AALBORG UNIVERSITY

Title:

A researching step on the WSN routing protocols' scope.

ProMultihoping proposal and his behavior

Project period:

S10, summer semester 2011

Participants:

Jordi Ferrer Martí

Supervisors:

Rasmus Hjorth Nielsen
Neeli R. Prasad

Number of copies: 3

Number of pages: 60

Attachments: ANNEX A and B

Submit date: 24/May/2011

Department of Electronic Systems
Electronics & IT
Fredrik BajersVej 7 B
9220 Aalborg Ost
Phone: 9940 8600
http://es.aau.dk

Abstract

This document speaks about the research made by the author over the different behavior of two main topologies of WSN routing protocols. These topologies are clustering and flat. There is a lot of literature of energy-efficient clustering routing protocols, also of energy-aware flat routing protocols. After reading are proved both performances by simulations.

As in Wireless Sensor Network the most important break point is energy management, knowing behavior of routing protocol it is possible to find the best energy management in each moment.

After that, a new energy-aware routing protocol are proposed and compared.

Also an aggregation mode is proposed and checked it by simulating.

The results open new lines for future works.

Content

1. Introduction to Wireless Sensor Network

1.1. Internet of Things (IoT)

The Internet of Things refers to uniquely identifiable objects (Things) and their virtual representations in an Internet-like structure. The term Internet of Things has first been used by Kevin Ashton in 1999. The next generation of Internet applications using Internet Protocol Version 6 (IPv6) would be able to communicate with devices attached to virtually all human-made objects because of the extremely large address space of IPv6. This system would therefore be able to identify any kind of object.
The Internet of objects would encode 50 to 100 trillion objects, and be able to follow the movement of those objects. Every human being is surrounded by 1000 to 5000 objects.
In this Internet of Things, made of billions of parallel and simultaneous events, time will no more be used as a common and linear dimension but will depend on each entity (object, process, information system, etc.). This Internet of Things will be accordingly based on massive parallel IT systems.[REF. 1]

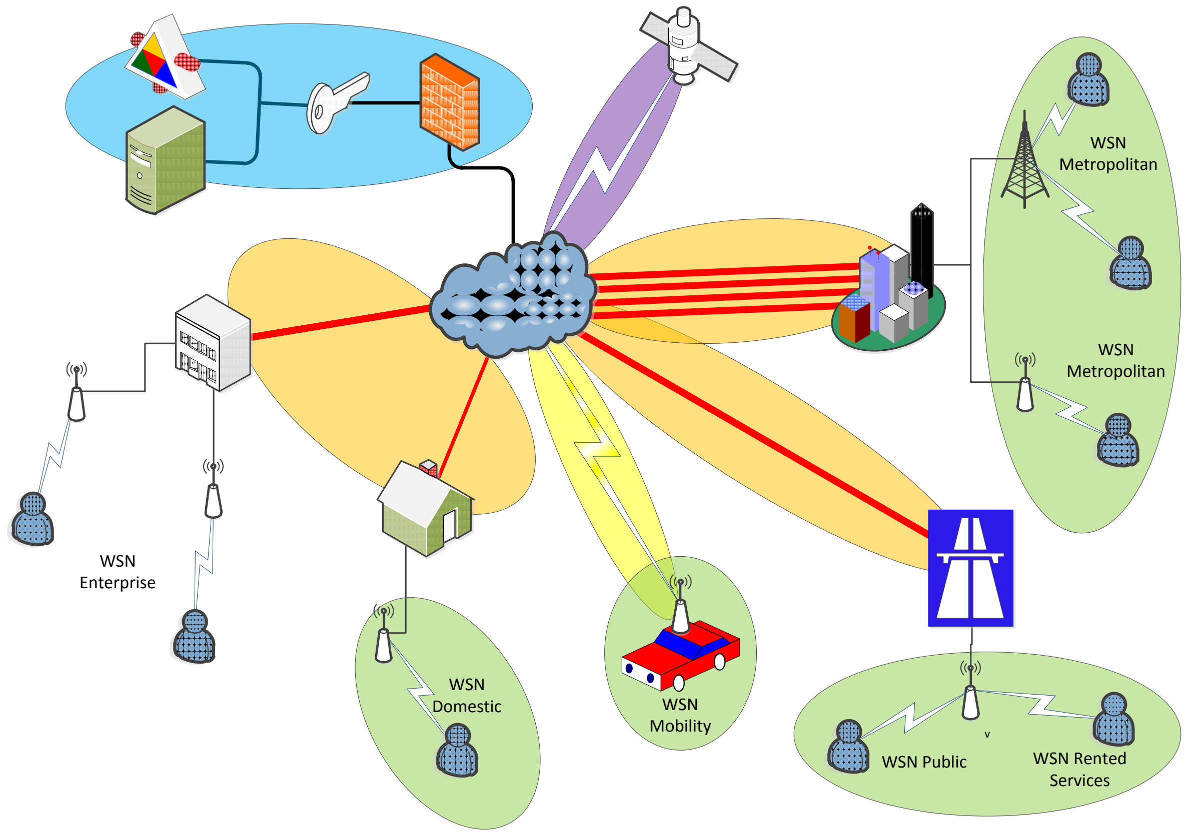
Alcatel-Lucent touchatag service and Violet's Mirror gadget provide a pragmatic consumer oriented approach to the Internet of Things by which a developer can link real world items to the online world using RFID tags and QR Codes. Arrayent, Inc. is a B2B company, internet-connecting consumer products (i.e. thermostats, security systems, toys, healthcare products) to smartphones, tablets, and web browsers. Pachube, built by Connected Environments Ltd, provides data management infrastructure for sensors, devices and environments, and describes itself as *"a convenient, secure & scalable platform that helps you connect to & build the 'internet of things"*. Nimbits, is an open source data historian server built on cloud computing architecture that provides connectivity between devices using data points.

This communication revolution is now extending to objects as well as people. Imagine if all the objects in the world had all the information that they needed to function optimally. Buildings would adjust themselves according to the temperature. Ovens would cook things for exactly the right time. The handles of umbrellas would glow when it was about to rain. We long ago inserted "intelligence" into objects in the form of thermostats and the like; the internet of everything will extend this principle exponentially, giving us unprecedented control over the objects that surround us.

The Internet of Things will help solve two of the biggest problems facing the world: energy and health care. Buildings currently waste more energy than they use effectively. We will be able to cut this waste down to almost nothing. Health care is currently delivered in lumps: we visit the doctor a couple of times a year at most, and get our blood pressure checked every now and again. The internet of everything will allow us to monitor our bodily functioning all the time. A few sensors discreetly attached to the body will keep you constantly informed about how your vital functions are doing. It will also help us to keep ourselves healthy. [REF. 2]

The Internet of Things, which will include connections over the public Internet as well as an unlimited number of private networks, comes with a host of challenges that don't exist with today's networks – those designed with the assumption of unlimited power for driving ever bigger and faster digital machines for massive, fiber optic-based broadband links.

There is a current of layering point of view, based on *"Internet of Things. A critique of ambient technology and the all-seeing network of RFID"* by Rob Van Kranenburg.



Picture 1. Layer view of WSN. From a human user to the Cloud.

Basic layer of Internet of Things is WSN. Watch at Picture 1 every chain ending is a wireless network of sensors.

Applications for WSN:

- Disabled people: shoes for blind people.
- Medicine: ICU control devices, drugs delivery devices.

- Road traffic safety.
- Naval & air control.
- Buildings, Office, Self-desk.

1.2. Beginning with WSN

As we could see in other networks, referred to OSI model, WSN could also have several layers with a specific job each of them.

Layer	Job	Design issues
Routing	Manage nodes and data end to end.	Save energy, scalability, node drops, mobility…
MAC	Manage the physic layer. Awake/sleep mode. Syncro TDMA, FDMA/CDMA assign parameters.	Few components. Low requirements.
Physic	Modulation of signal and launch to environment.	Strong. Multipath. Fading. Noise. Efficient BW.

Table 1 WSN stack protocols

Before going to the main topic of this paper, routing, I would like to write some words about if it is necessary a MAC layer in this kind of networks.
All of us may have OSI model very present. Well, in wired networks, we started with bus networks where it was needed a MAC protocol to avoid collision. After MAC layer has continued inside the evolution of networking to allow the integration of old network infrastructures in new ones. Always many authors refer to OSI model to design any communication protocol, but just now, a new infrastructure is built: nodes to Sink/BS or Gateways.
There are many modulations which avoid collisions, for example TDMA, FDMA, CDMA..., then it is possible to build, from the beginning, a new protocol with a new concept of MAC layer. This layer only manages the physical layer, as shown in Table 1, allowing save energy and computational cost. Then layer of OSI stack could be pretty reduced.

1.3. Energy problem

The main problem of WSN is the power consumption. Those kind of devices have a deep constrain of power supply capacity. Due to his size, weight, usually the power supply has a small capacity, moreover due to localization could be impossible to recharge. This is the main topic of research. Also there are several restrictions more, here below a summary.

Main node restrictions	Protocol research targets
Processing capability Battery power Bandwidth	Minimize power consumption. Scalable for any size. Minimize computational.

Storage Space	Simple and independent from hardware Limit number of transmissions Mobile nodes Data heterogeneity QoS

Table 2. WSN summarized requirements

2. Comparison of different solutions for WSN routing protocols

2.1. Introduction to WSN routing protocols

Reading many papers related with WSN in the IEEE web-side, I found a lot of authors who describes their own protocols which outperform older ones. In a first lecture I filtered some of them, but others give interesting ideas and point of view of how to solve the restrictions of WSN. Below there is a table describing the most representative protocols and, in blue color, evolutions of the main ones and the interesting changes they release.

NAME	Topology	Packet headers	Packet control	Energy Aware	Path setup	Schedule	Signaling
Flooding	Direct	No	None	No	None	None	All Broadcast
Gossiping	Direct	No	None	No	None	None	All Broadcast
SPIN	Direct	No	Negotiation	No	None	None	All Broadcast
SEER	Flat	73bits	Proactive	Yes	Flooding	None	Unicast
HEER	Flat. Schedule TDMA. Path Setup Broadcast.						
LEACH	Hierarachical Singlehop	ID CH	Proactive	No	Broadcast	TDMA CH / Event nodes	Unicast
M-LEACH	Hierarachical Multihop	ID CH+ ID level+ ID next	Proactive	Yes	Flooding	TDMA CH / Event nodes	Unicast
PEGASIS	Pseudo Hierarachical	No?	Proactive	No	Global knowledge	Random CH	Unicast
HEED	Hierarchical	ID CH	Proactive	Yes	Flooding	Random CH residual E	Unicast
CDC	GPS location.						
AROS	LEACH allowing Multihop						
EEMR	Ennergy Efficient CH closer BS smaller cluster size. Multihop CH Routing:more residual ECH						
EAAC	GPS location. Next CH round -> Energy Aware. Each cluster own TDMA Schedule.						
BCDCP	Balanced clusters, Enery Aware. Splitting clusters. Lowest Energy Multihop routing. TDMA non-CH						
TEEN	Tree	Not defined	Reactive	No	Not defined	TDMA?	Not defined
APTEEN	Tree	No	Hybrid	No	Broadcast	TDMA	Unicast
MSTBBN	Search Tree definition.						
EECLSTP	Energy Efficient Clustering based on Least Spanning Tree						
EAP	Flat	Yes	Proactive	Yes	Broadcast/ Flooding	Event	Unicast
E-WME	Flat	Yes	Proactive	Yes	Broadcast/ Flodding	Event	Unicast
HESS	Flat	Yes	Reactive	Yes	Broadcast next	Event	Unicast

Table 3. Shown the most relevant routing protocols followed in this work.

There are a lot of authors who beat for a clustering topology also other ones who prefers flat topology[1] let see some differences:

Hierarchical routing	Flat routing
Collision avoided	Collision overhead
Reduced duty cycle	Variable duty cycle
Data aggregation clustering	Multihop aggregates data from neighbors
Simple not optima routing	Routing accurate but add complexity
Routing Energy consumption made in setup.	Routing Energy consumption a long data routing.
Syncro	Asyncro
Low latency by multihop	Latency grows with network size
Energy dissipation not controlled	Energy dissipation adapts routing decisions
Hotspots	Hotspots avoided by routing complexity

Table 4. Comparasion among the most relevant features of Hierarchical and Flat routing protocol.

Neither of those topologies are the total solution to all questions launched for WSN. According their authors, for some applications clustering is able to offer the best solution, but many times not.
This compilation brings me to think in a mixed topology, also introduced for many authors I read.

Looking up the behavior of clustering and flat protocols it is possible to summarize that always the data goes from All to One, means from each sensor node to the sink[2]. That behavior always form a tree, doesn't mind its topology.
But the difference between both, now, it is significant over issues like management: scalability, failure detection, setup process, nodes memory resources, nodes computational resources and so on.
On the other hand, clustering topology expends a lot of energy setting up the network otherwise flat topology doesn't use much energy at the beginning but needs to use it later due to signaling required.

As a conclusion can be done an improved by features classification:

Hierarchical		Flat	
Advantages	Disadvantages	Advantages	Disadvantages
Data aggregation	Medium Hotspots	Scalability	Low Hotspot
Low Local power consumtion	Higher Hardware (clusters)	Simplicity	Nodes Unreachable due to table update.

[1] Some cases flat routing also is referred as hoping or multihoping.

[2] Sink also called BS (Base Station) or Gateway. In this work will be used BS.

Overhead	Complexity setup		Overhead
Low Signaling	High cost scalability		Signaling
	Static nodes	Asyncro	
	Homogeneous distribution		

Table 5 clustering vs flat constrains and advantages

2.2. WSN available routing protocols

2.2.1. For hierarchical or clustering protocols.

Hierarchical[3] network defines a specific topology very structured. For this goal, many kind of protocols are defined. All of them have the main target, to setup and manage the structure. By the way, a lot of modifications are needed in those protocols because the location of sensor, many times, makes difficult to spread a hierarchical sketch.

Here below the most representative algorithm are described with some of the upgrades.

LEACH. All the Clusterheaders can reach the sink. In setup every node looks for a Clusterhead and joins. To avoid hotspots Clusterheads turns. The schedule for data sending from nodes to the heads is TDMA. Some papers say 5% Clusterheads is a good proportion.
Setup steps: 1) Clusterheads is generated randomly among nodes 2) Each sensor node *i* generates a random number such that 0 < random < 1 and compares it to a pre-defined threshold T(*i*). If random < T(*i*), the sensor node becomes Clusterhead in that round, otherwise it is cluster member.·3) Clusterheads send broadcast messages to all nodes to inform the status of them. 4) Non-Clusterheads join to the cluster with the best signal strength message. [REF. 3]

M-LEACH. Multihop-LEACH. Avoid the necessary homogeneous distribution of nodes in LEACH. This protocol follows LEACH's algorithm, but allowing that not all the
Clusterheads reach the BS. That means, a 1st layer for Clusterheads are linked to BS, 2sd layer of Clusternodes are linked to 1st layer and nodes join to that 2sd layer.
Setup step: 1) BS broadcast (SETUP Control Packet) then all CH reply with ID_{chx} at default lowest signal. 2) BS broadcast Control Packet with list of layer 1 ID_{chx}. 3) Non layer 1 ID reply with ID_{chy} at default lowest signal. 4) Layer 1 CH_x relay (before reply non layer 1 CH_y) with ID_{chx} and ID_{chy}. 5) Sink store ID_{chx} as Layer 1 then broadcast Control Packet: ID_{chy}+ID_{level}+ID_{chw} (above level CH). Repeat process till no new CH discovered. [REF. 4]

[3] Also referres to clustering.

PEGASIS. Find a chain of lineal node to the BS. The nodes will be organized to form a chain using a greedy algorithm starting from some node, one node of chain send an aggregation of data, of all chain partners, to the BS. The PEGASIS's authors start with the furthest node from the BS. Alternative BS can compute this chain and send broadcast to all sensor nodes to save computational energy in nodes. It is necessary all nodes reach the BS. The node which sends data per chain to sink is randomized and also could use a token in the chain. For the furthest node of the chain (a threshold is set) protocol will avoid transmitting so as not spend much energy in far transmissions. That threshold is adaptive to the remaining energy levels in nodes.[REF. 5]

2.2.2. Flat protocols:

As the main different among flat and hierarchical protocols is the metric. In order to calculate the best route from node to the sink some parameters of network state are used. Evaluating these parameters a metric about routing path cost is kept. To update the table some signaling (Control Packet) are used. Those methods can be reactive o proactive. Proactive means make routes before send. Reactive means make route when it has something to send.
Here below, a description of the most representative algorithms.

E-WME. The algorithm defines for each node U the cost C which depends on the available energy E, the battery capacity B, the harvesting power rate P and the reception and transmission energy. The algorithm calculates the shortest path from the source to the destination with respect to this node cost.
To calculate the shortest path from the nodes to the base station Dijkstra's algorithm is used.[REF. 6]

HESS. This protocol is an improvement of EAP (Energy-Aware Protocol). These kinds of protocol sense the energy available in the node network to choice the path. HESS has the following tags in his routing table: Next ID_{node} to Hub, Metric or Cost Benefits.
(Cost-Benefits: means the metric of a path comes from calculates the high residual energy and high harvesting energy rates).
This protocol is reactive. When a node wants route a packet, it will use some energy to send a CFR (Cost Function Request) to his neighbors. [REF. 7]

SEER. It selects the next hop choosing a neighbor that has a smaller or equal hop count. If multiple neighbors satisfy this, remaining energy is chosen as next hop.
Each node has a table: $ID_{neighbor}$; hop count; Energy remaining.
This protocol is proactive: periodic signaling messages sent through the network for updating tables.

Also exist much more examples of routing protocols, would be impossible to describe all of them. Below a table with the main features of other ones.

Name	RRS	Energy Aware	Energy average	Setup/ Update	Schedule/ Signaling	Multi hop	Size cell	Simple
MR-LEACH	Yes		RE, T(i)	Rounded. Broadcast/ Flooding	TDMA syncro	Yes	Fix	BS flooding (over load)
HEED	AMRP/ Yes		RE, PQ Threshold	Time Trigger. Broadcast	DSSS	Yes	Variable	CH continous
EEMR	Yes		RE, PE.	BS Broadcast.	TDMA	Yes/ SPF	Variable/ Smaller close BS	Broadcast. BS Overhead
EECB			RE/ BS length		Not defined	Yes/ Chain Tree		Not defined
EAP	Yes	H. C+ PE+ PL		Broadcast/ Reactive Flooding	Control msg. Maintenance msg.	Lightweight metric.		Continuos. Proactive

Table 6. Legend: H.C.=Hop Count. PE=Power Tx Next Link. PL=Quality Link, RE=Residual Energy, HV=Harvesting, QU=Queue/Work Load. PQ= Power Capacity.

