implement than wireless communication, and for that reason, wired are more expensive to deploy. Wireless communication are also more suitable for remote end applications [19].

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[26]. Increases reliability and automatic network connectivity and 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 it quick and low-cost to obtain data communications coverage over a large geographic area [17]. There several solutions that uses 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[19]. 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[18] because it provides a standardized platform for exchanging data between smart metering devices and appliances located on costumer premises[26]. WirelessHART and ISA100.11a are other examples of wireless communications based on the IEEE 802.15.4 protocol.

Other examples of wireless communication are stallite communication, 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 been already installed, so the cost of deployment is way less than other wired solution, but has also big security disadvantages. Fiber-Optic has the advantage of being fast and more secure but is very expensive to deploy.

2.4 Smart Meters

Smart meters are devices that sense the energy consumption. They are installed in the costumer side , in households ou industrial facilities, depending on the type of costumer. Playing a major role in the information subsystem, smart meters present a number of challenges in sensing, analyzing, and communication[29]. Smart meters, more specifically, the Smart Mettering System has also the denomination of AMR(Automatic Metering Reder). In [24] the AMR is refered as the technology whose goal is to help collect the meter measurement automatically and possibly send commands to the meters.
As referred, the main function of a smart meter, and all meters, is sense the consumption in the costumer side. Plus that, this smart devices also have communication capabilities. So, every pre-defined period of time, they communicate the sensed consumption to a central device that aggregates the data. This feature, 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[33]. Jorge Vasconcelos [**e**]nlightens 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. [24] states some benefits os the smart mettering system: Real time pricing, power quality measurement, automated Billing, Load management,, Remote Connect/Disconnect, Outage notification and Bundlig with water and gas.
Privacy and security are important concerns when dealing with the sensed information. Several privacy issues appeared considering that external parties access the consumer energy consumption. Some are authorized parties, but if unauthorized entities access this data, some security and privacy dangers

could appear. For example, by analyzing the data, could determinate which devices are plugged in some specific time, giving for example information about if there is people in home or note. Many works 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.

The smart meters, as any desirable component of the SG, enable two way communication.

Smart Grid topology?

Chapter 3

Wireless Sensor Networks

Wireless Sensor Networks(WSN) are *ad-hoc* networks composed of tiny devices with limited computation and energy capacities. These tiny devices, sensors, are so 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[28] 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, also enabling control of sensor activity. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance.

Today such networks are used in many industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring and so on. Some of WSNs requirements are a large number os 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 which can be considered as an extension of the future internet towards smart devices, Internet of Things(IoT)[28].

3.1 WSN and The Smart Grid

Review this intro

Is the concept of a AMI being a specific case of WSN correct?

Considering the overall appliances, WSNs have several applications regarding the SG. Furthermore, the AMI(Automated Metering Infrastructure)could be considered as a specific appliance of WSN, where could be implement the proposed solutions for data aggregation.

Recently, WSN has been widely recognized as a vital component of the electric power system[27]. 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[21]. The collaborative and context-awareness nature of WSN brings several advantages over traditional sensing include great fault tolerance, improved accuracy, larger coverage area and extraction of localized features [27]. 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, smart sensors for transformer monitoring, high voltage line temperature and weather condition sensors, distributed generation, smart gird storage and, referenced before and more importantly for this work, WSN for AMI. A specific example is in [27] where a WSN could apply perfectly to a household or House Area Network(HAN) . In section **??** it is mentioned ZigBee as communication technology in Smart Grids. Due to its reliable wide area coverage and predictable latencies, ZigBee is a suitable choice for a Local Area Network such as a household or a neighborhood. As a example in [27], a WAMR(Wireless Automatic Meter Reading) can determinate real-time energy consumption of the customers by sensing each device that have a wireless sensor on it. The mart 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 founded in [27]. 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 [24]. For example, individual measurments must preserve its informations. 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 indetifier that links the smar meter to a household/costumer/producer. Other important difference is the fact that AMR must support bi-direccional communications, most of WSN only have one way communicaton. 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 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.

dont understand this assertion, in WSN is possible for a sensor to change the position?

Chapter 4

Distributed Data Aggregation

4.1 Definition

Data aggregation is a technique that consists, on its basis, in reduce the amount of data collected, reducing the resources needed to process it. According to [23], 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 4.1. 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. As 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 means that it is a task which is distributed among several nodes in the network. In contrary of a *centralized* arquitecture, 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.