Name	Remark
MR-LEACH	BS must research all nodes. Scalability limited.
HEED	Nodes Tx power adapted to cluster size. Node join to CH 1)(AMRP),2)Closest.
EEMR	BS make all. All CH reach BS.
EECB	Avoid Long Link.
EAP	Energy Aware basic.

Table 7. TP: Time processing: CPU+ signaling.

2.3. Topipcs of researching

After introducing in the WSN small size but depth research chances, a sketch of the main topics obtained.

So the most important constrains will be:

- Battery capacity.
- Power Tx/Rx consumption.

And the main targets of researching will be:

- Scalable for any size.
- Limit number of transmissions.
- Mobile nodes.
- Data heterogeneity.

With the list of constrains, the list of targets and taking references from the protocols improvements in Table 6, below a summary of the main topics to achieve a good WSN routing protocol.

Clusterize the nodes to take some advantage of data aggregation. Study efficiency aggregation over work-load, it is possible to reduce the data load of a network thanks to shrink data size.
Multi-hoping is the way to achieve a good scalability, no constrains to reach the BS directly, a node always can find the least energy cost link, on the other hand if a node can reach the BS might find another node which it can.
Study the effects of short multihop path loss against long links.
Use energy-aware metrics to choose the lightest shortest path. For this target use all parameters available to calculate an accurate metric. An important parameter is Received Signal Stretch (RSS) to get information about network topology and nodes distance instead of GPS technologies.
It will be important take into account the differences in energy consumption using proactive protocols in front reactive.
As power supply capacity is reduced; do not use ACKs, if it is required, use ACK packets to compute information of network like signaling, keep-alive or updating.
Finally try to put the heaviest signaling at the BS.

Then, as I can see, always there is a fight among both topics: focus the development in a clustered way or in a flat way.

2.4. Introducing in WSN development

Setting the rules to follow to build a new improved WSN protocol I do not forget the considerations made before. Here below a table with a summary obtained from literature and my own ideas.

Setup				Signaling	MultiHop	
CH assigment	CH distribution		Cell size	Broadcast CDMA Proactive	Routing	Metric
Energy average Energy efficiency	CASE 2 (seen after)	CASE 3 (seen after)	RRS Aggregation Work Load		SPF	Energy Aware
Residual Energy Power Capacity T(i) RSS					Hop count and metric value.	Hop Count Residual Energy Harvesting Energy Work Load RRS

Table 8. Summaty of main topics for WSN routing protocols development

Experience got before is the best base to build next. Table 8 will be used in whole work as a reference of ideas. Many decisions taken during the work are following the topics described there.

3. Suggested questions

First of all I will consider a mixed protocol structure which adapts their performance and behavior to the current scenario. Then it must be necessary to join a clustering behavior protocol with a flat protocol. The clustering part of the protocol give improvements over data aggregation, the flat part gives scalability.
But both of those parts have a compromise with:

1) How to build the clusters: CH distribution, CH assignation, Nodes joining.
2) How to discover the shortest path, in a flat topology, with a low data signaling.

A bad management of cluster building could be terrible. Also following the management of actual shortest path protocols, whose use a huge among of signaling so it is bad for WSN.
For the first point I would use the references of:

- HEED: over his well-distributed cluster building protocol (AMRP + intracluster CH negotiation).
- EEMR: idea of variable radio cell to optimize the power consumption of CH. Authors suggest small cells close to BS, maybe this concept should be reformulatedm. [REF. 8]
- EAP: Proactive protocol and references over signaling.
- HESS: Accurate use of parameters to calculate the SPF metrics.

In this way there are some studies [REF. 8][REF. 9][REF. 10] where the authors try to obtain formulas to determine the parameters of clustering settings. Those papers suggested me to go further, not only focus my effort finding a new behavior for a routing protocol but first research the performance of nodes, evaluate the main topics, find the parameters that change that performance, and based on that, define some rules to guide the design of a wireless sensor network routing protocol.
For getting a new knowledge, is needed to simplify the topics, also to split the problem in small ones, and after, add the solutions. In that way we have to pay attention to split the questions in uncorrelated smaller ones, if not, the solutions would be wrong.
Most of the clustering protocols authors uphold this kind of performance because exist a data aggregation improvement. I would like to fix this theories doing a little study I will call CASE 1.

In EEMR paper [REF. 8] authors suggest to build small clusters close to the BS with the main idea to spend few energy in clustering management a save this to relay the data from further CH. I would like to evaluate that conclusion and check the results. I will call CASE 2.

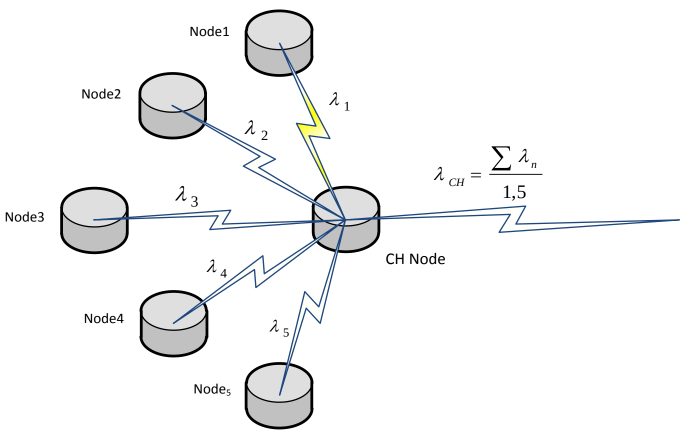
Other important question in wireless communications is the power energy required to send a signal is not lineal with the distance. Then in a WSN we will have lineal energy uses mixed with non lineal. It is necessary to split this problem and study in different ways also find if exits a breakpoint among the lineal energy behavior and non lineal.

For example, the free space path loss have a squared rising over the distance. There are many path losses models then I made, in this paper, a special point for talking about it, point 4.2.1.1 and 4.2.1.2].

Looking up other path losses models is possible to confirm that Free Space Path Loss is the less restrictive on long links.

3.1. CASE 1

The authors of [REF. 11] speak about the profits of aggregation. The idea is to combine the data coming from different sources: eliminating redundancy, minimizing the number of transmissions and thus saving energy. However since various sensor nodes often detect common phenomena, there is likely to be some redundancy in the data the various sources communicate to a particular BS. This paradigm shifts the focus from the traditional address-centric approaches for networking (finding short routes between pairs of addressable end-nodes) to a more data-centric approach (finding routes from multiple sources to a single destination that allows in-network consolidation of redundant data). Although data aggregation results in fewer transmissions, there is a tradeoff: potentially greater delay in work topology.
Then, in this case, will be very important to know the data shrink behavior. Data shrink behavior might not be lineal through all the network and the delay time network suffer due to the aggregation processing.



Picture 2. Simple aggregation example in clustering

If we consider the case where λ1= λ2= λ3= λ4= λ5 the result is:

$$\mathbf{CH_{Data}} = \frac{\mathbf{1}}{\mathbf{1,5}}\boldsymbol{\lambda}_1 + \frac{\mathbf{1}}{\mathbf{1,5}}\boldsymbol{\lambda}_2 + \frac{\mathbf{1}}{\mathbf{1,5}}\boldsymbol{\lambda}_3 + \frac{\mathbf{1}}{\mathbf{1,5}}\boldsymbol{\lambda}_4 + \frac{\mathbf{1}}{\mathbf{1,5}}\boldsymbol{\lambda}_5 = \mathbf{3,33}\boldsymbol{\lambda}_1$$

As it is possible to see in Picture 2, if the clusterhead node achieve an aggregation of 1,5, the data rate resulting from the aggregation of a cluster of 5 nodes it is 1/3. In this case I used simple values, like same traffic rate for all nodes, to make the point easier to understand.

Thanks to aggregation, a network can simplify among of energy used to transmit data, that is true, but must modify the topology and the spreading of the links between nodes? In [REF. 11], authors base their routing algorithm on data aggregation improvement, so I will discuss the relevance of it over others WSN parameters.

For this topic is needed to look over how the data compression works through the entire network, also compare the gain of aggregation with other parameters and evaluate his importance in network behavior.

3.2. CASE 2

As I introduced before [REF. 8] the authors suggest the best way to add nodes to a clusterhead is assign few nodes to the CH closer to the BS and a more to the furthest. In my first reading I considered that as a great idea, but later, thinking in that idea I guess may be the opposite way to spread nodes to CH is the best, that means make the biggest cell close to BS, so if a node reach to CH with the same metric but one closer to the BS than other, chose the first.

But all that idea is just a supposition then I discussed this point later in 4.1 with a sketch and some math.

3.3. CASE 3

One of the main problems of wireless telecommunications is the uncertainty of the environment losses. In a wired transmission the cable has parameterized almost all their features, but in a wireless environment exist too much randomized variable.

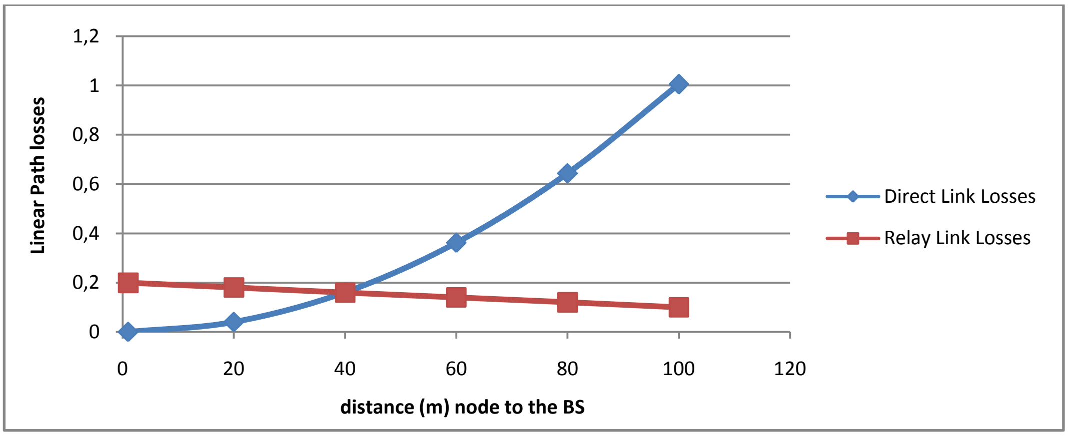
Thanks to many researchers who tried to study those environments and parameterized his behavior, we have propagation models to predict, with several certainties, variable path losses.

The main topic in this case is: path losses are never linear with the distance. Regarding the unified propagation models [REF. 12], those losses are exponential with distance.

According to report [REF. 10], above sentence is reinforced, who says that signal propagation follows an exponential law to the transmitting distance. Therefore, minimizing the amount of data communicated among sensors and reducing the long transmitting distance into a number of short ones are key elements to optimizing the communicating energy. This sentence becomes an important topic for this document. And the main issue of CASE 3: knows different performance among long links energy uses and short links with some hops.

Also there are more parameters in an environment that affect losses, diffractions, reflections, intermediate elements (walls, ceiling, trees…), all of them increase the slope of exponential losses.
In that way I use the simplest model, Free Path Loss.

Path losses can represent the energy spent by node to send a bit. More losses more Tx energy needed to arrive at destination. As it is possible to see in the above budget, every node in a flat topology has to use the same energy than a cluster node to send a 1 bit five times closer. But remember the FSPL it is not linear and relaying yes, then it is possible to arrive to a break-point.



Picture 3. Graph Break Point related Direct Link Losses VS Relay Link Losses

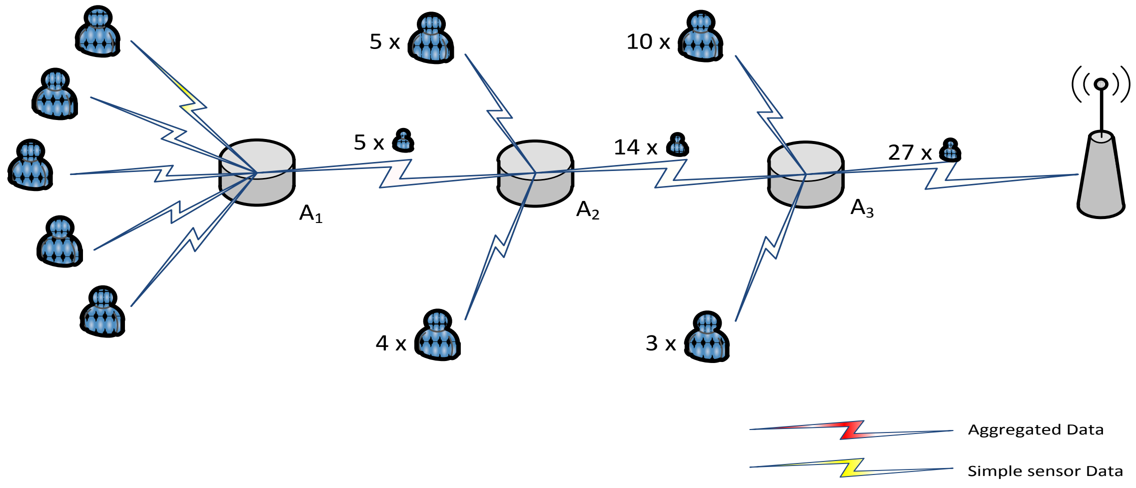
Summarizing, in the cells building, for optimizing the energy resources to send the data to the BS there are two main parameters, the exponential distance and the linear relaying energy uses.
The next discussing of those two parameters will reach the best and optimized networking shape. Also will guide to the best performance of a WSN, means, when is needed clustering or relaying, which has to be the size of those clusters, the most relevant parameters for node metrics, and so on. Shortly it will be showed deeply.

4. Discussing the CASE 2 & 3 another related topic.

Ones introduced the *CASEs*, below I will continue with the resolution of them. This resolution, also, give the first rules and ways to follow in a design of routing protocols.

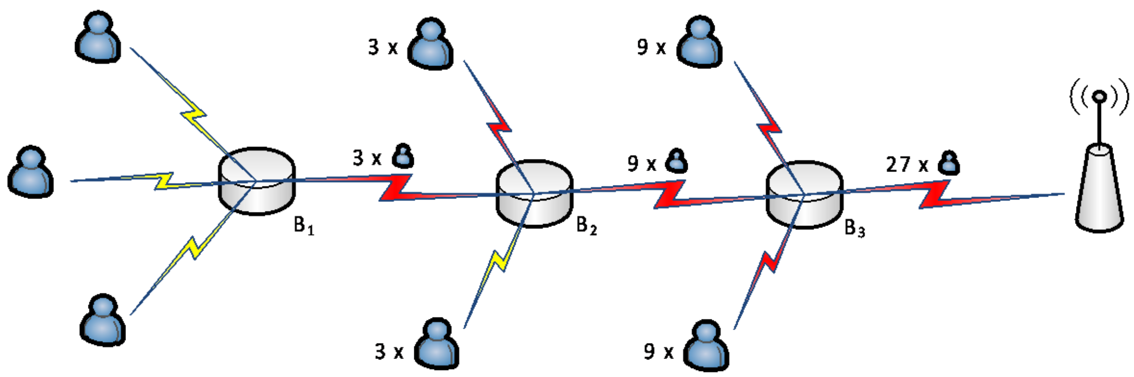
4.1. CASE 2:

As it is explained above may be should be reformulated some questions released by some authors when they declare that the cell have to be small closer to the BS and larger in further distance.



Picture 4. Simple situation where bigger cluster are further.

In Picture 4 a distribution where the biggest cluster are far from the BS. Lastly the chain drawn relay a total of 27 data rates. To simplify the study I assign the same data rate to each user. In case of a real world the data rate would be different but in this case it is possible to apply superposition.
Let see what happens in a scenario where the cluster are bigger close to the BS.



Picture 5. Simple situation where bigger cluster are closer.

Again, this chain relay 27 data rates. In this case the data rates are aggregated to the CH as close as possible to the BS.
Both cases, Picture 4, Picture 5, relay the same among of data. That is the main topic, both settings have the same behavior for the final node, but as it is possible to see, the performance is different for the intermediate nodes. Loads of nodes are different.

Setting largest cell close to BS allows unload B2 and B1. Comparing traffic load: B1<A1, B2<A2, B3=A3.
So, regarding after a slot of time, the links among nodes change, the B sketch offers more opportunities because the amount energy used for intermediate nodes are less.
It is true that node #3 in both sketches should use the same energy inside a defined slot time, but in a future next slot, the energy remaining metric will be higher. In this way it is possible to confirm that in the B sketch the time until the first dropped node will be longer.
More energy available for intermediate node means more chances to have energy to find nodes further away and extends the size of the network.
In the worst case the load of both distributions could be the same.
After discussing this case, the routing protocol should try to manage the link finding the nodes closer to the BS.

4.2. CASE 3: clustering & flat performance. cell built.

Now exist two main groups of routing protocols: clustering and flat. Clustering idea is older than use a flat performance of the network using multihoping. Then there are a lot of literature about clustering protocols and many studies of how it works. But some authors started a new branch based on flat and multihoping routing protocols which give different a nice improvements in environments where clustering protocols start to fail.
By the way, knowing that clustering protocols have a good behavior in specific situations and flat protocols in others, why do not try to add the best features of each one. So, protocols which mix ideas began to appear.
As is before comment there are many authors who tries to find the optimum way to spread the links among nodes, or how the network has manage the node connections to get the optimum energy efficiency
As well is explained at the beginning the energy efficiency is the philosopher's stone of the routing protocols. That is the reason always among of energy used or remaining in the node is the main value to manage the routing behavior.

Some authors focus their effort optimizing the way to spread the cluster efficiently. But maybe the question is if clustering is the best performance or the most efficient. After some else reading [REF. 9][REF. 10] started to grow idea of compare the clustering energy behavior against flat energy.
In another way, it is possible to consider a cluster like a multihop between a node and his clusterhead, plus a long link between clusterhead and BS.

Inside that idea of comparing also it is the topic of an existing break point where one kind of topology is more efficient than other. Then will be very important, if that point exist, to know how it works to focus the design of a protocol in that way.

REF. 10 it the main guide to develop all this point and I am based on it. The authors have made a deep mathematic work to present by equations the cluster behavior and multihop behavior.
In the same way and having as example above mentioned work, I developed a basic mathematics which leads me in the knowledge of clustering (long distance) against multi-hopping performance and how their features change.

4.2.1. Math approximation: Chain nodes vs long distance links

First of all and introduction of propagation losses will be made and a justification of why is chosen one of them.

4.2.1.1. Path Loss.

In the study of wireless communications, path loss can be represented by the path loss exponent, whose value is normally in the range of 2 to 4 (where 2 is for propagation in free space, 4 is for relatively lossy environments and for the case of full reflection from the earth surface, the called flat-earth model). In some environments, such as buildings, stadiums and other indoor environments, the path loss exponent can reach values in the range of 4 to 6. On the other hand, a tunnel may act as a waveguide, resulting in a path loss exponent less than 2.
Path loss is usually expressed in dB. In its simplest form, the path loss can be calculated using the formula

$$L = 10\, n\, log_{10}\,(d) + \ C.$$

Equation 1. General path losses

where L is the path loss in decibels, n is the path loss exponent, d is the distance between the transmitter and the receiver, usually measured in meters, and C is a constant which accounts for system losses.
Radio and antenna engineers use the following simplified formula (also known as the Friis transmission equation) for the path loss between two isotropic antennas in free space:
Path loss in dB:

$$L = 20\, log_{10} \left(\frac{4\pi d}{\lambda} \right)$$

Equation 2. Free Space signal spreading.

where L is the path loss in decibels, λ is the wavelength and d is the transmitter-receiver distance in the same units as the wavelength.

Modifications to the basic equation
Empirical adjustments are also sometimes made to the basic Friis equation. For example, in urban situations where there are strong multipath effects and there is frequently not a clear line-of-sight available, a formula of the following 'general' form can be used to estimate the 'average' ratio of the received to transmitted power:

$$\frac{P_r}{P_t} \propto G_t G_r \left(\frac{\lambda}{d}\right)^n$$

Equation 3. Friis ecuation; n=1 for Free-Space

where n is experimentally determined, and is typically in the range of 3 to 5, and Gt and Gr are taken to be the mean effective gain of the antennas. However, to get useful results further adjustments are usually necessary resulting in much more complex relations, such the Hata Model for Urban Areas.

Calculation of the path loss is usually called prediction. Exact prediction is possible only for simpler cases, such as the above-mentioned free space propagation or the flat-earth model. For practical cases the path loss is calculated using a variety of approximations.
Statistical methods (also called stochastic or empirical) are based on measured and averaged losses along typical classes of radio links.

4.2.1.2. Wireless environment model.

Easy approximations for calculating the path loss over distances significantly shorter than the distance to the radio horizon:

- In free space the path loss increases with 20 dB per decade (one decade is when the distance between the transmitter and the receiver increases ten times) or 6 dB per octave (one octave is when the distance between the transmitter and the receiver doubles). This can be used as a very rough first-order approximation for SHF (microwave) communication links;
- For signals in the UHF/VHF band propagating over the surface of the Earth the path loss increases with roughly 35 -- 40 dB per decade (10 -- 12 dB per octave). This can be used in cellular networks as a first guess.

This is a simple approximation of Free Path Loss Formula is ones of the most optimistic.

It is not a target of this thesis to study deeply the effects of different deterministic or empiric path losses models, but it is necessary to have some knowledge of it to evaluate how optimistic or pessimistic is the chosen model.

In all this thesis estimations and simulations the model chosen is Free-Space Path Loss. This is done in the way it is very simple equation and does not need references of the environment, means, if the estimation or simulation is indoor or outdoor, or in a urban or rural area.

Below FSPL model is compared with *“Unified Propagation Model for Wi-Fi, UMTS and WiMAX Planning in Mixed Scenarios”* [REF. 12]. This paper presents an

unified and empirical propagation model to obtain the received power in mixed scenarios, with outdoor and indoor environments, or in a scenario with only one kind of environment, either for an urban, sub-urban or rural scenario, with or without vegetation. This unified can be included into planning tools for wireless communication systems.
As the authors describe in paper, that work is aimed to WiFi, UMTS, WiMAX devices like can be wireless sensor, aimed to outdoor, indoor and rural to urban scenarios, the entire scenarios wireless sensor can be spread.

Proposed model: *"Unified Propagation Model for Wi-Fi, UMTS and WiMAX Planning in Mixed Scenarios"*

The average PL between transmitter and receiver as a function of the distance, is given by Equation 5, where NE is the total number of environments, ENV represents the environment, VEG the type of vegetation, u[.] and u(.) are unit step functions (discrete and continuous, respectively), d_0 is the reference distance, λ is the wavelength, d_{bp} is the breakpoint distance, γ_1 and γ_2 are the PL exponents at different distances (before and after the breakpoint), $\mathcal{X}_f$ and $\mathcal{X}_h$ are frequency correction factors, A_m, d_{depth} and β are parameters for the additional vegetation loss, W_e , W_{GE} , θ, θ_H and θ_V are parameters related to the buildings penetration loss, G_f is the floor height gain, NF is the number of floors, NW is the number of walls, FAF is the floor attenuation factor, WAF is the wall attenuation factor while floor and wall are variables that identify each floor and wall through the path, respectively. A detailed description for these parameters follows:

- The meaning for the breakpoint distance, d_{bp}, is twofold:
  - i. If only the indoor environment is included in a scenario, the propagation loss has two distinct regions, as a function of distance. In the first region, within the 5–20 metre range from the transmitter, the propagation loss is similar to the free space one. For higher distances. The distance at which this transition in propagation loss occurs is referred to as the breakpoint distance.
  - ii. If the transmitter antenna is mounted in an outdoor environment, the breakpoint distance represents the distance between transmitter antenna and the interface for the indoor environment.

Parameters	Wi-Fi	UMTS	WiMAX
f [Hz]	2.4×10^9	2.2×10^9	3.5 × 10 9
d_0 [m]	1	25	100

Table 9. Parameters for d_0

  - iii. When the prediction is performed for an outdoor-to-indoor environment, γ_1 represents the value for the outdoor environment while γ_2 refers to the indoor one; when the prediction is performed for only an indoor environment, γ_1 represents the PL exponent before the first obstacle, i.e., up to the breakpoint distance, while γ_2 represents the PL exponent after the breakpoint.

Parameters	Indoor	Mixed		
	Wi-Fi	Wi-Fi	UMTS	WiMAX
γ_1	3	3.85	From Equation 4	From Equation 4
γ_2	2.5	3.35	$\gamma_{1UMTS} - 0.5$	$\gamma_{1WIMAX} - 0.5$

Table 10. Parameter for distance exponents

When the prediction is performed for a single outdoor environment, only γ_1 is considered. In the case of UMTS and WiMAX technologies, γ_1 is given by the following equation:

$$\gamma_1 = a - b_{[m^{-1}]} * h_{b[m]} + \frac{c_{[m]}}{h_{b[m]}}$$

Equation 4. Value of Y1 for UMTS and WiMAX

where the parameters a, b & c, are constants that characterize the type of scenario involved, urban, suburban or rural, and their values are defined by the experimental results.

$$PL(d)_{[dB]} = \left\{ \begin{aligned} & 20 \cdot \log_{10}\left(\frac{4\pi \cdot d_0}{\lambda}\right) + 10 \cdot \gamma_1 \cdot \log_{10}\left(\frac{d}{d_0}\right) \cdot \mathrm{u}\left(d_{bp} - d\right) + \\ & + ENV \cdot \left[X_f + X_h + VEG \cdot A_m \cdot \left[1 - \exp\left(\frac{-d_{depth} \cdot \beta}{A_m}\right)\right]\right] + \\ & + \mathrm{u}[NE - 2] \cdot \left[W_e + W_{GE} \cdot (1 - \sin\theta)^2 - G_f \right] \cdot \mathrm{u}\left(d - d_{bp}\right) + \\ & + (1 - ENV) \cdot \left[\begin{aligned} & \mathrm{u}[1 - NE] \cdot \left(\begin{aligned} & 10 \cdot \gamma_1 \cdot \log_{10}\left(\frac{d_{bp}}{d_0}\right) + \frac{WAF_{bp}}{\sin\theta_H} + \\ & + \frac{FAF_{bp}}{\sin\theta_V} + \sum_{floor=2}^{NF} \frac{FAF(floor)}{\log_{10}(10 \cdot floor)} \end{aligned} \right) + \\ & + 10 \cdot \gamma_2 \cdot \log_{10}\left(\frac{d}{d_{bp}}\right) + \sum_{wall=2}^{NW} \frac{WAF(wall)}{\log_{10}(10 \cdot wall)} \end{aligned} \right] \cdot \mathrm{u}\left(d - d_{bp}\right) \end{aligned} \right\}$$

Equation 5. PL due to “Unified Propagation Model for Wi-Fi, UMTS and WiMAX Planning in Mixed Scenarios”

After this long description of the introduced model, the most relevant parameter aim of this discuss are shown. These parameters are all whose have influence upon distance. Those are: d_0, d_{bp}, γ_1 , γ_2.
Path Loss is the addition of: d_0 can be assumed as Free-Space Path Loss and more function (d) parameters whose have γ_1 and γ_2 exponent. As is possible to watch in Table 10 both are higher than 2.

Comparing Equation 3 and Equation 5 is easy to summarize that Free Space Path Loss is more optimistic than other. Just in case of using Free-Space Path Loss it must take into account in a real environment the result would be worse.

4.2.2. Mathematics modeling.

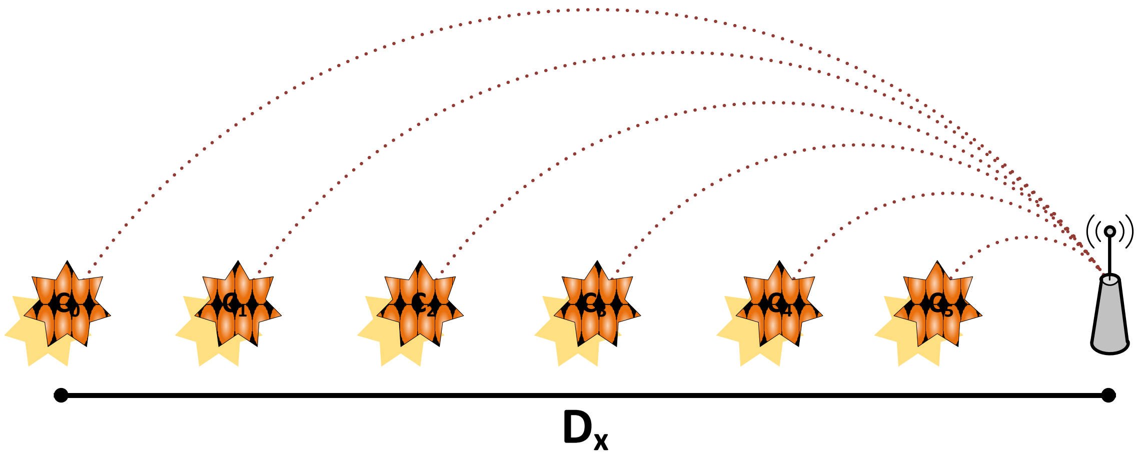
Starting the study of clustering against multihoping topology I will start with simple models, two chains. For every chain it will be formulated their performance. Following it has been writing before, for this modeling is used a path losses model. I decided FSPL because is less restrictive, so in a real environment all conclusions get will be better. Also for this topic it is not essential to know parameters of a real environment.

$$P_{Rx} = P_{Tx} G_{Tx} G_{Rx} \left(\frac{c}{4\pi f d}\right)^2$$

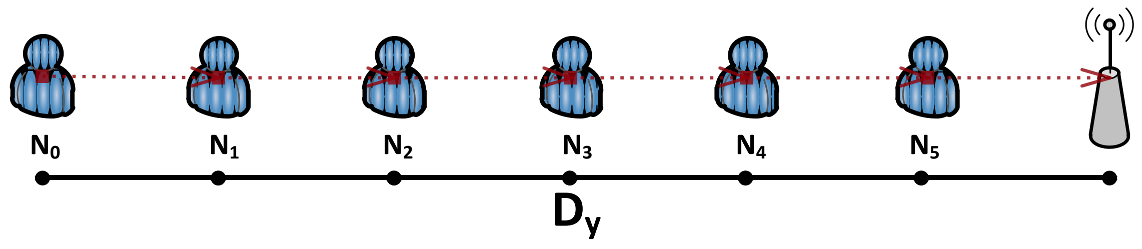
$$P_{Tx} = \frac{P_{Rx}}{G_{Tx} G_{Rx} \left(\frac{c}{4\pi f}\right)^2} * d^2$$

$$P_{Tx}(W) = \frac{P_{Tx}(J)}{s} * \frac{1}{Tasa\left({}^{Kb}/_{s}\right)} = \frac{P_{Tx}}{Tasa}\left({}^{J}/_{bit}\right)$$

Equation 6. Free-Space Path Loss model



Picture 6. Cluster links modeling



Picture 7. Multi-hopping links sketch of a Flat modeling

- Dx = Dy = Maximum length that a node can reach.

- N=5 (number of nodes in a chain) .
- k=2;

Direct Links: $P_{total\ Tx} = \sum_{n=0}^{N} P_t \left[\left(1 - \frac{n}{N}\right) D_x\right]^k * \eta * A$

Multi-hoping Links: $P_{total\ Tx} = \sum_{n=0}^{N} P_t \left[\frac{D_y}{N}\right]^k \left((n * (\eta + 1) + 1)\right)$

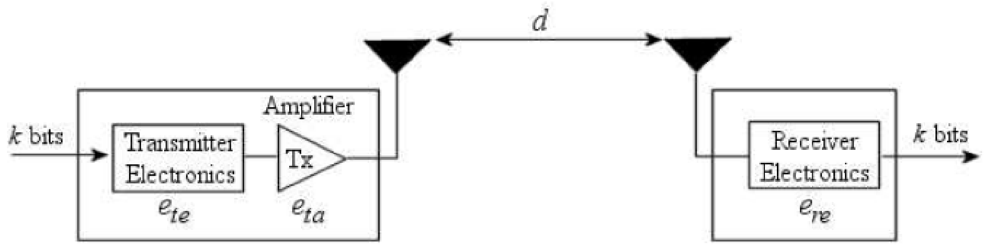
$$P_t = \frac{Device\ Sensitivity}{G_{Tx} G_{Rx} \left(\frac{c}{4\pi f}\right)}$$

$$\eta = change\ load\ due\ to\ Rx\ energy\ use$$

$$A = node\ belonging\ a\ Cluster$$

4.2.2.1. Previous point study of η.

It is very important to analyze the performance of typical wireless device.



Picture 8. Wireless device scketch

All of them have a Tx and Rx electronic. Both uses among of energy from the power supply.
In Picture 8 we can see that ***e***$_{te}$ and ***e***$_{ta}$ are the Tx energy values and ***e***$_{re}$ is Rx energy value. The main characteristic of those parameters are:

- ***e***$_{te}$ always is rather smaller than ***e***$_{re}$
- ***e***$_{ta}$ is the value of $\equiv \frac{Sentitibity}{G_{tx} G_{rx} \left(\frac{c}{4\pi f}\right)^k}$
- ***e***$_{re}$ is a fix value and not dependent localization of node

If a node is set in relaying work, have to switch on the receiving part of device, means ***e***$_{re}$ energy use. On the other hand a direct sending node just needs to activate Tx. This is very important because in a multihop node, as a difference with direct link nodes, must to add or take into account there is ***e***$_{re}$.
The main point of WSN an also this document is study behavior of energy node inside a network, so all energy uses or highest have to be counted. Energy to manage the receiver electronics is important.
In the previous point I introduced the parameter η as the relationship among Tx energy and Rx energy. Using η, receiving energy costs are counted. And the study of relaying behavior and direct link is accurate.

$$\eta = \frac{Power\ Electronics\ Rx}{Power\ Electronics\ Tx}$$

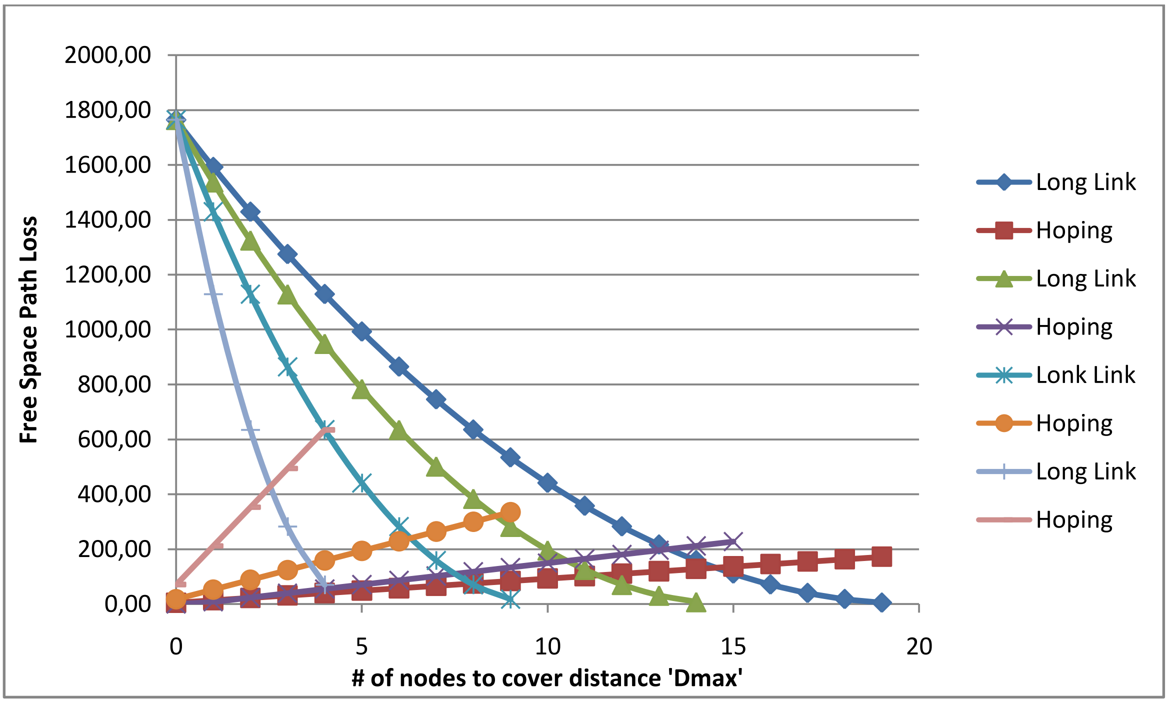
In the future there are some simulations to know the nodes behavior in a WSN it is important to lookup some values of real sensor devices.

Although this is a theoretical introduction of researching above idea is taken to do not be away from real world.