4.1.1 Decomposable functions

Correct definition? Need to specify more why decomposable functions could be computed in a parallel or distributed way

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 [23]:

Definition 4.2. An aggregation function $f : \mathbb{N}^I \rightarrow O$ 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 [23]. Some functions that are self-decomposable:

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

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

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

Definition 4.3. An aggregation function $f : \mathbb{N}^I \rightarrow O$ is said said to be decomposable if for some function g and a sel-decomposable aggregation function h, it can be expressed as:

$$f = g \circ h$$

As the definition above, stated in [23], self decomposable functions are a subset of the decomposable functions. One example of a decomposable functions $AVERAGE$:

$$\begin{aligned} AVERAGE(X) &= g(h(X)), \\ h(x) &= (x, 1), \\ h(X \uplus Y) &= h(X) + h(Y), \\ g((s, c)) &= s/c. \end{aligned}$$

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

4.1.2 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)[23]. 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 [23].

Definition 4.4. 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 [23] and it is presented below.

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

4.2 Distributed Data Aggregation Algorithms

In [23] is also presented a simple taxonomy of the existing algorithms that performe distributed data aggregation. First it is analyzed the algorithms from the communication prespective, i. e., the routing protocols and the intrisic topologies, afterwards, it is analyzed the computation issues, how the aggregation functions are computed by the algortihms.

Review DDA, Write the algorithms or just the type?

4.2.1 Communication

4.2.1.1 Hierarchy-based approaches

Traditionally, existing aggregation algorithms operate on a hierarchy-based communication scheme. This is *structed* 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 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 AGregation algorithm that suits for ad-hoc networks described in [1]. This algrotihm requires the previous creation of a tree-based routing topology, and the continous maintenance of such routing structure in order to operate over mobile networks. TAG provides a SQL-like declarative language to the users. The algortihms 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.

DAG An aggregation scheme fro WSN is proposed in [13] that aims to reduce the number of message losses.

resume this shitty algorithm

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

I-LEAG Cluster-based aggregation approach designated as I-LEAG is in [10]. The routing structure of this algortihm is composed by a hierarchy of clusters or partitions. A single pivot is desginated 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 tradional 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 tradiontal tree schemes and delta in regions with higher rate of message losses(mostly regions near the sink with the aggregate of several nodes).

resume other approachs

Ring Based Approachs, Flooding and Randomized necessary or relevant?, IDEA: state the hierarchy and gossip approach and then the hybrid

4.2.1.2 Gossip-based approaches

This type of approach is reffered as an *unstructed* approach, contrary of a aforementioned *structed* approachs. 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 accross the network without follwing any specifc topology, the informartion it is passed node to node, or nodes, like a infectious disease or a gossip,i.e., a ”infected” node sends a message to a random subset of nodes. This type scheme tend to allow a robust(fault tolerant) and scalable information dissemenation over all the network[23]

Push-Sum Protocol Push-sum protocol[3] is a gossip-based protocol to compute aggregation functions. 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 choosen uniformly at random and half of the actual values are sent to the target node and the other hald to the node itself. Upon reveived, the local values are updated, adding each value from a received pair to its local component[23].

review this definition, too much alike

4.2.1.3 Hybrid approaches

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

Chitnis et al, 2008 Chitnis et al.[14] 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 performe a hierachic communication with other leaders nodes regarding the aggregation results from the gossip group.

4.2.2 Computation

4.2.2.1 Hierarchical

The input is separted into groups so it can be computed in a distributed hirarchical way. 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 work as a normal *forward* node. This class of algorithms allow any decomposable function with high accuracy without the presence of faults. Algortithms of this class were aforementioned.

4.2.2.2 Averaging

This class of computation scheme is based on a 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. One example of algorithms of this class, are the ones with gossip base communication scheme, since the results of the aggregates could be share randomnly with the neighbor nodes. Due to its nature, Averaging algorithms tend to by highly robust, i.e., tolerant to faults on opposite of the structed algrotihms. Decomposable and duplciate sensitece functions can be be computed in this class.

specify the mass conservation concept?

Push-Pull Gossiping Similar to the aforementioned *push-sum protocol*, the push-pull gossiping[6] performes an averaging process. This algortihm executes an epidemic protocol to perform a pari-wise exchange of aggregated values between neighbor nodes[23]. 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 perfomed 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 [12] randomly creates groups across the network in which aggregates are sucessfully computed. There are three modes a node can performe:*leader,idle* and *member* which corresponds to three phases. First every node is in *idle* mode, them 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, respondes to the first one received with a JACK(Joining Acknowledgment) tagged with their aggregated value becoming a member of the group. Finally, the *leader* gather all the aggregated values, computing the aggregation function($AVERAGE$) and broadcasts a Group Assignment Message with the final result. Every group member waits unit it receives the result from the leader to update its local value and them returns to *idle* mode.