4.2.2.2. Comparing Multihop and Direct link node performance

Distance (Units)	η	k
5000	1	2

In a graph:



Graph 1. Path Loss comparing long distance vs relaying

This first approximation in Graph 1, there is a numerical comparison among a usual clustering (long links) topology against a flat multihoping (close multihoping links) topology.

This graph give a first overview of the behavior of wireless devices, it is just an approach, because every wireless device has his own behavior, constrains and features, this is only a simple view. Anyway let me explain some of an important point shown by Graph 1: in clustering routing protocol a clusterhead is selected

from many nodes of an area, the protocol never looks the length from this clusterhead to the BS, that makes, many times, long links. As has been introduced in CASE 3 more distance from a node to the BS higher is the effort in energy has to manage to reach his destination, due to exponential characteristic of Path Losses function. That is a lack for long length relays.
On the other hand in the same graph there is, over long link response, how many energy has to manage a node to relay data in a hop by hop. As it is possible to see his behavior is not exponential. Although hoping links has the same Path Loss function, shorter distance between nodes do not allow realizing that shape.
D_{max} means the maximum distance that a Tx electronics' node can reach. The number of nodes inside a hoping chain assuming the minimum Tx power set in a node allows to built those chains.
The most important issue of this graph is the point where the line with the same number of nodes overlaps.
I will call this as Q point. This point marks an area where directs links manage less energy than multi-hoping. I think this will be an interesting idea to study in a future, now is out of the targets of this work.

4.2.2.3. Simulation of networks structures

I think above comparison emphasize the first difference there are in the WSN protocols, following some strategies the result can be very different. So I decided to spend more resources in this way.
Trying to get a depth knowledge of which are the different among a clustering far distance and a hop by hop shorter distance performance, let's use a simulation where almost all topologies are defined an, after applying some data, shows how nodes uses the energy.
In this simulation any routing protocol are used. Links among nodes are spread balancing the number of joins on each of them. To share as well as possible the load, and obtain accurate results over path losses costs.
Once known the philosophy of this simulation is needed to select some value for the energy parameters, in this case as a difference of some other authors, I will look for some of the commercial devices released by company builders to get, as a reference, their features.
This has been a personal decision as a researcher trying to put this simulation close to real world.

Searching data sheet information of different products, first surprise comes up.
These devices can be classified in two sections:

- **Non shift Power Tx transmission**: This kind of products always delivers the same power out energy to send data. Then no improvements might be done over it by a routing protocol, to manage better the energy uses. In case to use these devices direct links are the

way to manage a network. If you must always use the same power to send all traffic, try always to reach the BS directly. Any other strategy is an energy waste. Just in case that a node can not reach the BS try to fins another node which relays data.
And example of this is: Imote2 platform (see ANNEX A).
Important to say and might be obvious this work is not oriented to these type of platforms.

- **Discrete setting up of Power Tx transmission**: This devices can manage his power Tx in some discrete values. That can be used by a routing protocol which selects the accurate power to send data. Also allows reducing the energy use due to the node location or how far is from another.
  Example is: Atmel ATmega platform (see ANNEX B).

	Power out (dBm)	I power supply (mA)
Tx	3,5	15
	1,5	10
	-2,5	9
	-16,5	7
Rx		12

Table 11. Atmel ATmega power features

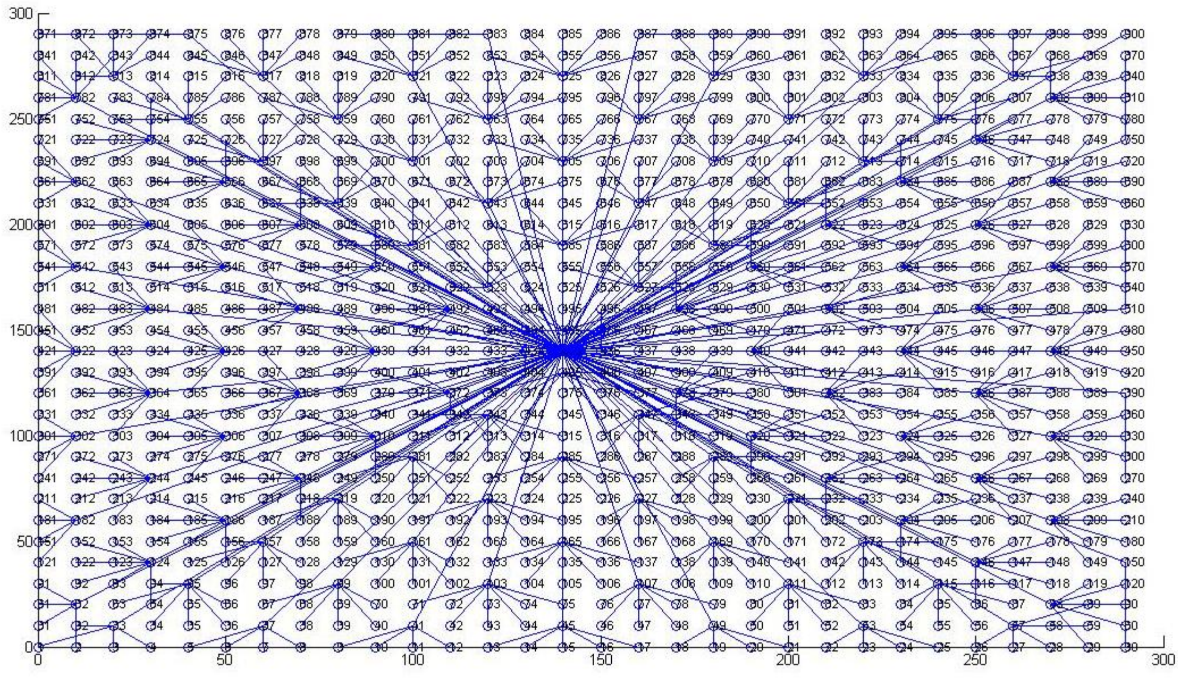
From Table 11 it is possible to get a reference of the parameter η introduced before. Based on it the range can be [0.8, 1.5]. This will be applied in next simulation.

Simulation will be done in MatLab, choice of the MATLAB simulator is made based in the author's previous experience. Other simulators, like OMNET++ or NS-2, were also suitable for the simulation development but their learning curves were out of the work time plan.

The algorithm wrote for this simulation will be trying different schemas of link nodes, from clusters to long chains.

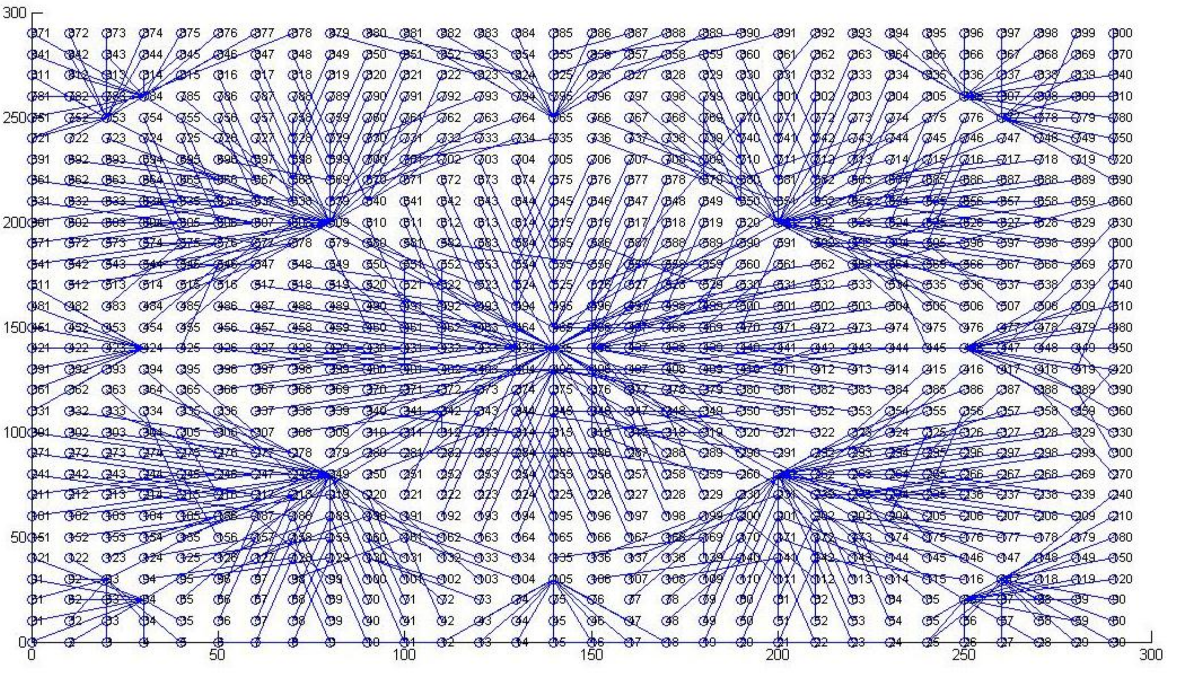
Node location is squared because it is the easiest to analyze the result because exits symmetries.
The number of nodes is 900 that allow to performance many structures in this space.
The results take values from path losses. Means how much power each node has to manage over path looses, every node has an initial value of 90000 and power

path loss to send a bit is subtracted from it. The length among nodes would symbolize the minimum reachable with the minimum set power Tx.

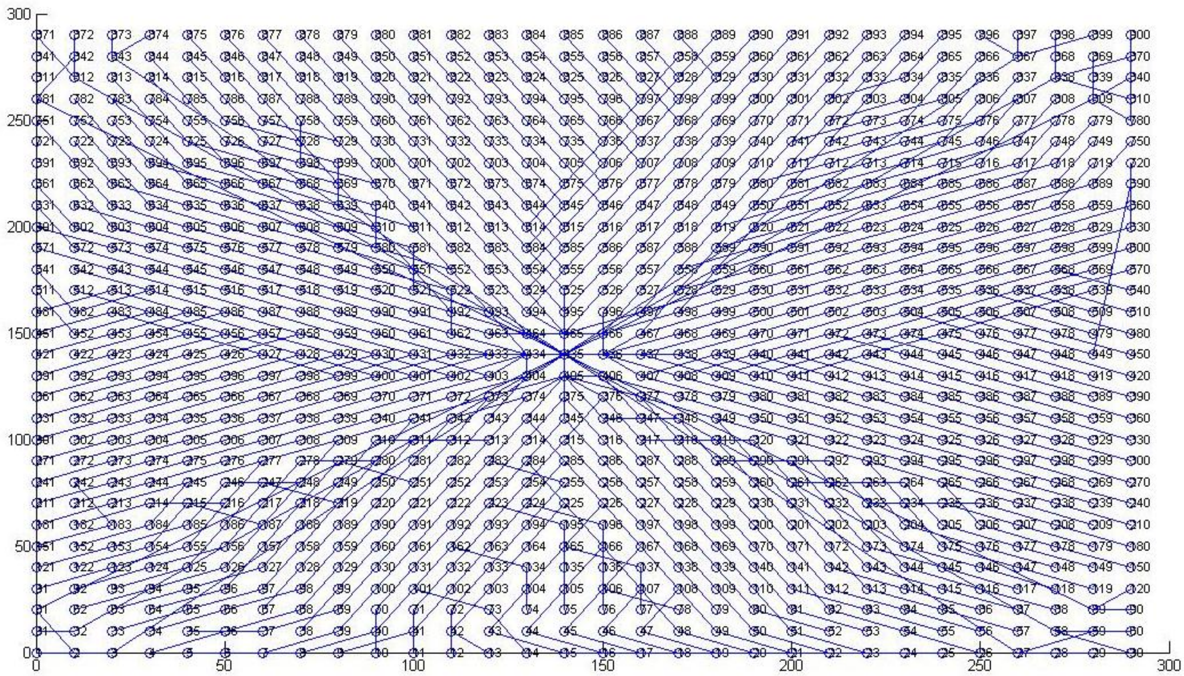


Picture 9. Clustering shape with twice Min Power Tx size

Let’s see pictures of node schemas to understand better how the algorithm works.



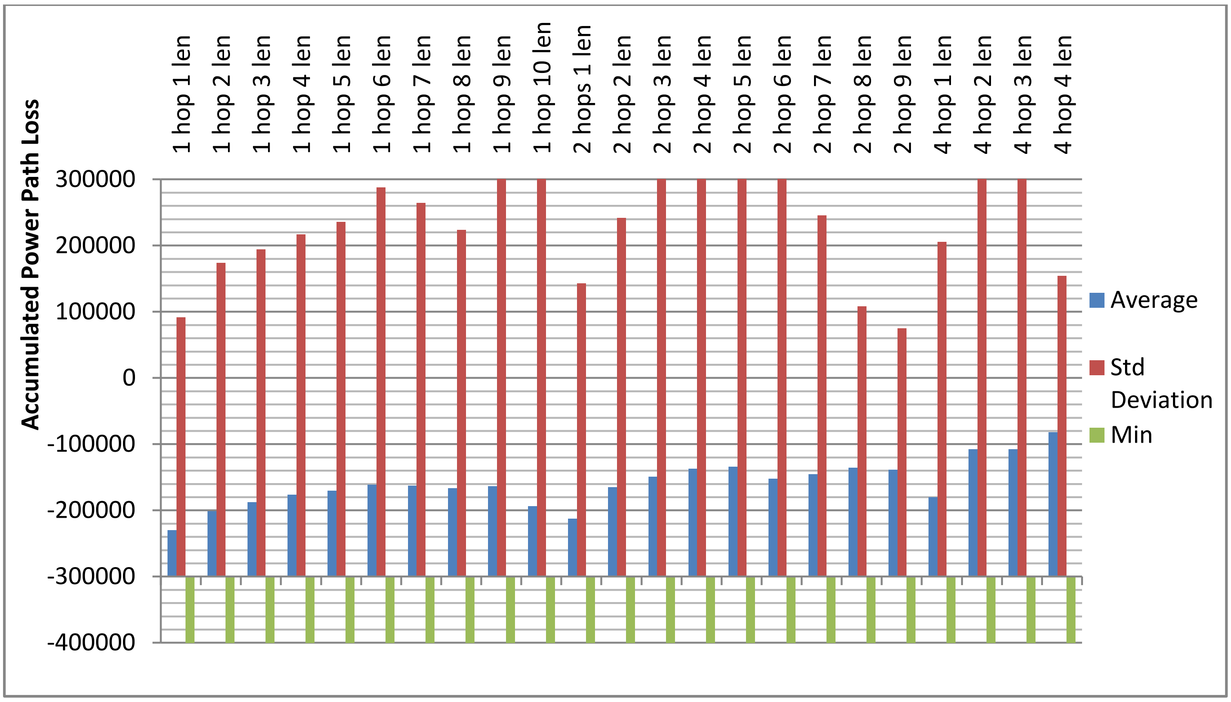
Picture 10. Clustering shape with four times Min Power Tx size



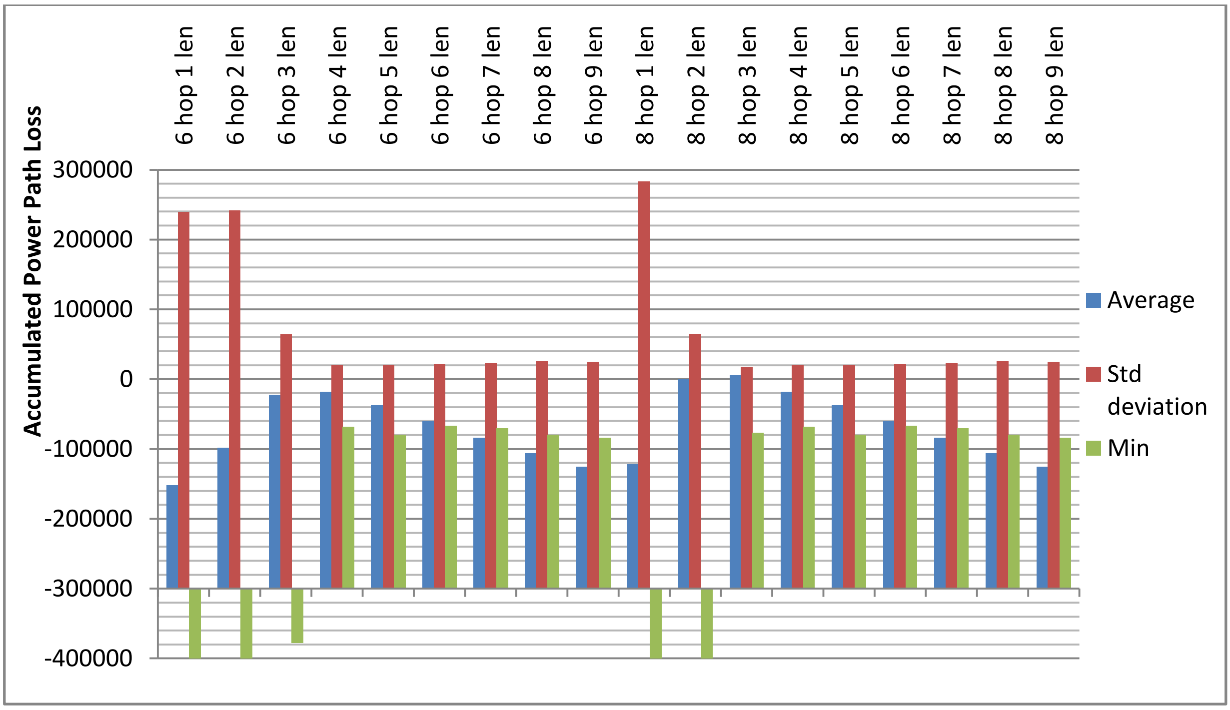
Picture 11. Flat multihop until 8 hop from BS

The results are huge because have been managed 900 nodes, then it is impossible to show a list node by node, instead of this I use statistics functions as average and standard deviation to show the results.

- **Average:** will mean the power used by most of the nodes and it is a reference of which is the effort of each one.
- **Standard deviation:** will show with is the different from the less loaded nodes over the average. This value is important because emphasize how the uses of power are spread. Big values will means that some nodes use a lot of power an others few.
- **Min (Minimum)**: value can show how deep the load is over a node.



Picture 12. Results of topologies close to Clusters



Picture 13. Results of topologies close to Flat Hoping

Comparing all structures shown in Picture 12 and Picture 13 the best results comes from *10 hops 2 len* and *10 hops 3 len*. Then a deep overview comparing cluster with long distance sent works worse than flat hoping topology. Also there is marked the influence of Rx/Tx ratio in *10 hops 2 len* where Rx power use it is a little lack.

This let me introduce a nice reading [REF. 13] of a great work about the optimum node radio according to his electronic features.

5. Study of Aggregation

A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants, at deferent locations. Sensor nodes take lesser power for computation than the power need for them to transmit sensed data. Sensor Nodes try to reduce amount of data needed to send to base station by using various techniques such as data aggregation, data compression etc. There are some application need some time aggregated data and some time compress data. Hence a common approach is propose which reduces overhead and deliver either compress data or/and aggregated data, or both in some ratio base on what base station is demanding. [REF. 22]

In recent years, there is a growing interest in methodologies aiming at combining different sources of information, usually from several surveys. Parallel questionnaires, panel survey, tentative of enriching basic surveys through specific questionnaires (Santini, 2001) may often require such techniques usually named as data fusion (Aluja-Banet et al. 2007).

The objective of sensor routing algorithm is then to jointly explore the data structure and network topology to provide the optimal strategy for data gathering.

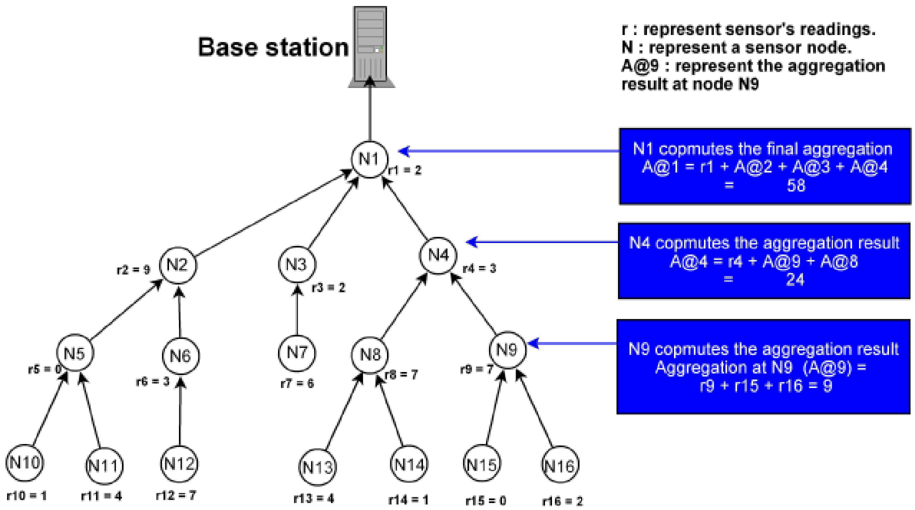
[REF. 20] Routing with data aggregation can be generally classified into two categories: routing-driven and aggregation-driven. In routing-driven algorithms, data is routed through shortest paths to the sink, with aggregation taking place opportunistically when data flows encounter. In aggregation-driven routing algorithms, routing paths are heavily dependent on data correlation in order to fully benefit from information reduction resulted from data aggregation. In this paper, we will use “aggregation” and “fusion” interchangeably, denoting the data reduction process on intermediate sensor nodes.

Regardless of the techniques employed, existing strategies miss one key dimension in the optimization space for routing correlated data, namely the data aggregation cost. An optimal routing algorithm needs to jointly optimize over the transmission and fusion costs in order to minimize the total energy consumption.

Data aggregation

Sensor nodes take less power for computation than the power need for them to transmit sensed data. Hence it is always preferable to reduce the amount of data that need to be sent to the base station by processing sense data inside the sensor network and then send the process data.

One such approach is data aggregation in which data sense by various sensor nodes are combine in some phasing and process data by applying one of the aggregate operator (max, min, avg, sum, div, count, and so on…) on sense data at the time of combining at deferent sensor nodes. This extracts certain statistical quantities from the sensory data, other information is thus lost and hence this aggregation technique only applies to particular applications that require limited information from a WSN. Data aggregation helps us to reduce amount of data that need to send to base station which consecutively improve life time of sensor nodes as well as sensor networks.



Picture 14. Data Aggregation with SUM function example.

The WSN in Picture 14 contains 16 sensor nodes and uses SUM function to minimize the energy consumption by reducing the number of bits reported to the base station. Node 7, 10-16 are normal nodes that are collecting data and reporting them back to the upper nodes whereas nodes 1-6, 8, 9 are aggregators that perform sensing and aggregating at the same time. In this example 16 packets traveled within the network and only one packet is transmitted to the base station. However, the number of traveling packets would increase to 50 packets if no data aggregation exists. This number of packets has been computed for one query.

Data compression

As is introduced in above point, where some aggregation cause the loosing of data, there are some applications which require knowledge of complete data for some analysis. In such applications data aggregation can not be use as base stations receives only process data. Another way to reduce amount of data need to send to base station without losing any knowledge of complete data is to use data compression techniques. In this technique data is gather at some intermediate node where size of data need to send is reduce by applying compression technique.

When the amount of data to be transmitted is reduced, the effect is that of increasing the capacity of the communication channel. If data compression is used in a data transmission application, the goal is speed. Speed of transmission depends upon the number of bits sent, the time required for the encoder to

generate the coded message, and the time required for the decoder to recover the original ensemble.
In WSN case speed it is not the target, is the degree of compression is the primary concern, it is nonetheless necessary that the algorithm be efficient in order for the scheme to be practical.

Inside the data compression exist two branch, Lossless and Lossy data.

- Lossless compression algorithms usually exploit statistical redundancy in such a way as to represent the sender's data more concisely without error. Lossless compression is possible because most real-world data has statistical redundancy. His target is erase the redundancy but not any information.
- Lossy data compression will be guided by research on how people perceive the data in question. Lossy data compression provides a way to obtain the best fidelity for a given amount of compression. His target erases some information that is less important for the receiver.

There are equally good data compression techniques them are mention below:

LossLess		Lossy
Technology	Algorithm	Technology
Dictionary coders	LZ77 LZW Statistical Lempel Ziv	Discrete cosine transform
Dynamic Markov Compression		A-law Mu-law
Entropy encoding	Adaptive Huffman coding Shannon-Fano coding Elias gamma coding Fibonacci coding	Linear predictive coding
Burrows–Wheeler transform		Modulo-N code for correlated data

Table 12. Reference for data compression. [REF. 27]

Even though this compression schemes are still under development, experimental results indicate that their compression rate and power reduction manners are quite impressive

Data Fusion

As it is introduced before several nodes in the network can collect data from neighboring nodes, aggregate the data into one packet and then transmit that packet to the management station. Management data can be compressed before transmission of less data and conserving both energy and bandwidth.

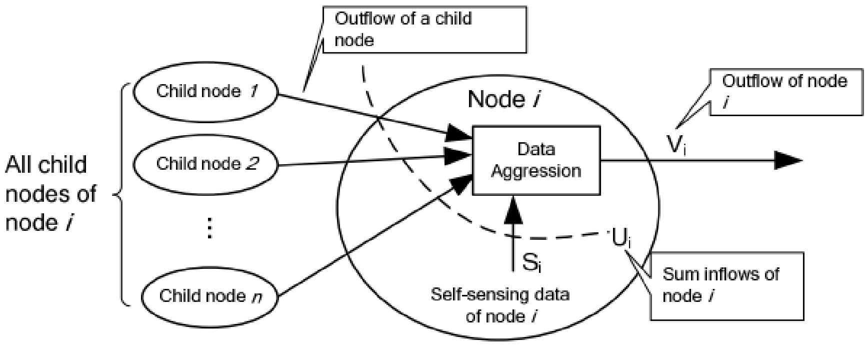
Up to here nothing new, but other techniques can be applied to achieve that compression. Data fusion has been used to eliminate redundancy in neighboring nodes. If multiple sources send the same data, the intermediate node will only forward one of them.
Fusion manner, the fusion point aggregates its own data with one input first, and next fuses the aggregation result with another input. This process will be repeated until all the inputs are aggregated.
Heinzelman et al. [REF. 21] proposed SPIN to realize traffic reduction for information dissemination using metadata negotiations between sensors to avoid redundant and or unnecessary data propagation through the network. But SPIN protocol it is not a Energy Efficient protocol as is a target of this document, then it is rejected.

The greedy aggregation approach can improve path sharing and attain significant energy savings when the network has higher node densities compared with the opportunistic approach.

There is other data fusion algorithm that is based on fuzzy logic methods to reduce traffic and enhance the performance of the sensor networks. Fuzzy logic methods are capable of fusing uncertain data from multiple sensor nodes to improve the quality of information. They require less computational power than conventional mathematical computational methods such as addition, subtraction, multiplication and division. In addition, only few data samples are required in order to extract accurate result at the end.[REF. 23]
The most common fuzzy logic inferences are the Mamdani and Tsukamoto-Sugeno methods. Both fuzzy logic Mamdani and Tsukamoto inference methods used by the proposed data fusion algorithm are completed in four phases: fuzzification, rule evaluation, combination or aggregation of rules, and deffuzification [REF. 24]
In [REF. 20] shows a general aggregation model where data aggregation may potentially occur at any point along a route. In particular, aggregated data may be fused again. Mathematically, the model only requires that the output data amount of the fusion function is not less than any of its inputs and not more than the summation of all inputs. This algorithm is called Minimum Fusion Steiner Tree (MFST), and achieve approximation ratio of ${}^{5}/_{4}\log(k+1)$, where k denotes the number of source nodes.
This is an example of a kind of protocols which based his performance on the data fusion enhancement.

5.1. Analysis of Aggregation parameters for WSN performance.

Following the last sentence, it is possible to find WSN routing algorithms which uses his energy remaining or number hops, others the signal strength. But we can add the data shrink ratios in the metrics to decide the network links or topologies.



Picture 15. Node Data Shrink model.

Whichever is the chosen data shrink these affects the performance of the node. The main parameters are CPU and Memory of Node features and Time Delaying through node.

Some of the proposed methods by some authors will be discussed below. With this analysis we can get a model of data shrink behavior to add in a simulator.

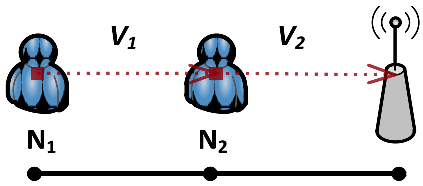
This model let compare how it has an impact on the network performance. Adding this to others parameters, as node joining and balance of sending data, it is possible to prevent the Energy uses of a node.

All this parameters could be correlated or not. If several parameters are correlated can be represented on the same metric. If not is needed a metric which shows that value and allows to neighbors nodes to evaluate it and change the performance of network.

When applying certain aggregation policies in WSN, the aggregation time for each node should be taken into consideration. It is a novel but significant problem to be studied for aggregation policies how to properly allocate the aggregation time among nodes to maximize the aggregation efficacy subject to the transfer delay bound.

[REF. 28] proposes a distributed and negotiated Aggregation Time Control (ATC) algorithm, which involves a dynamic balance process. During the runtime phase, every node independently increase the aggregation delay according to some rules, and BS is in charge of checking out the overtime packets. Once having detected such packets, sink starts broadcasting the overtime message, which makes specific nodes decrease their aggregation time. Although this method can effectively restrain the packets overtime, its aggregation effectiveness is ordinary and it is hard to control the network to maintain the stable state.

For example, considering a simple network, we now calculate the total network traffic load. Picture 16 shows a network with only two sensor nodes (Node1 and Node2) and one BS. Node 1 is a leaf node and its inflow equates with its self-sampling data flow S; Node 2 is the parent node and should forward data of node 1 to BS. If the two nodes are allocated with T_1 and T_2 for aggregation respectively, the network overall data outflow $f(T_1, T_2)$ can be computed as follows:



Picture 16. Aggregation delay.

$$f(T_1, T_2) = V_1 + V_2 = \frac{S}{1 + S \cdot T_1} + \frac{S + V_1}{1 + (S + V_1) \cdot T_2} =$$

$$= \frac{S}{1 + S \cdot T_1} + \frac{S + S/(1 + S \cdot T_1)}{1 + (S + S/(1 + S \cdot T_1)) \cdot T_2}$$

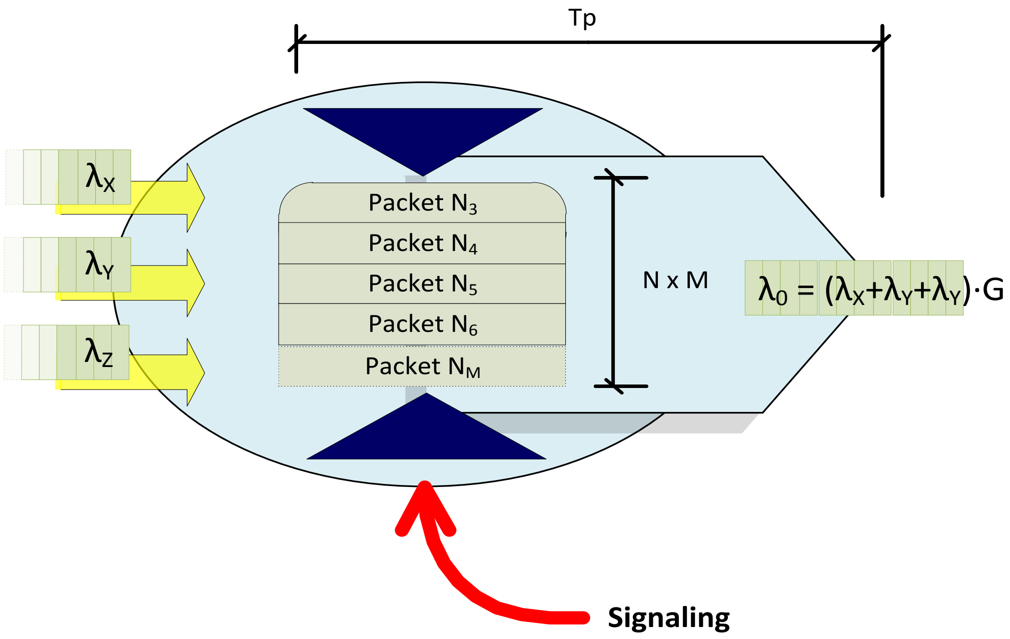
According to the above analysis, data traffic can be effectively reduced with the increase of node aggregation time, bringing about a performance enhancement in network transmission efficiency. But this increase can also lead to the deterioration of network delay.

5.1.1. Data Aggregation proposal

To show the effect of aggregation in the WSN performance I decided to use a model of data aggregation based on the theory of data compression. In this way is needed to define the main parameters of aggregation behavior in a device, across this definition is possible to get a model and set a simulation, then knows the aggregation role in a WSN.

Defining the parameters for simulations I emphasize:

- Aggregation Gain (G) that means the capacity to reduce the output data rate compared with the input data rate.
- Delay Time (Tp) due to compression process. This time include time of CPU but also, and more important, time to get the enough data to finish the compression.



Picture 17. Sketch showing parameters of proposed model.

Due to the correlation among the data input, G (aggregation gain) can change. This parameter is randomized, that means depends on the data input in each node. For example if the nodes sensor in a WSN are in a temperature environment, the data from them could be much correlated. That G parameter can be tends to 0. On the other hand, in case of a very heterogonous environment, data sensed can be much uncorrelated. So gain parameter can tend to 1.

In the next point I will check the capability of a typical WSN device to process the data shrink, also, at the same time, this checking, give a reference of the behavior of compression in a multi-hoping topology. Another of the mainly point to know it is estimate an average time that a WSN device CPU can use to compress data.

5.1.2. Computational capacity

I will start with a comparison among different real platforms of WSN devices to get a base reference of which can of performance can be applied to compression issue.

Name	Intel imote	Atmel ATmega	Jennic JN5148	Sensinode
CPU features	ARM 12Mhz 16b	AVR 16Mhz 8b	RISC 32Mhz 32b	MSP 18Mhz 16b
RAM (Kb)	64	16	96	16
ROM (Kb)	512	128	192	192
Tx (dBm)	0	3,5	2,5	4,5
Sensi (dBm)	-80	-100	-96	-97
MAC protocol	Bluetooth	ZigBee	6LoWPAN	6LoWPAN
Data Rate (kbps)	250	2000	250	250

Table 13. General features scope of WSN bundeled devices

As it is possible to see in Table 13 devices of WSN have similar features. This is due to all the devices shown above are a bundle of chips from different companies. Then the roots of WSN networks are the same, in this way it is possible to use above table as a reference if our aggregation proposal can be managed by real WSN devices

I have chosen Intel Imote series device to analyze, this choice is made as so use a devices of our department laboratory. As it is possible to see that device has enough features of memory space, and CPU computational capacity, but, low clock speed and 8 bits management.

5.1.3. Benchmarking of proposal compression

5.1.3.1. Gzip overview

Gzip is any of several software applications used for file compression and decompression. The term usually refers to the GNU Project's implementation, "Gzip" standing for GNU zip.

A good example of LZ77 and Huffman technique compression is Gzip. It is a lossless algorithm and freeware. License part is quite important to apply a technology in ours systems whose should have a commercial target.
GZIP is based on the DEFLATE [RFC 1951] algorithm, which is a combination of LZ77 and Huffman coding. DEFLATE was intended as a replacement for LZW and other patent-encumbered data compression algorithms, which, at the time, limited the usability of compress and other popular files.

Empirically, the deflate method is capable of compression factors exceeding 1000:1, The limit comes from the fact that one length/distance pair can represent at most 258 output bytes. A length requires at least one bit and a distance requires at least one bit, so two bits in can give 258 bytes out, or eight bits in give 1032 bytes out. A dynamic block has no length restriction, so you could get arbitrarily close to the limit of 1032:1.
To know better Gzip behavior, a benchmarking will be done with the typical data it is possible to find in a WSN. That benchmark will be the reference of Gzip behavior in this work.

The memory requirements for compression depend on two parameters, *windowBits* and *memLevel.*

$$\text{Deflate memery usage (bytes} = \big(1 \ll (\text{WindowsBits} + 2)\big) + (1 \ll (\text{memLevel} + 9))$$

Both windowBits and memLevel can be set to lower values at compile time via the MAX_WBITS and MAX_MEM_LEVEL macros, but only at a cost in compression efficiency.

The memory requirements for decompression depend only on windowBits, this case it is not too much important due to this process is done in BS.

As is shown the compression algorithm can be adapted to the wireless sensor resources. This is not a constraint to add on it.

5.1.3.2. Gzip benchmarking

Gzip performance:

- a 10-byte header, containing a magic number, a version number and a time stamp
- optional extra headers, such as the original file name,
- a body, containing a DEFLATE-compressed payload
- an 8-byte footer, containing a CRC-32 checksum and the length of the original uncompressed data

Just in case of a real application of this algorithm to a real WSN, as it is a GNU license code, we can modify the source and makes it lighter. As it is being shown above this algorithm adds some headers and footers to the data compressed, for this reason should be important to evaluate how it behaves with already gathered data. Could be, that in already processed data, the algorithm would not compress data anymore but adds so much headers and footers that the compressing gain becomes higher than 1.