Flow Updating

finish this shitty algorithm

4.2.2.3 Sketches

Algrorithms 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 agregation could be done using multiple paths. This type of algorithms enable operations of order and duplicate insentitive. 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 algorirthms 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 [15], a multi-path routing aggregation approach. The algrotihm consists of two phases. First a aggregation request is sent by the sink throughour the whole network, creating a multipath routinh hierarchy. Second, starting in the lower leves, each node generates a *sketch* correspondent to its current state and sent 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 send it to the upper node until the top is reached where the sink computes the aggregation estimate.

Extrema propagation This approach redices the computation of an aggregation function[23]. 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 os the computation of the pointwise minimum between all exchagend vectors. At each node, the obtaind vector is used as a sample to produce an aproximation of the aggregation result. This algorithm is focused ob obtaining a fast estimate, tather than an accurate one.

4.2.2.4 Digests

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

Q-Digest This aggregation scheme allows the approximation of complex aggregation function in WSN is proposed in [7]. Uses an hierarchical routing topology to build and disseminate quantile digests. Each node maintains a quantile digest of the data avilabe, which are built in a botton-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 acording to a specific criteria.

Equi-Depth Gossip-based approach described in [2]. The scheme executes a gossip protocol and merge 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 aproximation of the network distribution of values. There are four merging techniques *swap, concise, equi-with histograms* and *equi-depth histograms* that are detailed in [23]. ***Adam2*** Adam2 is a gossip based algorithm to estimate the statistical distribution of values across a decentralized system[20]. Each node can decide to start an instance of Adam2 where each instance is uniquely identified by its starting node. The statrting 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 seting 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 store a set of interpolaton points for each running algorithm instance. A new node that learning about the new instance performes a initializtaion ant then starts participating in the protocol. The sets are exchanged like push-pull, the sets are merged by averaging the fraction at each interpolaiton poin. After a predefined number round the CDF is approximated by interpolating the point of the resulting set.

4.3 WSN Data Aggregation

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 ate

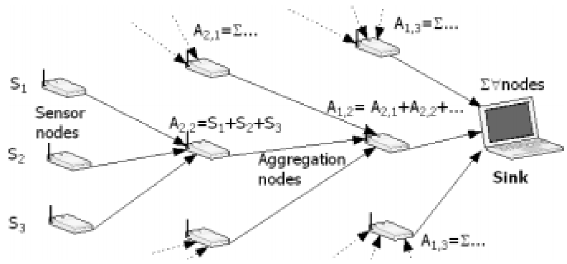

FIGURE 4.1: Principle of in-network aggregation

potential aggregators and that data gets aggregated as they propagate towards the sink. The aggregation is set as must being simple not involve 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 [11] 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 [16]. 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.

4.4 Smart Metering Aggregation Model

There are two main architectures for smart metering considering data aggregation are *centralized* and *distributed* or *decentralized*[29]. 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 [29]. As introduced before, the architecture chosen for this work is *de-centralized* due to the nature of the aggregation algorithms.

In the literature, it is possible to find some particular studies. Keita Suzuki *et al* [32] 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*[25]. 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. The overall scheme is presented in 4.2.

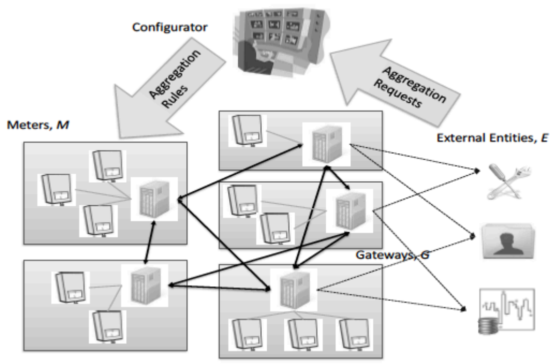

FIGURE 4.2: The functional nodes of the architecture

Chapter 5

Conclusion

What to write? Status of the work?

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] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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).

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

[19] 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.

[20] 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.

[21] 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`.

[22] 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`.

[23] 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`.

[24] 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.

[25] 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.

[26] 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`.

[27] 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`.

[28] 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.

[29] 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`.

[30] 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`.

[31] 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`.

[32] 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.

[33] 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`.