Known is the entropy of Gzip from English text which is 2,71 bits/character. If a ASCII text every character is 8 bits the gain reach: G = 0,33.

In [REF. 33][REF. 34] the authors make a benchmark using different kind of data sources. After I will compare the gain from theorist entropy.

Kind of data	Rate of compressing	GAIN	Data mapping
Random data	17.05%	0,83	High uncorrelated data
ASCII text sources	69%	0,31	Environment sensors
Binaries data	59%	0,41	Machine sensors
ASCII and Image	52%	0,48	Streaming sensors

Table 14. Benchmarking of different kind of data sources

As we can see in this benchmarking the text sources gain is pretty similar than theorist gain.

Another test I would like to do is what happens when a compressed data is recompressed.

As it is introduced before GZIP adds some headers and footers, if a compressed data has a very uncorrelated content, maybe those headers and footers, make the compressing gain worse.

For this test I will compress the same file many times.

Compressing loops	Origin size (Kbytes)	Compressed size (Kbytes)	GAIN
1	4.096	2.009	0,49
2	2.009	2.040	1,015
3	2.040	2.071	1,015
4	2.071	2.102	1,014
5	2.102	2.130	1,013
6	2.130	2.161	1,0145
7	2.161	2.192	1,014

Table 15. Value of gain when the same file is compressed many times

Table 15 is very representative of the Gzip behavior; in case that a data has been compressed never again will be done.

5.1.4. Effects of compression in network parameters = Energy + Delay

Compression Gain is a linear function with data throughput. This means that more Gain in a node due to the incoming data, will be a reduction of data output.
Gain parameter is a function of data correlation that arrives to a node, also is an output function then directly affects to the energy uses of a node. This is the reason why some authors use aggregation gain as metric for their routing protocol design.
Regarding CASE 1 now it can take first conclusions, always for this kind of aggregation:

- In a multi-hoping environment in each step there is not a data improvement due to compression.
- A gain compression of 0,5 in front another node which just reach a gain of 1, the first node can manage double data rate or can relay the traffic of two nodes.

Final conclusion of this point is, evaluating the importance of aggregation in a WSN, it has a proportional influence over energy node uses but this value is not correlated with any other inside the node, then must be send to network as a different metric.

Delay point is a very important feature for all data networks. In this benchmarking have been impossible to evaluate but is taken into account, for that reason next point introduce a solution to try to modify that delay.

5.1.5. Design rules: Aggregation signaling

After have seen an approximation of how compression performance works, let take some consideration for routing protocol design.

- Already compressed FLAG.: this FLAG should be sent inside the data relaying to advice next node that incoming data is compressed an then do not make it again. This feature helps WSN routing protocol to relay faster due to is not needed to use CPU time for data compressing.

- Packet ALERT change size of compression MATRIX: since a packet come into the node until it leaves, there is a randomized gap time, impossible to know in advance, I propose to use an ALERT signaling packet, from the BS, if it detects a huge delay in data receiving. This packet will be relayed to the node and this will use the number inside this packet to modify the length of his matrix compression. To reduce delay: shrinking it.

6. Proposed Routing Protocol design

Arrived at this point, has been seen some suggested questions coming from the reading done. One of *CASES, CASE 3,* has been deeply examined, it suggests very important points of view, because of it, I discarded to follow working with clustering topologies.

A cluster can be seen like a multihop, so, there is a link short link among node and a clusterhead and another longer (most of them) from CH to BS. Then a clustering even could be included in a multihop topology, which are focused in energy efficiency.

Looks that multihoping topologies has a better performance in most of the environment. Clustering protocols have some constrains to adapt his behavior to any kind of environment. Most of them need to have a direct sight to the BS, then, in case of an environment that some of the nodes do not reach the BS the protocol fails.

Type of devices used to set up the network, as is mentioned before they are non shift power transmission. In that case only one option it is a multihoping performance to manage the network, because reachable BS nodes will be linked directly and non-reachable will try to find a node which relays his data.

For all of this my choice is continue working with multihoping protocols and focus my effort to develop a new proposal.

The main idea would be energy efficiency, as is introduced in clustering lectures, using energy aware from flat lectures.

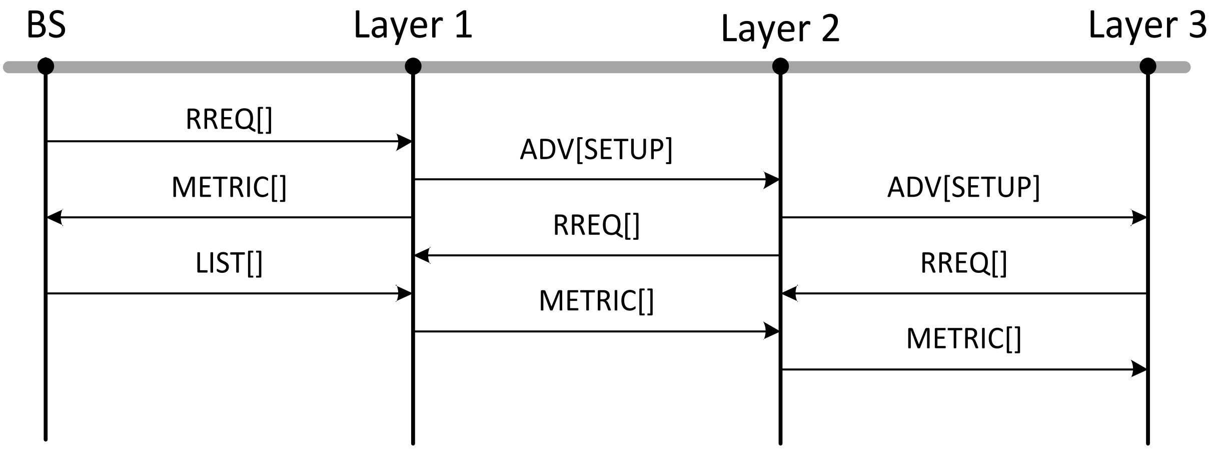
The last point for routing development is related with another of the *CASES* introduced: aggregation. This feature will be joined later in this new proposal.

In next points a routing protocol is described, simulated his behavior and finally compared with an existing one.

6.1. Flat Multihping routing protocols draft

There are some common features in flat routing protocols. First of all it is necessary communication among them to know which is the location and situation of every node. This information will be treated by the routing protocol algorithm to take his decisions.

Here below a sketch of how is the communication among nodes. Signaling named as well.



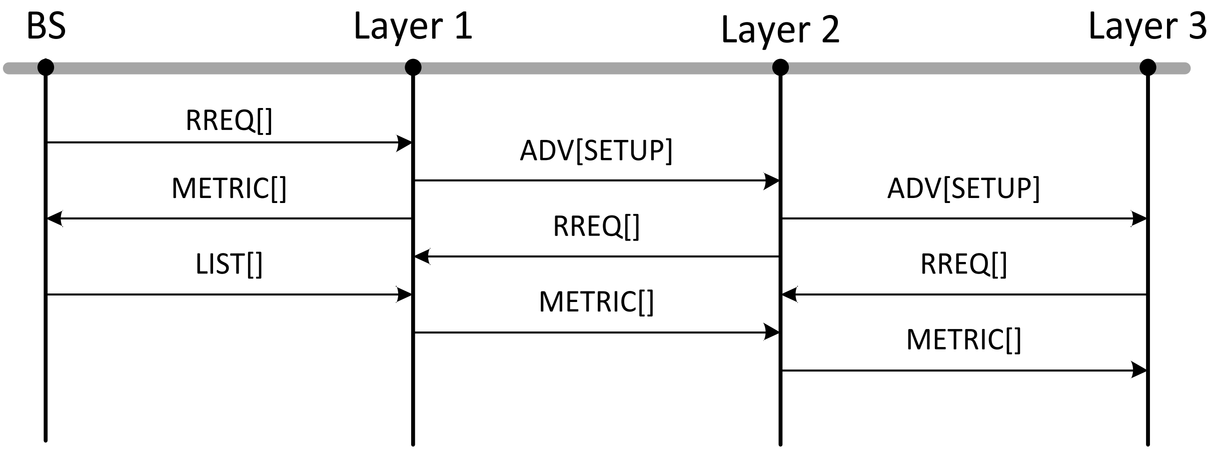
Picture 18. Scketch message of nodes

In my proposal every node will send message to discover which nodes are around him. This message can be:

- ADV: message of signaling used to send information though the network
- METRIC: main message used by algorithm to know the features of nodes to choose the next hop. This is a response to a request of a node.
- RREQ: message requesting information of metrics in set up process.

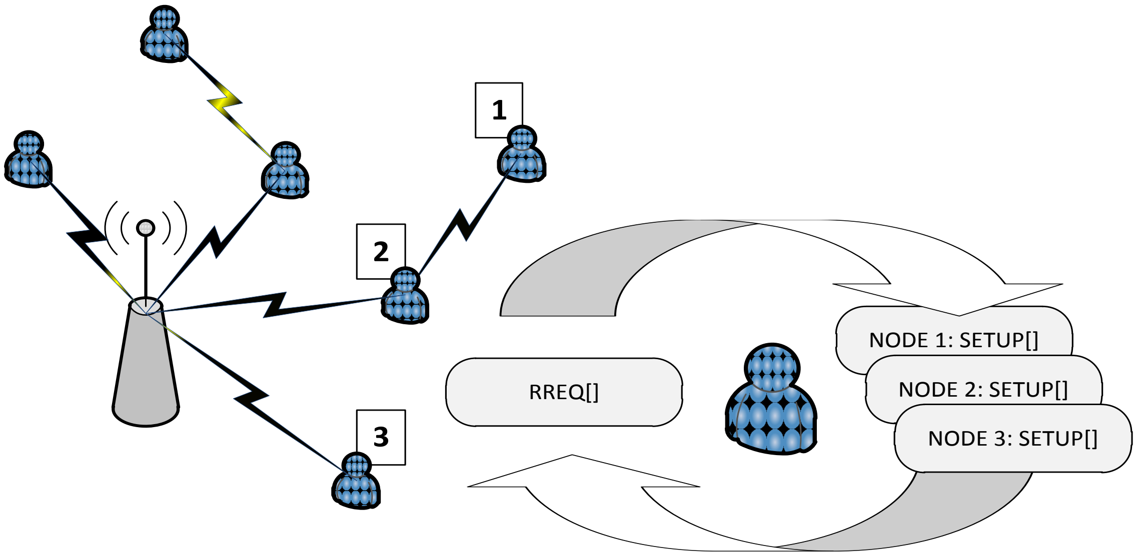
There is a special node which is BS, this node has no problems to manage energy, so this node can be load with a huge work because always would be assumed that it has enough resources.
In this way, routing protocol, as possible as routing it can, will load BS with the hardest works. That is the reason why whole layer one will be built by BS.
Thanks to this all nodes in this layer will be discharged of this work, which means save energy. Take into account that in flat routing protocols, nodes close to BS always are the most loaded.

What happen when a new network switch on, the BS will start a broadcasting of RREQ messages. Knowing every node that it is initializing his system the metric reply with be used to choose next hop like is shown in



Picture 18.

Another issue is where a new node is joined to the network.



Picture 19. Signaling of a new node joining to the network

As we can see in Picture 19 a new signaling is done when a new node joins to the network. This feature gives to protocol a high flexibility for that kind of networks with mobile nodes. My proposal network now is not oriented to mobile network but it is obvious that is a sooner future feature.

After a nodes has sent a request asking by his neighbor all nodes which can listen that request will send a setup packet. This includes his metric and hop-count from the BS. As is argued in CASE 2, the optimum topology is which has his nodes linked to the BS direction. Hop-count will help on that, on the other hand the nodes will manage the metric (described after) to find the less loaded node.
Every node use the RSS (received signal strength) to know how far or which is the cost to reach a neighbor, this value given by Rx electronics is added to metric to finish the table of links cost. Where nodes write this information is a table. On this

table the reachable nodes are sorted by his cost metric and the first one (less metric) is chosen as a relayer.
This table of cost only will content the direct neighbor or reachable nodes' *ID by direct link*. This is a difference compared with actual *Shortest Path First* [RFC 2328] which use a long table with all link cost.

After a briefing of how the structure of the protocol signaling is and the setup process, next step will be the core of this routing protocol, it is called metric.

6.2. Metric treatment and Frame definition

The basic idea in multihoping protocols is communication hop by hop of their state in order to realize their health. This idea is not new, comes from the wired networks.
All energy aware protocols like EAR [REF. 37] or HESS [REF. 38] use different parameters to calculate the metric. This is the main point that gives to the network a different behavior.
My metric proposal has these parameters:

$$METRIC = (\%Remain\ E)\gamma_R - (\%Joins)\gamma_J - (\%Work)\gamma_W - (RSS)\gamma_{RSS}$$

Equation 7. Prososed metric

- Remain Energy: it is the % of remaining energy in the power supply. This value is the ratio from the full power supply and actual state. This value always will be calculated taking into account any kind of harvesting system in a node. If this harvesting system always maintains the power supply full this node always can relay all data, then this node can be overloaded.
- Joins: number of nodes joined in that moment. To calculate the % of this parameter it is necessary to have some signaling, that is the reason of *SETUP* message design. Otherwise same threshold will be set up, by network manager, in all nodes as full joins. This never means that over this threshold a node doesn't except more joins.
- Work Load: % of the relay queue load.
- RSS: value of the signal stretch of the frame received by neighbor node in his request. This value would be from 0 to 100. 100 is the sensibility of the devices.

In Equation 7 there are other values, before mentioned, that are the weight of each parameter inside the metric. This weight shows how influence has each parameter in the metric value. In my proposal these weights are fixed. In future works can be customized depends the network behavior.

Now let's show a draft of how the proposed packets which carries information are built.

0 3 13 21 28 bits

0	ID Type	ID node source	Metric Value	Aggregation ratio

Picture 20. Metric frame structure

0 3 13 bits

0	ID Type	ID node source

Picture 21. RREQ frame structure

0 4 14 24 40 bits

1	ID Type	ID node destination	ID node source	DATA*

Picture 22. ADV frame structure

Trying to reduce the size of packet to the minimum, the *ID node* can be reduced if the WSN owner decides that his network will never be over a number of nodes. In this work case I have been working with a network of 900 nodes then the minimum of bits are 10.
The flag *ID type* will identify the kind of frame and helps to the node to know how the information has been assembled. This means a closed group of packet, four in case of 0 first bit, and 8 of *ADV*.
Neither *CRC* nor ending bits are included.

The most important frame is Picture 20. Metrics is the core of the well running of network because is the parameter used by the routing protocol algorithm to decide with is the best performance according with his design.

Inside *ADV* frame there is a section for *DATA*. Here in will be attached the signaling information that anyone wants to publish to network.
One of this is *SETUP* date frame. This frame is used by neighbor nodes which received a *RREQ* from a new node or a new network setup up. His structure is:

[DATA]

24 32 40 bits

Hop count	METRIC

Picture 23. ADV-Setup frame

Last important point of a routing protocol is if this protocol is reactive or proactive.
I proposed a proactive behavior that means a timeout that switch renewing of network topology and the upgrade of the metric tables.
Again this timeout is set at a concrete value in this proposal, but it can be a great work study the mobility of nodes or environment changes to modify this timeout value.

6.3. Aggregation: compression rules

I will add aggregation features to the routing protocol. Following the introduction at 5.1.1 my proposal is a compression function. To achieve this, a node has to fill a file, in this case a matrix of bits, and when this is full apply a compression algorithm. The result of this will be passed to the relay section waiting to be sent.
As is shown in 5.1.1 the algorithm (LZ77 + Hoffman), is taken as a reference, when a data is compressed anymore should be compressed again because that would make worse the compression results.
In this way some signaling will be added, if compression features are introduced before the data frame there will be a flag which tells if data is already compressed.

Due to this compression strategy needs to fill a matrix depend on matrix size or the ratio of filling, plus the CPU time processing of algorithm, time of data sending will suffer a delay.
To save some computing resources, would be the BS who manage the statistics of time delay and could be added by BS some signaling into the network which informs nodes change the size of matrix in order to fall down delay time of data sending.

7. Simulations

In order to simulate the routing protocols algorithms performance, the routing protocol was implemented in MATLAB. The choice of the MATLAB simulator is made based in the author's previous experience. As is mentioned before, other simulators, like OMNET or NS-2, were also suitable for the simulation development but their learning curves were out of the work time plan.

The simulator get information from Excel files where is defined the features of every node. These features are:

- Energy Power Supply capacity
- Location inside the network
- Relation among Tx power and Rx power.

From the other Excel file are introduced the traffic pattern. The data sent by the every node will follow this pattern. This pattern simulates the data got from sensor devices.

For this work a simple approach was chosen in order to interconnect the different nodes deployment results. A grid deployment seemed to be the perfect scenario to simulate and compare routing protocols.

Then a squared environment will be the reference. 900 nodes are deployed in that grid. The distance among them represents the minimum power energy set up in a node device. On this way no energy wasted by covering overlapping.
The BS node is placed in the middle of the node deployment.

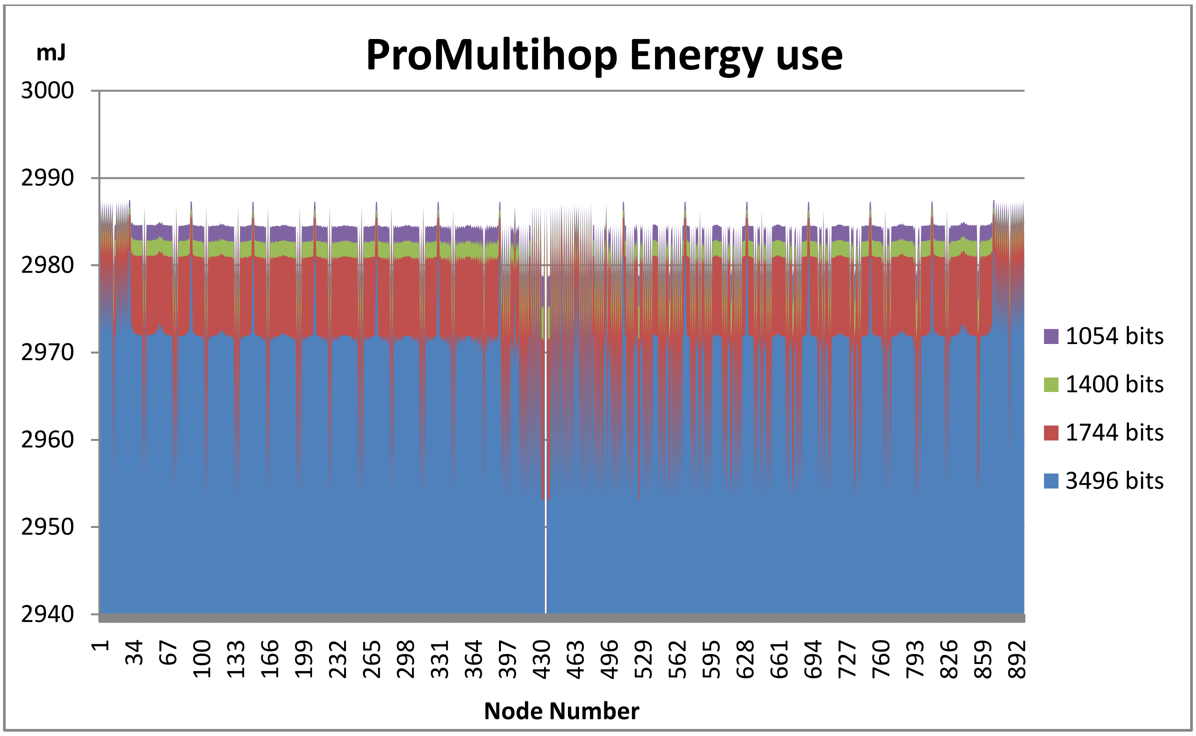
7.1. ProMultihoping behavior

Beginning energy		3000
Number nodes		900
Location shape		Grid
Path loss profile		Free Space Path Loss
η		1
Traffic pattern	Pattern	ON-OFF
	Duty	87,5%
Metric weight	γ_R	0,4
	γ_J	0,1
	γ_W	0,1
	γ_{RSS}	0,4

Table 16. Simulation parameters for ProMultihoping

The simulation will start with the setup process, on it, are counted in each node the energy used for setup frames. In section 6.2 you can see the length of the frame.

After that, starts a loop where every node is managed single. In this state the energy counted is the energy to send a single bit through the gap until the next hop.
Due to the size of network and the computing resources the simulation will be stopped as soon as possible there is a relevant result.
Here below the results of called “ProMultihoping” my proposed multihoping routing protocol.



Graph 2. Simulation results of ProMultihoping

The most important think of this result is not the concrete values of energy uses because, as is said before, these value are no real and do not match with a real devices. It Is just a reference to know the routing behavior over a network.

The shape of Graph 2 is the main issue to study. As we can see the shape is pretty flat that means the energy is really spread through all nodes. Although the nodes close to the BS looks overloaded, that is, because have to relay all the data of the network. But, it is better to relay a lot of data than send it from far away.

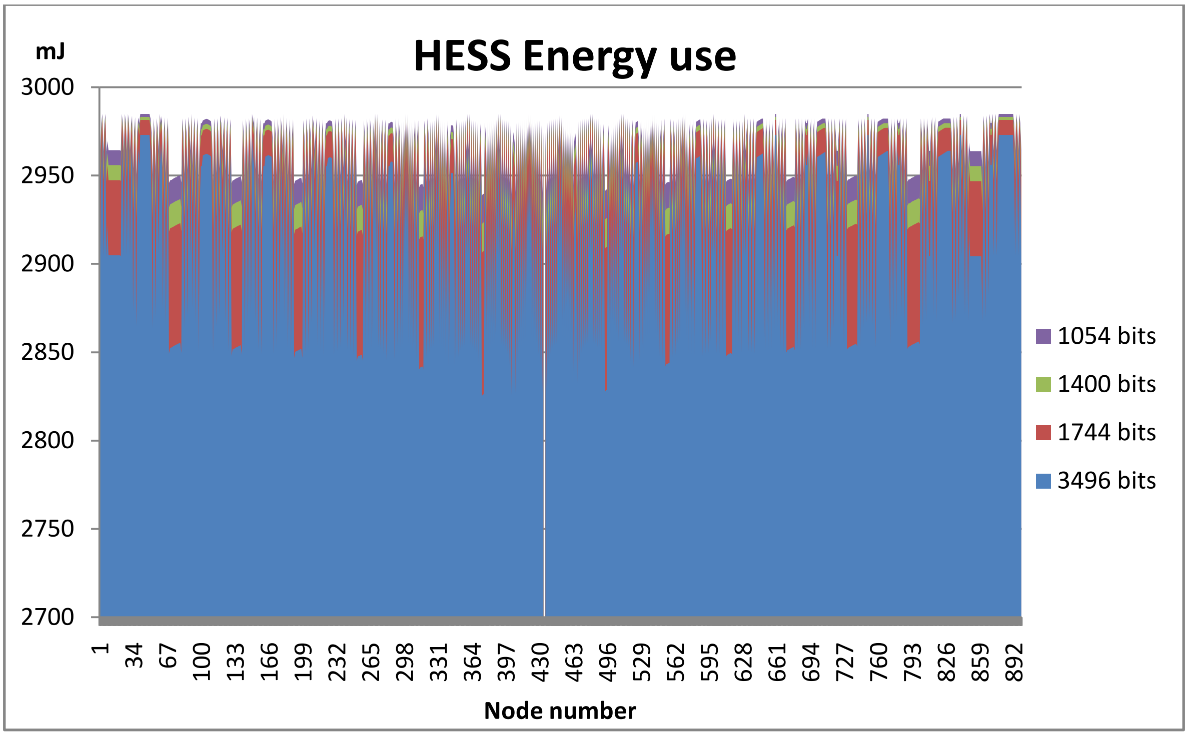
It is interesting to see over right part the shape change from node 463 to 661. That is due to decision taken for the routing protocol.

7.2. Compering with HESS

Beginning energy		3000
Number nodes		900
Location shape		Grid
Path loss profile		Free Space Path Loss
η		1
Traffic pattern	Pattern	ON-OFF
	Duty	87,5%
Metric weight	$\gamma_{SC} + \gamma_{RB}$	0,65
	$\gamma_{\gamma} + \gamma_{LC}$	0,35
	γ_{HC}	0,5
	γ_{OC}	0,5

Table 17. Simulation parameters for HESS

Metric weight of HESS protocol taken from his author work.[REF. 7]



Graph 3. Simulation results of HESS

Look up the shape of graph where HESS results are shown. A flat shape are found on it, but now it is possible to find overloaded nodes or deep point that means that node will be the first to arrive at his ending live. Closer to the BS this point are deeper. On the other hand the shape again is pretty flat as well as ProMultihop algorithm.

On both routing protocols the behavior is linear that means it is possible to predict at how many bits a node will reach bottom of his energy supply. Looking that at 1400 and 3496 shape, there is the double gap from the reference point. Then double bits sent double energy used.

Comparing energy uses, ProMultihoping reach better performance. The energy used is a little bit less (around 3,3%) also number of overloaded nodes are less too. This second feature, in my opinion is the most important in this comparison.

7.3. Aggregation: Compression data proposal

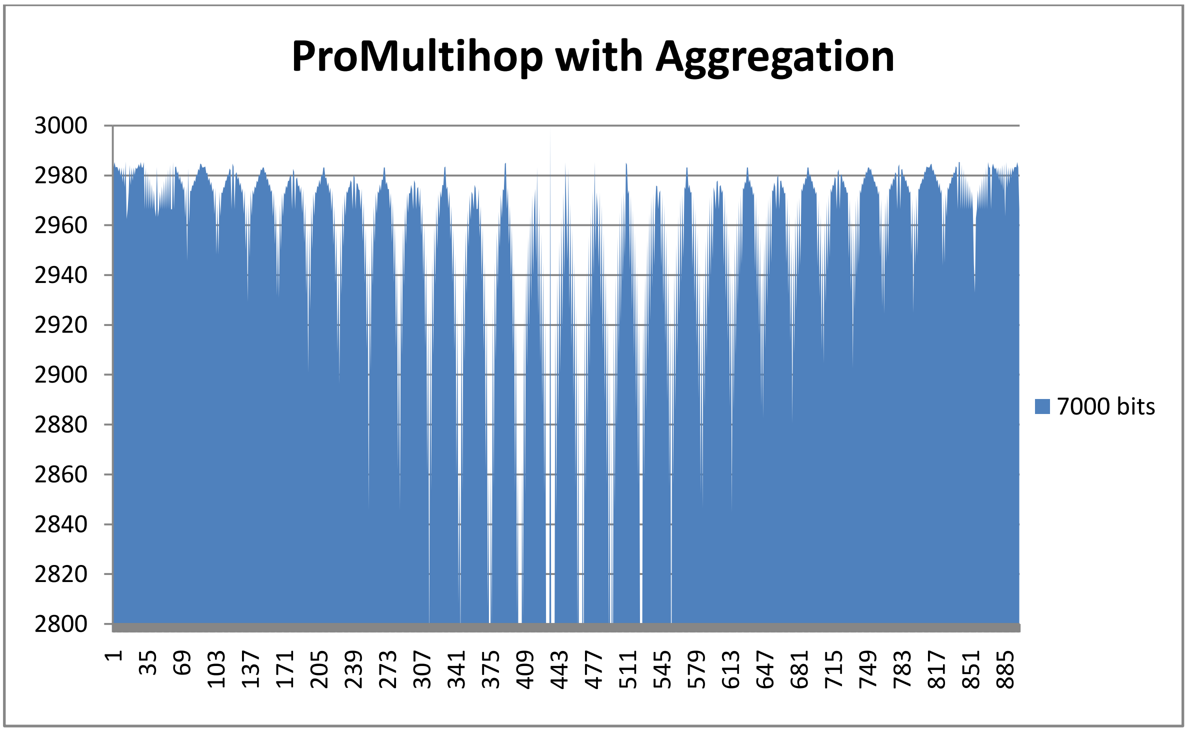
On this point I will apply proposed aggregation to ProMultihop routing protocol to know the behavior together. In a first point of view the energy managed has to be less.

How is done. First of all I have to introduce the table of traffic patterns shrink. Like is pointed before the compression algorithm is not implemented, instead of this I define a table of shrink among traffic pattern, this will be used by the simulator to know which is the shrink ratio of the data in each node.

Traffic patterns are distributed randomized then every node will relay data of nodes which may be have a high shrink ratio or not.

Traffic pattern A	Traffic patern B	Percentage
1	1	87
1	2	25
1	3	25
1	4	25
2	2	87
2	3	25
2	4	25
3	3	87
3	4	25
4	4	87
Average		40

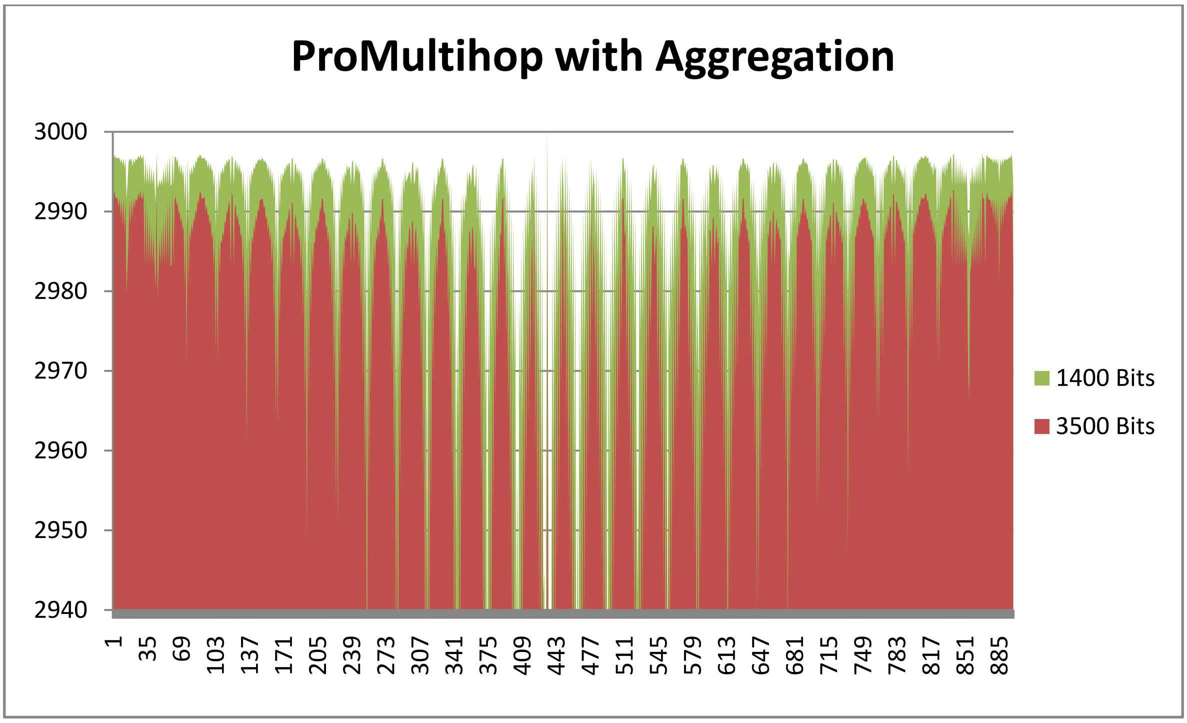
Table 18. Traffic pattern shrink ratios



Graph 4. ProMultihop behavior with Aggregation

This first simulation will be compared with the given in Graph 2. In this case double among of bits are sent, and the result for the furthest nodes is around 2970 mJ (average) and for this nodes in ProMultihoping without aggregation are almost same number. 2975.mJ.

Then, for those nodes, the double bits sent use almost the same energy. This is a great performance of compression feature. This value goes a bit beyond the average calculated in Table 18. Maybe because the random distribution of traffic patterns makes place some nodes with the same traffic pattern together.
But, there is a new shape in the results. Graph 4 shows a great overloading at nodes around BS.
This is due to the compressed data which are relayed to the BS can not be shrieked anymore, that means packet size are bigger than before (ProMultihoping without aggregation) so for the closer to the BS nodes that is an overload.



Picture 24. ProMultihop behavior with Aggregation.

At this second simulation is comparing the energy use when the same numbers of bits are sent, look up Graph 2.
In this case same conditions are done, again the same performance and shape. The inner nodes, close to BS, are overloaded. For the nodes furthest to the center of network the aggregation improvement are higher than the average calculated, but close to the BS this improvement becomes a fault.

From here appears a section in metric frame to place the aggregation ratio of a node. Due to the long time to finish more simulation, no more simulation could be done since time to write this report.

7.4. Aggregation: delay question

On the point 6.3 is described that the most important lack of compression is delaying on delivery.

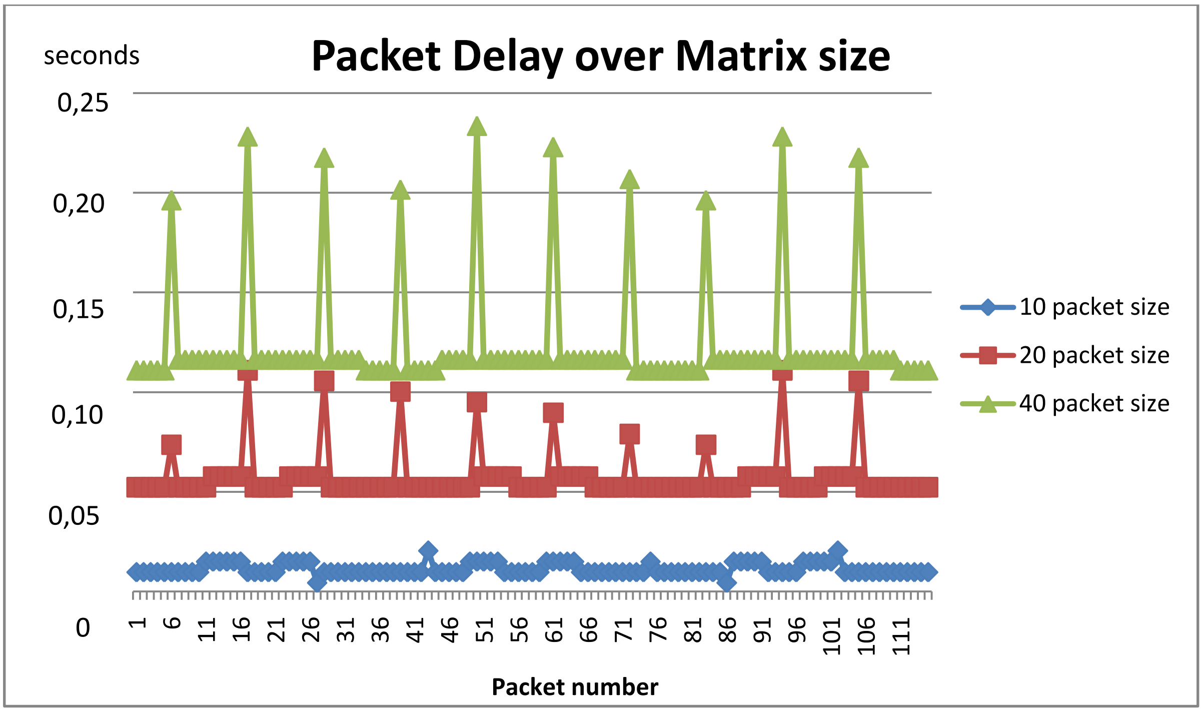


Table 19. Packet Delay due to compression

In this result time of CPU has not taken into account as is described very no compression algorithm are implemented.

This result shows the time delay due to the matrix filling. The most important topic is to realize that reducing the size of matrix the delay time is falling. That was the theoretical goal and, as it can see in the graph, is achieved.

8. Conclusions

Internet of Things would be the next step of the globalized interconnectivity. First were our computer, now our smartphones and laptops, tomorrow all.

In this work it is possible to get that the clustering protocols are close to be deprecated. Flat topologies give better results because the Path Loss is an exponential function then as far as is a node located more energy needs.
The sentence before will be stepped up with the future development of devices for wireless sensor networks, more over with spreading of Internet of things.

Based on the works done before, a new protocol is released. This protocol has some changes trying to keep the best of before done, and finally that target is achieved. Energy efficiency is emphasized following the energy aware treatment of predecessors. This new one can by including in flat routing protocol classification tree and as main features are a new metric treatment and proactive signaling.

Following the improvements and adding features to make routing protocols more efficient, and aggregation system is proposed (compression) and after check his behavior. As always in engineering there is a compromise related with this new feature. This is energy use, regards close to BS behavior, and time delay.
Compression helps network to download his data managing, saving some energy to extend his live. Also, as is shown at the end, changing the matrix size it is possible to change the time that a packet remains in a node, reducing the delay to the BS.

This work used simulations because many times is more useful and fast try to find the mathematical expression of a performance. Sometimes the mathematical expression it is almost impossible to find. At this point is needed a simulation.

During the work many unpredicted things appears, one of them it is close to the BS there is, for the author, called “Q area” inside the direct link could be the best choice. Although, it is better propose a future work for it.
Also about devices behavior, which ones can offer better performance if they are able to setup their power Tx and reduce electronics Rx power consumption.

9. Future work

This work is and step more in the research of Wireless Sensor Networks.
All test have been theoretical and based on many suppositions. Such in mathematic calculation than simulations all values are based on Path Losses. That gave a general scope and view of which are the continuous performance of power in a concrete scenario. But most of the actual devices use discrete power transmission, so will be a great improvement to adapt the algorithm to that.
Inside the ProMultihoping behavior metric weight are fixed and could be a nice work to simulate with different values of them to know which changes are.
Previous to fix metric weight some other values were tried but I focused in other results as goals of this work.

Concerning to above said, another improvement can be to allow working with real devices. In that way a function or discrete graph with the relation among power transmission and power supply consumption, can be passed to the algorithm.
With this it could be possible to develop systems based on real devices; even that simulator could become a product interesting for the companies. I am already working on it.

Above improvement can help to know which optimum power transmission settings are due to his electronics features.
In continuous research there is a topic from this work called “Q area”. That it is an interesting point referred along the work because is where the exponential behavior of path losses starts to grow strongly. Might be interesting to know if routing protocols could take into account this area and be more efficient, over all adding compression. In that point we saw an overloading of the BS area that is pretty much similar as Q area.

The proposed routing protocols use a proactive strategy where a timeout manage when the networks have to renew his settings or tables. I supposed that owner of the network will set this timeout value, but and study of random changes of environment as a stochastic process are done.
This study will help on the introduction of mobile nodes to this kind of networks.

REFERENCES

REF. 1. http://en.wikipedia.org/wiki/Internet_of_Things

REF. 2. http://www.economist.com/blogs/schumpeter/2010/12/internet_things

REF. 3 An Energy Efficient Routing Scheme for Mobile Wireless Sensor Networks. Lan Tien Nguyen , Xavier Defago , Razvan Beuran, Yoichi Shinoda, Japan Advanced Institute of Science and technology

REF. 4. MR-LEACH: Multi-hop Routing with Low Energy Adaptive Clustering Hierarchy. Muhamnmad Omer Farooq, Department of Computer Science, Virtual University of Pakistan. Abdul Basit Dogar Department of Computer Science Virtual University of Pakistan. Ghalib Asadullah Shah College of EME NUST, Pakistan.

REF. 5. PEGASIS: Power-Efficient Gathering in Sensor. Stephanie Lindsey. Cauligi S. Raghavendra. Computer Systems Research Department, The Aerospace Corporation. CA USA.

REF. 6. A Framework Architecture for Energy-Aware. Fachbereich Elektrotechnik und Informatik der Universitat Siegen.

REF. 7. Routing Protocol for Wireless Sensor Networks with Hybrid Energy Storage System. Nuno André Saraiva Pais. Institute of Electronic Systems . Aalborg University. Sep 2009.

REF. 8. EEMR: An Energy-Efficient Multi-hop Routing Protocol for Wireless Sensor Networks Jiguo Yu, enjun Liu, Jingjing Song, Baoxiang Cao School of Computer Science, Qufu Normal University.

REF. 9. A Clustering/Multi-hop Hybrid Routing Method for Wireless Sensor Networks with Heterogeneous Node Types Sampath Priyankara, Kazuhiko Kinoshita, Hideki Tode and Koso Murakami.

REF. 10. Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation by Vivek Mhatre, Catherine Rosenberg. School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285, USA.

REF. 11. The Impact of Data Aggregation in Wireless Sensor Networks by Bhaskar Krishnamachari Cornell niversity ECE,Deborah Estrin UCLA CS ,Stephen Wicker Cornell University ECE.

REF. 12. Unified Propagation Model for Wi-Fi, UMTS and WiMAX Planning in Mixed Scenarios. Frederico Varela, Pedro Sebastião, Américo Correia, Member, IEEE, Francisco Cercas, Member, IEEE ISCTE-LUI/ Instituto de Telecomunicações.

REF. 13. Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation. Vivek Mhatre, Catherine Rosenberg School of Electrical and Computer Engineering, Purdue University, West Lafayette,USA.

REF. 14. Rappaport T.S, Wireless Communications Principles and Practice, Prentice Hall, NJ, 1996.

REF. 15. An Application-Specific Protocol Architecture for Wireless Microsensor Networks Wendi B. Heinzelman, Member, IEEE, Anantha P. Chandrakasan, Senior Member, IEEE, and Hari Balakrishnan, Member IEEE.

REF. 16. http://en.wikipedia.org/wiki/Path_loss.

REF. 17. http://en.wikipedia.org/wiki/Friis_transmission_equation.

REF. 18. Data Fusion and Topology Control in Wireless Sensor Networks. Vrinda Gupta, Rajoo Pandey Department of Electronics and Communication Engg National Institute of Technology (Deemed University) Kurukshetra.

REF. 19. A double imputation method for Data Fusion. Alfonso Piscitelli. Dipartimento di Sociologia, Universit`a di Napoli Federico II

REF. 20. Routing Correlated Data with Fusion Cost in Wireless Sensor Networks Hong Luo, Jun Luo, Yonghe Liu, Sajal K. Das Center for Research in Wireless Mobility and Networking, Dept. of Computer Science and Engineering, University of Texas at Arlington.

REF. 21. Heinzelman W., A. Chandrakasan and H. Balakrishnan, Energy efficient communication protocol for wireless microsensor networks, in Proceedings on 33 rd Hawaii international conferences system sciences (HICSS), vol. 8.page 8020-8029, Jan 2000.

REF. 22. Demand Base Data Aggregation And Compression In Sensor Networks. M Nimesh Kumar Agarwal.

REF. 23. Data Fusion Algorithms in Cluster-based Wireless Sensor Networks Using Fuzzy Logic Theory. WEILIAN SU, Naval Postgraduate School, Department of Electrical Engineering. THEODOROS C. BOUGIOUKLIS, Naval Postgraduate School, Department of Electrical Engineering.

REF. 24. G.–J. Klir and B. Yuan, Fuzzy Sets and Fuzzy.

REF. 25. EATA: Effectiveness based Aggregation Time Allocation Algorithm for Wireless Sensor Networks. HUANGFU Wei 1,2 , LIU Yan 3+ , DUAN Bin 1,2 , SUN Limin 1,2 , MA Jian 4 , CHEN Canfeng 4, Beijing.

REF. 26 Adaptive Data Fusion for Energy Efficient Routing in Wireless Sensor Networks. Luo, Jun Luo, Yonghe Liu, Sajal K. Das Center for Research in Wireless Mobility and Networking, Dept. of Computer Science and Engineering, University of Texas at Arlington.

REF. 27 http://en.wikipedia.org/wiki/Data_compression

REF. 28. Jae Young Choi, Sunghyun Choi, Wook Hyun Kwon, and Hong Seong Park. Aggregation Time Control Algorithm for Time constrained Data Delivery in Wireless Sensor Networks. In: Proceedings of IEEE VTC 2006.

REF. 29. http://blog.mmassonnet.info/2010/09/benchmarking-compression-tools.html.

REF. 30. http://en.wikipedia.org/wiki/Gzip

REF. 31. http://www.gzip.org/zlib/zlib_tech.html

REF. 32. http://hackage.haskell.org/packages/archive/zlib/0.5.0.0/doc/html/Codec-Compression-GZip.html#t:MemoryLevel

REF. 33. http://stephane.lesimple.fr/wiki/blog/lzop_vs_compress_vs_gzip_vs_bzip2_vs_lzma_vs_lzma2-xz_benchmark_reloaded

REF. 34. http://tukaani.org/lzma/benchmarks.html

REF. 35. Cross-layer Energy Analysis of Multi-hop Wireless Sensor Networks Jussi Haapola, Zach Shelby, Carlos Pomalaza-Raez, University of Oulu, Finland. And Petri Mahonen Aachen University, Institute of Wireless Networks, Germany.

REF. 36. Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation Vivek Mhatre, Catherine Rosenberg School of Electrical and Computer Engineering, Purdue University, USA.

REF. 37 Huan L. and Yi P. Keong L., "An Efficient and Reliable Routing Protocol for Wireless Sensor Networks," Proceedings of the First International IEEE WoWMoM Workshop on Autonomic Communications and Computing (ACC'05), vol. 02, 2005.

REF. 38. Routing Protocol for Wireless Sensor Networks with Hybrid Energy Storage System. Nuno André Saraiva Pais Master Thesis September 2009. Aalborg Universitet. Denmark.

ANNEX A

Imote2

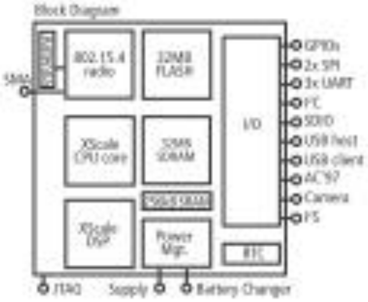
HIGH-PERFORMANCE WIRELESS SENSOR NETWORK NODE

- Marvell PXA271 XScale® Processor at 13 – 416MHz
- Marvell Wireless MMX DSP Coprocessor
- 256kB SRAM, 32MB FLASH, 32MB SDRAM
- Integrated 802.15.4 Radio
- Integrated 2.4GHz Antenna, Optional External SMA Connector
- Multi-color Status Indicator LED
- USB Client With On-board mini-B Connector and Host Adapters
- Rich Set of Standard I/O: 3xUART, 2xSPI, I²C, SDIO, GPIOs
- Application Specific I/O: I²S, AC97, Camera Chip Interface, JTAG
- Compact Size: 36mm x 48mm x 9mm

Applications

- Digital Image Processing
- Condition Based Maintenance
- Industrial Monitoring and Analysis
- Seismic and Vibration Monitoring

Block Diagram



Imote2

The Imote2 is an advanced wireless sensor node platform. It is built around the low-power PXA271 XScale CPU and also integrates an 802.15.4 compliant radio. The design is modular and stackable with interface connectors for expansion boards on both the top and bottom sides. The top connectors provide a standard set of I/O signals for basic expansion boards. The bottom connectors provide additional high-speed interfaces for application specific I/O. A battery board supplying system power can be connected to either side.

Processor

The Imote2 contains the Marvell PXA271 CPU. This processor can operate in a low voltage (0.85V), low frequency (13MHz) mode, hence enabling very low power operation. The frequency can be scaled from 13MHz to 416MHz with Dynamic Voltage Scaling. The processor has a number of different low power modes such as sleep and deep sleep. The PXA271 is a multi-chip module that includes three chips in a single package, the CPU with 256kB SRAM, 32MB SDRAM and 32MB of FLASH memory. It integrates many I/O options making it extremely flexible in supporting different sensors, A/Ds, radios, etc. These I/O features include I²C, 2 Synchronous Serial Ports (SPI) one of which is dedicated to the radio, 3 high speed UARTs, GPIOs, SDIO, USB client and host, AC97 and I²S audio codec interfaces, a fast infrared port, PWM, a Camera Interface and a high speed bus (Mobile Scaleable Link).

The processor also supports numerous timers as well as a real time clock. The PXA271 includes a wireless MMX coprocessor to accelerate multimedia operations. It adds 30 new media processor (DSP) instructions, support for alignment and video operations and compatibility with Intel MMX and SSE integer instructions. For more information on the PXA271, please refer to the Marvell datasheet.

Radio & Antenna

The Imote2 uses the CC2420 IEEE 802.15.4 radio transceiver from Texas Instruments. The CC2420 supports a 250kb/s data rate with 16 channels in the 2.4GHz band.

The Imote2 platform integrates a 2.4GHz surface mount antenna which provides a nominal range of about 30 meters. For longer range a SMA connector can be soldered directly to the board to connect to an external antenna.

Power Supply

The Imote2 can be powered by various means:

Primary Battery: This is typically accomplished by attaching a MEMSIC Imote2 Battery Board to either the basic or advanced connectors.

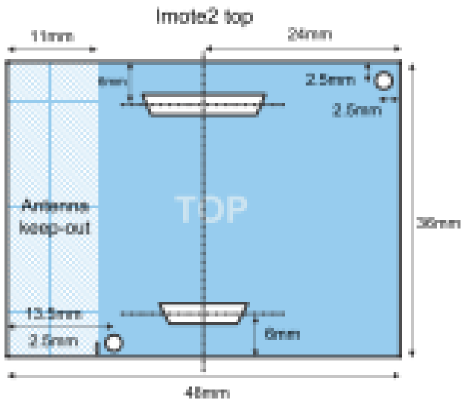
Phone: 408.964.9700 | Fax: 408.324.4841 | E-mail: infoca@memsic.com | www.memsic.com

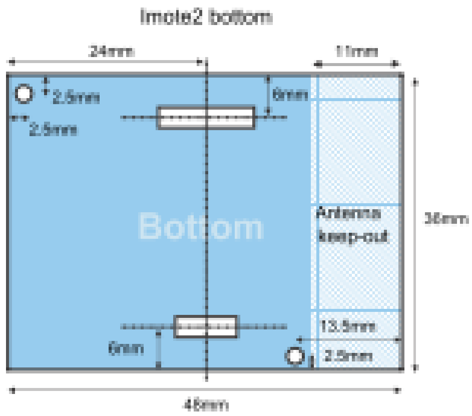
Rechargeable Battery: This requires a specially configured battery board attached to either the basic or advanced connectors. The Imote2 has a built-in charger for Li-Ion or Li-Poly batteries.

USB: The Imote2 can be powered via the on-board mini-B USB connector. This mode can also be used to charge an attached battery.

Battery Pads: A suitable primary battery or other power source can be connected via a dedicated set of solder pads on the Imote2 board.

Processor/Radio Board	IPR2400	Remarks
CPU		
Processor	Marvell PXA271	
SRAM Memory	256 kB	
SDRAM Memory	32MB	
FLASH Memory	32MB	
POWER CONSUMPTION		
Current Draw In Deep Sleep Mode	390 µA	
Current Draw In Active Mode	31 mA	13MHz, radio off
Current Draw In Active Mode	44 mA	13MHz, radio Tx/Rx
Current Draw In Active Mode	66 mA	104MHz, radio Tx/Rx
Radio		
Transceiver	TI CC2420	
Frequency Band (ISM)	2400.0 – 2483.5 MHz	
Data Rate	250 kb/s	
Tx Power	-24 – 0 dBm	
Rx Sensitivity	-94 dBm	
Range (line of sight)	~30 m	With integrated antenna
I/O		
USB Client (mini-B), USB Host		
UART 3x, GPIOs, I²C, SDIO, SPI 2x, I²S, AC97, Camera		
Power		
Battery Board	3x AAA	
USB Voltage	5.0 V	
Battery Voltage	3.2 – 4.5 V	
Li-Ion Battery Charger		
Mechanical		
Dimensions Imote2 Board	36mm x 48mm x 9mm	
Weight	12g	





Imote2 design licensed from Intel® Corporation.

ANNEX B

Features

- High Performance, Low Power AVR® 8-Bit Microcontroller
- Advanced RISC Architecture
  - 135 Powerful Instructions – Most Single Clock Cycle Execution
  - 32x8 General Purpose Working Registers
  - Fully Static Operation
  - Up to 16 MIPS Throughput at 16 MHz and 1.8V
  - On-Chip 2-cycle Multiplier
- Non-volatile Program and Data Memories
  - 128K Bytes of In-System Self-Programmable Flash
    - Endurance: 2000 Write/Erase Cycles @ 85°C
  - 4K Bytes EEPROM
    - Endurance: 2000 Write/Erase Cycles @ 85°C
  - 16K Bytes Internal SRAM
- JTAG (IEEE std. 1149.1 compliant) Interface
  - Boundary-scan Capabilities According to the JTAG Standard
  - Extensive On-chip Debug Support
  - Programming of Flash EEPROM, Fuses and Lock Bits through the JTAG interface
- Peripheral Features
  - Multiple Timer/Counter & PWM channels
  - Real Time Counter with Separate Oscillator
  - 10-bit, 330 ks/s A/D Converter; Analog Comparator; On-chip Temperature Sensor
  - Master/Slave SPI Serial Interface
  - Two Programmable Serial USART
  - Byte Oriented 2-wire Serial Interface
- Advanced Interrupt Handler
- Watchdog Timer with Separate On-Chip Oscillator
- Power-on Reset and Low Current Brown-Out Detector
- Advanced Power Save Modes
- Fully integrated Low Power Transceiver for 2.4 GHz ISM Band
  - Supported Data Rates: 250 kb/s and 500 kb/s, 1 Mb/s, 2 Mb/s
  - -100 dBm RX Sensitivity; TX Output Power up to 3.5 dBm
  - Hardware Assisted MAC (Auto-Acknowledge, Auto-Retry)
  - 32 Bit IEEE 802.15.4 Symbol Counter
  - Baseband Signal Processing
  - SFR-Detection, Spreading; De-Spreading; Framing ; CRC-16 Computation
  - Antenna Diversity and TX/RX control
  - TX/RX 128 Byte Frame Buffer
- Hardware Security (AES, True Random Generator)
- Integrated Crystal Oscillators (32.768 kHz & 16 MHz)
- I/O and Package
  - 38 Programmable I/O Lines
  - 64-pad QFN (RoHS/Fully Green)
- Temperature Range: -40°C to 85°C Industrial
- Supply voltage range 1.8V to 3.6V with integrated voltage regulators
- Ultra Low Power consumption (1.8 to 3.6V) for Rx/Tx & AVR: <18.6 mA
  - CPU Active Mode (16MHz): 4.1 mA
  - 2.4GHz Transceiver: RX_ON 12.5 mA / TX 14.5 mA (maximum TX output power)
  - Deep Sleep Mode: <250nA @ 25°C
- Speed Grade: 0 – 16 MHz @ 1.8 – 3.6V

Applications

- ZigBee® / IEEE 802.15.4-2006/2003™ – Full And Reduced Function Device (FFD/RFD)
- General Purpose 2.4GHz ISM Band Transceiver with Microcontroller
- RF4CE, SP100, WirelessHART™, ISM Applications and IPv6 / 6LoWPAN

8-bit AVR® Microcontroller with Low Power 2.4GHz Transceiver for ZigBee and IEEE 802.15.4

ATmega128RFA1

PRELIMINARY

8266AS-MCU Wireless-12/09